International Journal of Genomics

The Promise of Agriculture Genomics

Guest Editors: Wenqin Wang, Xuan H. Cao, Mihai Miclăuş, Jianhong Xu, and Wenwei Xiong

The Promise of Agriculture Genomics

International Journal of Genomics

The Promise of Agriculture Genomics

Guest Editors: Wenqin Wang, Xuan H. Cao, Mihai Miclaus, Jianhong Xu, and Wenwei Xiong

Editorial Board

Contents

Hindawi
International Journal of Genomics
Volume 2017, Article ID 9743749, 3 pages
http://dx.doi.org/10.1155/2017/9743749

Editorial

The Promise of Agriculture Genomics

Wenqin Wang,[1] Xuan H. Cao,[2] Mihai Miclăuș,[3] Jianhong Xu,[4] and Wenwei Xiong[5]

[1]*Shanghai Jiaotong University, Shanghai, China*
[2]*Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany*
[3]*National R&D Institute for Biological Sciences, Cluj-Napoca, Romania*
[4]*Institute of Crop Science, Zhejiang Key Laboratory of Crop Germplasm, Zhejiang University, Zhejiang 310058, China*
[5]*Montclair State University, Montclair, NJ, USA*

Correspondence should be addressed to Wenqin Wang; wang2015@sjtu.edu.cn

Received 31 January 2017; Accepted 1 February 2017; Published 5 March 2017

With the fact of growing global population, changing climate, and environmental pressure, there is an urgent need to accelerate breeding novel crops with higher production, drought or heat tolerance, and less pesticide usage. Advances in genomics offer the potential to speed up the process of developing crops with promising agronomic traits. Agriculture genomics is the application of genomics in agriculture to improve the productivity and sustainability in crop and livestock production. With the combination of traditional and high-throughput sequencing platforms, there has been a tremendous increase in genomic resources available, including expressed sequence tags (ESTs), BAC end sequence, genetic sequence polymorphisms, gene expression profiling, whole-genome (re)sequencing, and genome-wide association studies. Given the emergence of genomic sequencing and expansion of bioinformatic tools, we are shifting from single gene study to whole-genome analysis, which offers a broader view of how all genes work together. Therefore, this special issue was organized, comprising eleven original research papers and three reviews aimed to highlight recent advances in (1) comparative genomics for plant breeding, (2) transcriptome analysis for plant breeding, (3) genotyping and marker-assisted breeding, and (4) recombinant DNA technology (Figure 1).

Comparative Genomics for Plant Breeding. To date, the genome sequences for more than 55 plant species (mainly model plants, such as *Arabidopsis*, rice, and maize) have been produced. The 1000-plant (one KP or 1 KP) initiative is underway. As a result, scientists are not totally dependent on genomic sequences of model plants. Almost every species-specific genome can be sequenced with affordable price and thus offer great opportunities for targeted crop breeding. Furthermore, genomics is playing a very crucial role in biodiversity conservation. Advanced genomics helps in identifying the segments of the genome responsible for adaptation. It can also improve our understanding of microevolution through a better understanding of natural selection, mutation, and recombination, as summarized in S. Khan et al.'s article. Understanding the structure, organization, and dynamics of genomes in plant species can provide insights into how genes have been adapted by natural and artificial selection to respond to environmental constraints and the potential of their manipulation for crop improvement.

Conventional breeding in agriculture is based exclusively on phenotypic selection. Until the availability of genomic sequence for model plants, comparative genomics approaches were successful for identifying homologues/orthologues or cloning species-specific genes by using sequence conservation or synteny from model plant systems. The group of D. T. Le surveyed the genomes of *Arabidopsis* and soybean for genes encoding Met-rich proteins (MRPs) based on sequence similarity. Genes encoding MRPs were classified into functional categories including RNA transcription, protein modification, and calcium signaling. It was found that MRPs were mainly responsible for drought and salinity stress in *Arabidopsis* and soybean.

Predicting gene function solely based on homology to others can sometimes be difficult. Thus, proteomics (the large-scale analysis of proteins) will greatly contribute to our

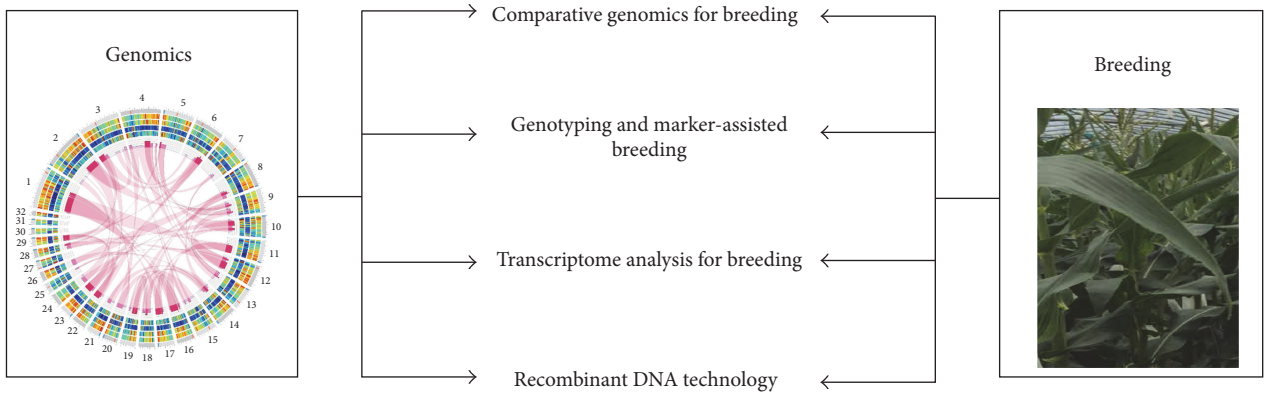


Figure 1: Advances in this special issue.

understanding of gene function in the postgenomic era. The group of J. Lee used shotgun proteomic analysis to quantify the protein in *Brassica rapa* under drought treatment. Their results showed that the levels of proteins associated with photosynthesis were decreased while the proteins involved in catabolic processes and stress responses were increased, rendering their genes as potential targets for engineering drought resistance in plants.

Transcriptome Analysis for Plant Breeding. With numerous genome sequences being deposited into the public databases at an accelerating pace, it is still challengeable to translate sequence into function directly. In this context, additional efforts are needed to understand the gene structure and quantify its expression. Y. Kong et al. took such an approach by molecularly cloning, characterizing, and quantifying the gene expression of hemocyanin subunit 1 (*MnHc-1*) in oriental river prawn. The full-length cDNA of *MnHc-1* was 2,163 bp with a 2,028 bp open reading frame (ORF) encoding a polypeptide of 675 amino acids. The *MnHc-1* gene was expressed in the hepatopancreas, gill, hemocytes, intestine, ovary, and stomach, with the highest level in the hepatopancreas. Their investigation indicated that the *MnHc-1* expression can be influenced by dietary copper and the hemocyanin may potentially participate in antibacterial defense. The method provides the foundation for further studies to understand gene function using loss-of-function mutant phenotype.

In the absence of the complete genome sequence, transcriptome analysis would improve our understanding of gene function. A global transcriptome study unveils the gene responses to a particular biological condition at the genome level. For example, aluminum and acid combination is the main factor limiting plant growth and crop production worldwide. P. Zhou et al. used microarray data in alfalfa seedlings to investigate how acid soils and aluminum toxicity impact the global gene expression. The main functional categories involved in phytohormone regulation, reactive oxygen species, and transporters were enriched after aluminum stress in alfalfa. Their results contribute towards understanding the key regulatory genes and pathways that would be advantageous for improving crop production not only in alfalfa but also in other crops under aluminum-acid stress. The transcript levels of MRP-coding genes under normal and stress conditions in *Arabidopsis* and soybean were also studied by microarray indicating that MRPs participate in various vital processes of plants under normal and stress conditions.

RNA-Seq, also called whole transcriptome shotgun sequencing, uses next-generation sequencing (NGS) to reveal the presence and quantity of RNA in a biological sample at a given time, which is a more sensitive and accurate way to investigate genome-wide gene expression than microarray. Using RNA-Seq in rainbow trout with high and low carcass fat content, the group of B. Wang identified 1,694 differentially expressed transcripts involved in lipid metabolism, such as L-FABP, adiponectin, PPAR-α, PPAR-β, and IGFBP1a. Their findings also indicated that PPAR-α and PPAR-β could be used as molecular markers for fat storage in rainbow trout liver.

Genotyping and Marker-Assisted Breeding. In addition to the generation of reference genome, high-throughput sequencing technology has facilitated resequencing of genomes of the same species but different accessions to identify genomic variation. The genotyping platforms have been used to generate large-scale marker segregation data on mapping populations and have led to comprehensive genetic maps. The genome sequence allows us to identify genome-wide molecular markers including functional markers, candidate genes, and predictive markers for breeding.

Duckweeds are promising plants to clean wastewater and to be digested into renewable biofuel. It was reported that different ecotypes of the same duckweed species exhibit a variety of biochemical and physiological properties. Developing sensitive markers to select desirable ecotypes is critical in plant breeding. The group of H. Zhao designed three molecular markers, PCR-amplified their products, and ran them through high-resolution capillary electrophoresis. Eleven haplotypes were found both in *Spirodela polyrhiza* and in *Landoltia punctata*. The marker system with multiple sequence polymorphisms is sensitive to intraspecies discrimination compared with interspecies identification and thus will promote large-scale identification at the ecotype level.

Cowpea is one of the most important legume crops in the world. It is also a major food crop in Africa, Latin America, and India because of its high protein content. Genetic diversity is the greatest resource for plant breeders to select lines that could potentially enhance food quality and quantity. The group of E. N. Wamalwa evaluated genetic diversity in 19 cowpea accessions from the Kenyan national gene bank, which they classified into two major groups. Breeders can now cross genetically distant accessions from those two groups for the improvement of cowpea crop, harnessing the power of heterosis.

Molecular markers such as simple sequence repeats (SSR) and single nucleotide polymorphism (SNP) from genomic and transcriptomic studies are great resources in plant breeding, used for trait dissection and for enhancing precision in selecting functional genes. Tomato is a thermophilic vegetable and is sensitive to low temperature. To map QTLs conferring cold tolerance in tomato, the group of H. Chen developed 120 SSR markers from a population of 146 RILs (Recombinant Inbred Lines) that was derived from a cross between a cold-sensitive cultivated *Solanum lycopersicum* and a cold-tolerant wild *Solanum pimpinellifolium*. The study resulted in nine QTLs providing references for further fine mapping of cold tolerance. Meanwhile, the polymorphic markers that have been developed can be used in selecting desirable traits and aid in developing new tomato varieties by marker-assisted breeding. The group of T. D. Khanh analyzed five typical rice landraces including three *indica* and two *japonica* by using 30-fold coverage of short paired-end reads from NGS. Compared with reference genomes, they determined more than two million SNPs and INDELs that would provide informational resources to help further marker-assisted selection in rice breeding programs.

Genome-wide association study is another method to find the relationship between molecular markers and QTL based on linkage disequilibrium. Association mapping on founder parents and their derivatives can find some important QTL and favorable allelic variations, which can be further used for marker-assisted selection to produce more favorable varieties. The group of H.-X. Ma screened Ningmai 9 wheat and 117 of its derivatives for their solvent retention capacity (SRC) and associated this phenotypic trait with 29 QTL markers. Their study was aimed at improving breeding soft wheat flour for cookie quality.

Recombinant DNA Technology. Recombinant DNA technology is a milestone in plant science and crop breeding that can help to design almost any desirable characteristic by controlled targeted gene expression. The team of H. Hou reviewed the history, the current research progress, and applications of recombinant DNA technology. They summarize that the genetically modified plants have been shown to possess improved resistance to harmful agents, enhanced product yield, and increased adaptability for abiotic stress.

One of the most powerful tools of recombinant DNA technology is CRISPR-Cas9-induced genome editing. H. X. Cao et al. overviewed the CRISPR-Cas9 system and the major technical advances for manipulation of model and crop plant genomes. They also discussed the future perspectives of CRISPR-Cas technology in molecular plant breeding. It is believed that the promise of next green revolution with new crops meeting long-standing requests is soon to be achieved greatly aided by the rapid development of CRISPR-Cas technology.

Genetically modified crops are plants that had their DNA genetically modified using recombinant DNA technology. For example, transgenic maize (WL-73) plants overexpressing the *BcWRKY1* gene were generated by *Agrobacterium*-mediated transformation, which were able to resist 300 mM NaCl stress. Still, the crop safety needs to be further evaluated. X. Zeng et al. measured the effects on rhizosphere soil in terms of enzyme activities, physicochemical properties, and microbial populations between the genetically modified maize, overexpressing the *BcWRKY1* gene, and nontransgenic maize. They reported that salinity-tolerant transgenic maize had no adverse impact on soil rhizosphere during a period of three consecutive years, which paves the way for further commercialization.

Acknowledgments

We thank all the authors who contributed their original research articles to this special issue. We also thank all the reviewers for their hard work and timely response.

Wenqin Wang
Xuan H. Cao
Mihai Miclăuș
Jianhong Xu
Wenwei Xiong

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2017, Article ID 9272363, 11 pages
http://dx.doi.org/10.1155/2017/9272363

Research Article

Whole-Genome Characteristics and Polymorphic Analysis of Vietnamese Rice Landraces as a Comprehensive Information Resource for Marker-Assisted Selection

Hien Trinh,[1] Khoa Truong Nguyen,[2] Lam Van Nguyen,[1] Huy Quang Pham,[1] Can Thu Huong,[3] Tran Dang Xuan,[3] La Hoang Anh,[3] Mario Caccamo,[4] Sarah Ayling,[4] Nguyen Thuy Diep,[2] Cuong Nguyen,[1] Khuat Huu Trung,[2] and Tran Dang Khanh[2]

[1]*Laboratory of Bioinformatics, Institute of Biotechnology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Cau Giay, Hanoi, Vietnam*
[2]*Department of Genetic Engineering, Agricultural Genetics Institute, Vietnam Academy of Agricultural Sciences, Km2 Pham Van Dong, Tuliem, Hanoi, Vietnam*
[3]*Graduate School for International Development and Cooperation, Hiroshima University, Hiroshima 739-8529, Japan*
[4]*Genetics and Breeding, National Institute of Agricultural Botany, Cambridge, UK*

Correspondence should be addressed to Cuong Nguyen; cuongnguyen@ibt.ac.vn and Tran Dang Khanh; khanhkonkuk@gmail.com

Received 24 July 2016; Revised 21 November 2016; Accepted 20 December 2016; Published 7 February 2017

Academic Editor: Hieu Xuan Cao

Next generation sequencing technologies have provided numerous opportunities for application in the study of whole plant genomes. In this study, we present the sequencing and bioinformatic analyses of five typical rice landraces including three *indica* and two *japonica* with potential blast resistance. A total of 688.4 million 100 bp paired-end reads have yielded approximately 30-fold coverage to compare with the Nipponbare reference genome. Among them, a small number of reads were mapped to both chromosomes and organellar genomes. Over two million and eight hundred thousand single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels) in *indica* and *japonica* lines have been determined, which potentially have significant impacts on multiple transcripts of genes. SNP deserts, contiguous SNP-low regions, were found on chromosomes 1, 4, and 5 of all genomes of rice examined. Based on the distribution of SNPs per 100 kilobase pairs, the phylogenetic relationships among the landraces have been constructed. This is the first step towards revealing several salient features of rice genomes in Vietnam and providing significant information resources to further marker-assisted selection (MAS) in rice breeding programs.

1. Introduction

Whole-genome sequencing (WGS) has revealed genetic information and genome structure and facilitated the identification of gene function of different plant species including some major crops such as rice, wheat, tomato, and soybean [1–5]. Next generation sequencing (NGS) methods are rapid and cost-effective, providing many promising applications for plant genomics and affording further insight into massive genomic variations [6, 7]. Combining NGS with bioinformatics is a powerful approach to detect DNA polymorphisms for quantitative trait loci (QTLs) analyses, marker-assisted selection, genome-wide association studies (GWAS), and linkage disequilibrium analysis in plants [8–11]. Moreover, DNA polymorphisms have been widely applied as DNA markers in genetic crop research [12].

Rice (*Oryza sativa* L.) is a staple crop and provides daily food for over half of the world's population. It contains two major groups, *indica* and *japonica*, which diverged more than one million years ago [13]. The recent sequencing of one representative from each group facilitated the study of the genomic structure of rice. In 2002, the first draft sequence of the *indica* genome was established by shotgun sequencing and homologous genes were predicted by comparison

Table 1: Abbreviation list of Vietnamese rice genomes used in the study.

Abbreviation	Name of rice landraces (name in Vietnamese)	Subspecies	Origin
indica 12	Chiem nho Bac Ninh 2	*indica*	Bac Ninh province
indica 13	Nep lun	*indica*	Ha Giang province
indica 15	OM6377	*indica*	Can Tho province
japonica 11	Bletelo	*japonica*	Lang Son province
japonica 14	Khau mac buoc	*japonica*	Nghe An province

with the *Arabidopsis thaliana* genome by Yu et al. [14]. Subsequently, the complete genome sequence of *japonica* was generated in 2005 [15] and then updated using NGS and optical mapping in 2013 [1]. This genome version has been widely utilized as the reference rice genome for DNA polymorphism findings reported in some previous studies [16, 17].

Asian rice is one of the major worldwide cereal crops. Among Asian countries, Vietnam is one of the world's leading rice exporters, accounting for 16% of the world trade volume of rice [18]. To the best of our knowledge, the lack of whole-genome sequencing has raised concerns in relation to the characteristics of Vietnamese rice landraces; therefore, the objective of the current work was to perform genome analysis of five rice landraces, three *indica* and two *japonica*, collected in different ecological areas of Vietnam, by applying Illumina's paired-end sequencing. The generated reads were then mapped to the Nipponbare reference sequence to analyze the genomic features and discover and annotate candidate single nucleotide polymorphisms (SNPs) and insertions and deletions (InDels). The results presented may help to unravel the genetic basis of our rice genomes and polymorphic resources for molecular marker identification in the future.

2. Materials and Methods

2.1. Plant Materials and Whole-Genome Sequencing. The five Vietnamese rice landraces were collected from different ecological areas in Vietnam: Ha Giang (104°58′51″E, 22°49′00″N) and Lang Son (106°45′40″E, 21°51′14″N) in the northeast, Bac Ninh (106°04′24″E, 21°11′15″N) in the Red River Delta, Nghe An (104°58′38″E, 19°10′35″N) in the North Central Coast, and Can Tho (105°47′03″E, 10°01′57″N) in the Mekong River Delta. For convenience, the genomes were labeled (Table 1, Table S1 in the Supplementary Material available online at https://doi.org/10.1155/2017/9272363) as indicated in the list of Vietnamese native rice landraces according to the report of Trung and Ham [19]. Total DNA of each landrace was extracted from young leaf tissue using Qiagen DNeasy kit (Qiagen, Germany). The library preparation and sequencing of the rice genomes were carried out using Illumina HiSeq 2000 by applying Illumina pipeline 1.9 at the Genome Analysis Centre (TGAC), UK. The obtained FASTQ files were further assessed by FastQC software and deposited in the NCBI sequence read archive (SRA) with accession numbers SRP064171 (*indica* 12: SRR2529343, *indica* 13: SRR2543299, *indica* 15: SRR2543338, *japonica* 11: SRR2543336, and *japonica* 14: SRR2543337).

2.2. Mapping and Identification of SNPs and InDels. The paired-end reads were first aligned with the nuclear reference genome (*O. sativa* L. cv. Nipponbare, MSU release 7.0, GenBank accession PRJDB1747) and the organellar genomes (chloroplast genome, GenBank accession NC_001320.1; mitochondria genome, GenBank accession BA00029.3) using the alignment software BWA (version 0.6.2) with default parameters. The mapping quality was assessed by Qualimap (version 2.1) and BEDtools. The duplicated sequences were marked and removed by Picard tools (version 1.79). SNPs and InDels were then called and qualified by SAMtools (version 1.1) and VarScan (version 2.3.7) with the following parameters: mapping quality of 20, depth of coverage of 10, average of base quality of 30, and variant frequency of 0.1 with SNPs and 0.3 with InDels. The distribution of SNPs and InDels per 100 kb along the chromosome was used to determine SNP-poor regions (<1 SNP/1 kb) (Figure S1). Pearson correlation coefficients of SNP density among the landraces were calculated using R. The common reads mapped to both chromosomes and organelles were removed from BAM files using SAMtools (version 1.1) and BWA (version 0.6.2).

2.3. Annotation of SNPs and InDels. The SNPs and InDels were annotated using SnpEff (version 3.6) with the GFF file of the reference genome containing positional information of rice genomic regions, exons, 5′UTR, 3′UTR, and CDS. To identify outlier genes, the cutoff values of nonsynonymous SNPs were identified by using the five-number summary of box-and-whisker plot.

3. Results

3.1. Mapping of Whole-Genome Sequencing Reads. The whole genomes of three Vietnamese *indica* and two *japonica* rice landraces were resequenced and produced 688.4 million 100 bp paired-end reads in total. Among them, 621.8 million reads (90.32%) were successfully aligned to the Nipponbare nuclear reference genome. The alignment rates of each landrace were relatively high, ranging from 86.33% to 93.87%, yielding 30x–40x coverage in depth (Table 2) (Table S2a–e for chromosome coverage in detail).

The genome coverage ranged from 89.18% to 96.49% of the reference (Table 2 and Tables S2a–S2e). The organelle

Table 2: Mapping and coverage of the reads from the landraces to Nipponbare reference genome.

Landraces	*indica* 12	*indica* 13	*indica* 15	*japonica* 11	*japonica* 14
Total reads[a]	129,251,696	112,867,645	151,576,754	133,602,832	161,141,384
# mapped reads[b]	112,667,155	101,814,936	130,851,417	125,253,671	151,261,159
Mapped reads (%)	87.17%	90.21%	86.33%	93.75%	93.87%
# unmapped reads[c]	16,584,541	11,052,709	20,725,337	8,349,161	9,880,225
Unmapped reads (%)	12.83%	9.79%	13.67%	6.25%	6.13%
Mean mapping quality[d]	42.21	44.59	42.03	47.81	48.14
Coverage[e]	91.30%	89.18%	92.56%	95.50%	96.49%
Depth[f]	29.94	27.09	34.74	33.46	40.41

[a]The raw reads in the FASTQ files. [b]The number of reads mapped to 12 chromosomes in the nucleus. [c]The number of unmapped reads to both nuclear and organellar genomes. [d]The error probability of read mapping scaled by Phred quality. [e]The breadth of coverage across the nuclear genome. [f]The sequencing depth of reads.

Table 3: Number of SNP effects annotated by SnpEff. SNPs were classified into various genomic regions (intergenic regions, intron, and exon) and effect terms by types. The total of intergenic, intronic, and exonic SNPs was more than the total number of SNPs due to overlapping gene models in the GFF file.

	indica 12		*indica* 13		*indica* 15		*japonica* 11		*japonica* 14	
Intergenic	1,374,419	68.45	1,329,647	69.46	1,527,754	68.16	488,888	66.10	454,811	65.83
Genic	633,408	31.55	584,505	30.54	713,664	31.84	250,766	33.90	236,104	34.17
Intron & regulatory sequences	288,375	45.53	279,350	47.79	320,956	44.97	109,191	43.54	101,187	42.86
UTRs	85,749	13.54	82,611	14.13	96,948	13.58	29,587	11.80	28,138	11.92
CDS	259,284	40.93	222,544	38.07	295,760	41.44	111,988	44.66	106,779	45.23
Nonsynonymous	151,998	58.62	131,754	59.20	172,767	58.41	66,645	59.51	63,962	59.90
Synonymous	107,286	41.38	90,790	40.80	122,993	41.59	45,343	40.49	42,817	40.10

genomes (mitochondrial and chloroplast) had coverage of 94%–100% (Tables S3a and S3b). There was a portion of reads (0.17%~0.26%) that could be aligned to both nuclear DNA genome and organelle genomes, mitochondrial and chloroplast (Tables S4a–S4e).

3.2. Detection and Distribution of SNPs and InDels. By using SAMtools and VarScan software, the qualified SNPs and InDels were called. The distributions of SNPs and InDels by chromosomes of each landrace are shown in Figures 1 and 2.

For the *indica* lines, the numbers of SNPs and InDels were approximately two million and three hundred thousand, respectively (Tables S5a–S5c). Accordingly, for the *japonica* lines, the numbers of SNPs and InDels were approximately seven hundred thousand and one hundred thousand, respectively (Tables S5d and S5e). For the *indica* lines, chromosomes 1 and 9 had the highest and lowest variation rates in terms of both SNP and InDel, respectively. However, there was an exception in *indica* 12, and the lowest InDel rate was on chromosome 10 instead of chromosome 9 (Tables S5a–S5c). For the *japonica* lines, the highest SNP rate was on chromosome 8, while the lowest SNP rate was on chromosomes 2 (*japonica* 11) and 3 (*japonica* 14). The highest and lowest InDel rates were on chromosomes 1 and 5, respectively, for both (Tables S5d and S5e).

The distribution of DNA polymorphisms has been examined on each 100 kb nonoverlapping window to obtain average densities of SNP and InDels of chromosomes. The average densities of SNPs and InDels of *indica* landraces were about 2.5 times that of *japonica* landraces (Tables S5a–S5e). The average densities of deletions tended to be higher than those of insertions on all chromosomes (Tables S5a–S5e). Moreover, SNP deserts where the SNP densities were below 1 SNP/kb have been identified with differing sizes (100 kb to 6.7 Mb, average of 300 kb) and chromosomal locations. Of all landraces, there were three SNP deserts of larger sizes (2.6 MB, 0.8 MB, and 0.7 Mb) on chromosomes 5, 4, and 1, respectively (Figure 3). Moreover, within all five lines, there is a small SNP desert of 0.1 Mb located from position 12.4 to 12.5 Mb on chromosome 11 in which there are no genes found. We have used the SNP densities per 100 kb interval for reconstructing the relationships among the rice landraces. By calculating the Pearson correlation coefficient, the relationship between the rice lines was observed because the patterns grouped the rice landraces into two major subspecies, confirming the classification of landraces based on the chloroplast DNA presence/absence of a deletion in the Pst-12 fragment [19]. In detail, the landraces in the same subspecies had a positive correlation close to one, whereas there was no linear relationship between *indica* and *japonica* lines with the coefficients around zero (Figure 4).

3.3. Annotation of SNPs and InDels. SNPs and InDels were annotated against the GFF file of the Nipponbare reference genome using SnpEff. SNPs mostly occurred in the intergenic regions (approximately 68.0% for *indica*, 66.0% for *japonica*), respectively (Figure 5, Table 3). For SNPs within genic

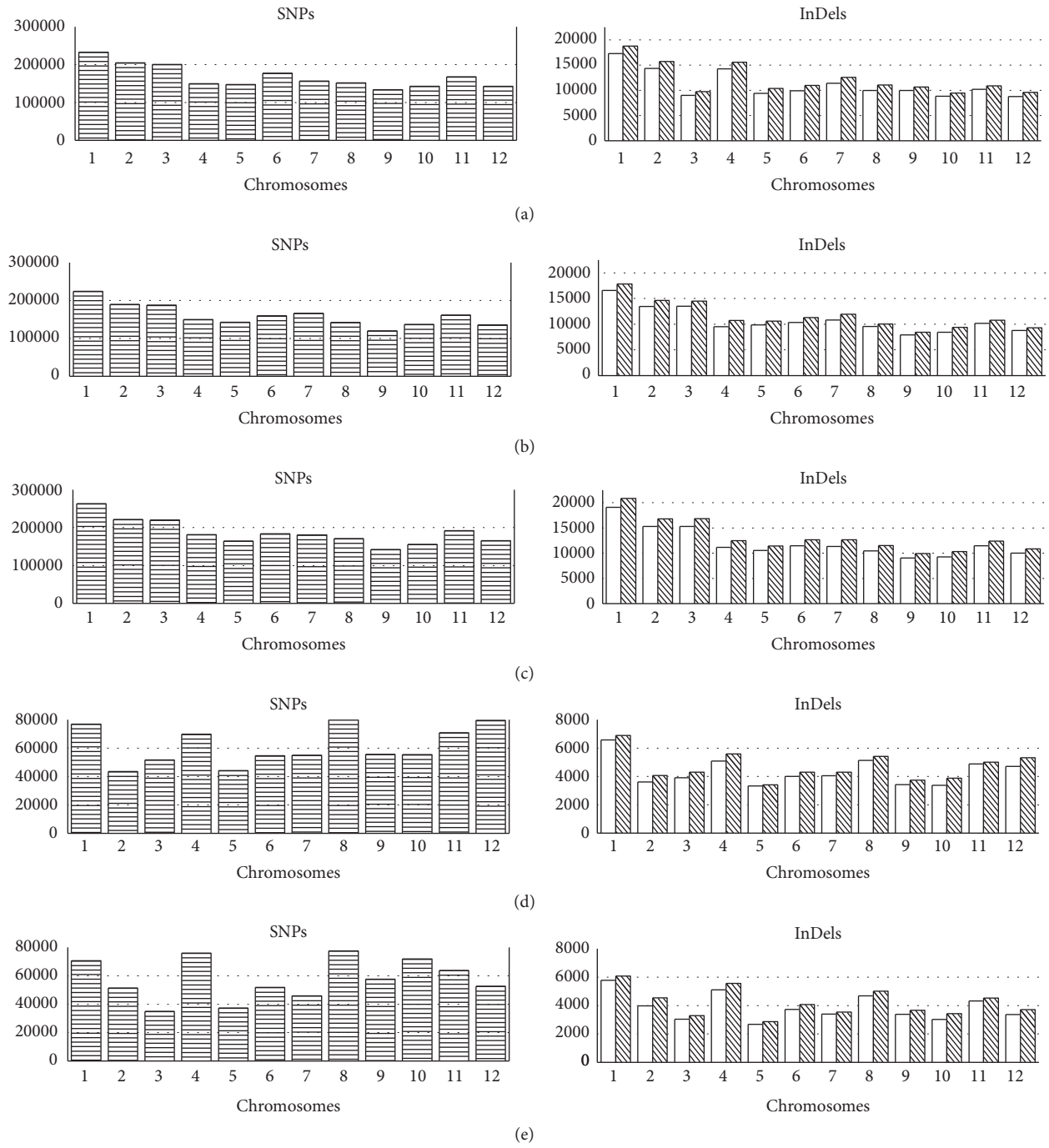


Figure 1: The number of SNPs and InDels on each chromosome in five landraces in comparison with the Nipponbare reference. The x-axis represents the chromosomes. The y-axis represents the number of SNPs (horizontal-line bars), insertions (white bars), and deletions (downward-diagonal bars). The five rice lines were (a) *indica* 12, (b) *indica* 13, (c) *indica* 15, (d) *japonica* 11, and (e) *japonica* 14.

regions (approximately 32.0% for *indica*, 34.0% for *japonica*), SNPs occurred in introns and regulatory sequences (44.9% to 47.8% for *indica*, about 43.0% for *japonica*) and UTR regions (approximately 13.75% for *indica* and 12% for *japonica*).

Among CDS regions, the split between nonsynonymous and synonymous SNPs was 58.75%, 41.25% for *indica*, and 59.7%, 40.3% for *japonica* (Figure 5, Table 3). Similarly, of the InDels, more than 73.0% were detected in intergenic regions. Most of the InDels within genic regions were within InDels or regulatory sequences (more than 62.0%), with 9.33% to 11.95% within coding sequences (Figure S2). The length of insertions ranged from one to 27 bp while the length of deletions detected was up to 41 bp. The majority of InDels were mononucleotide (≈55.5%) and dinucleotide (≈16.65%). In order to provide more insights into the effects of nonsynonymous SNPs on the genes, the distribution and skewness

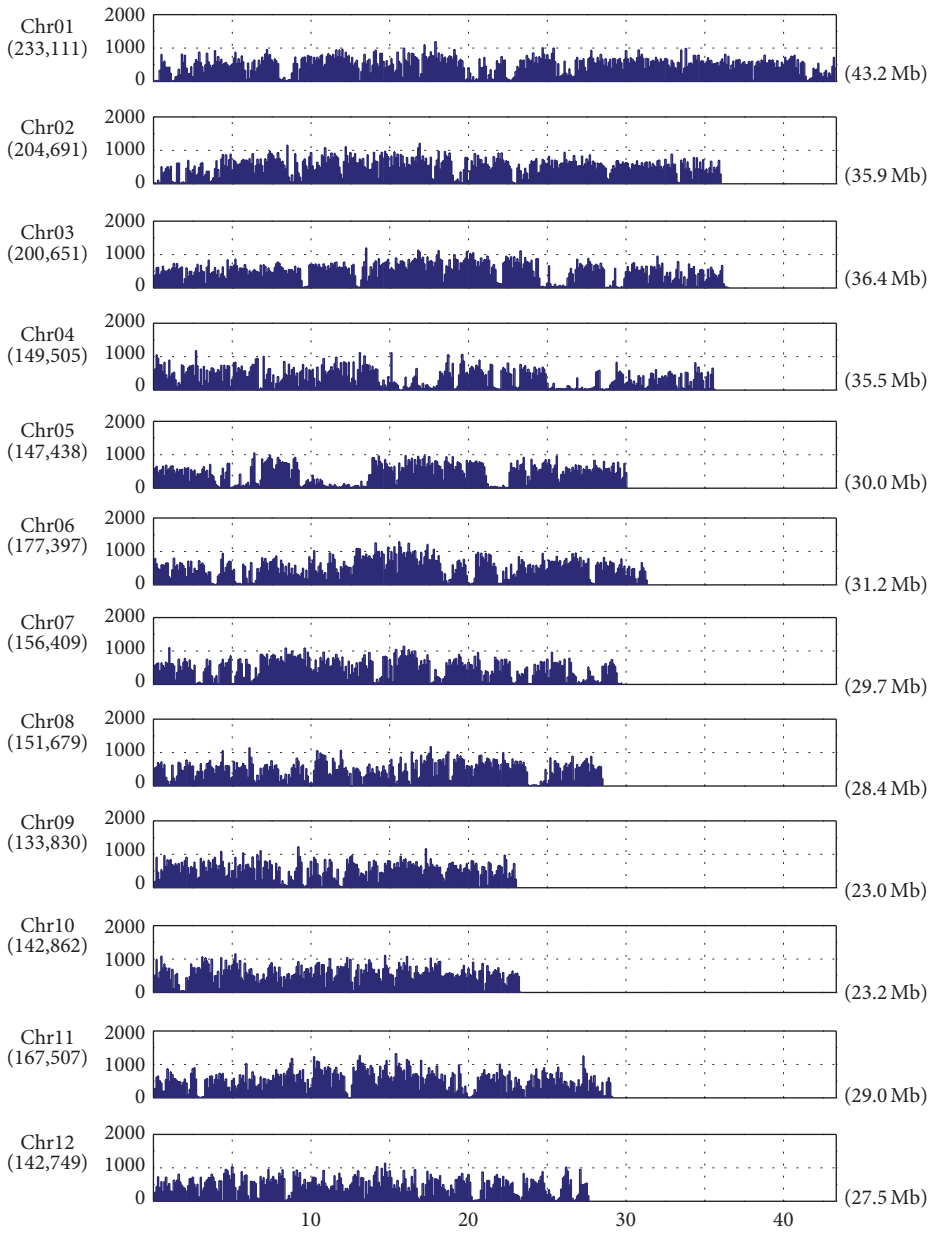


Figure 2: Distribution of SNPs between *indica* 12 and Nipponbare on the 12 chromosomes. The x-axis shows the physical distance of chromosome into 100 kb windows. The chromosome size is indicated in brackets. The y-axis represents the number of SNPs per 100 kb. The total of SNPs in each chromosome is shown in parentheses.

were calculated to identify outlier genes, which possess very high numbers of nonsynonymous SNPs as shown in Figure 6. According to the Nipponbare reference genome, nearly half of the 149 outlier genes were retrotransposon proteins (Table S6).

We have also observed that the number of transitional SNPs (A/G and C/T) was much higher than that of transversions (A/C, A/T, C/G, and G/T) for all of the five landraces with the ratios (Ts/Tv) ranging from 2.19 (*japonica* 14) to 2.37 (*indica* 12). Within transitions, the frequency of C/T was slightly higher than those of A/T and much larger than transversion SNPs (Table 4).

4. Discussion

In this study, we have sequenced five Vietnamese rice genomes consisting of three *indica* and two *japonica* landraces. The sequence datasets were aligned to the Nipponbare reference genome to identify genetic variations. The genetic variation annotation and analysis provided novel insights into the specific Vietnamese rice landraces which should be a good resource for further molecular breeding in rice. We have analyzed the sequence datasets of two major groups of rice landraces, *indica* and *japonica*, through five elite landraces in Vietnam. This provides significant information

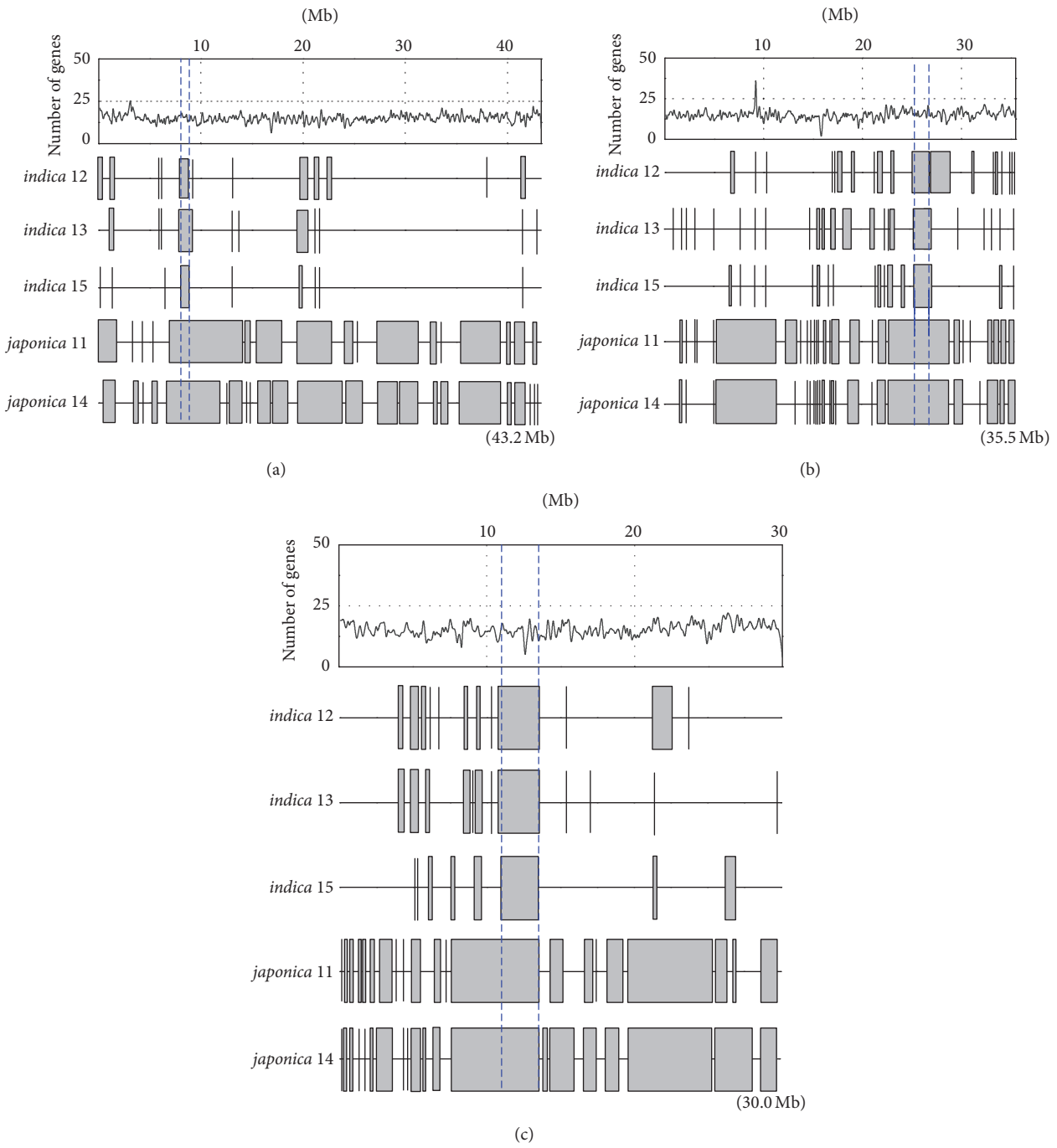


Figure 3: Distribution of "SNP deserts" on (a) chromosome 1, (b) chromosome 4, and (c) chromosome 5 in the five landraces. The low-SNP density regions are shown as gray bars or vertical lines. The vertical-dash intervals indicate the "SNP deserts" occurring in all five landraces. The line charts represent the distribution of genes across the chromosomes. The chromosome sizes are indicated in brackets.

as to the genetic diversity of the two types of rice lines in the domestication process. The results have further contributed additional evidence for the transfer of DNA regions in the organelle into the nucleus in the rice genome. Our study has also strengthened the classification of the relationship among the landraces. The detection of DNA polymorphisms can be used to identify novel genes that differentiated between our landraces and will serve as a reference for studies relating to specialty characteristics of Vietnamese rice landraces. These polymorphisms are also available for use in marker-assisted selection in rice breeding programs.

The whole genomes of five rice landraces have yielded high-quality reads, from 112 to 161 million reads per line. Most of the reads (≥86.33%) were successfully mapped to the Nipponbare reference genome with the breadth of coverage of more than 89.18%, proving that the selected rice

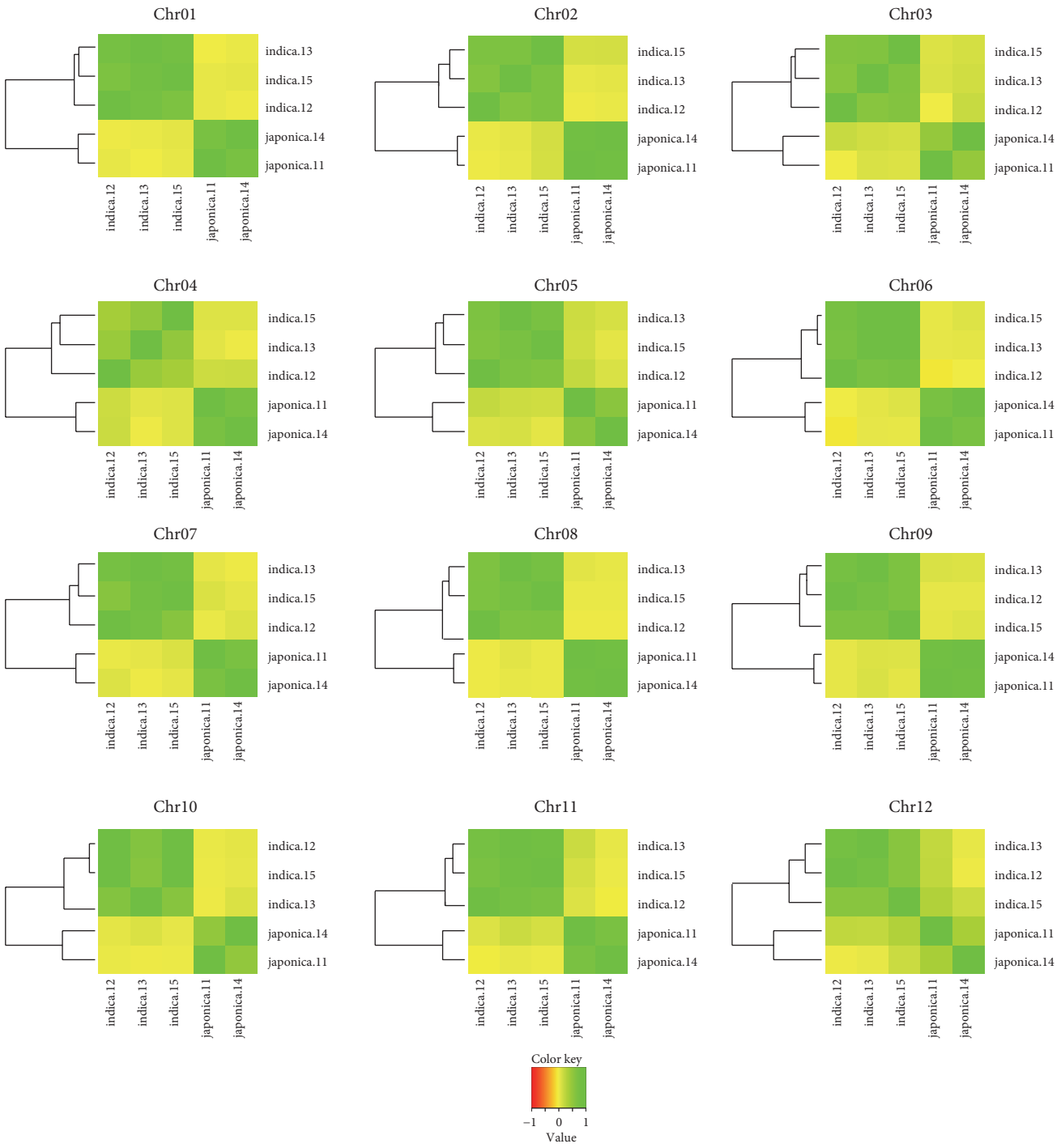


Figure 4: The heatmap and cluster view showing the correlation among the five landraces on the 12 chromosomes. Pearson correlation coefficients were calculated based on the SNP densities per 100 kb on the chromosome.

genomes are similar to the reference. Interestingly, we found a small number of reads aligned with both chromosomes and organelle genomes including chloroplast and mitochondria. This phenomenon is known as "organellar insertion" in the nuclear genome, which has been previously reported by the International Rice Genome Sequencing Project [20]. The reads mapped to the mitochondria concentrated on chromosome 12 (1.0%), and the reads mapped to the chloroplast located on chromosome 10 (0.8%). Therefore, using NGS, these data are consistent with some previous reports and also reconfirmed that chromosomes 10 and 12 have had more insertions than the others [20].

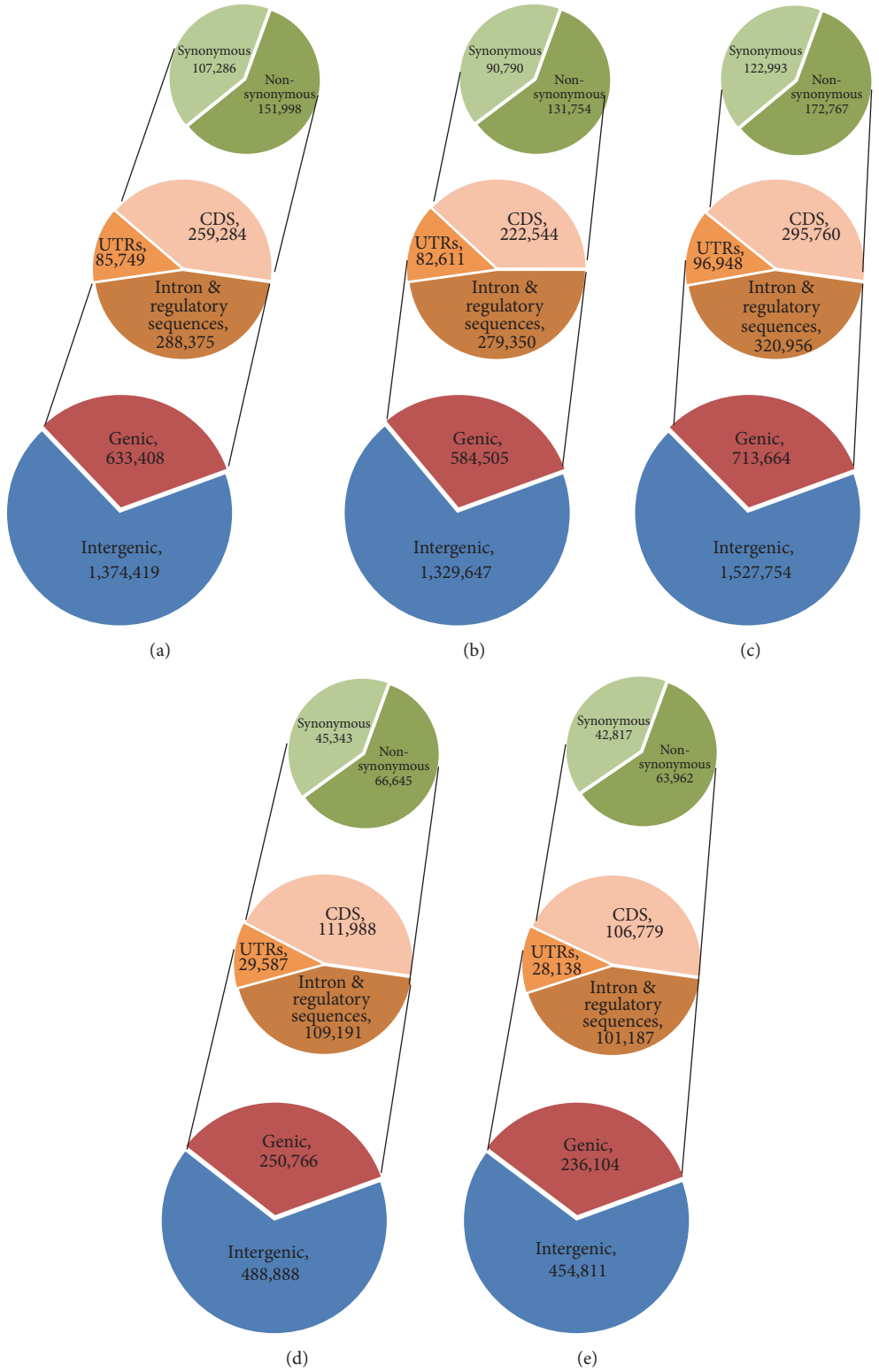


Figure 5: Annotation of single nucleotide polymorphisms (SNPs) between five Vietnamese rice landraces ((a) *indica* 12, (b) *indica* 13, (c) *indica* 15, (d) *japonica* 11, and (e) *japonica* 14) and Nipponbare based on the annotations of Nipponbare reference genome.

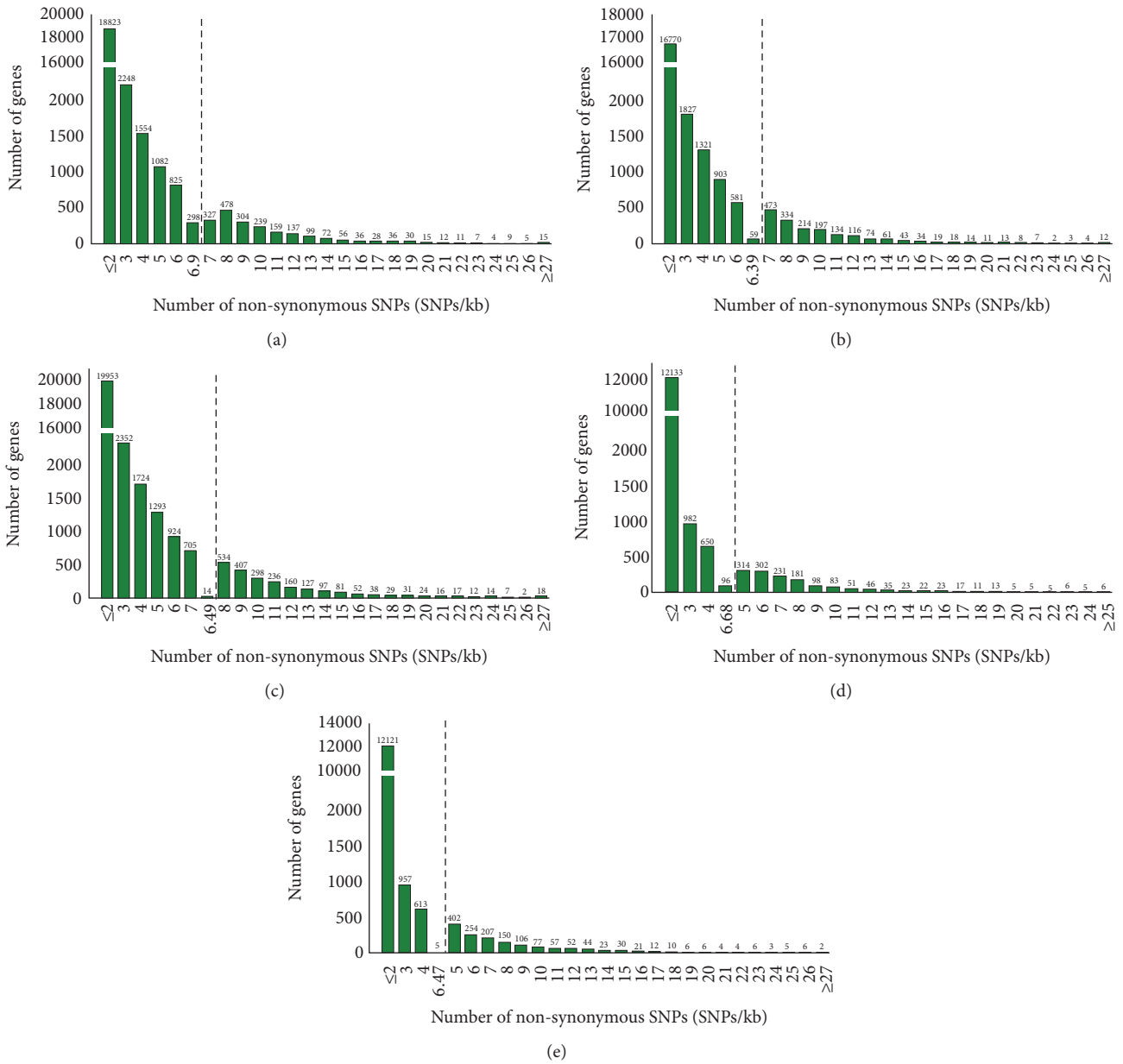


Figure 6: The distribution and skewness of the nonsynonymous SNPs. SNPs per 1 kb in (a) *indica* 12, (b) *indica* 13, (c) *indica* 15, (d) *japonica* 11, and (e) *japonica* 14. The black dash indicates the outlier value cutoff.

Table 4: Base changes in five rice genomes.

Landraces	*indica* 12	*indica* 13	*indica* 15	*japonica* 11	*japonica* 14
Transitions					
A/G	705,908	671,815	787,002	255,473	236,847
C/T	707,496	672,959	787,976	255,542	237,620
Transversions					
A/C	155,310	149,265	174,014	59,111	55,410
A/T	182,882	176,243	205,343	70,949	68,165
G/C	101,541	95,523	113,736	39,460	37,120
G/T	154,639	148,304	173,286	59,011	55,639
Ts/Tv	2.378	2.362	2.364	2.236	2.193

The sequencing depth of approximately 30x has been sufficient for detecting DNA polymorphisms. Our results of variant calling of *indica* lines (1,914,152 to 2,241,418 SNPs, 268,400 to 303,039 InDels) are in agreement with the previous study reporting that the 93-11 *indica* possesses about 1.7 million SNPs and 480 thousand InDels [21]. However, the numbers of SNPs and InDels (about 714 thousand SNPs and 102 thousand InDels) of *japonica* lines were about five and three times those of Omachi line (132,462 SNPs and 35,766 InDels) but were similar to those of Moroberekan (827,448 SNPs and 159,597 InDels), a tropical *japonica* cultivar [12, 22]. The distinct difference between tropical and temperate *japonica* landraces has been reported by Arai-Kichise et al. [22]. Further validation is underway to obtain SNP markers for rice breeding selection.

SNP deserts, genome regions of SNP rate less than 1 SNP/kb, were identified in all 5 Vietnamese landraces with sizes varying from 100 kb to 6.7 MB. The *japonica* SNP deserts are longer than those of *indica*. Mostly, SNP deserts have been previously found on chromosome 5 [23, 24]; however, we have identified two additional SNP deserts within all five Vietnamese landraces with the size of 0.7 Mb on chromosome 1 and 0.8 Mb on chromosome 4. Cheng et al. [25] have reported that SNP deserts were in the vicinity of the centromere on chromosome 5 but far from the centromere on chromosomes 1 and 4. Therefore, the current results have supported the hypothesis that SNP-low regions have not been correlated with low recombination [26]. The common SNP deserts might include highly conserved regions among the five Vietnamese rice landraces. They could result from selective sweeps reducing the variants during human selection and rice domestication [27]. These SNP deserts are able to raise fascinating questions for future studies as to whether the persistence of chromosomal regions is random or special for the individual landrace.

The SNP distribution correlation analysis of the five landraces has demonstrated the distinct divergence between *indica* and *japonica* and disclosed the phylogenetic relationships among the landraces (Figure 4); thus, it is possible to be exploited for the verification of landrace classifications. The relationship among the rice landraces in the correlation of chromosomes could be utilized for the genetic linkage disequilibrium studies.

Additionally, the phenomenon "transition bias," which means the ratios of transitions (Ts) and transversions (Tv) larger than 1 : 2, occurred among the five landraces. The transitional SNPs are more tolerated than transversional ones during mutation and natural selection because they are more likely synonymous in coding protein, resulting in conserving the protein structure [28]. Within transitions, a number of both A/G and C/T changes have been rather similar. Among transversions, the A/T transversions have been shown to be higher than others which was also reported in the genomes of citrus and rice [17, 28, 29].

In summary, by sequencing and aligning five Vietnamese rice genomes with the reference genome, Nipponbare, our results have disclosed interesting genetic information such as SNP and InDel distributions and effects and SNP deserts. Three "SNP desert" regions, which might result from selective sweeps in the domestication of rice landraces, were also observed in the different chromosomes. Furthermore, the SNP distribution analysis has revealed the phylogeny of *indica* and *japonica* with distinct classification. Further SNP validations need to be examined to identify accurate SNP markers for molecular breeding.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Hien Trinh and Khoa Truong Nguyen contributed equally to this work. Mario Caccamo, Cuong Nguyen, and Khuat Huu Trung contributed to the conception and design of the study. Sarah Ayling, Khoa Truong Nguyen, and Hien Trinh carried out the quality control of sequencing data and data analysis. Tran Dang Xuan, La Hoang Anh, Can Thu Huong, Nguyen Thuy Diep, Lam Van Nguyen, and Huy Quang Pham conceived the study and designed and performed the data analysis. Tran Dang Khanh, Hien Trinh, and Cuong Nguyen drafted and revised the manuscript. All authors contributed to the editing of the final version of the manuscript.

Acknowledgments

The authors would like to thank Hung Nguyen from MU Informatics Institute, University of Missouri, Columbia, MO 65211, USA, for his help in developing the homemade Python script to split 100 kb window sizes across the chromosomes. This work was supported by the Ministry of Science and Technology (MOST), Vietnam, and the Genome Analysis Centre (TGAC), Biotechnology and Biological Sciences Research Council (BBSRC), through a collaboration program between Vietnam and the UK. The authors are grateful to colleagues in this program that have made significant contributions to the collection, analysis, or interpretation of samples and results.

References

[1] Y. Kawahara, M. Bastide, J. P. Hamilton et al., "Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data," *Rice*, vol. 6, no. 1, article 4, 2013.

[2] R. Brenchley, M. Spannagl, M. Pfeifer et al., "Analysis of the bread wheat genome using whole-genome shotgun sequencing," *Nature*, vol. 491, no. 7426, pp. 705–710, 2012.

[3] M. Kobayashi, H. Nagasaki, V. Garcia et al., "Genome-wide analysis of intraspecific dna polymorphism in 'Micro-Tom', a model cultivar of tomato (*solanum lycopersicum*)," *Plant and Cell Physiology*, vol. 55, no. 2, pp. 445–454, 2014.

[4] C. B. Yadav, P. Bhareti, M. Muthamilarasan et al., "Genome-wide SNP identification and characterization in two soybean cultivars with contrasting mungbean yellow mosaic india virus disease resistance traits," *PLoS ONE*, vol. 10, no. 4, Article ID e0123897, 2015.

[5] M. Shimomura, H. Kanamori, S. Komatsu et al., "The *Glycine max* cv. enrei genome for improvement of Japanese soybean

cultivars," *International Journal of Genomics*, vol. 2015, Article ID 358127, 8 pages, 2015.

[6] D. R. Bentley, "Whole-genome re-sequencing," *Current Opinion in Genetics and Development*, vol. 16, no. 6, pp. 545–552, 2006.

[7] O. Morozova and M. A. Marra, "Applications of next-generation sequencing technologies in functional genomics," *Genomics*, vol. 92, no. 5, pp. 255–264, 2008.

[8] K. L. McNally, K. L. Childs, R. Bohnert et al., "Genomewide SNP variation reveals relationships among landraces and modern varieties of rice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 30, pp. 12273–12278, 2009.

[9] T. Yamamoto, H. Nagasaki, J.-I. Yonemaru et al., "Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms," *BMC Genomics*, vol. 11, article 267, 2010.

[10] X. Huang, X. Wei, T. Sang et al., "Genome-wide association studies of 14 agronomic traits in rice landraces," *Nature Genetics*, vol. 42, no. 11, pp. 961–967, 2010.

[11] Y. Liu, X. Qi, N. D. Young, K. M. Olsen, A. L. Caicedo, and Y. Jia, "Characterization of resistance genes to rice blast fungus *Magnaporthe oryzae* in a "Green Revolution" rice variety," *Molecular Breeding*, vol. 35, article no. 52, 2015.

[12] Y. Arai-Kichise, Y. Shiwa, H. Nagasaki et al., "Discovery of genome-wide DNA polymorphisms in a landrace cultivar of japonica rice by whole-genome sequencing," *Plant and Cell Physiology*, vol. 52, no. 2, pp. 274–282, 2011.

[13] J. L. Bennetzen, "Comparative sequence analysis of plant nuclear genomes: microcolinearity and its many exceptions," *Plant Cell*, vol. 12, no. 7, pp. 1021–1029, 2000.

[14] J. Yu, S. Hu, J. Wang et al., "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*)," *Science*, vol. 296, no. 5565, pp. 79–92, 2002.

[15] S. A. Goff, D. Ricke, T. H. Lan et al., "A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)," *Science*, vol. 296, no. 5565, pp. 92–100, 2005.

[16] M. Jain, K. C. Moharana, R. Shankar, R. Kumari, and R. Garg, "Genomewide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance," *Plant Biotechnology Journal*, vol. 12, no. 2, pp. 253–264, 2014.

[17] P. Rathinasabapathi, N. Purushothaman, V. L. Ramprasad, and M. Parani, "Whole genome sequencing and analysis of Swarna, a widely cultivated *indica* rice variety with low glycemic index," *Scientific Reports*, vol. 5, Article ID 11303, 2015.

[18] K. Tsukada, "Vietnam: food security in a rice-exporting country," in *The World Food Crisis and the Strategies and Asian Rice Exporter*, S. Shigetomi, K. Kubo, and K. Tsukada, Eds., Spot Survey 32, IDE-JETRO, Chiba, Japan, 2011.

[19] K. H. Trung and L. H. Ham, "Sequencing the genomes of a number of native Vietnamese rice line," in *Proceedings of the 1st National Proceeding of Crop Science, VAAS*, Hanoi, Vietnam, September 2013.

[20] Sequencing Project IRG, "The map-based sequence of the rice genome," *Nature*, vol. 436, pp. 793–800, 2005.

[21] Y.-J. Shen, H. Jiang, J.-P. Jin et al., "Development of genome-wide DNA polymorphism database for map-based cloning of rice genes," *Plant Physiology*, vol. 135, no. 3, pp. 1198–1205, 2004.

[22] Y. Arai-Kichise, Y. Shiwa, K. Ebana et al., "Genome-wide DNA polymorphisms in seven rice cultivars of temperate and tropical *japonica* groups," *PLoS ONE*, vol. 9, no. 1, Article ID e86312, 2014.

[23] G. K. Subbaiyan, D. L. E. Waters, S. K. Katiyar, A. R. Sadananda, S. Vaddadi, and R. J. Henry, "Genome-wide DNA polymorphisms in elite *indica* rice inbreds discovered by whole-genome sequencing," *Plant Biotechnology Journal*, vol. 10, no. 6, pp. 623–634, 2012.

[24] S. G. Krishnan, D. L. E. Waters, and R. J. Henry, "Australian wild rice reveals pre-domestication origin of polymorphism deserts in rice genome," *PLoS ONE*, vol. 9, no. 6, Article ID e98843, 2014.

[25] Z. Cheng, F. Dong, T. Langdon et al., "Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon," *The Plant Cell*, vol. 14, no. 8, pp. 1691–1704, 2002.

[26] F. A. Feltus, J. Wan, S. R. Schulze, J. C. Estill, N. Jiang, and A. H. Paterson, "An SNP resource for rice genetics and breeding based on subspecies Indica and Japonica genome alignments," *Genome Research*, vol. 14, no. 9, pp. 1812–1819, 2004.

[27] D. L. Waters and R. J. Henry, "Australian wild rice reveals pre-domestication origin of polymorphism deserts in rice genome," *PLoS ONE*, vol. 9, no. 6, Article ID e98843, 2014.

[28] J. Wakeley, "The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance," *Trends in Ecology & Evolution*, vol. 11, no. 4, pp. 158–162, 1996.

[29] J. Terol, M. A. Naranjo, P. Ollitrault, and M. Talon, "Development of genomic resources for *Citrus clementin*a: characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences," *BMC Genomics*, vol. 9, article 423, 2008.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2017, Article ID 2413150, 8 pages
http://dx.doi.org/10.1155/2017/2413150

Research Article

Genetic Diversity and Association Analysis for Solvent Retention Capacity in the Accessions Derived from Soft Wheat Ningmai 9

Peng Jiang, Ping-Ping Zhang, Xu Zhang, and Hong-Xiang Ma

Provincial Key Lab for Agrobiology, Jiangsu Academy of Agricultural Sciences/Jiangsu Collaborative Innovation Center for Modern Crop Production, 50 Zhongling Street, Nanjing, Jiangsu 210014, China

Correspondence should be addressed to Hong-Xiang Ma; hxma@jaas.ac.cn

Received 2 August 2016; Revised 14 October 2016; Accepted 16 January 2017; Published 5 February 2017

Academic Editor: Mihai Miclăuş

Solvent retention capacity (SRC) test is an effective method for quality evaluation of soft wheat. Ningmai 9 is a founder in soft wheat breeding. The SRC and genotype of Ningmai 9 and its 117 derivatives were tested. Association mapping was employed to identify the quantitative trait loci (QTL) associated with SRCs. Ningmai 9 had the allele frequency of 75.60% and 67.81% to its first- and second-generation derivatives, respectively, indicating higher contribution than theoretical expectation. Neighbor-joining cluster based on the genotyping data showed that Ningmai 9 and most of its first-generation derivatives were clustered together, whereas its second-generation derivatives were found in another group. The variation coefficients of SRCs in the derivatives ranged from 5.35% to 8.63%. A total of 29 markers on 13 chromosomes of the genome were associated with the SRCs. There were 6 markers associated with more than one SRC or detected in two years. The results suggested that QTL controlling SRCs in Ningmai 9 might be different from other varieties. Markers *Xgwm44*, *Xbarc126*, *Xwmc790*, and *Xgwm232* associated with SRCs in Ningmai 9 might be used for quality improvement in soft wheat breeding.

1. Introduction

Soft wheat flour of low protein content is usually associated with the cookie quality [1], which produces good quality cookies with a large spread factor, such as low thickness, tender texture with smaller particle size, and low water absorption. Soft wheat yields less flour with a smaller average particle size and less damaged starch [2].

In comparing with hard wheat, solvent retention capacity (SRC) is used more often for evaluating the quality of soft wheat in cookie making [3]. SRC is the weight of solvent held by flour after centrifugation and draining. SRC tests were developed by Slade and Levine (1994) to estimate grain and end-use quality in soft wheat [4]. They are all based on a mixture of flour plus one of four different solutions: water, 5% sodium carbonate ($NaCO_3$), 5% lactic acid, and 50% sucrose to predict water-holding capacity, damaged starch, gluten strength, and water soluble pentosan (arabinoxylan), respectively [3]. SRC was mainly determined by genotype [5–7]; however, most of the previous studies concerned the evaluation of SRC in different genotypes with various treatments, whereas the genetic mechanism of SRC received little attention.

Understanding genetic mechanisms and the identification of quantitative trait loci (QTL) associated with the components regulating end-use traits are the basis for quality improvement in wheat. Several mapping studies have been conducted to locate QTL associated with baking quality in wheat. However, most of them were conducted using hard wheat population. In soft wheat, Smith et al. (2011) reported large effect QTL for quality on chromosomes 1B and 2B [8]. Cabrera et al. (2015) identified 26 regions as potential QTL in a diversity panel and 74 QTL in all 5 biparental mapping populations [9].

Association mapping is a method to test the association between molecular markers and QTL based on linkage disequilibrium [10]. In recent years, it has been widely used for QTL detection in main crops, such as maize, wheat, and rice [11–13]. Generally, natural populations with wide genetic basis were used for association mapping [14, 15]. In soft wheat, Cabrera et al. (2015) identified 26 regions as potential QTL in a diversity panel from the soft wheat breeding program in

USA by using an association mapping approach [9]. Zhang et al. (2015) discovered several favorable allelic variations for SRCs by association mapping with a natural population including 176 varieties (lines) from China [16].

Association mapping on founder parents and its derivatives can find some important QTL and favorable allelic variations in founder parent, which can be further used for marker assisted selection to produce more favorable varieties [17]. Ningmai 9 is a soft wheat cultivar with desirable quality and has been widely used in wheat production and as parent in the Yangtze River to Huai River valley area in China. A total of 20 cultivars derived from Ningmai 9 have been released in the past 10 years. Ningmai 9 has high general combining ability in SRCs [18]; however, the QTL and chromosome regions associated with SRCs in Ningmai 9 were unclear. In this study, the genotypes of Ningmai 9 and its derivatives were screened with SSR molecular markers covering whole-genome; meanwhile the phenotypes of SRCs were analyzed in two consecutive growth seasons. The genetic structure, genetic similarity, and association mapping were analyzed to reveal the relationship between Ningmai 9 and its derivatives and to identify molecular markers associated with SRCs.

2. Materials and Methods

2.1. Plant Materials and Phenotyping. Ningmai 9 and its 117 derivatives including 39 lines of first generation and 78 lines of second generation were used in this study (Table 1). The materials were planted in 2014 and 2015 at the experimental farm of Jiangsu Academy of Agricultural Sciences in Nanjing, China. Each line was planted in a plot comprising 3 rows with two replications. Each row was sowed with 50 seeds with the length of 1.3 m and a row-to-row distance of 0.25 m. After harvest and milling, the flour was tested for SRCs according to AACC 56-11 [19, 20]. The SRC of water, sodium carbonate, lactic acid, and sucrose were described as WSRC, SCSRC, LaSRC, and SuSRC, respectively.

2.2. Genotype Analysis. DNA was extracted from fresh leaves using a CTAB procedure according to Saghai-Maroof et al. (1984) [21]. One hundred and eighty-five polymorphic simple sequence repeat (SSR) primer pairs were used to screen Ningmai 9 and its 117 derived lines in the study. These markers were randomly distributed across the wheat genome, and each chromosome included 5–14 markers with an average of 8.8 markers. Map positions of markers were based on the linkage map reported by Somers et al. (2004) [22].

Each 10 μL PCR contained 1 μL 10 × PCR buffer, 0.6 μL 15 mM $MgCl_2$, 0.8 μL 2 mMdNTP, 1 μL 0.02 μM of each primer, 0.1 μLTaq DNA polymerase, 1 μL 0.02 μM template DNA, and 3.5 μL deionized water. The cycling system consisted of an initial denaturation step of 94°C/5 min, followed by 36 cycles of 94°C/45 s, 50~60°C/45 s, 72°C/60 s, and a final extension of 72°C/10 min. Amplification bands were electrophoretically separated through nondenaturating 6% polyacrylamide gels and visualized by silver staining.

2.3. Data Processing and Analysis. Excel 2007 was used for data preparation; ANOVA was performed using SPSS 17.0.

Table 1: List of Ningmai 9 and its derivatives.

Generation	Number	Variety/lines
Parent	L1	Ningmai 9
1st generation	L2	Ningmai 13
	L3	Ningmai 14
	L4	Ningmai 16
	L5	Shengxuan 6
	L6	Yangmai 18
	L7	Yangfumai 4
	L8	3E/158
	L9	Nannong 0686
	L10	Ningmai 18
	L11	Ning 0556
	L12	Ning 07123
	L13	Ning 07119
	L14	Ning 0853
	L15	Ning 0866
	L16	Ning 0894
	L17	Ning 08105
	L18	Ning 0561
	L19	Ning 0564
	L20	Ning 0565
	L21	Ning 0417
	L22	Ning 0418
	L23	Ning 0422
	L24	Ning 0311
	L25	Ning 0316
	L26	Ning 0319
	L27	Ning 0320
	L28	Ning 0327
	L29	Ning 0331
	L35	Ning 9-11
	L36	Ning 9-36
	L37	Ning 9 Large 41
	L38	Ning 9 Large 44
	L39	Ning 9 Large 76
	L40	Ning 9 Large 78
	L41	Ning 9 Large 80
	L60	71666
	L61	6E/123
	L62	09-654
	L64	09-444
	L30	Ning 0798
	L31	Ning 07117
	L32	F307
	L33	F308
	L34	Ning 0797
	L42	Ning 0862
	L43	Ning 0869
	L44	Ning 0872
	L45	Ning 0880
	L46	Ning 0882
	L47	Ning 0884

Table 1: Continued.

Generation	Number	Variety/lines
	L48	Ning 0887
	L49	Ning 0893
	L50	Ning 0895
	L51	Ning 0897
	L52	Ning 0898
	L53	Ning 0899
	L54	Ning 08102
	L55	Ning 08104
	L56	Ning 08108
	L57	Ning 08110
	L58	Ning 08115
	L59	Ning 08116
	L63	09-569
	L65	Zhenmai 166
	L66	Ning 0867
	L67	Ning 0881
	L68	Ning 0883
	L69	Ning 0886
	L70	Ning 0896
	L71	Ning 08109
	L72	Ning 08111
2nd generation	L73	Ning 08112
	L74	Ning 08113
	L75	08F331
	L76	08F333
	L77	08F337
	L78	08F353
	L79	08F362
	L80	08F386
	L81	08F387
	L82	08F396
	L83	08F397
	L84	08F399
	L85	08F406
	L86	08F407
	L87	08F408
	L88	08F409
	L89	08F410
	L90	08F411
	L91	08F417
	L92	08F418
	L93	08F423
	L94	08F424
	L95	08F426
	L96	08F432
	L97	08F433
	L98	08F434
	L99	08F435
	L100	08F436
	L101	08F437
	L102	08F442
	L103	08F443
	L104	08F444
	L105	08F445
	L106	08F446
	L107	08F448
	L108	08F449
	L109	08F450
	L110	08F451
	L111	08F453
	L112	08F454
	L113	08F457
	L114	08F458
	L115	08F468
	L116	08F459
	L117	08F516
	L118	08F517

Neighbor-joining cluster was performed with Mega 6.0 [23]. Both the Q matrix and K matrix were determined using STRUCTURE v2.3.4 [24]. Five independent simulations were processed for each k, ranging from 1 to 8, with a 10,000 burn-in length and 100,000 iterations. The association analysis was calculated using the mixed linear model (MLM) method incorporated into the TASSEL 3.0 software [25]. The significant marker-trait associations were declared for $P \leq 0.01$.

3. Results

3.1. Genetic Contribution of Ningmai 9 to Its Derivatives. A total of 490 alleles were detected with 1–7 and an average of 2.6 alleles per locus. The ratio of allele frequency between Ningmai 9 and its derivatives on the chromosomes ranged from 55.71% to 88.29% with an average of 75.60% for first generation and from 56.33% to 83.50% with an average of 67.81% for second generation (Table 2), which indicated that Ningmai 9 had a higher contribution to its derivatives than theoretically expected. Both first and second generation had highest allele frequency on chromosome 4A. The first generation possesses the higher allele frequency compared to the second on all chromosomes except for chromosome 6D.

3.2. Population Structure Analysis and Cluster Analysis. In order to eliminate the spurious association caused by population structure of the materials, the number of populations was calculated according to the method by Evanno et al. (2005) [26]. Two populations in the materials were previously reported in our research [27].

Neighbor-joining cluster based on the genotyping data also showed that there were 2 groups in the materials (Figure 1). Ningmai 9 and most of its first-generation derivatives were clustered together, whereas its second-generation derivatives were found in another group. Yangfumai 4 was distantly clustered with those two groups since it was a

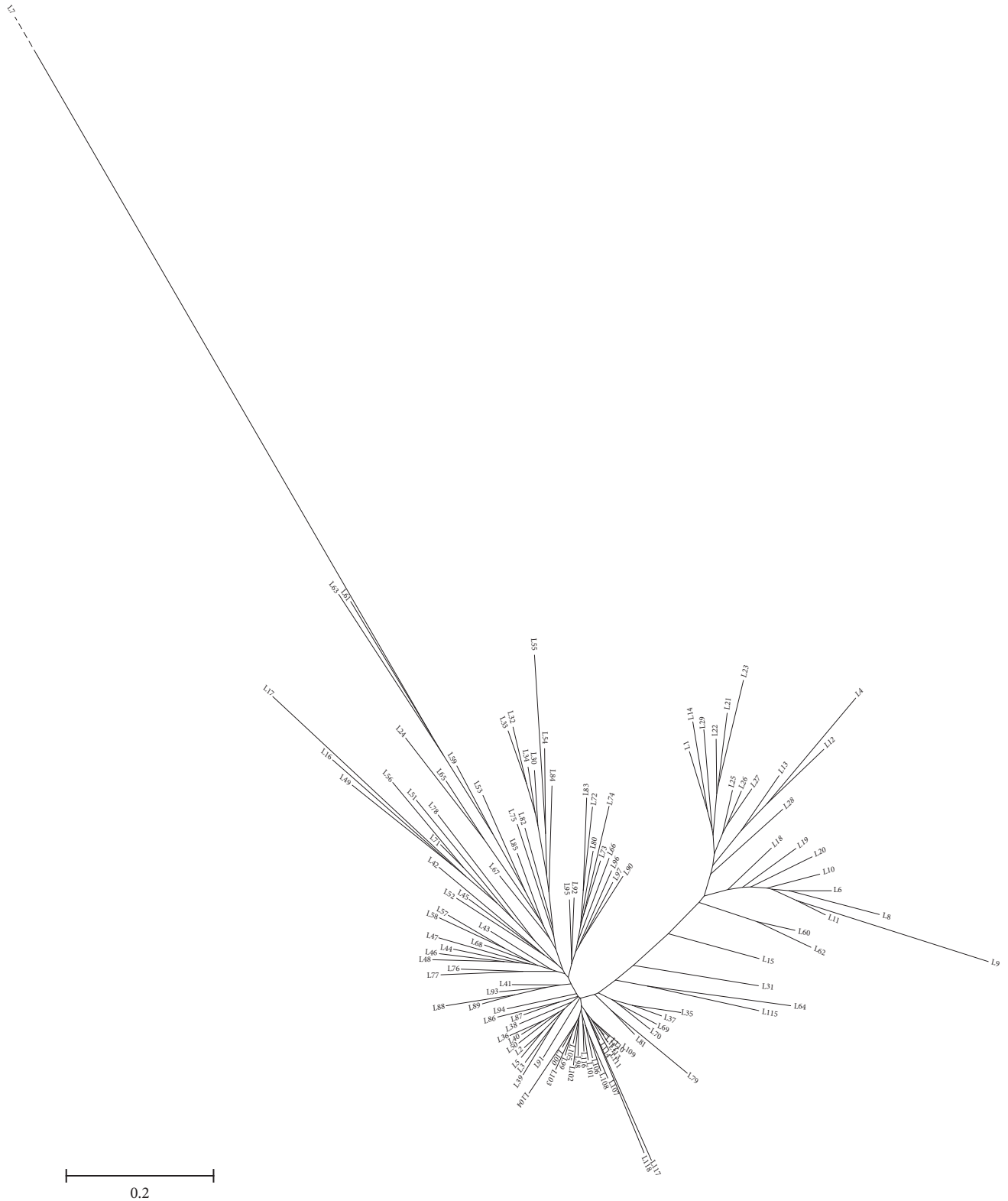


Figure 1: Neighbor-joining cluster of Ningmai 9 and its derivatives. Note: the genetic distance of L7 is so large that it is marked by dashed line.

Table 2: The allele frequency between Ningmai 9 and its derivatives on chromosomes.

Chromosome	Allele frequency (%)	
	1st generation	2nd generation
1A	71.28	57.33
2A	80.98	66.65
3A	71.23	64.16
4A	88.29	83.50
5A	77.09	75.34
6A	77.61	77.08
7A	74.47	60.23
Mean	77.28	69.18
1B	68.61	61.01
2B	73.08	70.61
3B	79.40	72.37
4B	76.71	64.02
5B	79.39	67.69
6B	69.97	62.60
7B	80.14	73.98
Mean	75.33	67.47
1D	75.30	66.54
2D	78.97	72.75
3D	64.86	56.33
4D	87.95	73.46
5D	72.49	64.24
6D	55.71	59.79
7D	84.09	74.27
Mean	74.19	66.77
Genome wide allele frequency with Ningmai 9 (%)		
1st generation	75.60	
2nd generation	67.81	

mutant induced from hybrid seed treated with ^{60}Co radiation.

3.3. Phenotype Analysis. There were significant variations among the derivatives of Ningmai 9 for all SRCs. The value of each SRC of the derivatives was higher, on average, than that of Ningmai 9, and the variations were high with coefficients of variation (CV) ranging from 5.35% in SuSRC (2014) to 8.63% in WSRC (2015) (Table 3).

ANOVA revealed significant effects of genotype for all SRCs (Table 4). Year effect was also significant for SCSRC. ANOVA showed that the effect of genotype by year was not significant for each SRC. There was no significant difference among generations for SRCs except for SCSRC, though the values of all the SRCs in second generation were larger than that in first generation and in Ningmai 9 except for the value of LaSRC.

There was significant positive correlation between the two years for all SRCs (Table 5). The correlation between different SRCs was identical in the two years; SuSRC was significantly positively correlated with WSRC, SCSRC, and LaSRC, and there was also significant correlation between WSRC and SCSRC.

3.4. Association Analysis. A total of 29 markers on 13 wheat chromosomes were associated with the SRCs (Table 6). Five markers associated with WSRC were identified on chromosomes 4A, 4D, 7B, and 7D, 21 markers on chromosomes 1B, 1D, 2A, 2B, 2D, 3B, 4A, 6A, 6B, 7A, 7B, and 7D were associated with SCSRC, two markers on chromosome 3B were associated with LaSRC, and four markers on chromosomes 1D, 2D, and 3B were associated with SuSRC. The QTL related to such markers could explain 5.12%~12.05% of the phenotypic variation. *Xgwm44* was associated with WSRC and SCSRC, and *Xwmc754* and *Xwmc326* were associated with both LaSRC and SuSRC. *Xbarc126*, *Xwmc517*, *Xgwm484*, *Xwmc754*, *Xwmc326*, and *Xgwm232* were detected in both years, and *wmc754* and *wmc326* associated with LaSRC presented different alleles in two years. Most of the alleles of the marker associated with more than one SRC or detected in two years had negative effects on their corresponding SRCs, which were a benefit for soft wheat quality.

4. Discussion

Ningmai 9 is a soft wheat variety with stable soft wheat quality, high yield, wide adaptation, and resistance to multiple diseases including *Fusarium* head blight, soil born mosaic virus, and sharp eye spots released in 1997. Since 2006, 20 wheat cultivars derived from Ningmai 9 have been released to wheat production of the Yangtze River to the Huai River regions in China. As a founder parent, Ningmai 9 has a high general combining ability in sterile spikelet number (negative effect), grain weight per spike, protein content (negative effect), SRC (negative effect), and *Fusarium* head blight resistance, which means that it is easy to produce desirable traits in progenies [18]. At genomic level, founder parents have more favorable allelic variations than other varieties, and the genetic composition of new varieties is more similar to founder parent rather than the average value of parents. In this study, the genetic contribution of Ningmai 9 to its first and second generation was 75.60% and 67.81%, respectively, which were both significantly higher than theoretical expectation of 50% and 25%. The result was consistent with previous reports on other founder parents, such as Triumph/Yanda 1817 [28], Orofen [29], Bima 4 [30], and Zhou 8425B [31].

Solvent retention capacity (SRC) has been considered as an important breeding tool for predicting flour functionality of different wheat for different end uses ever since it has been developed [4, 32, 33]. SRC addresses the relative contributions to water absorption of each flour component using four different solvents: water, lactic acid, sodium carbonate, and

Table 3: Phenotype variation for 4 SRCs of Ningmai 9 and its derivatives.

Index	Year	Ningmai 9	Mean	Stdev	Min	Max	CV (%)
WSRC	2014	59.49	64.03	5.36	49.43	79.40	8.37
	2015	59.60	63.72	5.50	49.98	77.25	8.63
SCSRC	2014	78.96	86.73	6.86	71.11	98.91	7.91
	2015	79.85	85.66	6.97	70.38	100.15	8.13
LaSRC	2014	108.52	116.88	9.24	93.98	141.71	7.91
	2015	109.05	117.23	9.57	96.21	146.57	8.16
SuSRC	2014	108.85	116.05	5.97	99.15	131.09	5.15
	2015	109.51	117.06	6.45	104.39	132.75	5.51

Table 4: ANOVA and multiple comparisons among generations for the SRCs of Ningmai 9 and its derivatives.

Index	F value			Multiple comparison test (S-N-K method)		
	Genotype	Year	Genotype × year	Ningmai 9	1st generation	2nd generation
WSRC	5.24**	0.52	0.28	59.55[a]	61.15[a]	65.29[a]
SCSRC	8.04**	6.02*	0.77	79.41[a]	81.14[ab]	88.81[b]
LaSRC	4.75**	0.21	0.54	108.79[a]	117.72[a]	116.83[a]
SuSRC	3.43**	3.00	0.51	109.18[a]	115.66[a]	117.10[a]

** and * show significant difference at 0.01 and 0.5 level, respectively; different small letters in the same row show significant difference at 0.05 level.

Table 5: Correlation analysis for the SRCs over two years in Ningmai 9 and its derivatives.

	WSRC	SCSRC	LaSRC	SuSRC
WSRC	0.897**	0.663**	−0.012	0.343**
SCSRC	0.640**	0.825**	−0.007	0.525**
LaSRC	0.037	0.085	0.809**	0.403**
SuSRC	0.311**	0.582**	0.398**	0.740**

** shows significant difference at 0.01 level. The correlation analysis for the same trait between 2014 and 2015 is marked on the diagonal; the correlation analysis among different traits in 2014 is marked below the diagonal, whereas the correlation analysis among different traits in 2015 is marked above.

sucrose. While WSRC has been associated with the overall water-holding capacity of all flour constituents, LaSRC is associated more specifically with the glutenin network formation and gluten elasticity or strength of flour. SCSRC is closely related to the amount of damaged starch of the flour, while SuSRC relates more specifically to the concentration of arabinoxylan and gliadin [19]. In this study, SRCs of Ningmai 9 and its derivatives were measured in two consecutive years, and all the SRCs of Ningmai 9 were lower than those of the derivatives on average, as wheat breeders did not take soft wheat as the only goal in wheat breeding. Therefore, genetic improvement for soft wheat quality would be strengthened in the future.

Identification of molecular marker associated with desired traits is a basis for marker assisted selection in wheat breeding. Association mapping is an effective method for identifying related markers. In this study, a total of 29 markers on 13 chromosomes were associated with the SRCs. Five markers associated with WSRC were identified on chromosomes 4A, 4D, 7B, and 7D. Cabrera et al. [9] and Carter et al. [34] discovered QTL related to WSRC on chromosomes 4A and 4D, respectively, and the QTL on 4A was close to *Xwmc468* detected in this study. Twenty-one markers on chromosomes 1B, 1D, 2A, 2B, 2D, 3B, 4A, 6A, 6B, 7A, 7B, and 7D were associated with SCSRC in the study. Wmc751 on 3B reported by Carter et al. [34] was located at the interval between *Xwmc777* and *Xwmc653*, and *Xgwm44* on 7D was also reported by Zhang et al. [16]. Smith et al. (2011) found that a QTL on 2B associated with SCSRC was close to *Xgwm257* by using 171 families from the cross Foster/Pioneer "25R26" [8]. Some markers on chromosomes 1A, 1B, 3A, 3B, 6A, and 7A related to SCSRC were also reported [8, 9, 16], but a little far from the ones we detected, as the markers on multiple chromosomes including chromosomes 1D, 2D, and 3B associated with LaSRC and SuSRC. There was high correlation between two years for all SRCs, but only a few markers were repeatedly detected, which might be due to a limited number of markers used in this study. The association mapping in Ningmai 9 and its derivatives showed that SRCs were determined by lots of minor QTL effects but also the environment, which suggest that the genetic mechanism of SRCs was complex in Ningmai 9 and QTL controlling SRCs might differ from other varieties. The favorable allelic variations of *Xgwm44*, *Xbarc126*, *Xwmc790*, and *Xgwm232* associated with SRCs in Ningmai 9 may be used for quality improvement in soft wheat breeding.

Competing Interests

The authors declare that they have no competing interests.

Table 6: Association analysis for SRCs.

Traits	Marker	Chromosome	2014			2015		
			P	R^2 (%)	Effect (allele)	P	R^2 (%)	Effect (allele)
WSRC	*Xwmc468*	4AL	4.74×10^{-3}	6.61	− (134)			
	Xwmc89	4DS	8.10×10^{-3}	5.94	− (140)			
	Xwmc517	7BL				9.68×10^{-3}	5.71	+ (183)
	Xgwm44	7DS	8.11×10^{-3}	5.78	− (196)	1.37×10^{-3}	8.84	− (196)
	Xbarc126	7DS	6.48×10^{-3}	6.13	− (170)	4.34×10^{-4}	10.80	− (170)
SCSRC	*Xgwm153*	1BL				5.65×10^{-3}	6.02	− (188)
	Xcfd72	1DL				4.55×10^{-3}	6.22	+ (310)
	Xgwm232	1DL	7.64×10^{-4}	8.57	+ (144)			
	Xwmc658	2AL	6.82×10^{-3}	5.44	+ (250)			
	Xgwm257	2BS				9.54×10^{-3}	5.16	− (186)
	Xgwm539	2DL	6.32×10^{-3}	5.56	+ (160)			
	Xgwm102	2DS	3.04×10^{-3}	6.57	+ (142)			
	Xgwm484	2DS	1.94×10^{-4}	10.81	− (179)	2.90×10^{-3}	6.96	− (179)
	Xwmc231	3B	1.09×10^{-3}	8.17	+ (240)			
	Xwmc777	3B	4.13×10^{-4}	9.48	− (100)			
	Xwmc653	3B				6.05×10^{-3}	5.83	− (160)
	Xwmc219	4AL	6.68×10^{-3}	5.47	+ (160)			
	Xgwm169	6AL				9.54×10^{-3}	5.68	− (190)
	Xwmc397	6BL	9.68×10^{-5}	12.05	—			
	Xwmc790	7A	1.68×10^{-3}	7.42	− (108)			
	Xwmc809	7A				6.04×10^{-3}	5.81	− (180)
	Xwmc311	7BL	9.64×10^{-3}	5.12	+ (120)			
	Xwmc634	7DL	4.16×10^{-4}	10.17	+ (210)			
	Xgwm437	7DL				6.04×10^{-3}	5.81	− (110)
	Xgwm44	7DS				6.61×10^{-3}	5.88	− (183)
	Xcfd14	7DS				1.80×10^{-3}	7.62	− (100)
LaSRC	*Xwmc754*	3B	2.09×10^{-3}	8.77	− (160)	7.89×10^{-3}	6.39	+ (152)
	Xwmc326	3B	7.20×10^{-3}	7.15	+ (186)	8.29×10^{-3}	7.00	+ (186)
SuSRC	*Xgwm232*	1DL	5.17×10^{-4}	10.44	− (144)	4.27×10^{-3}	7.20	− (144)
	Xgwm349	2DL	5.97×10^{-3}	6.47	+ (310)			
	Xwmc754	3B	5.96×10^{-3}	6.68	− (160)			
	Xwmc326	3B				4.38×10^{-3}	7.35	+ (186)

The number in brackets following "+" or "−" represents the allele of markers, and "+" and "−" indicate a positive or negative effect by the allele of markers.

Acknowledgments

This work was partially supported by the national key project for the research and development of China (2016YFD0100500) and the indigenous innovation foundation of Jiangsu provincial agricultural science and technology (CX[14]2002), China Agricultural Research System Program (CARS-03).

References

[1] M. Moiraghi, L. Vanzetti, C. Bainotti, M. Helguera, A. León, and G. Pérez, "Relationship between Soft wheat flour physicochemical composition and cookie-making performance," *Cereal Chemistry*, vol. 88, no. 2, pp. 130–136, 2011.

[2] A. Abboud, R. C. Hoseney, and G. Rubenthater, "Effect of fat and sugar in sugar-snap cookies and evaluation of tests to measure cookie flour quality," *Cereal Chemistry Journal*, vol. 62, pp. 124–129, 1985.

[3] M. Kweon, L. Slade, and H. Levine, "Solvent retention capacity (SRC) testing of wheat flour: principles and value in predicting flour functionality in different wheat-based food processes and in wheat breeding—a review," *Cereal Chemistry*, vol. 88, no. 6, pp. 537–552, 2011.

[4] L. Slade and H. Levine, "Structure-function relationships of cookie and cracker ingredients," *Cereal Chemistry Journal*, vol. 81, pp. 261–266, 1994.

[5] M. J. Guttieri, D. Bowen, D. Gannon, K. O'Brien, and E. Souza, "Solvent retention capacities of irrigated soft white spring wheat flours," *Crop Science*, vol. 41, no. 4, pp. 1054–1061, 2001.

[6] M. J. Guttieri and E. Souza, "Sources of variation in the solvent retention capacity test of wheat flour," *Crop Science*, vol. 43, no. 5, pp. 1628–1633, 2003.

[7] Q. J. Zhang, Y. Zhang, Z. H. He, and R. J. Pena, "Relationship between soft wheat quality traits and cookie quality parameters," *Acta Agronomica Sinica*, vol. 31, no. 9, pp. 1125–1131, 2005.

[8] N. Smith, M. Guttieri, E. Souza, J. Shoots, M. Sorrells, and C. Sneller, "Identification and validation of QTL for grain quality traits in a cross of soft wheat cultivars pioneer brand 25r26 and foster," *Crop Science*, vol. 51, no. 4, pp. 1424–1436, 2011.

[9] A. Cabrera, M. Guttieri, N. Smith et al., "Identification of milling and baking quality QTL in multiple soft wheat mapping populations," *Theoretical and Applied Genetics*, vol. 128, no. 11, pp. 2227–2242, 2015.

[10] S. A. Flint-Garcia, J. M. Thornsberry, and E. S. Buckler, "Structure of linkage disequilibrium in plants," *Annual Review of Plant Biology*, vol. 54, pp. 357–374, 2003.

[11] J. R. Andersen, T. Schrag, A. E. Melchinger, I. Zein, and T. Lübberstedt, "Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.)," *Theoretical and Applied Genetics*, vol. 111, no. 2, pp. 206–217, 2005.

[12] F. Breseghello and M. E. Sorrells, "Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars," *Genetics*, vol. 172, no. 2, pp. 1165–1177, 2006.

[13] H. A. Agrama, G. C. Eizenga, and W. Yan, "Association mapping of yield and its components in rice cultivars," *Molecular Breeding*, vol. 19, no. 4, pp. 341–356, 2007.

[14] J. Bordes, C. Ravel, J. Le Gouis, A. Lapierre, G. Charmet, and F. Balfourier, "Use of a global wheat core collection for association analysis of flour and dough quality traits," *Journal of Cereal Science*, vol. 54, no. 1, pp. 137–147, 2011.

[15] J. C. Reif, M. Gowda, H. P. Maurer et al., "Association mapping for quality traits in soft winter wheat," *Theoretical and Applied Genetics*, vol. 122, no. 5, pp. 961–970, 2011.

[16] Y. Zhang, X. Zhang, J. Guo, D. Gao, and B. Zhang, "Association analysis of solvent retention capacity in soft wheat," *Acta Agronomica Sinica*, vol. 41, no. 2, pp. 251–258, 2015.

[17] X. Y. Zhang, Y. P. Tong, G. X. You et al., "Hitchhiking effect mapping: a new approach for discovering agronomic important genes," *Scientia Agricultura Sinica*, vol. 39, pp. 1526–1535, 2006.

[18] J. B. Yao, H. X. Ma, P. P. Zhang et al., "Research of wheat elite parent Ningmai 9 and its utilization," *Acta Agriculturae Nucleatae Sinica*, vol. 26, pp. 17–21, 2012.

[19] C. Gaines, "Report of the AACC committee on soft wheat flour. Method 56-11, solvent retention capacity profile," *Cereal Foods World*, vol. 45, pp. 303–306, 2000.

[20] L. C. Haynes, A. D. Bettge, and L. Slade, "Soft wheat and flour products methods review: solvent retention capacity equation correction," *Cereal Foods World*, vol. 54, no. 4, pp. 174–175, 2009.

[21] M. A. Saghai-Maroof, K. M. Soliman, R. A. Jorgensen, and R. W. Allard, "Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 81, no. 24, pp. 8014–8018, 1984.

[22] D. J. Somers, P. Isaac, and K. Edwards, "A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.)," *Theoretical and Applied Genetics*, vol. 109, no. 6, pp. 1105–1114, 2004.

[23] 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.

[24] M. J. Hubisz, D. Falush, M. Stephens, and J. K. Pritchard, "Inferring weak population structure with the assistance of sample group information," *Molecular Ecology Resources*, vol. 9, no. 5, pp. 1322–1332, 2009.

[25] P. J. Bradbury, Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler, "TASSEL: software for association mapping of complex traits in diverse samples," *Bioinformatics*, vol. 23, no. 19, pp. 2633–2635, 2007.

[26] 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.

[27] P. Jiang, P.-P. Zhang, X. Zhang, X. Chen, and H.-X. Ma, "Association analysis for mixograph properties in Ningmai 9 and its deriva-tives," *Acta Agronomica Sinica*, vol. 42, no. 8, pp. 1168–1175, 2016.

[28] J. Han, L. Zhang, J. Li et al., "Molecular dissection of core parental cross "Triumph/Yanda1817" and its derivatives in wheat breeding program," *Acta Agronomica Sinica*, vol. 35, no. 8, pp. 1395–1404, 2009.

[29] X. J. Li, X. Xu, W. H. Liu, X. Q. Li, and L. H. Li, "Genetic diversity of the founder parent Orofen and its progenies revealed by SSR markers," *Scientia Agricultura Sinica*, vol. 42, pp. 3397–3404, 2009.

[30] Y.-Y. Yuan, Q.-Z. Wang, F. Cui, J.-T. Zhang, B. Du, and H.-G. Wang, "Specific loci in genome of wheat milestone parent bima 4 and their transmission in derivatives," *Acta Agronomica Sinica*, vol. 36, no. 1, pp. 9–16, 2010.

[31] Y. G. Xiao, G. H. Yin, H. H. Li et al., "Genetic diversity and genome-wide association analysis of stripe rust resistance among the core wheat parent zhou 8425b and its derivatives," *Scientia Agricultura Sinica*, vol. 44, pp. 3919–3929, 2011.

[32] A. Colombo, G. T. Pérez, P. D. Ribotta, and A. E. León, "A comparative study of physicochemical tests for quality prediction of Argentine wheat flours used as corrector flours and for cookie production," *Journal of Cereal Science*, vol. 48, no. 3, pp. 775–780, 2008.

[33] C. Guzmán, G. Posadas-Romano, N. Hernández-Espinosa, A. Morales-Dorantes, and R. J. Peña, "A new standard water absorption criteria based on solvent retention capacity (SRC) to determine dough mixing properties, viscoelasticity, and bread-making quality," *Journal of Cereal Science*, vol. 66, pp. 59–65, 2015.

[34] A. H. Carter, K. Garland-Campbell, and K. K. Kidwell, "Genetic mapping of quantitative trait loci associated with important agronomic traits in the spring wheat (*Triticum aestivum* L.) Cross "Louise" × "Penawawa"," *Crop Science*, vol. 51, no. 1, pp. 84–95, 2011.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2017, Article ID 5196763, 8 pages
http://dx.doi.org/10.1155/2017/5196763

Research Article

Development of a New Marker System for Identification of *Spirodela polyrhiza* and *Landoltia punctata*

Bo Feng,[1] Yang Fang,[1,2] Zhibin Xu,[1] Chao Xiang,[1] Chunhong Zhou,[1] Fei Jiang,[1] Tao Wang,[1] and Hai Zhao[1,2]

[1]*Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China*
[2]*Chengdu University, Chengdu 610106, China*

Correspondence should be addressed to Hai Zhao; zhaohai@cib.ac.cn

Received 12 August 2016; Revised 30 October 2016; Accepted 16 November 2016; Published 12 January 2017

Academic Editor: Wenqin Wang

Lemnaceae (commonly called duckweed) is an aquatic plant ideal for quantitative analysis in plant sciences. Several species of this family represent the smallest and fastest growing flowering plants. Different ecotypes of the same species vary in their biochemical and physiological properties. Thus, selecting of desirable ecotypes of a species is very important. Here, we developed a simple and rapid molecular identification system for *Spirodela polyrhiza* and *Landoltia punctata* based on the sequence polymorphism. First, several pairs of primers were designed and three markers were selected as good for identification. After PCR amplification, DNA fragments (the combination of three PCR products) in different duckweeds were detected using capillary electrophoresis. The high-resolution capillary electrophoresis displayed high identity to the sequencing results. The combination of the PCR products containing several DNA fragments highly improved the identification frequency. These results indicate that this method is not only good for interspecies identification but also ideal for intraspecies distinguishing. Meanwhile, 11 haplotypes were found in both the *S. polyrhiza* and *L. punctata* ecotypes. The results suggest that this marker system is useful for large-scale identification of duckweed and for the screening of desirable ecotypes to improve the diverse usage in duckweed utilization.

1. Introduction

The Lemnaceae (duckweed) comprise a widespread family of monocotyledonous plants growing in water [1–3]. It is an ideal plant with several unique properties such as fast reproduction through gemmation and high protein content and it can absorb large amounts of nutrients such as nitrogen (N) and phosphorus (P) [4–6]. As a result, duckweed has shown great potential in recovering nutrients from wastewater.

Spirodela polyrhiza and *Landoltia punctata* are two species of Lemnaceae which are widely distributed around the world. On the morphological basis, the identification of these two species (*L. punctata* also is a genus) is commonly based on the number of roots. Normally, *S. polyrhiza* has seven to 21 roots with one root (rarely two) perforating the prophyllum. *L. punctata* has two to seven roots per frond; however, one and up to 12 roots were rarely observed. Sometimes, a spot is found in *S. polyrhiza* because of the accumulation of anthocyanin under limitation of nutrients. Fronds of *L. punctata* often have a red dorsal surface.

S. polyrhiza is considered as a potential energy crop which can be used for bioethanol production due to its fast growth rate and starch accumulation capability. The protein content of some ecotypes of *S. polyrhiza* grown on anaerobically treated swine wastewater was found to reach as high as 45% of the dry weight [7, 8]. With the removal of nutrients, the starch content could be increased by 59.3% within four days at 5°C [9]. After the fermentation of the enzymatic hydrolysis of the duckweed biomass, the ethanol content was corresponding to 50.9% of the original dry duckweed biomass [8].

S. polyrhiza and *L. punctata* can be used as a stable and efficient gene expression system [10]. The rapid clonal growth and simple axenic culture have made them suitable laboratory subjects for researching such diverse topics as photoperiod,

leaf morphogenesis, and toxicology on plants [11]. Aprotinin, a small serine protease inhibitor used in human health, has been stably transformed and expressed in *Spirodela* [12].

Being extremely reduced in morphology, miniaturization of organs, and its wide distribution, as well as high phenotypic plasticity in response to environmental conditions, have made taxonomy of Lemnaceae a great challenge [3, 13]. As a result, the employment of genetic markers for identification of duckweed at the inter- and intraspecies-level is used to confirm the morphological classification results [14–16]. Previously, genetic markers such as RAPD, AFLP, and DNA barcode were employed for identification of the phylogenetic relationship of duckweed [2, 17]. The DNA barcode markers based on cp-DNA sequences were for species identification [18]. The marker atpF-atpH noncoding spacer was considered to be able to serve as a universal marker for species-level identification.

However, no marker was reported to be useful for ecotype (intraspecies) level identification so far. Most significant diagnostic value is at specie level. For example, the number of roots: *L. punctata* normally possesses two to seven roots per frond. Seven to 21 roots are present in *S. polyrhiza* (Martirosyan et al. 2009). Different ecotypes of the same species vary significantly in their biochemical and physiological properties. Under the low temperature, some ecotypes show several times higher turion formation capacity than other ecotypes from *S. polyrhiza*. By using 27 ecotypes, the range of number of turions formed per frond was ranged from 0.22 to 5.9 (Kuehdorf et al. 2014). Under standardised cultivation conditions, the relative growth rate and weekly yields of 12 ecotypes from *L. punctata* were determined. Relative growth rate ranged from 0.253 to 0.509 days (−1) and relative yields from 5.9 to 17.8 weeks (−1). Meanwhile, the result shows that relative growth rate does not vary primarily at the level of the species but rather reflects the adaptation of individual ecotypes to specific condition (Ziegler et al. 2014). Under the same treatment, the starch content of the different ecotypes from the same species varies significantly. For *L. punctata*, the starch content of ecotypes ZH1010 and ZH1031 [19] was 27.3% and 18.18%, respectively. For *S. polyrhiza*, the starch content of ecotypes ZH1045 and ZH1027 was 14.9% and 30.1%, respectively. Meanwhile, the component and amount of the flavonoids with the same condition vary greatly (unpublished result, data not shown).

In this study, we present the development of new markers for intraspecies identification of duckweed. We establish a simple and accessible protocol to construct a database against the individual duckweed which could be validated. Many new ecotypes were found by using this method.

2. Materials and Methods

2.1. Plant Materials. A worldwide collection of duckweed has been characterized by morphologic classification. Duckweed was classified according to Les et al. [14]. From this collection, 48 ecotypes of *L. punctata* and 49 ecotypes of *S. polyrhiza* were selected. 18 ecotypes were collected from the Institut für Integrative Biologie (Zürich, Switzerland).

Table 1: Number of ecotypes per country and geographical area for samples characterized.

Country	*Landoltia* Sample number	*Spirodela* Sample number
China	39	23
Vietnam	3	8
South America	1	2
Asia	1	6
Europe	1	9
Australia	2	0
Pacific	1	0
Africa	0	1

The ecotypes used encompass the worldwide geographic distribution ranging from 5 m to 1890 in altitude (Table 1; Table S1, in Supplementary Material available online at https://doi.org/10.1155/2017/5196763). After collection, these ecotypes were maintained in LB plates. A summary of all ecotypes included in this study was listed in Table S1.

2.2. Analysis Using DNA Barcoding Markers. Previously, three noncoding spacers (atpF-atpH, psbK-psbI, and trnH-psbA) were used for genetic analysis of duckweed [18]. To validate the efficiency of the three markers, the PCR amplification and products sequencing were conducted. All the ecotypes of *L. punctata* and *S. polyrhiza* were used. Total DNA was extracted using CTAB. The primers sequences and PCR amplification condition were conducted according to Wang et al. [18]. The PCR products were fractionated using 2% agarose gels. The fragments were cloned into pGEM-T vector and sequenced by automatic DNA sequencing. Each product was sequenced at least three times.

2.3. Development and Validation of SSR Marker. 48 pairs of primer designed by the sequence of the genomic and chloroplast DNA were synthesized by Invitrogen Company. For validation of the primers, two ecotypes of duckweed from *Landoltia* and *Spirodela* were used as template. The primers showing good ability to detect the polymorphism among the accessions were selected for further analysis (Table 2).

PCR was carried on the Mastercycler Thermal Cycler (Eppendorf). For amplification, a total of 50 μL reaction contained 50 ng of genomic DNA, 25 μL of 2x Buffer, 0.5 mM of each of the dNTPs, 0.25 mM of $MgCl_2$, 0.5 μM of forward and reverse primers, and 2 units of KOD Plus Polymerase (TOYOBO). The PCR conditions were one cycle of 95°C for four min and 28 cycles of 94°C for 30 sec, 52°C for 30 sec, and 72°C for 30 sec, followed by a final extension of 72°C for 10 min.

2.4. Analysis of PCR Products Using the Applied Biosystems 3730 DNA Analyzer. PCR amplification by using the selected primers was conducted and the products were diluted 1 : 30 in water. One μL of the diluted PCR products was added to 7 μL of the HIDI-formamide and 0.1 μL of GeneScan 500 LIZ size standard (Applied Biosystems, Forster City, CA). The

Table 2: Three pairs of primers designed and used well for duckweed ecotype identification. "T (°C)" represents the degenerate temperature.

Primers	Primer sequences (5′-3′)	T (°C)	PCR product length range
SC09/10	TTAGTATTGTGCGACATTCG TTTCTTTGATTTGAACTCCC	52	170–256 bp
SC19/20	GCGTTCTGTTTCTTTACCTA CGGAGTAGAGCAGTTTGG	53	212–284 bp
SC35/36	ACCCTGGAGCATACCTTG AGGATTAGGAATGGGCGT	53	224–298 bp

Table 3: Success ratios of PCR amplification and sequencing results of three pairs of designed markers.

	atpF-atpH		psbK-psbI		trnH-psbA	
	Spirodela	*Landoltia*	*Spirodela*	*Landoltia*	*Spirodela*	*Landoltia*
Length of product 1	662	683	522	531	489	273
Length of product 2	—	—	—	501	484	—
Length of product 3	—	—	—	—	273	—
% success of PCR	100%	100%	100%	100%	100%	98%

mixtures underwent denaturation at 95°C for 10 minutes and then were analyzed in the Applied Biosystems 3730 DNA Analyzer. The patterns of the DNA fragment were analyzed with GeneMarker V2.2.0 software.

3. Results

3.1. Length Polymorphisms among Different Duckweeds by DNA Barcoding Markers. Previously, three DNA barcoding markers (atpF-atpH, psbK-psbI, and trnH-psbA) were detected to be easy for amplification and good for identification of different type of duckweeds. In this study, 97 ecotypes were used (48 *L. punctata* and 49 *S. polyrhiza*) and the PCR products length was detected. The results show only one type of length was found for primers atpF-atpH (683 bp) and trnH-psbA (273 bp) in *S. polyrhiza*. psbK-psbI acquired two types of products length (501 bp and 531 bp) in *S. polyrhiza*. For *L. punctata*, one type of product length was found in atpF-atpH (662 bp) and psbK-psbI (522 bp) primers PCR reaction. Three types of products length (489 bp, 484 bp, and 273 bp) were detected in trnH-psbA primer PCR reaction. All the three pairs of primers show high success for PCR amplification, with 100% success for atpF-atpH and psbK-psbI and 98% success for trnH-psbA (Table 3).

3.2. Validation of Markers Designed. For identification of different ecotypes, the primers designed should meet two basic requirements: first is high success rate of PCR amplification and the second is high PCR products polymorphism to permit different ecotypes to be distinguished. To evaluate these 48 primers, genomic DNA extracted from ecotypes (12 *S. polyrhiza* and 12 *L. punctata*) was used for PCR amplification. To meet the two requirements, three markers were found suitable (both high PCR amplification success and high polymorphism) for all ecotypes identification.

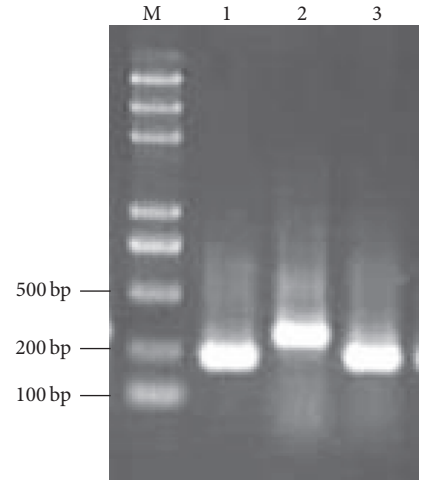


Figure 1: Electrophoresis of the PCR products amplified from duckweed (*Landoltia punctata* ecotype ZH0001-S-0) with the designed three pairs of primers in an agarose gel. M: DNA ladder Marker III (100, 200, 500, 750, 1,000, 2,000, 3,000, and 5,000 bp; Tiangen Biotech Co., Ltd.). Line 1: primers SC19/20. Line 2: primers SC35/36. Line 3: primers SC09/10.

For *L. punctata*, the PCR product length of primers SC09/10, SC19/20, and SC35/36 was quite variable, with 186–256 bp, 222–284 bp, and 234–298 bp, respectively (Figure 1; Table 2). All the three pairs of primers show high success for PCR amplification, with 95% success for SC09/10, 90% success for SC19/20, and 100% success for SC35/36. Meanwhile, some ecotypes gave more than one fragment (mostly two fragments). For *S. polyrhiza*, the PCR product length of primers SC09/10, SC19/20, and SC35/36 also showed significant variability, with 170–254 bp, 212–266 bp, and 224–298 bp, respectively (Figure 1; Table 2). All the three pairs of primers

show high success for PCR amplification, with 92% success for SC09/10, 95% success for SC19/20, and 100% success for SC35/36. Two fragments were detected in some ecotypes.

3.3. Length Polymorphism of Duckweeds Revealed by Applied Biosystems 3730 DNA Analyzer. To efficiently identify the polymorphism of each ecotype, the PCR products of three pairs of primers were mixed and then measured by Applied Biosystems 3730 DNA Analyzer. As a result, each ecotype product mixture contained fragments from three pairs of primers. This made the detection more efficient and cost lower. Theoretically, at least three fragments could be found from each ecotype except for some unsuccessful PCR amplification which resulted in two fragments. 11 haplotypes were detected in *Landoltia* and *Spirodela* ecotypes, and most of them contained three fragments. Two and three haplotypes were found with only two fragments in *Landoltia* and *Spirodela* ecotypes, respectively. Four and three haplotypes were detected with four fragments in *Landoltia* and *Spirodela*, respectively. Only one haplotype was comprised of five fragments in both species (Table 3).

Table 4: Multilocus haplotypes defined with three pairs of primers. Alleles codes correspond to the size of the PCR products.

Landoltia haplotype number	SC09/10, SC19/20, SC35/36
1	221, 236, 260, 270
2	221, 260, 269
3	221, 232, 256, 270
4	221, 236, 256, 260, 270
5	222, 256, 270
6	222, 227, 251, 270
7	222, 252, 270
8	222, 284
9	217, 223, 278
10	222, 278
11	217, 223, 261, 285
Spirodela haplotype number	SC09/10, SC19/20, SC35/36
1	187, 214, 238, 268
2	214, 260, 270, 282, 298
3	214, 275, 282, 298
4	214, 282, 298
5	214, 298
6	237, 268
7	186, 267
8	214, 287, 298
9	195, 226, 268
10	212, 272, 279
11	187, 211, 221, 274

4. Discussion

Duckweed has been researched intensively for its promising usage in bioenergy, biomedicine, and waste water treatment. A *Spirodela polyrhiza* (160 Mb) has been selected for whole genome sequencing by DOE-JGI community sequencing program. With genome sequence information, the gene discovery and functional verification could be conducted in this aquatic monocot family. Meanwhile, from a systematic view, the morphological character combined with the DNA sequence method could resolve the classification problem. However, until the sequencing of many ecotypes of duckweed, the identification of this family could be conducted by genetic markers. Indeed, many genetic markers have been designed and used for the phylogenetic study [18, 20].

In this study, we validated the most useful DNA barcoding markers for duckweed previously reported and designed new markers for ecotype identification [18]. Ecotypes from two species of *S. polyrhiza* and *L. punctata* were selected for analysis. These ecotypes represent a worldwide collection which resulted highly accessible for phylogenetic and genomic research (Table 1). Meanwhile, ecotypes from these two species are easy to be collected and morphological classified.

First, the DNA barcoding markers were found useful for interspecies identification but not suitable for intraspecies distinguishing. Previously, the DNA barcoding markers atpF-atpH, psbK-psbI, and trnH-psbA were selected for duckweed species identification [18]. To validate these three markers for ecotypes identification, PCR amplification and products sequencing were conducted as reported. High success of PCR amplification was acquired with only three PCR amplifications failed (Table 3). For DNA barcoding marker atpF-atpH, 48 ecotypes of *L. punctata* and 49 ecotypes of *S. polyrhiza* were found both with only one length of PCR products. As a result, this marker was not suitable for discriminating the different ecotype from these two species because none of the polymorphism was detected (Table 3). The similar results were found in DNA barcoding marker psbK-psbI. Two length polymorphisms were found of the marker psbK-psbI for ecotypes from *L. punctata* and three length polymorphisms were found of the marker trnH-psbA for ecotypes from *L. punctata* (Table 3). As a result, these three markers were good for species identification not suitable for ecotype discrimination.

Because of the low polymorphism detected from DNA barcoding markers, we designed new markers for intraspecies identification of duckweed. 48 pairs of primers were designed and validated by two types of ecotypes. According to two criterions (the high success of PCR amplification and polymorphism between the two ecotypes), three markers were found. The length of PCR products was highly polymorphic for all these three markers, with a range of 170–254 bp, 212–266 bp, and 224–298 bp, respectively (Table 2; Figure 2). Furthermore, for detecting of more polymorphisms, the products of three markers were mixed and fractionated by Applied Biosystems 3730 DNA Analyzer which is a capillary electrophoresis. This electrophoresis can discriminate one base pair of the PCR products with several fragments. This method is highly effective and accurate and costs much less than sequencing for polymorphism detection (Figure 2).

By using this method, 11 haplotypes of *S. polyrhiza* and 11 haplotypes of *L. punctata* were found in 97 ecotypes (Table 4). On average, four ecotypes can find one haplotype, which is much efficient compared to DNA barcoding markers used. As a result, more and more haplotypes could

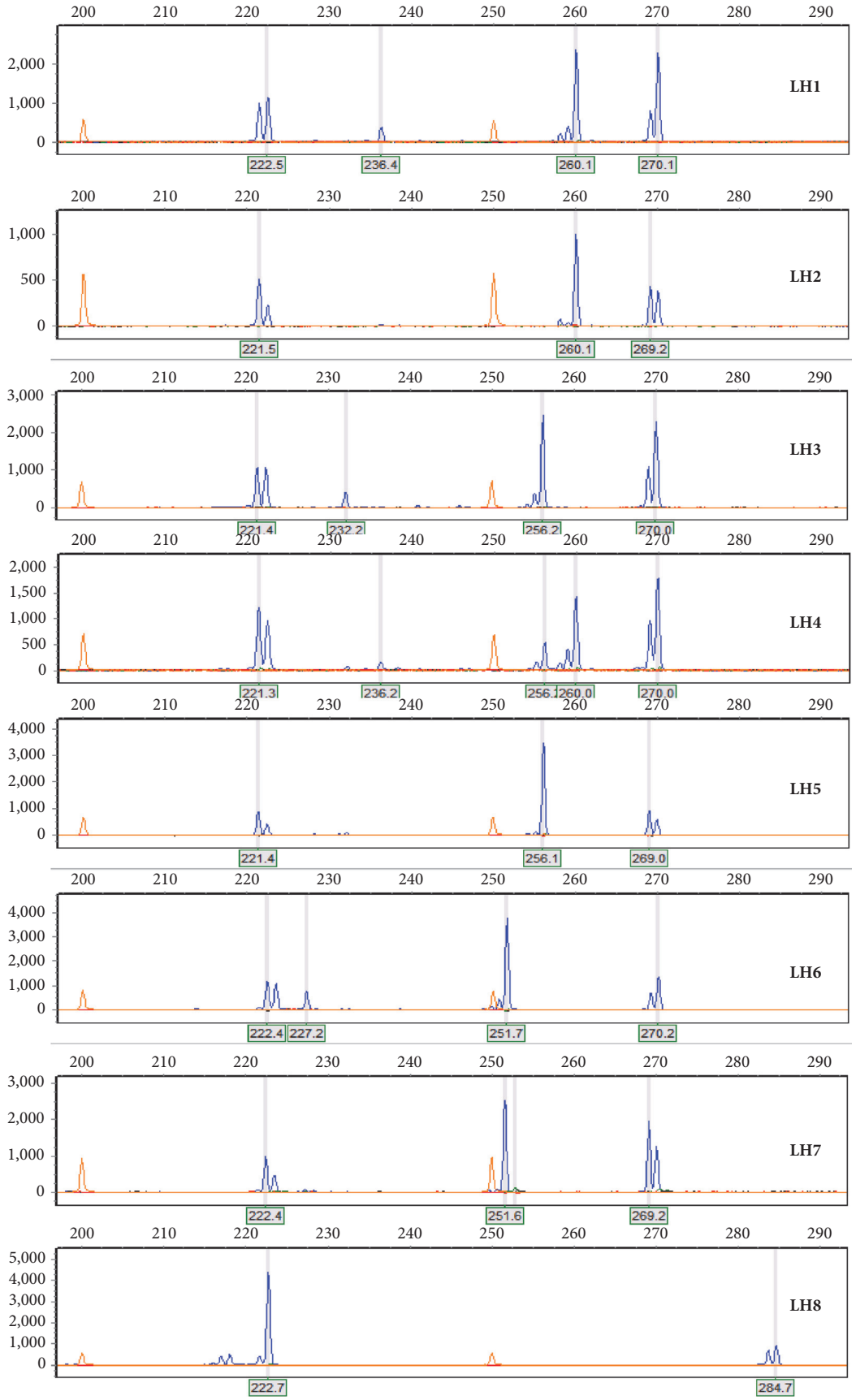


Figure 2: Continued.

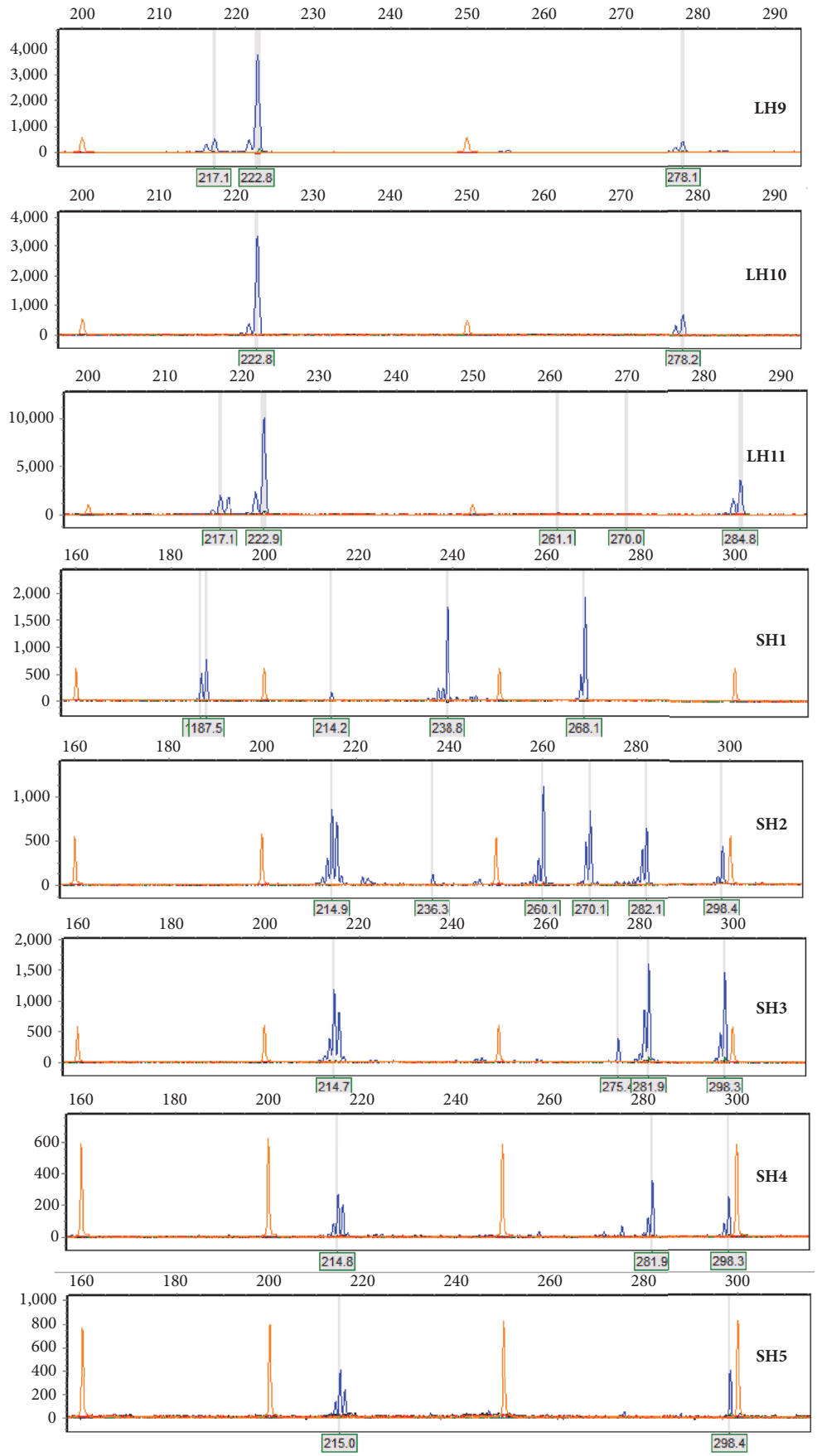


Figure 2: Continued.

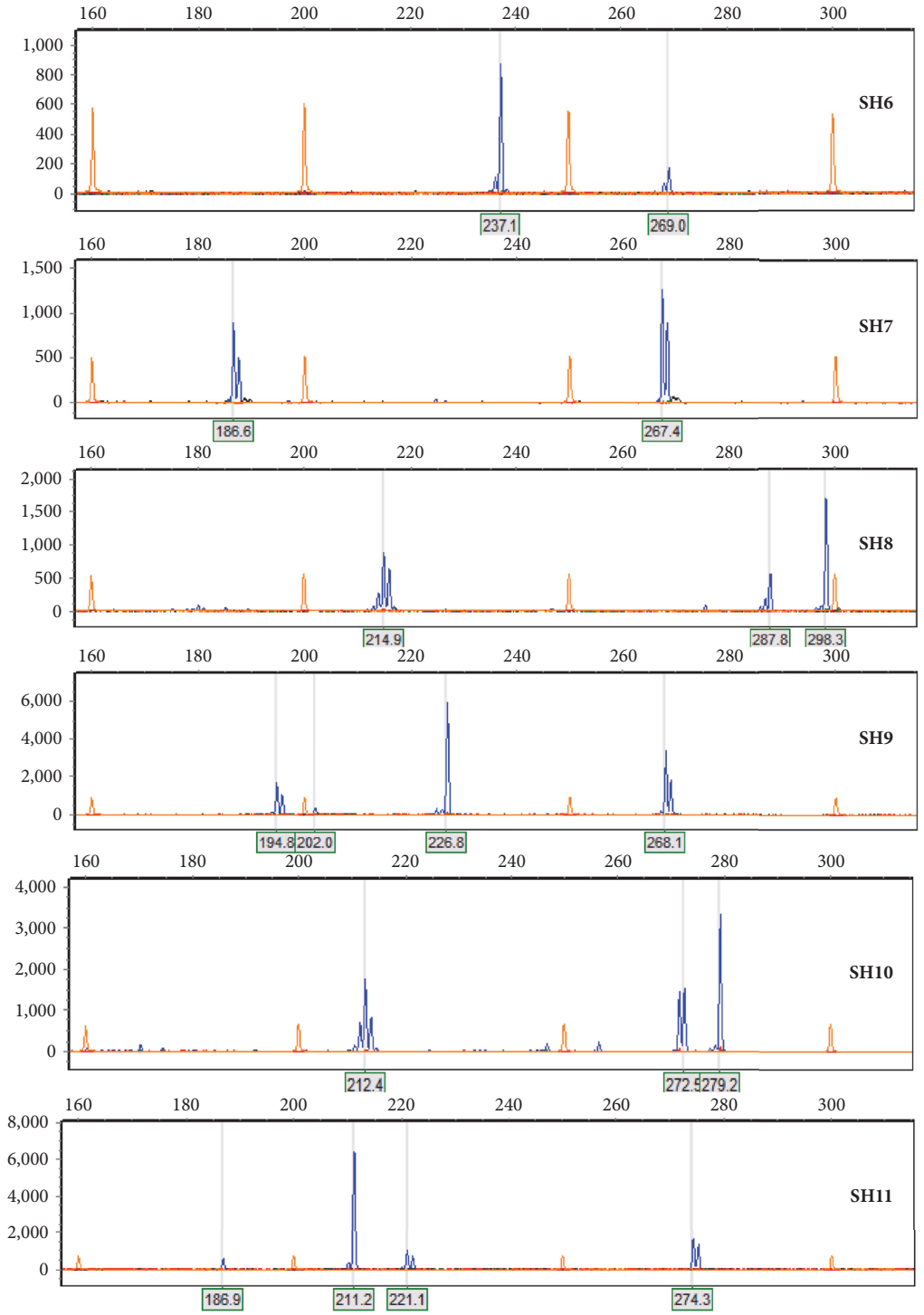


Figure 2: Electropherograms showing capillary electrophoresis separation of the PCR product fragments amplified from ecotypes with the three primers. The horizontal axis displays the size of the detected PCR product fragments, while the vertical axis presents the intensity of the signal (i.e., the indicator of concentration of fragments in the PCR products). The orange peaks match the standard fragments in the GeneScan 500 LIZ size standard, while the blue ones represent the PCR products fragments amplified from different ecotypes. The numbers on the horizontal axis represent the size of the corresponding peak in the GeneScan 500 LIZ size standard (orange). "LH" represents *Landoltia punctata* haplotype number and "SH" represents *Spirodela polyrhiza* haplotype number. Different haplotypes displayed different types of blue peaks (DNA fragments) and combinations.

be found and discriminated. The geographic differentiation showed significant influence on the genetic haplotypes in our study. For instance, haplotype 7 was found in Australia and pacific area in ecotypes of *S. polyrhiza*. Meanwhile, several haplotypes were detected worldwide. Haplotype 2 was in ecotypes of *L. punctata* found around the world with high presence. These two findings suggested that both the geographic differentiation and conservation were the feature of duckweed.

In conclusion, we designed new markers which could serve as a universal marker for inter- or intraspecies-level identification of *S. polyrhiza* and *L. punctata*. These markers combined with the capillary electrophoresis will significantly lower the cost and improve the efficiency of duckweed distinguishing especially at ecotype level. Thus, many new ecotypes with different physiological properties could be screened. Therefore, this new marker system is a significant contribution to the identification of duckweed.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Key Technology R&D Program of China (no. 2015BAD15B01), the Projects of International Cooperation of Ministry of Science and Technology of China (no. 2014DFA30680), and Science and Technology Service Network Initiative (no. KFJ-EW-STS-121).

References

[1] E. Landolt, *The Family of Lemnaceae—A Monographic Study, Vol.1*, Veroff Geobot Inst ETH, Stiftung Rubel, Zurich, 1986.

[2] D. H. Les, D. J. Crawford, E. Landolt, J. D. Gabel, and R. T. Kimball, "Phylogeny and systematics of Lemnaceae, the duckweed family," *Systematic Botany*, vol. 27, no. 2, pp. 221–240, 2002.

[3] G. D. Lemon and U. Posluszny, "Comparative shoot development and evolution in the Lemnaceae," *International Journal of Plant Sciences*, vol. 161, no. 5, pp. 733–748, 2000.

[4] E. Landolt, "The family of Lemnaceae monographic study," *Veroeffentlichungen des Geobotanischen Institutes der ETH, Stiftung Rubel, Zurich*, vol. 71, no. 1, pp. 15–71, 1986.

[5] G. Oron and H. Willers, "Effect of wastes quality on treatment efficiency with duckweed," *Water Science and Technology*, vol. 21, no. 6-7, pp. 639–645, 1989.

[6] J. E. Vermaat and M. K. Hanif, "Performance of common duckweed species (*Lemnaceae*) and the waterfern *Azolla filiculoides* on different types of waste water," *Water Research*, vol. 32, no. 9, pp. 2569–2576, 1998.

[7] D. Porath, B. Hepher, and A. Koton, "Duckweed as an aquatic crop: evaluation of clones for aquaculture," *Aquatic Botany*, vol. 7, pp. 273–278, 1979.

[8] J. J. Cheng and A.-M. Stomp, "Growing Duckweed to recover nutrients from wastewaters and for production of fuel ethanol and animal feed," *Clean—Soil, Air, Water*, vol. 37, no. 1, pp. 17–26, 2009.

[9] W. H. Cui, J. L. Xu, J. J. Cheng, and A. M. Stomp, "Growing duckweed for bioethanol production," An ASABE Meeting Presentation, 2010.

[10] R. Vunsh, J. Li, U. Hanania et al., "High expression of transgene protein in Spirodela," *Plant Cell Reports*, vol. 26, no. 9, pp. 1511–1519, 2007.

[11] M. Edelman, R. Vunsh, J. Li, U. Hanania, M. Flaishman, and A. Perl, "Transgenic Spirodela: a unique, low-risk, plant biotechnology system," in *Plant Biology 2003, Section: Biotech Risk Assessment. Proceedings of the Annual Meeting of the American Society of Plant Biologists, Honululu, Hawaii*, abstract 901, pp. 25–30, 2003.

[12] S. Rival, J.-P. Wisniewski, A. Langlais et al., "Spirodela (duckweed) as an alternative production system for pharmaceuticals: a case study, aprotinin," *Transgenic Research*, vol. 17, no. 4, pp. 503–513, 2008.

[13] D. Vaughan and R. G. Baker, "Influence of nutrients on the development of gibbosity in fronds of the duckweed *Lemna gibba* L," *Journal of Experimental Botany*, vol. 45, no. 270, pp. 129–133, 1994.

[14] D. H. Les, E. Landolt, and D. J. Crawford, "Systematics of the Lemnaceae (duckweeds): inferences from micromolecular and morphological data," *Plant Systematics and Evolution*, vol. 204, no. 3-4, pp. 161–177, 1997.

[15] D. J. Crawford, E. Landolt, D. H. Les, and R. T. Kimball, "Allozyme studies in *Lemnaceae*: variation and relationships in *Lemna* sections *Alatae* and *Biformes*," *Taxon*, vol. 50, no. 4, pp. 987–999, 2001.

[16] D. J. Crawford, E. Landolt, D. H. Les, J. K. Archibald, and R. T. Kimball, "Allozyme variation within and divergence between *Lemna gibba* and *L. disperma*: systematic and biogeographic implications," *Aquatic Botany*, vol. 83, no. 2, pp. 119–128, 2005.

[17] E. V. Martirosyan, N. N. Ryzhova, K. G. Skryabin, and E. Z. Kochieva, "RAPD analysis of genome polymorphism in the family Lemnaceae," *Russian Journal of Genetics*, vol. 44, no. 3, pp. 360–364, 2008.

[18] W. Wang, Y. Wu, Y. Yan, M. Ermakova, R. Kerstetter, and J. Messing, "DNA barcoding of the *Lemnaceae*, a family of aquatic monocots," *BMC Plant Biology*, vol. 10, pp. 205–216, 2010.

[19] http://www.internationallemnaassociation.org/uploads/Intl_Duckweed_Committee_Letter_No._1.pdf.

[20] H. Xue, Y. Xiao, Y. Jin et al., "Genetic diversity and geographic differentiation analysis of duckweed using inter-simple sequence repeat markers," *Molecular Biology Reports*, vol. 39, no. 1, pp. 547–554, 2012.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 5078796, 10 pages
http://dx.doi.org/10.1155/2016/5078796

Review Article

The Power of CRISPR-Cas9-Induced Genome Editing to Speed Up Plant Breeding

Hieu X. Cao,[1] Wenqin Wang,[2] Hien T. T. Le,[3] and Giang T. H. Vu[1]

[1]*Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, Gatersleben, 06466 Stadt Seeland, Germany*
[2]*School of Agriculture and Biology, Shanghai Jiaotong University, 800 Dong Chuan Road, Shanghai 200240, China*
[3]*Institute of Genome Research, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay, Hanoi, Vietnam*

Correspondence should be addressed to Hieu X. Cao; cao@ipk-gatersleben.de

Received 10 August 2016; Revised 17 October 2016; Accepted 1 November 2016

Academic Editor: Graziano Pesole

Genome editing with engineered nucleases enabling site-directed sequence modifications bears a great potential for advanced plant breeding and crop protection. Remarkably, the RNA-guided endonuclease technology (RGEN) based on the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) is an extremely powerful and easy tool that revolutionizes both basic research and plant breeding. Here, we review the major technical advances and recent applications of the CRISPR-Cas9 system for manipulation of model and crop plant genomes. We also discuss the future prospects of this technology in molecular plant breeding.

1. Introduction

Under pressure of rapid population growth, climate change, and agricultural pests and diseases, the next green evolution with new technologies is required to address and provide novel genetic variations to improve yield, quality, and resistance against biotic and abiotic stresses in crop plants. During the last two decades several crop genomes have been altered by introduction of one or more foreign genes of high agronomic values to overcome the limitations of conventional breeding techniques. Promises as well as critics on such genetically modified crops have been discussed intensively in other reviews (e.g., [1–3]) and are beyond the scope of this paper. Alternatively and more powerful, genome editing allows precise and predictable changes to be made to the crop genetic materials and currently revolutionizes crop breeding (e.g., [4–7]).

Genome editing with site-specific nucleases introduces DNA double-strand breaks (DSBs) at a target site, stimulating cellular DNA repair mechanisms and subsequently resulting in various types of genome modifications such as targeted mutagenesis, gene insertion, or gene replacement. The two main DSB repair pathways in eukaryotic cells are nonhomologous end-joining (NHEJ) and homologous recombination (HR). NHEJ often can cause insertions or deletions, potentially producing a gene knockout. When repair templates (regions of homology to the sequence surrounding the DSB) are available, the HR machinery can be recruited to achieve precise modifications (homology-directed repair, HDR), such as gene replacement or gene insertion (Figure 1). Apparently, NHEJ is the most commonly employed DSB repair mechanism in many organisms, including higher plants [8, 9].

The latest ground-breaking technology for genome editing is the CRISPR-Cas system which was inspired by the bacterial adaptive immunity against invading bacteriophages. In August 2012, the groups of Jennifer A. Doudna at the University of California, Berkeley, and Emmanuelle Charpentier at the Umea University in Sweden (now at the Max Planck Institute of Infection Biology in Berlin) [10] showed for the first time that a monomeric DNA endonuclease, known as Cas9, from *Streptococcus pyogenes* can be easily programmed to cut double-stranded DNA at a specific genomic sequence using complementary base pairing of a single-guide RNA (sgRNA, Figure 1). The potential to exploit this simple system for genome editing in eukaryotic systems (human,

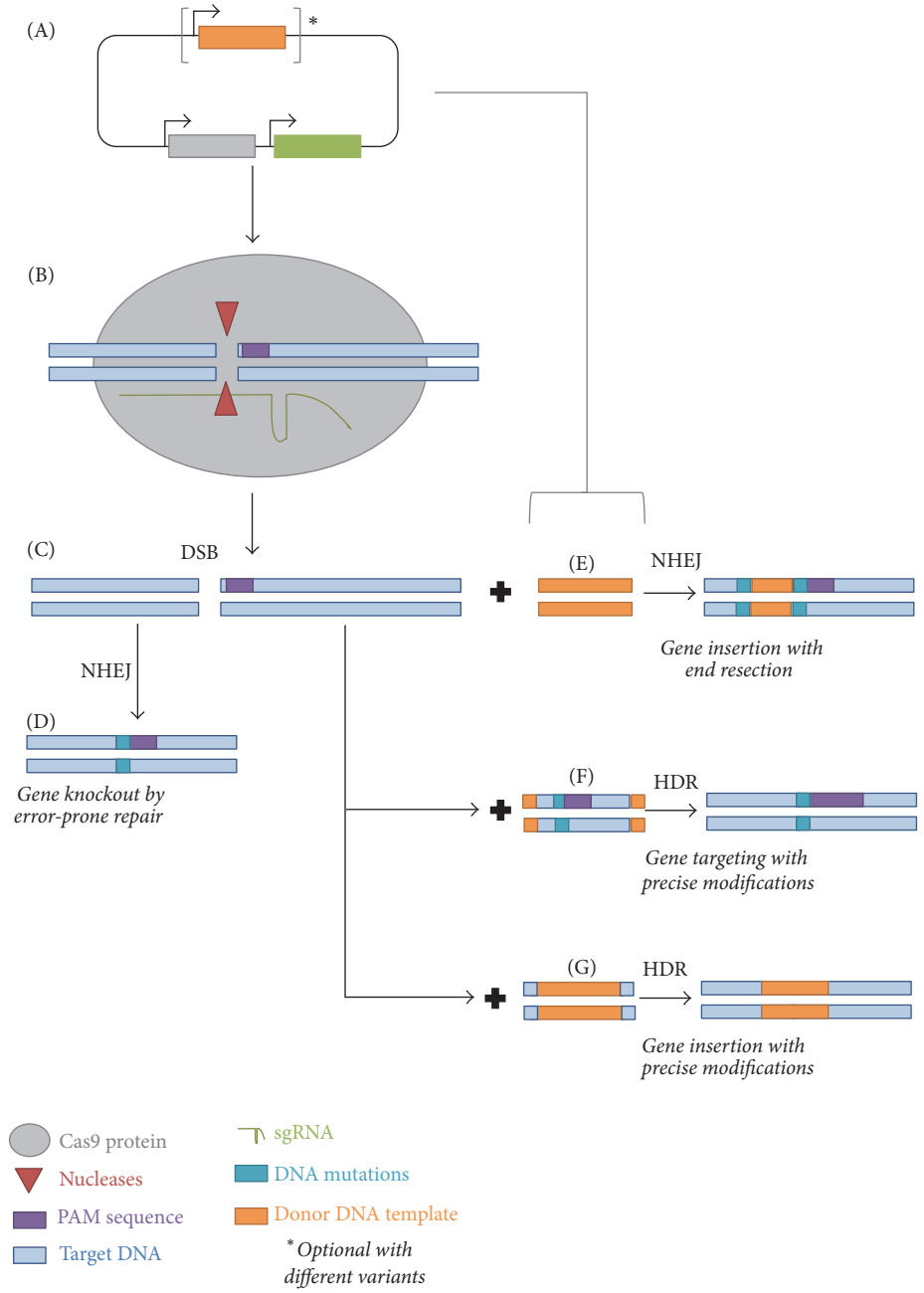


Figure 1: Overview of CRIPR-Cas9 technology for plant genome editing. (A) The most widely used engineered CRISPR-Cas9 system in plants utilizes a plant-codon-optimized Cas9 protein and (could be more than one) single-guide RNA (sgRNA). Optionally, the gene targeting system with geminivirus replicons includes an additional donor DNA template. (B) In plant cells, sgRNA associated with Cas9 nuclease mediates cleavage of target DNA sites that are complementary to the sgRNA and locate next to a PAM sequence. (C) Cas9-induced double-strand DNA breaks (DSBs) can be repaired by nonhomologous end-joining (NHEJ) or homology-directed repair (HDR) pathways. (D) Imprecise NHEJ-mediated repair can generate insertion and/or deletion mutations with variable length at the site of the DSB. These InDels can cause out-of-frame mutations in the coding sequences of the target genes, resulting in gene knockout. (E) In the presence of donor DNA, NHEJ can insert the donor DNA into the site of the DSB together with possibly additional InDel mutations. HR-driven repair can produce precise modifications, including point mutations (F) or insertions from double-/single-strand DNAs as donor templates (G).

mouse) was demonstrated few months later by the work of Feng Zhang's group at Massachusetts Institute of Technology (MIT) [11] and George Church's group at Harvard University [12]. In these studies, only a single construct for expression of Cas9 nuclease and a specifically designed sgRNA are needed for transformation. Since then, due to its ease of implementation and robustness the CRISPR-Cas9 system has been utilized widely for genome engineering in various organisms, including plants [13–15], insects [16], fish [17], rabbits [18], pigs [19], mice [20], monkeys [21], and human cells

[22, 23]. A large number of publications using the CRISPR-Cas9 technology came up rapidly since the first reports and promoted our understanding and applications of the system. Here we review the major advances in plant genome editing technology using Cas9 RNA-guided endonuclease (RGEN) and discuss its applications as well as future prospects in molecular plant breeding.

2. Overview of the Genome Editing CRISPR-Cas9 System

The immense versatility of the CRISPR-Cas9 technology in the field of genome editing is due to its simplicity, efficiency, and robustness. Basically, the CRISPR-Cas9 tool consists of two main components, deliverable as a single plasmid (Figure 1(A)): a bacterial Cas9 endonuclease protein and a specifically designed sgRNA containing a 20-bp sequence homologous to the target DNA (called protospacer). A prerequisite for cleavage of the target DNA is the presence of a sequence 5′-NGG-3′ [10] or 5′-NAG-3′ [24] as the conserved protospacer-adjacent motif (PAM). Importantly, it has been shown that multiple sgRNA targeting to different genomic loci can be simultaneously exploited to achieve high-efficiency multiplex genome engineering without requiring additional Cas9 proteins [11, 12]. Moreover, some initial in vitro and in vivo evidence suggested that Cas9 endonuclease activity is not affected by DNA CpG methylation [24]. However, sgRNA preferentially binds to open chromatin regions including off-target sites [25, 26]. Further efforts to improve our understanding of the effect of chromatin accessibility and epigenetic environment at the target site on the efficiency of the CRISPR-Cas9 system are needed.

The main concern regarding the implementation of the CRISPR-Cas9 system for genome editing is occasional off-target modifications reported in some studies [11, 12, 24, 27–29]. Although a 20 bp recognition sequence in the sgRNA was initially considered necessary to determine specificity, it was later shown that a perfect match between the 7–12 bp at the 3′ end of the sgRNA (called the seed region) and the equivalent region of the target DNA confers target site recognition and cleavage, whereas multiple mismatches in the PAM-distal region are generally tolerated [24, 27–29]. Several strategies have been developed to control the specificity of CRISPR-Cas9, in which the design of the sgRNA is considered as an important and easily implementable one. A number of guidelines and online tools have been developed to facilitate the selection of unique target sites in organisms for which high quality whole genome sequences are available [24, 30, 31]. Truncated sgRNA with length of 17 bp or elongated sgRNA with 2 additional guanidine residues at the 5′ end could reduce nontarget mutations [32, 33]. Low expression level of Cas9 nuclease is another way to reduce off-target activities [29, 34].

The most widely used Cas9 nuclease originates from the type II (class 2) CRISPR-Cas9 system of *Streptococcus pyogenes* (SpCas9). However, Cas9 orthologues from other bacterial species are also applicable and may offer further optimization of the current CRIPSR-Cas9 system. For instance, Cas9 gene of *Staphylococcus aureus* (SaCas9), which is 1 kb shorter than that of *S. pyogenes,* could improve its stability in transformation vectors [35]. Interestingly, SaCas9 targets another distinct PAM 5′NNGGGT3′. On the other hand, a new endonuclease of the class 2 CRISPR-Cas systems, Cpf1 (CRISPR from *Prevotella* and *Francisella* 1), has been reported to require a T-rich PAM motif upstream of the target site and generates a DSB with 5′ overhangs [36]. Finding of further Cas9 nucleases which require different PAM motifs might allow targeting of more diverse genomic positions and enable harnessing more complex applications for genome engineering by using combination of these Cas9.

Detailed understanding of the molecular structure of the CRISPR-Cas9 systems guides us to rationally redesign and customize variants of Cas9 enzymes. Crystal structures of SpCas9, SaCas9, *Francisella novicida* Cas9 (FnCas9), or Cpf1 in complex with their sgRNA and double-stranded target DNAs [37–41] were solved, revealing distinct mechanisms of PAM recognition and of RNA-guided DNA targeting by Cas9 nucleases. Several engineered CRISPR-Cas9 variants were produced to increase Cas9 specificity or to alter the PAM recognition patterns [42]. Cas9 nickase variants (Cas9-D10A or Cas9-H840A) containing a single inactive nuclease domain cleave only one DNA strand to create a single-strand break at the target sites. A pair of induced nicks, one on each strand and up to 100 bp apart from each other, can result in a DSB with overhang. This approach using Cas9 nickase could significantly reduce the off-target mutation rate [43, 44]. In addition, fusion of catalytically inactive Cas9 (Cas9-D10A-H840A, dCas9) and FokI nuclease, which functions only as a dimer, showed comparable results when they are guided by a pair of sgRNAs [45, 46]. Interestingly, dCas9 could be exploited not only in genome editing but also in many other applications, such as modifications of gene expression [47], epigenetic editing [48], and visualization of specific DNA sequences in living cells [49].

3. Major Advances of Plant and Crop Genome Editing Technology Using Cas9 RNA-Guided Endonucleases (RGENs)

The CRISPR-Cas9 system with the ability to precisely cut DNA of essentially any organism provides an unprecedented tool for genomic engineering. Soon after the evidence that the CRISPR-Cas9 system works in animal models, three papers reported expression and activities of the plant-codon-optimized CRISPR-Cas9 in plant model species of *Arabidopsis* (*Arabidopsis thaliana*) and tobacco (*Nicotiana benthamiana*) as well as in crops such as rice and wheat [50–52]. Those first groups demonstrated the versatility of the technology by using different transient or stable transformation platforms (protoplast transfection, leaf agroinfiltration, and particle bombardment of callus) in order to generate small deletions and/or insertions, targeted insertions, and multiplex genome modifications. Furthermore, the transmission to progeny and Mendelian heritability of CRISPR-Cas9-induced mutations was shown by using the *Agrobacterium*-mediated germ line transformation in *Arabidopsis* [14, 15, 53, 54] and rice [55–58], suggesting that the CRISPR-Cas9 system could become a powerful tool in crop genome editing. Subsequent work

reported successful applications of the CRISPR-Cas9 tool for sorghum [13], wheat [59], maize [60], sweet orange [61], tomato [62], potato [63], liverwort *Marchantia polymorpha* L. [64], barley and *Brassica oleracea* [65], soybean [66], melon [67], and poplar [68]. Summarizing information about transformation/delivery methods and expression systems for CRISPR-Cas9-based applications in plants can be found in recent reviews [69, 70]. It is needed to emphasise that spreading of this technology is highly promoted by the CRISPR research community, providing open access to plasmids, web tools, and active discussion groups (Table 1).

Since plant genomes are large, complex, and often polyploid, off-target mutations can be expected to happen during genome engineering. When introducing a gRNA-Cas9 cleavage in rice, Xie and Yang [71] reported a mutation rate of 1.6% at a single off-target sequence which has a single mismatch at position 15 bp proximal to the PAM. Off-target effects of CRISPR-Cas9 have been observed in other plant species, including soybean [72], maize [73], and barley and *B. oleracea* [65]. In contrast, no off-target mutation events could be detected at the putative off-target sites in studies on *Arabidopsis*, tobacco, wheat, rice, or sweet orange [14, 50–52, 58, 59, 61], even using whole-genome sequencing [14]. Notably, off-target events are less problematic in plant breeding than for clinical research because off-target mutations can be segregated away from the mutation of the target by crossing mutants with wild-type plants. However, the crossing procedure can be laborious, time-consuming, or even impossible for perennial plants and vegetatively propagated crops, such as potatoes, bananas, and cassava. The off-target problem can be overcome either by optimizing sgRNA design [74, 75] or by using high fidelity CRISPR-Cas9 approaches with more precise Cas9 variants (e.g., [76]), paired Cas9 nickases, or dCas9:FokI fusions. The application of the Cas9-D10A nickase in *Arabidopsis* suggests off-target effects might be avoided by using a pair of nickases [15, 77].

The primary application of CRISPR-Cas9 technology in genome editing (or reverse genetics studies) is gene knockout because the host cells preferentially repair the Cas9-induced DSBs via the "error-prone" NHEJ pathway which often results in short insertions or deletions. The size of these modifications and the ratio between insertions and deletions could have an impact on genome size and direct genome size evolution [78, 79]. The flexibility of the CRISPR-Cas9 tool enables targeting of adjacent sites in chromosome for specific removal of a large unwanted DNA sequences, from several kb in *Arabidopsis* [52, 53], tobacco [30], and tomato [62] up to 245 kb in rice [58]. CRISPR-Cas9 can also be utilized to knockout multiple genes of gene family in rice [80] or homoeologous genes in hexaploid bread wheat [81, 82]. A CRISPR-Cas9 toolbox for multiplexed genome editing was demonstrated in model plants such as *Arabidopsis*, tobacco, and rice [83]. In order to perform highly multiplexing genome manipulations, CRISPR-Cas systems using Cas9 orthologues of *Staphylococcus aureus* (SaCas9) and *Streptococcus thermophilus* (St1Cas9) have been adapted for use in *Arabidopsis* [84]. In addition, modified Cas9 variants enable targeting to noncanonical PAM sites in rice [85], providing a wider range of genome editing.

For site-specific gene insertion ("trait stacking") or replacement it is needed to exploit the HR pathway for repairing the Cas9-induced DSB. HDR events require as template a sequence homologous to the target gene (Figure 1, [8]). However, HDR frequency in CRISPR-Cas9-mediated gene targeting (GT) is rather low as shown in rice [50], in tobacco [52], and in *Arabidopsis* [53, 55]. By using approximately 670 bp homology on either side of the break, Schiml et al. [77] could insert a 1.8 kb marker gene into an endogenous gene of *Arabidopsis* with a frequency of 0.14%. In order to overcome the low HDR efficiency of targeted genome manipulation in mammalian cells, components of the NHEJ pathway were inhibited [86]. Similarly, by manipulating of DNA ligase IV (a member of the NHEJ pathway), CRISPR-Cas9-induced HDR-mediated GT can work more efficiently in rice, resulting in biallelic plants [87]. Alternatively, rational design of orientation, polarity, and length of the donor ssDNA to match the properties of the Cas9-DNA complex could increase the HDR events [88]. Ideally, the efficiency of HDR-mediated genome modifications would be improved by delivery of sufficient quantities of the donor sequence for HDR repair at the Cas9-targeted site. A transformation method using a nuclear replicating DNA virus [89] which produces multiple copies of the donor sequences for HDR inside plant cells has recently been demonstrated to generate high-frequency, precise genome modifications in tomato [90].

4. Future Perspectives of the CRISPR-CAS Technology for Plant Breeding

The CRISPR-Cas9 technology is revolutionizing genome engineering and equipping scientists and breeders with the ability to precisely modify the DNA of crop plants. Importantly, CRISPR-Cas9 enables genome modifications also in potential crop plants for which genetic manipulation has been a challenge (e.g., duckweed [91]), provided that high quality whole genome sequences [92, 93] and efficient transformation procedures are available [94]. This review does not cover ethical, legal, and social issues of this revolutionary tool (for such aspects, see [2–4, 95]). In this context, it is necessary to note that the common white button mushroom (*Agaricus bisporus*) that has been modified to resist browning using CRISPR-Cas9 became the first CRISPR-edited organism that can be cultivated and sold without further oversight of US regulations [96]. Interestingly, the CRISPR-Cas9 approach offers genetic manipulation of crops without transgenic footprints by delivering preassembled Cas9-sgRNA ribonucleoproteins [97] or by transient expression of the in vitro transcripts of Cas9-coding sequence and sgRNA [82] and thus might not be classified as genetically modified organisms and regulated by existing biosafety regulations.

A major power of CRISPR/Cas9-induced genome editing is to provide an opportunity for targeting multiple sites simultaneously. New application of this technology is conferring multiple pathogen resistances to crop plants. By establishing CRISPR-Cas9-like immune systems, tobacco and *Arabidopsis* were made resistant to the beet severe curly top virus [98], the bean yellow dwarf virus [99], and the tomato yellow leaf curl virus, respectively [100]. The recently developed CRISPR-Cas

Table 1: Useful resources/tools for CRISPR-Cas9 research in plants.

Site	Purpose	Authority
http://www.addgene.org	Access to plasmid resource and tutorial documents [115]	Addgene
https://www.protocols.io	Access to detailed protocol resource	Protocols.io
http://cbi.hzau.edu.cn/cgi-bin/CRISPR	Design optimal sgRNA with 43 plant genomes from Ensembl Plants [24]	Huazhong Agricultural University
http://www.genome.arizona.edu/crispr/CRISPRsearch.html	Predict high specific sgRNA of 8 plant genomes [116]	University of Arizona
http://www.rgenome.net/cas-offinder/	Search for potential off-target sites of sgRNA from 37 plant genomes [117]	Institute for Basic Science, Korea
http://www.e-crisp.org/E-CRISP/index.html	Design sgRNA for genome-libraries projects or individual sequences with 11 plant genomes [118]	German Cancer Research Center (DKFZ)
http://crispr.mit.edu/	Find the CRISPR-Cas9 target sites within an input sequence with *Arabidopsis* genome [24]	Zhang Lab, MIT
http://chopchop.cbu.uib.no/	Select CRISPR target sites and predict off-target sites with *Arabidopsis* genome [119]	University of Bergen
http://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design	Design highly active sgRNAs for the provided targets [74]	Broad Institute of MIT and Harvard, Cambridge University
http://eendb.zfgenetics.org/casot/	Open-sourced tool for finding potential off-target sites of any user-specific genome [120]	Peking University
https://groups.google.com/forum/#!forum/crispr	Active discussion groups	Google

system with programmable RNA recognition and cleavage [101] would be exciting to apply in plants because the majority of plant viruses are RNA viruses [102]. However, further studies will be required to monitor the stability of such resistances over generations and in diverse habitats [103].

CRISPR-Cas9 has triggered innovative applications in several fields, including agriculture. CRISPR-edited individual organisms could spread a positively selectable gene throughout a wild population in a so-called gene drive process. In principle, such CRISPR-based gene drive systems could be beneficial to mankind, for example, by potentially preventing the spread of diseases, or supporting agriculture by reversing pesticide and herbicide resistance in insects and weeds, and control damaging invasive species [104]. Gene drives will work only in sexually reproducing species and spread significantly only in species that reproduce quickly. The gene drive model has been tested in yeast [105] and the first CRISPR-Cas9-engineered mosquitoes have recently developed to fight malaria [106]. However, because of low efficient homologous recombination, gene drive application to either eliminate or reduce invasive plant species in a given area is still challenging. In addition, since such technology might pose tremendous alterations to wild populations, biosafety precautions and measures are needed (for more details, see [107, 108]). Importantly, the CRISPR-Cas9-mediated gene drive technology (called as mutagenic chain reaction (MCR) [109]) can be used to produce stable homozygous (biallelic) mutant lines by using HDR-driven propagation of the CRISPR-Cas9-cassette to the companion chromosome. Moreover, such concept can also be applied for editing organelle genomes (e.g., chloroplast) in order to overcome the high copy number of genomes and reversion of mutations [110].

Until now synthetic biology is limited to bacteria models to engineer completely new metabolic pathways. The CRISPR-Cas9 technology opens the way to an easier use of synthetic biology in more complex systems, for example, for agronomical traits in crop plants [111]. Since many complex metabolic pathways in plants interact with each other and are controlled by multiple tissue- or development-specific regulators, metabolic engineering in plants could require not only multiple gene targeting but also probably fine-tuning multiple gene expression level at different tissues or developmental stages. For such sophisticated applications, modifications or customizations of the CRISPR-Cas9 systems including (1) specific Cas9-/sgRNA expression promoters (e.g., [112, 113]), (2) modified Cas9 for alterations of gene expression and epigenetic changes (e.g., [47, 48]), (3) combinations of different Cas9 variants (e.g., [35, 36]) for expanding the target range in the genomes, and (4) efficient technology for increasing HDR-driven precise gene replacement will be needed to be further developed or optimized for particular cell types or organisms. With the rapid development of CRISPR-Cas9 technology during the last 4 years, the promise of a next green revolution with new crops meeting long-standing requests for metabolic engineering (e.g., plants that can fix their own nitrogen, have better nutritious values, can be efficiently utilized for biofuel production, or display enhanced photosynthetic capacity [114]) could be realized in the near future.

Abbreviations

CRISPR: Clustered regularly interspaced short palindromic repeats
Cas9: CRISPR-associated protein 9
DSB: DNA double-strand break
GT: Gene targeting
HR: Homologous recombination
NHEJ: Nonhomologous end-joining
sgRNA: Single-guide RNA
PAM: Protospacer-adjacent motif
RGEN: RNA-guided endonuclease.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors thank Ingo Schubert from the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) for critical reading of the manuscript.

References

[1] H. Chen and Y. Lin, "Promise and issues of genetically modified crops," *Current Opinion in Plant Biology*, vol. 16, no. 2, pp. 255–260, 2013.

[2] R. R. Aldemita, I. M. Reaño, R. O. Solis, and R. A. Hautea, "Trends in global approvals of biotech crops (1992–2014)," *GM Crops & Food*, vol. 6, no. 3, pp. 150–166, 2015.

[3] M. Araki and T. Ishii, "Towards social acceptance of plant breeding by genome editing," *Trends in Plant Science*, vol. 20, no. 3, pp. 145–149, 2015.

[4] S. Huang, D. Weigel, R. N. Beachy, and J. Li, "A proposed regulatory framework for genome-edited crops," *Nature Genetics*, vol. 48, no. 2, pp. 109–111, 2016.

[5] S. Schiml and H. Puchta, "Revolutionizing plant biology: multiple ways of genome engineering by CRISPR/Cas," *Plant Methods*, vol. 12, article 8, 2016.

[6] K. Belhaj, A. Chaparro-Garcia, S. Kamoun, N. J. Patron, and V. Nekrasov, "Editing plant genomes with CRISPR/Cas9," *Current Opinion in Biotechnology*, vol. 32, pp. 76–84, 2015.

[7] D. Carroll, "Genome engineering with targetable nucleases," *Annual Review of Biochemistry*, vol. 83, pp. 409–439, 2014.

[8] G. T. H. Vu, H. X. Cao, K. Watanabe et al., "Repair of site-specific DNA double-strand breaks in barley occurs via diverse pathways primarily involving the sister chromatid," *Plant Cell*, vol. 26, no. 5, pp. 2156–2167, 2014.

[9] H. Puchta, "The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution," *Journal of Experimental Botany*, vol. 56, no. 409, pp. 1–14, 2005.

[10] 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.

[11] L. Cong, F. A. Ran, D. Cox et al., "Multiplex genome engineering using CRISPR/Cas systems," *Science*, vol. 339, no. 6121, pp. 819–823, 2013.

[12] P. Mali, L. Yang, K. M. Esvelt et al., "RNA-guided human genome engineering via Cas9," *Science*, vol. 339, no. 6121, pp. 823–826, 2013.

[13] W. Jiang, H. Zhou, H. Bi, M. Fromm, B. Yang, and D. P. Weeks, "Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice," *Nucleic Acids Research*, vol. 41, no. 20, article e188, 2013.

[14] Z. Feng, Y. Mao, N. Xu et al., "Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 12, pp. 4632–4637, 2014.

[15] F. Fauser, S. Schiml, and H. Puchta, "Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*," *Plant Journal*, vol. 79, no. 2, pp. 348–359, 2014.

[16] A. R. Bassett, C. Tibbit, C. P. Ponting, and J.-L. Liu, "Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System," *Cell Reports*, vol. 4, no. 1, pp. 220–228, 2013.

[17] W. Y. Hwang, Y. Fu, D. Reyon et al., "Efficient genome editing in zebrafish using a CRISPR-Cas system," *Nature Biotechnology*, vol. 31, no. 3, pp. 227–229, 2013.

[18] A. Honda, M. Hirose, T. Sankai et al., "Single-step generation of rabbits carrying a targeted allele of the tyrosinase gene using CRISPR/Cas9," *Experimental Animals*, vol. 64, no. 1, pp. 31–37, 2015.

[19] K. M. Whitworth, K. Lee, J. A. Benne et al., "Use of the CRISPR/Cas9 system to produce genetically engineered pigs from in vitro-derived oocytes and embryos," *Biology of Reproduction*, vol. 91, no. 3, article no. 78, 2014.

[20] D. Seruggia, A. Fernández, M. Cantero, P. Pelczar, and L. Montoliu, "Functional validation of mouse tyrosinase non-coding regulatory DNA elements by CRISPR-Cas9-mediated mutagenesis," *Nucleic Acids Research*, vol. 43, no. 10, pp. 4855–4867, 2015.

[21] Y. Niu, B. Shen, Y. Cui et al., "Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos," *Cell*, vol. 156, no. 4, pp. 836–843, 2014.

[22] S. W. Cho, S. Kim, J. M. Kim, and J.-S. Kim, "Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease," *Nature Biotechnology*, vol. 31, no. 3, pp. 230–232, 2013.

[23] M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna, "RNA-programmed genome editing in human cells," *Elife*, vol. 2013, no. 2, Article ID e00471, 2013.

[24] P. D. Hsu, D. A. Scott, J. A. Weinstein et al., "DNA targeting specificity of RNA-guided Cas9 nucleases," *Nature Biotechnology*, vol. 31, no. 9, pp. 827–832, 2013.

[25] X. Wu, D. A. Scott, A. J. Kriz et al., "Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells," *Nature Biotechnology*, vol. 32, no. 7, pp. 670–676, 2014.

[26] C. Kuscu, S. Arslan, R. Singh, J. Thorpe, and M. Adli, "Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease," *Nature Biotechnology*, vol. 32, no. 7, pp. 677–683, 2014.

[27] Y. Fu, J. A. Foden, C. Khayter et al., "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells," *Nature Biotechnology*, vol. 31, no. 9, pp. 822–826, 2013.

[28] W. Y. Jiang, D. Bikard, D. Cox, F. Zhang, and L. A. Marraffini, "RNA-guided editing of bacterial genomes using CRISPR-Cas systems," *Nature Biotechnology*, vol. 31, no. 3, pp. 233–239, 2013.

[29] V. Pattanayak, S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna, and D. R. Liu, "High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity," *Nature Biotechnology*, vol. 31, no. 9, pp. 839–843, 2013.

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

[31] J. G. Doench, E. Hartenian, D. B. Graham et al., "Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation," *Nature Biotechnology*, vol. 32, no. 12, pp. 1262–1267, 2014.

[32] Y. F. Fu, J. D. Sander, D. Reyon, V. M. Cascio, and J. K. Joung, "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs," *Nature Biotechnology*, vol. 32, no. 3, pp. 279–284, 2014.

[33] S. W. Cho, S. Kim, Y. Kim et al., "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases," *Genome Research*, vol. 24, no. 1, pp. 132–141, 2014.

[34] W. Fujii, K. Kawasaki, K. Sugiura, and K. Naito, "Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease," *Nucleic Acids Research*, vol. 41, no. 20, article e187, 2013.

[35] 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.

[36] B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh et al., "Cpf1 Is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system," *Cell*, vol. 163, no. 3, pp. 759–771, 2015.

[37] H. Nishimasu, F. A. Ran, P. D. Hsu et al., "Crystal structure of Cas9 in complex with guide RNA and target DNA," *Cell*, vol. 156, no. 5, pp. 935–949, 2014.

[38] H. Nishimasu, L. Cong, W. X. Yan et al., "Crystal Structure of *Staphylococcus aureus* Cas9," *Cell*, vol. 162, no. 5, pp. 1113–1126, 2015.

[39] C. Anders, O. Niewoehner, A. Duerst, and M. Jinek, "Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease," *Nature*, vol. 513, no. 7519, pp. 569–573, 2014.

[40] T. Yamano, H. Nishimasu, B. Zetsche et al., "Crystal structure of Cpf1 in complex with guide RNA and target DNA," *Cell*, vol. 165, no. 4, pp. 949–962, 2016.

[41] H. Hirano, J. Gootenberg, T. Horii et al., "Structure and Engineering of Francisella novicida Cas9," *Cell*, vol. 164, no. 5, pp. 950–961, 2016.

[42] S. Hirano, H. Nishimasu, R. Ishitani, and O. Nureki, "Structural basis for the altered PAM specificities of engineered CRISPR-Cas9," *Molecular Cell*, vol. 61, no. 6, pp. 886–894, 2016.

[43] F. A. Ran, P. D. Hsu, C.-Y. Lin et al., "Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity," *Cell*, vol. 154, no. 6, pp. 1380–1389, 2013.

[44] B. Shen, W. Zhang, J. Zhang et al., "Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects," *Nature Methods*, vol. 11, no. 4, pp. 399–402, 2014.

[45] J. P. Guilinger, D. B. Thompson, and D. R. Liu, "Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification," *Nature Biotechnology*, vol. 32, no. 6, pp. 577–582, 2014.

[46] S. Q. Tsai, N. Wyvekens, C. Khayter et al., "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing," *Nature Biotechnology*, vol. 32, no. 6, pp. 569–576, 2014.

[47] L. S. Qi, M. H. Larson, L. A. Gilbert et al., "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression," *Cell*, vol. 152, no. 5, pp. 1173–1183, 2013.

[48] I. B. Hilton, A. M. D'Ippolito, C. M. Vockley et al., "Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers," *Nature Biotechnology*, vol. 33, no. 5, pp. 510–517, 2015.

[49] B. Chen, L. A. Gilbert, B. A. Cimini et al., "Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system," *Cell*, vol. 155, no. 7, pp. 1479–1491, 2013.

[50] Q. Shan, Y. Wang, J. Li et al., "Targeted genome modification of crop plants using a CRISPR-Cas system," *Nature Biotechnology*, vol. 31, no. 8, pp. 686–688, 2013.

[51] V. Nekrasov, B. Staskawicz, D. Weigel, J. D. G. Jones, and S. Kamoun, "Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease," *Nature Biotechnology*, vol. 31, no. 8, pp. 691–693, 2013.

[52] J.-F. Li, J. E. Norville, J. Aach et al., "Multiplex and homologous recombination-mediated genome editing in *Arabidopsis*and *Nicotiana benthamiana*using guide RNA and Cas9," *Nature Biotechnology*, vol. 31, no. 8, pp. 688–691, 2013.

[53] Y. Mao, H. Zhang, N. Xu, B. Zhang, F. Gou, and J.-K. Zhu, "Application of the CRISPR-Cas system for efficient genome engineering in plants," *Molecular Plant*, vol. 6, no. 6, pp. 2008–2011, 2013.

[54] W. Jiang, B. Yang, and D. P. Weeks, "Efficient CRISPR/Cas9-mediated gene editing in *Arabidopsis thaliana* and inheritance of modified genes in the T2 and T3 generations," *PLoS ONE*, vol. 9, no. 6, Article ID e99225, 2014.

[55] Z. Feng, B. Zhang, W. Ding et al., "Efficient genome editing in plants using a CRISPR/Cas system," *Cell Research*, vol. 23, no. 10, pp. 1229–1232, 2013.

[56] H. Zhang, J. Zhang, P. Wei et al., "The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation," *Plant Biotechnology Journal*, vol. 12, no. 6, pp. 797–807, 2014.

[57] J. Miao, D. S. Guo, J. Z. Zhang et al., "Targeted mutagenesis in rice using CRISPR-Cas system," *Cell Research*, vol. 23, no. 10, pp. 1233–1236, 2013.

[58] H. Zhou, B. Liu, D. P. Weeks, M. H. Spalding, and B. Yang, "Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice," *Nucleic Acids Research*, vol. 42, no. 17, pp. 10903–10914, 2014.

[59] 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.

[60] Z. Liang, K. Zhang, K. Chen, and C. Gao, "Targeted mutagenesis in *Zea mays* using TALENs and the CRISPR/Cas system," *Journal of Genetics and Genomics*, vol. 41, no. 2, pp. 63–68, 2014.

[61] H. Jia and N. Wang, "Targeted genome editing of sweet orange using Cas9/sgRNA," *PLoS ONE*, vol. 9, no. 4, Article ID e93806, 2014.

[62] C. Brooks, V. Nekrasov, Z. B. Lippman, and J. Van Eck, "Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system," *Plant Physiology*, vol. 166, no. 3, pp. 1292–1297, 2014.

[63] N. M. Butler, P. A. Atkins, D. F. Voytas, and D. S. Douches, "Generation and inheritance of targeted mutations in potato (*Solanum tuberosum* L.) using the CRISPR/Cas System," *PLoS ONE*, vol. 10, no. 12, Article ID e0144591, 2015.

[64] S. S. Sugano, M. Shirakawa, J. Takagi et al., "CRISPR/Cas9-mediated targeted mutagenesis in the liverwort *Marchantia polymorpha L*," *Plant and Cell Physiology*, vol. 55, no. 3, pp. 475–481, 2014.

[65] T. Lawrenson, O. Shorinola, N. Stacey et al., "Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease," *Genome Biology*, vol. 16, article 258, 2015.

[66] T. B. Jacobs, P. R. LaFayette, R. J. Schmitz, and W. A. Parrott, "Targeted genome modifications in soybean with CRISPR/Cas9," *BMC Biotechnology*, pp. 1–10, 2015.

[67] G. Tzuri, X. Zhou, N. Chayut et al., "A 'golden' SNP in CmOr governs the fruit flesh color of melon (*Cucumis melo*)," *Plant Journal*, vol. 82, no. 2, pp. 267–279, 2015.

[68] D. Fan, T. Liu, C. Li et al., "Efficient CRISPR/Cas9-mediated targeted mutagenesis in populus in the first generation," *Scientific Reports*, vol. 5, Article ID 12217, 2015.

[69] X. Ma, Q. Zhu, Y. Chen, and Y. G. Liu, "CRISPR/Cas9 platforms for genome editing in plants: developments and applications," *Molecular Plant*, vol. 9, no. 7, pp. 961–974, 2016.

[70] S. M. Schaeffer and P. A. Nakata, "The expanding footprint of CRISPR/Cas9 in the plant sciences," *Plant Cell Reports*, vol. 35, no. 7, pp. 1451–1468, 2016.

[71] K. Xie and Y. Yang, "RNA-guided genome editing in plants using a CRISPR-Cas system," *Molecular Plant*, vol. 6, no. 6, pp. 1975–1983, 2013.

[72] X. Sun, Z. Hu, R. Chen et al., "Targeted mutagenesis in soybean using the CRISPR-Cas9 system," *Scientific Reports*, vol. 5, article 10342, 2015.

[73] C. Feng, J. Yuan, R. Wang, Y. Liu, J. A. Birchler, and F. Han, "Efficient targeted genome modification in maize using CRISPR/Cas9 system," *Journal of Genetics and Genomics*, vol. 43, no. 1, pp. 37–43, 2016.

[74] J. G. Doench, N. Fusi, M. Sullender et al., "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9," *Nature Biotechnology*, vol. 34, no. 2, pp. 184–191, 2016.

[75] S. Q. Tsai and J. K. Joung, "Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases," *Nature Reviews Genetics*, vol. 17, no. 5, pp. 300–312, 2016.

[76] B. P. Kleinstiver, V. Pattanayak, M. S. Prew et al., "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects," *Nature*, vol. 529, no. 7587, pp. 490–495, 2016.

[77] S. Schiml, F. Fauser, and H. Puchta, "The CRISPR/Cas system can be used as nuclease for *in planta* gene targeting and as paired nickases for directed mutagenesis in *Arabidopsis* resulting in heritable progeny," *Plant Journal*, vol. 80, no. 6, pp. 1139–1150, 2014.

[78] G. T. H. Vu, T. Schmutzer, F. Bull et al., "Comparative genome analysis reveals divergent genome size evolution in a carnivorous plant genus," *Plant Genome*, vol. 8, no. 3, 2015.

[79] I. Schubert and G. T. Vu, "Genome stability and evolution: attempting a holistic view," *Trends in Plant Science*, vol. 21, no. 9, pp. 749–757, 2016.

[80] M. Endo, M. Mikami, and S. Toki, "Multigene knockout utilizing off-target mutations of the CRISPR/cas9 system in rice," *Plant and Cell Physiology*, vol. 56, no. 1, pp. 41–47, 2015.

[81] 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.

[82] 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.

[83] L. G. Lowder, D. Zhang, N. J. Baltes et al., "A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation," *Plant Physiology*, vol. 169, no. 2, pp. 971–985, 2015.

[84] J. Steinert, S. Schiml, F. Fauser, and H. Puchta, "Highly efficient heritable plant genome engineering using Cas9 orthologues from *Streptococcus thermophilus* and *Staphylococcus aureu*," *Plant Journal*, vol. 84, no. 6, pp. 1295–1305, 2015.

[85] X. Hu, C. Wang, Y. Fu, Q. Liu, X. Jiao, and K. Wang, "Expanding the range of CRISPR/Cas9 genome editing in rice," *Molecular Plant*, vol. 9, no. 6, pp. 943–945, 2016.

[86] T. Maruyama, S. K. Dougan, M. C. Truttmann, A. M. Bilate, J. R. Ingram, and H. L. Ploegh, "Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining," *Nature Biotechnology*, vol. 33, no. 5, pp. 538–542, 2015.

[87] M. Endo, M. Mikami, and S. Toki, "Biallelic gene targeting in rice," *Plant Physiology*, vol. 170, no. 2, pp. 667–677, 2016.

[88] C. D. Richardson, G. J. Ray, M. A. DeWitt, G. L. Curie, and J. E. Corn, "Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA," *Nature Biotechnology*, vol. 34, no. 3, pp. 339–344, 2016.

[89] N. J. Baltes, J. Gil-Humanes, T. Cermak, P. A. Atkins, and D. F. Voytas, "DNA replicons for plant genome engineering," *Plant Cell*, vol. 26, no. 1, pp. 151–163, 2014.

[90] T. Čermák, N. J. Baltes, R. Čegan, Y. Zhang, and D. F. Voytas, "High-frequency, precise modification of the tomato genome," *Genome Biology*, vol. 16, article 232, 2015.

[91] K.-J. Appenroth, D. J. Crawford, and D. H. Les, "After the genome sequencing of duckweed—how to proceed with research on the fastest growing angiosperm?" *Plant Biology*, vol. 17, supplement 1, pp. 1–4, 2015.

[92] H. X. Cao, G. T. H. Vu, W. Wang, K. J. Appenroth, J. Messing, and I. Schubert, "The map-based genome sequence of Spirodela polyrhiza aligned with its chromosomes, a reference for karyotype evolution," *New Phytologist*, vol. 209, no. 1, pp. 354–363, 2016.

[93] W. Wang, G. Haberer, H. Gundlach et al., "The *Spirodela polyrhiz*a genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle," *Nature Communications*, vol. 5, article 3311, 2014.

[94] A. Cantó-Pastor, A. Mollá-Morales, E. Ernst et al., "Efficient transformation and artificial miRNA gene silencing in *Lemna minor*," *Plant Biology*, vol. 17, no. 1, pp. 59–65, 2015.

[95] J. D. Wolt, K. Wang, and B. Yang, "The regulatory status of genome-edited crops," *Plant Biotechnology Journal*, vol. 14, no. 2, pp. 510–518, 2016.

[96] E. Waltz, "Gene-edited CRISPR mushroom escapes US regulation," *Nature*, vol. 532, no. 7599, pp. 293–293, 2016.

[97] J. W. Woo, J. Kim, S. I. Kwon et al., "DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins," *Nature Biotechnology*, vol. 33, no. 11, pp. 1162–1164, 2015.

[98] X. Ji, H. Zhang, Y. Zhang, Y. Wang, and C. Gao, "Establishing a CRISPR-Cas-like immune system conferring DNA virus resistance in plants," *Nature Plants*, vol. 1, Article ID 15144, 2015.

[99] N. J. Baltes, A. W. Hummel, E. Konecna et al., "Conferring resistance to geminiviruses with the CRISPR-Cas prokaryotic immune system," *Nature Plants*, vol. 1, no. 10, Article ID 15145, 2015.

[100] Z. Ali, A. Abulfaraj, A. Idris, S. Ali, M. Tashkandi, and M. M. Mahfouz, "CRISPR/Cas9-mediated viral interference in plants," *Genome Biology*, vol. 16, no. 1, article 238, 2015.

[101] M. R. O'Connell, B. L. Oakes, S. H. Sternberg, A. East-Seletsky, M. Kaplan, and J. A. Doudna, "Programmable RNA recognition and cleavage by CRISPR/Cas9," *Nature*, vol. 516, no. 7530, pp. 263–266, 2014.

[102] D. D. Zhang, Z. X. Li, and J. F. Li, "Genome editing: new antiviral weapon for plants," *Nature Plants*, vol. 1, no. 10, Article ID 15146, 2015.

[103] A. Chaparro-Garcia, S. Kamoun, and V. Nekrasov, "Boosting plant immunity with CRISPR/Cas," *Genome Biology*, vol. 16, article 254, 2015.

[104] K. M. Esvelt, A. L. Smidler, F. Catteruccia, and G. M. Church, "Concerning RNA-guided gene drives for the alteration of wild populations," *eLife*, vol. 3, Article ID e03401, 2014.

[105] J. E. DiCarlo, A. Chavez, S. L. Dietz, K. M. Esvelt, and G. M. Church, "Safeguarding CRISPR-Cas9 gene drives in yeast," *Nature Biotechnology*, vol. 33, no. 12, pp. 1250–1255, 2015.

[106] A. Hammond, R. Galizi, K. Kyrou et al., "A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*," *Nature Biotechnology*, vol. 34, no. 1, pp. 78–83, 2016.

[107] O. S. Akbari, H. J. Bellen, E. Bier et al., "Safeguarding gene drive experiments in the laboratory," *Science*, vol. 349, no. 6251, pp. 927–929, 2015.

[108] J. Champer, A. Buchman, and O. S. Akbari, "Cheating evolution: engineering gene drives to manipulate the fate of wild populations," *Nature Reviews Genetics*, vol. 17, no. 3, pp. 146–159, 2016.

[109] V. M. Gantz, N. Jasinskiene, O. Tatarenkova et al., "Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 49, pp. E6736–E6743, 2015.

[110] E. Martin Avila, M. F. Gisby, and A. Day, "Seamless editing of the chloroplast genome in plants," *BMC Plant Biology*, vol. 16, no. 1, article 168, 2016.

[111] G. Farré, R. M. Twyman, P. Christou, T. Capell, and C. Zhu, "Knowledge-driven approaches for engineering complex metabolic pathways in plants," *Current Opinion in Biotechnology*, vol. 32, pp. 54–60, 2015.

[112] Y. Hyun, J. Kim, S. W. Cho, Y. Choi, J.-S. Kim, and G. Coupland, "Site-directed mutagenesis in *Arabidopsis thaliana* using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles," *Planta*, vol. 241, no. 1, pp. 271–284, 2015.

[113] Z.-P. Wang, H.-L. Xing, L. Dong et al., "Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in *Arabidopsis* in a single generation," *Genome Biology*, vol. 16, article 144, 2015.

[114] W. Lau, M. A. Fischbach, A. Osbourn, and E. S. Sattely, "Key applications of plant metabolic engineering," *PLoS biology*, vol. 12, no. 6, Article ID e1001879, 2014.

[115] J. Kamens, "The Addgene repository: an international nonprofit plasmid and data resource," *Nucleic Acids Research*, vol. 43, no. 1, pp. D1152–D1157, 2015.

[116] K. Xie, J. Zhang, and Y. Yang, "Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated

genome editing in model plants and major crops," *Molecular Plant*, vol. 7, no. 5, pp. 923–926, 2014.

[117] S. Bae, J. Park, and J.-S. Kim, "Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases," *Bioinformatics*, vol. 30, no. 10, pp. 1473–1475, 2014.

[118] F. Heigwer, G. Kerr, and M. Boutros, "E-CRISP: fast CRISPR target site identification," *Nature Methods*, vol. 11, no. 2, pp. 122–123, 2014.

[119] K. Labun, T. G. Montague, J. A. Gagnon, S. B. Thyme, and E. Valen, "CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering," *Nucleic Acids Research*, vol. 44, no. W1, pp. W272–W276, 2016.

[120] A. Xiao, Z. Cheng, L. Kong et al., "CasOT: a genome-wide Cas9/gRNA off-target searching tool," *Bioinformatics*, vol. 30, no. 8, pp. 1180–1182, 2014.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 2095195, 13 pages
http://dx.doi.org/10.1155/2016/2095195

Research Article

Gene Expression Analysis of Alfalfa Seedlings Response to Acid-Aluminum

Peng Zhou,[1] Liantai Su,[1] Aimin Lv,[1] Shengyin Wang,[1] Bingru Huang,[2] and Yuan An[1,3]

[1]*School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*Department of Plant Biology and Pathology, Rutgers University, New Brunswick, NJ 08901, USA*
[3]*Key Laboratory of Urban Agriculture (South), Ministry of Agriculture, Shanghai 201101, China*

Correspondence should be addressed to Yuan An; anyuan@sjtu.edu.cn

Received 27 July 2016; Accepted 12 October 2016

Academic Editor: Wenwei Xiong

Acid-Aluminum (Al) is toxic to plants and greatly affects crop production worldwide. To understand the responses of plants to acid soils and Aluminum toxicity, we examined global gene expression using microarray data in alfalfa seedlings with the treatment of acid-Aluminum. 3,926 genes that were identified significantly up- or downregulated in response to Al^{3+} ions with pH 4.5 treatment, 66.33% of which were found in roots. Their functional categories were mainly involved with phytohormone regulation, reactive oxygen species, and transporters. Both gene ontology (GO) enrichment and KEGG analysis indicated that phenylpropanoid biosynthesis, phenylalanine metabolism, and flavonoid biosynthesis played a critical role on defense to Aluminum stress in alfalfa. In addition, we found that transcription factors such as the MYB and WRKY family proteins may be also involved in the regulation of reactive oxygen species reactions and flavonoid biosynthesis. Thus, the finding of global gene expression profile provided insights into the mechanisms of plant defense to acid-Al stress in alfalfa. Understanding the key regulatory genes and pathways would be advantageous for improving crop production not only in alfalfa but also in other crops under acid-Aluminum stress.

1. Introduction

Aluminum (Al) combined with acid is the main factor limiting plant growth and crop production worldwide [1]. Al in soils is solubilized into ionic forms, especially when the soil pH falls to lower than 5. Roots are the primary targets of acid-Al toxicity in plants. Several studies have reported Al inhibition of cell elongation and cell division in plant roots [2–4]. The root apex (particularly the distal transition zone of the root) has been shown to be a critical site for the perception of Al toxicity [5]. Zhou et al. [6] reported the presence of Al ions in cell walls, intracellular membranes, and the center of the nucleus in alfalfa root cells. Furthermore, extensive research has demonstrated that Al^{3+} alters physiological processes (i.e., cytosolic Ca^{2+} homeostasis and cytoskeleton dynamics) and modifies the levels of endogenous nitric oxide in the root tips [7–9].

Al-induced toxicity is caused by the high binding affinity of Al to various extracellular and intracellular substances. Most reports have suggested that organic acids (OAs) play an important role in the mechanism by which plants tolerate Al stress [10]. Plants also have other mechanisms to cope with Al stress. Phenolic compounds such as flavonoids, alkaloids, terpenoids, and glycosides form strong complexes with Al ions, and these compounds have been implicated in internal Al detoxification in *Camellia sinensis* and other Al-accumulating species [11, 12]. Kidd et al. [13] reported that differential Al-tolerance in *Zea mays* genotypes showed a better correlation with the rate of Al-stimulated root exudation of flavonoids (catechin and quercetin) than with Al-activated exudation of OAs. Other studies showed that the induction of antiperoxidation enzymes could ameliorate the oxidative damage caused by Al stress and lead to Al-tolerance phenotypes in various plants [14, 15].

Many genes and signaling pathways have been proposed to be involved in the Al stress response in plants [16–19]. A group of Al-induced genes, such as *wali1–5* in wheat *(Triticum aestivum)*, *Sali5-4a* and *Sali3-2* in soybeans *(Glycine max)*, and *ALS3* in *Arabidopsis*, have been identified and characterized [20–22]. *Medicago sativa* L. (alfalfa) is very

sensitive to acid and Al ions. The alfalfa yield in acidic soils was inhibited due to reduced nitrogen fixation and destroying symbiotic bacteria [23]. However, the underlying mechanism of Aluminum phytotoxicity on root growth at the molecular level remains unclear. Here, we used microarray analysis to investigate genome-wide transcriptional profiling and bioinformatics data mining to examine the enriched gene ontology and metabolic pathways. The identified genes, which is differentially expressed under Al stress, together with the metabolic pathway information obtained from microarray analysis, will provide an informative platform for cultivating Al-tolerant species with improved agronomic features in the future.

2. Materials and Methods

2.1. Plant Material and Treatment. Alfalfa (WL-525), which is an Al-tolerant cultivar [24, 25], was obtained from the National Seed Corporation (New Delhi, India). Healthy seeds of uniform size were surface-sterilized with 0.5% (v/v) sodium hypochlorite solution and repeatedly washed with double-distilled water. After drying with a blotting paper, the seeds were placed on two layers of filter paper in a petri dish. The filter paper was soaked in 2 mL of 0.2 mM $CaCl_2$ solution containing 0 (pH 6.0), 0 (pH 4.5), 0.8 (pH 4.5), or 3.2 (pH 4.5) mM $AlCl_3$. The pH was adjusted by the addition of 1 M HCl. The experiments were conducted in an environmentally controlled growth room with 14 h/27°C day and 10 h/25°C night cycles, light intensity of 480 μmol·m^2·s^{-1}, and relative humidity of 70 ± 5%. After germination for 60 h, the seedlings with green cotyledons and formed roots were defined as being successfully germinated and survived. Seedlings without green cotyledons or formed roots were considered dead and failed germination. We then calculated the germination and survival rate according to these definitions. All the experiments were repeated three times. The whole seedlings (with roots, stems and leaves) germinated in the presence of 0 (pH 6.0), 0 (pH 4.5), 0.8 (pH 4.5), and 3.2 (pH 4.5) mM of $AlCl_3$ were collected, frozen in liquid nitrogen for 7 min, and then stored at −80°C for microarray analysis.

2.2. Microarray Analysis. Seedlings germinated in the presence of 0 (pH 6.0), 0 (pH 4.5), 0.8 (pH 4.5), or 3.2 (pH 4.5) mM $AlCl_3$ for 60 h were collected and used for microarray analysis. Total RNA was extracted using the TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) from the germinated alfalfa samples. The quality and integrity of the total RNA were evaluated with an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc. Santa Clara, USA) (with $OD_{260\,nm}/OD_{280\,nm} \geq 1.8$ and RIN ≥ 9.0). RNA purification and microarray hybridization were performed according to the one-color microarray-based gene expression analysis protocol. RNA samples (3 μg) extracted from each individual were pooled to form four sets (referred to as seedlings germinated under 0 (pH 6.0), 0 (pH 4.5), 0.8 (pH 4.5), or 3.2 (pH 4.5) mM $AlCl_3$ solution for 60 h). The RNA pool was used as a template for cDNA preparation. cDNA was further transcribed into cRNA and double-labeled using an Agilent low RNA input fluorescent linear amplification kit (Agilent Technologies, Santa Clara, CA, USA). Then, 0.5 μg of labeled cRNA samples was purified, mixed with hybridization buffer, and hybridized to oligonucleotide microarrays (Medicago Gene Expression 4 × 44K; Agilent, Santa Clara, CA, USA; http://www.genomics.agilent.com/) for 17 hours at 65°C. The microarrays were designed based on RefSeq (Release 32), UniGene (Build 33), TIGR Plant Transcript Assemblies (Release 2), and TIGR Gene Indices (Release 9) and contained a total of 43,803 oligonucleotide probes (60-mer). After hybridization, the slide glass was washed using a gene expression wash buffer kit (Agilent) and scanned with the Genepix 400B (Axon Instruments, Foster City, CA, USA). The fluorescence intensity was calculated using Feature Extraction software version 9.5 (Agilent), and the data were analyzed with GeneSpring GX software version 11.0 (Agilent). The whole experiments were biologically repeated three times, and the microarray data were normalized by GeneSpring GX 11.0.

Transcripts with more than twofold differences between the seedlings grown under Al stress at a specified statistical cutoff (fold change [FC] ≥ 2.0 and $P < 0.05$ according to the t test) were defined as differentially expressed genes.

2.3. Quantitative Real-Time RT-PCR (qPCR). To validate our microarray results, total RNA was extracted from the alfalfa seedlings germinated with different concentrations of $AlCl_3$ solution using the TRIzol Reagent (Invitrogen, Carlsbad, CA, USA). First-strand cDNA was generated with a RETROscript kit (Invitrogen) using oligo (dT) primers and the RETROscript RTase. Then, we examined 17 genes that were differentially up- or downregulated following Al^{3+} treatments (FC ≥ 2.0 and $P < 0.05$ t test) using qPCR. 17 genes which were mainly involved in the phenylpropanoid and flavonoid biosynthesis pathways (based on AgriGO and KEGG analysis) or transcription factors (TFs) that might be related to metabolic pathways [26, 27]. The specific primers used for qPCR were listed in Table 1. These primers were used to validate differentially expressed genes (FC ≥ 2.0 and $P < 0.05$; t test) obtained from microarray analysis. Briefly, the 25 μL qPCR amplification mixture contained 25 ng of template cDNA, 12.5 μL of 2x SYBR Green I Master Mix buffer (Applied Biosystems), and 300 nM each of the forward and reverse primers. The reactions were run on an ABI Prism 5700 sequence detector (Applied Biosystems). The PCR protocol was as follows: polymerase activation and predenaturation for 4 min at 94°C, followed by 40 cycles at 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s. We selected three genes (EF-α, 18S rRNA, and ubiquitin; primers listed in Table 1, accession numbers: XM_003618727 in Genbank, DQ311983 in Genbank, and TC174254 in the Dana-Farber Cancer Institute [DFCI], resp.) as internal controls. The geometric mean of their C_t was used as the endogenous control. Each qPCR reaction was repeated three times. All PCR efficiencies were above 95%. Results of the sequence detection software (version 1.3, Applied Biosystems) were exported as tab-delimited text files and imported into Microsoft Excel for further analysis. The median coefficient of variation (based on calculated quantities) of the duplicate samples was 6%.

Table 1: Primer sequences used for qPCR.

Probe Name	Description	Forward sequence	Reverse sequence	Accession number
A_27_P077456	Peroxidase-456	5′-AGGAAATCTAAGGTGGCAACTG-3′	5′-TTTAGGTAAGCCAGGAATGTGG-3′	TC192119 (DFCI)
A_27_P091936	Resistance protein-936	5′-ATCGTGGAATGGGAAAGACAAC-3′	5′-CCAGGACCAAACCAATCAAGT-3′	TC181653 (DFCI)
A_27_P015415	WRKY 11	5′-TCATTTCCTCTGGCAAGCCT-3′	5′-TCAGCGACCTTTGAACTTATCG-3′	TC175297 (DFCI)
A_27_P274912	CCCH29-LIKE	5′-CAAGAGGGAAGTAGATGAGAAGGA-3′	5′-CAACAGCACAATGAAGAGCAG-3′	TC199223 (DFCI)
A_27_P133981	Myb-like	5′-AGAGGACAATGGGAAAGAAGAC-3′	5′-CAGCACTTGATGCCTAAGACA-3′	TC183988 (DFCI)
A_27_P155316	Nbs-containing resistance-like protein	5′-AACGCTCTGAACAACGAGGA-3′	5′-CACCGAAATCACACTCCGAG-3′	XM_003588950
A_27_P055706	Peroxiredoxin-chloroplastic-like	5′-GAATCCACCTTCTCCTACCTCG-3′	5′-CTTCAAATCCTCCTTCCACGC-3′	TC185056 (DFCI)
A_27_P181166	MYB an2	5′-GGACATACGAGGAAGACAACTTAC-3′	5′-CCTTCCAGCAATCAATGACCAT-3′	XM_003621524
A_27_P050191	Pathogenesis-related protein 4	5′-TGGTTACGGGATGTCTCAAGG-3′	5′-TTTGGTGTTGGTGTTGGTGC-3′	TC183745 (DFCI)
A_27_P162931	bHLH120	5′-TCAACCACCACCAACATCAC-3′	5′-TGAAGGGTAGCCATTTCTTGTC-3′	TC190718 (DFCI)
A_27_P111336	Peroxidase 43-like	5′-GTCCAGGAGTGGTTTCTTGTG-3′	5′-CTGTGAGACCCTTGTTTAGGAAC-3′	TC190043 (DFCI)
A_27_P036981	Peroxidase 12-like	5′-TTCCTCTGTTCTGGCTAATGGT-3′	5′-AGCACCTGAAAGGGCAACTA-3′	TC196923 (DFCI)
A_27_P104616	bZIP60	5′-ATCCTTCTGTTTCCGTCGCA-3′	5′-TCCCTGTTCCTCATCTGCCT-3′	TC180070 (DFCI)
A_27_P136211	WRKY 22	5′-CCCTAAAGAGCCTGAACAAGTC-3′	5′-GCTTCGTGGATAAGGTGAACC-3′	TC196399 (DFCI)
A_27_P123906	DREF1	5′-CCTTCCTATTCCAGCAACTTCC-3′	5′-CCTGTTCATCAACTTCCACACA-3′	TC182024 (DFCI)
A_27_P045576	Anthocyanidin reductase-like	5′-CATTTACCGACCCTGCTGGT-3′	5′-TCTCTGCCCTCATCTTGCCT-3′	TC186981 (DFCI)
A_27_P060501	P450 83b1-like	5′-GTTTGTAGGACTGCGTTCGG-3′	5′-TTCGTGTGGAGGCAACTTCT-3′	TC173827 (DFCI)
N	EF-α	5′-GCACCAGTGCTCGATTGC-3′	5′-TCGCCTGTCAATCTTGGTAACAA-3′	XM_003618727
N	18S ribosomal RNA gene	5′-TCAGAGGATGGCGACGAAG-3′	5′-CCGTTGCCGAGAGTCATTCT-3′	DQ311983
N	Ubiquitin	5′-CTCACTGGAAAGACAATCACCC-3′	5′-GAAGTCGCAACACAAGATGGA-3′	TC174254 (DFCI)

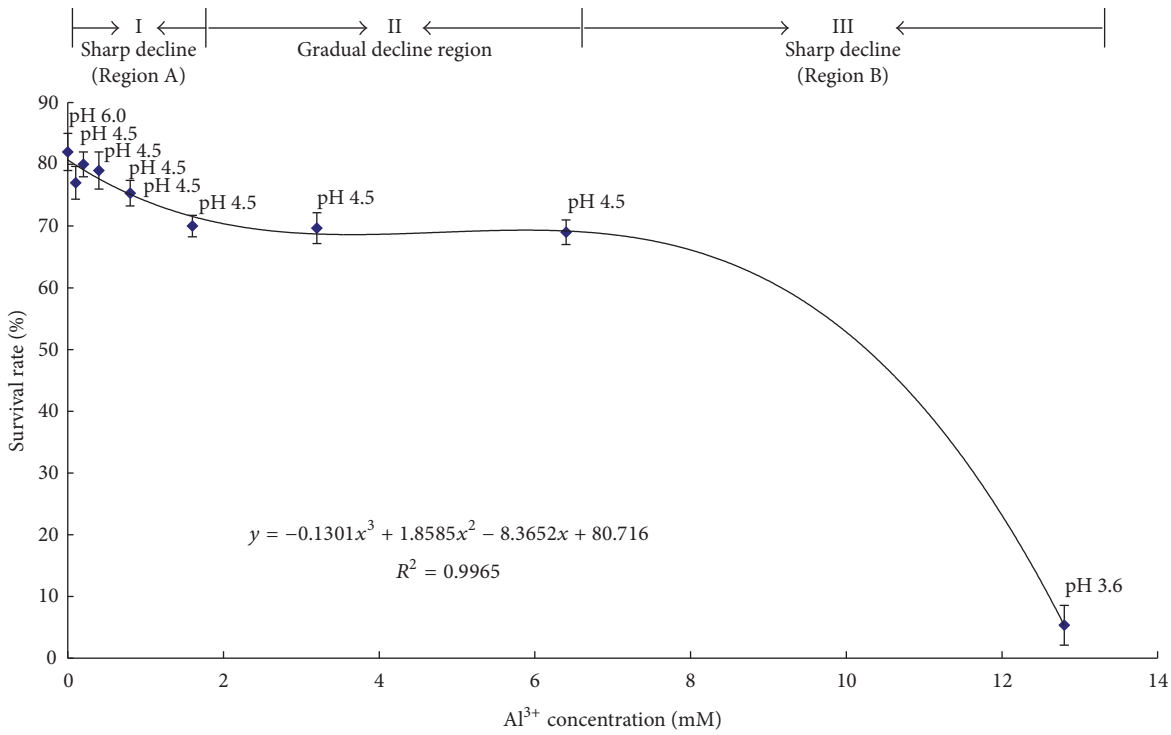


FIGURE 1: Survival rate of alfalfa under Al stress. The regression equation and the square of correlation coefficient are presented. Based on the regression equation curve, the Al concentration affecting alfalfa germination can be divided into three regions, I of sharp decline (region A); II of gradual decline; III of sudden decline (region B).

2.4. GO Enrichment and KEGG Analysis. GO functional enrichment analysis was performed using singular enrichment analysis (SEA) on AgriGO (http://bioinfo.cau.edu.cn/agriGO/) [26]. The MAGA data were used as the background reference, and the hypergeometric test was used for statistical analysis. GO terms could be divided into three categories: biological process, cellular component, and molecular function. All significant GO secondary level terms could be used to generate a flash bar charts showing the overrepresented terms in all three categories. The genes with significant GO categories were subjected to hierarchical clustering analysis using Genesis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis.

2.5. Statistical Analyses. All results shown in the figures were the mean ± SE of at least three replicates. Significant differences between and among treatments were statistically evaluated by analysis of variance using SAS version 9.0 (SAS Institute Inc., Cary, NC, USA) with statistical significance set at $P = 0.05$. Analysis of correlation coefficients among 226 genes with significantly enriched GO terms was performed using SAS 9.0. The relationship between the TFs and metabolic genes was also studied. Genes that had correlation coefficients (r) ≥ 0.7 ($P \leq 0.01$) or (r) ≤ -0.7 ($P \leq 0.01$) (especially between genes encoding TFs and metabolism-associated genes) were selected and analyzed (Table 3).

3. Results

3.1. Seedling Survival Rate. The germination of alfalfa declined with increasing concentration of Al ions from 0 mM to 1.6 mM, and the seedling survival rates gradually decreased from 1.6 to 6.4 mM. However, when the concentration of Al ions was increased to 12.8 mM with the pH value of 3.6, the survival rate declined to 5% ($r^2 = 0.997$; Figure 1). Based on the regression equation curve shown in Figure 1, the Al concentrations could be divided into three ranges based on its survival rate in alfalfa: concentration of 0–1.6 mM Al caused a sharp decline in the germination and survival rates; 1.6–6.4 mM Al caused a gradual decline; 6.4–12.8 mM Al resulted in most steep decline. Although we tried to represent each stage for our experimental design, it is challengeable to get enough tissue under 6.4–12.8 mM Al. Thus, the treatments of alfalfa seeds germinated in the presence of 0 μM Al^{3+} (pH 6.0), 0 μM Al^{3+} (pH 4.5), 800 μM Al^{3+} (pH 4.5), and 3.2 mM Al^{3+} (pH 4.5) were used for further microarray analysis.

3.2. Microarray Data Quality Assessment. The data quality was assessed using two measurements. First, a correlation > 0.96 was obtained among three biological replicates of all of the treatments analyzed. A principal component analysis demonstrated that seedlings treated with or without Al^{3+} at pH 6.0 or 4.5 were distributed in distinct groups (Figure 2). The distinction among data points from different sample treatments validated our experimental pipeline. Second, we validated the expression profiles of 17 genes using qPCR. The results were highly consistent with the data from the microarray analysis ($r = 0.76$; $P < 0.01$) (Figure 3). Taken together, the microarray data obtained in this study were reliable for further study.

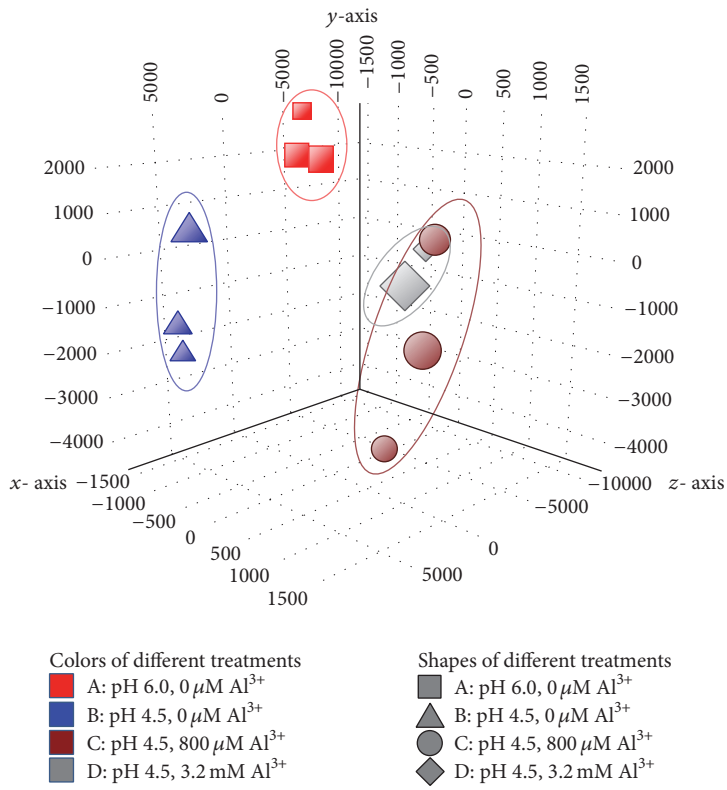


FIGURE 2: Principle component analysis for microarray data. Principle component analysis of intact genes detected by microarray from samples under different Al treatments for 60 h. The colored graph shows all data points projected in the three-dimensional space formed by three coordinates after rotation. Each data point represents an independent Al treatment, with the red colored square representing samples treated with pH 6.0 and 0 mM Al, the blue triangle representing samples treated with pH 4.5 and 0 mM Al, the brown circle representing samples treated with pH 4.5 and 0.8 mM Al and the gray rhombus representing samples treated with pH 4.5 and 3.2 mM Al. Samples with the same treatment are closely related and can be encircled together.

TABLE 2: Differentially expressed genes across all treatments.

	B versus A	C versus A	D versus A	C versus B	D versus B	D versus C
Downregulated	1126	340	566	735	1111	595
Upregulated	1379	1037	912	1381	1893	304
Total	4146					

All of the genes mapped to the reference sequence and genome sequences were examined for differences in expression across the different libraries. Numbers of differentially expressed genes were analyzed across sense transcripts using a threshold value FC ≥ 2. A, B, C, and D represent germinated seeds treated with 0 (pH 6.0), 0 (pH 4.5), 0.8 (pH 4.5), and 3.2 (pH 4.5) mM $AlCl_3$ solution for 60 h, respectively.

3.3. Features of the Expressed Genes. Of all the 43,803 probe sets measured by RNA hybridization, 43,651 were found expressed in the seedlings treated with Al at different pH values. The total number was reduced to about 65% of the transcripts after the probe sets with ambiguous signals and those that were not called "present" in at least two replicates were removed (see Table S1 in Supplementary Material available online at http://dx.doi.org/10.1155/2016/2095195). Of these, 4,146 transcripts were up- or downregulated with at least two times of FC and P value of 0.05 in the paired t-tests (Table 2).

There are 1037 and 912 genes being upregulated in the 800 μM and 3.2 mM Al^{3+} (pH 4.5) groups, respectively, in comparison with the corresponding gene expression levels in the 0 μM Al^{3+} (pH 6.0) group. Among the upregulated genes, 747 were common in both groups (Table 2). 340 and 566 genes were downregulated in the groups treated with 800 μM and 3.2 mM Al^{3+} (pH 4.5), respectively. 242 genes were downregulated in both groups (Table 2). Furthermore, in comparison with the corresponding gene expression levels in the 0 μM Al^{3+} (pH 4.5) group, 1381 and 1893 genes were upregulated in the 800 μM and 3.2 mM Al^{3+} (pH 4.5) groups, respectively, and 1052 were upregulated in both groups (Table 2). 735 and 1,111 genes were downregulated in those two groups, respectively, with 563 genes common in both groups (Table 2). Thus, a total of 3926 genes were found to be either up- or downregulated following exposure to Al^{3+} (FC ≥ 2.0; and $P < 0.05$, t-test).

Table 3: Correlation coefficient analysis of TFs with metabolic pathway genes.

TF	Probe	Genes having highly positive correlation coefficients with TF	Genes having significant negative correlation with TF
MYB 305	A_27_P172741 (unknown)	A_27_P036981 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.91$, $P < 0.001$ A_27_P053551 (r&s) (peroxidase; EC: 1.11.1.7), $r = 0.88$, $P < 0.001$ A_27_P111336 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.81$, $P = 0.0013$ A_27_P003941 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.81$, $P = 0.0013$	A_27_P042756 (unknown) (flavonoid 3′-monooxygenase; EC: 1.14.13.21), $r = -0.73$, $P = 0.006$
MYB apl	A_27_P129206 (unknown)	A_27_P077456 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.75$, $P = 0.005$ A_27_P036981 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.87$, $P < 0.001$ A_27_P241817 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.88$, $P < 0.001$ A_27_P111336 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.77$, $P = 0.003$ A_27_P170851 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.79$, $P = 0.002$ A_27_P053691 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.80$, $P = 0.0015$ A_27_P003941 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.86$, $P < 0.001$ A_27_P273952 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.72$, $P = 0.008$	
MYB an2	A_27_P181166 (unknown)	A_27_P241317 (r&s&l) (peroxidase; EC: 1.11.1.7), $r = 0.70$, $P = 0.01$	
MYB-like	A_27_P133981 (sr)	A_27_P053551 (r&s) (peroxidase; EC: 1.11.1.7), $r = 0.70$, $P = 0.01$ A_27_P111336 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.83$, $P < 0.001$ A_27_P273952 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.80$, $P = 0.0015$ A_27_P045576 (r&s&l) (cinnamoyl-CoA reducatase; dihydrokaempferol 4-reducatse; EC: 1.2.1.44; 1.1.1.219) $r = 0.87$, $P < 0.001$	A_27_P149301 (unknown) (4-coumarate-CoA ligase; EC: 6.2.1.12), $r = -0.76$, $P = 0.004$
MYC2	A_27_P057826 (r&s&l)	A_27_P273952 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.76$, $P = 0.003$ A_27_P045576 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219) $r = 0.83$, $P < 0.001$	
MYC2-like	A_27_P348932 (r&s&l)	A_27_P273952 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.95$, $P < 0.001$	
OCS TF	A_27_P065811 (sr)	A_27_P077456 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.89$, $P < 0.001$ A_27_P036981 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.72$, $P = 0.008$ A_27_P053551 (r&s) (peroxidase; EC: 1.11.1.7), $r = 0.81$, $P = 0.0013$ A_27_P111336 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.82$, $P = 0.001$ A_27_P170851 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.79$, $P = 0.002$ A_27_P003941 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.83$, $P < 0.001$ A_27_P273952 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.78$, $P = 0.0025$	A_27_P149301 (unknown) (4-coumarate-CoA ligase; EC: 6.2.1.12), $r = -0.72$, $P = 0.008$
WRKY 40	A_27_P031021 (sr)	A_27_P065476 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.77$, $P = 0.003$ A_27_P045576 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219) $r = 0.89$, $P < 0.001$ A_27_P273952 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.72$, $P = 0.008$	
WRKY 11	A_27_P015415 (r&s&l)	A_27_P065476 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.85$, $P < 0.001$ A_27_P045576 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219) $r = 0.83$, $P < 0.001$	A_27_P149301 (unknown) (4-coumarate-CoA ligase; EC: 6.2.1.12), $r = -0.81$, $P = 0.001$ A_27_P263014 (r&s&l) (Shikimate O-hydroxycinnamoyl transferase; EC: 2.3.1.133), $r = -0.71$, $P = 0.009$

Table 3: Continued.

TF	Probe	Genes having highly positive correlation coefficients with TF	Genes having significant negative correlation with TF
			A_27_P034176 (r&s) (3-hydroxyisobutyryl-hydrolase 1; EC: 4.2.1.17), $r = -0.85$, $P < 0.001$
ERF13	A_27_P008251 (sr)	A_27_P077456 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.73$, $P = 0.007$ A_27_P111336 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.79$, $P = 0.002$ A_27_P065476 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.71$, $P = 0.0096$ A_27_P273952 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.88$, $P < 0.001$ A_27_P045576 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.82$, $P = 0.001$	
bHLH35	A_27_P048601 (unknown)	A_27_P036981 (sr) (peroxidase; EC: 1.11.1.7), $r = 0.76$, $P = 0.004$ A_27_P053551 (r&s) (peroxidase; EC: 1.11.1.7), $r = 0.76$, $P = 0.004$ A_27_P111336 (unknown) (peroxidase; EC: 1.11.1.7), $r = 0.87$, $P < 0.001$ A_27_P273952 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219), $r = 0.83$, $P < 0.001$ A_27_P045576 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219) $r = 0.81$, $P = 0.001$	A_27_P149301 (unknown) (4-coumarate-CoA ligase; EC: 6.2.1.12), $r = -0.70$, $P = 0.01$
CCCH 29	A_27_P274912 (r&s&l)	A_27_P045576 (r&s&l) (cinnamoyl-CoA reductase; dihydrokaempferol 4-reductase; EC: 1.2.1.44; 1.1.1.219) $r = 0.81$, $P = 0.001$	A_27_P149301 (unknown) (4-coumarate-CoA ligase; EC: 6.2.1.12), $r = -0.84$, $P < 0.001$ A_27_P034176 (r&s) (3-hydroxyisobutyryl-hydrolase 1; EC: 4.2.1.17), $r = -0.81$, $P = 0.001$

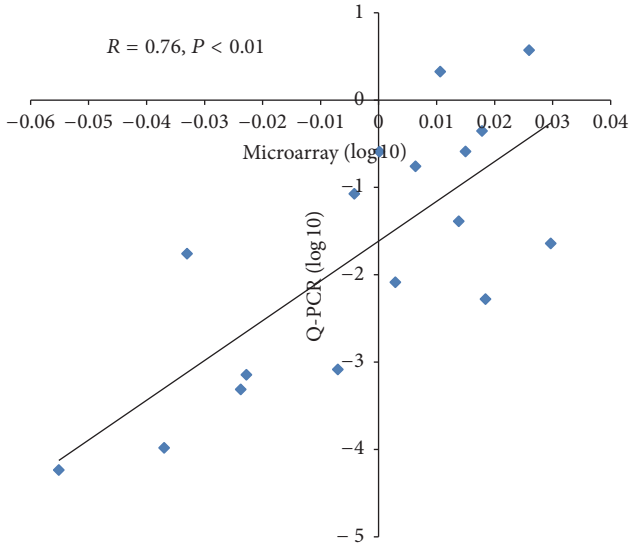


Figure 3: Relationship between microarray and qPCR data. The expression profiles of each gene based on the microarray data and qPCR were $\log_{10}$ transformed. The microarray data were plotted against the qPCR data.

To identify gene expression patterns, the EST sources in NCBI (National Center for Biotechnology Information) databases were used, and the results indicated the expression patterns of the 3926 genes were divided into nine types (Figure 4). Among those nine types, the most noticeable patterns include sr (genes expressed specifically in roots), r&l (genes expressed in roots and leaves), and r&s&l (genes expressed in roots, stems, and leaves), with proportions of 29.98%, 15.37%, and 14.14%, respectively. Genes expressed in root could account for 66.33% of all the genes, which indicated that most gene expressions were induced or inhibited in root under acid-Al stress. Although EST may be incomplete as an indication of gene activity, the comprehensive gene expression patterns did give us hint that gene expressions response to Al stress mainly happened in alfalfa root. In our study, the location of gene expression was shown in a parenthesis following the probe ID. For example, A_27_P051336 (r&s&l) meant the gene A_27_P051336 was expressed in root, stem, and leaf.

We summarized gene lists for response to plant hormone, stress defense, and membrane transporters. A total of 26 genes related to plant hormones (i.e., IAA, ABA, and ethylene) were found (Table S2), in which 9 genes were related to auxin, such as the auxin response protein genes, cationic peroxidase genes with IAA oxidase activity, and auxin conjugate hydrolase gene; 9 genes encoded ethylene responsive transcription factors; 2 genes were related to ABA; 2 genes were related to cytokinin; and 4 genes were related to gibberellin. A total of 28 genes were related to stress defense (Table S3), including an Aluminum sensitive protein,

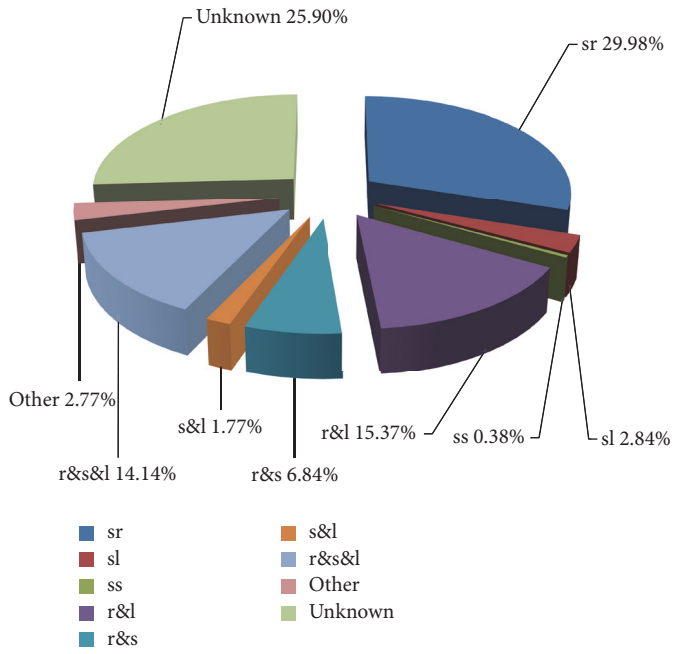


Figure 4: The percentages of expression patterns for differentially expressed genes. The percentages of expression patterns for the 3926 differentially expressed genes were shown in the pie chart. Basically, they were divided into nine types: (1) "sr" meant specifically expressing in roots; (2) "sl" meant specifically expressing in leaves; (3) "ss" meant specifically expressing in stems; (4) "r&l" indicated genes expressed in roots and leaves; (5) "r&s" indicated genes expressed in roots and stems; (6) "s&l" indicated genes expressed in stems and leaves; (7) "r&s&l" indicated genes expressed in roots, stems, and leaves; (8) "other" indicated genes expressed in other plant tissues; (9) "unknown" indicated genes which cannot identify their expression patterns in NCBI databases.

DREB (dehydration responsive element binding) protein, heat shock proteins, LEA (late embryogenesis abundant) proteins, and a universal stress protein. A total of 20 genes were membrane transporters (Table S4), including 8 ABC transporter genes, 6 nitrate transporter genes, peptide transporters, a potassium transporter, a sulfate transporter, a zinc transporter, and a mitochondrial phosphate transporter.

3.4. AgriGO Functional Enrichment Analysis of All Differentially Expressed Genes. GO functional enrichment was done against the total of 3926 differentially expressed genes. In the biological process category, the percentage of regulation of biological processes (GO:0050789) was 29.5% (12% in MAGA), biological regulation (GO:0065007) was 34% (14% in MAGA), cellular process (GO:0009987) was 63% (41.5% in MAGA), metabolic process (GO:0008152) was 58% (36% in MAGA), and response to stimulus (GO:0050896) was 53% (17.5% in MAGA). In the cellular component category, the percentage of cell (GO:0005623) was 62.3% (52.5% in MAGA). In the molecular function category, the percentage of transcription regulator activity (GO:0030528) was 21% (4% in MAGA) (Figure 5). 79 GO terms were found highly significantly (from the tertiary level to the bottom level) with the threshold of P values < 0.001 and false discovery rates (FDR) < 0.05 (Table S5). The most enriched GO terms included metabolic process, response to stimulus, and transcription regulator activity (Figure 5). The highest significance enriched GO terms with P values $\leq 5 \times 10^{-10}$ were related to the phenylpropanoid biosynthetic process (GO:0009699), flavonoid metabolic process (GO:0009812), and flavonoid biosynthetic process (GO:0009813), revealing that phenylpropanoid and flavonoid metabolism may be involved in the response to Al stress (Figure S1).

3.5. Hierarchical Clustering Analysis. When all of the genes in the 79 significant enrichment GO terms were combined, 226 genes were obtained. The Cluster program generated four different clusters according to the gene expression patterns (Figure 6). Cluster A included 32 genes that were downregulated by acid treatment. The genes in this cluster included those encoding the TFs WRKY 22 (A_27_P136691 (unknown)), MYB-like (A_27_P234867 (unknown)), OCSB factor 1 (A_27_P065811 (sr)), and MYC2 (A_27_P348932 (r&s&l)). Cluster B included 75 genes that were upregulated by Al-acid treatment, including genes encoding the TFs MYB an2 (A_27_P181166 (unknown)) and bHLH120 (A_27_P162931 (sr)) and genes involved in flavonoid biosynthesis (GO:0009813), such as dihydroflavonol 4-reductase (A_27_P258277 (other)) and the HCT (A_27_P263014 (r&s&l)). Cluster C included 3 genes that were upregulated by acid treatment. Cluster D included 116 genes that were downregulated by Al-acid treatment, including genes encoding the TFs WRKY 11 (A_27_P015415 (r&s&l)) and CCCH29-like (A_27_P274912 (r&s&l)) and genes involved in phenylpropanoid metabolism (GO:0009698) terms, such as TT7 (A_27_P042756 (unknown)) and cinnamoyl-CoA reductase (A_27_P045576 (r&s&l)). Comparison of gene

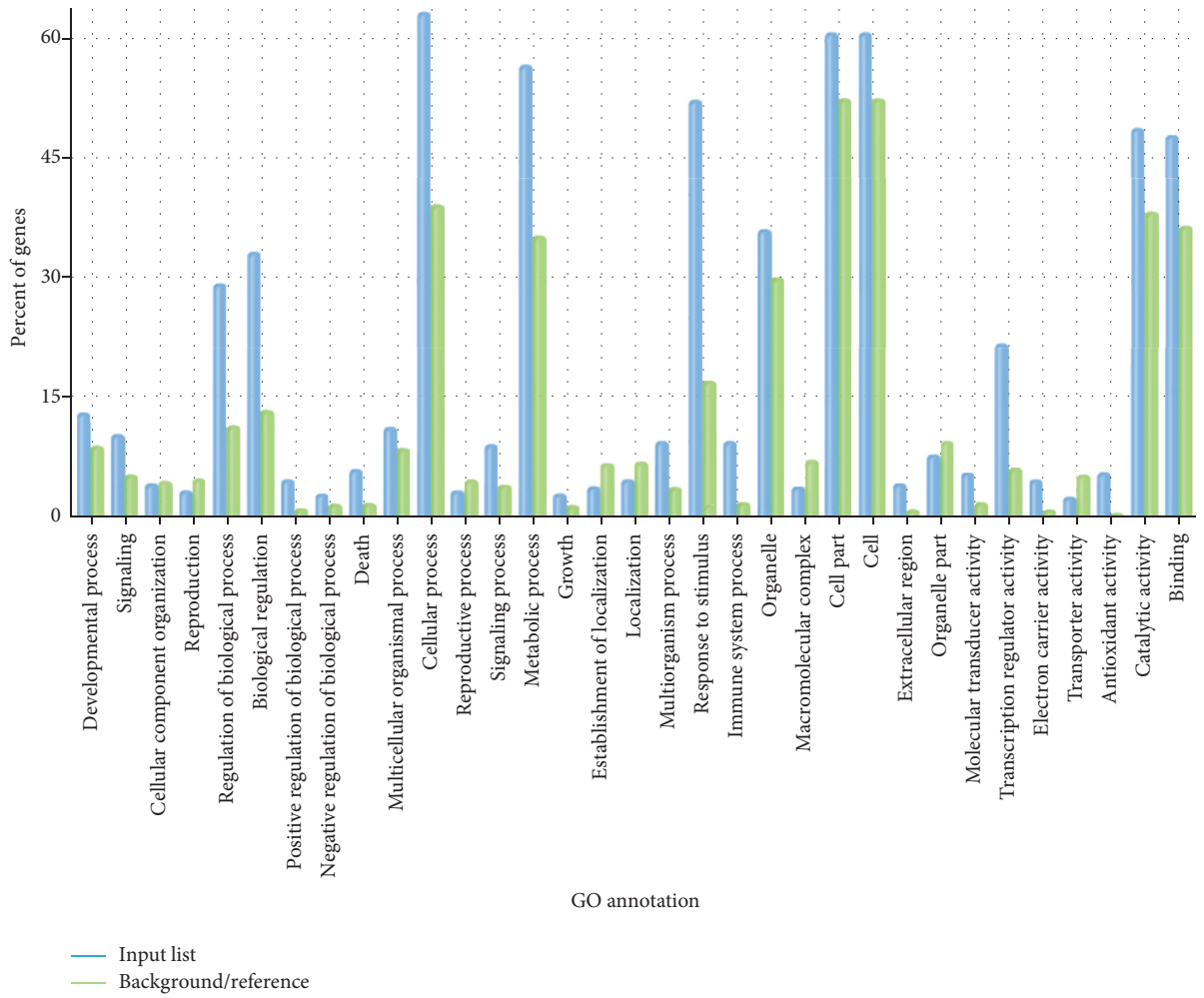


Figure 5: Flash bar chart of overrepresented terms in all three categories. The y-axis is the percentage of genes mapped by the term and represents the abundance of the GO term. The percentage for the input list is calculated by the number of genes mapped to the GO term divided by the total number of genes in the input list. The same calculation was applied to the reference list to generate its percentage. These two lists are represented using different colors. The x-axis is the definition of the GO terms.

expression under acid-Al treatment indicated that gene sets are differentially expressed under each stress treatment.

3.6. Kyoto Encyclopedia of Genes and Genomes (KEGG) Analysis of the Pathways. KEGG analysis revealed that phenylpropanoid biosynthesis, phenylalanine metabolism, and starch and sucrose metabolism were the three main pathways in 226 acid-Al-responsive genes (Table S6). Therefore, we focused on the genes involved in these pathways to demonstrate the utility of these data in understanding their specific functions in the response of alfalfa to Al stress.

Because the phenylpropanoid biosynthesis, phenylalanine metabolism, and flavonoid biosynthesis pathways were closely linked in plant metabolism, they were combined together in this analysis. The results showed 17 up- or downregulated genes were involved in phenylpropanoid biosynthesis. They encoded 6 key enzymes. Furthermore, 14 differentially expressed genes encoding 3 enzymes were involved in phenylalanine metabolism, and 4 genes encoding 3 enzymes were involved in flavonoid biosynthesis. Totally, 19 genes encoding 9 enzymes were involved in those three linked pathways (Figure S2). 11 genes encoded peroxidases, which were related to reactive oxygen species (ROS) scavenging and lignin synthesis. Based on the correlation coefficient analysis, many TFs, such as MYB305, MYB ap1, WRKY 40, and WRKY 11, were found to have high positive correlation coefficients with genes in this metabolic pathway (Table 3), indicating that these MYB family genes and WRKY family genes and others genes showed in Table 3 may be related to metabolic pathway regulation. Specifically, MYB apl, WRKY 11, and WRKY 40 had high positive correlation coefficients with genes related to ROS scavenging and lignin and flavonoid synthesis ($r > 0.7$, $P < 0.01$; Figure 7).

Finally, 12 genes encoding 7 enzymes were involved in starch and sucrose metabolism. Most enzymes can catalyze the hydrolysis of disaccharides to monosaccharides (Figure S3). However, we failed to find TFs with high positive correlation coefficients with genes involved in starch and sucrose metabolism.

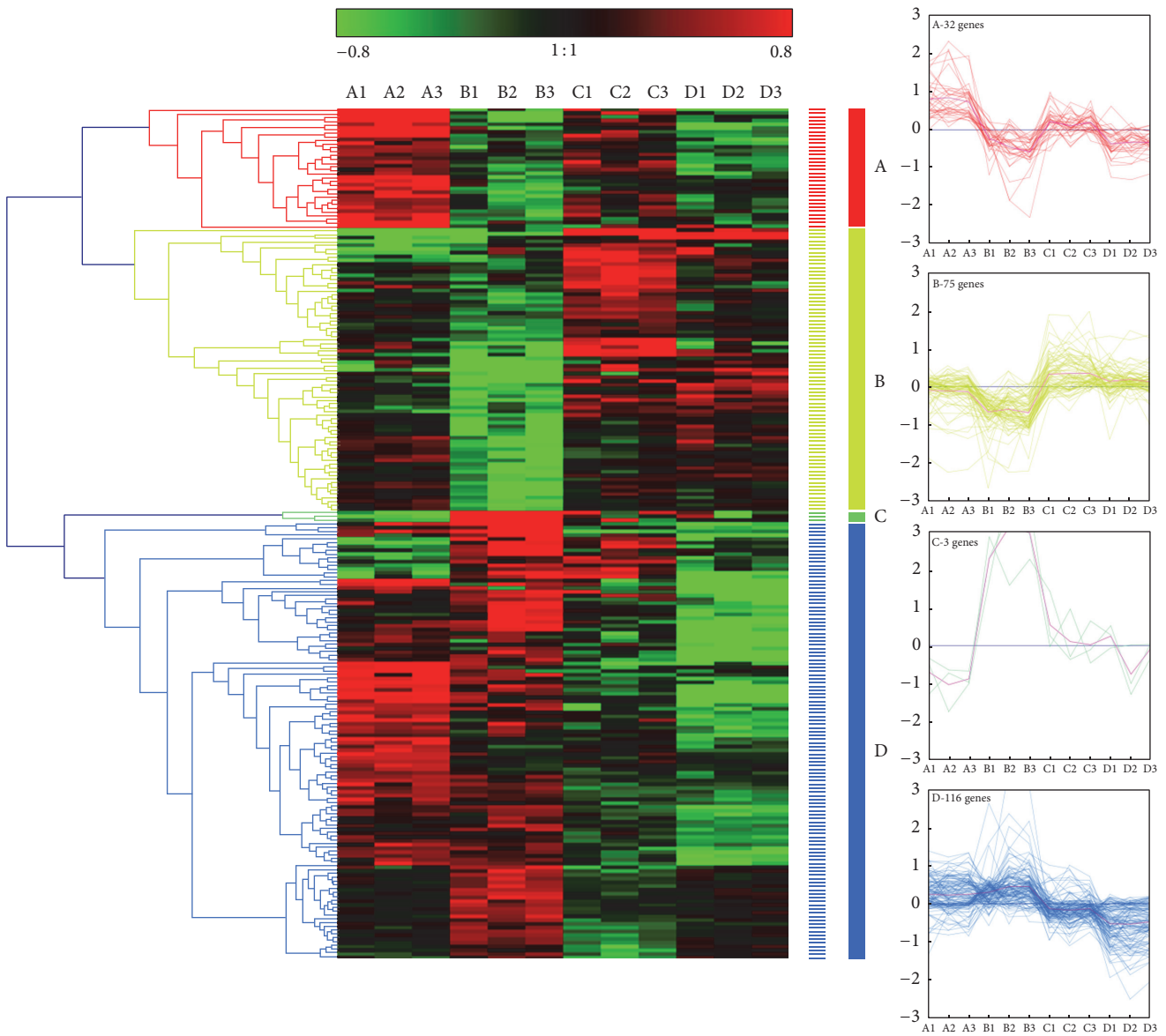


Figure 6: Cluster analysis of the expression profiles. Cluster analysis of the expression profiles of 226 genes in the highly significantly enriched GO terms. Cluster analysis for each group of genes was performed using hierarchical clustering with Genesis 1.7.5 with average linkage and Euclidian distance measurements. Rows represent differentially expressed genes, while columns represent different independent treatments (A, B, C, and D represent the germinated seeds treated with 0 (pH 6.0), 0 (pH 4.5), 0.8 (pH 4.5), and 3.2 (pH 4.5) mM $AlCl_3$ solution for 60 h, resp.; each treatment was repeated three times). The color scale shown at the top illustrates the relative expression ratios of genes across all samples. Four line charts of the four groups are shown on the right.

4. Discussion

4.1. Global Gene Expression Analysis of Alfalfa Exposed to Acid and High Concentrations of Al Ions. Global gene expression analysis using the Agilent gene expression microarray revealed the expression of 3926 genes that were changed by acid and Al ions, in which 66.33% can be detected in root including those specifically expressing in root which accounted for 29.98% of all the genes. Many genes related to phytohormones and ROS-metabolism were identified in the gene expression analysis. Potters et al. [28] hypothesized that Al toxicity can induce SIMR (stress-induced morphogenic response). Interestingly, phytohormones and ROS production are the main regulatory interactions controlling stress-induced SIMR of plants [28]. Thus, genes related to phytohormones and ROS-metabolism may be involved in SIMR following exposure to Al. In this study, the expression of 9 genes related to auxin and 9 genes related to ethylene were found to be up- or downregulated, revealing for the response of auxin and the ethylene to Al toxicity. Auxin and ethylene were found to play an important role in regulating Aluminum-induced inhibition of root growth [6, 29, 30]. Sun et al. [2] found that Al-induced ethylene may act as a signal to alter auxin distribution in roots, thereby inhibiting root

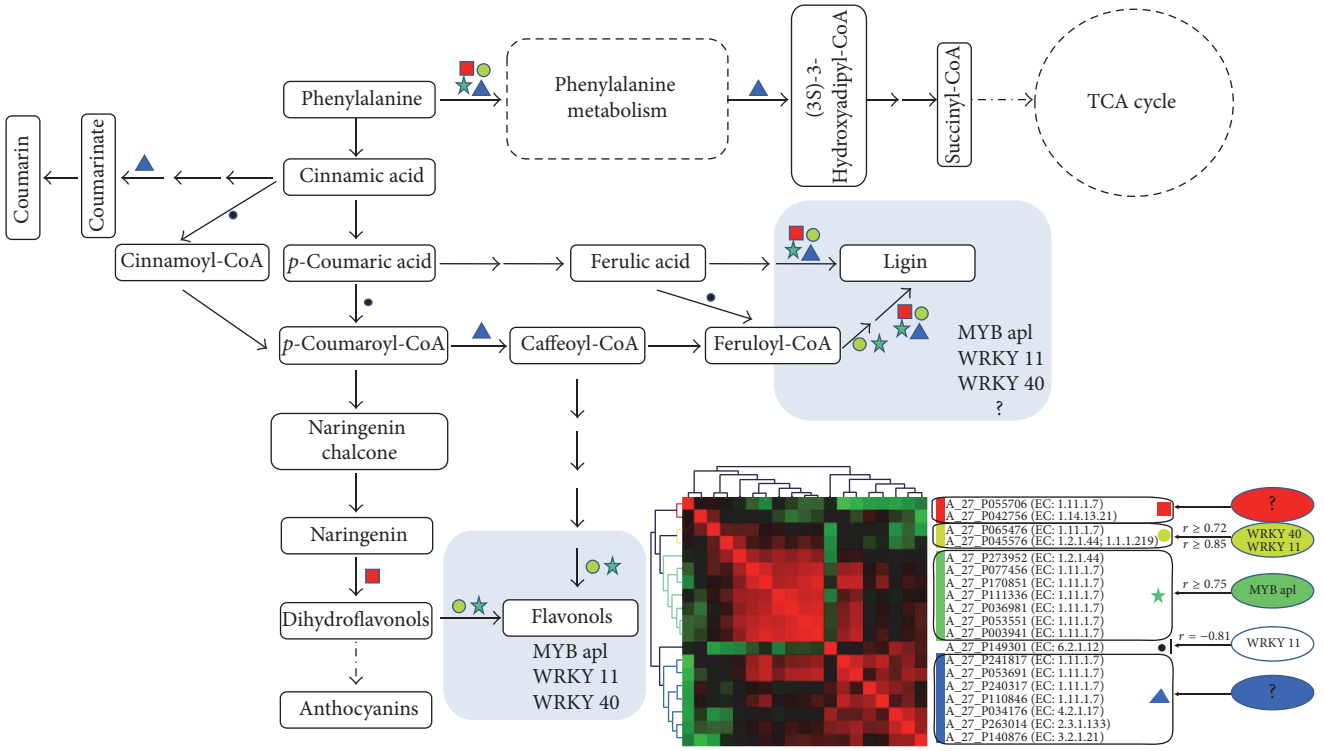


FIGURE 7: Pathways in response to acid-Al ions based on KEGG analysis. Overview of the combinations of the phenylpropanoid biosynthesis, phenylalanine metabolism, and flavonoid biosynthesis pathways in response to acid-Al ions based on KEGG analysis. Nineteen genes were differentially expressed in response to acid-Al ions. Hierarchical clustering of the results of the correlation coefficient analysis was shown in the lower right corner. The divided groups were marked by different shapes and color icons, which were also marked in the metabolic pathway. High positive correlation coefficient with genes in the divided groups were shown near the image of the hierarchical clustering analysis and also marked in the light shadow zone of the metabolic pathway.

elongation. Twelve genes encoding peroxidase (GO:0004601) could eliminate ROS production, leading to the plant morphological changes induced by Al.

A total of 20 genes encoding transporters were found in our Agilent gene expression microarray analysis, in which 8 genes encoded ABC transporters. ABC transporters contain an ATP-binding cassette (ABC) and participate directly in the transport of a wide range of molecules across membranes [31]. ABC transporters also participated in transporting phytohormones such as auxin and abscisic acid and were involved in the detoxification of toxic minerals, such as cadmium (Cd), arsenic (As), and Aluminum (Al) [32].

The AgriGO and KEGG analyses revealed 19 genes encoding 9 key enzymes involved in phenylpropanoid biosynthesis, phenylalanine metabolism, and flavonoid biosynthesis metabolism after the addition of acidic pH and high Al^{3+} concentrations (Figure S2). The activation of these pathways indicated that they were associated with response to Aluminum stress. Phenylpropanoid compounds produced by the phenylpropanoid and flavonoid biosynthesis metabolism pathways could play a variety of roles in plant defense [33, 34]. The phenylpropanoid compound formed complexes with Al ions and led to internal Al detoxification in Al-accumulating species [11, 12]. Kováčik et al. [35] suggested that phenolic compounds might also affect shoot Al uptake. The POD (EC 1.11.1.7) in the phenylpropanoid pathway can scavenge ROS produced by Al stress [36, 37]. Furthermore, the lignin produced by phenylpropanoid biosynthesis would affect the cell wall composition and the SIMR under Al stress [28, 38].

A total of 14 genes encoding 3 enzymes (POD (EC 1.11.1.7), enoyl-CoA hydratase, and 4-coumarate-CoA ligase) that took part in the phenylalanine metabolism pathway were affected by Al. The gene expression changes induced by Al would affect the synthesis of succinyl-CoA and in turn influenced OA synthesis in the TCA (tricarboxylic acid) cycle. An et al. [10] investigated that Al could affect the synthesis of OAs by influencing the TCA cycle, and the foliar application of succinic acid could increase the accumulation of organic acids (including oxalic acid, malic acid, citric acid, and succinic acid) and alleviate Al toxicity.

4.2. TFs Regulating the Metabolic Pathways Were Activated by Acid-Al Stress. Although the phenylpropanoid and flavonoid biosynthesis pathways in plants have been extensively studied [39], limited information is available for gene regulations involved in Al stress. Many studies have discussed the transcriptional regulation of the phenylpropanoid and flavonoid biosynthesis pathways [40]. One of genes regulating the phenylpropanoid pathway was AmMYB308, which

was found in *Antirrhinum majus* [41]. AmMYB308 repressed several phenylpropanoid pathway genes when overexpressed in transgenic tobacco. Jin et al. [42] demonstrated that AtMYB4 could inhibit the general phenylpropanoid pathway in *Arabidopsis*. This regulator was the first example of an MYB protein that functioned as a transcriptional repressor in the phenylpropanoid pathway in *Arabidopsis* [43]. Hichri et al. [40] proposed a model in which the phenylpropanoid and flavonoid biosynthesis pathways were controlled by a complex of the MYB, basic helix-loop-helix (bHLH), and WD40 proteins.

In this study, we used correlation coefficient analysis to study the possible TFs regulating the metabolic pathways activated by Al stress. MYB family genes such as MYB 305, MYB an2, and MYB apl were found to have high positive correlation coefficients with genes involved in the phenylpropanoid and flavonoid biosynthesis pathways ($r \geq 0.7$, $P \leq 0.01$; Table 3). Furthermore, MYB 305 was found to have a significant negative correlation with TT7 (flavonoid 3′-monooxygenase, A_27_P042756 (unknown); $r = -0.73$ and $P < 0.01$), revealing that this gene may suppress the expression of TT7. In *Arabidopsis*, the MYB family gene PAP1, which was homologous to MYB 305, could suppress the expression of TT7; this result was consistent with our data [44]. Many previous reports have suggested that bHLH may be a cofactor of MYB [40, 45, 46]. In our study, we found that the bHLH gene bHLH35 had a high positive correlation coefficient with many genes in the phenylpropanoid and flavonoid biosynthesis pathways (Table 3). The genes of the WRKY family (WRKY 40 and WRKY 11) were also found to have high positive correlation coefficients with genes related to lignin and flavonoid synthesis ($r > 0.7$, $P < 0.01$), indicating the possible regulatory roles of WRKY 40 and WRKY 11 in the phenylpropanoid and flavonoid biosynthesis pathways. However, no study has discussed the relationship between the WRKY family and the phenylpropanoid biosynthesis pathway to date.

5. Conclusion

Global gene expression analysis showed that acid-Al could significantly affect 3,926 gene expression. The fact of 66.33% of differentially expressed genes from roots verified that the primary target of acid-Al toxicity in plants was the root. GO enrichment and KEGG study indicated that the phenylpropanoid, flavonoid biosynthesis, and transcription factors of MYB and WRKY families were mainly involved in the response to acid-Al stress in alfalfa. Understanding the key regulatory genes and pathways would be advantageous to produce a better crop yield on acid soils and Al stress not only in alfalfa but also in other crops.

Competing Interests

There is no conflict of interests.

Authors' Contributions

Peng Zhou and Liantai Su contributed equally to this work.

Acknowledgments

The authors thank Dr. Lida Zhang (Shanghai Jiao Tong University, Shanghai) for good advice on bioinformatic analysis. This research was funded by grants from the Chinese Natural Science Foundation General Projects (nos. 31272198 and 31572451).

References

[1] B. Narasimhamoorthy, E. B. Blancaflor, J. H. Bouton, M. E. Payton, and M. K. Sledge, "A comparison of hydroponics, soil, and root staining methods for evaluation of aluminum tolerance in Medicago truncatula (barrel medic) germplasm," *Crop Science*, vol. 47, no. 1, pp. 321–328, 2007.

[2] P. Sun, Q.-Y. Tian, J. Chen, and W.-H. Zhang, "Aluminium-induced inhibition of root elongation in Arabidopsis is mediated by ethylene and auxin," *Journal of Experimental Botany*, vol. 61, no. 2, pp. 347–356, 2010.

[3] E. Klimashevskii and V. Dedov, "Localization of the mechanism of growth inhibiting action of Al3+ in elongating cell walls," *Soviet Plant Physiology*, pp. 1040–1046, 1975.

[4] S. Doncheva, M. Amenós, C. Poschenrieder, and J. Barceló, "Root cell patterning: a primary target for aluminium toxicity in maize," *Journal of Experimental Botany*, vol. 56, no. 414, pp. 1213–1220, 2005.

[5] M. Sivaguru and W. J. Horst, "The distal part of the transition zone is the most aluminum-sensitive apical root zone of maize," *Plant Physiology*, vol. 116, no. 1, pp. 155–163, 1998.

[6] P. Zhou, F. Yang, X. Ren, B. Huang, and Y. An, "Phytotoxicity of aluminum on root growth and indole-3-acetic acid accumulation and transport in alfalfa roots," *Environmental and Experimental Botany*, vol. 104, no. 1, pp. 1–8, 2014.

[7] Z. Rengel and W.-H. Zhang, "Role of dynamics of intracellular calcium in aluminium-toxicity syndrome," *New Phytologist*, vol. 159, no. 2, pp. 295–314, 2003.

[8] H. Matsumoto, "Cell biology of aluminum toxicity tolerance in higher plants," *International Review of Cytology*, vol. 200, pp. 1–46, 2000.

[9] Q.-Y. Tian, D.-H. Sun, M.-G. Zhao, and W.-H. Zhang, "Inhibition of nitric oxide synthase (NOS) underlies aluminum-induced inhibition of root elongation in Hibiscus moscheutos," *New Phytologist*, vol. 174, no. 2, pp. 322–331, 2007.

[10] Y. An, P. Zhou, Q. Xiao, and D. Shi, "Effects of foliar application of organic acids on alleviation of aluminum toxicity in alfalfa," *Journal of Plant Nutrition and Soil Science*, vol. 177, no. 3, pp. 421–430, 2014.

[11] P. Ofei-Manu, T. Wagatsuma, S. Ishikawa, and K. Tawaraya, "The plasma membrane strength of the root-tip cells and root phenolic compounds are correlated with Al tolerance in several common woody plants," *Soil Science and Plant Nutrition*, vol. 47, no. 2, pp. 359–375, 2001.

[12] B. Ezaki, K. Jayaram, A. Higashi, and K. Takahashi, "A combination of five mechanisms confers a high tolerance for aluminum to a wild species of Poaceae, *Andropogon virginicus* L," *Environmental and Experimental Botany*, vol. 93, pp. 35–44, 2013.

[13] P. S. Kidd, M. Llugany, C. Poschenrieder, B. Gunsé, and J. Barceló, "The role of root exudates in aluminium resistance and silicon-induced amelioration of aluminium toxicity in three varieties of maize (Zea mays L.)," *Journal of Experimental Botany*, vol. 52, no. 359, pp. 1339–1352, 2001.

[14] D. A. Watt, "Aluminium-responsive genes in sugarcane: identification and analysis of expression under oxidative stress," *Journal of Experimental Botany*, vol. 54, no. 385, pp. 1163–1174, 2003.

[15] 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.

[16] O. A. Hoekenga, T. J. Vision, J. E. Shaff et al., "Identification and characterization of aluminum tolerance loci in Arabidopsis (Landsberg erecta x Columbia) by quantitative trait locus mapping. A physiologically simple but genetically complex trait," *Plant Physiology*, vol. 132, no. 2, pp. 936–948, 2003.

[17] S. B. Goodwin and T. R. Sutter, "Microarray analysis of Arabidopsis genome response to aluminum stress," *Biologia Plantarum*, vol. 53, no. 1, pp. 85–99, 2009.

[18] L. Mattiello, M. Kirst, F. R. da Silva, R. A. Jorge, and M. Menossi, "Transcriptional profile of maize roots under acid soil growth," *BMC Plant Biology*, vol. 10, article 196, 2010.

[19] A. Niedziela, P. T. Bednarek, H. Cichy, G. Budzianowski, A. Kilian, and A. Anioł, "Aluminum tolerance association mapping in triticale," *BMC Genomics*, vol. 13, no. 1, article 67, pp. 1–16, 2012.

[20] K. C. Snowden and R. C. Gardner, "Five genes induced by aluminum in wheat (*Triticum aestivum* L.) roots," *Plant Physiology*, vol. 103, no. 3, pp. 855–861, 1993.

[21] 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," *Plant Journal*, vol. 41, no. 3, pp. 353–363, 2005.

[22] M. Ragland and K. Soliman, "Sali5-4a and Sali3-2, two genes induced by aluminum in soybean roots," *Plant Physiology*, vol. 114, pp. 555–560, 1997.

[23] P. G. Hartel and J. H. Bouton, "Rhizobium meliloti inoculation of alfalfa selected for tolerance to acid, aluminum-rich soils," *Plant and Soil*, vol. 116, no. 2, pp. 283–285, 1989.

[24] X.-B. Pan, C. Zhu, and C. Cheng, "Assessment of techniques for screening alfalfa cultivars for aluminum tolerance," *Euphytica*, vol. 164, no. 2, pp. 541–549, 2008.

[25] S. Wang, X. Ren, B. Huang, G. Wang, P. Zhou, and Y. An, "Aluminium-induced reduction of plant growth in alfalfa (Medicago sativa) is mediated by interrupting auxin transport and accumulation in roots," *Scientific Reports*, vol. 6, Article ID 30079, 2016.

[26] Z. Du, X. Zhou, Y. Ling, Z. Zhang, and Z. Su, "agriGO: a GO analysis toolkit for the agricultural community," *Nucleic Acids Research*, vol. 38, no. 2, pp. W64–W70, 2010.

[27] M. Kanehisa, "Toward pathway engineering: a new database of genetic and molecular pathways," *Science & Technology Japan*, vol. 59, pp. 34–38, 1996.

[28] G. Potters, T. P. Pasternak, Y. Guisez, K. J. Palme, and M. A. K. Jansen, "Stress-induced morphogenic responses: growing out of trouble?" *Trends in Plant Science*, vol. 12, no. 3, pp. 98–105, 2007.

[29] M. Kollmeier, H. H. Felle, and W. J. Horst, "Genotypical differences in aluminum resistance of maize are expressed in the distal part of the transition zone. Is reduced basipetal auxin flow involved in inhibition of root elongation by aluminum?" *Plant Physiology*, vol. 122, no. 3, pp. 945–956, 2000.

[30] Z.-B. Yang, X. Geng, C. He et al., "TAA1-regulated local auxin biosynthesis in the root-apex transition zone mediates the aluminum-induced inhibition of root growth in Arabidopsis," *Plant Cell*, vol. 26, no. 7, pp. 2889–2904, 2014.

[31] C. F. Higgins, "ABC Transporters: from microorganisms to man," *Annual Review of Cell Biology*, vol. 8, no. 1, pp. 67–113, 1992.

[32] J. Kang, J. Park, H. Choi et al., *Plant ABC Transporters*, vol. 9 of *Arabidopsis Book*, 2011.

[33] R. A. Dixon, L. Achnine, P. Kota, C.-J. Liu, M. S. S. Reddy, and L. Wang, "The phenylpropanoid pathway and plant defence—a genomics perspective," *Molecular Plant Pathology*, vol. 3, no. 5, pp. 371–390, 2002.

[34] A. Zamboni, L. Zanin, N. Tomasi et al., "Genome-wide microarray analysis of tomato roots showed defined responses to iron deficiency," *BMC Genomics*, vol. 13, no. 1, article 101, pp. 1–14, 2012.

[35] J. Kováčik, F. Štork, B. Klejdus, J. Grúz, and J. Hedbavny, "Effect of metabolic regulators on aluminium uptake and toxicity in Matricaria chamomilla plants," *Plant Physiology and Biochemistry*, vol. 54, pp. 140–148, 2012.

[36] F. Ghanati, A. Morita, and H. Yokota, "Effects of aluminum on the growth of tea plant and activation of antioxidant system," *Plant and Soil*, vol. 276, no. 1-2, pp. 133–141, 2005.

[37] M. Dahajipour Heidarabadi, F. Ghanati, and T. Fujiwara, "Interaction between boron and aluminum and their effects on phenolic metabolism of *Linum usitatissimum* L.roots," *Plant Physiology and Biochemistry*, vol. 49, no. 12, pp. 1377–1383, 2011.

[38] R. Bhardwaj, N. Handa, R. Sharma et al., "Lignins and abiotic stress: an overview," in *Physiological Mechanisms and Adaptation Strategies in Plants Under Changing Environment*, pp. 267–296, Springer, New York, NY, USA, 2014.

[39] T. Vogt, "Phenylpropanoid biosynthesis," *Molecular Plant*, vol. 3, no. 1, pp. 2–20, 2010.

[40] I. Hichri, F. Barrieu, J. Bogs, C. Kappel, S. Delrot, and V. Lauvergeat, "Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway," *Journal of Experimental Botany*, vol. 62, no. 8, pp. 2465–2483, 2011.

[41] L. Tamagnone, A. Merida, A. Parr et al., "The AmMYB308 and AmMYB330 transcription factors from antirrhinum regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco," *The Plant Cell*, vol. 10, no. 2, pp. 135–154, 1998.

[42] H. Jin, E. Cominelli, P. Bailey et al., "Trancriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis," *The EMBO Journal*, vol. 19, no. 22, pp. 6150–6161, 2000.

[43] M. R. Hemm, K. M. Herrmann, and C. Chapple, "AtMYB4: a transcription factor general in the battle against UV," *Trends in Plant Science*, vol. 6, no. 4, pp. 135–136, 2001.

[44] A. Wu, A. D. Allu, P. Garapati et al., "*JUNGBRUNNENI*, a reactive oxygen species-responsive NAC transcription factor, regulates longevity in Arabidopsis," *The Plant Cell*, vol. 24, no. 2, pp. 482–506, 2012.

[45] E. Grotewold, M. B. Sainz, L. Tagliani, J. M. Hernandez, B. Bowen, and V. L. Chandler, "Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 25, pp. 13579–13584, 2000.

[46] A. Feller, K. MacHemer, E. L. Braun, and E. Grotewold, "Evolutionary and comparative analysis of MYB and bHLH plant transcription factors," *The Plant Journal*, vol. 66, no. 1, pp. 94–116, 2011.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 2405954, 14 pages
http://dx.doi.org/10.1155/2016/2405954

Review Article

Role of Recombinant DNA Technology to Improve Life

Suliman Khan,[1] Muhammad Wajid Ullah,[2] Rabeea Siddique,[3] Ghulam Nabi,[1] Sehrish Manan,[4] Muhammad Yousaf,[5] and Hongwei Hou[1]

[1]*The Key Laboratory of Aquatic Biodiversity and Conservation of Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China*
[2]*Department of Biomedical Engineering, Huazhong University of Science and Technology, Wuhan 430074, China*
[3]*Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Peshawar 25000, Pakistan*
[4]*National Key Laboratory of Crop Genetic Improvement, College of Plant Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, China*
[5]*Center for Human Genome Research, Cardio-X Institute, Huazhong University of Science and Technology, Wuhan 430074, China*

Correspondence should be addressed to Hongwei Hou; houhw@ihb.ac.cn

Received 10 August 2016; Revised 21 October 2016; Accepted 6 November 2016

Academic Editor: Wenqin Wang

In the past century, the recombinant DNA technology was just an imagination that desirable characteristics can be improved in the living bodies by controlling the expressions of target genes. However, in recent era, this field has demonstrated unique impacts in bringing advancement in human life. By virtue of this technology, crucial proteins required for health problems and dietary purposes can be produced safely, affordably, and sufficiently. This technology has multidisciplinary applications and potential to deal with important aspects of life, for instance, improving health, enhancing food resources, and resistance to divergent adverse environmental effects. Particularly in agriculture, the genetically modified plants have augmented resistance to harmful agents, enhanced product yield, and shown increased adaptability for better survival. Moreover, recombinant pharmaceuticals are now being used confidently and rapidly attaining commercial approvals. Techniques of recombinant DNA technology, gene therapy, and genetic modifications are also widely used for the purpose of bioremediation and treating serious diseases. Due to tremendous advancement and broad range of application in the field of recombinant DNA technology, this review article mainly focuses on its importance and the possible applications in daily life.

1. Introduction

Human life is greatly affected by three factors: deficiency of food, health problems, and environmental issues. Food and health are basic human requirements beside a clean and safe environment. With increasing world's population at a greater rate, human requirements for food are rapidly increasing. Humans require safe-food at reasonable price. Several human related health issues across the globe cause large number of deaths. Approximately 36 million people die each year from noncommunicable and communicable diseases, such as cardiovascular diseases, cancer, diabetes, AIDS/HIV, tuberculosis, malaria, and several others according to http://GlobalIssues.org/. Despite extensive efforts being made, the current world food production is much lower than human requirements, and health facilities are even below standard in the third-world countries. Rapid increase in industrialization has soared up the environmental pollution and industrial wastes are directly allowed to mix with water, which has affected aquatic marines and, indirectly, human-beings. Therefore, these issues urge to be addressed through modern technologies.

Unlike tradition approaches to overcome agriculture, health, and environmental issues through breeding, traditional medicines, and pollutants degradation through conventional techniques respectively, the genetic engineering utilizes modern tools and approaches, such as molecular cloning and transformation, which are less time consuming and yield more reliable products. For example, compared to conventional breeding that transfers a large number of both specific and nonspecific genes to the recipient, genetic engineering only transfers a small block of desired genes to

the target through various approaches, such as biolistic and Agrobacterium-mediated transformation [1]. The alteration into plant genomes is brought either by homologous recombination dependent gene targeting or by nuclease-mediated site-specific genome modification. Recombinase mediated site-specific genome integration and oligonucleotide directed mutagenesis can also be used [2].

Recombinant DNA technology is playing a vital role in improving health conditions by developing new vaccines and pharmaceuticals. The treatment strategies are also improved by developing diagnostic kits, monitoring devices, and new therapeutic approaches. Synthesis of synthetic human insulin and erythropoietin by genetically modified bacteria [3] and production of new types of experimental mutant mice for research purposes are one of the leading examples of genetic engineering in health. Likewise, genetic engineering strategies have been employed to tackle the environmental issues such as converting wastes into biofuels and bioethanol [4–7], cleaning the oil spills, carbon, and other toxic wastes, and detecting arsenic and other contaminants in drinking water. The genetically modified microbes are also effectively used in biomining and bioremediation.

The advent of recombinant DNA technology revolutionized the development in biology and led to a series of dramatic changes. It offered new opportunities for innovations to produce a wide range of therapeutic products with immediate effect in the medical genetics and biomedicine by modifying microorganisms, animals, and plants to yield medically useful substances [8, 9]. Most biotechnology pharmaceuticals are recombinant in nature which plays a key role against human lethal diseases. The pharmaceutical products synthesized through recombinant DNA technology, completely changed the human life in such a way that the U.S. Food and Drug Administration (FDA) approved more recombinant drugs in 1997 than in the previous several years combined, which includes anemia, AIDS, cancers (Kaposi's sarcoma, leukemia, and colorectal, kidney, and ovarian cancers), hereditary disorders (cystic fibrosis, familial hypercholesterolemia, Gaucher's disease, hemophilia A, severe combined immunodeficiency disease, and Turnor's syndrome), diabetic foot ulcers, diphtheria, genital warts, hepatitis B, hepatitis C, human growth hormone deficiency, and multiple sclerosis. Considering the plants develop multigene transfer, site-specific integration and specifically regulated gene expression are crucial advanced approaches [10]. Transcriptional regulation of endogenous genes, their effectiveness in the new locations, and the precise control of transgene expression are major challenges in plant biotechnology which need further developments for them to be used successfully [11].

Human life is greatly threatened by various factors, like food limitations leading to malnutrition, different kinds of lethal diseases, environmental problems caused by the dramatic industrialization and urbanization and many others. Genetic engineering has replaced the conventional strategies and has the greater potential to overcome such challenges. The current review summarized the major challenges encountered by humans and addresses the role of recombinant DNA technology to overcome aforementioned issues. In line with this, we have detailed the limitations of genetic engineering and possible future directions for researchers to surmount such limitations through modification in the current genetic engineering strategies.

2. Recombinant DNA Technology

Recombinant DNA technology comprises altering genetic material outside an organism to obtain enhanced and desired characteristics in living organisms or as their products. This technology involves the insertion of DNA fragments from a variety of sources, having a desirable gene sequence *via* appropriate vector [12]. Manipulation in organism's genome is carried out either through the introduction of one or several new genes and regulatory elements or by decreasing or blocking the expression of endogenous genes through recombining genes and elements [13]. Enzymatic cleavage is applied to obtain different DNA fragments using restriction endo-nucleases for specific target sequence DNA sites followed by DNA ligase activity to join the fragments to fix the desired gene in vector. The vector is then introduced into a host organism, which is grown to produce multiple copies of the incorporated DNA fragment in culture, and finally clones containing a relevant DNA fragment are selected and harvested [11]. The first recombinant DNA (rDNA) molecules were generated in 1973 by Paul Berg, Herbert Boyer, Annie Chang, and Stanley Cohen of Stanford University and University of California San Francisco. In 1975, during "The Asilomar Conference" regulation and safe use of rDNA technology was discussed. Paradoxically to the view of scientists at the time of Asilomar, the recombinant DNA methods to foster agriculture and drug developments took longer than anticipated because of unexpected difficulties and barriers to achieve the satisfactory results. However, since the mid-1980s, the number of products like hormones, vaccines, therapeutic agents, and diagnostic tools has been developed continually to improve health [13].

A quick approach is offered by recombinant DNA technology to scrutinize the genetic expression of the mutations that were introduced into eukaryote genes through cloned insulin genes insertion inside a simian virus fragment [3]. In a similar way, tumor growth was inhibited by adenoviral vector that encodes endostain human secretory form through antiangiogenic effects. Antiangiogenic effect can be enhanced by *dl*1520 through rescuing replication of Ad-Endo [14]. Targeted gene disruption has been used to produce antitumor derivatives in other hosts which were structurally similar for the production pathways [15]. Besides, longer acting therapeutic proteins have been developed through recombinant DNA technologies; for example, sequences containing additional glycosylation site are one of the most followed approaches. A new chimeric gene has been developed through this technique which contains the FSH β-subunit coding sequences and the C-terminal peptide of the hCG β-subunit coding sequences [16]. Researchers have also developed vectors and combined vectors for gene therapy and genetic modification approaches. Presently, viral vectors have received immense consideration in clinical settings, some of which have also been commercialized. In principle, viruses are modified to be safe for clinical purposes. They have several

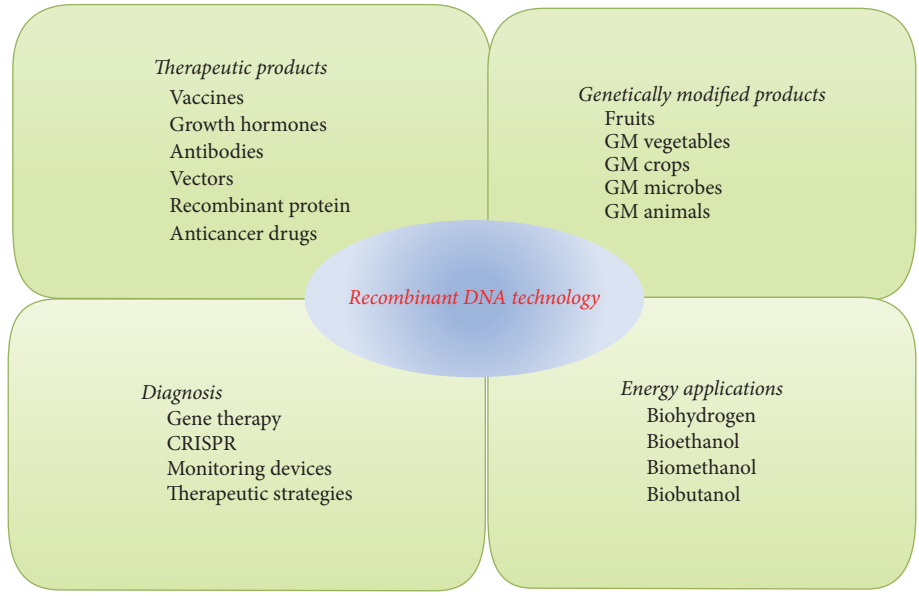


Figure 1: Illustration of various applications of recombinant DNA technology.

applications including treatment of severe diseases including cancer either through in vivo or gene therapy (ex vivo), vaccination, and protein transduction approaches [17]. The production of clinical grade viral vectors improvement has become possible due to advance manufacturing technologies [18]. At present, due to the severe adverse effects, retroviral vectors are losing their importance although the viral entities transfer genes quickly and correctly into a number of species. The simplest nonviral gene delivery system uses "naked" DNA, when injected directly into certain tissues, particularly muscles, produces significant levels of gene expression with least side effects [19]. More recently, a P1 vector has been designed to introduce the recombinant DNA into *E. coli* through electroporation procedures. This new cloning system is used for establishing 15,000 clone library initially averagely 130–150 kb pairs insert size. PAC cloning system is considered useful for complex genome analysis and in mapping [20]. The construction of low copy number vectors, for example, pWSK29, pWKS30, pWSK129, and pWKS130, was carried out using PCR and recombinant DNA technology. These vectors can also be used for generating unidirectional deletions with exonuclease, complementation analysis, DNA sequencing, and run-off transcription [21]. A broad range of applications of recombinant DNA technology has been summarized in Figure 1.

3. Current Research Progress

Recombinant DNA technology is a fast growing field and researchers around the globe are developing new approaches, devices, and engineered products for application in different sectors including agriculture, health, and environment. For example, Lispro (Humalog), in comparison with regular human insulin, is a well effective and fast acting recombinant insulin [3]. Similarly, Epoetin alfa is a novel and well-recognized recombinant protein that can be effectively used in curing of anemia [22]. Recombinant hGH was found with a great improvement in treating children lacking the ability to produce hGH in a required quantity. Clinical testing approval by the FDA in December 1997 for a recombinant version of the cytokine myeloid progenitor inhibitory factor-1 (MPIF-1) was an achievement to give recognition to this technology. With its help anticancer drug's side effects can be mitigated whereas it has the ability to mimic the division of immunologically important cells [23, 24]. The following section summarizes the most recent developments of recombinant DNA technology.

Clustered regularly interspaced short palindromic repeats (CRISPR), a more recent development of recombinant DNA technology, has brought out solutions to several problems in different species. This system can be used to target destruction of genes in human cells. Activation, suppression, addition, and deletion of genes in human's cells, mice, rats, zebrafish, bacteria, fruit flies, yeast, nematodes, and crops proved the technique a promising one. Mouse models can be managed for studying human diseases with CRISPR, where individual genes study becomes much faster and the genes interactions studies become easy by changing multiple genes in cells [25]. The CRISPR of *H. hispanica* genome is capable of getting adapted to the nonlytic viruses very efficiently. The associated Cas operon encodes the interfering Cas3 nucleases and other Cas proteins. The engineering of a strain is required with priming CRISPR for priming crRNAs production and new spacers acceptance. CRISPR-cas system has to integrate new spacers into its locus for adaptive immunity generation [26]. Recognition of foreign DNA/RNA and its cleavage is a controlled process in sequence-specific manner. Information related to the intruder's genetic material is stored by the host system with the help of photo-spacer incorporation into the CRISPR system [27]. Cas9t (gene

editing tool) represents DNA endonucleases which use RNA molecules to recognize specific target [28]. Class 2 CRISPR-Cas system with single protein effectors can be employed for genome editing processes. Dead Cas9 is important for histone modifying enzyme's recruitment, transcriptional repression, localization of fluorescent protein labels, and transcriptional activation [29]. Targeting of genes involved in homozygous gene knockouts isolation process is carried out by CRISPR-induced mutations. In this way, essential genes can be analyzed which in turn can be used for "potential antifungal targets" exploration [30]. Natural CRISPR-cas immunity exploitation has been used for generation of strains which are resistant to different types of disruptive viruses [31].

CRISPR-Cas, the only adaptive immune system in prokaryotes, contains genomic locus known as CRISPR having short repetitive elements and spacers (unique sequences). CRISPR array is preceded by AT-rich leader sequence and flanked by cas genes which encode Cas proteins [32, 33]. In *Escherichia coli* cas1 and cas2 catalases promote new spacers through complex formation. Photo-spacer adjacent motif (PAM) is required for interference and acquisition because the target sequence selection is not random. The memorization of the invader's sequence starts after CRISPR array transcription into long precursor crRNA. During the final stages of immunity process, target is degraded through interference with invaded nucleic acids. Specific recognition prevents the system from self-targeting [32, 34]. In different species of *Sulfolobus*, the CRISPR loci contain multiple spacers whose sequence matches conjugative plasmids significantly while in some cases the conjugative plasmids also contain small CRISPR loci. Spacer acquisition is affected by active viral DNA replication in *Sulfolobus* species whereas the DNA breaks formation at replication forks causes the process to be stimulated [35]. According to the above information, CRISPR-Cas system has obtained a unique position in advanced biological systems because of its tremendous role in the stability and enhancement of immunity.

Zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) are chimeric nucleases composed of programmable, sequence-specific DNA-binding modules linked to a nonspecific DNA cleavage domain. Therapeutic potential of ZFNs and TALENs is more specified and targeted [25, 36, 37]. Similarly, recombinant protein fibroblast growth factor (FGF-1) has been developed which functions in inducing the formation of new blood vessels in myocardium. Its injection (biologic bypass) into a human myocardium cause an increased blood supply to the heart. Apligraf, an FDA approved product, which serves as a recombinant skin replacer, specified for the leg ulcer's treatment and Derma-Graft, is effective in the treatment of diabetic ulcers [38–40]. After successful production of insulin from *E. coli* through recombinant DNA technology, currently several animals, notably cattle and pigs, have been selected as insulin producing source, which however, triggered immune responses. The recombinant human insulin is identical to human porcine insulin and comparatively infrequently elicits immunogenic responses. Furthermore, it is more affordable and can satisfy medical needs more readily. Human growth hormone was the first protein expressed in tobacco plants [41, 42]. Besides insulin, several new drugs related to recombinant DNA technology have undergone developmental improvements and a number of protein production systems have been developed. Several engineered microbial strains have been developed to carry out the formulation of drugs [41, 43, 44]. Molecular medicine formation that is specifically based on proteins faces serious issues including methods and biology of the cells which function to produce medically important compounds through recombinant DNA techniques. To overcome these obstacles, there is intense need to improve quality and quantity of medicines based on a molecular phenomenon. Cell factories are considered important in recombinant DNA technologies, but these needed to be explored with more details and in depth as the conventional factories are not fulfilling the needs [42]. Similarly, the endothelial growth factor and Notch signaling were used to engineer oncolytic adenovirus which acts as a breast cancer selective agent for the antagonist's expression. This further, through tumor angiogenesis disruption acts as anticancer agent. This decreases the total blood vessels numbers and causes a dramatic change along with the perfused vessels which indicates the improved efficacy against the tumor and vascular effects [13]. Efforts have been made to modify the influenza virus genome using recombinant DNA technology for development of vaccines. The modifications are based on engineering of vectors to expression of foreign genes. In practical, the NS gene of the influenza virus was replaced with foreign gene, commonly chloramphenicol acetyltransferase gene. Thereafter, the RNA previously recombined is expressed and packaged into virus particles after transfection with purified influenza A virus in the presence of helper virus. It has been clarified that $5'$ terminal and the $3'$ terminal bases are sufficient from influenza A virus RNA to produce signals for RNA replication, RNA transcription, and RNA packaging into influenza virus [15].

The abovementioned new production systems enhance pipelines for development of various vaccines and drugs and so forth. Production of high quality proteins depends on physiology of a cell and the conditions provided to it. The expression of proteins becomes retarded if a cell goes under stressful conditions, which may also favor the production in some cases. Thus, further improvements are required for the better and safe production at genetic and metabolic levels. Microorganisms are considered the most convenient hosts to produce molecular medicines. These cells allow the incorporation of foreign genes with less resistant barriers and expression is easily controlled. Compared to plant and mammalian cells to be taken as hosts, microbial systems provide less complicated machinery which ultimately enhances the performance and quality of proteins production. The use of common microbial species, including bacteria and yeasts, is promising but the less common strains have also been observed promising as being cellular factories to produce recombinant molecular drugs. The increasing demands of drugs and the needs of quality can be fulfilled with better results if these cellular factories of microorganisms get incorporated into productive processes of pharmaceuticals (Table 1) [41, 45, 46].

Table 1: Current DNA assembly methods for the synthesis of large DNA molecules. The table has been reproduced from Nature reviews 14: 781–793, with permission from Nature Publishing Group.

Method	Mechanism	Overhang (bp)	Scar (bp)	Comments	Examples of applications
BioBricks	Type IIP restriction endonuclease	8	8	Sequentially assembles small numbers of sequences	Construction of a functional gene expressing enhanced cyan fluorescent protein
BglBricks	Type IIP restriction endonuclease	6	6	Uses a highly efficient and commonly used restriction endonuclease, the recognition sequences of which are not blocked by the most common DNA methylases	Construction of constitutively active gene-expression devices and chimeric, multidomain protein fusions
Pairwise selection	Type IIS restriction endonuclease	65	4	Requires attachment tags at each end of fragments to act as promoters for antibiotic resistance markers; rapid, as a liquid culture system is used	Assembly of a 91 kb fragment from 1-2 kb fragments
GoldenGate	Type IIS restriction endonuclease	4	0	Allows large-scale assembly; ligations are done in parallel one-step assembly of 2-3 fragment	One-step assembly of 2-3 fragments
Overlapping PCR	Overlap	0	0	Uses overlapping primers for the PCR amplification of 1–3 kb-long fragments	Usually used for 1–3 kb-long fragments, for example, for gene cassette construction
CPEC	Overlap	20–75	0	Uses a single polymerase for the assembly of multiple inserts into any vector in a one-step reaction in vitro	One-step assembly of four 0.17–3.2 kb-long PCR fragments
Gateway	Overlap	20	0	Uses a specific recombinase for small-scale assembly	One-step assembly of three 0.8–2.3 kb-long fragments
USER	Overlap	Up to 708	0	Replaces a thymidine with a uracil in the PCR primers, which leaves 3′ overhangs for cloning after cleaving by a uracil exonuclease	One-step assembly of three 0.6–1.5 kb-long fragments
InFusion	Overlap	15	0	Uses an enzyme mix for parallel assembly through a "chew-back-and-anneal" method	One-step assembly of three 0.2–3.8 kb-long fragments
SLIC	Overlap	>30	0	(i) Uses a T4 DNA polymerase through a chew-back method in the absence of dNTPs (ii) Uses Recombinase A* to stabilize the annealed fragments and avoid in vitro ligation (iii) Allows the parallel assembly of several hundred base-long fragments	Generation of a ten-way assembly of 300–400 bp-long PCR fragments
Gibson	Overlap	40–400	0	Uses enzymatic "cocktails" to chew back and anneal for the parallel assembly of several kilobase-long fragments	Assembly of the 1.08 Mb *Mycoplasma mycoides* JCVI-syn1.0 genome

4. Applications of Recombinant DNA Technology

4.1. Food and Agriculture. Recombinant DNA technology has major uses which made the manufacturing of novel enzymes possible which are suitable in conditions for specified food-processing. Several important enzymes including lipases and amylases are available for the specific productions because of their particular roles and applications in food industries. Microbial strains production is another huge achievement that became possible with the help of recombinant DNA technology. A number of microbial strains have been developed which produce enzyme through specific engineering for production of proteases. Certain strains of fungi have been modified so that their ability of producing toxic materials could be reduced [47]. Lysozymes are the effective agents to get rid of bacteria in food industries. They prevent the colonization of microbial organisms. It is suitable agent for food items including fruits, vegetables, cheese, and meat to be stored as it increases their shelf life. The inhibition of food spoiling microorganisms can be carried out through immobilized lysozyme in polyvinyl alcohol films and cellulose. Lysozyme impregnation of fish skin gelatin gels increase the shelf life of food products and inhibit different food spoiling bacterial growth [48–50]. Exopolysaccharides of *Staphylococcus* and *E. coli* can be hydrolyzed with the use of DspB which is engineered from T7. This ability of DspB causes a declination in the bacterial population [50]. Biofilms related to food industries can be removed by the combining activity of serine proteases and amylases [51]. *S. aureus*, *Salmonella infantis*,

Clostridium perfringens, *B. cereus*, *Campylobacter jejuni*, *L. monocytogenes*, *Yersinia enterocolitica*, and some other food spoiling microorganisms can be inhibited by glucose oxidase. It is also considered one of the most important enzymes in food industry to kill wide range of foodborne pathogens [50].

Derivation of recombinant proteins being used as pharmaceuticals came into practice from first plant recently and many others are through to be used for more production of similar medically important proteins [52].

Wide range of recombinant proteins have been expressed in different plant species to be used as enzymes in industries, some majorly used proteins in research are proteins present in milk which play a role in nutrition, and new polymeric proteins are being used in industries and medical field [52]. With the invention of HBV vaccine production in plants, the oral vaccination concept with edible plants has gained popularity. Plants have been used to produce several therapeutic protein products, such as casein and lysozyme for improving health of child and polymers of protein for tissue replacement and surgery. Furthermore, tobacco plants can be engineered genetically to produce human collagen. High yielding molecular proteins is one of the major tasks under consideration in field of recombinant DNA technology [52]. Traditional breeding and quantitative trade locus (QTL) analysis assisted in the identification of a rice variety with protein kinase known as PSTOL1 (*phosphorus starvation tolerance1*) help in enhancing root growth in early stages and tolerates phosphorus deficiency [53]. Overexpression of this enzyme enables root to uptake nutrients in sufficient amount in phosphorus deficient soil which ultimately enhances the grain yield [54]. Chloroplast genome sequences are important in plant evolution and phylogeny. *Rpl22* is considered to be transferred from chloroplast into nuclear genome. This gene contains a peptide which plays role in delivery of protein from cytosol to chloroplast. A number of important genes deleted from chloroplast have been observed to be transferred into nucleus, except ycf1 and ycf2, in order to avoid disruptions in photosynthesis and other necessary processes. Trans-genesis into chloroplast is considered stable as the nuclear transgenic plants face the problems of lower expression and transgene escape via pollen. Almost ten thousand copies of transgenes have been incorporated into the genome of chloroplast [55–57]. Transgene expression is dependent on heterologous regulatory sequences but independent of cellular control. T7gene10 engineering against salt stress has been found successful but with lower expression rate into nongreen tissues. γ-tmt gene insertion into chloroplast genome results in multiple layer formation of the inner chloroplast envelope. Lycopene β-cyclase genes introduction into the plastid genome of tomato enhances the lycopene conversion into provitamin A [57, 58].

Organ or tissue specific genes identification can be carried out through gene expression profiles. cDNAs with full lengths are the main resources for expression profiling of genes. 44 K Agilent Oligonucleotide microarray is used for field grown rice transcriptome analysis. Gene expression fluctuation and transcriptome dynamics can be predicted by transcriptomic data and meteorological information. These processes and predictions are helpful to improve crop production and resistance to either environmental or microbial stresses. Resistance to fungal and bacterial infections can be enhanced by WRKY45 gene in rice which is induced by plant activator benzothiadiazole that activates innate immune system of plant. The larger grain size can be achieved by inserting qSW5 gene. qSH1 causes the loss of seed shattering by preventing the abscission layer formation. Kala4 gene is responsible for the black color of rice which makes the rice resistant to attacking pathogens [59, 60]. Genetic modification is needed in facilitating gene by gene introduction of well-known characters. It allows access to extended range of genes from an organism. Potato, beans, eggplant, sugar beet, squash, and many other plants are being developed with desirable characters, for example, tolerance of the herbicide glyphosate, resistance to insects, drought resistance, disease and salt tolerance. Nitrogen utilization, ripening, and nutritional versatility like characters have also been enhanced [61].

4.2. Health and Diseases. Recombinant DNA technology has wide spectrum of applications in treating diseases and improving health conditions. The following sections describe the important breakthroughs of recombinant DNA technology for the improvement of human health:

4.2.1. Gene Therapy. Gene therapy is an advanced technique with therapeutic potential in health services. The first successful report in field of gene therapy to treat a genetic disease provided a more secure direction toward curing the deadliest genetic diseases [62, 63]. This strategy shows good response in providing treatment for adenosine deaminase-deficiency (ADA-SCID), which is a primary immunodeficiency. At the beginning of this technology, several challenges including maintenance of patients on PEGylated ADA (PEG-ADA) during gene therapy and the targeting of gene transfer to T-lymphocytes were the reasons for unsuccessful results [64, 65]. However, later on successful results were obtained by targeting haematopoietic stem cells (HSCs) by using an improved gene transfer protocol and a myeloablative conditioning regime [66].

Adrenoleukodystrophy (X-ALD) and X-linked disorder are is possible through the expression of specific genes transferred by lentiviral vector, based on HIV-1 [67]. X-ALD protein expression indicates that gene-correction of true HSCs was achieved successfully. The use of lentiviral vector was made successful for the first time to treat genetic human disease [68]. Metastatic melanoma was treated through immunotherapy by enhancing the specific proteins expression during 2006. This success in the field of health sciences opened up new doors to extend the research to treat serious death causing diseases through immunotherapy [69]. Highly sustained levels of cells that were engineered for tumor recognition in blood using a retrovirus encoding a T-cell receptor in two patients up to 1 year after infusion resulted in regression of metastatic melanoma lesions. This strategy was later used to treat patients with metastatic synovial cell carcinoma [70]. Autologous T-cells were genetically modified to express a Chimeric Antigen Receptors (CAR) with specificity for the B-cell antigen CD19 for the treatment of chronic lymphocytic leukemia. Genetically modified cells undergo

selective expansion for diseases such as SCID-X1 and ADA-SCID as a consequence of in vivo selection conferred by the disease pathophysiology despite the correction of only a modest number of progenitors. Combination of gene and drug therapy's potential has recently been highlighted in a trial seeking to confer chemoprotection on human HSCs during chemotherapy with alkylating agents for glioblastoma [71].

Gene transfer to a small number of cells at anatomically discrete sites is a targeted strategy that has the potential to confer therapeutic benefit. It showed impressive results for incurable autosomal recessive dystrophies such as congenital blindness and Leber congenital amaurosis (LCA). Swiss–German phase I/II gene therapy clinical trial aimed to treat chronic granulomatous disease in April 2006 that came up with success [72]. Mobilized CD34+ cells isolated from peripheral blood were retrovirally transduced and infused into the patient where two-thirds of the patients showed clear benefit from this treatment. After the treatment silencing of the transgene as a result of methylation of the viral promoter caused the severity of infection that leaded to the death of patient [73].

Many different cancers including lung, gynecological, skin, urological, neurological, and gastrointestinal tumors, as well as hematological malignancies and pediatric tumors, have been targeted through gene therapy. Inserting tumor suppressor genes to immunotherapy, oncolytic virotherapy and gene directed enzyme prodrug therapy are different strategies that have been used to treat different types of cancers. The p53, a commonly transferred tumor suppressor gene, is a key player in cancer treating efforts. In some of the strategies, p53 gene transfer is combined with chemotherapy or radiotherapy. The most important strategies that have been employed until now are vaccination with tumor cells engineered to express immunostimulatory molecules, vaccination with recombinant viral vectors encoding tumor antigens and vaccination with host cells engineered to express tumor antigens [19]. New fiber chimeric oncolytic adenoviruses vectors (Ad5/35-EGFP) offer an affective new anticancer agent for the better cure of hepatocellular carcinoma. A demonstration of these vectors through proper assaying was significant for transduction improvement and more progeny of the virus were produced in HCC. A higher level of transgenic expression was mediated and an enhanced antitumor effect was observed on in vitro HCC cells while keeping the normal cells protected against cytotoxicity. Tumor growth was also inhibited by utilizing this technology [74]. Cancer gene therapy has become more advanced and its efficacy has been improved in recent years [75].

Treatment of cardiovascular diseases by gene therapy is an important strategy in health care science. In cardiovascular field, gene therapy will provide a new avenue for therapeutic angiogenesis, myocardial protection, regeneration and repair, prevention of restenosis following angioplasty, prevention of bypass graft failure, and risk-factor management. Mutation in gene encoding WASP, a protein regulating the cytoskeleton, causes Wiskott-Aldrich Syndrome (inherited immunodeficiency). Its treatment requires stem cells transplantation; in case matched donors are unavailable the treatment is carried out through infusion of autologous HSPCs modified ex vivo by gene therapy [76]. Metastatic cancer can be regressed through immunotherapy based on the adoptive transfer of gene-engineered T-cells. Accurate targeting of antigens expressed by tumors and the associated vasculature and the successful use of gene engineering to retarget T-cells before their transfer into the patient are mainly focused on in this therapy [77]. Cancer cells often make themselves almost "invisible" to the immune system and its microenvironment suppresses T-cells survival and migration but genetic engineering of T-cells is the solution to these challenges. T-cells in cancer patients can be modified by recombining the genes responsible for cancer-specific antigens recognition, resistance to immunosuppression, and extending survival and facilitating migration to tumors [78]. Fusion between the genes echinoderm microtubule-associated protein like 4 (*EML4*) and anaplastic lymphoma kinase (*ALK*) is generated by an inversion on the short arm of chromosome confers sensitivity to ALK inhibitors. Vial-mediated delivery of the CRISPR/Cas9 system to somatic cells of adult animals induces specific chromosomal rearrangements [79].

Wnt signaling is one of the key oncogenic pathways in multiple cancers. Targeting the Wnt pathway in cancer is an attractive therapeutic approach, where LGK974 potently inhibits Wnt signaling, has strong efficacy in rodent tumor models, and is well-tolerated. Head and neck cancer cell lines with loss-of-function mutations in the Notch signaling pathway have a high response rate to LGK974 [80]. Codon-optimized gene, on the basis of coding sequence of the influenza virus hemagglutinin gene, was synthesized and cloned into a recombinant modified vaccinia virus Ankara (MVA). Immunization with MVA-H7-Sh2 viral vector in ferrets proved to be immunogenic as unprotected animals that were mock vaccinated developed interstitial pneumonia and loss of appetite and weight but vaccination with MVA-H7-Sh2 protected the animals from severe disease [81]. Viral gene therapy is one of the leading and important therapies for head and neck cancer. Tumor-associated genes are targeted by viruses, and p53 gene function was targeted through such therapy at first. Cancer cells can be destroyed by oncolytic viruses through viral replication and by arming with therapeutic transgenes [82].

High density lipoprotein gene *ABCA1* mutation in cells can make the cells be differentiated into macrophages. Gene knockouts in embryonic stem cells enhance the capability of cells to be differentiated into macrophages and specifically target the desired pathogens. The allele replacements in this case will assist in studying protein coding changes and regulatory variants involved in alteration of mRNA transcription and stability in macrophages [83].

4.2.2. Production of Antibodies and Their Derivatives. Plant systems have been recently used for the expression and development of different antibodies and their derivatives. Most importantly, out of many antibodies and antibody derivatives, seven have reached to the satisfactory stages of requirements. Transgenic tobacco plants can be used for the production of chimeric secretory IgA/G known as CaroRx, CaroRx. Oral pathogen responsible for decay of a tooth known as Streptococcus mutants, can be recognized

by this antibody. A monoclonal antibody called T84.66 can affectively function to recognize antigen carcinoembryonic, which is still considered an affectively characterized marker in cancers of epithelia [84, 85]. A full-length humanized IgG1 known as anti-HSV and anti-RSV, which can function as the recognizing agent for herpes simplex virus (HSV)-2-glycoprotein B, has been expressed in transgenic soybean and Chinese Hamster Ovary (CHO) cells. Antibodies from both sources have been shown to prevent vaginal HSV-2 transmission in mice after applying topically; if worked similarly in humans it would be considered as inexpensive and affective prevention against diseases transmitted through sexual interactions [86–88]. *38C13* is scFv antibody based on the idiotype of malignant B lymphocytes in the well-characterized mouse lymphoma cell line 38C13. Administration of the antibody to mice resulted in the production of anti-idiotype antibodies that are able to recognize 38C13 cells, which help to protect the mice against with injected lymphoma cells, is a lethal challenge [89, 90]. Unique markers recognizing enzymes could be produced through this system, most affectively the surface markers of a malignant B-cells to work as an effective therapy for non-Hodgkin lymphoma like diseases in human [61]. A monoclonal antibody known as PIPP is specific for human chorionic gonadotropin recognition. The production of full-length monoclonal antibody and scFv and diabody derivatives was made possible in plants through transgenesis and agroinfiltration in tobacco transformed transiently [91]. Testosterone production by stimulated hCG can be inhibited by each of these antibodies in cells cultured by LEYDIG and uterine weight gain could be delayed in mice, through which hCG activity is checked. Diagnosis and therapy of tumors can be carried out with the help of antibodies [61].

4.2.3. Investigation of the Drug Metabolism. Complex system of drug metabolizing enzymes involved in the drug metabolism is crucial to be investigated for the proper efficacy and effects of drugs. Recombinant DNA approaches have recently contributed its role through heterologous expression, where the enzyme's genetic information is expressed in vitro or in vivo, through the transfer of gene [92, 93].

4.2.4. Development of Vaccines and Recombinant Hormones. Comparatively conventional vaccines have lower efficacy and specificity than recombinant vaccine. A fear free and painless technique to transfer adenovirus vectors encoding pathogen antigens is through nasal transfer which is also a rapid and protection sustaining method against mucosal pathogens. This acts as a drug vaccine where an anti-influenza state can be induced through a transgene expression in the airway [74].

In vitro production of human follicle-stimulating hormone (FSH) is now possible through recombinant DNA technology. FSH is considerably a complex heterodimeric protein and specified cell line from eukaryotes has been selected for its expression. Assisted reproduction treatment through stimulating follicular development is an achievement of recombinant DNA technology. A large number of patients are being treated through r-FSH. Most interestingly r-FSH and Luteinizing Hormone (LH) recombination was made successful to enhance the ovulation and pregnancy [94, 95].

4.2.5. Chinese Medicines. As an important component of alternative medicine, traditional chines medicines play a crucial role in diagnostics and therapeutics. These medicines associated with theories which are congruent with gene therapy principle up to some extent. These drugs might be the sources of a carriage of therapeutic genes and as coadministrated drugs. Transgenic root system has valuable potential for additional genes introduction along with the Ri plasmid. It is mostly carried with modified genes in *A. rhizogenes* vector systems to enhance characteristics for specific use. The cultures became a valuable tool to study the biochemical properties and the gene expression profile of metabolic pathways. The intermediates and key enzymes involved in the biosynthesis of secondary metabolites can be elucidated by the turned cultures [96, 97].

4.2.6. Medically Important Compounds in Berries. Improvement in nutritional values of strawberries has been carried through rolC gene. This gene increases the sugar content and antioxidant activity. Glycosylation of anthocyanins requires two enzymes glycosyl-transferase and transferase. Some nutrition related genes for different components in strawberry including proanthocyanidin, l-ascorbate, flavonoid, polyphenols, and flavonoid are important for improving the component of interest through genetic transformation. In case of raspberry, bHLH and FRUITE4 genes control the anthocyanin components whereas ERubLRSQ072H02 is related to flavonol. By specific transformation, these genes can enhance the production and improve the quality. All these mentioned compounds have medical values [98].

4.3. Environment. Genetic engineering has wide applications in solving the environmental issues. The release of genetically engineered microbes, for example, *Pseudomonas fluorescens* strain designated HK44, for bioremediation purposes in the field was first practiced by University of Tennessee and Oak Ridge National Laboratory by working in collaboration [99, 100]. The engineered strain contained naphthalene catabolic plasmid pUTK21 [101] and a transposon-based bioluminescence-producing *lux* gene fused within a promoter that resulted in improved naphthalene degradation and a coincident bioluminescent response [102]. HK44 serves as a reporter for naphthalene bioavailability and biodegradation whereas its bioluminescence signaling ability makes it able to be used as an online tool for in situ monitoring of bioremediation processes [102]. The production of bioluminescent signal is detectable using fiber optics and photon counting modules [101].

4.3.1. Phytoremediation and Plant Resistance Development. Genetic engineering has been widely used for the detection and absorption of contaminants in drinking water and other samples. For example, At*PHR1* gene introduction into garden plants *Torenia*, *Petunia*, and *Verbena* changed their ability for Pi absorption. The At*PHR1* transgenic plants with enhanced Pi absorption ability can possibly facilitate effective phytoremediation in polluted aquatic environments [103]. A fragment of the At*PHR1* gene was inserted into binary vector pBinPLUS, which contains an enhanced cauliflower mosaic

virus 35S promoter. This plasmid was named pSPB1898 and was used for transformation [104] in Petunia and Verbena using *Agrobacterium tumefaciens* [105]. At*PHR1* is effective in other plant species, such as *Torenia*, *Petunia*, and *Verbena* [103] but posttranscriptional modification of the endogenous At*PHR1* counterpart might be inhibited by overexpression of At*PHR1* [103].

Plant metabolism processes identify their importance to use for remediating the environmental pollutants. Some of the chemicals are not prone to be degraded or digested. TNT is only partially digested in which the nitrogen further reacts with oxygen to form toxic superoxide. To overcome this issue, the gene responsible for monodehydroascorbate reductase is knocked out which increases the plant tolerance against TNT. Fine-tuning enzymatic activity and knockout engineering together enhance the plant responses to toxic metals. Phytochelatin synthase, a heavy metal binding peptides synthesizing enzyme, revealed a way to enhance tolerance against heavy metals through enzymatic activity attenuation [106]. Recombinant DNA technology has proven to be effective in getting rid of arsenic particles that are considered as serious contaminants in soil. PvACR3, a key arsenite [As(III)] antiporter was expressed in *Arabidopsis* which showed enhanced tolerance to arsenic. Seeds of plants genetically engineered with *PvACR3* can germinate and grow in the presence of higher than normal quantity of arsenate [As(V)] which are generally lethal to wild-type seeds. Arsenic (As) is reduced by As reductase present in *A. thaliana*. Phytochelatins restrict the arsenic movement in root cells and phloem companion cells. OsNramp5 and OsHMA3 represent the transporters to uptake cadmium (Cd) and its retention [107]. In plants, brassino-steroid (BR) is involved in regulating physiological and developmental processes. Its activity is started with triggering phosphorylation or dephosphorylation cascade [108].

Recent biotechnological approaches for bioremediation include biosorption, phytostabilization, hyperaccumulation, dendroremediation, biostimulation, mycoremediation, cyanoremediation, and genoremediation, which majorly depend on enhancing or preventing specified genes activities. However, the challenges in adopting the successful technique cannot be ignored [109].

4.3.2. Energy Applications. Several microorganisms, specifically cyanobacteria, mediate hydrogen production, which is environmental friendly energy source. The specific production is maintained by utilizing the required enzymes properly as these enzymes play a key role in the product formation. But advanced approaches like genetic engineering, alteration in nutrient and growth conditions, combined culture, metabolic engineering, and cell-free technology [110–112] have shown positive results to increase the hydrogen production in cyanobacteria and other biofuels [3, 4]. The commercialization of this energy source will keep the environment clean which is not possible by using conventional energy sources releasing CO_2 and other hazardous chemicals [113]. Also cyanobacteria can be engineered to make them able to convert of CO_2 into reduced fuel compounds. This will make the carbon energy sources harmless to environment. This approach has been successful for vast range of commodity chemicals, mostly energy carriers, such as short chain and medium chain alcohols [114].

The conductive biofilms of *Geobacter sulfurreducens* are potential sources in the field in renewable energy, bioremediation, and bioelectronics. Deletion of PilZ genes encoding proteins in *G. sulfurreducens* genome made the biofilm more active as compared to wild-type. CL-1ln is specified for the strain in which the gene GSU1240 was deleted. Biofilm production was enhanced along with the production of pili and exopolysaccharide. The electron acceptor CL-1 produced biofilms that were 6-fold more conductive than wild-type biofilms when they were grown with electrode. This high fold conductivity lowered the potential losses in microbial fuel cells, decreasing the charge transfer resistance at the biofilm-anode surface and lowering the formal potential. Potential energy was increased by lower losses [115].

5. Current Challenges and Future Prospects

The fact that microbial cells are mostly used in the production of recombinant pharmaceutical indicates that several obstacles come into their way restricting them from producing functional proteins efficiently but these are handled with alterations in the cellular systems. Common obstacles which must be dealt with are posttranslational modifications, cell stress responses activation, and instability of proteolytic activities, low solubility, and resistance in expressing new genes. Mutations occurring in humans at genetic levels cause deficiencies in proteins production, which can be altered/treated by incorporation of external genes to fill the gaps and reach the normal levels. The use of *Escherichia coli* in recombinant DNA technology acts as a biological framework that allows the producers to work in controlled ways to technically produce the required molecules through affordable processes [41, 116].

Recombinant DNA research shows great promise in further understanding of yeast biology by making possible the analysis and manipulation of yeast genes, not only in the test tube but also in yeast cells. Most importantly, it is now possible to return to yeast by transformation with DNA and cloning the genes using a variety of selectable marker systems developed for this purpose. These technological advancements have combined to make feasible truly molecular as well as classical genetic manipulation and analysis in yeast. The biological problems that have been most effectively addressed by recombinant DNA technology are ones that have the structure and organization of individual genes as their central issue [117, 118]. Recombinant DNA technology is recently passing thorough development which has brought tremendous changes in the research lines and opened directions for advanced and interesting ways of research for biosynthetic pathways though genetic manipulation. *Actinomycetes* are being used for pharmaceutical productions, for example, some useful compounds in health sciences and the manipulation of biosynthetic pathways for a novel drugs generation. These contribute to the production of a major part of biosynthetic compounds and thus have received immense

considerations in recombinant drugs designing. Their compounds in clinical trials are more applicable as they have shown high level activity against various types of bacteria and other pathogenic microorganisms. These compounds have also shown antitumor activity and immunosuppressant activity [119].

Recombinant DNA tech as a tool of gene therapy is a source of prevention and cure against acquired genetic disorders collectively. DNA vaccines development is a new approach to provide immunity against several diseases. In this process, the DNA delivered contains genes that code for pathogenic proteins. Human gene therapy is mostly aimed to treat cancer in clinical trials. Research has focused mainly on high transfection efficacy related to gene delivery system designing. Transfection for cancer gene therapy with minimal toxicity, such as in case of brain cancer, breast cancer, lung cancer, and prostate cancer, is still under investigation. Also renal transplantation, Gaucher disease, hemophilia, Alport syndrome, renal fibrosis, and some other diseases are under consideration for gene therapy [120].

6. Conclusions

Recombinant DNA technology is an important development in science that has made the human life much easier. In recent years, it has advanced strategies for biomedical applications such as cancer treatment, genetic diseases, diabetes, and several plants disorders especially viral and fungal resistance. The role of recombinant DNA technology in making environment clean (phytoremediation and microbial remediation) and enhanced resistace of plants to different adverse acting factors (drought, pests, and salt) has been recognized widely. The improvements it brought not only in humans but also in plants and microorganisms are very significant. The challenges in improving the products at gene level sometimes face serious difficulties which are needed to be dealt for the betterment of the recombinant DNA technology future. In pharmaceuticals, especially, there are serious issues to produce good quality products as the change brought into a gene is not accepted by the body. Moreover, in case of increasing product it is not always positive because different factors may interfere to prevent it from being successful. Considering health issues, the recombinant technology is helping in treating several diseases which cannot be treated in normal conditions, although the immune responses hinder achieving good results.

Several difficulties are encountered by the genetic engineering strategies which needed to be overcome by more specific gene enhancement according to the organism's genome. The integration of incoming single-stranded DNA into the bacterial chromosome would be carried out by a RecA-dependent process. This requires sequence homology between both entities, the bacterial chromosome and incoming DNA. Stable maintenance and reconstitution of plasmid could be made easy. The introduction of genetic material from one source into the other is a disaster for safety and biodiversity. There are several concerns over development of genetically engineered plants and other products. For example, it is obvious that genetically engineered plants can cross-breed with wild plants, thus spreading their "engineered" genes into the environment, contaminating our biodiversity. Further, concerns exist that genetic engineering has dangerous health implications. Thus, further extensive research is required in this field to overcome such issues and resolve the concerns of common people.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contributions

Suliman Khan, Muhammad Wajid Ullah, and Ghulam Nabi contributed equally to this work.

Acknowledgments

The authors are thankful to Chinese Academy of Science and The World Academy of Science (CAS-TWAS) scholarship program. The corresponding author is thankful to Xuan H. Cao, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany, the guest editor for the special issue "The Promise of Agriculture Genomics" of "International Journal of Genomics," for his kind invitation.

References

[1] S. Kumar and A. Kumar, "Role of genetic engineering in agriculture," *Plant Archives*, vol. 15, pp. 1–6, 2015.

[2] T. Cardi and C. N. Stewart Jr., "Progress of targeted genome modification approaches in higher plants," *Plant Cell Reports*, vol. 35, no. 7, pp. 1401–1416, 2016.

[3] P. T. Lomedico, "Use of recombinant DNA technology to program eukaryotic cells to synthesize rat proinsulin: a rapid expression assay for cloned genes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 19, pp. 5798–5802, 1982.

[4] M. W. Ullah, W. A. Khattak, M. Ul-Islam, S. Khan, and J. K. Park, "Encapsulated yeast cell-free system: a strategy for cost-effective and sustainable production of bio-ethanol in consecutive batches," *Biotechnology and Bioprocess Engineering*, vol. 20, no. 3, pp. 561–575, 2015.

[5] M. W. Ullah, W. A. Khattak, M. Ul-Islam, S. Khan, and J. K. Park, "Bio-ethanol production through simultaneous saccharification and fermentation using an encapsulated reconstituted cell-free enzyme system," *Biochemical Engineering Journal*, vol. 91, pp. 110–119, 2014.

[6] W. A. Khattak, M. Ul-Islam, M. W. Ullah, B. Yu, S. Khan, and J. K. Park, "Yeast cell-free enzyme system for bio-ethanol production at elevated temperatures," *Process Biochemistry*, vol. 49, no. 3, pp. 357–364, 2014.

[7] W. A. Khattak, M. W. Ullah, M. Ul-Islam et al., "Developmental strategies and regulation of cell-free enzyme system for ethanol production: a molecular prospective," *Applied Microbiology and Biotechnology*, vol. 98, no. 23, pp. 9561–9578, 2014.

[8] L. Galambos and J. L. Sturchio, "Pharmaceutical firms and the transition to biotechnology: a study in strategic innovation," *Business History Review*, vol. 72, no. 2, pp. 250–278, 1998.

[9] F. M. Steinberg and J. Raso, "Biotech pharmaceuticals and biotherapy: an overview," *Journal of Pharmacy and Pharmaceutical Science*, vol. 1, no. 2, pp. 48–59, 1998.

[10] W. Liu, J. S. Yuan, and C. N. Stewart Jr., "Advanced genetic tools for plant biotechnology," *Nature Reviews Genetics*, vol. 14, no. 11, pp. 781–793, 2013.

[11] M. Venter, "Synthetic promoters: genetic control through cis engineering," *Trends in Plant Science*, vol. 12, no. 3, pp. 118–124, 2007.

[12] A. Berk and S. L. Zipursky, *Molecular Cell Biology*, vol. 4, WH Freeman, New York, NY, USA, 2000.

[13] M. Bazan-Peregrino, R. C. A. Sainson, R. C. Carlisle et al., "Combining virotherapy and angiotherapy for the treatment of breast cancer," *Cancer Gene Therapy*, vol. 20, no. 8, pp. 461–468, 2013.

[14] L.-X. Li, Y.-L. Zhang, L. Zhou et al., "Antitumor efficacy of a recombinant adenovirus encoding endostatin combined with an E1B55KD-deficient adenovirus in gastric cancer cells," *Journal of Translational Medicine*, vol. 11, no. 1, article 257, 2013.

[15] C. Méndez and J. A. Salas, "On the generation of novel anticancer drugs by recombinant DNA technology: the use of combinatorial biosynthesis to produce novel drugs," *Combinatorial Chemistry — High Throughput Screening*, vol. 6, no. 6, pp. 513–526, 2003.

[16] B. C. J. M. Fauser, B. M. J. L. Mannaerts, P. Devroey, A. Leader, I. Boime, and D. T. Baird, "Advances in recombinant DNA technology: corifollitropin alfa, a hybrid molecule with sustained follicle-stimulating activity and reduced injection frequency," *Human Reproduction Update*, vol. 15, no. 3, pp. 309–321, 2009.

[17] O. Merten and B. Gaillet, "Viral vectors for gene therapy and gene modification approaches," *Biochemical Engineering Journal*, vol. 108, pp. 98–115, 2016.

[18] O.-W. Merten, M. Schweizer, P. Chahal, and A. A. Kamen, "Manufacturing of viral vectors for gene therapy: part I. Upstream processing," *Pharmaceutical Bioprocessing*, vol. 2, no. 2, pp. 183–203, 2014.

[19] S. L. Ginn, I. E. Alexander, M. L. Edelstein, M. R. Abedi, and J. Wixon, "Gene therapy clinical trials worldwide to 2012—an update," *Journal of Gene Medicine*, vol. 15, no. 2, pp. 65–77, 2013.

[20] A. Rivero-Müller, S. Lajić, and I. Huhtaniemi, "Assisted large fragment insertion by Red/ET-recombination (ALFIRE)—an alternative and enhanced method for large fragment recombineering," *Nucleic Acids Research*, vol. 35, no. 10, article e78, 2007.

[21] L. E. Metzger IV and C. R. H. Raetz, "Purification and characterization of the lipid A disaccharide synthase (LpxB) from *Escherichia coli*, a peripheral membrane protein," *Biochemistry*, vol. 48, no. 48, pp. 11559–11571, 2009.

[22] E. A. Masson, J. E. Patmore, P. D. Brash et al., "Pregnancy outcome in Type 1 diabetes mellitus treated with insulin lispro (Humalog)," *Diabetic Medicine*, vol. 20, no. 1, pp. 46–50, 2003.

[23] A. K. Patra, R. Mukhopadhyay, R. Mukhija, A. Krishnan, L. C. Garg, and A. K. Panda, "Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from Escherichia coil," *Protein Expression and Purification*, vol. 18, no. 2, pp. 182–192, 2000.

[24] D. C. Macallan, C. Baldwin, S. Mandalia et al., "Treatment of altered body composition in HIV-associated lipodystrophy: comparison of rosiglitazone, pravastatin, and recombinant human growth hormone," *HIV Clinical Trials*, vol. 9, no. 4, pp. 254–268, 2008.

[25] E. Pennisi, "The CRISPR craze," *Science*, vol. 341, no. 6148, pp. 833–836, 2013.

[26] R. Wang, M. Li, L. Gong, S. Hu, and H. Xiang, "DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*," *Nucleic Acids Research*, vol. 44, no. 9, pp. 4266–4277, 2016.

[27] S. Shmakov, O. O. Abudayyeh, K. S. Makarova et al., "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems," *Molecular Cell*, vol. 60, no. 3, pp. 385–397, 2015.

[28] G. Gasiunas and V. Siksnys, "RNA-dependent DNA endonuclease Cas9 of the CRISPR system: holy grail of genome editing?" *Trends in Microbiology*, vol. 21, no. 11, pp. 562–567, 2013.

[29] P. Mohanraju, K. S. Makarova, B. Zetsche, F. Zhang, E. V. Koonin, and J. van der Oost, "Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems," *Science*, vol. 353, no. 6299, 2016.

[30] V. K. Vyas, M. I. Barrasa, and G. R. Fink, "A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families," *Science Advances*, vol. 1, no. 3, Article ID e1500248, 2015.

[31] A. P. Hynes, S. J. Labrie, and S. Moineau, "Programming native CRISPR arrays for the generation of targeted immunity," *mBio*, vol. 7, no. 3, p. e00202-16, 2016.

[32] F. Hille and E. Charpentier, "CRISPR-Cas: biology, mechanisms and relevance," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 371, no. 1707, Article ID 20150496, 2016.

[33] K. S. Makarova, Y. I. Wolf, O. S. Alkhnbashi et al., "An updated evolutionary classification of CRISPR-Cas systems," *Nature Reviews Microbiology*, vol. 13, no. 11, pp. 722–736, 2015.

[34] D. Rath, L. Amlinger, A. Rath, and M. Lundgren, "The CRISPR-Cas immune system: biology, mechanisms and applications," *Biochimie*, vol. 117, pp. 119–128, 2015.

[35] G. Liu, Q. She, and R. A. Garrett, "Diverse CRISPR-Cas responses and dramatic cellular DNA changes and cell death in pKEF9-conjugated *Sulfolobus* species," *Nucleic Acids Research*, vol. 44, no. 9, pp. 4233–4242, 2016.

[36] T. Gaj, C. A. Gersbach, and C. F. Barbas, "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering," *Trends in Biotechnology*, vol. 31, no. 7, pp. 397–405, 2013.

[37] P. R. Blackburn, J. M. Campbell, K. J. Clark, and S. C. Ekker, "The CRISPR system—keeping zebrafish gene targeting fresh," *Zebrafish*, vol. 10, no. 1, pp. 116–118, 2013.

[38] G. D. Yancopoulos, S. Davis, N. W. Gale, J. S. Rudge, S. J. Wiegand, and J. Holash, "Vascular-specific growth factors and blood vessel formation," *Nature*, vol. 407, no. 6801, pp. 242–248, 2000.

[39] R. K. Jain, P. Au, J. Tam, D. G. Duda, and D. Fukumura, "Engineering vascularized tissue," *Nature Biotechnology*, vol. 23, no. 7, pp. 821–823, 2005.

[40] K. Naoto, F. Dai, G. Oliver, A. Patrick, S. S. Jeffrey, and K. J. Rakesh, "Tissue engineering: creation of long-lasting blood vessels," *Nature*, vol. 428, pp. 138–139, 2004.

[41] N. Ferrer-Miralles, J. Domingo-Espín, J. Corchero, E. Vázquez, and A. Villaverde, "Microbial factories for recombinant pharmaceuticals," *Microbial Cell Factories*, vol. 8, article 17, 2009.

[42] M. Kamionka, "Engineering of therapeutic proteins production in *Escherichia coli*," *Current Pharmaceutical Biotechnology*, vol. 12, no. 2, pp. 268–274, 2011.

[43] S. Eriksson, "Enzymatic synthesis of nucleoside triphosphates," in *Nucleoside Triphosphates and their Analogs: Chemistry, Biotechnology, and Biological Applications*, vol. 23, 2016.

[44] D. J. Urban and B. L. Roth, "DREADDs (designer receptors exclusively activated by designer drugs): chemogenetic tools with therapeutic utility," *Annual Review of Pharmacology and Toxicology*, vol. 55, pp. 399–417, 2015.

[45] J. S. Tzartos, M. A. Friese, M. J. Craner et al., "Interleukin-17 production in central nervous system-infiltrating T cells and glial cells is associated with active disease in multiple sclerosis," *The American Journal of Pathology*, vol. 172, no. 1, pp. 146–155, 2008.

[46] K. Rabe, M. Lehrke, K. G. Parhofer, and U. C. Broedl, "Adipokines and insulin resistance," *Molecular Medicine*, vol. 14, no. 11-12, pp. 741–751, 2008.

[47] Z. S. Olempska-Beer, R. I. Merker, M. D. Ditto, and M. J. DiNovi, "Food-processing enzymes from recombinant microorganisms—a review," *Regulatory Toxicology and Pharmacology*, vol. 45, no. 2, pp. 144–158, 2006.

[48] Z.-X. Lian, Z.-S. Ma, J. Wei, and H. Liu, "Preparation and characterization of immobilized lysozyme and evaluation of its application in edible coatings," *Process Biochemistry*, vol. 47, no. 2, pp. 201–208, 2012.

[49] S. H. Bang, A. Jang, J. Yoon et al., "Evaluation of whole lysosomal enzymes directly immobilized on titanium (IV) oxide used in the development of antimicrobial agents," *Enzyme and Microbial Technology*, vol. 49, no. 3, pp. 260–265, 2011.

[50] B. Thallinger, E. N. Prasetyo, G. S. Nyanhongo, and G. M. Guebitz, "Antimicrobial enzymes: an emerging strategy to fight microbes and microbial biofilms," *Biotechnology Journal*, vol. 8, no. 1, pp. 97–109, 2013.

[51] C. E. Torres, G. Lenon, D. Craperi, R. Wilting, and Á. Blanco, "Enzymatic treatment for preventing biofilm formation in the paper industry," *Applied Microbiology and Biotechnology*, vol. 92, no. 1, pp. 95–103, 2011.

[52] J. K.-C. Ma, P. M. W. Drake, and P. Christou, "The production of recombinant pharmaceutical proteins in plants," *Nature Reviews Genetics*, vol. 4, no. 10, pp. 794–805, 2003.

[53] R. Gamuyao, J. H. Chin, J. Pariasca-Tanaka et al., "The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency," *Nature*, vol. 488, no. 7412, pp. 535–539, 2012.

[54] K. Hiruma, N. Gerlach, S. Sacristán et al., "Root endophyte *Colletotrichum tofieldiae* confers plant fitness benefits that are phosphate status dependent," *Cell*, vol. 165, no. 2, pp. 464–474, 2016.

[55] S. Jin and H. Daniell, "The engineered chloroplast genome just got smarter," *Trends in Plant Science*, vol. 20, no. 10, pp. 622–640, 2015.

[56] D. J. Oldenburg and A. J. Bendich, "DNA maintenance in plastids and mitochondria of plants," *Frontiers in Plant Science*, vol. 6, article 883, 2015.

[57] D. Henry, L. Choun-Sea, Y. Ming, and C. Wan-Jung, "Chloroplast genomes: diversity, evolution, and applications in genetic engineering," *Genome Biology*, vol. 17, article 134, 2016.

[58] W. Apel and R. Bock, "Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin A conversion," *Plant Physiology*, vol. 151, no. 1, pp. 59–66, 2009.

[59] T. Oikawa, H. Maeda, T. Oguchi et al., "The birth of a black rice gene and its local spread by introgression," *Plant Cell*, vol. 27, no. 9, pp. 2401–2414, 2015.

[60] Y. Oono, T. Yazawa, Y. Kawahara et al., "Genome-wide transcriptome analysis reveals that cadmium stress signaling controls the expression of genes in drought stress signal pathways in rice," *PLoS ONE*, vol. 9, no. 5, Article ID e96946, 2014.

[61] European Commission, Restrictions of geographical scope of GMO applications/authorisations: Member States demands and outcomes, 2015, http://ec.europa.eu/food/plant/gmo/authorisation.

[62] M. Cavazzana-Calvo, S. Hacein-Bey, G. De Saint Basile et al., "Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease," *Science*, vol. 288, no. 5466, pp. 669–672, 2000.

[63] S. Hacein-Bey-Abina, F. Le Deist, F. Carlier et al., "Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy," *The New England Journal of Medicine*, vol. 346, no. 16, pp. 1185–1193, 2002.

[64] S. J. Howe, M. R. Mansour, K. Schwarzwaelder et al., "Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients," *Journal of Clinical Investigation*, vol. 118, no. 9, pp. 3143–3150, 2008.

[65] R. M. Blaese, K. W. Culver, A. D. Miller et al., "T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years," *Science*, vol. 270, no. 5235, pp. 475–480, 1995.

[66] A. Aiuti, S. Vai, A. Mortellaro et al., "Immune reconstitution in ADA-SCID after PBL gene therapy and discontinuation of enzyme replacement," *Nature Medicine*, vol. 8, no. 5, pp. 423–425, 2002.

[67] N. Cartier, S. Hacein-Bey-Abina, C. C. Bartholomae et al., "Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy," *Science*, vol. 326, no. 5954, pp. 818–823, 2009.

[68] E. Montini, A. Biffi, A. Calabria et al., "Integration site analysis in a clinical trial of lentiviral vector based haematopoietic stem cell gene therapy for meatchromatic leukodystrophy," *Human Gene Therapy*, vol. 23, article A13, 2012.

[69] R. A. Morgan, M. E. Dudley, J. R. Wunderlich et al., "Cancer regression in patients after transfer of genetically engineered lymphocytes," *Science*, vol. 314, no. 5796, pp. 126–129, 2006.

[70] P. F. Robbins, R. A. Morgan, S. A. Feldman et al., "Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1," *Journal of Clinical Oncology*, vol. 29, no. 7, pp. 917–924, 2011.

[71] J. E. Adair, B. C. Beard, G. D. Trobridge et al., "Extended survival of glioblastoma patients after chemoprotective HSC gene therapy," *Science Translational Medicine*, vol. 4, no. 133, Article ID 133ra57, 2012.

[72] M. G. Ott, M. Schmidt, K. Schwarzwaelder et al., "Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1," *Nature Medicine*, vol. 12, no. 4, pp. 401–409, 2006.

[73] S. Stein, M. G. Ott, S. Schultze-Strasser et al., "Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease," *Nature Medicine*, vol. 16, no. 2, pp. 198–204, 2010.

[74] J. Zhang, E. B. Tarbet, H. Toro, and D.-C. C. Tang, "Adenovirus-vectored drug-vaccine duo as a potential driver for conferring mass protection against infectious diseases," *Expert Review of Vaccines*, vol. 10, no. 11, pp. 1539–1552, 2011.

[75] P. Lam, G. Khan, R. Stripecke et al., "The innovative evolution of cancer gene and cellular therapies," *Cancer Gene Therapy*, vol. 20, no. 3, pp. 141–149, 2013.

[76] A. Aiuti, L. Biasco, S. Scaramuzza et al., "Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome," *Science*, vol. 341, no. 6148, Article ID 1233151, 2013.

[77] N. P. Restifo, M. E. Dudley, and S. A. Rosenberg, "Adoptive immunotherapy for cancer: harnessing the T cell response," *Nature Reviews Immunology*, vol. 12, no. 4, pp. 269–281, 2012.

[78] M. H. Kershaw, J. A. Westwood, and P. K. Darcy, "Gene-engineered T cells for cancer therapy," *Nature Reviews Cancer*, vol. 13, no. 8, pp. 525–541, 2013.

[79] M. Danilo, M. Eusebio, P. C. Carla et al., "*In vivo* engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system," *Nature*, vol. 516, no. 7531, pp. 423–427, 2014.

[80] L. Jun, P. Shifeng, H. H. Mindy, and N. Nicholas, *Targeting Wnt-Driven Cancer through the Inhibition of Porcupine by LGK974*, MRC Laboratory of Molecular Biology, Cambridge, UK, 2013.

[81] J. H. C. M. Kreijtz, L. C. M. Wiersma, H. L. M. De Gruyter et al., "A single immunization with modified vaccinia virus Ankara-based influenza virus H7 vaccine affords protection in the influenza A(H7N9) pneumonia ferret model," *The Journal of Infectious Diseases*, vol. 211, no. 5, pp. 791–800, 2015.

[82] J. P. Hughes, G. Alusi, and Y. Wang, "Viral gene therapy for head and neck cancer," *The Journal of Laryngology & Otology*, vol. 129, no. 4, pp. 314–320, 2015.

[83] J. D. Smith, "Human Macrophage Genetic Engineering," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 36, no. 1, pp. 2–3, 2016.

[84] E. Stöger, C. Vaquero, E. Torres et al., "Cereal crops as viable production and storage systems for pharmaceutical scFv antibodies," *Plant Molecular Biology*, vol. 42, no. 4, pp. 583–590, 2000.

[85] C. Vaquero, M. Sack, F. Schuster et al., "A carcinoembryonic antigen-specific diabody produced in tobacco," *The FASEB journal*, vol. 16, no. 3, pp. 408–410, 2002.

[86] E. Karrer, S. H. Bass, R. Whalen, and P. A. Patten, U.S. Patent No. 8,252,727, 2012.

[87] R. M. Ionescu, J. Vlasak, C. Price, and M. Kirchmeier, "Contribution of variable domains to the stability of humanized IgG1 monoclonal antibodies," *Journal of Pharmaceutical Sciences*, vol. 97, no. 4, pp. 1414–1426, 2008.

[88] C. W. Adams, D. E. Allison, K. Flagella et al., "Humanization of a recombinant monoclonal antibody to produce a therapeutic HER dimerization inhibitor, pertuzumab," *Cancer Immunology, Immunotherapy*, vol. 55, no. 6, pp. 717–727, 2006.

[89] A. A. McCormick, S. Reddy, S. J. Reinl et al., "Plant-produced idiotype vaccines for the treatment of non-Hodgkin's lymphoma: safety and immunogenicity in a phase I clinical study," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 29, pp. 10131–10136, 2008.

[90] M. Bendandi, S. Marillonnet, R. Kandzia et al., "Rapid, high-yield production in plants of individualized idiotype vaccines for non-Hodgkin's lymphoma," *Annals of Oncology*, vol. 21, no. 12, pp. 2420–2427, 2010.

[91] S. Kathuria, R. Sriraman, R. Nath et al., "Efficacy of plant-produced recombinant antibodies against HCG," *Human Reproduction*, vol. 17, no. 8, pp. 2054–2061, 2002.

[92] A. Rostami-Hodjegan and G. T. Tucker, "Simulation and prediction of in vivo drug metabolism in human populations from in vitro data," *Nature Reviews Drug Discovery*, vol. 6, no. 2, pp. 140–148, 2007.

[93] J. K. Nicholson, E. Holmes, and I. D. Wilson, "Gut microorganisms, mammalian metabolism and personalized health care," *Nature Reviews Microbiology*, vol. 3, no. 5, pp. 431–438, 2005.

[94] Q. R. Fan and W. A. Hendrickson, "Structure of human follicle-stimulating hormone in complex with its receptor," *Nature*, vol. 433, no. 7023, pp. 269–277, 2005.

[95] M. Assidi, I. Dufort, A. Ali et al., "Identification of potential markers of oocyte competence expressed in bovine cumulus cells matured with follicle-stimulating hormone and/or phorbol myristate acetate in vitro," *Biology of Reproduction*, vol. 79, no. 2, pp. 209–222, 2008.

[96] Z.-B. Hu and M. Du, "Hairy root and its application in plant genetic engineering," *Journal of Integrative Plant Biology*, vol. 48, no. 2, pp. 121–127, 2006.

[97] C.-Q. Ling, L.-N. Wang, Y. Wang et al., "The roles of traditional Chinese medicine in gene therapy," *Journal of integrative medicine*, vol. 12, no. 2, pp. 67–75, 2014.

[98] L. Mazzoni, P. Perez-Lopez, F. Giampieri et al., "The genetic aspects of berries: from field to health," *Journal of the Science of Food and Agriculture*, vol. 96, no. 2, pp. 365–371, 2016.

[99] S. Ripp, D. E. Nivens, Y. Ahn et al., "Controlled field release of a bioluminescent genetically engineered microorganism for bioremediation process monitoring and control," *Environmental Science and Technology*, vol. 34, no. 5, pp. 846–853, 2000.

[100] G. S. Sayler, C. D. Cox, R. Burlage et al., "Field application of a genetically engineered microorganism for polycyclic aromatic hydrocarbon bioremediation process monitoring and control," in *Novel Approaches for Bioremediation of Organic Pollution*, pp. 241–254, Springer, New York, NY, USA, 1999.

[101] J. M. H. King, P. M. DiGrazia, B. Applegate et al., "Rapid, sensitive bioluminescent reporter technology for naphthalene exposure and biodegradation," *Science*, vol. 249, no. 4970, pp. 778–781, 1990.

[102] J. Chatterjee and E. A. Meighen, "Biotechnological applications of bacterial bioluminescence (*lux*) genes," *Photochemistry and Photobiology*, vol. 62, no. 4, pp. 641–650, 1995.

[103] K. Matsui, J. Togami, J. G. Mason, S. F. Chandler, and Y. Tanaka, "Enhancement of phosphate absorption by garden plants by genetic engineering: a new tool for phytoremediation," *BioMed Research International*, vol. 2013, Article ID 182032, 7 pages, 2013.

[104] R. B. Horsch, J. E. Fry, N. L. Hoffmann, D. Eichholtz, S. G. Rogers, and R. T. Fraley, "A simple and general method for transferring genes into plants," *Science*, vol. 227, no. 4691, pp. 1229–1230, 1985.

[105] M. Tamura, J. Togami, K. Ishiguro et al., "Regeneration of transformed verbena (*verbena* × *hybrida*) by *Agrobacterium tumefaciens*," *Plant Cell Reports*, vol. 21, no. 5, pp. 459–466, 2003.

[106] J. M. Jez, S. G. Lee, and A. M. Sherp, "The next green movement: plant biology for the environment and sustainability," *Science*, vol. 353, no. 6305, pp. 1241–1244, 2016.

[107] S. Clemens and J. F. Ma, "Toxic heavy metal and metalloid accumulation in crop plants and foods," *Annual Review of Plant Biology*, vol. 67, no. 1, pp. 489–512, 2016.

[108] E.-J. Kim, J.-H. Youn, C.-H. Park et al., "Oligomerization between BSU1 family members potentiates brassinosteroid signaling in *Arabidopsis*," *Molecular Plant*, vol. 9, no. 1, pp. 178–181, 2016.

[109] D. Mani and C. Kumar, "Biotechnological advances in bioremediation of heavy metals contaminated ecosystems: an overview with special reference to phytoremediation," *International Journal of Environmental Science and Technology*, vol. 11, no. 3, pp. 843–872, 2014.

[110] M. W. Ullah, M. Ul-Islam, S. Khan, Y. Kim, and J. K. Park, "Structural and physico-mechanical characterization of bio-cellulose produced by a cell-free system," *Carbohydrate Polymers*, vol. 136, pp. 908–916, 2016.

[111] M. W. Ullah, M. Ul-Islam, S. Khan, Y. Kim, and J. K. Park, "Innovative production of bio-cellulose using a cell-free system derived from a single cell line," *Carbohydrate Polymers*, vol. 132, pp. 286–294, 2015.

[112] M. W. Ullah, W. A. Khattak, M. Ul-Islam, S. Khan, and J. K. Park, "Metabolic engineering of synthetic cell-free systems: strategies and applications," *Biochemical Engineering Journal*, vol. 105, pp. 391–405, 2016.

[113] A. Tiwari and A. Pandey, "Cyanobacterial hydrogen production—a step towards clean environment," *International Journal of Hydrogen Energy*, vol. 37, no. 1, pp. 139–150, 2012.

[114] P. Savakis and K. J. Hellingwerf, "Engineering cyanobacteria for direct biofuel production from CO_2," *Current Opinion in Biotechnology*, vol. 33, pp. 8–14, 2015.

[115] C. Leang, N. S. Malvankar, A. E. Franks, K. P. Nevin, and D. R. Lovley, "Engineering *Geobacter sulfurreducens* to produce a highly cohesive conductive matrix with enhanced capacity for current production," *Energy and Environmental Science*, vol. 6, pp. 1901–1908, 2013.

[116] Z. Vajo, J. Fawcett, and W. C. Duckworth, "Recombinant DNA technology in the treatment of diabetes: insulin analogs," *Endocrine Reviews*, vol. 22, no. 5, pp. 706–717, 2001.

[117] J. M. DeJong, Y. Liu, A. P. Bollon et al., "Genetic engineering of taxol biosynthetic genes in *Saccharomyces cerevisiae*," *Biotechnology and Bioengineering*, vol. 93, no. 2, pp. 212–224, 2006.

[118] G. M. Walker, "Yeasts," in *Desk Encyclopedia of Microbiology*, Elsevier, 2nd edition, 2009.

[119] C. Méndez, G. Weitnauer, A. Bechthold, and J. A. Salas, "Structure alteration of polyketides by recombinant DNA technology in producer organisms prospects for the generation of novel pharmaceutical drugs," *Current Pharmaceutical Biotechnology*, vol. 1, no. 4, pp. 355–395, 2000.

[120] A. Misra, *Challenges in Delivery of Therapeutic Genomics and Proteomics*, Elsevier, Amsterdam, Netherlands, 2010.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 3460416, 8 pages
http://dx.doi.org/10.1155/2016/3460416

Review Article

Overview on the Role of Advance Genomics in Conservation Biology of Endangered Species

Suliman Khan,[1] Ghulam Nabi,[1] Muhammad Wajid Ullah,[2] Muhammad Yousaf,[3] Sehrish Manan,[4] Rabeea Siddique,[5] and Hongwei Hou[1]

[1]*The Key Laboratory of Aquatic Biodiversity and Conservation, Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China*
[2]*Department of Biomedical Engineering, Huazhong University of Science and Technology, Wuhan 430074, China*
[3]*Center for Human Genome Research, Cardio-X Institute, Huazhong University of Science and Technology, Wuhan 430074, China*
[4]*National Key Laboratory of Crop Genetic Improvement, College of Plant Sciences and Technology, Huazhong Agricultural University, Wuhan 430070, China*
[5]*Institute of Biotechnology and Genetic Engineering, The University of Agriculture, Peshawar 25000, Pakistan*

Correspondence should be addressed to Hongwei Hou; houhw@ihb.ac.cn

Received 21 July 2016; Revised 23 October 2016; Accepted 8 November 2016

Academic Editor: Hieu Xuan Cao

In the recent era, due to tremendous advancement in industrialization, pollution and other anthropogenic activities have created a serious scenario for biota survival. It has been reported that present biota is entering a "sixth" mass extinction, because of chronic exposure to anthropogenic activities. Various *ex situ* and *in situ* measures have been adopted for conservation of threatened and endangered plants and animal species; however, these have been limited due to various discrepancies associated with them. Current advancement in molecular technologies, especially, genomics, is playing a very crucial role in biodiversity conservation. Advance genomics helps in identifying the segments of genome responsible for adaptation. It can also improve our understanding about microevolution through a better understanding of selection, mutation, assertive matting, and recombination. Advance genomics helps in identifying genes that are essential for fitness and ultimately for developing modern and fast monitoring tools for endangered biodiversity. This review article focuses on the applications of advanced genomics mainly demographic, adaptive genetic variations, inbreeding, hybridization and introgression, and disease susceptibilities, in the conservation of threatened biota. In short, it provides the fundamentals for novice readers and advancement in genomics for the experts working for the conservation of endangered plant and animal species.

1. Introduction

Anthropogenic activities have changed the global environment, reducing the biodiversity through extinction and also reducing the population size of already surviving species. Due to man-made activities and interruptions, the current rate of species extinction is 1,000 times higher than natural background rates of extinction and future rates are likely to be 10,000 times higher [1]. *According to IUCN 2015 report, currently 79,837 species were assessed, of which* 23,250 are threatened with extinction. Only one-third of the world's freshwater fishes are at risk from hydropower dam expansion [2]. According to various estimates, each year few thousands to 100,000 species extinct, most without ever having been scientifically described [3]. Due to these tremendous anthropogenic activities, the notion has been emerged that earth biota is entering a "sixth" mass extinction [4] which is based on the facts that recent rates of species extinction are very high than prehuman background rates [5, 6]. *Only in the Island of Tropical Oceana, 1800 bird species were reported to extinct* in approximately 2000 years, since human colonization [7]. Even in the scientifically advanced *19th and 20th centuries, numerous species of birds, mammals, reptiles, fresh water fishes, amphibians, and other organisms extinction*

have been documented [5, 8, 9]. *If species extinction persists at such a tremendous speed*, future generation will occupy a planet with significantly reduce biodiversity, diminished ecosystem services, reduced evolutionary potential, and ultimately higher extinction rate and collapse ecosystem [3, 10].

It is a major challenge for biologists and ecologists to protect endangered species. Several measures have been taken and efforts done in this regard which is extensively described in literature such as population viability analysis, formulation of metapopulation theory, species conservation, contribution of molecular biology, development of global position system, geographical information system, and remote sensing [11]. *In the recent era*, *genomics is a key part of all the biological sciences and* a quickly changing approach to conservation biology. The genomes of many thousands of organisms including plants, vertebrates, and invertebrates have been sequenced and the results augmented, are annotated, and are refined through the use of new approaches in metabolomics, proteomics, and transcriptomics that enhance the characterization of metabolites, messenger RNA, and protein [12]. The genomic approaches *can provide detail information about the* present and past demographic parameters, phylogenetic issues, the molecular basis for inbreeding, understanding genetic diseases, and detecting hybridization/introgression in organisms [13]. It can also provide information to understand the mechanisms that relate low fitness to low genetic variation, for integrating genetic and environmental methodologies to conservation biology and for designing latest, fast monitoring tools. The rapid financial and technical progress in genomics currently makes conservation genomics feasible and will improve the feasibility in the very near future even [14]. The objective of this review is to describe recently advanced molecular technologies and their role in species conservation. We have described the effectiveness and possibility of conservation technology using the advance genomic approaches along with their limitations and future development. We hope that this review will provide fundamentals and new insights to both new readers and experienced biologists and ecologists in formulating new tools and establishing technologies to prevent endangered species.

2. Biodiversity and Conservation

Biodiversity refers to the variety of all forms of life on this planet, including various microorganisms, plants, animals, the ecosystem they form, and the genes they contain. Biodiversity within an area, biome, or planet is therefore considered at three levels including species diversity, genetic diversity, and ecosystem diversity [15]. As the names indicate, species diversity refers to the variety of species; genetic diversity is the variation of genes within species and populations and ecosystem diversity relates to the variety of habitats, ecological processes, and biotic communities in the biosphere [15]. Today's biodiversity about 9.0 to 52 million species is the result of billions of years of evolution, shaped by natural phenomena, and forms the web of life of which we are an integral part and upon which we are so fully [15, 16]. For species adaptation and survival, genetic diversity is the basic element and all the evolutionary achievement and to some degree survival depend on it. Though both adaptation and survival can be viewed in terms of space, time, and fitness but fitness further includes adaptation, genetic variability, and stability. The phenomenon of extinction can be the result of either abiotic or biotic stresses, caused by various factors such as disease, parasitism, predation, and competition or due to habitat alteration or isolation due to human activities, natural catastrophes, and slow climatic and geological changes. Considering these persistent threats, it is very crucial that genetic diversity in species should be appropriately understood and efficiently conserved and used [17].

At present, several species are in retreat, losing localities, and increasingly threatened with extinction by various factors mainly human intervention, and thus conservation biology has become a major file in recent times. A "threatened" designation generally recognizes a significant risk of becoming endangered throughout all or a portion of a species' range. Although extinction is a natural process, the human understanding of the value of the endangered species and its realization to intervene the stability of the environment is rapidly increasing. Human interferes in the natural environment of species in different ways, such as destruction of natural habitat, the introduction of nonnative organisms, and direct killing of natural components of a population [18]. Maintaining natural variation of species is beneficial from an economical, ecological, and social perspective. Several combinations of benefit occur for any particular species, and some species are obviously more valuable than the others.

Currently, the maintenance of rare and endangered species is a main focus of interest of biologists and geneticists. The impact of extinction is not always apparent and difficult to predict, and thus several parameters have been set and different technologies are being developed. For example, population viability analysis (PVA) quantitatively predicts the probability of extinction and prioritizes the conservation needs. It takes into account the combined impact of both stochastic (including the demography, environment, and genetics) and terministic (including habitat loss and overexploitation) factors [11]. Mandujano and Escobedo-Morales using PVA method for howler monkeys (*Alouatta palliata mexicana*) to simulate a group trend and local extinction and to investigate the role of demographic parameters to population growth under two landscape scenarios isolated populations and metapopulation [19]. They found that the rate of relative reproductive success and fecundity is directly linked with the number of adult females per fragment. As a result, the finite growth rate depended mainly on the survival of adult females while in both isolated populations and metapopulation the probability of extinction was exponentially dependent on fragment size. Further, it establishes a minimum viable population, predicts population dynamics, establishes conservation management programs, and evaluates its strategies. However, it is limited by several factors; for example, it is often very difficult to measure small-population parameters which need to be used in PVA models. This necessitates the development of more comprehensive and well-established approaches that can not only predict the extinction but also predict rather at a very early stage.

3. Role of Genomics Analysis Tools in Species Conservation

The term genome is about 75 years old and refers to the total set of genes on chromosomes or refers to the organism complete genetic material [20]. Together with the effect of an environment, it forms the phenotype of an individual. Thomas Roderick in 1986 coined the term genomics as a scientific discipline which refers to the mapping, sequencing, and analysis of the genome [21]. Now due to universal acceptance of genomics, it expands and is generally divided into functional and structural genomics. Structural genomics refers to the evolution, structure, and organization of the genome while functional genomics deals with the expression and function of the genome. Functional genomics needs assistance from structural genomics, mathematics, computer sciences, computational biology, and all areas of biology [22].

Genome analysis was once limited to model organisms [23] but now the genomes of thousands of organisms including plants, invertebrates, and vertebrates have been sequenced and the results annotated are further refined and augmented by using new approaches in metabolomics, proteomics, and transcriptomics [12]. Nowadays, it is quite easier to investigate the population structure, genetic variations, and recent demographic events in threatened species, using population genomic approaches. With recent developments, hints for becoming endangered species can be found in their genome sequences. For example, any deleterious mutations in the genes for brain function, metabolism, immunity, and so forth can be easily detected by advanced genomic approaches. Conversely, these can also detect any changes in their genome which may result in enhanced functions of some genes, for example, related to enhanced brain function and metabolism that may lead to the abnormal accumulation of toxins [24–26]. Specific genetic tools and analytical techniques are used to assess the genome of various species to detect genetic variations associated with specific conservation and population structure. Currently, most commonly used genetic tools for detection of genetic variations in both plant and animal species include random fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), random amplification of polymorphic DNA (RAPD), single strand conformation polymorphism (SSCP), minisatellites, microsatellites, single nucleotide polymorphisms (SNPs), DNA and RNA sequence analysis, and DNA finger printing. Analysis of genetic variation in species or population using these tools is carried out either using current DNA of individuals or historic DNA [27]. These tools target different variables within the genome of target species and selection of the specific tools and gnome part to be analyzed is carried out based on the available information. For example, mitochondrial DNA in animals possessing a high substitution rate is a useful marker for the determination of genetic variations in individuals of the same species. However, these techniques have several limitations associated with them. For instance, genetic high substitution rate in animal mitochondrial DNA is only inherited in female lines. Similarly, the mitochondrial DNA in plants has a very high rate of structural mutations and thus can rarely be used as genetic marker for detection of genetic variation. Various genomic tools used for the detection of genetic variations in species and limitations associated with them are summarized in Figure 1. Genome-wide association studies (GWAS), development of genome-wide genetic markers for DNA profiling and marker assisted breeding, and quantitative trait loci (QTL) analysis in endangered and threatened species can give us information about the role of natural selection at the genome level and identification of loci linked with the disease susceptibility, inbreeding depression, and local adaptations. For example, most of the QTLs have been detected using linkage mapping and cover large segments of the genome in different species. Currently, due to the availability of high-density SNP chips and genome-wide analysis techniques, GWAS has proven to be effective in identification of important genomic regions more precisely within the genome of species, for example, those associated with genetic variations and important qualitative and quantitative traits [28]. Further, use of population genetics and phylogenomics can help us in identifying conservation units for recovery, management, and protections [23]. As the genome of more species is sequenced, the rescue of more endangered species will become easier. The applications of advance genomics in the conservation of threatened biota are illustrated in Figure 2.

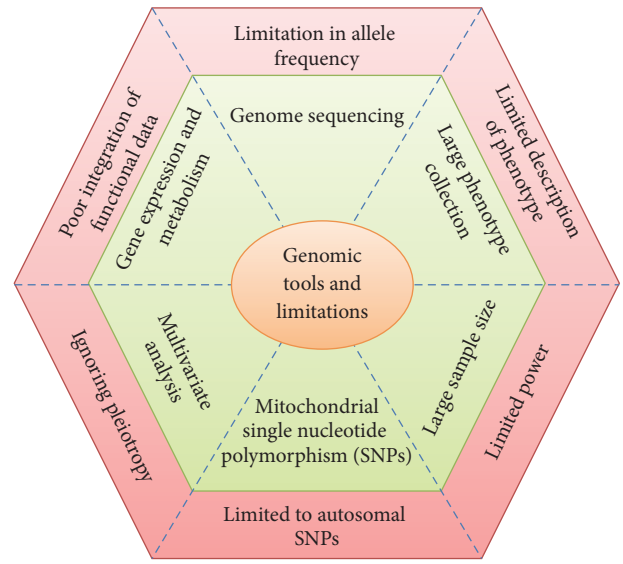


Figure 1: Illustration of various genetic tools for detection of genetic variations in species and their limitations in broad spectrum applications.

3.1. Demography. To identify recent and historic demographic events such as geographic population structure, gene flow, admixture, and population size fluctuations, specific genetic markers such as silent sites and microsatellites have been traditionally used. Although traditional molecular approaches have successfully analyzed and modeled complex demography histories, effective population size, nucleotide diversity, and recombination, genomics have provided a greater statistical and analytical power [29]. Genomics can

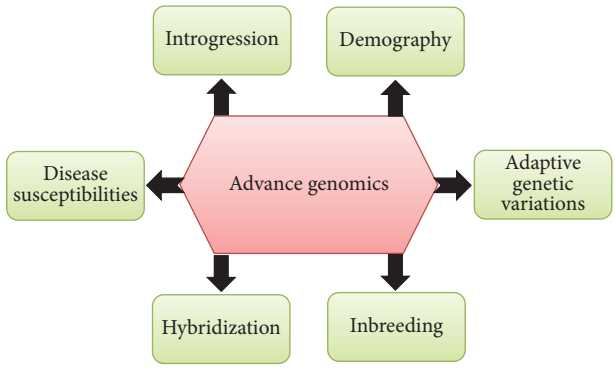


Figure 2: Illustration of advance genomic approaches for the conservation of species.

also provide information about speciation time, recombination rates, origin, relationship, and estimation of current and ancestral effective population size [30, 31]. Similarly, population genomics can improve our understanding about microevolution through a better understanding of recombination, assertive matting, mutation, and selection which helps us in identifying genes that are crucial for adaptation and fitness [32]. Future genetic analysis using SNPs can be of more advantage in determination of genome structure in regions with high linkage disequilibrium (LD) and low haplotype number in order to accelerate and optimize gene mapping based on genetic association, for example, finding relatively frequent variants associated with complex traits. However, this requires extensive knowledge of the LD patterns in the genome. It has been suggested that LD in genomes can be organized as a pattern of blocks of different length possessing limited diversity and separated by regions of low LD. Such structure can be the result of a number of possible mechanisms, one of which is recombination hotspots [33].

3.2. Adaptive Genetic Variations. Selective forces shape adaptive variations and identification of these adaptive loci is one of the most crucial focuses of genomics in conservation and evolution [34]. Genomics can help us to identify genetic changes resulting from local adaptation and the way these alterations influence fitness, through access to genome-wide data and annotated genomes in wild species. This information will not only help in defining conservation units [35] but also provide information about population potentials to respond to changing environmental condition [36]. Similarly, understanding the relationship between local adaptation and geographic distribution of loci will also benefit to evaluate habitat requirements for population persistence and the ecological exchangeability of divergent populations [37].

Various techniques are used to identify genetic regions associated with the adaptive traits. The most frequently used method is QTL [38], which has been used for many wild species such as cave tetra fish [39], deer mouse [40], and the zebra finch [41]. For example, the yield improvement of several crops such as wheat and maize has been made possible through the indirect manipulation of QTLs that control the heritable variability of the traits and physiological mechanisms [42, 43]. The conventional approaches of crop improvement such as breeding were based on little or no knowledge of the factors governing the genetic variability [44]. However, the conventional approaches for determining the genetic diversity are currently insufficient as the factors, for example, abiotic factors, including heat, stress, drought, water logging, and salinity, are becoming more prevalent in certain areas. Consequently, the genetic dissection of quantitative traits controlling the adaptive response in important crops to abiotic stresses is essential to allow cost-effective applications of genomic-based approaches to breeding programs aimed at improving the sustainability and stability of yield under adverse conditions [43]. Due to limited life history availability, nowadays, RAD-sequence [13, 45], GWAS [46, 47], and genome scan [48–50] are also used for identification of genetic regions associated with the adaptive traits. For example, GWAS was applied across the whole genome in several crops to detect the nonrandom association between the genomic markers scattered across the genome and the adaptive trait of interest [51]. Historical recombination increases the resolution in the detection of the locus controlling the adaptive traits [52], and thus GWAS identifies the nonrandom association of alleles among a locus with the adaptive traits (i.e., LD) as a result of action of natural selection [53]. The Major Histocompatibility Complex (MHC) has a role in kin recognition, intraspecific territoriality, and mate choice [54] and identification of polymorphism in MHC loci through genomics can give us information about the immunological fitness of the population [55] and further as advances are made can help us in conservation managements.

3.3. Inbreeding. Inbreeding of a species results in inbreeding depression, which can cause reduction of evolutionary adaptive potential, and ultimately increases the risk of extinction [56] but the exact mechanism of how this leads to the inbreeding depression is poorly understood. Only it is the genomics, which can shed light on the genetic architecture of inbreeding depression, a number of loci that contribute to inbreeding, and some underlying genetic mechanisms such as epistasis, overdominance, dominance, and/or gene-environment interactions [23]. For example, small scabious, a perennial plant in Netherlands, is an endangered species with highly fragmented and genetically eroded populations. Various transcriptomics and epigenetic analyses of inbreeding and inbreeding depression have been used to analyze this plant in the context of conservation. Various methods such as GWAS [57], gene expression profiles [58], and sequencing the whole genome of both parents and offspring [59] are used to identify loci related to inbreeding depression. The most immediate effect of inbreeding in a population is to reduce the frequency of heterozygotes, lowered fitness of individuals, inbreeding depression [60], and loss of diversity due to genetic drift by reducing effective population size [61]. For example, Mooney and McGraw studied *Panax quinquefolius* (American ginseng), a rare plant for outcrossing and inbreeding [62]. For inbreeding, the *Panax quinquefolius were either self-pollinated or* either cross-pollinated within the population. On the other hand for outcrossing, *Panax quinquefolius were* either cross-pollinated within the

population or with cultivated plants. Offspring resulted from all the crosses were followed for 4 years. Seedlings from self-pollinated plants showed 33% smaller heights and 45% smaller leaf areas relative to those from cross-pollination. On the other hand, Seedlings from crosses with cultivated plants showed 165% greater root biomass and 127% greater leaf area relative to outcrosses within the population. This example shows how inbreeding accelerates population extinction.

3.4. Hybridization and Introgression. Hybridization in some plant taxonomic group requires molecular markers at the genome level due to the peculiar characteristics of their genome architecture. For example, the wild form of sunflower, a noxious weed, can serve as a weed to the crop form. Hybridization can take place through pollinating insects which can move some of the crop's pollens into the weed populations. Ongoing hybridization between closely related species appears to be common in nature [63]. Genomics can provide better insight in the roles of hybridization and introgressive gene flow in natural populations and also can clear our concept of how species can maintain their genetic distinctiveness and reproductive isolation. Because introgressive gene flow may decrease or increase fitness, a better capability to identify the timing and occurrence of gene flow between species is relevant to population management and sustainability [23]. Translocated populations sometimes can hybridize with closely related or native populations of the same species, compromising the genetic purity of each species. For example, when *Cervus nippon* (Sika deer) were introduced to Western Europe, they readily inbreed with native *C. elaphus* (red deer) and as a result in Great Britain, there are no pure red deer [64]. Some extent of genetic flow is through a normal and evolutionarily constructive process, as the entire constellations of genotypes and genes cannot be preserved. However, hybridization with or without introgression in a rare or threatened species may compromise their existence. In this regard, only advance molecular technologies can play a significant role to understand the underappreciated problem that is not always evident from morphological observations alone [65].

3.5. Disease Susceptibilities. Infectious diseases, especially viral ones, are generally considered as a cause of decline in population [66] and are seldom considered a cause of extinction. In conservation biology, except in unusual circumstances, infectious diseases have a contributory or marginal influence on extinction [67, 68]. Recently it has been found that long term exposure to infectious diseases may alter the constitution of genome [69] which has a role in evolution and shaping of our biochemical individuality [70]. Advanced genomics can identify relevant susceptible genes and can provide better comprehensions into protective and pathogenic mechanisms and can pinpoint new molecular targets for therapeutic and prophylactic interventions [71]. Genome-wide SNP studies and whole genome sequencing can provide better understanding in wild life species managements and treatment of diseases [72] as the immediate goal for conservation management is to assess carrier status and to provide the basis for species recovery [73]. Currently, there are various examples under threat for various reasons being severely impacted by infectious diseases such as canine distemper in lions and black-footed ferrets [74, 75], Marburg and Ebola hemorrhagic diseases in anthropoids [76], transmissible facial tumor disease in Tasmanian devils [77], Kola retrovirus [78], and *Chlamydia pecorum* in Koalas [79]. In conservation biology, though host-pathogen interaction is a subject of particular interest, the possibility that pathogen causes extinction in certain context is rarely understood. However, increasing developments in the molecular technologies can provide substantial contribution to precisely understand the microbiological processes in wildlife [80].

4. Future Hope from Advance Genomics

Approximately, one-quarter of all avian species are either nearly threatened or threatened. Only 73 species of which are rescued from extinction. One of them was *Nipponia nippon* (crested ibis), which only from seven individuals was recovered, using high-quality genome sequences [25]. Even scientists for the first time created a viable clone of the world smallest endangered sheep, *European mouflon*, providing a hope to save them from extinction. Similarly, the original gene pool of any extinct population can be regenerated *via* cloning, by preserving their genetic diversity, through collection of cell samples. Even if cloning is managed properly, it may expand the genetic pool, can help us bring back genetic materials from dead animals, infertile animals, and even young animals that were too immature to breed [81]. For those extinct organisms for which no living cell exists, cloning is impossible; however, genome editing is the only means to bring extinct species or more accurately extinct traits back to life [82].

5. Limitations of Genomics

The most important impediment in conservation genomics is either lack of availability of samples or difficulty in sample collections of endangered species. Similarly, production of genomics data is easier and faster, but data analysis technique mostly lags. In addition, many statistical programs for population genetics need to be adapted to large data sets and require significant advances in bioinformatics and computational biology. Application of genetic data may result in defining units of conservation too narrowly, may impede conservation actions, and may stand in the way of endangered species management [83]. Further, genomes of some endangered species have not been sequenced yet and this requires not only heavy funds but considerable time. Some policies relating to sample exchange among countries also retard the speed of biological conservation.

6. Conclusions and Future Prospects

Conservation genetics is mainly focused on to determine the relationship between species or population, study the cross-species variation, and describe the interactions between

species and their threatening processes. In the current manuscript, we have overviewed the problems of conservation of endangered species and possible solutions and genetic and genomic approaches to apprehend them. Besides preventing the threatened species, diversity can benefit from looking beyond these and considering the genome of rare species and others that share a common environment. By identifying the factors or processes that influence the genomic composition of the threatened or extinction species, we can predict and identify the ecologically and genetically unique species. In the future, we hope that as advancement continues in genomics, we will be able to accurately predict the viability of local population and also to predict the ability of populations to adapt to climatic change and other anthropogenic challenges. Both climatic changes and anthropogenic activities due to a population explosion will be increasing day by day. Therefore, both these factors are a serious menace to biodiversity loss. The only hope to prevent their loss is expected from advance genomics. Further studies are also needed to appropriately understand and utilize environmental and genomic data and better ways to integrate them with multidisciplines, including policy analysis for effective conservation. Further, special policies should be established, to exchange the samples and genetic data of endangered species, in order to enhance species survival by the efforts of multinational groups.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contributions

Suliman Khan, Ghulam Nabi, and Muhammad Wajid Ullah contributed equally to this work.

Acknowledgments

The authors are thankful to the Academy of Science and The World Academy of Science (CAS-TWAS) scholarship program.

References

[1] J. M. De Vos, L. N. Joppa, J. L. Gittleman, P. R. Stephens, and S. L. Pimm, "Estimating the normal background rate of species extinction," *Conservation Biology*, vol. 29, no. 2, pp. 452–462, 2015.

[2] IUCN, The IUCN red list of threatened species, 2015, http://www.iucnredlist.org/current-news.

[3] N. S. Sodhi, B. W. Brook, and J. A. Corey Bradshaw, "Causes and consequences of species extinctions," in *The Princeton Guide to Ecology*, vol. 1, pp. 514–520, Princeton University Press, 2009.

[4] C. Gerardo, R. E. Paul, D. B. Anthony, G. Andrés, M. P. Robert, and M. P. Todd, "Accelerated modern human–induced species losses: entering the sixth mass extinction," *Science Advances*, vol. 1, no. 5, Article ID e1400253, 2015.

[5] R. M. May, "Ecological science and tomorrow's world," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, no. 1537, pp. 41–47, 2010.

[6] R. Dirzo, H. S. Young, M. Galetti, G. Ceballos, N. J. B. Isaac, and B. Collen, "Defaunation in the anthropocene," *Science*, vol. 345, no. 6195, pp. 401–406, 2014.

[7] D. W. Steadman, *Extinction and Biogeography of Tropical Pacific Birds*, Chicago University Press, Chicago, Ill, USA, 2006.

[8] H. M. Pereira, P. W. Leadley, V. Proença et al., "Scenarios for global biodiversity in the 21st century," *Science*, vol. 330, no. 6010, pp. 1496–1501, 2010.

[9] IUCN, "The IUCN Red List of Threatened Species, Version 2014.3," http://www.iucnredlist.org.

[10] J. Yule, R. Fournier, and P. Hindmarsh, "Biodiversity, extinction, and humanity's future: the ecological and evolutionary consequences of human population and resource use," *Humanities*, vol. 2, no. 2, pp. 147–159, 2013.

[11] H. Y. Jun, L. Zhong, C. G. Fa, W. W. Chao, and F. Z. Wei, "Advances in conservation methods of endangered species," *Acta Ecologica Sinica*, vol. 24, pp. 338–346, 2004.

[12] W. E. Johnson and K. Koepfli, "The role of genomics in conservation and reproductive sciences," *Advances in Experimental Medicine and Biology*, vol. 753, pp. 71–96, 2014.

[13] M. R. Miller, J. P. Brunelli, P. A. Wheeler et al., "A conserved haplotype controls parallel adaptation in geographically distant salmonid populations," *Molecular Ecology*, vol. 21, no. 2, pp. 237–249, 2012.

[14] N. J. Ouborg, F. Angeloni, and P. Vergeer, "An essay on the necessity and feasibility of conservation genomics," *Conservation Genetics*, vol. 11, no. 2, pp. 643–653, 2010.

[15] U. S. Rawat and N. K. Agarwal, "Biodiversity: concept, threats and conservation," *Environment Conservation Journal*, vol. 16, pp. 19–28, 2015.

[16] C. Mora, D. P. Tittensor, S. Adl, A. G. B. Simpson, and B. Worm, "How many species are there on earth and in the ocean?" *PLOS Biology*, vol. 9, no. 8, Article ID e1001127, 2011.

[17] V. R. Rao and T. Hodgkin, "Genetic diversity and conservation and utilization of plant genetic resources," *Plant Cell, Tissue and Organ Culture*, vol. 68, no. 1, pp. 1–19, 2002.

[18] F. J. Mazzotti, *The Value of Endangered Species: The Importance of Conserving Biological Diversity*, Wildlife Ecology and Conservation department, University of Florida, Cooperative Extension Service, Institute of Food and Agriculture Sciences, UF/IFAS Extension, Gainesville, Fla, USA, 2014.

[19] S. Mandujano and L. A. Escobedo-Morales, "Population viability analysis of howler monkeys (Alouatta palliata mexicana) in a highly fragmented landscape in Los Tuxtlas," *Tropical Conservation Science*, vol. 1, pp. 43–62, 2008.

[20] P. Hieter and M. Boguski, "Functional genomics: it's all how you read it," *Science*, vol. 278, no. 5338, pp. 601–602, 1997.

[21] Y. Xu, "Molecular plant breeding," in *Molecular Breeding Tool: Omics and Arrays*, p. 68, CABI, 2012.

[22] V. A. McKusick, "HUGO news. The human genome organisation: history, purposes, and membership," *Genomics*, vol. 5, no. 2, pp. 385–387, 1989.

[23] C. C. Steiner, A. S. Putnam, P. E. A. Hoeck, and O. A. Ryder, "Conservation genomics of threatened animal species," *Annual Review of Animal Biosciences*, vol. 1, pp. 261–281, 2013.

[24] C. E. Grueber, "Comparative genomics for biodiversity conservation," *Computational and Structural Biotechnology Journal*, vol. 13, pp. 370–375, 2015.

[25] S. Li, B. Li, C. Cheng et al., "Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species," *Genome Biology*, vol. 15, no. 12, pp. 557–572, 2014.

[26] B. J. Mcmahon, E. C. Teeling, and J. Höglund, "How and why should we implement genomics into conservation?" *Evolutionary Applications*, vol. 7, no. 9, pp. 999–1007, 2014.

[27] R. K. Wayne and P. A. Morin, "Conservation genetics in the new molecular age," *Frontiers in Ecology and the Environment*, vol. 2, no. 2, pp. 89–97, 2004.

[28] M. E. Goddard and B. J. Hayes, "Mapping genes for complex traits in domestic animals and their use in breeding programmes," *Nature Reviews Genetics*, vol. 10, no. 6, pp. 381–391, 2009.

[29] J. C. Avise, "Perspective: conservation genetics enters the genomics era," *Conservation Genetics*, vol. 11, no. 2, pp. 665–669, 2010.

[30] C. S. Wilding, R. K. Butlin, and J. Grahame, "Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers," *Journal of Evolutionary Biology*, vol. 14, no. 4, pp. 611–619, 2001.

[31] D. P. Locke, L. W. Hillier, W. C. Warren et al., "Comparative and demographic analysis of orang-utan genomes," *Nature*, vol. 469, no. 7331, pp. 529–533, 2011.

[32] W. C. Black IV, C. F. Baer, M. F. Antolin, and N. M. DuTeau, "Population genomics: genome-wide sampling of insect populations," *Annual Review of Entomology*, vol. 46, pp. 441–469, 2001.

[33] M. P. H. Stumpf, "Haplotype diversity and the block structure of linkage disequilibrium," *Trends in Genetics*, vol. 18, no. 5, pp. 226–228, 2002.

[34] F. W. Allendorf, P. A. Hohenlohe, and G. Luikart, "Genomics and the future of conservation genetics," *Nature Reviews Genetics*, vol. 11, no. 10, pp. 697–709, 2010.

[35] S. Manel, S. Joost, B. K. Epperson et al., "Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field," *Molecular Ecology*, vol. 19, no. 17, pp. 3760–3772, 2010.

[36] A. A. Hoffmann and C. M. Sgró, "Climate change and evolutionary adaptation," *Nature*, vol. 470, no. 7335, pp. 479–485, 2011.

[37] K. A. Crandall, O. R. R. Bininda-Emonds, G. M. Mace, and R. K. Wayne, "Considering evolutionary processes in conservation biology," *Trends in Ecology and Evolution*, vol. 15, no. 7, pp. 290–295, 2000.

[38] J. R. Stinchcombe and H. E. Hoekstra, "Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits," *Heredity*, vol. 100, no. 2, pp. 158–170, 2008.

[39] M. Protas, I. Tabansky, M. Conrad et al., "Multi-trait evolution in a cave fish, Astyanax mexicanus," *Evolution and Development*, vol. 10, no. 2, pp. 196–209, 2008.

[40] C. C. Steiner, J. N. Weber, and H. E. Hoekstra, "Adaptive variation in beach mice produced by two interacting pigmentation genes," *PLOS Biology*, vol. 5, no. 9, pp. 1880–1889, 2007.

[41] H. Schielzeth, W. Forstmeier, B. Kempenaers, and H. Ellegren, "QTL linkage mapping of wing length in zebra finch using genome-wide single nucleotide polymorphisms markers," *Molecular Ecology*, vol. 21, no. 2, pp. 329–339, 2012.

[42] M. W. Ullah, J. Hussain, R. Ahmad, A. Hassan, and M. M. Shah, "Molecular evaluation of wheat genotypes for vernalization response based on the Intron 1 deletion in VRN gene," *Minerva Biotecnologica*, vol. 27, no. 4, pp. 201–209, 2015.

[43] N. C. Collins, F. Tardieu, and R. Tuberosa, "Quantitative trait loci and crop performance under abiotic stress: where do we stand?" *Plant Physiology*, vol. 147, no. 2, pp. 469–486, 2008.

[44] N. E. Borlaug, "Sixty-two years of fighting hunger: personal recollections," *Euphytica*, vol. 157, no. 3, pp. 287–297, 2007.

[45] B. K. Peterson, J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra, "Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species," *PLOS ONE*, vol. 7, no. 5, Article ID e37135, 2012.

[46] T. M. Anderson, B. M. vonHoldt, S. I. Candille et al., "Molecular and evolutionary history of melanism in North American gray wolves," *Science*, vol. 323, no. 5919, pp. 1339–1343, 2009.

[47] H. E. Johnson, L. S. Mills, J. D. Wehausen, T. R. Stephenson, and G. Luikart, "Translating effects of inbreeding depression on component vital rates to overall population growth in endangered bighorn sheep," *Conservation Biology*, vol. 25, no. 6, pp. 1240–1249, 2011.

[48] C. R. Linnen, E. P. Kingsley, J. D. Jensen, and H. E. Hoekstra, "On the origin and spread of an adaptive allele in deer mice," *Science*, vol. 325, no. 5944, pp. 1095–1098, 2009.

[49] P. A. Hohenlohe, S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko, "Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags," *PLoS Genetics*, vol. 6, no. 2, Article ID e1000862, 2010.

[50] S. Renaut and L. Bernatchez, "Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (*Coregonus* spp. Salmonidae)," *Heredity*, vol. 106, no. 6, pp. 1003–1011, 2011.

[51] B. Hunter, K. M. Wright, and K. Bomblies, "Short read sequencing in studies of natural variation and adaptation," *Current Opinion in Plant Biology*, vol. 16, no. 1, pp. 85–91, 2013.

[52] E.-Y. Hwang, Q. Song, G. Jia et al., "A genome-wide association study of seed protein and oil content in soybean," *BMC Genomics*, vol. 15, article 1, 12 pages, 2014.

[53] B. E. Stranger, E. A. Stahl, and T. Raj, "Progress and promise of genome-wide association studies for human complex trait genetics," *Genetics*, vol. 187, no. 2, pp. 367–383, 2011.

[54] B. Ujvari and K. Belov, "Major histocompatibility complex (MHC) markers in conservation biology," *International Journal of Molecular Sciences*, vol. 12, no. 8, pp. 5168–5186, 2011.

[55] A. Aguilar, G. Roemer, S. Debenham, M. Binns, D. Garcelon, and R. K. Wayne, "High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 10, pp. 3490–3494, 2004.

[56] L. F. Keller and D. M. Waller, "Inbreeding effects in wild populations," *Trends in Ecology and Evolution*, vol. 17, no. 5, pp. 230–241, 2002.

[57] C. Charlier, W. Coppieters, F. Rollin et al., "Highly effective SNP-based association mapping and management of recessive defects in livestock," *Nature Genetics*, vol. 40, no. 4, pp. 449–454, 2008.

[58] K. N. Paige, "The functional genomics of inbreeding depression: a new approach to an old problem," *BioScience*, vol. 60, no. 4, pp. 267–277, 2010.

[59] D. Charlesworth and J. H. Willis, "The genetics of inbreeding depression," *Nature Reviews Genetics*, vol. 10, no. 11, pp. 783–796, 2009.

[60] D. Charlesworth and B. Charlesworth, "Inbreeding depression and its evolutionary consequences," *Annual Review of Ecology and Systematics*, vol. 18, pp. 237–268, 1987.

[61] S. C. H. Barrett and J. R. Kohn, "Genetic and evolutionary consequences of small population size in plants: implications for conservation," in *Genetics and Conservation of Rare Plants*, D. A. Falk and K. E. Holsinger, Eds., pp. 3–30, Oxford University Press, New York, NY, USA, 1991.

[62] E. H. Mooney and J. B. McGraw, "Effects of self-pollination and outcrossing with cultivated plants in small natural populations of american ginseng, *Panax quinquefolius* (Araliaceae)," *American Journal of Botany*, vol. 94, no. 10, pp. 1677–1687, 2007.

[63] L. H. Rieseberg, "Evolution: replacing genes and traits through hybridization," *Current Biology*, vol. 19, no. 3, pp. R119–R122, 2009.

[64] V. P. Lowe and A. S. Gardiner, "Hybridization between Red deer (*Cervus elaphus*) and Sika deer (*Cervus nippon*) with particular reference to stocks in N.W. England," *Journal of Zoology*, vol. 177, no. 4, pp. 553–566, 1975.

[65] J. M. Rhymer and D. Simberloff, "Extinction by hybridization and introgression," *Annual Review of Ecology and Systematics*, vol. 27, pp. 83–109, 1996.

[66] R. D. MacPhee and A. D. Greenwood, "Infectious Disease, Endangerment, and Extinction," *International Journal of Evolutionary Biology*, vol. 2013, Article ID 571939, 9 pages, 2013.

[67] K. F. Smith, D. F. Sax, and K. D. Lafferty, "Evidence for the role of infectious disease in species extinction and endangerment," *Conservation Biology*, vol. 20, no. 5, pp. 1349–1357, 2006.

[68] A. B. Pedersen, K. E. Jones, C. L. Nunn, and S. Altizer, "Infectious diseases and extinction risk in wild mammals," *Conservation Biology*, vol. 21, no. 5, pp. 1269–1279, 2007.

[69] D. Weatherall, J. Clegg, and D. Kwiatkowski, "The role of genomics in studying genetic susceptibility to infectious disease," *Genome Research*, vol. 7, no. 10, pp. 967–973, 1997.

[70] A. E. Garrod, *The Inborn Factors in Disease*, Oxford University Press, Oxford, UK, 1931.

[71] A. V. S. Hill, "Genetics and genomics of infectious disease susceptibility," *British Medical Bulletin*, vol. 55, no. 2, pp. 401–413, 1999.

[72] A. E. Savage and K. R. Zamudio, "MHC genotypes associate with resistance to a frog-killing fungus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 40, pp. 16705–16710, 2011.

[73] M. N. Romanov, M. Koriabine, M. Nefedov, P. J. de Jong, and O. A. Ryder, "Construction of a California condor BAC library and first-generation chicken-condor comparative physical map as an endangered species conservation genomics resource," *Genomics*, vol. 88, no. 6, pp. 711–718, 2006.

[74] M. E. Roelke-Parker, L. Munson, C. Packer et al., "A canine distemper virus epidemic in Serengeti lions (*Panthera leo*)," *Nature*, vol. 379, pp. 441–445, 1998.

[75] E. S. Williams, E. T. Thorne, M. J. Appel, and D. W. Belitsky, "Canine distemper in black-footed ferrets (Mustela nigripes) from Wyoming," *Journal of wildlife diseases*, vol. 24, no. 3, pp. 385–398, 1988.

[76] E. M. Leroy, P. Telfer, B. Kumulungui et al., "A serological survey of ebola virus infection in central African nonhuman primates," *Journal of Infectious Diseases*, vol. 190, no. 11, pp. 1895–1899, 2004.

[77] A.-M. Pearse and K. Swift, "Allograft theory: transmission of devil facial-tumour disease," *Nature*, vol. 439, article 549, 2006.

[78] J. J. Hanger, L. D. Bromham, J. J. Mckee, T. M. O'Brien, and W. F. Robinson, "The nucleotide sequence of koala (*Phascolarctos cinereus*) retrovirus: a novel type C endogenous virus related to gibbon ape leukemia virus," *Journal of Virology*, vol. 74, no. 9, pp. 4264–4272, 2000.

[79] M. Jackson, N. White, P. Giffard, and P. Timms, "Epizootiology of Chlamydia infections in two free-range koala populations," *Veterinary Microbiology*, vol. 65, no. 4, pp. 255–264, 1999.

[80] R. D. E. MacPhee and A. D. Greenwood, "Infectious disease, endangerment, and extinction," *International Journal of Evolutionary Biology*, vol. 2013, Article ID 571939, 9 pages, 2013.

[81] B. P. Trivedi, *Cientists Clone First Endangered Species: a Wild Sheep*, National Geographic New, 2001.

[82] B. Shapiro, "Mammoth 2.0: will genome engineering resurrect extinct species?" *Genome Biology*, vol. 16, no. 1, article 228, 2015.

[83] R. Frankham, J. D. Ballou, M. R. Dudash et al., "Implications of different species concepts for conserving biodiversity," *Biological Conservation*, vol. 153, pp. 25–31, 2012.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 8956412, 5 pages
http://dx.doi.org/10.1155/2016/8956412

Research Article

Genetic Diversity of Cowpea (*Vigna unguiculata* (L.) Walp.) Accession in Kenya Gene Bank Based on Simple Sequence Repeat Markers

Emily N. Wamalwa,[1] John Muoma,[1] and Clabe Wekesa[2]

[1]*Department of Biological Sciences, Masinde Muliro University of Science and Technology, P.O. Box 190-50100, Kakamega, Kenya*
[2]*Department of Biochemistry and Biotechnology, Kenyatta University, P.O. Box 43844-00100, Nairobi, Kenya*

Correspondence should be addressed to Emily N. Wamalwa; emndakwa@gmail.com

Received 4 August 2016; Revised 18 October 2016; Accepted 1 November 2016

Academic Editor: Wenqin Wang

Increased agricultural production is an urgent issue. Projected global population is 9 million people by mid of this century. Estimation projects death of 1 million people for lack of food quality (micronutrient deficit) and quantity (protein deficit). Majority of these people will be living in developing countries. Other global challenges include shrinking cultivable lands, salinity, and flooding due to climate changes, new emerging pathogens, and pests. These affect crop production. Furthermore, they are major threats to crop genetic resources and food security. Genetic diversity in cultivated crops indicates gene pool richness. It is the greatest resource for plant breeders to select lines that enhance food security. This study was conducted by Masinde Muliro University to evaluate genetic diversity in 19 cowpea accessions from Kenya national gene bank. Accessions clustered into two major groups. High divergence was observed between accessions from Ethiopia and Australia and those from Western Kenya. Upper Volta accessions were closely related to those from Western Kenya. Low variation was observed between accessions from Eastern and Rift Valley than those from Western and Coastal regions of Kenya. Diversity obtained in this study can further be exploited for the improvement of cowpea in Kenya as a measure of food security.

1. Introduction

Cowpea, *Vigna unguiculata* (L.) Walp. ($2n = 22$), is one of the most ancient human food sources [1]. It is one of the most important legume crops in the world and it is a major food crop in Africa, Latin America, and India because of its high protein content [2, 3]. As a result, cowpea is referred to as a poor man's meat [4]. Cowpea is primarily a self-pollinating crop and its genetic base is considered to be narrow [5–7].

The world's cowpea production as at 2013 was estimated at 5,718,144.66 tonnes of which 5,421,561 tonnes were from Africa, with East Africa contributing 532,901 tonnes [8]. In Kenya, yields remain extremely low, ranging from 150 to 500 kg/ha which is attributed to abiotic and biotic stresses, lack of high yielding cultivars, and poor crop management practices [9]. The area under cowpea in Kenya is estimated at 1800 ha excluding the area under the crop in home gardens [10].

Traditionally, diversity in cowpea is estimated by measuring variation in phenotypic or qualitative traits such as flower colour, growth habit, or quantitative agronomic traits such as yield potential and stress tolerance [11] that do not necessarily reflect real genetic relationships [12]. Furthermore, the expression of quantitative traits is subject to strong environmental influence and therefore limits knowledge of the germplasm structure for development of hybrids with specific ecological adaptations [11]. DNA markers are the most popular and widely used techniques to differentiate among genotypes at species and subspecies level [13]. Comparative studies in plants have shown that simple sequence repeat (SSR) markers, which are single locus markers with multiple alleles, are more valuable and provide an effective means for discriminating between genotypes [14, 15]. This study characterized 19 cowpea accessions from different sources that had been preserved at the national gene bank of Kenya using two SSR markers (Table 2) that had highest

Table 1: List of the names of all 19 cowpea accessions used in this study and their geographic distributions (locations).

Sample number	Accession name	Genebank accession number	District and locality of collection	Latitude	Longitude	Date of collection
1	Rift Valley 040539	GBK-040539	Turkana; Nadoto	2.7333°N	35.11667°E	18.9.1994
2	Australia 016157	GBK-016157	Uasin Gishu	—	—	5.1.1989
3	Coast 032338	GBK-032338	Kwale; Mwachanda	—	—	11.2.1992
4	Coast 032344	GBK-032344	Kwale; Marenje village	4.46167°S	39.12833°E	11.3.1992
5	Coast 032723	GBK-032723	—	—	—	1.1.1976
6	Rift Valley 040472	GBK-040472	Kabarnet	—	—	9.2.1994
7	Coast 031913	GBK-031913	Busia	0.45694°N	34.191389°E	7.10.1992
8	Western 047102	GBK-047102	Kakamega; Bunyala East	0.44172°N	034.68136°E	22.11.2004
9	Western 047111	GBK-047111	Vihiga; Mudete	0.11785°N	034.76527°E	24.11.2004
10	Eastern 046585	GBK-046585	Mwingi; Nzelune-Makilungi	1.284167°N	38.258611°E	29.8.2003
11	Western 044082	GBK-044082	Meru; Nkubu market	0.066667°S	37.666667°E	2.11.1997
12	Eastern 033061	GBK-033066	Embu; Embu research station	3.508889°S	37.454722°E	1.12.1992
13	Eastern 033066	GBK-033061	Embu; Embu research station	3.508889°S	37.454722°E	1.12.1992
14	Ethiopia 015141	GBK-015141	Siaya; Kigilo	—	—	5.1.1989
15	Eastern 033060	GBK-033060	Embu; Embu research station	3.508889°S	37.454722°E	1.12.1992
16	Upper Volta 022436	GBK-022436	Kilifi	—	—	10.1.1975
17	Western 047048	GBK-047048	Busia; Ageng'a	0.22152°N	034.08540°E	19.11.2004
18	Western 047119	GBK-047119	Vihiga; Serem Tiriki East	0.07745°N	034.8548°E	24.11.2004
19	Rift Valley 032108	GBK-032108	Nandi; Kaptumo location	0.067500°N	35.067500°E	27.8.1992
20	Western 047082	GBK-047082	Busia; Bumala	0.30394°N	034.20103°E	19.11.2004

Data obtained from the national gene bank of Kenya.

Table 2: Primer sequences used.

Primer code	Primer sequence 5′ to 3′
SSR-6540	5′-GGACATTTAGGATTGGGTGG-3′
	5′-CCATAGGTTAAACTTATTGTACTC-3′
SSR-6652	5′-CAAAATTCCACGGTCACC-3′
	5′-CGGGACTTGAGGTAGCGCG-3′

polymorphic amplification of both local and inbred lines of cowpea as reported by Badiane et al. [16].

2. Materials and Methods

2.1. DNA Extraction. Seedlings of each cowpea accession (Table 1) were grown in pots of sterile soil in a greenhouse with 3 plants per accession. Leaf samples were purposively sampled from three plants per accession from 15-day-old seedlings [16] and frozen in liquid nitrogen and genomic DNA extracted according to the prescribed protocol of the DNeasy Plant Mini Kit (Qiagen).

2.2. PCR Amplification of DNA and Electrophoresis. PCR was carried out in 0.2 mL tubes with a reaction volume of 25 μL containing 2.5 μL 10x PCR buffer, 1 μL of both primers, 1 mM of each dNTPs, 0.5 U Taq DNA polymerase, and 50 ng DNA. The tubes were placed in an Eppendorf Mastercycler Gradient thermocycler programmed for initial denaturation at 94°C for 1 minute, followed by 35 cycles for 30 seconds at 94°C, 30 seconds at 55°C, 1 minute at 75°C, and final extension for 10 minutes at 72°C.

The PCR products were resolved on a agarose gel (1%) using 0.5x TBE containing 1 mg/mL ethidium bromide with a vertical electrophoresis apparatus. The gel was photographed using Alphaimager 2200 under UV transilluminator (Figure 1). The resolved products were extracted from the gel and purified using the Qiagen DNA purification kit according to the prescribed protocol. DNA quantification was done by using a DNA NanoDrop 2000/2000c Spectrophotometer.

3. Phylogenetic Analysis

The sequences obtained were first edited by BioEdit version (version 7) and then nucleotide alignments were generated using ClustalW software. The evolutionary history was inferred using the Neighbor-Joining method. Analyses were conducted using the Jukes-Cantor model [17]. The analysis involved 19 nucleotide sequences. Codon positions included were the 1st, 2nd, 3rd, and noncoding positions. All ambiguous positions were removed for each sequence pair. Evolutionary analyses were conducted in MEGA6. The confidence of the branches was measured by bootstrapping with 1,000 replicates [18].

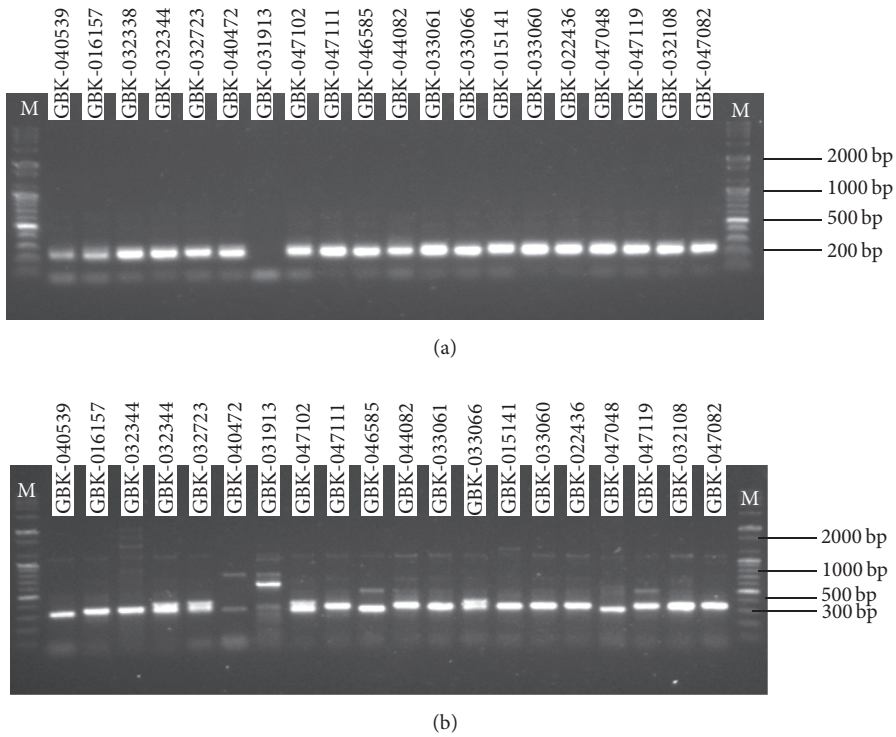


Figure 1: (a) PCR products of the 20 accessions of cowpea amplified using SSR primer 6652: M indicates the ladder. (b) PCR products of the 20 accessions of cowpea amplified using SSR primer 6540: M indicates the ladder.

4. Results

4.1. PCR Products. Nineteen cowpea accessions gave PCR products of ~300 bp (Figure 1(b)) for the SSR-6540 marker, while SSR-6652 gave PCR products of ~200 bp (Figure 1(a)) and these were not able to give any significant similarity from the NCBI and hence were not considered useful for this study.

4.2. Phylogenetic Analysis. These accessions were clustered into two main groups: A and B, as indicated in Figure 2. Cluster A is comprised of 16 accessions most diverged that form seven subclusters with bootstrap support of 79 (Eastern_033066, Rift_Valley_032108, Eastern_033060, and Western_047102), 59 (Rift_Valley_040472), 62 (Eastern_033061), 96 (Ethiopia_015141), 98 (Coast_032344 and Coast_032373), 63 (Western_047048), and 88 (Eastern_046585, Coast_032338). Three accessions in this cluster (Western_044082, Rift_Valley_040539, and Australia_016157) had their bootstrap less than 50; hence, their branches were not reliable.

Cluster B is comprised of four accessions that exhibit moderate level of divergence, forming four subclusters with bootstrap support of 91 (Western_047119), 88 (Western_047111), and 99 (Western_047082 and Upper_Volta_022436).

5. Discussion

Genetic diversity is the extent to which material differs within a group of plants [19]. The low genetic variability in the cowpea accessions used in this study is consistent with the findings of previous studies due to the fact that a single domestication event is involved in the origin of this crop [7, 9, 20–24].

The low genetic divergence observed in this study is in agreement with the findings of Padulosi and Ng [25], who attributed it to the self-pollination nature of this crop. Given that the accessions were from different regions, it could also indicate high-gene flow within regions and limited time for significant genetic differentiation along geographical lines as indicated by Karuma et al. 2008 [9]. Highest levels of divergence between the accession from Western Kenya, Australia, and Ethiopia could be attributed to the fact that the accessions have been popularly cultivated in the respective regions over time giving enough time for significant genetic differentiation along geographical lines [9]. At the same time, it could indicate genetic evidence of cowpea being a very diverse taxon as reported by Huynh et al. 2013 [26]. This therefore would mean that the studied germplasm from Australia and Ethiopia has some amount of diversity that can be used for cowpea improvement in Western Kenya. In the same manner, it can be argued that Upper_Volta_022436 has some amount of diversity that can be used for improvement of cowpea at the Rift Valley and Eastern and Coastal region of Kenya. Eastern Kenya constitutes 85% of the total production of cowpea in Kenya [27]. The comparison of the genetic distances between accessions from Western Kenya (Western_047111, Western_047102, Western_044082, Western_047119, Western_047048, and Western_047082) to those from Eastern Kenya revealed a closer genetic relationship with Eastern_046585 than all other accessions from the same region (Eastern_033060, Eastern_033061, and

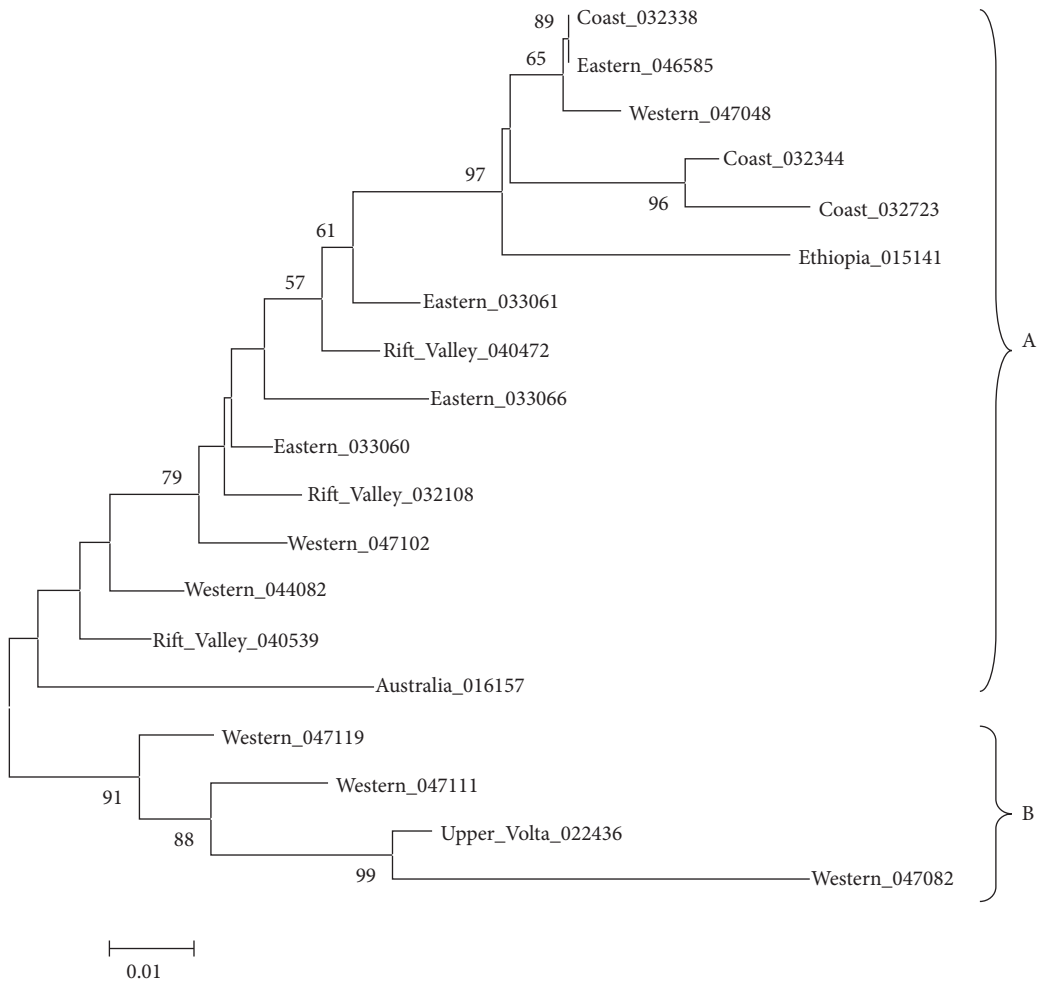


Figure 2: Phylogenetic tree of cowpea varieties: varieties having similar distances are genetically similar.

Eastern_033066). It could be argued that these accessions could have common origin with accession Eastern_046585 and the observed variability is attributed to hybridization as indicated by Adewale et al. [28].

Similarly, two Western Kenya accessions (Western_047102 and Western_044082) exhibited a closer relationship with Rift Valley_032108. The low genetic distances observed between these accessions could possibly reflect the initial bottleneck during domestication maintained by the inherent self-pollination mechanism of cowpea crop [7]. On the contrary higher genetic distance observed between Western_047082 and Rift Valley_040539 and between Western_047048 and both Rift_Valley_040539 and Rift Valley_040472 could possibly indicate that the diversity in these accessions could be used for the improvement of cowpea crop in these two regions [26].

On the other hand, the genetic distances of the accessions from Western Kenya to the accessions from Australia, Ethiopia, and Upper Volta revealed that there was greater divergence between Western_047082 and the accession from Australia and Ethiopia, but this accession was genetically closely related with the accession from Upper Volta which is in agreement with Ba et al. [23]. This could mean that these two accessions (Western_047082 and Upper_Volta_022436) are actually the same accession, despite being collected from different regions. This study has also observed that there is higher diversity between Australia_016157 and the accessions from the Coastal region of Kenya which could be exploited to improve the cowpea germplasm in the Coastal region of Kenya. This study has proved that although the genetic base of cowpea is narrow, there exists some level of diversity between cowpea accessions in the Kenya gene bank that can be exploited for the improvement of cowpea crop as a measure for food security in Kenya.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This study was made possible through the financial support of ASARECA. The cowpea accessions were provided by the gene bank in Kenya. The authors acknowledge Mary Silingi

for assisting them in accessing the gene bank data for the accessions used in this work.

References

[1] R. I. Summerfield, P. A. Huxley, and W. Steele, "Cowpea (*Vigna unguiculata* (L) Walp)," *Field Crop Abstracts*, vol. 27, pp. 301–312, 1974.

[2] D. Diouf and K. W. Hilu, "Microsatellites and RAPD markers to study genetic relationships among cowpea breeding lines and local varieties in Senegal," *Genetic Resources and Crop Evolution*, vol. 52, no. 8, pp. 1057–1067, 2005.

[3] K. T. Kareem and M. A. Taiwo, "Interactions of viruses in cowpea: effects on growth and yield parameters," *Virology Journal*, vol. 4, pp. 15–21, 2007.

[4] S. Sharmar, H. D. Upadhyaya, M. Rootkiwal, R. K. Varshney, and C. L. L. Gowda, "Chickpea," in *Genetic and Genomic Resources of Grain Legume Improvement*, M. Singh, H. D. Upadhyaya, and I. S. Bisht, Eds., pp. 81–104, Elsevier, London, UK, 2013.

[5] W. M. Sharawy and Z. A. El-Fiky, "Characterization of cowpea (*Vigna unguiculata* L.) accessions based on yield traits and RAPD-PCR analyses," *Arab Journal of Biotechnology*, vol. 6, pp. 67–78, 2002.

[6] J. Fang, P. S. Devanand, and C. C. T. Chao, *Genetic Diversity of Cowpea [*Vigna unguiculata *(L.) Walp] Breeding Lines from Different Countries*, AFLP Markers, Las Vegas, Nev, USA, 2005.

[7] A. T. Asare, B. S. Gowda, I. K. A. Galyuon, L. L. Aboagye, J. F. Takrama, and M. P. Timko, "Assessment of the genetic diversity in cowpea (*Vigna unguiculata* L. Walp.) germplasm from Ghana using simple sequence repeat markers," *Plant Genetic Resources*, vol. 8, no. 2, pp. 142–150, 2010.

[8] FAO 2014 FAOSTAT, http://faostat3.fao.org/download /Q/QC/E 1/04/2014.

[9] R. W. Karuma, O. Kiplagat, E. Ateka, and G. Owuoche, "Genetic diversity of Kenyan cowpea accessions based on morphological and microsatellite markers," *East African Agricultural and Forestry Journal*, vol. 76, pp. 3–4, 2008.

[10] J. G. N. Muthamia and F. K. Kanampiu, "On-farm cowpea evaluation in the marginal areas of eastern Kenya," in *Focus on Agricultural Research for Sustainable Development in a Changing Economic Environment: Proceedings of the 5th KARI Scientific Conference, 14th to 16th October, 1996, KARI Headquarters, Kaptagat Road, Loresho, Nairobi, Kenya*, Kenya Agricultural Research Institute, 1996.

[11] R. N. Kameswara, *Biotechnology for Plant Resources Conservation and Use*, Principles of Seed Handling in Genebanks Training Course, Kampla, Uganda, 2004.

[12] D. M. Patil, S. V. Sawardekar, N. B. Gokhale et al., "Genetic diversity analysis in cowpea [*VignaUnguiculata* (L.) Walp.] by using RAPD markers," *International Journal of Innovative Biotechnology and Biochemistry*, vol. 1, pp. 15–23, 2013.

[13] L. S. Kumar, "DNA markers in plant improvement: an overview," *Biotechnology Advances*, vol. 17, no. 2-3, pp. 143–182, 1999.

[14] W. Powell, M. Morgante, C. Andre et al., "The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis," *Molecular Breeding*, vol. 2, no. 3, pp. 225–235, 1996.

[15] C.-D. Li, C. A. Fatokun, B. Ubi, B. B. Singh, and G. J. Scoles, "Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers," *Crop Science*, vol. 41, no. 1, pp. 189–197, 2001.

[16] F. A. Badiane, B. S. Gowda, N. Cissé, D. Diouf, O. Sadio, and M. P. Timko, "Genetic relationship of cowpea (Vigna unguiculata) varieties from Senegal based on SSR markers.," *Genetics and molecular research : GMR*, vol. 11, no. 1, pp. 292–304, 2012.

[17] T. H. Jukes and C. R. Cantor, "Evolution of protein molecules," in *Mammalian Protein Metabolism*, H. N. Munro, Ed., pp. 21–132, Academic press, New York, NY, USA, 1969.

[18] J. Felsenstein, "Confidence limits on phylogenies: an approach using the bootstrap," *Evolution*, vol. 39, no. 4, pp. 783–791, 1985.

[19] Th. J. L. Van Hintum, "Hierarchical approaches to analysis of genetic diversity in crop plants," in *Core Collections of Plant Genetic Resources*, pp. 23–34, 1995.

[20] J. Doebley, "Isozymic evidence and the evolution of crop plants," in *Isozymes in Plant Biology*, D. E. Soltis and P. S. Soltis, Eds., pp. 165–191, Dioscorides Press, Portland, Ore, USA, 1989.

[21] R. S. Pasquet, "Allozyme diversity of cultivated cowpea *Vigna unguiculata* (L.) Walp," *Theoretical and Applied Genetics*, vol. 101, no. 1-2, pp. 211–219, 2000.

[22] S. Coulibaly, R. S. Pasquet, R. Papa, and P. Gepts, "AFLP analysis of the phenetic organization and genetic diversity of *Vigna unguiculata* L. Walp. reveals extensive gene flow between wild and domesticated types," *Theoretical and Applied Genetics*, vol. 104, no. 2-3, pp. 358–366, 2002.

[23] F. S. Ba, R. S. Pasquet, and P. Gepts, "Genetic diversity in cowpea [*Vigna unguiculata* (L.) Walp.] as revealed by RAPD markers," *Genetic Resources and Crop Evolution*, vol. 51, no. 5, pp. 539–550, 2004.

[24] E. M. Magembe, "Genetic diversity analysis among cowpea [*vignaunguiculata* (l.) walp] accessions from Sub-Saharan Africa using simple sequence repeats (ssr's)," University of Nairobi Publications, 2008, https://www.uonbi.ac.ke/.

[25] S. Padulosi and N. Q. Ng, "A useful and unexploited herb, *Vigna marina* (Leguminosae-Papilionoideae) and the taxonomic revision of its genetic diversity," *Bulletin du Jardin botanique National de Belgique*, vol. 62, no. 1–4, pp. 119–126, 1993.

[26] B.-L. Huynh, T. J. Close, P. A. Roberts et al., "Gene pools and the genetic architecture of domesticated cowpea," *Plant Genome*, vol. 6, article 3, 2013.

[27] A. M. Alghali, "On–farm evaluation of control strategies for insect pests in cowpea with emphasis on flower thrips, *Megalurothrips sjostedti* Trybom (Thysanoptera: Thripidae)," *Tropical Pest Management*, vol. 38, no. 4, pp. 420–424, 1992.

[28] B. D. Adewale, O. O. Adeigbe, and C. O. Aremu, "Genetic distance and diversity among some cowpea (*Vigna unguiculata* L. Walp)genotypes," *International Journal of Research in Plant Science*, vol. 1, pp. 9–14, 2011.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 6019046, 13 pages
http://dx.doi.org/10.1155/2016/6019046

Research Article

Effect on Soil Properties of *BcWRKY1* Transgenic Maize with Enhanced Salinity Tolerance

Xing Zeng, Yu Zhou, Zhongjia Zhu, Hongyue Zu, Shumin Wang, Hong Di, and Zhenhua Wang

College of Agronomy, Northeast Agricultural University, Mucai Street, Xiangfang District, Harbin, Heilongjiang 150030, China

Correspondence should be addressed to Hong Di; dihongdh@163.com and Zhenhua Wang; zhenhuawang_2006@163.com

Received 11 July 2016; Revised 28 August 2016; Accepted 15 September 2016

Academic Editor: Wenqin Wang

Maize (*Zea mays* L.) is the most important cereal crop in the world. However, soil salinity has become a major problem affecting plant productivity due to arable field degradation. Thus, transgenic maize transformed with a salinity tolerance gene has been developed to further evaluate its salt tolerance and effects on agronomic traits. It is necessary to analyze the potential environmental risk of transgenic maize before further commercialization. Enzyme activities, physicochemical properties, and microbial populations were evaluated in saline and nonsaline rhizosphere soils from a transgenic maize line (WL-73) overexpressing *BcWRKY1* and from wild-type (WT) maize LH1037. Measurements were taken at four growth stages (V3, V9, R1, and R6) and repeated in three consecutive years (2012–2014). There was no change in the rhizosphere soils of either WL-73 or WT plants in the four soil enzyme activities, seven soil physicochemical properties, and the populations of three soil organisms. The results of this study suggested that salinity tolerant transgenic maize had no adverse impact on soil properties in soil rhizosphere during three consecutive years at two different locations and provided a theoretical basis for environmental impact monitoring of salinity tolerant transgenic maize.

1. Introduction

In China, maize (*Zea mays* L.) is the most important cereal crop, and the production of this crop is affected by soil salinity. This problem has become ubiquitous in many countries. Thus, many salinity-tolerant crops, such as maize and rice, have been developed using transgenic technology [1]. Researchers have found that the microbial communities in the rhizosphere are influenced by the plant. Questions have been raised about whether antibiotic resistance genes, as selective markers, can transfer from genetically modified GM plants to indigenous microbes in the soil rhizospheres. Another question is whether certain GM plants differentially affect soil microbial communities compared to non-GM plants [2, 3].

Previous studies have shown that GM plants, including GM maize, potato, soybean, rice, and triticale, are equivalent to non-GM crops in terms of nutrition and are safe as food or feed [4]. The effects of *Bacillus thuringiensis* (*Bt*) transgenic cotton ("Mech 162") and non-*Bt* plants of the same cultivar on the ecology of many organisms in the soil were evaluated over three years in a subtropical environment. The authors concluded that the *Bt* cotton "Mech 162" did not have any negative effects on the organisms or biochemical characteristics of the soil [5]. The bacterial communities in the rhizosphere were studied using GM and non-GM maize in another study. Plant growth can promote rhizobacterial multiplication associated with both GM and non-GM plants, which indicates the mutually beneficial relationship between rhizobacteria and maize. No significant differences in the isolated rhizospheres were found during plant growth in GM or non-GM plants [6]. Using transgenic, salinity-tolerant *SUV3* and *PDH45* rice, the communication between rhizobacteria and rice was studied, and no significant effect was found [3, 7]. However, there have been few reports of the influence of GM and non-GM maize on rhizosphere soils [8]. The *BcWRKY1* gene was cloned from *Boea crassifolia* Hemsl, and it encodes a 444-amino acid WRKY-like protein

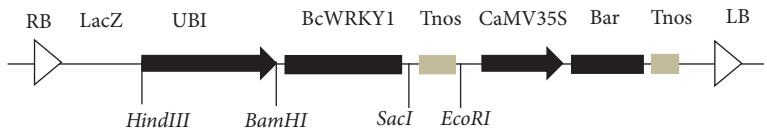


Figure 1: Schematic of the expression vector p3300-Ubi-*BcWRKY1*. RB, right border; LB, left border; UBI, ubiquitin promoter; Tnos, nopaline synthase terminator; *BcWRKY1*, *Boea crassifolia WRKY1* gene; Bar, Bialaphos resistance selectable marker gene; *CaMV35S*, cauliflower mosaic virus 35S promoter. *HindIII*, *BamHI*, *SacI*, and *EcoRI* are restriction endonuclease recognition sites.

containing two conserved domains: WRKYGQK and C2H2 motifs. The full-length *BcWRKY1* cDNA was 1,803 bp, and its expression could be induced by abiotic stresses, including soil salinity, low temperature, and drought. In addition, *BcWRKY1* transcription was accompanied by changes in plant hormones, including abscisic acid (ABA), salicylic acid (SA), and jasmonic acid (JA) [9]. Then, the *Boea crassifolia* DNA helicase of *BcWRKY1* was overexpressed under salt stress in maize. The NaCl stress-tolerant phenotype appeared even when plants were irrigated continuously with 150–200 mM NaCl, with no effect on their yield. Furthermore, other salt-induced genes, including *GmWRKY54*, *TaWRKY2*, and *HvWRKY38*, in *Arabidopsis thaliana* also promoted strong NaCl stress-tolerant phenotypes [10–12].

In this study, GM maize (WL-73) plants overexpressing the *BcWRKY1* gene and control non-GM maize LH1037 were used to evaluate the effects of WL-73 growth on the microbial populations in saline or nonsaline soils in Harbin, China. *BcWRKY1*-transgenic maize, which carries a kanamycin resistance gene, was compared to non-GM LH1037 maize to determine its effects on rhizosphere soil in terms of enzyme activities (including dehydrogenase, alkaline phosphatase, urease, and sucrase activities), physicochemical properties, and microbial populations.

2. Materials and Methods

2.1. Plant Sample Treatment. The line WL-73 was derived from the maize inbred line LH1037 transformed with the vector pCAMBIA 3300-*Ubi-WRKY1* (Figure 1). Seeds were obtained from T0 antisaline transgenic maize selected by saline stress and self-crossed to T6; WL-73 plants overexpressing *BcWRKY1* were able to survive under 300 mM NaCl stress [13].

The seeds of WL-73 and its receptor line LH1037 (WT) were grown in a three-row field with simulated saline-alkaline soil derived from a natural saline-alkaline field in Heilongjiang Province, China. The saline soil contained 29.42 g·kg^{-1} organic matter, 0.31 g·kg^{-1} total N, 127.27 mg·kg^{-1} available N (AN), 23.54 mg·kg^{-1} available P (AP), and 178.91 mg·kg^{-1} available K (K_2O) (AK) and had a pH of 8.65. The nonsaline control soil contained 58.02 g·kg^{-1} organic matter, 0.31 g·kg^{-1} total N, 119.26 mg·kg^{-1} AN, 26.01 mg·kg^{-1} AP, and 267.14 mg·kg^{-1} AK and had a pH of 7.67.

The seeds were salt and mock treated following Di's method with modifications [14]. WT and WL-73 seeds were germinated in sterilized vermiculite in a greenhouse with a humidity of 40–50% at 22°C and a light cycle of 16 h light/8 h darkness. The plants were well watered until the three-leaf stage. In addition, 0.5x Hoagland's nutrient solution with 300 mmol NaCl was applied to the salt treatment plants daily for 7 days, while the same solution without NaCl was applied to the control plants at the same frequency. Both the salt-treated and control plants in the experiment were then watered with 0.5x Hoagland's nutrient solution every 3 days to prevent excessive NaCl accumulation in the vermiculite.

2.2. Molecular Characterization and Salt Tolerance of Transgenic Maize. Leaves of the salt-treated seedlings were collected, and DNA and RNA were isolated. The CTAB method was used to isolate genomic DNA from the two youngest leaves of each plant [15]. Total RNA was isolated using TRIzol following the manufacturer's protocol (Tiangen Biotech, Beijing, China) under the requirement of 100 mg of young seeding leaves per mL of TRIzol. Exogenous *BcWRKY1* gene transcription was analyzed by RT-PCR.

The plant height and fresh weights were measured according to the methods of Di [14]. In addition, the membrane integrity parameters of the plants were determined by detecting superoxide dismutase (SOD) and peroxidase (POD) activity, proline (Pro) and malondialdehyde (MDA) content, relative electrical conductivity (REC), and chlorophyll content in leaves following the methods of Arnon [16] and Bates et al. [17].

2.3. Rhizosphere Soil Sampling. WL-73 and WT maize plants were grown in saline or conventional soil in triplicate from 2012 to 2014 at the Transgenic Experiment Station of Northeast Agricultural University, Harbin, Heilongjiang, China (longitude 126°73′, latitude 45°75′). Soil samples were isolated from the rhizosphere of WL-73 and WT at the V3 (the three lowest leaves have a visible collar), V9 (nine leaves have collars present), R1 (silking), and R6 (physiological maturity) stages. After removing the surface leaves, three soil samples from each plot were collected according to a checkerboard method. The soil volumes between 0 and 20 cm in depth were extracted using a soil auger with a 4 cm diameter, and the bulk soil on the root was shaken off. The soil from the root was stripped using a sterilizing brush and constituted the rhizosphere soil samples. We mixed three rhizosphere soil samples from each plot into one sample and then divided this sample into two. One of the samples was stored at 4°C until microbial analysis. Another sample was air-dried at room temperature, homogenized by sieving through a 2 mm mesh, and stored at 4°C until analysis.

2.4. Measurement of Soil Enzyme Activities. Dehydrogenase activity was analyzed as described by Min et al. [18]. Alkaline phosphatase activity was measured spectrophotometrically as described by Tabatabai and Bremner [19]. Urease enzyme activity was estimated as previously described [20]. Soil sucrase activity was measured using the 3,5-dinitrosalicylic acid method [21, 22].

2.5. Quantification of Physicochemical Properties. The physicochemical characteristics of soil and nutrient constituents, including soil type, pH, electrical conductivity (Ec) (mS $\cdot cm^{-1}$), organic carbon (OC) (%), AN (kg$\cdot ha^{-1}$), AP (kg $\cdot ha^{-1}$), and AK (kg$\cdot ha^{-1}$), were determined.

The soil pH and Ec were analyzed through the following steps. A 50 g soil sample was suspended in 100 mL of distilled, deionized water and stirred for 1 h at 100 rpm on a rotary shaker. The supernatant was collected by centrifugation at 10,000 ×g for 5 min. The Ec was recorded using a conductivity meter against 0.01 N KCl, and the pH was measured [23].

The available carbon, nitrogen, phosphorus, and potassium contents in the soils were determined following standard methods [24–27].

The calcium (Ca^{2+}), sodium (Na^{+}), and magnesium (Mg^{2+}) ion concentrations were determined using an atomic absorption method to determine the sodium adsorption ratio (SAR) [26]. The SAR was then calculated using the following formula:

$$\mathrm{SAR} = \frac{\mathrm{Na}^{+}}{\sqrt{(1/2)\left(\mathrm{Ca}^{2+} + \mathrm{Mg}^{2+}\right)}}. \quad (1)$$

2.6. Isolation of Rhizospheric Bacteria, Actinomycetes, and Fungi. To obtain isolated colonies, serial dilutions (10^{-4} dilution) prepared from 1 g soil samples were streaked onto nutrient agar medium in plates. Colonies were then selected, diluted, and spread onto plates containing beef extract peptone agar to detect bacteria, Gause's agar to detect actinomycetes, and Rose Bengal agar to detect fungi. Three replicates of the inoculated agar plates were incubated at 30°C, 28°C, or 28°C for 3 d for bacteria, 3 d for actinomycetes, and 5 d for fungi, after which the number of various types of colonies was recorded. The total populations of bacteria, actinomycetes, and fungi in each Petri dish were counted as colony forming units (cfu)$\cdot g^{-1}$ dry soil.

2.7. Data Processing Methods. This study was designed as a randomized complete block. The block treatments were the four growth stages (V3, V9, R1, and R6), the two maize materials (WL-73 and WT), and the two soil types (saline and nonsaline). All of the experiments were performed with three biological replicates over three years from 2012 to 2014. The data were analyzed statistically, and the standard error was calculated. An analysis of variance (ANOVA) was performed on treatment means using a generalized linear mixed model (GLMM), including treatment and sample time, in SAS 9.1 (Copyright 2008, SAS Institute, Cary, NC). Mean separations were performed using a least significant difference (LSD) test.

Table 1: PCR and RT-PCR primers for amplifying *BcWRKY1*.

Primer name	Sequence (5′-3′)
Primer I	ATGTCGTCTCTCGGCTCATC
Primer II	GAGCCCAACTGATTTTCTTG

3. Results and Discussion

3.1. Molecular Characterization of Transgenic Maize Plants (WL-73). The *BcWRKY1* gene was cloned from *Boea crassifolia*, which has the ability to tolerate salt stress [9]. WL-73 transgenic plants overexpressing the *BcWRKY1* gene were successfully generated by *Agrobacterium*-mediated transformation with the binary vector pCAMBIA3300-Ubi-*BcWRKY1* (Figure 1) introduced into the inbred line LH1037 [28].

The 1308-bp *BcWRKY1* PCR product was amplified from WL-73 transgenic plants with *BcWRKY1* gene-specific primers (Figure 2(a)). The transcription of the *BcWRKY1* gene in plant leaves was detected by RT-PCR (Table 1, Figure 2(b)). As expected, PCR and RT-PCR bands characteristic of *BcWRKY1* were detected in WL-73 but not in WT plants.

3.2. Salt Tolerance Evaluation of Transgenic Maize Plants (WL-73). When we treated maize plants with 300 mM NaCl solution for 7 days, the WL-73 plants were 5.3 cm taller and 60% heavier (fresh weight) than WT plants (Table 2). WT seedlings became almost entirely yellow on the 7th day after salt stress (Figure 3). The membrane integrity of the plants was measured in terms of parameters such as SOD, POD, Pro, MDA content, REC, and chlorophyll content following salt stress (Figure 4). The SOD, POD, Pro, MDA, and REC of WL-73 plants were significantly lower than those of WT plants ($P < 0.01$), while the chlorophyll content of WL-73 plants was higher than that of WT plants under the 300 mM NaCl treatment (Table 2). However, no significant difference was found in these values between WL-73 and WT plants under control conditions. These results suggest that the membranes of WL-73 plants were less damaged than those of the WT plant.

Our results suggest that *BcWRKY1* enhanced the tolerance of WL-73 plants to salinity stress via membrane stabilization and reduced REC and MDA contents compared with WT plants under salt stress (Figure 4, Table 2).

The *BcWRKY1* gene has been analyzed with other stress-related genes in transgenic plants, where it enhances the tolerance to salt and drought stress [10–12]. Therefore, increasing the expression level of *BcWRKY1* via transgenic technology should be critical for engineering crop plants with improved tolerance under multiple environmental stresses.

3.3. Activities of Four Enzymes in Rhizosphere Soil. The effects of WL-73 maize compared to control maize on rhizosphere soil enzyme activity, including the activities of alkaline phosphatase, urease, dehydrogenase, and sucrase, were studied in saline or control soil environments at four maize growth stages (V3, V9, R1, and R6) from 2012 to 2014. These four enzymes are the main enzymes in soil and significantly

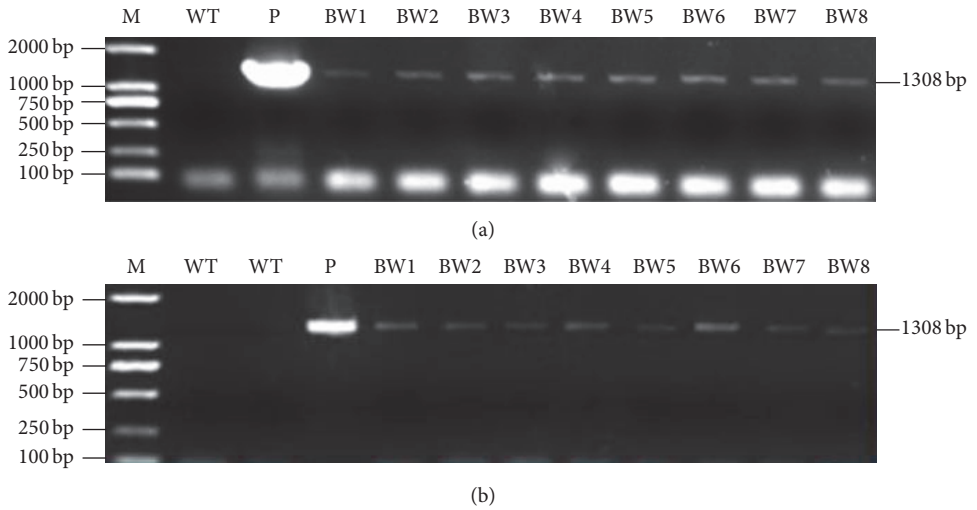


Figure 2: PCR and RT-PCR analyses of transformed and WT plants using *BcWRKY1* primers. (a) PCR analysis of WL-73 and WT; M, DNA marker DL2000 (TaKaRa); WT, wild-type plants of LH1037; P, pCAMBIA3300-*Ubi-WRKY1* plasmid; BW1-8 derived from plants of WL-73. (b) RT-PCR analysis of WL-73 and WT1; M, DNA marker DL2000 (TaKaRa); WT, wild-type plants of LH1037; P, pCAMBIA3300-*Ubi-WRKY1* plasmid; BW1-8 derived from WL-73 plants.

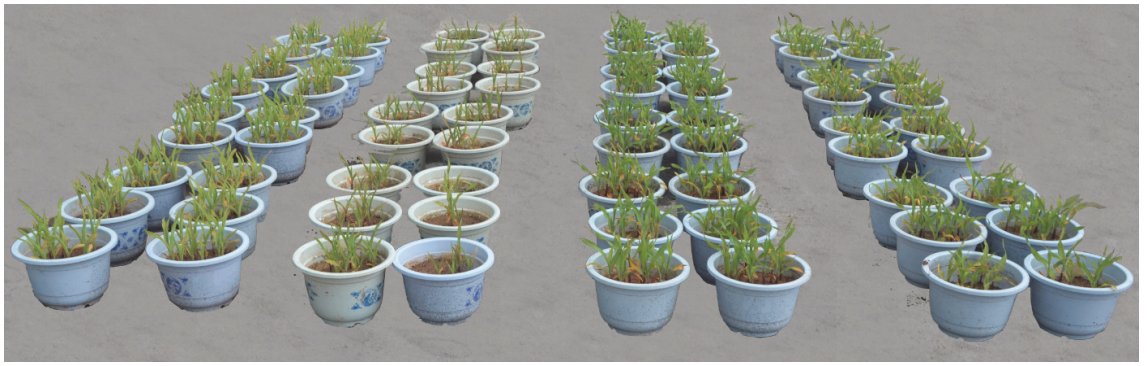


Figure 3: Enhanced salt tolerance of transgenic versus wild-type seedlings under 300 mM NaCl. Wild-type and transgenic maize plants were treated with 0.5x Hoagland's nutrient solution and either 0 mM NaCl (control) or 300 mM NaCl for 7 days. WL-73, LH1037 plants transformed with *BcWRKY1*; WT, wild-type LH1037 plants. Note: salt tolerance of transgenic maize compared to WT. Photographs were taken after salt treatment. *BcWRKY1*-overexpressing T6 transgenic and WT maize plants under salt-stressed (300 mM NaCl) and nonstressed conditions after 7 days.

Table 2: Height and fresh weight of maize WL-73 and WT under 300 mM NaCl stress.

Parameter	7 days of non-NaCl stress		7 days of NaCl stress	
	WL-73	WT	WL-73	WT
Height (cm)	25.7 ± 0.42^a	25.6 ± 0.34^a	24.4 ± 0.70^a	19.1 ± 0.81^b
Fresh weight per plant (g)	5.95 ± 0.21^a	5.89 ± 0.14^a	5.41 ± 0.63^a	3.38 ± 0.22^b
SOD ($U \cdot g^{-1}$ FW)	260.60 ± 1.35^a	259.80 ± 1.48^a	283.59 ± 3.38^a	332.02 ± 1.73^b
POD ($U \cdot g^{-1}$ FW)	813.03 ± 0.19^a	812.48 ± 0.47^a	927.62 ± 1.40^a	$1{,}316.92 \pm 1.83^b$
Pro ($mg \cdot g^{-1}$)	138.00 ± 0.48^a	137.58 ± 0.57^a	168.17 ± 2.64^a	225.80 ± 3.14^b
MDA ($nmol \cdot g^{-1}$)	78.41 ± 0.72^a	78.33 ± 0.75^a	102.44 ± 2.11^a	170.89 ± 4.06^b
REC (%)	0.17 ± 0.01^a	0.17 ± 0.01^a	0.23 ± 0.01^a	0.35 ± 0.02^b
Chlorophyll content ($mg \cdot g^{-1}$)	2.71 ± 0.03^a	2.72 ± 0.01^a	2.67 ± 0.05^a	2.05 ± 0.03^b

Wild-type or transgenic maize plants were treated with 0.5x Hoagland's nutrient solution and 0 mM NaCl (control) or 300 mM NaCl for 7 days. Then, the height, fresh weight, and SOD, POD, Pro, MDA, REC, and chlorophyll contents were measured. WT, wild-type maize LH1037; WL-73, LH1037 plant transformed with *BcWRKY1*.

Different letters following the numbers in the same column indicate a significant ($P \leq 0.05$) difference between treatments according to the LSD test.

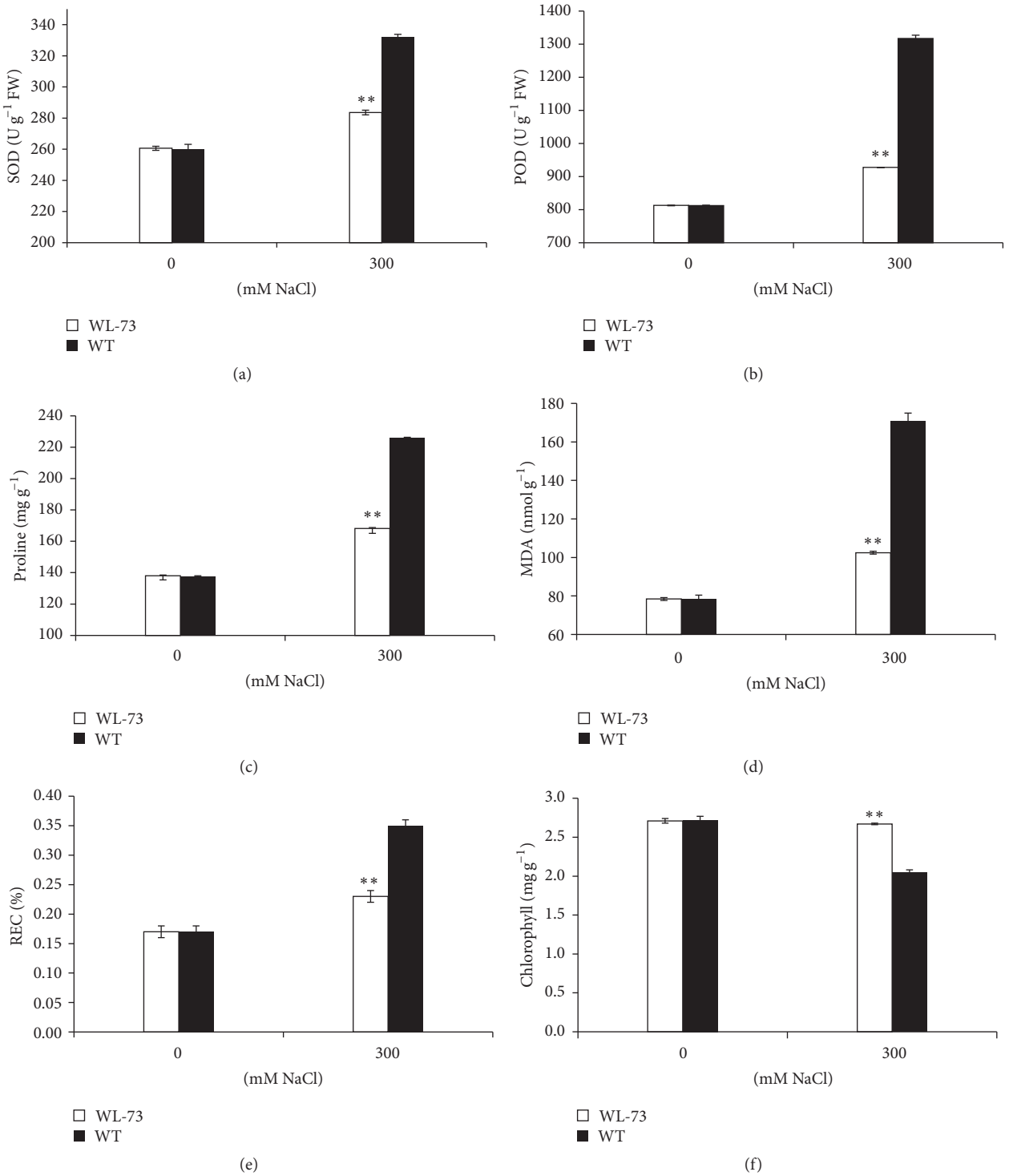


Figure 4: Effects of NaCl stress on the SOD, POD, Pro, MDA, REC, and chlorophyll contents of WT maize and WL-73 transgenic maize plants. WT and WL-73 were treated with 0.5x Hoagland's nutrient solution and either 0 mM NaCl (control) or 300 mM NaCl for 7 days; then, the SOD, POD, Pro, MDA, REC, and chlorophyll contents were measured. (a) SOD content; (b) POD content; (c) Pro content; (d) MDA content; (e) REC; (f) chlorophyll content. WT, wild-type maize LH1037; WL-73, LH1037 plant transformed with *BcWRKY1*. ** indicates a significant difference at 0.01 according to the LSD test ($n = 3$). The standard error is based on the average of three biological replicates. Note: the SOD, POD, Pro, MDA, and REC of WL-73 were significantly lower than those of WT ($P < 0.01$), while the chlorophyll content of WL-73 was higher than that of WT under 300 mM NaCl treatment for 7 days.

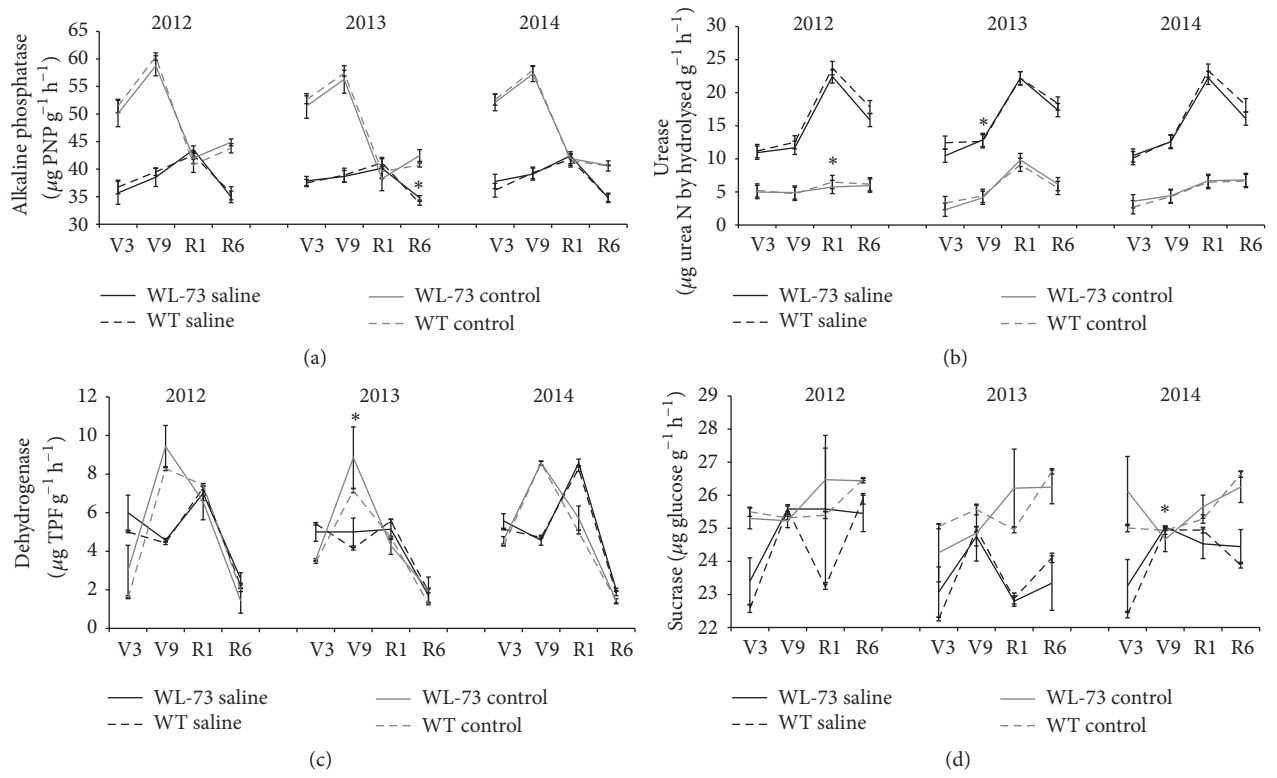
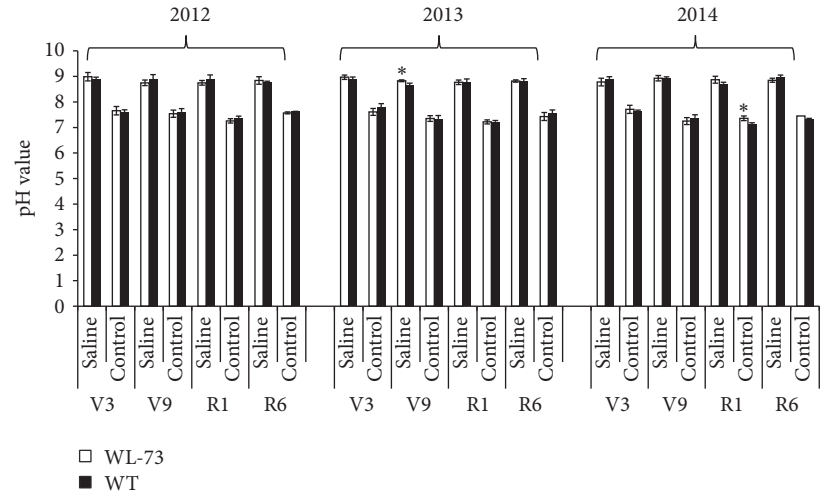


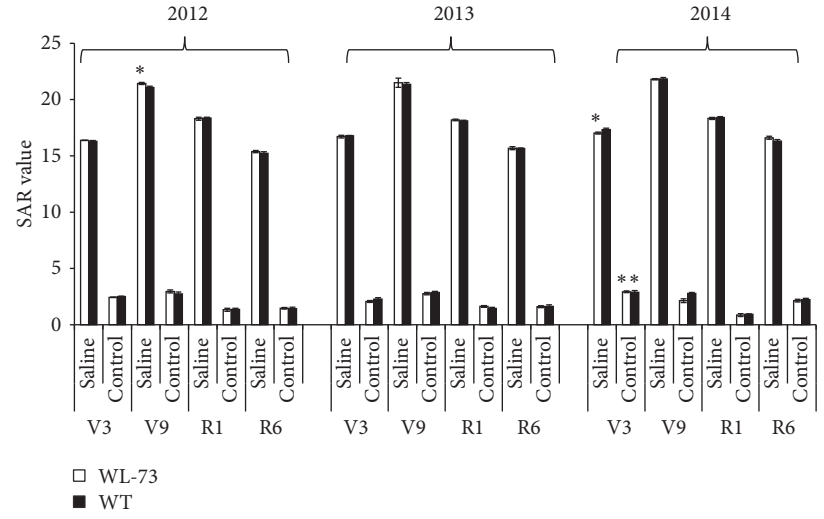
Figure 5: Activities of four enzymes in rhizosphere soil at different sampling times. (a) Alkaline phosphatase activity; (b) urease activity; (c) dehydrogenase activity; (d) sucrase activity. WT, wild-type maize LH1037; WL-73, LH1037 plant transformed with *BcWRKY1*; V3, the three lowest leaves have a visible collar; V9, nine leaves have collars present; R1, silking; R6, physiological maturity. $*$ indicates a significant difference at $P < 0.05$ according to the LSD test ($n = 3$). The standard error is based on the average of three biological replicates. Note: during the three years of this study, there were no overall significant differences ($P > 0.05$) in the alkaline phosphatase, urease, dehydrogenase, or sucrase activity in rhizosphere soil of WL-73 and WT plants in two soil environments (saline or nonsaline) or at four growth stages (V3, V9, R1, and R6). In addition, the activities of these four soil enzymes were not significantly different between WL-73 and WT plants in some soil environments, years, and growth stages.

affect the growth and yield of maize. Alkaline phosphatase is mainly involved in the soil phosphorus cycle. Urease is associated with the nitrogen cycle. Dehydrogenase is the main oxidoreductase, and sucrose is the major hydrolase enzyme. During the three years of this study, the alkaline phosphatase, urease, dehydrogenase, and sucrase activities in the rhizosphere soil of WL-73 and WT plants were not different ($P > 0.05$) in the two soil environments (saline or nonsaline) or at any of the four growth stages (V3, V9, R1, and R6). The ANOVA results showed that the dehydrogenase and alkaline phosphatase activities in the rhizosphere soil of WL-73 and WT maize were similar, with values ranging from 34.84 to 39.04 μg PNP g^{-1}·h^{-1} in saline soil and from 40.74 to 57.44 μg PNP g^{-1}·h^{-1} in control soil for WL-73 and from 34.74 to 41.85 μg PNP g^{-1}·h^{-1} in saline soil and from 40.62 to 59.26 μg PNP g^{-1}·h^{-1} in control soil for WT maize (Tables 3 and 4; Figures 5(a)–5(d)). However, the activities of these four soil enzymes were significantly different between WL-73 and WT plants in some soil environments, years, and growth stages. For example, there were significant differences in alkaline phosphatase activity in saline soil at R6 in 2013 ($P = 0.03$); in urease activity in control soil at R1 in 2012 and in saline soil at V9 in 2013 ($P < 0.05$); in dehydrogenase activity in control soil at V9 in 2013; and in sucrase activity in saline soil at V9 in 2014 (Table 3, Figures 5(a)–5(d)). However, no consistent trends in enzyme activity were detected in the two soil environments, over the three-year study, or in the four growth stages analyzed here. These results agree with those of previous studies in other regions and for a variety of crops [29–32].

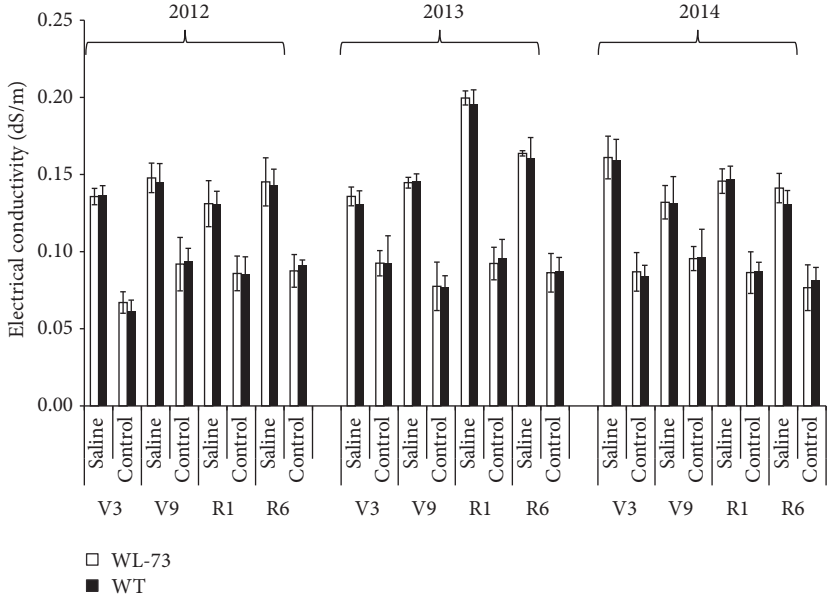
Soil enzymes play an important role in maintaining soil ecology, physicochemical properties, fertility, and health [33, 34]. The overexpression of *PDH45* and *SUV3* in transgenic rice has no adverse effect on rhizosphere soil or its microflora [3, 7]. In other studies of transgenic crops, the only consistent significant differences in soil enzymes and physicochemical properties between transgenic and nontransgenic plants were due to seasons and crop varieties. There were no significant differences in the enzyme activities of rhizospheric microbes from soils in which *Bt* or non-*Bt* cotton was grown [5]. The results of the present study indicated few significant differences in the alkaline phosphatase, urease,



(a)

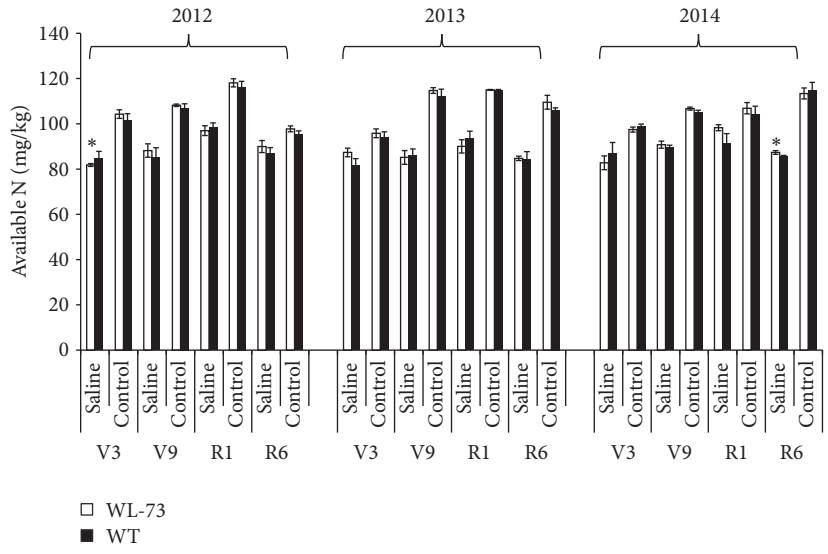


(b)

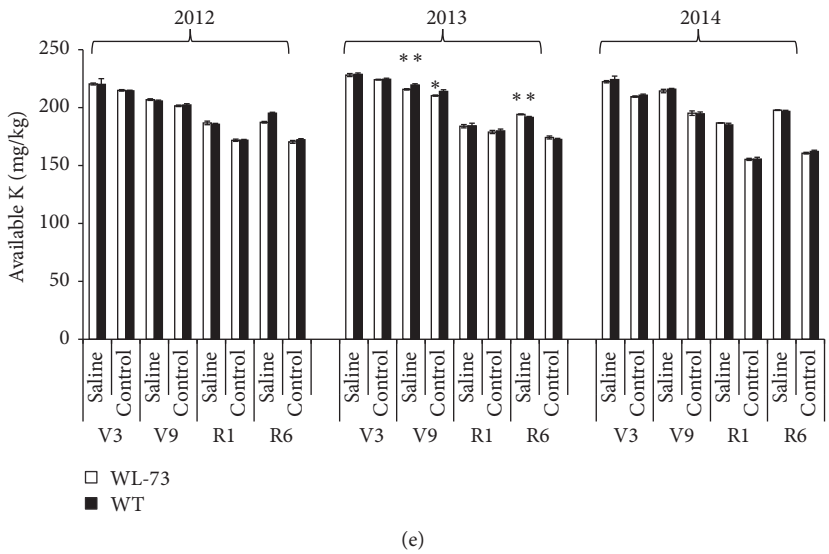


(c)

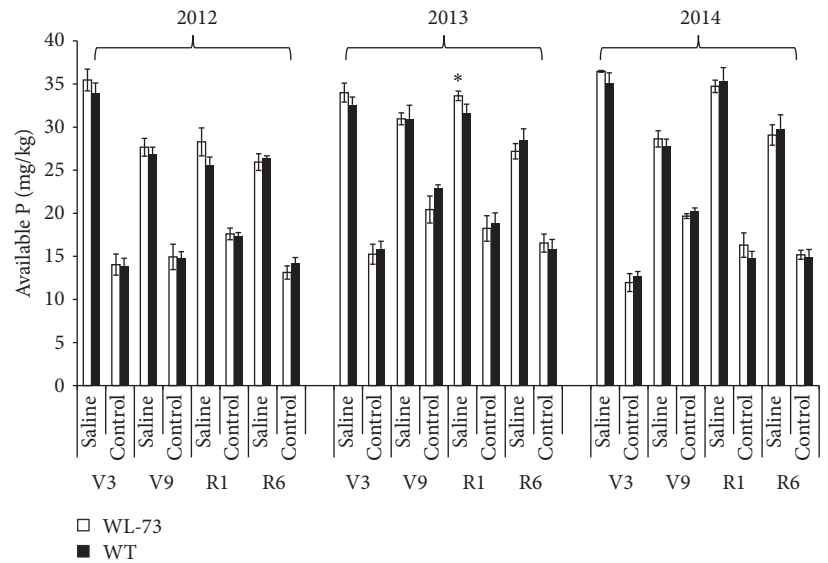
Figure 6: Continued.



(d)



(e)



(f)

Figure 6: Continued.

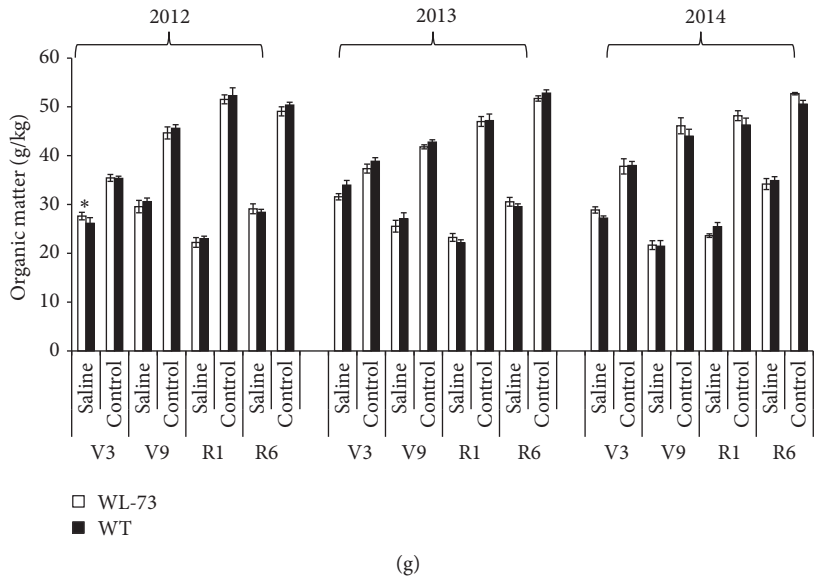


Figure 6: Assays of seven physicochemical properties of rhizosphere soil. *BcWRKY1* transgenic maize (WL-73) and nontransgenic maize (WT; LH1037) plants were grown in saline or control nonsaline soil from 2012 to 2014. Seven physicochemical parameters were measured, and a statistical analysis was performed using the LSD test with $P < 0.05$ and $n = 3$. (a) pH; (b) SAR value; (c) electrical conductivity; (d) available N; (e) available K; (f) available P; (g) organic matter. WT, wild-type maize LH1037; WL-73, LH1037 plant transformed with *BcWRKY1*; V3, the three lowest leaves have a visible collar; V9, nine leaves have collars present; R1, silking; R6, physiological maturity. $*$ indicates a significant difference at $P < 0.05$ according to the LSD test ($n = 3$). $**$ indicates a significant difference at $P < 0.01$ according to the LSD test ($n = 3$). The standard error is based on the average of three biological replicates. Note: no overall significant differences were observed between WL-73 and WT plants for most of the physicochemical properties ($P > 0.05$) at four maize growth periods (R1, R9, V1, and V6) in saline and control soil environments from 2012 to 2014.

dehydrogenase, and sucrase activities in the rhizosphere soil between WL-73 and WT plants, as in the five studies cited above.

3.4. Physicochemical Properties of Rhizosphere Soil. The physicochemical properties, including the pH, SAR, Ec, AN, AK, AP, and OC, of rhizosphere soil from WL-73 and WT maize plants are shown in Figures 6(a)–6(g). There were no overall significant differences between WL-73 and WT plants for most of the physicochemical properties ($P > 0.05$) at four maize growth periods (R1, R9, V1, and V6) in saline and control soil environments from 2012 to 2014 (Table 5). However, there were significant differences between WL-73 and WT plants ($P < 0.05$) for nine combinations of factors and very significant differences between the genotypes for three combinations of factors. There were significant differences in specific parameters under some conditions, including pH in saline soil at V9 in 2013; pH in control soil at R1 in 2014; SAR in saline soil at V9 in 2012 and 2014; AN in saline soil at V9 in 2013 and in saline soil at R6 in 2014; AK in control soil at V9 in 2013; AP in saline soil at R1 in 2013; and organic matter in saline soil at V9 in 2013. There were very significant differences in SAR in control soil at V9 in 2014; AK in saline soil at V9 in 2013; and AK in saline soil at R6 in 2013 (Table 3; Figures 6(a)–6(g)). However, no overall consistent trends in physicochemical properties were detected in the two soil environments, over the three-year study, or in the four growth stages. This result is consistent with previous results in other plants and fields [32].

3.5. Culturable Microbial Populations in Rhizosphere Soil. The number of total actinomycetes, bacteria, and fungi per gram of dry rhizosphere soil from WL-73 and WT plants over the crop developmental cycle is shown in Figure 7. No significant differences were found in the total number of bacteria, fungi, and actinomycetes between the two maize rhizosphere soils at any of the plant growth stages, except for actinomycetes in saline soil at V9 and in control soil at R6 in 2014 (Table 3). The variation in actinomycete, bacterial, and fungal populations was consistent. The three microbial populations increased from growth stages V3 through V9. The total number of each of the three kinds of microbes in the rhizosphere soil peaked at the V1 stage. Subsequently, the populations of all three types of microorganisms decreased at the V6 stage (Figures 7(a)–7(c)).

Soil microbial analysis is a common method used to detect the effect of exogenous chemicals or environmental pollutants on soil fertility and crop yields. Similarly, monitoring soil microbial populations in response to transgenic plants will reveal the risks of exogenous genes in soil. Investigations of the microbial populations of rhizosphere

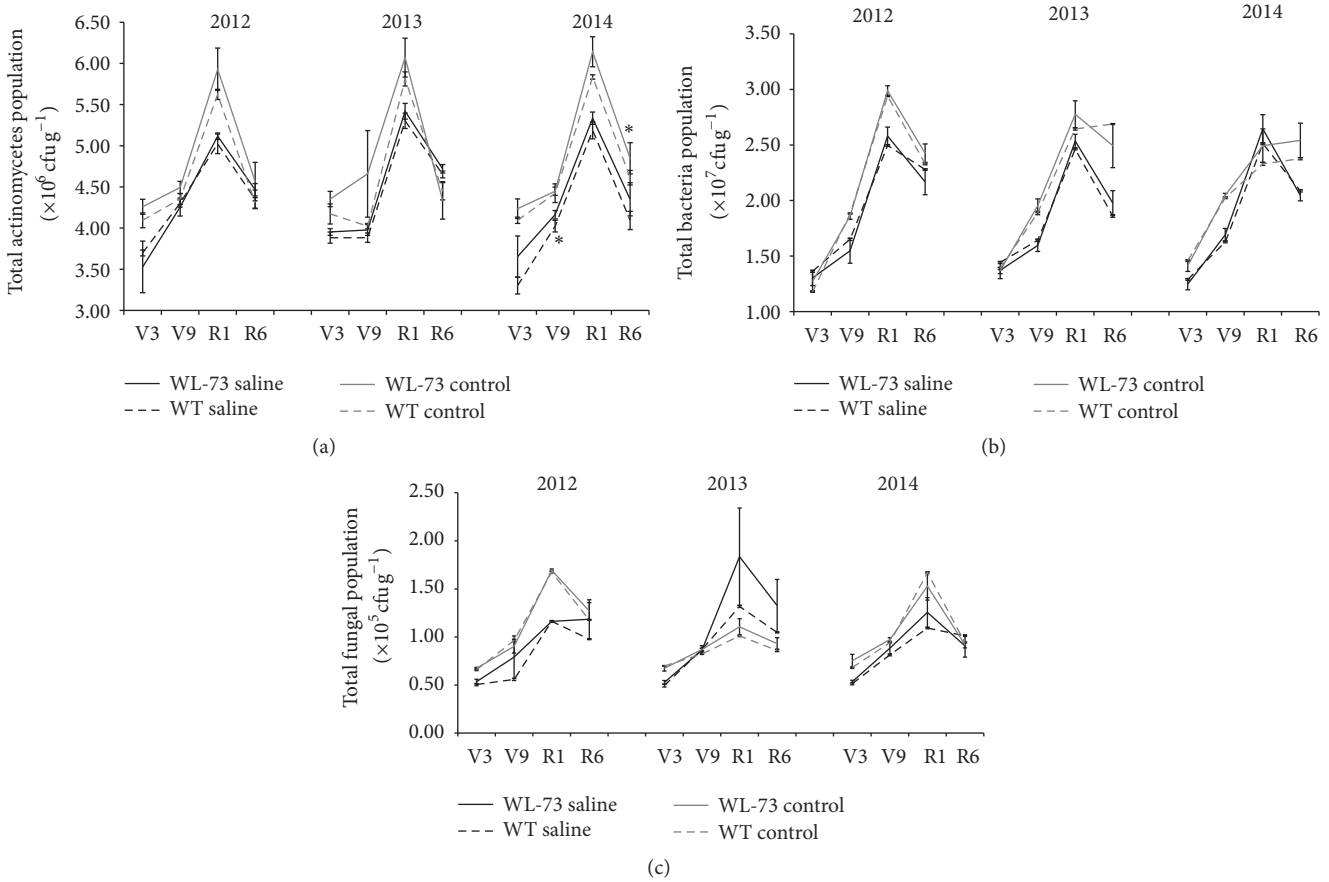


Figure 7: Assays of three microbial communities in rhizosphere soil. *BcWRKY1* transgenic maize (WL-73) and nontransgenic maize (WT; LH1037) plants were grown in saline or control nonsaline soil from 2012 to 2014 ($^*P < 0.05, n = 3$). We investigated the microbial populations in their rhizosphere soil. (a) Actinomycete population, (b) bacterial population, and (c) fungal population in terms of cfu (colony forming units) per g of dry soil. WT, wild-type maize LH1037; WL-73, LH1037 plant transformed with *BcWRKY1*; V3, the three lowest leaves have a visible collar; V9, nine leaves have collars present; R1, silking; R6, physiological maturity. $*$ indicates a significant difference at $P < 0.05$ according to the LSD test ($n = 3$). The standard error is based on the average of three biological replicates from 2012 to 2014. Note 1: the variation in the actinomycete, bacterial, and fungal populations was consistent. The populations of the three microbial communities increased from growth stages V3 through V9. The total number of each of the three types of microbes in the rhizosphere soil peaked at the V1 stage. Subsequently, the population levels of all three types of microorganisms decreased at the V6 stage. Note 2: to ensure consistency in the experimental data, transgenic and nontransgenic material were planted in the same pot for the three years, and each pot was handled at the same time to minimize the impact of human factors on the experiment.

soil found that *Bt* maize had no direct effect on soil ecology [34]. Both the number and the diversity of microorganisms exhibit only significant seasonal variation, with no long-term effect on the cultivation of Cry1Ac-transgenic cotton [32]. No significant effects were found on the populations of various soil microorganisms with the growth of transgenic insect-resistant maize, *Bt* maize, and cotton compared to nontransgenic plants under field conditions [30, 31, 35–37]. There was no adverse effect on soil enzymatic activities or rhizosphere microbial communities by the cultivation of transgenic plants, such as *MCM6* transgenic tobacco, *PDH45* transgenic rice, and *SUV3*-overexpressing transgenic rice [3, 7, 37].

In our study, significant variation was detected in actinomycete populations in saline soil at V9 and in control soil at R6 in 2014 ($P > 0.05$). However, we did not find a significant effect on enzyme activities, physicochemical properties, or populations of soil microbes due to the long-term cultivation of WL-73 compared to WT. Our results are consistent with previous studies showing that the long-term cultivation of salt-tolerant GM plants has no effect on soil microbial populations. The effects that we observed were due to particular individual plants, techniques, exogenously expressed proteins, or environmental conditions.

4. Conclusions

In the present study, the minor significant differences in the rhizosphere soil between transgenic and nontransgenic maize plants were not as large as the effects associated with

Table 3: ANOVA of the effects of year, growth stage, soil type, and maize genotype.

Number	Source	Year	Growth stage	Soil type	WL-73	WT	*P* value
1	pH	2013	V9	Saline	8.83	8.66	0.03*
2	pH	2014	R1	Control	7.36	7.13	0.02*
3	SAR	2012	V9	Saline	21.43	21.07	0.01*
4	SAR	2014	V9	Saline	17.02	17.33	0.03*
5	SAR	2014	V9	Control	2.14	2.78	0.01**
6	AN	2013	V9	Saline	87.33	81.69	0.05*
7	AN	2014	R6	Saline	87.33	85.60	0.03*
8	AK	2013	V9	Saline	215.83	219.58	0.01**
9	AK	2013	V9	Control	210.45	213.99	0.02*
10	AK	2013	R6	Saline	194.23	191.55	0.01**
11	AP	2013	R1	Saline	33.63	31.66	0.04*
12	OC	2013	V9	Saline	31.59	33.96	0.03*
13	Alkaline phosphatase	2013	R6	Saline	34.79	33.80	0.03*
14	Urease activity	2012	R1	Control	5.76	6.49	0.04*
15	Urease activity	2013	V9	Saline	12.87	12.65	0.05*
16	Dehydrogenase activity	2013	V9	Control	8.86	7.14	0.03*
17	Sucrase activity	2014	V9	Saline	25.04	24.96	0.02*
18	Actinomycetes	2014	V9	Saline	4.16	4.02	0.05*
19	Actinomycetes	2014	R6	Control	4.85	4.61	0.05*

SAR, sodium adsorption ratio; AN, available nitrogen; AK, available potassium; OC, organic carbon; WT, wild-type maize LH1037; WL-73, LH1037 plant transformed with *BcWRKY1*; V3, the three lowest leaves have a visible collar; V9, nine leaves have collars present; R1, silking; R6, physiological maturity.
* indicates a significant difference at $P < 0.05$ according to the LSD test ($n = 3$).
** indicates a significant difference at $P < 0.01$ according to the LSD test ($n = 3$).
Note: only those traits with significant differences in a specific period are listed here.

Table 4: ANOVA of four enzyme activities in rhizosphere soils.

Enzyme activities	Source of variation	*P* value
Alkaline phosphatase	Year	0.87
	Soil variety	0.01**
	Maize variety	0.90
	Stage × maize variety	0.00**
Urease	Year	0.81
	Soil variety	0.00**
	Maize variety	0.52
	Stage × maize variety	0.00**
Dehydrogenase	Year	0.96
	Soil variety	0.87
	Maize variety	0.77
	Stage × maize variety	0.00**
Sucrase	Year	0.56
	Soil variety	0.01**
	Maize variety	0.88
	Stage × maize variety	0.00**

Years: 2012, 2013, and 2014.
Soil varieties: saline soil and nonsaline soil.
Maize varieties: WT (wild-type maize LH1037) and WL-73 (LH1037 plant transformed with *BcWRKY1*).
Stages: V3 (the three lowest leaves have a visible collar), V9 (nine leaves have collars present), R1 (silking), and R6 (physiological maturity).
** Significant source of variation ($P < 0.01$).

plant growth stages. These results indicated that the effects of *BcWRKY1* maize WL-73 on rhizosphere soil ecology are within the variation expected in conventional agriculture. The long-term planting (3 years) of WL-73 plants had no detectable effects on the enzymatic activities, physicochemical properties, or microbial populations of the rhizosphere soil compared with the WT at any of four maize growth stages (V3, V9, R1, and R6).

Competing Interests

The authors declare no conflict of interests.

Authors' Contributions

Xing Zeng and Yu Zhou contributed equally to this research.

Acknowledgments

The authors would like to thank Professor Zhongping Lin at the College of Life Science, Peking University, for his kind gift of the recombinant plasmid pCAMBIA3300-*Ubi-WRKY1*. This work was supported by the Major Genetically Modified Organism Breeding Project of China (2016ZX08011003); the Scientific Research Foundation for Returned Scholars of Heilongjiang Province, China (LC201411); and the Science Research Foundation of Northeast Agricultural University (2012RCB21).

Table 5: ANOVA of seven physicochemical properties.

Physicochemical properties	Source of variation	P value
EC	Year	0.9972
	Soil variety	0.012*
	Maize variety	0.156
pH value	Year	0.84
	Soil variety	0.001**
	Maize variety	0.742
SAR value	Year	0.56
	Soil variety	0.001**
	Maize variety	0.742
AN	Year	0.87
	Soil variety	0.052
	Maize variety	0.038*
AP	Year	0.62
	Soil variety	0.002**
	Maize variety	0.791
AK	Year	0.84
	Soil variety	0.033*
	Maize variety	0.935
OC	Year	0.95
	Soil variety	0.012*
	Maize variety	0.513

Years: 2012, 2013, and 2014.
Soil varieties: saline soil and nonsaline soil.
Maize varieties: WT (wild-type maize LH1037) and WL-73 (LH1037 plant transformed with *BcWRKY1*).
*Significant source of variation ($P < 0.05$).
**Significant source of variation ($P < 0.01$).

References

[1] B. B. Patnaik, S. Y. Park, S. W. Kang et al., "Transcriptome profile of the Asian giant hornet (*Vespa mandarinia*) using Illumina hiseq 4000 sequencing: *De novo* assembly, functional annotation, and discovery of SSR markers," *International Journal of Genomics*, vol. 2016, Article ID 4169587, 15 pages, 2016.

[2] J. M. Lynch, A. Benedetti, H. Insam et al., "Microbial diversity in soil: ecological theories, the contribution of molecular techniques and the impact of transgenic plants and transgenic microorganisms," *Biology and Fertility of Soils*, vol. 40, no. 6, pp. 363–385, 2004.

[3] R. K. Sahoo and N. Tuteja, "Effect of salinity tolerant PDH45 transgenic rice on physicochemical properties, enzymatic activities and microbial communities of rhizosphere soils," *Plant Signaling & Behavior*, vol. 8, no. 8, pp. 113–116, 2013.

[4] C. Snell, A. Bernheim, J.-B. Bergé et al., "Assessment of the health impact of GM plant diets in long-term and multigenerational animal feeding trials: a literature review," *Food and Chemical Toxicology*, vol. 50, no. 3-4, pp. 1134–1148, 2012.

[5] U. Mina, A. Chaudhary, and A. Kamra, "Effect of *Bt* cotton on enzymes activity and microorganisms in rhizosphere," *Journal of Agricultural Science*, vol. 3, no. 1, pp. 96–104, 2011.

[6] N. Ahmad, Z. K. Shinwari, S. Bashir et al., "Function and pulogenetic characterization of rhizospheric bacteria associated with GM and non GM maize," *Pakistan Journal of Botany*, vol. 45, no. 5, pp. 1781–1788, 2013.

[7] R. K. Sahoo, M. W. Ansari, R. Tuteja, and N. Tuteja, "Salt tolerant SUV3 overexpressing transgenic rice plants conserve physicochemical properties and microbial communities of rhizosphere," *Chemosphere*, vol. 119, pp. 1040–1047, 2015.

[8] S. Baumgarte and C. C. Tebbe, "Field studies on the environmental fate of the Cry1Ab Bt-toxin produced by transgenic maize (MON810) and its effect on bacterial communities in the maize rhizosphere," *Molecular Ecology*, vol. 14, no. 8, pp. 2539–2551, 2005.

[9] G. Liu, Y. Hu, F. Zhao, J. Zhu, and Z. Lin, "Molecular cloning of *BcWRKY1* transcriptional factor gene from *Boea crassifolia* hemsl and its preliminary functional analysis," *Acta Scientiarum Naturalium Universitatis Pekinensis*, vol. 43, no. 4, pp. 446–452, 2007.

[10] Q.-Y. Zhou, A.-G. Tian, H.-F. Zou et al., "Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants," *Plant Biotechnology Journal*, vol. 6, no. 5, pp. 486–503, 2008.

[11] C.-F. Niu, W. Wei, Q.-Y. Zhou et al., "Wheat WRKY genes *TaWRKY2* and *TaWRKY19* regulate abiotic stress tolerance in transgenic *Arabidopsis* plants," *Plant, Cell & Environment*, vol. 35, no. 6, pp. 1156–1170, 2012.

[12] C. Marè, E. Mazzucotelli, C. Crosatti et al., "Hv-WRKY38: a new transcription factor involved in cold and drought-response in barley," *Plant Molecular Biology*, vol. 55, no. 3, pp. 399–416, 2004.

[13] C. J. Shan, X. Liu, S. Y. Liu et al., "Identification of salt-tolerance of transgenic maize with *BcWRKY1* Gene," *Journal of Maize Sciences*, vol. 20, pp. 27–32, 2012.

[14] H. Di, Y. Tian, H. Zu, X. Meng, X. Zeng, and Z. Wang, "Enhanced salinity tolerance in transgenic maize plants expressing a BADH gene from Atriplex micrantha," *Euphytica*, vol. 206, no. 3, pp. 775–783, 2015.

[15] 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.

[16] D. I. Arnon, "Copper enzymes in isolated chloroplasts. Polyphenoloxidase in *Beta vulgaris*," *Plant Physiology*, vol. 24, no. 1, pp. 1–15, 1949.

[17] L. S. Bates, R. P. Waldren, and I. D. Teare, "Rapid determination of free proline for water-stress studies," *Plant and Soil*, vol. 39, no. 1, pp. 205–207, 1973.

[18] H. Min, Y.-F. Ye, Z.-Y. Chen, W.-X. Wu, and D. Yufeng, "Effects of butachlor on microbial populations and enzyme activities in paddy soil," *Journal of Environmental Science and Health, Part B*, vol. 36, no. 5, pp. 581–595, 2001.

[19] M. A. Tabatabai and J. M. Bremner, "Use of p-nitrophenyl phosphate for assay of soil phosphatase activity," *Soil Biology and Biochemistry*, vol. 1, no. 4, pp. 301–307, 1969.

[20] R. W. Weaver and S. H. Mickelson, *Microbiological and Biochemical Properties*, Soil Science Society of America, 1994.

[21] G. L. Miller, "Use of dinitrosalicylic acid reagent for determination of reducing sugar," *Analytical Chemistry*, vol. 31, no. 3, pp. 426–428, 1959.

[22] A. A. Nizova, "Study of sucrase activity in turf-podzol heavy-loamy soil," *Mikrobiologiia*, vol. 38, no. 2, pp. 336–339, 1969.

[23] P. K. Gupta, *Methods in Environmental Analysis: Water, Soil and Air*, Agrobios, 2004.

[24] A. Walkley and I. A. Black, "An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method," *Soil Science*, vol. 37, no. 1, pp. 29–38, 1934.

[25] Rathje, "Jackson, M. L.: *Soil Chemical Analysis*. Verlag: Prentice Hall, Inc., Englewood Cliffs, NJ. 1958, 498 S. DM 39.40," *Journal of Plant Nutrition & Soil Science*, vol. 85, no. 3, pp. 251–252, 1959.

[26] H. Rahimi, E. Pazira, and F. Tajik, "Effect of soil organic matter, electrical conductivity and sodium adsorption ratio on tensile strength of aggregates," *Soil and Tillage Research*, vol. 54, no. 3-4, pp. 145–153, 2000.

[27] A. L. Page, "Methods of Soil Analysis, Part 2: Chemical and Mineralogical Properties," *American Society of Agronomy*, 1982.

[28] G. D. Liang, C. H. Lu, Z. H. Wang et al., "Introduction of *BcWRKY1* gene mediated by agrobacterium tumefacien into shoot tip of maize," *Journal of Maize Sciences*, vol. 19, no. 3, pp. 56–58, 2011.

[29] S. Flores, D. Saxena, and G. Stotzky, "Transgenic *Bt* plants decompose less in soil than non-*Bt* plants," *Soil Biology and Biochemistry*, vol. 37, no. 6, pp. 1073–1082, 2005.

[30] B. S. Griffiths, S. Caul, J. Thompson et al., "Soil microbial and faunal community responses to *Bt* maize and insecticide in two soils," *Journal of Environmental Quality*, vol. 35, no. 3, pp. 734–741, 2006.

[31] R. F. Shen, H. Cai, and W. H. Gong, "Transgenic *Bt* cotton has no apparent effect on enzymatic activities or functional diversity of microbial communities in rhizosphere soil," *Plant and Soil*, vol. 285, no. 1-2, pp. 149–159, 2006.

[32] X. Li, B. Liu, J. Cui et al., "No evidence of persistent effects of continuously planted transgenic insect-resistant cotton on soil microorganisms," *Plant and Soil*, vol. 339, no. 1-2, pp. 247–257, 2011.

[33] R. L. Sinsabaugh, R. K. Antibus, and A. E. Linkins, "An enzymic approach to the analysis of microbial activity during plant litter decomposition," *Agriculture, Ecosystems and Environment*, vol. 34, no. 1–4, pp. 43–54, 1991.

[34] A. P. Oliveira, M. E. Pampulha, and J. P. Bennett, "A two-year field study with transgenic *Bacillus thuringiensis* maize: effects on soil microorganisms," *Science of the Total Environment*, vol. 405, no. 1–3, pp. 351–357, 2008.

[35] B. S. Griffiths, S. Caul, J. Thompson et al., "A comparison of soil microbial community structure, protozoa and nematodes in field plots of conventional and genetically modified maize expressing the *Bacillus thuringiens* is CryIAb toxin," *Plant and Soil*, vol. 275, no. 1-2, pp. 135–146, 2005.

[36] I. Icoz and G. Stotzky, "*Cry3Bb1* protein from *Bacillus thuringiensis* in root exudates and biomass of transgenic corn does not persist in soil," *Transgenic Research*, vol. 17, no. 4, pp. 609–620, 2008.

[37] V. Chaudhry, H. Q. Dang, N. Q. Tran et al., "Impact of salinity-tolerant MCM6 transgenic tobacco on soil enzymatic activities and the functional diversity of rhizosphere microbial communities," *Research in Microbiology*, vol. 163, no. 8, pp. 511–517, 2012.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 6404817, 9 pages
http://dx.doi.org/10.1155/2016/6404817

Research Article

Molecular Cloning, Characterization, and mRNA Expression of Hemocyanin Subunit in Oriental River Prawn *Macrobrachium nipponense*

Youqin Kong,[1,2] Liqiao Chen,[1] Zhili Ding,[1,2] Jianguang Qin,[3] Shengming Sun,[1] Ligai Wang,[1] and Jinyun Ye[2]

[1]*School of Life Sciences, East China Normal University, Shanghai 200062, China*
[2]*School of Life Sciences, Huzhou University, Huzhou, Zhejiang 313000, China*
[3]*School of Biological Sciences, Flinders University, Adelaide, SA 5001, Australia*

Correspondence should be addressed to Liqiao Chen; lqchen@bio.ecnu.edu.cn

Received 6 May 2016; Accepted 14 June 2016

Academic Editor: Wenwei Xiong

Hemocyanin is a copper-containing protein with immune function against disease. In this study, a hemocyanin subunit named MnHc-1 was cloned from *Macrobrachium nipponense*. The full-length cDNA of MnHc-1 was 2,163 bp with a 2,028-bp open reading frame (ORF) encoding a polypeptide of 675 amino acids. The MnHc-1 mRNA was expressed in the hepatopancreas, gill, hemocytes, intestine, ovary, and stomach, with the highest level in the hepatopancreas. In the infection trial, the MnHc-1 mRNA transcripts in the hemocytes were significantly downregulated at 3 h after injection of *Aeromonas hydrophila* and then upregulated at 6 h and 12 h, followed by a gradual recovery from 24 to 48 h. The MnHc-1 transcriptional expression in the hepatopancreas was measured after *M. nipponense* were fed seven diets with 2.8, 12.2, 20.9, 29.8, 43.1, 78.9, and 157.1 mg Cu kg^{-1} for 8 weeks, respectively. The level of MnHc-1 mRNA was significantly higher in the prawns fed 43.1–157.1 mg Cu kg^{-1} diet than in that fed 2.8–29.8 mg Cu kg^{-1} diet. This study indicated that the MnHc-1 expression can be affected by dietary copper and the hemocyanin may potentially participate in the antibacterial defense of *M. nipponense*.

1. Introduction

Crustaceans like other invertebrates only have innate immunity, including many immune molecules to eliminate exogenous pathogens [1]. To date, many studies have verified that hemocyanin is an important nonspecific innate immune defense molecule and can provide an effective immune defense in arthropods [2–8]. Hemocyanin is a copper-containing multifunctional protein in mollusks and arthropods [9]. Its primary function is to transport and store oxygen and also to participate in osmoregulation, molt cycle, exoskeleton formation, and melanin synthesis [10–12]. So far, studies on the immunologic function in crustacean are mainly focused on the hemocyanin itself or its degraded peptides, which have hemolytic activity, agglutination property, and antiviral function [2, 4, 5, 7, 13]. Hemocyanin genes have been cloned and characterized in certain crustacean species such as *Litopenaeus vannamei* [14], *Eriocheir sinensis* [15], *Cherax quadricarinatus* [16], *Homarus americanus* [17], and *Caridina multidentata* and *Atyopsis moluccensis* [18].

Copper is a central component of hemocyanin and a cofactor of many other enzymes, like superoxide dismutase, cytochrome oxidase, and lysyl oxidase [19, 20]. The dietary copper level can affect crustacean immune responses in *Penaeus monodon* [21] and *E. sinensis* [22]. It was found that the expression of hemocyanin mRNA in *E. sinensis* was affected by dietary copper level [15]. Investigation of the functional relationship between dietary copper and hemocyanin mRNA expressions can provide better understanding on crustacean innate immunity and offer insight into disease control through dietary management in shrimp farming. We hypothesize that copper as a major component of hemocyanin can impact prawn innate immunity through dietary copper manipulation.

Oriental river prawn (*Macrobrachium nipponense*) is an important aquaculture species in China and other Southeast Asian countries [23]. Various diseases have been found in the *M. nipponense* farming population due to intensive culture and environmental pollution [24]. For example, the bacterial disease induced by *Aeromonas hydrophila* is one of the major diseases which can cause 30% death of prawn, sometimes as high as 70% [25]. As a result, the investigation on the *M. nipponense* innate immune mechanism against *A. hydrophila* has become a key issue of health management in crustacean farming. The hemocyanin subunit is a functional group for crustacean immunity, and a hemocyanin subunit has been cloned and characterized in the freshwater prawn *M. nipponense* [26]. In this study, we discovered another hemocyanin subunit in *M. nipponense* (MnHc-1). To understand the role of MnHc-1 in *M. nipponense* immunization, (1) the full-length cDNAs of MnHc-1 was cloned, (2) the distribution of MnHc-1 in different tissues was examined, (3) the mRNA expression of MnHc-1 in immune defense after prawn challenge with *A. hydrophila* was examined, and (4) the relative expression level of MnHc-1 in the hepatopancreas of *M. nipponense* fed different levels of copper was quantified.

2. Materials and Methods

2.1. Sample Collection and Infection Test. All adult oriental river prawns (*M. nipponense*) were obtained from a local farm in Shanghai. The hepatopancreas, muscle, gill, ovary, intestine, heart, and stomach were collected from healthy prawn, flash-frozen in liquid nitrogen, and stored at −80°C until RNA extraction. A total of 1 mL hemolymph was collected from the ventral sinus using a sterile syringe and diluted using half volume of anticoagulant solution [27], then centrifuged at 8000 ×g for 10 min at 4°C to collect the hemocyte, and stored at −80°C immediately until RNA extraction.

Prior to the challenge experiment, the adult prawns were acclimatized in the laboratory for 2 weeks. A total of 300 healthy prawns were randomly divided into two groups with five replicates. According to the preliminary experiment, the prawn in the bacterial challenge trial was injected with 100 μL *A. hydrophila* in saline suspension (1×10^7 CFU/mL) obtained from Shanghai Ocean University, while each prawn in the control group received the same volume of saline injection. After injection, prawns were put back to the rearing tanks, and hemocyte samples were collected at 0, 3, 6, 12, 24, and 48 h after injection, centrifuged, and stored at −80°C for RNA extraction.

2.2. Experimental Diets and Farming. The juvenile *M. nipponense* were obtained from the same farm as the adults and acclimated for two weeks in the laboratory conditions prior to the feeding trial. The basal diet was supplemented with copper sulphate (Analytical Reagent, Shanghai Chemical Co., Shanghai, China) at 0, 10, 20, 30, 40, 80, and 160 mg Cu kg^{-1} diet, respectively. Procedure of diet preparation was similar to that described by Li et al. [28], and all other required nutrients for *M. nipponense* were included. The actual copper concentrations in the feeds were analyzed to be 2.8, 12.2, 20.9, 29.8, 43.1, 78.9, and 157.1 mg kg^{-1}, respectively, by the flame atomic absorption photometry [29]. The compositions of the experimental diet were showed in Table 1.

Table 1: Ingredients and compositions of experimental diets (%).

Ingredient	Percentage of dry weight
Casein[a]	30
Fish meal[b]	20
Corn starch	26
Fish oil[c]	4
Soybean oil[d]	2
Vitamin mix[e]	2
Cu-free mineral mix[f]	3
Attractant[g]	3
Cholesterol[h]	0.5
Choline chloride[h]	0.5
Lecithin[h]	0.5
Cellulose[h]	6.5
Sodium carboxymethylcellulose[h]	2
Proximate composition	
Crude protein	40.6
Crude lipid	7.47
Crude ash	7.02

[a]Sigma-Aldrich Co., Shanghai, China.
[b]Tecnologica De Alimentos USA.
[c]Xiamen Xinsha Pharmaceutical Co. Ltd., Xiamen, China.
[d]National Golden Dragon Fish Co. Ltd., Shanghai, China.
[e]Vitamin mixture (mg/100 g mixture): vitamin A 420000 IU; vitamin C 6000 mg; α-tocopherol acetate 2000 mg; vitamin D3 120000 IU; vitamin K 1000 mg; vitamin B1 1000 mg; vitamin B2 1000 mg; vitamin B6 1600 mg; vitamin B12 2 mg; niacin 5000 mg; folic acid 400 mg; inositol 6000 mg; biotin 10 mg; and calcium pantothenic 3500 mg, Hangzhou Minsheng Bio-Tech Co., Ltd., China.
[f]Composition of mineral mixture (g/kg diet): KCl 0.84, $MgSO_4 \cdot 7H_2O$ 3, NaH_2PO_4 6.45, KH_2PO_4 3, $Ca(H_2PO_4)_2 \cdot H_2O$ 7.95, $CaCO_3$ 3.15, $C_6H_{10}CaO_6 \cdot 5H_2O$ 4.95, $FeC_6H_5O_7 \cdot 5H_2O$ 0.36, $ZnSO_4 \cdot 7H_2O$ 0.1428, $MnSO_4 \cdot H_2O$ 0.0321, Na_2SeO_3 0.0009, $AlCl_3 \cdot 6H_2O$ 0.0045, $CoCl_2 \cdot 6H_2O$ 0.042, and KI 0.0069.
[g]Alanine 0.6%, glycine 0.6%, glutamic acid 0.6%, and betaine 1.2%.
[h]China National Medicine Corporation Co., Ltd., Beijing, China.

Prawn juveniles (0.101 ± 0.001 g) were randomly placed in 21 of 300-L tanks with 30 prawns per tank in triplicate. Prawns were fed to apparent satiation twice daily (8:00 and 17:00 h) for 56 days. To maintain water quality, one-third of the tank water was exchanged daily. During the feeding period, water temperature was 27–30°C, dissolved oxygen >6.5 mg L^{-1}, and total ammonia nitrogen <0.1 mg L^{-1}. The Cu concentration in rearing water was 1.4–1.7 μg Cu L^{-1}. At the end of the feeding trial, all prawns were counted and gained survival rate (Survival rate = 100 × (final prawn number)/(initial prawn number)). Prawns were fasted for 24 h before the hepatopancreas was collected. These samples were stored at −80°C until RNA extraction.

2.3. RNA Extraction and Reverse Transcription. Total RNA was isolated using RNA extraction kit (Aidlab Biotech, Beijing, China) following manufacture protocol. In reverse transcription reaction, 3 μg of total RNA was used for synthesis of first-strand cDNA by the PrimeScript™ RT-PCR

Table 2: Primers used in our study.

Name	Sequence (5′-3′)
MnHc-s1	GTCGACTCTACTCCTCTTGG
MnHc-a1	TCGGTTATCCTTCAGCTC
MnHc-s2	TTCTGCTGATGCCTCCAA
MnHc-a2	TTCTTCACGGTGCCTGTC
MnHc-5′GSP	GGCGGTGAACAGCTTCTCCATTATC
MnHc-3′GSP	GCATGATGTGAACTTCCTCCTGTGG
5′-RACE CDS primer A	$(T)_{25}$V N (N = A, C, G or T; V = A, G or C)
3′-RACE CDS primer A	AAGCAGTGGTATCAACGCAGAGTAC$(T)_{30}$ V N
UPM	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT CTAATACGACTCACTATAGGGC
β-Actin-s	GTGCCCATCTACGAGGGTTA
β-Actin-a	CGTCAGGGAGCTCGTAAGAC

Kit (TaKaRa, Dalian, China). Acquired cDNA was stored at −20°C for subsequent quantitative real-time PCR (qRT-PCR).

2.4. 3′-RACE and 5′-RACE Amplification of MnHc-1 Gene, Cloning, and Sequencing. To get the full length of MnHc-1 cDNA sequence, SMART™ RACE cDNA Amplification Kit (Clontech, USA) was used to conduct the 3′-RACE and 5′-RACE. Gene-specific primers of MnHc-1 were designed based on the known fragments initially identified from the EST cDNA library of *M. nipponense* [30]. The total RNA of the mixed hepatopancreas was used as a template. PCR was performed in a Bio-Rad thermal cycler. Reaction volume was 50 μL containing 5.0 μL 10x Ex Taq Buffer (Mg^{2+} Plus), 5.0 μL 10x UPM, 4.0 μL cDNA template, 4.0 μL dNTPs mix (2.5 mM each), 1.0 μL GSP (10 μM), 0.26 μL Ex Taq HS (TaKaRa, Dalian, China), and 30.74 μL of sterile deionized water. The PCR conditions were set as follows: 5 cycles of 94°C for 30 s and 72°C for 3 min, then 5 cycles of 94°C for 30 s and 70°C for 30 s, and 25 cycles of 94°C for 30 s, 68°C for 30 s, and 72°C for 3 min. All used primers were summarized in Table 2. Primers of MnHc-s1 and MnHc-a1 were used to confirm the MnHc-1 fragment from the EST cDNA library. 5′-CDS primer A (as the RT primer), gene-specific primer of MnHc-5′GSP, and the UPM (universal primer A mix) were used for the 5′-RACE. 3′-CDS primer A (as the RT primer), gene-specific primer of MnHc-3′GSP, and the UPM were used for the 3′-RACE.

The PCR products were purified by the UNIQ-10 Gel Extraction Kit (Sangon, Shanghai, China) and cloned into the pUCm−T vector (Sangon, Shanghai, China). The transformed bacteria were identified, confirmed by blue/white screening, and validated by PCR. More than two recombinant plasmids were sequenced using BigDye Terminator v3.1 Cycle Sequencing Kit operating on an Automatic DNA Sequencer (ABI 3730xl DNA Analyzer).

2.5. Sequence Analysis. The online ORF finder program (http://www.ncbi.nlm.nih.gov/gorf/) was used to predict the gene's putative open reading frame. The BLAST algorithm at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/) was applied to search for sequence homology. SMART (Simple Modular Architecture Research Tool, http://smart.embl-heidelberg.de/) was used to predict conserved motifs. Signal sequence was carried out using SignalP 4.1 program (http://www.cbs.dtu.dk/services/SignalP/). The deduced amino acid sequence of MnHc-1 from *M. nipponense* and other invertebrate hemocyanin sequences acquired from NCBI database were aligned by software ClustalX. A neighbor-joining (NJ) phylogenetic relationship was established based on amino acid sequences of the hemocyanin using the MEGA 5.1 program (http://www.megasoftware.net/).

2.6. Analysis of the MnHc-1 Expression in Tissues. The mRNA expression of MnHc-1 in the hepatopancreas, muscle, gill, ovary, intestine, heart, and stomach was detected by qRT-PCR with β-actin as internal control. The synthesis of the first-strand cDNA was the same as that described above. Gene-specific primers of MnHc-s2, MnHc-a2, β-actin-s, and β-actin-a were used in qRT-PCR (Table 2). The qRT-PCR liquid compositions and conditions were according to the manufacturer instructions of SYBR Premix Ex Taq (TaKaRa, Dalian, China). The qRT-PCR was conducted on the CFX96™ Real-Time System (Bio-Rad, USA) according to the manufacturer's protocols. During the detection, each sample was run in triplicate. The melt curve of the amplification products was analyzed to ensure that only one PCR product was amplified and detected at the end of each PCR. Expression level of MnHc-1 was measured by $2^{-\Delta\Delta Ct}$ method [31].

2.7. Analysis of the MnHc-1 Expression after A. hydrophila Challenge. The mRNA expression of MnHc-1 in the hemocytes of *M. nipponense* injected with *A. hydrophila* or saline water was, respectively, detected at 0, 3, 6, 12, 24, and 48 h after injection by a quantitative real-time RT-PCR.

2.8. Analysis of the MnHc-1 Expression after the Feeding Treatment at Different Dietary Copper Levels. At the feeding experiment, mRNA expression of MnHc-1 in hepatopancreas of *M. nipponense* from the feeding treatment at seven copper levels was analyzed by qRT-PCR.

2.9. Statistical Analysis. All the data are expressed as means ± SD. SPSS software (version 16.0) was used for statistical analysis. The results of relative mRNA expression in challenge test were analyzed by t-test, while other results were subjected to one-way ANOVA and post hoc Duncan multiple range tests. Differences were regarded as significant at $P < 0.05$.

3. Results

3.1. Sequence Analysis of MnHc-1. The full length of MnHc-1 cDNA from the *M. nipponense* was 2,163 bp (GenBank accession number JX456149.1), containing a 20 bp 5′-untranslated region, a 115 bp 3′-untranslated region with poly A tail, and a 2,028 bp open reading frame (ORF) (Figure 1). The ORF encoded a polypeptide of 675 amino acids with a calculated molecular weight of 78.060 kDa and an isoelectric point of 5.45. A putative signal peptide of 21 amino acids was found in the N-terminus by using SignalP 4.1 program [32]. SMART program predicted that the MnHc-1 belongs to the hemocyanin family, including N terminal domain (Ser24-Val150), typical copper-containing domain (Pro154-Glu411), and C terminal domain (Pro417-His667 amino acid). Six histidine residues (H212, H216, H242, H362, H366, and H402) of copper-binding sites were identified [17].

3.2. Homology and Phylogenetic Analysis of MnHc-1. Homology analysis with BLAST algorithm showed that MnHc-1 amino acid sequences had 75%, 73%, 70%, and 68% similarity to that of the *A. moluccensis* gamma subunit, *P. monodon*, *C. quadricarinatus,* and *E. sinensis*, respectively. The MnHc-1 exhibited 64% identity with another subunit of hemocyanin in *M. nipponense* (MnHc-2, GenBank accession number JF683437.1). A phylogenetic tree was constructed with the 24 full-length hemocyanin sequences from arthropod based on the neighbor-joining method (Figure 2). The result of phylogenetic analysis revealed that MnHc-1 was more closely related to the hemocyanin gamma subunit 1 of freshwater shrimps *A. moluccensis* and *C. multidentata*.

3.3. Analysis of MnHc-1 mRNA Expressions in Tissues. The mRNA transcripts of MnHc-1 were analyzed in the tissues of hepatopancreas, gill, muscle, hemocytes, intestine, ovary, and stomach. In prawn, the highest expression was found in hepatopancreas ($P < 0.05$); the expression value is 20.32, which is 20 times higher or more compared with other tissues. The value is as low as 0 so that the gene is not expressed in muscle (Figure 3). It looks to us that its expression is tissue-specific.

3.4. Analysis of MnHc-1 mRNA Expression after Challenge with A. hydrophila. Figure 4 exhibits the expression profile of MnHc-1 in the hemocytes challenged by *A. hydrophila* from 0 to 48 h. The expression of MnHc-1 showed a distinct time-dependent pattern. The mRNA expression level significantly dropped at 3 h after injection (0.6 times) ($P < 0.05$) and then started to significantly increase and reached the peak in 12 h (6.2 times) ($P < 0.05$), followed by a recovery to the initial level in 24 h and 48 h.

3.5. Analysis of MnHc-1 mRNA Expression in Response to Graded Levels of Dietary Copper. Survival rate (70–81%) of prawns was not affected by the dietary copper levels. As shown in Figure 5, the response of MnHc-1 expression in the hepatopancreas was affected by the dietary copper level. With the dietary copper level increasing from 43.1 to 157.1 mg Cu kg^{-1} diet, the level of MnHc-1 mRNA in hepatopancreas of prawns significantly increased (1.9-fold to 5.8-fold) ($P < 0.05$) and was significantly higher compared to that fed 2.8–29.8 mg Cu kg^{-1} diet ($P < 0.05$), but differences in 2.8–29.8 groups were not significant ($P > 0.05$).

4. Discussion

Hemocyanin is an extracellular, multisubunit protein in crustacean [14, 33]. Those subunits differ considerably in their primary structures and are encoded by distinct genes [34]. In this study, we cloned and characterized the expression pattern of one hemocyanin subunit from *M. nipponense* (MnHc-1). MnHc-1 was a polypeptide of 675 amino acids with a 21-amino acid putative signal peptide. The signal peptide ends Ala-X-Ala motif, which is a frequent accordance prior to the cleavage site of signal peptides, suggesting that a cleavage site is located at the 21-22 amino acids [32, 35]. Structurally, MnHc-1 has conservative copper-binding domains including six histidine residues (H212, H216, H242, H362, H366, and H402) of the copper-binding sites; this domain agrees with other crustaceans [4, 14, 15]. Based on immunological methods, the crustacean hemocyanin subunits are classified into three distinct subunit types: alpha, beta, and gamma [9]. The phylogenetic analysis showed that MnHc-1 and MnHc-2 belong to separate clade; MnHc-1 belongs to the gamma subunit which has evolved at a later time compared to alpha and beta subunits in freshwater shrimps [18].

The present study showed that the highest level of MnHc-1 mRNA expression occurred in the hepatopancreas. The result is the same as the findings in other crustaceans, such as *H. americanus* [17], *Fenneropenaeus chinensis* [36], and *E. sinensis* [15], and consistent with the report that hemocyanin synthesis occurs mainly in the hepatopancreas [37]. MnHc-1 expression was detected in all the examined tissues except for muscle, which is different from another subunit of hemocyanin in *M. nipponense* (MnHc-2) expressed in the muscle [26]. We also found that MnHc-1 was expressed in ovary, whereas MnHc-2 was hardly expressed in ovary [26]. The discrepant expression patterns of these two subunits of hemocyanin may be owing to their functional specialization in different tissues.

Hemocyanin is an important multifunctional protein in mollusks and arthropods. Besides its role as an oxygen carrier, its immune functions including antibacterial activities, agglutination property, and PO activity have become hot topics of immunological research [2, 4, 5, 7]. Zhang et al. [38] found that the main protein directly bound to the *Vibrio alginolyticus*, *Vibrio harveyi*, *A. hydrophila*, and *Staphylococcus aureus* in *L. vannamei* serum was hemocyanin, suggesting that hemocyanin possesses antibacterial functions. The C-terminus of *L. vannamei* hemocyanin is possibly related to

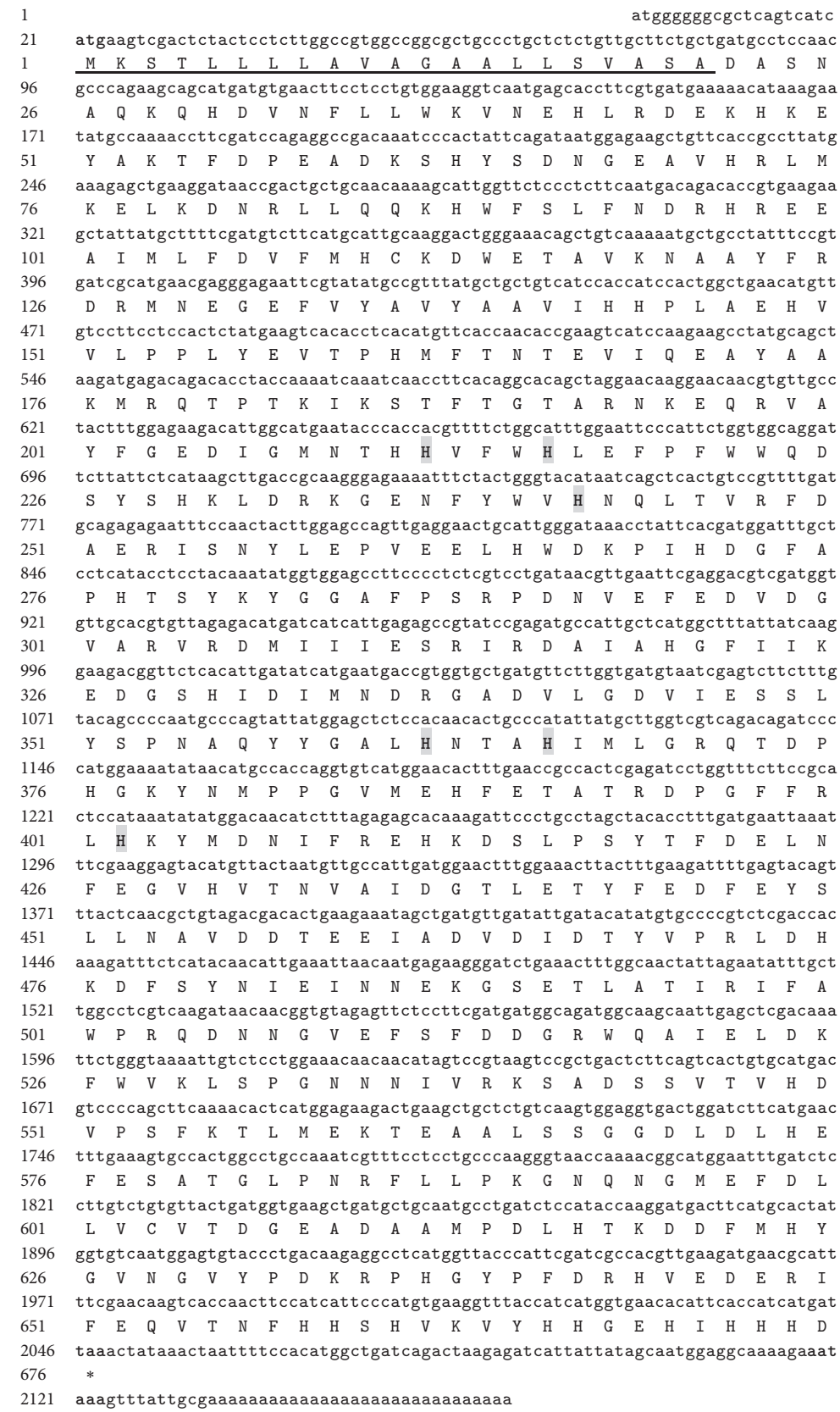


Figure 1: *M. nipponense* hemocyanin subunit 1 cDNA and deduced amino acid sequences. Putative signal peptide sequences were underlined. The start codon (atg), stop codon (taa), and the polyadenylation signal sequence (aataaa) were marked in bold and the six histidine residues within the copper-binding sites were marked in bold and shadow background.

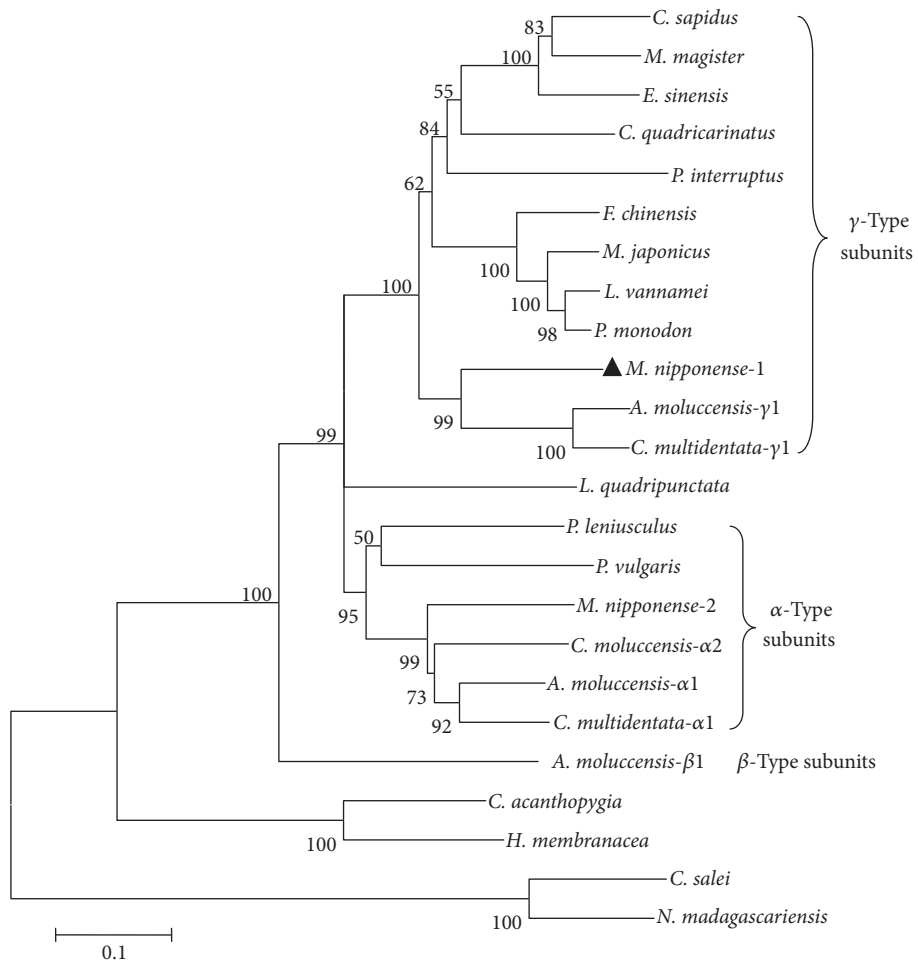


Figure 2: The phylogenetic tree based on the sequences of hemocyanin from different species. The amino acid sequences were derived from the GenBank as the following accession numbers: *A. moluccensis*-α1 (CCF55379.1); *A. moluccensis*-β1 (CCF55382.1); *A. moluccensis*-γ1 (CCF55383.1); *C. acanthopygia* (CAR85694.1); *C. multidentata*-α1 (CCF55384.1); *C. multidentata*-α2 (CCF55385.1); *C. multidentata*-γ1 (CCF55387.1); *C. quadricarinatus* (AFP23115.1); *C. salei* (CAC44753.1); *C. sapidus* (AAF64305.1); *E. sinensis* (AEG64817.1); *F. chinensis* (ACM61982.1); *H. membranacea* (CAR85695.1); *L. quadripunctata* (ADE58571.1); *L. vannamei* (ADZ15149.1); *M. japonicus* (ABR14693.1); *M. magister* (AAW57893.1); *M. nipponense*-1 (AGA17871.1); *M. nipponense*-2 (AEC46861.1); *N. madagascariensis* (CAD68057.1); *P. interruptus* (AAB22190.1); *P. leniusculus* (AAO47336.1); *P. monodon* (AEB77775.1); and *P. vulgaris* (CAC69244.1).

the immunity in shrimp to different pathogens [39]. Gram-negative bacteria *A. hydrophila* are a common species of the *Aeromonas* genus in water and water habitants [40]. The infection of *A. hydrophila* in fish and prawn including *M. nipponense* has been one of the major diseases under farming conditions [25, 41]. Sun et al. [15] showed that the hemocyanin gene expression of *E. sinensis* was significantly upregulated by *A. hydrophila* infection. In our study, temporal and spatial expressions of MnHc-1 in the hemocytes of prawn infected with *A. hydrophila* showed a clear time-dependent pattern. The level of MnHc-1 mRNA expression significantly decreased at 3 h after injection, then started to significantly increase after 6 h and 12 h, and then reached the peak at 12 h, implying that the hemocyanin is involved in the antibacterial defense of prawn. In another study, the level of MnHc-2 mRNA expression in prawn significantly increased over time and peaked at 3 h after the *A. hydrophila* challenge [26]. Clearly, these two hemocyanin subunits respond quite differently in defense against bacterial infection. Transcriptional upregulation in MnHc-1 was found after 6 h of *A. hydrophila* injection. Lei et al. [4] found a similar result in *P. japonicus* that *PjHcL* transcriptional upregulation occurred after 4 h of injection of the active WSSV. It is possible that *PjHcL* may be triggered by the fast expressed proteins in virus. It is also likely that MnHc-1 may be induced by fast expressed protein in bacteria, but the detailed defensing mechanism of hemocyanin against bacterial infection needs further study.

The expression of hemocyanin subunits varies with environmental or nutritional changes [42, 43]. Copper is the metal in the center of a hemocyanin molecule [42]. In *E. sinensis*, expression of hemocyanin mRNA was affected by the level of dietary copper [15]. Our present study showed that the level of dietary copper affected the hemocyanin gene expression in prawn. The level of MnHc-1 mRNA

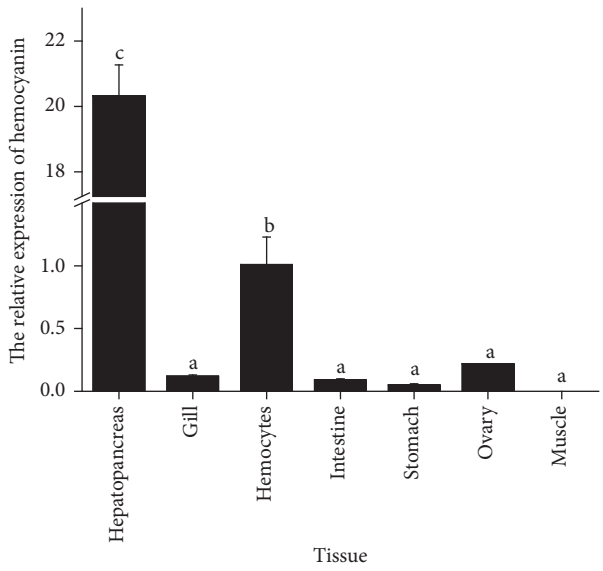


Figure 3: Quantitative real-time PCR analysis of hemocyanin expressions in various tissues of *M. nipponense*. The β-actin gene was used as the internal control. Different letters in each index indicated significant differences ($P < 0.05$).

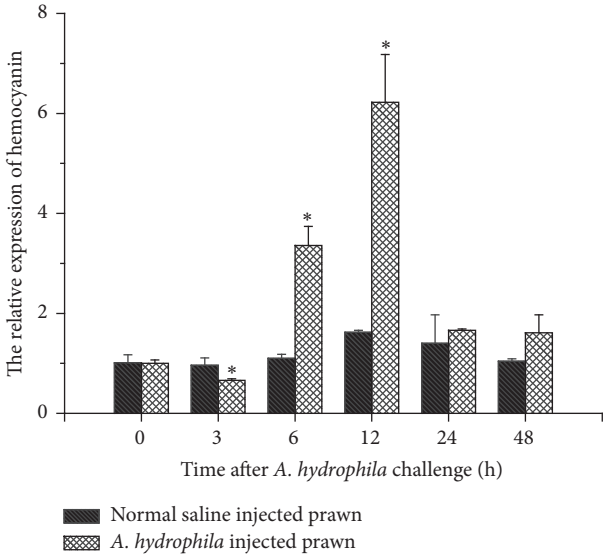


Figure 4: The hemocyanin gene expression in hemocytes of *M. nipponense*. The statistical test was performed by t-test after challenge with *A. hydrophila* compared to the control at the same time points (0, 3, 6, 12, 24, and 48 h). The internal standard was β-actin gene. Asterisks indicate being significantly different ($P < 0.05$).

in hepatopancreas of the prawn fed 43.1–157.1 mg Cu kg^{-1} diet was significantly higher compared to that fed 2.8–29.8 mg Cu kg^{-1}. The increase of dietary Cu concentration also increased the Cu content in hepatopancreas, especially in the 43.1–157.1 mg Cu kg^{-1} groups. Hemocyanin synthesis mainly occurs in the hepatopancreas [37]. Therefore, we suggest that the high level of copper content in the hepatopancreas can trigger MnHc-1 mRNA expression, whereas

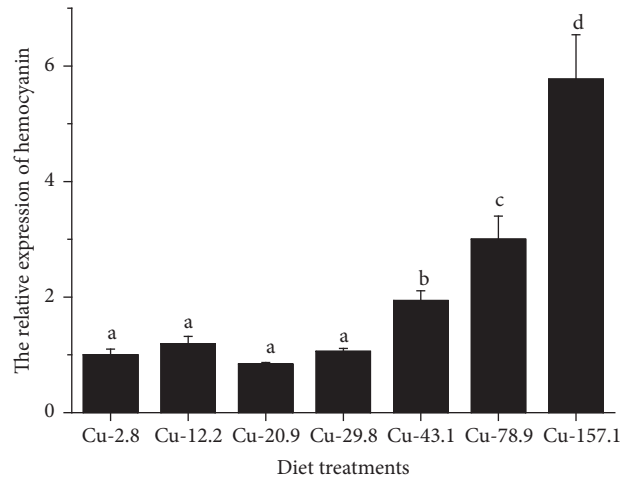


Figure 5: Relative hemocyanin mRNA levels in hepatopancreas of *M. nipponense*. *M. nipponense* was fed diet with different levels of copper for 8 weeks. Hemocyanin mRNA levels were evaluated by qRT-PCR and expressed relatively to the level of β-actin mRNA. Different letters in each index indicated significant differences ($P < 0.05$).

the optimal level of dietary copper (20–40 mg kg^{-1} diet) can increase hemocyanin mRNA expression in the hepatopancreas and hemocytes of crab [15]. These studies suggest that the effect of dietary copper level on hemocyanin mRNA expression is species-specific. It seems that the two subunits of hemocyanin may have structural functions in their own hemocyanin and show different response to the level of dietary copper. Hemocyanin is a copper-containing protein that mainly carries oxygen in crustaceans [9]. Dietary copper may be first used for hemocyanin synthesis as low dietary copper did not reduce the level of MnHc-1 mRNA.

In conclusion, we cloned the hemocyanin subunit gene (MnHc-1) from the hepatopancreas of *M. nipponense*. Our results suggest that MnHc-1 may play a critical role in antibacterial defense in prawn. Accumulation of high copper levels in hepatopancreas of prawn triggers MnHc-1 gene expression. These results provide the foundation for further studies in biological function and regulation of hemocyanin in crustacean.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research was supported by grants from the Special Fund for Agro-Scientific Research in the Public Interest (nos. 201003020 and 201203065), Shanghai Committee of Science and Technology, China (10JC1404100, 09ZR1409800), National "Twelfth Five-Year" Plan for Science & Technology Support (2012BAD25B03), the National Basic Research Program (973 Program, no. 2009CB118702), National Natural Science Foundation of China (nos. 31172422, 31001098, and 31402308), and Shanghai Technology System for Chinese

Mitten-Handed Crab Industry and partly by the E-Institute of Shanghai Municipal Education Commission (no. E03009) and Zhejiang Provincial Natural Science Foundation of China (nos. LQ14C190004 and LY16C190006).

References

[1] S. Iwanaga and L. L. Bok, "Recent advances in the innate immunity of invertebrate animals," *Journal of Biochemistry and Molecular Biology*, vol. 38, no. 2, pp. 128–150, 2005.

[2] H. Decker and E. Jaenicke, "Recent findings on phenoloxidase activity and antimicrobial activity of hemocyanins," *Developmental and Comparative Immunology*, vol. 28, no. 7-8, pp. 673–687, 2004.

[3] H. Decker, M. Ryan, E. Jaenicke, and N. Terwilliger, "SDS-induced Phenoloxidase Activity of Hemocyanins from *Limulus polyphemus*, *Eurypelma californicum*, and *Cancer magister*," *Journal of Biological Chemistry*, vol. 276, no. 21, pp. 17796–17799, 2001.

[4] K. Y. Lei, F. Li, M. C. Zhang, H. Yang, T. Luo, and X. Xu, "Difference between hemocyanin subunits from shrimp *Penaeus japonicus* in anti-WSSV defense," *Developmental and Comparative Immunology*, vol. 32, no. 7, pp. 808–813, 2008.

[5] X. B. Zhang, C. H. Huang, and Q. W. Qin, "Antiviral properties of hemocyanin isolated from shrimp *Penaeus monodon*," *Antiviral Research*, vol. 61, no. 2, pp. 93–99, 2004.

[6] Y. Zhang, S. Wang, A. Xu, J. Chen, B. Lin, and X. Peng, "Affinity proteomic approach for identification of an IgA-like protein in *Litopenaeus vannamei* and study on its agglutination characterization," *Journal of Proteome Research*, vol. 5, no. 4, pp. 815–821, 2006.

[7] Y. L. Zhang, F. Yan, Z. Hu et al., "Hemocyanin from shrimp *Litopenaeus vannamei* shows hemolytic activity," *Fish and Shellfish Immunology*, vol. 27, no. 2, pp. 330–335, 2009.

[8] N. Jiang, N. S. Tan, B. Ho, and J. L. Ding, "Respiratory protein-generated reactive oxygen species as an antimicrobial strategy," *Nature Immunology*, vol. 8, no. 10, pp. 1114–1122, 2007.

[9] J. Markl, "Evolution and function of structurally diverse subunits in the respiratory protein hemocyanin from arthropods," *Biological Bulletin*, vol. 171, no. 1, pp. 90–115, 1986.

[10] R. J. Paul and R. Pirow, "The physiological significance of respiratory proteins in invertebrates," *Zoology*, vol. 100, no. 4, pp. 298–306, 1998.

[11] E. Jaenicke, R. Föll, and H. Decker, "Spider hemocyanin binds ecdysone and 20-OH-ecdysone," *Journal of Biological Chemistry*, vol. 274, no. 48, pp. 34267–34271, 1999.

[12] K. Adachi, K. Wakamatsu, S. Ito et al., "An oxygen transporter hemocyanin can act on the late pathway of melanin synthesis," *Pigment Cell Research*, vol. 18, no. 3, pp. 214–219, 2005.

[13] F. Yan, Y. L. Zhang, R. P. Jiang et al., "Identification and agglutination properties of hemocyanin from the mud crab (*Scylla serrata*)," *Fish & Shellfish Immunology*, vol. 30, no. 1, pp. 354–360, 2011.

[14] D. Sellos, S. Lemoine, and A. Van Wormhoudt, "Molecular cloning of hemocyanin cDNA from *Penaeus vannamei* (Crustacea, Decapoda): Structure, evolution and physiological aspects," *FEBS Letters*, vol. 407, no. 2, pp. 153–158, 1997.

[15] S. M. Sun, L. Q. Chen, J. G. Qin et al., "Molecular cloning, characterization and mRNA expression of copper-binding protein hemocyanin subunit in Chinese mitten crab, *Eriocheir sinensis*," *Fish & Shellfish Immunology*, vol. 33, no. 5, pp. 1222–1228, 2012.

[16] D.-L. Wang, T. Sun, D. Zuo, L.-M. Wang, Q. Wang, and Y.-L. Zhao, "Cloning and tissue expression of hemocyanin gene in *Cherax quadricarinatus* during white spot syndrome virus infection," *Aquaculture*, vol. 410-411, pp. 216–224, 2013.

[17] K. Kusche and T. Burmester, "Molecular cloning and evolution of lobster hemocyanin," *Biochemical and Biophysical Research Communications*, vol. 282, no. 4, pp. 887–892, 2001.

[18] J. C. Marxen, C. Pick, M. Kwiatkowski, and T. Burmester, "Molecular characterization and evolution of haemocyanin from the two freshwater shrimps *Caridina multidentata* (Stimpson, 1860) and *Atyopsis moluccensis* (De Haan, 1849)," *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, vol. 183, no. 5, pp. 613–624, 2013.

[19] S. P. Lall, "The minerals," in *Fish Nutrition*, J. E. Halver and R. W. Hardy, Eds., pp. 259–308, Academic Press, New York, NY, USA, 2002.

[20] T. Watanabe, V. Kiron, and S. Satoh, "Trace minerals in fish nutrition," *Aquaculture*, vol. 151, no. 1–4, pp. 185–207, 1997.

[21] M.-H. Lee and S.-Y. Shiau, "Dietary copper requirement of juvenile grass shrimp, *Penaeus monodon*, and effects on non-specific immune responses," *Fish & Shellfish Immunology*, vol. 13, no. 4, pp. 259–270, 2002.

[22] S. M. Sun, J. G. Qin, N. Yu, X. Ge, H. Jiang, and L. Chen, "Effect of dietary copper on the growth performance, non-specific immunity and resistance to *Aeromonas hydrophila* of juvenile Chinese mitten crab, *Eriocheir sinensis*," *Fish & Shellfish Immunology*, vol. 34, no. 5, pp. 1195–1201, 2013.

[23] Y. Yang, S. Q. Xie, W. Lei, X. Zhu, and Y. Yang, "Effect of replacement of fish meal by meat and bone meal and poultry by-product meal in diets on the growth and immune response of *Macrobrachium nipponense*," *Fish and Shellfish Immunology*, vol. 17, no. 2, pp. 105–114, 2004.

[24] X. Y. Pan, J. Y. Shen, J. Y. Li et al., "Identification and biological characteristics of the pathogen causing *Macrobrachium nipponense* soft-shell syndrome," *Microbiology*, vol. 36, no. 10, pp. 1571–1576, 2009.

[25] J. Y. Shen, D. Qian, W. Liu et al., "Studies on the pathogens of bacterial diseases of *Macrobrachium nipponense*," *Journal of Zhejiang Ocean University (Nature Science)*, vol. 19, no. 3, pp. 222–224, 2000.

[26] W. F. Wang, X. C. Xia, F. Liu, X. Chen, H. Yang, and Q. Ning, "Cloning and characterization of the hemocyanin gene of prawn *Macrobrachium nipponense*," *Turkish Journal of Biochemistry*, vol. 37, no. 4, pp. 348–355, 2012.

[27] D. Zhao, S. Song, Q. Wang, X. Zhang, S. Hu, and L. Chen, "Discovery of immune-related genes in Chinese mitten crab (*Eriocheir sinensis*) by expressed sequence tag analysis of haemocytes," *Aquaculture*, vol. 287, no. 3-4, pp. 297–303, 2009.

[28] E. Li, N. Yu, L. Chen, C. Zeng, L. Liu, and J. G. Qin, "Dietary vitamin B_6 requirement of the pacific white shrimp, *Litopenaeus vannamei*, at low salinity," *Journal of the World Aquaculture Society*, vol. 41, no. 5, pp. 756–763, 2010.

[29] Association of Official Analytical Chemists (AOAC), *Official Methods of Analysis*, Association of Official Analytical Chemists (AOAC), Arlington, Va, USA, 13th edition, 1980.

[30] P. Wu, D. Qi, L. Chen et al., "Gene discovery from an ovary cDNA library of oriental river prawn *Macrobrachium nipponense* by ESTs annotation," *Comparative Biochemistry and Physiology—Part D: Genomics and Proteomics*, vol. 4, no. 2, pp. 111–120, 2009.

[31] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method," *Methods*, vol. 25, no. 4, pp. 402–408, 2001.

[32] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen, "SignalP 4.0: discriminating signal peptides from transmembrane regions," *Nature Methods*, vol. 8, no. 10, pp. 785–786, 2011.

[33] S. A. Lehnert and S. E. Johnson, "Expression of hemocyanin and digestive enzyme messenger RNAs in the hepatopancreas of the black tiger shrimp *Penaeus monodon*," *Comparative Biochemistry and Physiology-B Biochemistry and Molecular Biology*, vol. 133, no. 2, pp. 163–171, 2002.

[34] W. Voll and R. Voit, "Characterization of the gene encoding the hemocyanin subunit e from the tarantula *Eurypelma californicum*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 14, pp. 5312–5316, 1990.

[35] G. von Heijne, "How signal sequences maintain cleavage specificity," *Journal of Molecular Biology*, vol. 173, no. 2, pp. 243–251, 1984.

[36] J. Sun, B. J. Wang, Z. J. Sun et al., "cDNA cloning and sequence analysis of hemocyanin in *Fenneropenaeus chinensis*," *Progress in Fisheries Science*, vol. 31, no. 1, pp. 80–89, 2010.

[37] J. Rainer and M. Brouwer, "Hemocyanin synthesis in the blue crab *Callinectes sapidus*," *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, vol. 104, no. 1, pp. 69–73, 1993.

[38] Y. L. Zhang, Z. J. Lin, Z. J. Li et al., "Identification of the main proteins binding with pathogen directly in *Litopenaeus vannamei* serum," *Journal of Fisheries of China (China)*, vol. 32, no. 1, pp. 105–110, 2008.

[39] X. Zhao, L. Guo, Y. Zhang et al., "SNPs of hemocyanin C-terminal fragment in shrimp *Litopenaeus vannamei*," *FEBS Letters*, vol. 586, no. 4, pp. 403–410, 2012.

[40] M. Longshaw, "Diseases of crayfish: a review," *Journal of Invertebrate Pathology*, vol. 106, no. 1, pp. 54–70, 2011.

[41] G. Vivekanandhan, A. A. M. Hatha, and P. Lakshmanaperumalsamy, "Prevalence of *Aeromonas hydrophila* in fish and prawns from the seafood market of Coimbatore, South India," *Food Microbiology*, vol. 22, no. 1, pp. 133–137, 2005.

[42] J. P. Truchot, "Respiratory function of arthropod hemocyanins," in *Blood and Tissue Oxygen Carriers*, C. P. Mangum, Ed., vol. 13 of *Advances in Comparative and Environmental Physiology*, pp. 377–410, Springer, Berlin, Germany, 1st edition, 1992.

[43] A. Bellelli, B. Giardina, M. Corda et al., "Sexual and seasonal variability of lobster hemocyanin," *Comparative Biochemistry and Physiology-Part A: Physiology*, vol. 91, no. 3, pp. 445–449, 1988.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 7281585, 10 pages
http://dx.doi.org/10.1155/2016/7281585

Research Article

Transcriptome Analyses Reveal Lipid Metabolic Process in Liver Related to the Difference of Carcass Fat Content in Rainbow Trout (*Oncorhynchus mykiss*)

Guo Hu, Wei Gu, Peng Sun, Qingli Bai, and Bingqian Wang

National and Local Joint Engineering Laboratory for Freshwater Fish Breeding, Heilongjiang River Fishery Research Institute, Chinese Academy of Fishery Sciences, Harbin 150070, China

Correspondence should be addressed to Bingqian Wang; bingqianwang@yeah.net

Received 2 March 2016; Revised 11 June 2016; Accepted 13 July 2016

Academic Editor: Mihai Miclăuș

Excessive accumulation of carcass fat in farm animals, including fish, has a significant impact on meat quality and on the cost of feeding. Similar to farmed animals and humans, the liver can be considered one of the most important organs involved in lipid metabolism in rainbow trout (*Oncorhynchus mykiss*). RNA-seq based whole transcriptome sequencing was performed to liver tissue of rainbow trout with high and low carcass fat content in this study. In total 1,694 differentially expressed transcripts were identified, including many genes involved in lipid metabolism, such as *L-FABP*, *adiponectin*, *PPAR-α, PPAR-β,* and *IGFBP1a*. Evidence presented in this study indicated that lipid metabolic process in liver may be related to the difference of carcass fat content. The relevance of *PPAR-α* and *PPAR-β* as molecular markers for fat storage in liver should be worthy of further investigation.

1. Introduction

Excessive fat accumulation in farm animals, including fish, has a significant impact on meat quality and the cost of feeding [1, 2]. The management of fat deposition in farm animals has drawn increasing attention from a growing number of researchers and the public. As we know, body fatness is a heritable, quantitative trait in rainbow trout (*Oncorhynchus mykiss*), and genetic selection has been successfully used to change the muscle fat content in rainbow trout [3, 4]. A good understanding of the physiological mechanisms of fat deposition could be essential to genetic enhancement of the production performance of rainbow trout.

Rainbow trout originated in North America and was introduced to China from the Democratic People's Republic of Korea (North Korea) in 1959 [5, 6]. According to statistical data provided by the Food and Agriculture Organization (FAO), rainbow trout is the major cold-water fish farmed in China, and the production amount was approximately 23,000 tons in 2013 [7]. Additionally, in many research areas, such as immunology and environmental carcinogenesis, this organism has been used as a model animal [8]. The draft genome has been sequenced and several genetic materials and genomics tools have been developed for research [9], facilitating our understanding of the whole genomic physiological mechanisms of the fatness trait in rainbow trout.

Similar to farm animals and humans, lipogenesis, triglyceride accumulation, and metabolic homeostasis occur essentially in the liver of rainbow trout [10]. The liver can be considered one of the most important organs involved in lipid metabolism, and the different expression of functional genes involved in lipid synthesis, degradation, transportation, and storage has been analyzed in the livers of trout strains divergently selected for muscle fat content [11–13]. The research showed that genes involved in lipid metabolism might be influenced by divergent selection and lead to a different model of fat storage in the genetically selected lean and fat trout strains [13]. Another later study indicated that lipid and glucose metabolism was regulated by insulin in two experimental rainbow trout lines [14]. In addition, some evidence has suggested that TOR signaling pathway-associated lipogenesis could be overactivated and the utilization of

glucose in the liver might be also improved by the genetic selection [11, 15, 16].

In the present study, RNA-seq based transcriptome sequencing was performed to liver tissue of rainbow trout with extremely high and low carcass fat content. The objectives were to identify the differential expressed transcripts and get meaningful information about the molecular mechanism of fat deposition. The results may be helpful in determining more powerful target genes associated with carcass fat content and would potentially be applied in the development of molecular genetic markers for fatness traits in rainbow trout.

2. Materials and Methods

2.1. Experimental Animals and Sampling. In this study, a freshwater strain of rainbow trout, selected for large-size body weight and measurement traits using family based BLUP method, was used as experimental material [17]. The breeding program was performed at the Bohai cold-water fish experimental station, located near Jingbo Lake (latitude 44.02°N, longitude 128.74°E), in Heilongjiang Province, China. Water was supplied to the tanks and ponds at the Bohai experimental station using natural flowing spring water (5.2–18.0°C) with a water flow of 20–30 L/sec and a dissolved oxygen concentration ranging from 7 to 11 mg/L. Seventy-five full-sibling families were established using 75 male and 75 female fish for each generation based on passive integrated transponder (PIT) tag. After each family was separately hatched from eggs, the fish absorbed their yolk sacs (30 to 35 days) before swimming up and beginning to feed on a commercial trout fry diet. Each candidate family was then reared separately in a tank (1-meter diameter, 0.5-meter water depth, average water temperature ranging from approximately 6.5°C to 12°C, and water flow at 6 to 8 L/min) for 10 months until reaching a size of >50 g. Fifty fish with no deformities were randomly selected from each tank with PIT tags and were then deposited together in cemented pools (5-meter width, 30-meter length, and 0.7-meter water depth) for one year. The fish beyond two years of age were cultured in earth ponds (15-meter width, 120-meter length, and 0.8-meter water depth). It takes approximately 3–3.5 years to reach sexual maturity for rainbow trout at the Bohai station. All of the fish were fed according to the feeding program for trout from BioMar. All of the animal work was conducted according to the guidelines for the care and use of experimental animals established by the Ministry of Science and Technology of China (document number: 2006-398) and was approved by the Academic Ethics Committee of Heilongjiang River Fishery Research Institute.

Before the traits were measured, the fish were anaesthetized using 0.5 mg/L 2-phenoxyethanol ($C_8H_{10}O_2$) to avoid injury from handling. Each fish was measured for carcass fat content using a Fish Fatmeter (Distell Company, Old Levenseat, Scotland, UK) (Model FFM-692; calibration: TROUT-2; representing fat content of whole carcass including belly cavity and fish roe) [18], for body weight using electronic scales (0.1 g), and for total length using Vernier calipers (0.1 cm); each fish was sexed using secondary sexual characteristics, and at the same time, body measurement traits were assessed. All of phenotypic records of carcass fat content related to 1,422 individuals at 2.5 years of age were used for the calculation of the population mean (7.14 ± 1.52%). Six female fish were used for transcriptome sequencing. To ensure that every individual within and between the high and low carcass fat content groups used for RNA sequencing had the same genetic background but no direct kinship, a coancestry analysis was performed based on PIT tag ID of the individuals and the pedigree records. Three female rainbow trout with low carcass fat content (4.20 ± 0.31%) and the same number of high carcass fat content female ones (11.27 ± 0.31%) were randomly selected from the population for transcriptome sequencing. There were extremely significant differences (approximately 3-fold) in carcass fat content ($p < 0.001$) but no significant difference in body weight between the two groups ($p > 0.05$).

2.2. Total RNA Extraction, RNA-Seq Library Construction, and Sequencing. The livers of three low carcass fat content female fish and three high carcass fat content female fish were sampled under RNAse-free conditions to perform RNA extraction for sequencing. All of the tissue samples were frozen in liquid nitrogen and stored at −80°C until analysis. Total RNA was isolated using Trizol reagent (Invitrogen, Carlsbad, CA, USA), and RNA degradation and contamination were monitored on 1% agarose gels. Then, the clustering and sequencing were performed by the Experimental Department of Novogene Ltd. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using a TruSeq SR Cluster Kit v3-cBot-HS (Illumina), according to the manufacturers' instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq 2500 platform, and 125 bp paired-end reads were generated.

2.3. RNA-Seq Data Processing, Annotation, and Differential Expression Identification. Raw counts of RNA-seq reads for each transcript and in each sample were derived and normalized to fragments per kilobase of exon per million fragments mapped reads (FPKM). All transcriptome data from six individuals were aligned to the *Oncorhynchus mykiss* mRNA dataset (*Oncorhynchus mykiss* genome project accession via http://www.genoscope.cns.fr/trout/) using the TopHat [19], and the expression of trout transcripts was evaluated using the software Cufflinks [20]; TopHat and Cufflinks were performed with default parameters. Differentially expressed transcripts (fold changes ≥2 or fold changes ≤0.5 and adjusted p value ≤0.01) between high and low carcass fat content fish were identified with the Cufflink package based on FPKM; visualization of transcripts expression was performed using R software [21]. GO terms were assigned to the transcripts according to their corresponding homologs in the trout mRNA database (http://www.genoscope.cns.fr/trout/). The GO annotation result was visualized in WEGO [22]. Lastly, the GO terms enrichment analysis for the differentially expressed transcripts was carried out by topGO package [23].

Table 1: The information of the primer pairs used to analyze gene expression by real-time quantitative RT-PCR.

Standard nomenclature*	Primer	Product size (bp)
GSONMT00021034001	5′ CTGGGCACCACCTACTCAT 3′ 5′ GTGTCTGTGAAGGCAACG 3′	110
GSONMT00026574001	5′ AAACTGCTGGTGAAGACG 3′ 5′ TGTTGAGGACAAAGAGGG 3′	136
GSONMT00028411001	5′ CCCTGGGCGTACAGTTTGAC 3′ 5′ CTTGGCATCCACTCCATCGT 3′	188
GSONMT00029574001	5′ TTCAAGGGTCGTGGGAGA 3′ 5′ CGGAGGGTGTTGGAAGTG 3′	159
GSONMT00032762001	5′ AGCAGAACGGCAATGACT 3′ 5′ TCCTGGACGCTGGAGAAT 3′	181
GSONMT00051137001	5′ GGAGCCTGGTTGTGGATG 3′ 5′ GCTGTGGCCGTGGAGATA 3′	286
GSONMT00051158001	5′ CCTGGCTCTGTTTGTGGC 3′ 5′ TGTCGTCTGGGTGGTTGG 3′	137
GSONMT00054447001	5′ CTGGAGTAAAGATGGGTGAC 3′ 5′ GTCCTGTTCTGGGATTGG 3′	169
GSONMT00060652001	5′ CCGACCACCAACCCTAAT 3′ 5′ CACTGGCAGCGGTAGAAC 3′	272
GSONMT00062349001	5′ GCCTGGATGAGAATGA 3′ 5′ GATACCGCAGGACAAT 3′	335
GSONMT00064682001	5′ CCTACTGAGCCCATTCCT 3′ 5′ ATGGAGACTAAGCGAGGC 3′	360
GSONMT00070501001	5′ CGCCCACCTCTAAACAAGCC 3′ 5′ GCAGCGTCATCCAGCCCATC 3′	459
GSONMT00077785001	5′ CGTGCCCATCCGTTTCAATA 3′ 5′ CCCGAGCATCTTTGGTGTAG 3′	137
GSONMT00082167001	5′ ACCGTGGGAGTAGTTCTTGC 3′ 5′ TAGACACCGTTGTAGACCAG 3′	425
EF1α	5′ TCCTCTTGGTCGTTTCGCTG 3′ 5′ ACCCGAGGGACATCCTGTG 3′	159

*Note: the standard nomenclature for rainbow trout according to GBrowse-2.54 of rainbow trout genome project database.

2.4. Validation of Differentially Expressed Genes by Real-Time Quantitative RT-PCR. Fourteen differentially expressed genes identified by the transcriptome sequencing were validated by real-time quantitative RT-PCR, using the same fish sample, and EF1α was used as a reference control [11]. Real-time RT-PCR was performed using SYBR (R) GREEN I NUCLEIC A (Life Technologies) on the LightCycler 480II Real-Time System (Roche, Switzerland). The reaction was performed using the following conditions: denaturation at 95°C for 3 min, followed by 40 cycles of amplification (95°C for 30 s, 60°C for 30 s, and 60°C for 45 s). Relative expression was calculated using the delta-delta-Ct method; primer sequences can be found in Table 1.

3. Results

3.1. Sequencing of Liver Transcriptome in Rainbow Trout. The liver transcriptome expression of six fish at 2.5 years was analyzed by RNA sequencing. After removing low-quality reads, adaptor, and barcode sequences, nearly 35 million clean reads for each individual were obtained; the descriptive statistics of the RNA-seq reads for each individual are shown in Table 2. All of the raw data were submitted to the NCBI database (accession numbers SRX1067664 and SRR2072562).

3.2. Gene Functional Annotation for the Differentially Expressed Transcripts. A total of 1,694 differentially expressed transcripts were identified from liver tissue between the high and low carcass fat content rainbow trout (Figure 1). The standard nomenclature, expression, and gene annotation for the differentially expressed transcripts were shown in Table S1 in Supplementary Material available online at http://dx.doi.org/10.1155/2016/7281585. Among these transcripts, 912 transcripts were highly expressed in high carcass fat content fish, including many key factors of lipogenesis and fat deposition in rainbow trout, such as peroxisome proliferator activated receptor alpha (*PPAR-α*), *PPAR-β*, insulin-like growth factor binding protein 1a (*IGFBP1a*),

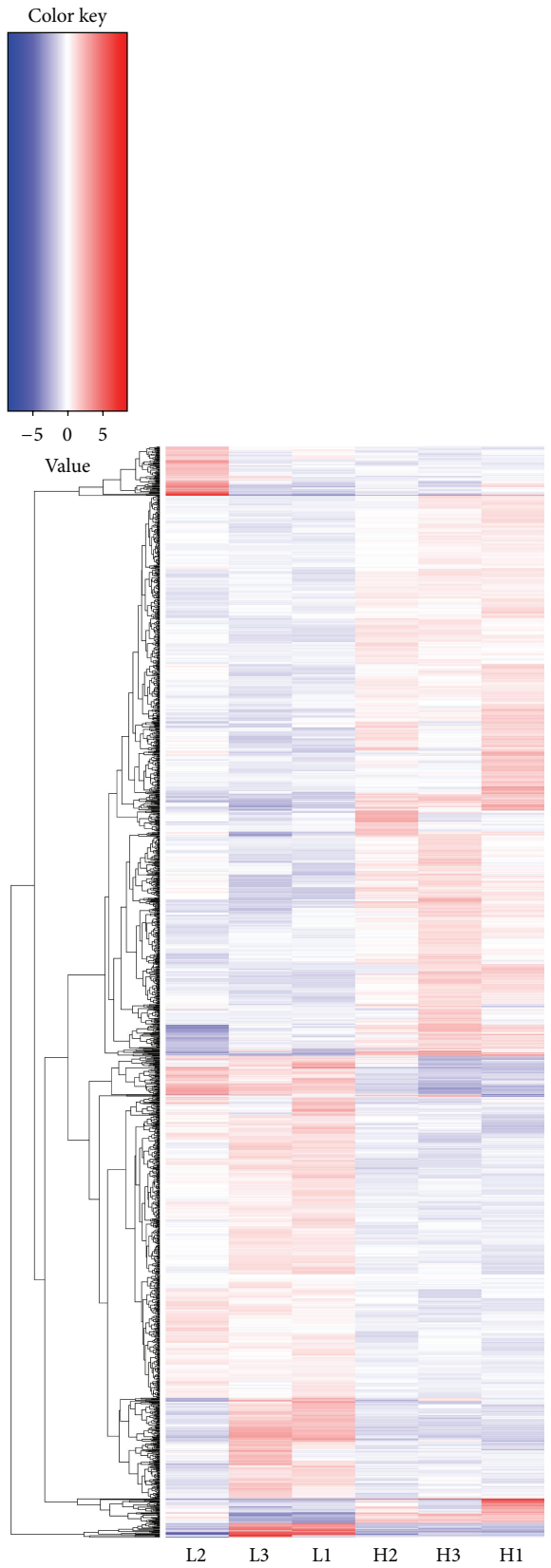


Figure 1: Cluster image of 1,694 significant differentially expressed transcripts between the high and low carcass fat content rainbow trout. A total of 1,694 differentially expressed transcripts were identified with Cufflink package based on FPKM: 912 were highly expressed in rainbow trout with high carcass fat content, and 782 showed higher expression level in those fish with low carcass fat content. H1, H2, and H3 represent the first, second, and third fish in the high carcass fat content group, respectively; and L1, L2, and L3 represent the first, second, and third fish in the low carcass fat content group, respectively. Colored bars indicate relative expression levels. Transcripts expressed at higher levels were assigned red, while transcripts expressed at low levels were assigned blue.

Table 2: The descriptive statistics of the raw RNA-seq datasets for all of the six individuals sequenced in the present study.

Sample	Raw reads	Clean reads	Clean bases	Error (%)	Q20 (%)	Q30 (%)	GC (%)
H1_1	17429887	16907974	2.11G	0.03	96.27	92.45	50.25
H1_2	17429887	16907974	2.11G	0.04	93.31	87.75	50.2
H2_1	18051923	17574994	2.2G	0.03	96.18	92.3	50.37
H2_2	18051923	17574994	2.2G	0.04	93.56	88.15	50.33
H3_1	16777106	16241788	2.03G	0.03	96.16	92.23	50.07
H3_2	16777106	16241788	2.03G	0.04	93.33	87.71	50.04
L1_1	18535584	17957775	2.24G	0.03	96.41	92.72	49.7
L1_2	18535584	17957775	2.24G	0.04	93.97	88.84	49.66
L2_1	18083733	17537741	2.19G	0.03	96.26	92.43	49.72
L2_2	18083733	17537741	2.19G	0.04	94.03	88.89	49.68
L3_1	19420929	18772655	2.35G	0.03	96.24	92.41	50.08
L3_2	19420929	18772655	2.35G	0.04	93.66	88.3	50.05

Notes: sample name: H1 is the first fish in high carcass fat content group and so on; L1 is the first fish in high carcass fat content group and so on; H1_1 is the left-end reads, H1_2 is the right-end reads; the descriptive statistics for the individual H1 is the sum of H1_1 and H1_2 and so on.

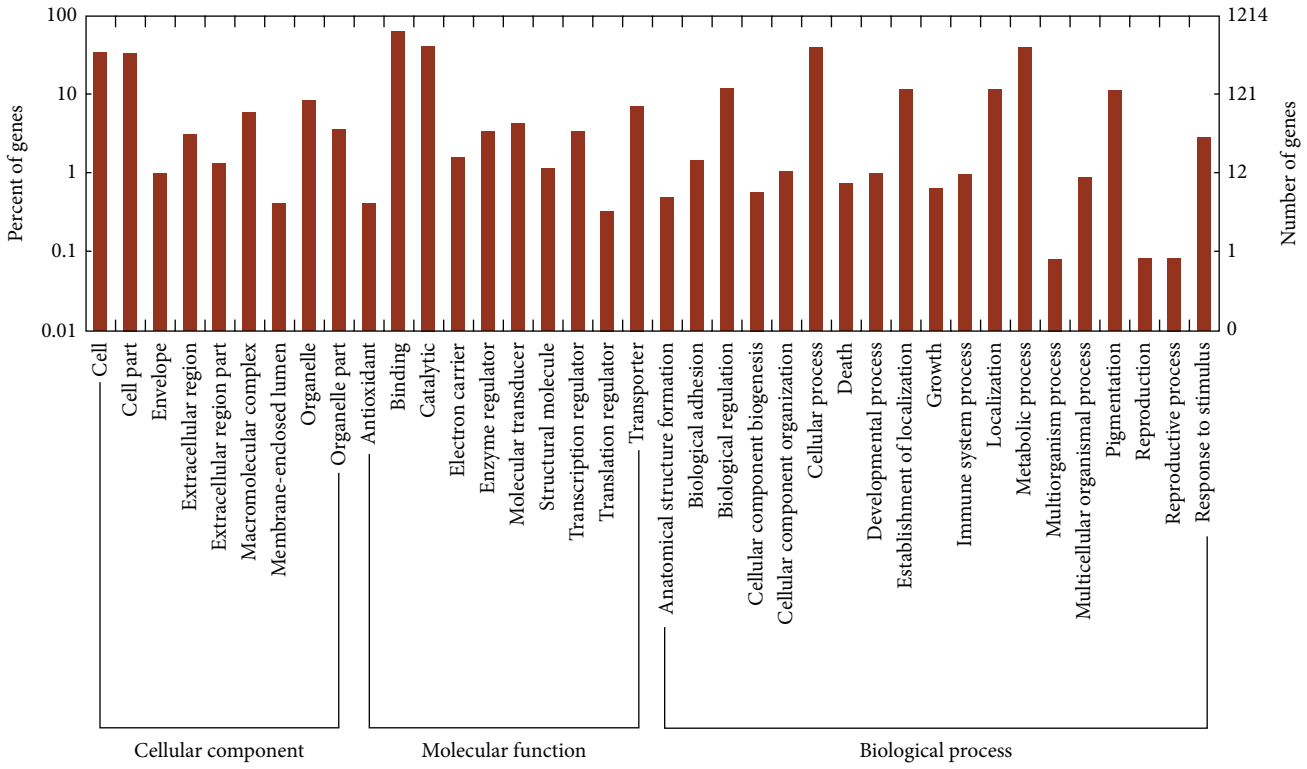


Figure 2: GO annotation results of the differently expressed transcripts between the high and low carcass fat content rainbow trout.

and fatty acid-binding protein 10-A, liver (*L-FABP*). At the same time, 782 transcripts showed higher expression levels in low carcass fat content fish, including many transcription factors with important effects on growth, differentiation, lipid metabolism, carbohydrate metabolism, and immunity, for example, growth hormone receptor isoform 1 (*GHR1*), insulin-induced gene 1 (*INSIG1*), and heat shock 70 kDa protein (*HSP70*).

GO annotations provided in the trout mRNA database (*Oncorhynchus mykiss* genome project accession via http://www.genoscope.cns.fr/trout/) were used in this study. At least one in three main categories of GO terms (biological process, molecular function, and cellular component) was assigned to all 1,694 differentially expressed transcripts and then was classified into different functional categories, as shown in Figure 2. The results showed that metabolic process (GO: 0008152) was the most abundant group in the biological process category, followed by cellular process (GO: 0009987), while biological regulation (GO: 0065007) and establishment of localization (GO: 0051234) were also common, consistent

with the central role of liver in the basal and intermediary metabolism. In the molecular function category, the most abundant group was binding (GO: 0005488), and the second most abundant one was catalytic activity (GO: 0003824). There were also transcripts classified into specific groups, such as transporter activity (GO: 0005215), molecular transducer activity (GO: 0060089), enzyme regulator activity (GO: 0030234), and transcription regulator activity (GO: 0030528), indicating that many genes were involved in important physiological functions of synthesis, transport, and catabolism in the livers of rainbow trout. In the cellular component category, the most abundant groups were related to cell (GO: 0005623), cell part (GO: 0044464), organelle (GO: 0043226), organelle part (GO: 0044422), and macromolecular complex (GO: 0032991).

3.3. GO Enrichment Analysis for Differentially Expressed Transcripts. The results of GO enrichment analysis for 1,694 differentially expressed transcripts were shown in Table S2 in the Supplementary Material. GO terms in the biological process category were highly enriched, with 84 GO terms attaining significant levels, including GO: 0044710 (single-organism metabolic process), GO: 0055114 (oxidation-reduction process), GO: 0006520 (cellular amino acid metabolic process), GO: 0006629 (lipid metabolic process), GO: 0006006 (glucose metabolic process), GO: 0019318 (hexose metabolic process), GO: 0005996 (monosaccharide metabolic process), and GO: 0006541 (glutamine metabolic process), consistent with the physiological functions of the livers of rainbow trout. At the same time, the GO terms in the molecular function category were also highly enriched, and 88 GO terms reached the significant level, including GO: 0016491 (oxidoreductase activity), GO: 0016298 (lipase activity), GO: 0048037 (cofactor binding), GO: 0050662 (coenzyme binding), GO: 0005319 (lipid transporter activity), and GO: 0004465 (lipoprotein lipase activity), indicating that the lipid metabolism mode might be different between high and low carcass fat content fish.

Moreover, we analyzed the expression of transcripts grouped as "lipid metabolic process" GO term (0006629) based on the transcriptome data, revealing that 52 of 372 transcripts (14%) were differentially expressed between the high and low carcass fat content fish groups (Figure 3), providing meaningful information about the difference in lipid metabolism between the two groups. The standard nomenclature, expression, and gene annotation for the 52 transcripts were shown in Table S3.

3.4. Validation of Differentially Expressed Genes by Real-Time Quantitative RT-PCR. To validate the results of differentially expressed genes in transcriptome sequencing, we performed real-time quantitative RT-PCR for 14 differentially expressed genes involved in lipid metabolism, growth regulation, and other important functions. Among these genes, *L-FABP*, putative *I-FABP*, *IGFBP1a*, insulin-like growth factor binding protein, acid labile subunit (*IGFBP-ALS*), *adiponectin, cytochrome P450, PPAR-α, PPAR-β*, and lipopolysaccharide-binding protein (*LBP*) were highly expressed in the high carcass fat content fish group, while heat shock 70 kDa protein (*HSP70*), acetyl-CoA acetyltransferase, cytosolic (*ACAT2*), StAR-related lipid transfer protein (*SATRT*), *GHR1*, and *INSIG1* showed lower expression level in the high carcass fat content fish group. The results of differentially expressed genes identified by real-time quantitative RT-PCR had good consistency with the transcriptome data. For transcriptome data, the evaluation criteria for differentially expressed genes between the high and low carcass fat content fish groups were fold changes $\geq$ 2 or fold changes $\leq$ 0.5 and adjusted p value $\leq$ 0.01, while for real-time quantitative RT-PCR we used Student's t-test ($p < 0.05$). All fourteen genes detected by real-time RT-PCR reached significance. The fold change results of RNA sequencing and real-time RT-PCR for 14 differentially expressed genes were compared in Table 3.

4. Discussion

As is well known, the Chinese believe that the best tasting and most nutritious aquatic animals must be alive immediately prior to cooking, and they eat whole fish from beginning to end (with head and tail); thus, the fat content of the whole carcass of the fish body is an economically important trait in China. Fatty acid de novo synthesis in rainbow trout occurs mainly in the liver [10, 24], and several studies have explored the correlations between gene expression in liver tissues and lipid metabolism [25–27].

4.1. Differentially Expressed Functional Genes Involved in Lipid Metabolism and Energy Balance. Among the 1,694 differentially expressed genes, *L-FABP* was detected to be more than 30-fold higher expressed in the fat group than in the lean group by both transcriptome sequencing and real-time RT-PCR methods. Liver type fatty acid-binding protein is a functional protein with a small molecular weight that belongs to a superfamily of lipid-binding proteins and participates in intracellular fatty acid transportation in human, farm animals, birds, and fish [28–30]. Higher expression of the *L-FABP* gene in the livers of the high carcass fat content group than in the low carcass fat content group might be the result of increased fatty acid transportation into white muscle, skin, and other parts, leading to more fat deposition in the whole carcass. The results indicated that the expression of the *L-FABP* gene should have a positive effect on fat deposition in the rainbow trout.

The PPAR gene superfamily belongs to the nuclear receptor superfamily, including three different genes—*PPAR-α, PPAR-β,* and *PPAR-γ*. *PPAR*s are a type of ligand-activated transcription factors that control gene expression by binding to specific response elements (PPREs) within promoters [31]. *PPAR*s play critical physiological roles as lipid sensors and regulators of lipid metabolism and can transactivate multiple target genes and interact with other transcription factors involved in lipid metabolic pathways [32, 33]. In the current study, *PPAR-α* and *PPAR-β* were expressed significantly more in the high carcass fat content fish group than in the low carcass fat content fish group by transcriptome sequencing and as validated by real-time RT-PCR. Our results regarding

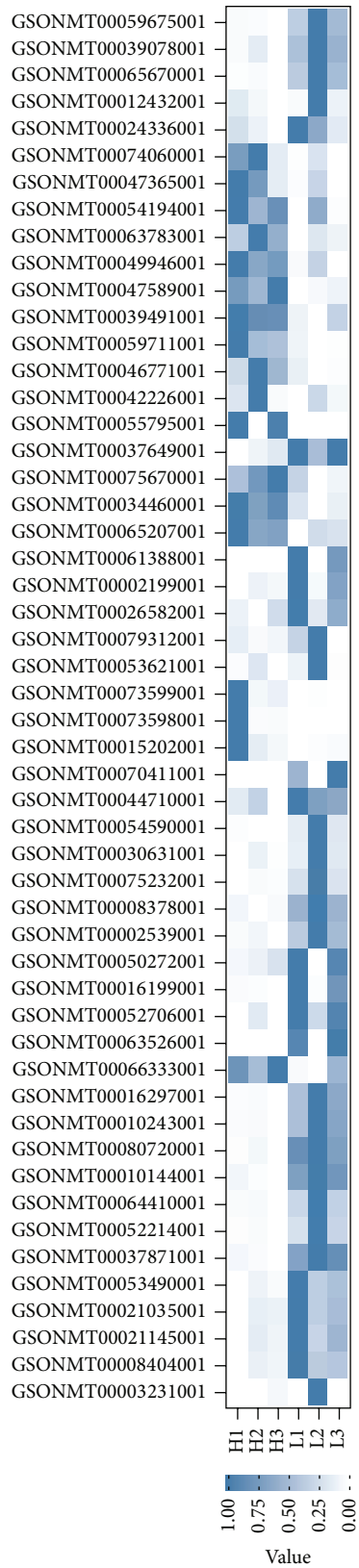


Figure 3: Heat map showing the expression profiles of differentially significant transcripts involved in lipid metabolic processes. A total of 52 of 372 transcripts involved in lipid metabolic processes (GO: 0006629) were differentially expressed between the high and low carcass fat content fish groups. H1, H2, and H3 represent the first, second, and third fish in the high carcass fat content group, respectively; L1, L2, and L3 represent the first, second, and third fish in the low carcass fat content group, respectively. Colored bars indicate relative expression levels. Transcripts expressed at higher levels were assigned blue, while transcripts expressed at low levels were assigned white.

Table 3: Comparison of fold differences in 14 differentially expressed genes detected by transcriptome sequencing and real-time RT-PCR analyses.

Standard nomenclature*	Fold change by real-time RT-PCR	Fold change by RNA-seq	Gene annotation
GSONMT00021034001	0.25 ± 0.03	0.18	Heat shock 70 kDa protein
GSONMT00026574001	1.40 ± 0.12	2.70	Lipopolysaccharide-binding protein/bactericidal permeability-increasing protein
GSONMT00028411001	4.84 ± 0.35	4.44	Fatty acid-binding protein, intestinal putative mRNA
GSONMT00029574001	36.83 ± 2.30	17.70	Cytochrome P450
GSONMT00032762001	33.50 ± 1.95	35.73	Fatty acid-binding protein 10-A, liver
GSONMT00051137001	6.05 ± 0.50	4.27	Insulin-like growth factor binding protein 1a
GSONMT00051158001	1.50 ± 0.15	3.03	Peroxisome proliferator activated receptor alpha
GSONMT00054447001	0.18 ± 0.04	0.18	Acetyl-CoA acetyltransferase, cytosolic
GSONMT00060652001	2.22 ± 0.15	2.66	Peroxisome proliferator activated receptor beta
GSONMT00062349001	0.13 ± 0.02	0.19	StAR-related lipid transfer protein
GSONMT00064682001	2.80 ± 0.15	2.76	Insulin-like growth factor binding protein, acid labile subunit
GSONMT00070501001	0.26 ± 0.03	0.11	Growth hormone receptor isoform 1
GSONMT00077785001	3.22 ± 0.19	17.14	Adiponectin
GSONMT00082167001	0.71 ± 0.05	0.29	Insulin-induced gene 1

*Note: the standard nomenclature for rainbow trout according to GBrowse-2.54 of rainbow trout genome project database.

PPAR-α expression were different from Kolditz et al.'s results on the divergent selection of lean and fat fish for muscle fat content by cDNA microarray in rainbow trout [13]. In their work, *PPAR-α* expression was lower in the fat line than in the lean line. The differences between our results and the result of Kolditz et al. might have been caused in part by different ages of the fish and varying aquaculture environmental conditions. Other potential reasons for the difference in PPAR-*α* expression between the high and low fat content fish in the two populations included the different histories of the two populations. In addition, in contrast to *PPAR-α* and *PPAR-β*, the expression of *PPAR-γ* in the liver was very low and was not differentially expressed between the fat and lean group, consistent with the expression level of *PPAR-γ* gene in liver tissue being only 10% to 30% of the level in adipose tissue in mammals [34, 35] and similar to that in the chicken [36]. In the current study, the gene expression of *PPAR-α* and *PPAR-β* was highly correlated with fat deposition, suggesting that they were also key factors of fat accumulation in the rainbow trout carcass.

Cytochrome P-450 is transcriptionally modulated by glucocorticoid binding to glucocorticoid response elements in the promoter region, playing important roles in the oxidative metabolism of endogenous and exogenous compounds [37]. *Cytochrome P-450* was detected to be more than 30-fold higher expressed in the fat group than in the lean group by transcriptome sequencing, and it was validated to be nearly 20-fold higher by real-time RT-PCR. This result was in agreement with the currently known oxidation-reduction process of gene expression in the liver. The insulin regulated lipid metabolism of fish does not seem to follow mammalian patterns; the insulin-like growth factors (IGFs) are polypeptides with high sequence similarity to insulin, and they are able to modulate lipolysis and lipogenesis in fish hepatocytes [14–16, 38]. Insulin-like growth factor binding proteins (IGFBPs) regulate the biological functions of IGFs, and IGFs can stimulate the growth of multiple tissue cell types [39]. In the current study, the gene expression of two *IGFBP* isoforms and gene expression of *INSIG1* were differently expressed between the lean and fat fish groups, suggesting potential roles of insulin and insulin-like growth factors in lipid storage or utilization of fat accumulation in rainbow trout.

In total, many functional genes involved in lipid and carbohydrate metabolism might contribute to phenotypic variations in carcass fat content, and they were found to be expressed in both high and low carcass fat content fish groups. However, some of these genes were expressed significantly different between these two groups. On the one hand, this phenomenon could partly occur because of potential positive genetic correlation between growth traits and carcass fat content and the growth traits being under strong artificial selection in this experimental population. On the other hand, there was a large difference in the group mean of carcass fat content between the lean and fat groups, suggesting that the dynamics of genetics effects and the gene expression mode might be influenced by genetic heterogeneity between the lean and fat groups.

4.2. Metabolic Process Related to the Difference of Carcass Fat Content Based on GO Enrichment Analysis. The GO approach could help to predict the functions of genes based on the existing architecture and prior knowledge of molecular biological mechanisms for nonmodel organisms. In the present study, a large number of functional genes were significantly

enriched in lipid and carbohydrate metabolism related GO terms, such as GO: 0006629 (lipid metabolic process), GO: 0006006 (glucose metabolic process), GO: 0019318 (hexose metabolic process), and GO: 0005996 (monosaccharide metabolic process), which was consistent with liver being an important tissue in the biological process of lipid and carbohydrate metabolism and metabolic energy balance. In detail, we identified 52 differentially expressed transcripts significantly enriched in the lipid metabolic process (GO: 0006629) and 20 differentially expressed transcripts significantly enriched in the glucose metabolic process (GO: 0006006). In farm animals, fatty acids and glucose are the two main metabolic substrates oxidized by animals for energy production. The mobilization of fat stores could be decreased by glucose, using both inhibiting lipolysis and stimulating primary reesterification approaches. Our results were consistent with this information and also suggested that the expression and regulation of lipid and carbohydrate metabolism genes should play key roles in the ontogenesis of carcass fat content in rainbow trout.

4.3. Consistency between Transcriptome Sequencing and Real-Time RT-PCR. As is known, transcriptome sequencing and real-time RT-PCR analysis are both powerful tools for obtaining a view of gene expression. However, quantifying gene expression by sequencing and real-time RT-PCR is not necessarily consistent. Although much meaningful information can be obtained from transcriptome sequencing alone, this technology still has limitations in accuracy and in fairness for RNA-seq data processing [40]. Although fold changes were different for some genes between these two methods, all fourteen of the genes detected by real-time RT-PCR attained significant levels of differential expression ($p < 0.05$), conferring robustness to our transcriptome dataset.

5. Conclusion

In this study, the transcript expression profile of the liver tissue in rainbow trout was investigated comprehensively using RNA sequencing and was confirmed by real-time RT-PCR. A total of 1,694 transcripts were differentially expressed between high and low carcass content fish group. Evidence presented in this study indicated that lipid metabolic process in liver may be related to the difference of carcass fat content. The relevance of *PPAR-α* and *PPAR-β* as molecular markers for fat storage in liver commands further investigation.

Competing Interests

The authors declare that they have no competing interests regarding the publication of this paper.

Acknowledgments

This study was supported by the Special Fund for Agro-Scientific Research in the Public Interest (201403012), National Natural Science Foundation of China (31502157), and the Central-Level Non-Profit Scientific Research Institutes Special Funds (HSY201501). The authors acknowledge Dr. Youyi Kuang and Dr. Xianhu Zheng for their technical support in the phenotypic data collection of the carcass fat content.

References

[1] J. D. Wood, M. Enser, A. V. Fisher et al., "Fat deposition, fatty acid composition and meat quality: a review," *Meat Science*, vol. 78, no. 4, pp. 343–358, 2008.

[2] D. Tobin, A. Kause, E. A. Mäntysaari et al., "Fat or lean? The quantitative genetic basis for selection strategies of muscle and body composition traits in breeding schemes of rainbow trout (Oncorhynchus mykiss)," *Aquaculture*, vol. 261, no. 2, pp. 510–521, 2006.

[3] C. Weil, A.-S. Goupil, E. Quillet, L. Labbe, and F. Le Gac, "Two-way selection for muscle lipid content modifies puberty and gametogenesis in rainbow trout," *Cybium*, vol. 32, no. 2, p. 198, 2008.

[4] E. Quillet, S. Le Guillou, J. Aubin, and B. Fauconneau, "Two-way selection for muscle lipid content in pan-size rainbow trout (*Oncorhynchus mykiss*)," *Aquaculture*, vol. 245, no. 1–4, pp. 49–61, 2005.

[5] D. Sun and B. Wang, "Aquaculture of salmonids in China," *Chinese Journal Fisheries*, vol. 23, no. 2, pp. 56–62, 2010.

[6] Z. Wang and Y. Yang, "Cold water fish culture in China," in *Cold Water Fisheries in the Trans-Himalayan Countries*, T. Petr and D. B. Swar, Eds., vol. 431, FAO, Fisheries Technical Paper, Rome, Italy, 2002.

[7] FAO, "Aquaculture production: quantities 1950–2013," FISHSTAT Plus.

[8] G. H. Thorgaard, G. S. Bailey, D. Williams et al., "Status and opportunities for genomics research with rainbow trout," *Comparative Biochemistry and Physiology—B Biochemistry and Molecular Biology*, vol. 133, no. 4, pp. 609–646, 2002.

[9] C. Berthelot, F. Brunet, D. Chalopin et al., "The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates," *Nature Communications*, vol. 5, article 3657, 2014.

[10] R. J. Henderson and J. R. Sargent, "Lipid biosynthesis in rainbow trout, *Salmo gairdnerii*, fed diets of differing lipid content," *Comparative Biochemistry and Physiology Part C: Comparative Pharmacology*, vol. 69, no. 1, pp. 31–37, 1981.

[11] S. Skiba-Cassy, M. Lansard, S. Panserat, and F. Médale, "Rainbow trout genetically selected for greater muscle fat content display increased activation of liver TOR signaling and lipogenic gene expression," *American Journal of Physiology—Regulatory Integrative and Comparative Physiology*, vol. 297, no. 5, pp. R1421–R1429, 2009.

[12] J. Jin, F. Médale, B. S. Kamalam, P. Aguirre, V. Véron, and S. Panserat, "Comparison of glucose and lipid metabolic gene expressions between fat and lean lines of rainbow trout after a glucose load," *PLoS ONE*, vol. 9, no. 8, Article ID e105548, 2014.

[13] C.-I. Kolditz, G. Paboeuf, M. Borthaire et al., "Changes induced by dietary energy intake and divergent selection for muscle fat content in rainbow trout (*Oncorhynchus mykiss*), assessed by transcriptome and proteome analysis of the liver," *BMC Genomics*, vol. 9, article 506, 2008.

[14] J. Jin, S. Panserat, B. S. Kamalam, P. Aguirre, V. Véron, and F. Médale, "Insulin regulates lipid and glucose metabolism similarly in two lines of rainbow trout divergently selected for

muscle fat content," *General and Comparative Endocrinology*, vol. 204, pp. 49–59, 2014.

[15] W. Dai, S. Panserat, J. A. Mennigen et al., "Post-prandial regulation of hepatic glucokinase and lipogenesis requires the activation of TORC1 signalling in rainbow trout (*Oncorhynchus mykiss*)," *Journal of Experimental Biology*, vol. 216, no. 23, pp. 4483–4492, 2013.

[16] M. Lansard, S. Panserat, E. Plagnes-Juan, I. Seiliez, and S. Skiba-Cassy, "Integration of insulin and amino acid signals that regulate hepatic metabolism-related gene expression in rainbow trout: role of TOR," *Amino Acids*, vol. 39, no. 3, pp. 801–810, 2010.

[17] G. Hu, W. Gu, Z. Jiang, Q. Bai, and B. Wang, "Heritability estimates and genetic correlations among body weight and measurement traits of large-size rainbow trout, *Oncorhynchus mykiss*, in China," *Journal of the World Aquaculture Society*, vol. 45, no. 6, pp. 691–698, 2014.

[18] Distell.com: Distell fish fatmeter, User and technical manual, version 2.8., 2011.

[19] C. Trapnell, L. Pachter, and S. L. Salzberg, "TopHat: Discovering splice junctions with RNA-Seq," *Bioinformatics*, vol. 25, no. 9, pp. 1105–1111, 2009.

[20] C. Trapnell, A. Roberts, L. Goff et al., "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks," *Nature Protocols*, vol. 7, no. 3, pp. 562–578, 2012.

[21] R Core Team, *R: a Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014.

[22] J. Ye, L. Fang, H. Zheng et al., "WEGO: a web tool for plotting GO annotations," *Nucleic Acids Research*, vol. 34, supplement 2, pp. W293–W297, 2006.

[23] A. Alexa, J. Rahnenführer, and T. Lengauer, "Improved scoring of functional groups from gene expression data by decorrelating GO graph structure," *Bioinformatics*, vol. 22, no. 13, pp. 1600–1607, 2006.

[24] D. J. Cowley and M. A. Sheridan, "Insulin stimulates hepatic lipogenesis in rainbow trout, *Oncorhynchus mykiss*," *Fish Physiology and Biochemistry*, vol. 11, no. 1–6, pp. 421–428, 1993.

[25] B. S. Kamalam, F. Médale, L. Larroquet, G. Corraze, and S. Panserat, "Metabolism and fatty acid profile in fat and lean rainbow trout lines fed with vegetable oil: effect of carbohydrates," *PLoS ONE*, vol. 8, no. 10, Article ID e76570, 2013.

[26] J. Sánchez-Gurmaches, L. Cruz-Garcia, J. Gutiérrez, and I. Navarro, "mRNA expression of fatty acid transporters in rainbow trout: in vivo and in vitro regulation by insulin, fasting and inflammation and infection mediators," *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, vol. 163, no. 2, pp. 177–188, 2012.

[27] J. S. Ings, M. R. Servos, and M. M. Vijayan, "Hepatic transcriptomics and protein expression in rainbow trout exposed to municipal wastewater effluent," *Environmental Science & Technology*, vol. 45, no. 6, pp. 2368–2376, 2011.

[28] R. L. Smathers and D. R. Petersen, "The human fatty acid-binding protein family: evolutionary divergences and functions," *Human Genomics*, vol. 5, no. 3, pp. 170–191, 2011.

[29] A. Chmurzyńska, "The multigene family of fatty acid-binding proteins (FABPs): function, structure and polymorphism," *Journal of Applied Genetics*, vol. 47, no. 1, pp. 39–48, 2006.

[30] R. L. Londraville, "Intracellular fatty acid-binding proteins: putting lower vertebrates in perspective," *Brazilian Journal of Medical and Biological Research*, vol. 29, no. 6, pp. 707–720, 1996.

[31] Y. Guan, Y. Zhang, and M. D. Breyer, "The role of PPARs in the transcriptional control of cellular processes," *Drug News and Perspectives*, vol. 15, no. 3, pp. 147–154, 2002.

[32] M. Ahmadian, J. M. Suh, N. Hah et al., "PPARγ signaling and metabolism: the good, the bad and the future," *Nature Medicine*, vol. 99, no. 5, pp. 557–566, 2013.

[33] J. Berger and D. E. Moller, "The mechanisms of action of PPARs," *Annual Review of Medicine*, vol. 53, pp. 409–435, 2002.

[34] J. M. Peters, I. Rusyn, M. L. Rose, F. J. Gonzalez, and R. G. Thurman, "Peroxisome proliferator-activated receptor α is restricted to hepatic parenchymal cells, not Kupffer cells: implications for the mechanism of action of peroxisome proliferators in hepatocarcinogenesis," *Carcinogenesis*, vol. 21, no. 4, pp. 823–826, 2000.

[35] P. Tontonoz, E. Hu, R. A. Graves, A. I. Budavari, and B. M. Spiegelman, "mPPARγ2: tissue-specific regulator of an adipocyte enhancer," *Genes and Development*, vol. 8, no. 10, pp. 1224–1234, 1994.

[36] H. B. Wang, Q. G. Wang, X. Y. Zhang et al., "Microarray analysis of genes differentially expressed in the liver of lean and fat chickens," *Animal*, vol. 4, no. 4, pp. 513–522, 2010.

[37] C. A. Hasemann, R. G. Kurumbail, S. S. Boddupalli, J. A. Peterson, and J. Deisenhofer, "Structure and function of cytochromes P450:a comparative analysis of three crystal structures," *Structure*, vol. 3, no. 1, pp. 41–62, 1995.

[38] E. Plagnes-Juan, M. Lansard, I. Seiliez et al., "Insulin regulates the expression of several metabolism-related genes in the liver and primary hepatocytes of rainbow trout (*Oncorhynchus mykiss*)," *Journal of Experimental Biology*, vol. 211, no. 15, pp. 2510–2518, 2008.

[39] D. R. Clemmons, "Role of insulin-like growth factor binding proteins in controlling IGF actions," *Molecular and Cellular Endocrinology*, vol. 140, no. 1-2, pp. 19–24, 1998.

[40] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 5427062, 8 pages
http://dx.doi.org/10.1155/2016/5427062

Research Article

Genome-Wide Analysis of Genes Encoding Methionine-Rich Proteins in *Arabidopsis* and Soybean Suggesting Their Roles in the Adaptation of Plants to Abiotic Stress

Ha Duc Chu,[1] Quynh Ngoc Le,[1,2] Huy Quang Nguyen,[2] and Dung Tien Le[1]

[1]*National Key Laboratory of Plant and Cell Technology, Agricultural Genetics Institute, Vietnam Academy of Agricultural Sciences, Pham Van Dong Road, Hanoi, Vietnam*
[2]*Department of Biochemistry and Plant Physiology, Faculty of Biology, VNU-University of Science, Vietnam National University-Hanoi, Nguyen Trai Street, Hanoi, Vietnam*

Correspondence should be addressed to Dung Tien Le; research@letiendung.info

Received 12 June 2016; Accepted 19 July 2016

Academic Editor: Xuan H. Cao

Oxidation and reduction of methionine (Met) play important roles in scavenging reactive oxygen species (ROS) and signaling in living organisms. To understand the impacts of Met oxidation and reduction in plants during stress, we surveyed the genomes of *Arabidopsis* and soybean (*Glycine max* L.) for genes encoding Met-rich proteins (MRPs). We found 121 and 213 genes encoding MRPs in *Arabidopsis* and soybean, respectively. Gene annotation indicated that those with known function are involved in vital cellular processes such as transcriptional control, calcium signaling, protein modification, and metal transport. Next, we analyzed the transcript levels of MRP-coding genes under normal and stress conditions. We found that 57 AtMRPs were responsive either to drought or to high salinity stress in *Arabidopsis*; 35 GmMRPs were responsive to drought in the leaf of late vegetative or early reproductive stages of soybean. Among the MRP genes with a known function, the majority of the abiotic stress-responsive genes are involved in transcription control and calcium signaling. Finally, *Arabidopsis* plant which overexpressed an MRP-coding gene, whose transcripts were downregulated by abiotic stress, was more sensitive to paraquat than the control. Taken together, our report indicates that MRPs participate in various vital processes of plants under normal and stress conditions.

1. Introduction

Under elevated ROS levels, free and protein-based Met are converted to methionine sulfoxide (MetO) which occurs in a diastereomeric mixture of methionine-*S*-sulfoxide (Met-*S*-O) and methionine-*R*-sulfoxide (Met-*R*-O) [1]. Oxidation of Met was reported to occur in various signaling proteins, thereby modulating their functions [1–9]. For example, calmodulin (CAM), a versatile protein involved in various signaling pathways, including ROS homeostasis in *Arabidopsis* [10], is well known to have its methionine residue oxidation linked to protein dysfunction [11] and loss of protein stability [12]. Nevertheless, efforts to systematically identify all the proteins whose methionine residues are susceptible to oxidation yielded limited results due to the lack of viable tools, including an antibody specific to methionine sulfoxide [13–16]. Recently, Tarrago and colleagues proposed an affinity chromatography approach employing methionine sulfoxide reductase (MSR) to catch interacting partners. Using AtMSRB1 as bait, the authors isolated 24 interacting partners functioning in photosynthesis, translation, and protection against oxidative stress from *Arabidopsis*. The authors found a preference of proteins with higher Met content to bind to the bait, becoming isolated by this approach [17]. Quite recently, Jacques et al. used a newly developed technique called COFRADIC [18] for proteome-wide identification of Met oxidation sites in *Arabidopsis* proteins. Their work revealed 500 sites of Met oxidation in 400 proteins of the plant [19].

Organisms evolved two enzyme families to repair oxidized Met in proteins: methionine-*S*-sulfoxide reductase (MSRA) to reduce Met-*S*-O and methionine-*R*-sulfoxide

reductase (MSRB) to reduce Met-*R*-O. *In vivo* modulation of MSR activities has been reported in yeast [20, 21], fruit flies [22], and mammals [23], which in turn affected their tolerance to oxidative stress and lifespan. In plants, Romero et al. demonstrated the role of *Arabidopsis* plastidial MSRAs in the defense against oxidative stress [24]. In another study, transgenic tomato overexpressing pepper *MSR* gene *CaMSRB2* was found to protect against oxidative stress and *Phytophthora* pathogen infection [25]. Very recently, the role of *Arabidopsis* cytosolic AtMSRB7 and AtMSRB8 in conferring tolerance to oxidative stress was also demonstrated [26], whereas overexpression of AtMSRB1 and AtMSRB2 in plastids did not improve tolerance to high light stress [27].

A number of studies have documented the role of enhancing expression of MSR-coding gene(s) in conferring stress tolerance to plants, but still little is known about the MSR targets which provide such tolerance. A notable study by Laugier et al. provided indirect evidence that *Arabidopsis* plastidial MSRBs confer tolerance to high light stress by acting on cpSRP43 and cpSRP54, thereby maintaining the integrity of the photosystem antenna under environmental constraints [28]. Another recent study by Lee and colleagues on *Arabidopsis* cytosolic MSRB indicated that the enzyme conferred stress tolerance to the plants by acting on two glutathione transferases, GSTF2 and GSTF3 [29]. This study also suggested a list of potential substrates of AtMSRB7. Despite the fact that oxidation and reduction of Met residues in CAM and other calcium signaling proteins were experimentally verified to be involved in regulating the protein's functions [2–4, 11, 12], they were not found among the potential candidates acquired by either the affinity chromatography approach or approaches that employed mass spectroscopy [17, 19, 29]. This line of evidence offers opportunity to argue that either the current approaches for proteome-wide identification of MSR targets pose technical limitations or the oxidation and reduction of Met in many proteins like CAM happen transiently, such that these techniques were unable to help in identifying them. To provide a complementary approach to identify possible targets of MSR in plants, in this work, we surveyed genomes of two dicots, *Arabidopsis* and soybean, to obtain polypeptides of more than 95 residues in length with more than 6% of Met in their sequences. We analyzed these genes in terms of functions, transcriptional responsiveness to stresses, and the conservation of Met residues in HMM profiles (a hidden Markov model-based profile of amino acid residues in a protein domain). A list of genes transcriptionally responsive to stresses with an HMM profile containing highly conserved Met is provided for experimental confirmation by the research community.

2. Materials and Methods

2.1. Materials. Unless otherwise stated, *Arabidopsis thaliana* studied in this work is Columbia ecotype and soybean is of *Williams 82* cultivar. *Arabidopsis* seed overexpressing MRP gene(s) was obtained from *Arabidopsis* FOX line library (RIKEN BioResource Center, Tsukuba, Japan). Briefly, full length cDNA of *Arabidopsis* were placed under the control of 35S promoter and then inserted into a hygromycin-resistant plasmid. The plasmids were then transformed into *Arabidopsis* plants using flower-dipping technique. The development of FOX line library and the line carrying At3g55240 were previously reported [30].

2.2. Growth of Arabidopsis and Stress Treatment. *Arabidopsis* seeds were germinated on 0.5x MS media with or without antibiotics. At 2 weeks old the plants were transferred to soil and allowed to grow at 24 ± 2°C with 16-hour lighting. For paraquat leaf disc assay, 3-week-old rosette leaves were excised and placed on paraquat solutions of various concentrations; after keeping in the dark for 1 hour, the plates were kept at 24°C for 24 hours. The experiments were done in triplicate; each replicate consists of 3 plants.

2.3. Computational Analyses. Protein sequences were downloaded from the PHYTOZOME database (http://www.phytozome.net/) and searched for proteins of 95 residues or more whose sequences contain 6% or more Met by using a java script. The cutoff value for protein length was chosen after consulting reports on the distribution of protein sizes [31, 32]. Genes satisfying these conditions were named Met-rich proteins (MRP): *AtMRPs* for genes from *Arabidopsis* and *GmMRPs* for genes from soybean. Genes encoding MRPs were classified into functional categories using MAPMAN [33]. The PFAM database (http://pfam.xfam.org/) was used to search for HMM profiles as well as possible protein domains.

To obtain transcription levels, microarray data from previous studies were mined, including datasets for *Arabidopsis* under drought and salinity stress [34, 35] and soybean under experimental drought conditions [36]. In that study, for drought treatment of *Arabidopsis*, 2-week-old plants were transferred to soil and allowed to grow for one more week; the 3-week-old plants were then withheld from watering for 10 days. After the tenth day, rosette leaves were collected from both well-watered and drought-stressed plants, frozen in liquid nitrogen, and stored at −80°C until RNA extraction. For high salinity treatment, 10-day-old plants grown on GM media were transferred onto 0.5x MS plates without sucrose, containing either 0 mM (untreated) or 200 mM NaCl and maintained for a period of 24 h. Samples were collected in three biological replicates, frozen in liquid nitrogen, and stored at −80°C until used for RNA extraction. The drought treatment of soybean plants and data acquisition was described previously [36]. Data analyses were performed with functions integrated in MS EXCEL.

3. Results and Discussions

3.1. Occurrence of Genes Encoding MRPs in Arabidopsis and Soybean. An exhaustive search of genes encoding proteins longer than 95 residues and containing 6% Met or more resulted in 121 and 213 genes from *Arabidopsis* and soybean genomes, respectively. Functions of about 50% of those genes were not known. RNA transcription, protein modification, and calcium signaling were the three major functional categories of the MRPs analyzed (Figure 1), indicating the important roles of MRP-coding genes in overall cellular function. Smaller categories include RNA processing and metal

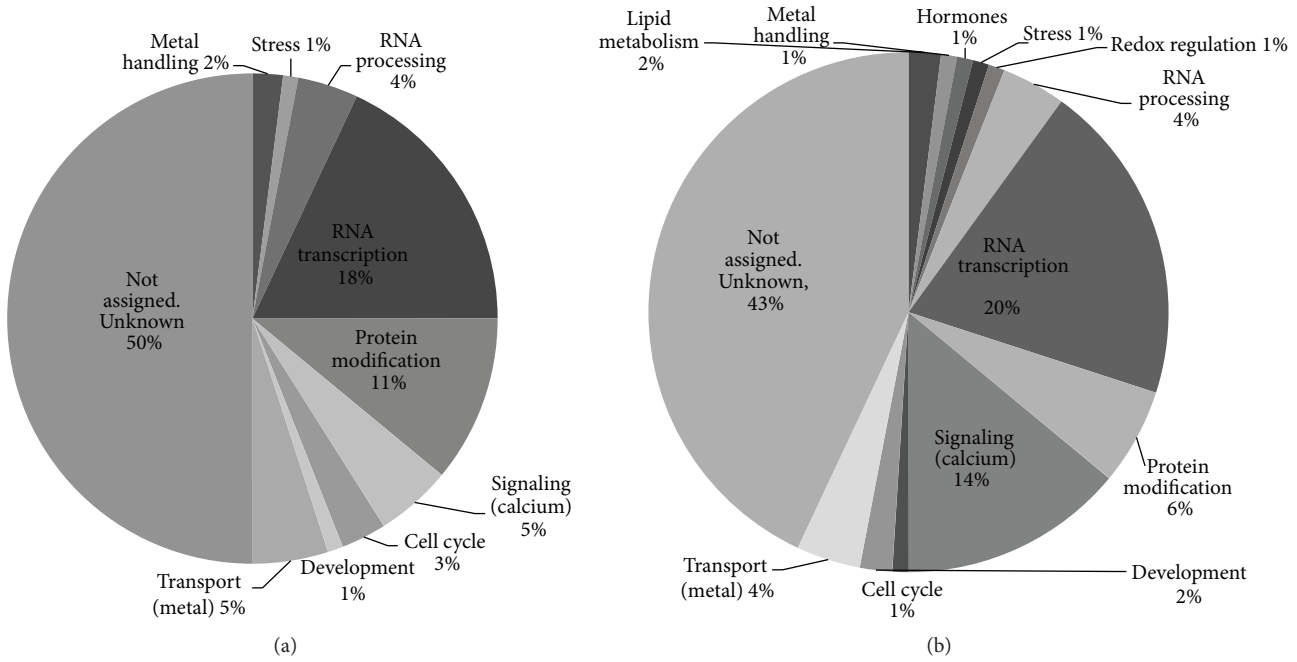


Figure 1: Distribution of genes encoding MRPs into various biological processes in *Arabidopsis* (a) and soybean (b).

transport. Specifically, there were 20 and 44 MRP-coding genes functioning in RNA transcription in *Arabidopsis* and soybean, respectively, making it the most abundant category. The second most abundant category was protein modification, of which 12 and 15 genes were found in *Arabidopsis* and soybean, respectively. There were 6 MRP-coding genes of *Arabidopsis* functioning in calcium signaling, whereas 31 MRP-coding genes found in soybean genome belonged to this category. Among MRP-coding genes identified, soybean has 10 genes distributed in 4 unique categories that were not presented in *Arabidopsis*, namely, lipid metabolism (4 genes), amino acid metabolism (2 genes), hormones (2 genes), and redox regulation (2 genes).

3.2. Stress-Responsive MRP-Coding Genes in Arabidopsis. To obtain MRP-coding genes transcriptionally responsive to abiotic stresses, our published microarray data of drought and high salinity treatments of wild type *Arabidopsis* [34, 35] were analyzed and data mining was performed. Of the 121 *Arabidopsis* genes, expression data of 108 genes were available (Table S1 in the Supplementary Material available online at http://dx.doi.org/10.1155/2016/5427062). Further analyses indicated that 23 and 16 genes were induced and repressed more than 2-fold, respectively, by drought treatment. Under treatment by high salinity, 11 and 17 genes were induced and repressed more than 2-fold, respectively. Among drought- and salt-responsive MRP-coding genes, several genes were previously confirmed to be stress-inducible, of which the expressions of the 10 genes were responsive to both drought and high salinity (Table 1). Most of the genes were responsive to both stressors' code for plant-specific proteins. AT4G33467, encoding an unknown protein, was the most induced gene. Its transcript level was upregulated more than 330-fold by drought. Among stress-repressed genes, AT3G55240 was the most downregulated. Its transcript level was repressed by 60- and 26-fold under drought and high salinity, respectively. The function of this gene is not yet known; however, it was reported that overexpression of this gene in *Arabidopsis* led to the phenotype "pseudo-etiolation in light" [30].

3.3. Stress-Responsive MRP-Coding Genes in Soybean. To identify stress-responsive genes among MRP-coding genes, we performed data mining with the microarray experiments conducted earlier. In these experiments, drought treatments were carried out by withholding water. Leaves of V6 (vegetative) and R2 (reproductive) stages were collected and microarray analyses performed as reported [36]. Transcript levels of all 213 MRP-coding genes were obtained, of which 11 were repressed and 12 genes were induced under drought in V6 *trifolia*. In reproductive leaves (R2 *trifolia*), drought treatment induced 24 genes, whereas only 6 genes were repressed (Table S2). A total of 13 MRP-coding genes were found to be transcriptionally responsive to drought in both vegetative and reproductive-stage leaves (Table 2). The gene Glyma04g37040, which encodes a calmodulin-binding protein CML38, was the most induced gene by drought and its transcript levels were induced 40- and 15-fold in R2 and V6 *trifolia*, respectively. The most repressed gene by drought was Glyma02g10620, encoding a 98-residue protein of unknown function whose transcript levels were repressed 44- and 4-fold in V6 and R2 *trifolia*, respectively.

3.4. Common Stress-Responsive MRP-Coding Genes in Arabidopsis and Soybean. In *Arabidopsis*, 7 AtMRPs were upregulated and 3 other AtMRPs were downregulated under both

Table 1: Genes encoding AtMRPs whose expression levels were responsive to both drought and high salinity.

Number	Locus IDs	Met (%)	Length (a.a.)	Drought versus untreated[1]		Salinity versus untreated[2]		Gene descriptions
				Fold change[3]	q-value	Fold change[3]	q-value	
1	AT1G32560	6.02	134	135.33	0.002	3.31	0.005	LEA group 1 domain-containing protein
2	AT1G33860	8.55	153	2.37	0.092	2.16	0.003	Unknown protein
3	AT3G55240	6.12	95	−60.29	0.007	−26.88	0.001	Overexpression leads to pseudo-etiolation in light phenotype
4	AT3G59900	6.20	130	10.70	0.011	−2.57	0.015	(ARGOS); unknown protein [AT3G59900.1]
5	AT3G62090	6.38	346	64.56	0.020	2.28	0.002	PHYTOCHROME INTERACTING FACTOR 3-LIKE 2
6	AT4G12334	6.25	113	−9.79	0.003	−3.04	0.005	Pseudogene of cytochrome P450 family protein
7	AT4G33467	8.91	102	337.51	0.002	6.16	0.023	Unknown protein [AT4G33467.1]
8	AT4G34590	6.33	159	8.26	0.004	3.27	0.002	GBF6 (*A. thaliana* BASIC LEUCINE-ZIPPER 11)
9	AT5G42325	6.03	233	2.70	0.028	2.45	0.049	Transcription elongation factor-related
10	AT5G67390	7.43	176	−4.17	0.015	−4.15	0.001	Similar to unknown proteins (TAIR:AT1G69360.1)

[1]Two-week-old plants were transferred to soil and allowed to grow for an additional week; the plants were then withheld from watering for 10 days. After 10 d of water with holding, rosette leaves were collected from both well watered and drought-stressed plants in three biological replicates, frozen in liquid nitrogen, and stored at −80°C until used for RNA extraction [35]. [2]10-day-old plants grown on GM media were transferred onto 0.5 × MS plates without sucrose, containing either 0 mM (control) or 200 mM NaCl, and maintained for a period of 24 h. Three independent experiments were performed for each condition. The samples were collected as three biological replicates (10 plants/replicate), frozen in liquid nitrogen, and stored at −80°C until used for RNA extraction [34]. [3]Stress responsive AtMRPs were defined as genes encoding MRP whose expression levels were induced or repressed 2-fold or more with an FDR corrected p value of less than 0.05.

Table 2: Genes encoding GmMRPs whose expression is responsive to drought stress in V6 and R2 leaves.

Number	Glyma ID	Met (%)	Length (a.a.)	V6 *trifolia*		R2 *trifolia*		Gene descriptions	Arabidopsis homologs
				Fold change	q-value	Fold change	q-value		
1	Glyma01g15910	8.08	100	3.63	0.045	4.96	0.042	No original description	
2	Glyma01g15930	6.56	458	−20.34	0.007	−3.87	0.015	UNE10; transcription factor	AT4G00050
3	Glyma02g10620	7.22	98	−44.63	0.007	−4.04	0.053	Overexpression leads to pseudo-etiolation in light	AT3G55240
4	Glyma03g32740	6.04	481	−2.19	0.007	−2.02	0.030	PIF1, PIL5; transcription factor	AT2G20180
5	Glyma04g37040	7.91	140	15.03	0.012	40.08	0.005	Calmodulin-binding protein CML38	AT1G76650
6	Glyma06g39910	10.34	117	3.12	0.067	4.14	0.013	Calcium-binding EF hand family protein	AT4G27280
7	Glyma10g30380	7.43	149	7.53	0.013	5.27	0.026	calmodulin 5; calcium ion binding	AT2G27030
8	Glyma15g05510	7.37	96	2.93	0.025	2.41	0.023	No original description	
9	Glyma16g02510	7.26	125	2.05	0.028	4.63	0.023	Calcium-binding protein, putative	AT2G46600
10	Glyma19g43580	6.7	210	−2.01	0.160	2.42	0.078	GIF, GIF1, AN3 (ANGUSITFOLIA3)	AT5G28640
11	Glyma20g00780	6.69	285	−3.03	0.046	−2.36	0.027	Contains homeodomain (InterPro:IPR009057)	AT1G10820
12	Glyma20g22280	6.59	426	2.25	0.046	2.99	0.056	PIF3, POC1, PAP3, transcription factor	AT1G09530
13	Glyma20g36730	7.89	153	3.06	0.042	2.29	0.129	calmodulin 5; calcium ion binding	AT2G27030

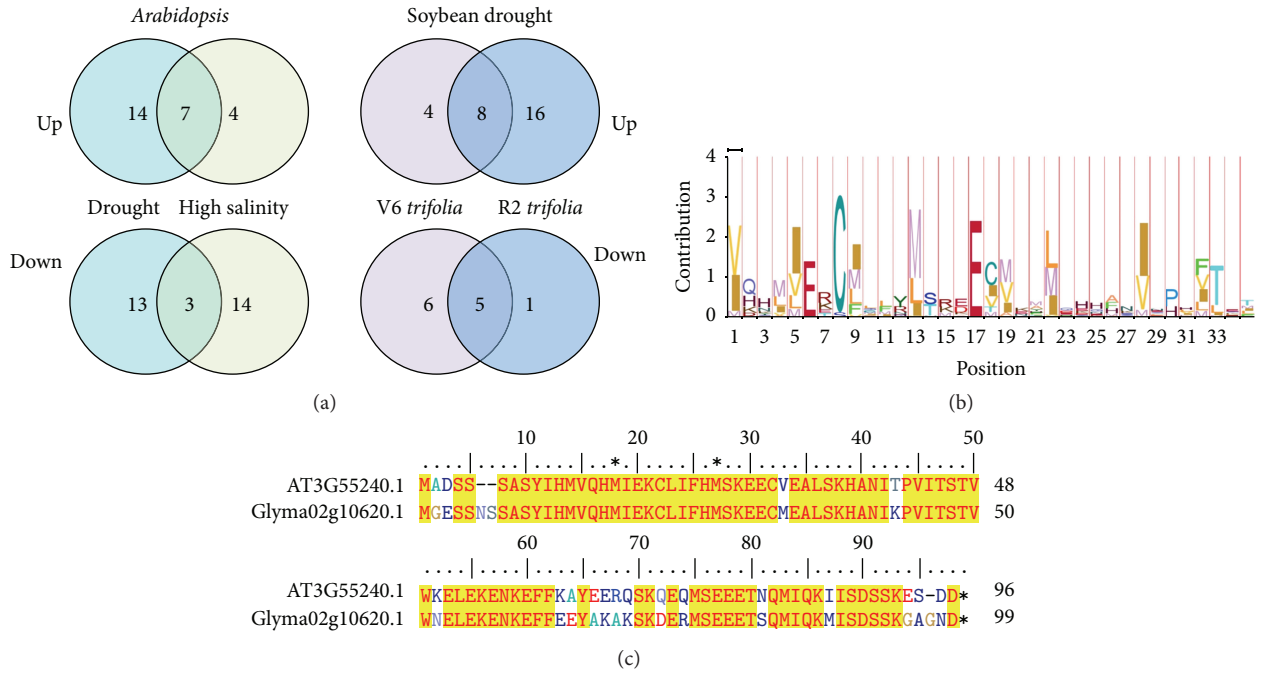


FIGURE 2: Venn diagram analyses of the expression of MRP-coding genes in *Arabidopsis* and soybean under abiotic stresses (a), and HMM profile of *Arabidopsis* and soybean homologs share common responsiveness to drought (b) and their peptide sequence alignment (c).

drought and high salinity. In soybean drought, 8 and 5 GmMRPs were up- or downregulated, respectively, in both V6 and R2 *trifolia* (Figure 2(a)). Among 6 *AtMRPs* encoding calcium signaling proteins, three genes were transcriptionally responsive to either drought or salt (Table S1). At the same time, 14 *GmMRPs* encoding calmodulin-like proteins were transcriptionally responsive to drought in either V6 or R2 *trifolia* or both. This data suggested that calcium signaling plays an important role in the plant's signaling during abiotic stress exposure. In light of previous studies, it is very likely that Met oxidation and reduction of calmodulin may also contribute significantly to the plant's signaling response to abiotic stresses. A number of *AtMRPs* and *GmMRPs* encoding transcription factors were also responsive to abiotic stresses in *Arabidopsis* and soybean (Tables 1 and 2), indicating the involvement of the MRPs in the important cellular activities.

Further analysis of stress-responsive MRP-coding genes identified a gene coding for a plant-specific protein that has homologs in both *Arabidopsis* (AT3G55240) and soybean (Glyma02g10620). These genes encoded highly homologous proteins (>70% identity) of about 100 amino acid residues that share an HMM profile with several conserved Met residues in the N-terminal (Figures 2(b) and 2(c)). In a previous study, it was found that overexpression of this gene in *Arabidopsis* caused a phenotype of pseudo-etiolation in light or leaf bleaching [30], although how such phenotype occurred was not explained. When *Arabidopsis* was transformed with an RNAi construct to downregulate this gene, most of the transgenic plants died at a very early stage and the plants that survived did not show any reduction in the transcription levels, suggesting a vital function of this gene [30]. To determine if the gene At3g55240 is involved in redox stress responses, we acquired the overexpressor line from the *Arabidopsis* FOX line library (RIKEN BioResource Center) and analyzed them. Growing of the plants on MS media and soil confirmed the pseudo-etiolation phenotype (Figure 3(a)). When treated with paraquat in a leaf disc assay, the overexpressor line exhibited higher sensitivity than the wild type control (Figure 3(b)), suggesting the gene may be involved in mediating redox stress responses. This gene and its soybean homolog were both repressed under abiotic stress; thus, increasing its expression may not provide benefit under stress.

3.5. Stress-Responsive cis-Elements of the MRPs' Promoter. To provide further evidence of the stress-responsiveness of MRP-coding genes, we searched for the presence of known stress-responsive *cis*-elements in the promoters of the genes in *Arabidopsis* 1 kbs upstream of the transcriptional start sites. We found that promoters of *AtMRPs* contain 23, 26, and 16 *cis*-elements of ABRE, MYBR, and MYCR, respectively. On average there are 0.54 *cis*-elements per AtMRP and 0.86 *cis*-elements for each stress-responsive *AtMRP*, indicating a slight enrichment of stress-responsive *cis*-elements among drought- and/or salt-responsive *AtMRPs*.

3.6. Subcellular Localization of MRPs. Chloroplast and mitochondria are the two types of cellular organelles which generate high levels of ROS. Thus, identification of MRPs localizing to these organelles may shed light on their functions. Using amino acid sequences of AtMRPs and the prediction tools such as TargetP [37], pSORT [38], and CELLO [39], we

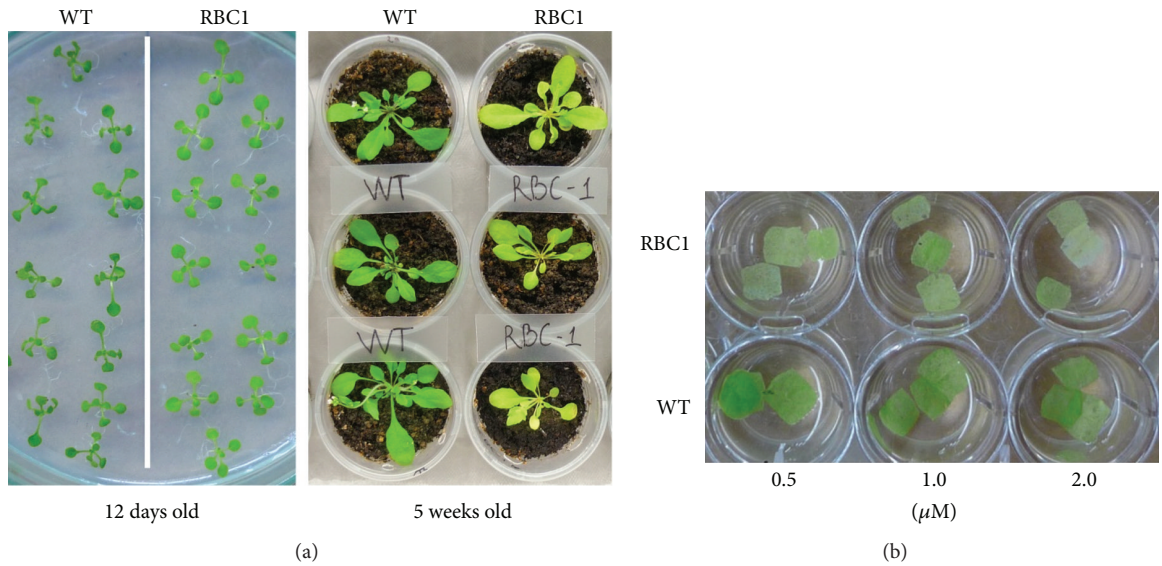


FIGURE 3: Phenotypes of an *Arabidopsis* overexpressed a gene encoding a MRP under normal physiological condition (a) and under a paraquat leaf disc assay (b). WT, wild type control; RBC1, *Arabidopsis* plant overexpressed At3g55240 gene.

identified all AtMRPs targeted to either chloroplast or mitochondria (Table S1). Among 121 AtMRPs, 21 were predicted to target chloroplast by either ChloroP or pSORT or both; 9 were predicted to localize to mitochondria by either pSORT or CELLO or both.

In conclusion, as a complementary approach to other studies on the identification of targets of Met oxidation and reduction in plants, here we found a large number of MRPs involved in important cellular processes such as RNA transcription control and calcium signaling. Several genes encoding these MRPs were transcriptionally responsive to abiotic stresses, such as drought and high salinity, suggesting their roles in the adaptation of plants to these stressors. The fact that promoters of the genes encoding stress-responsive MRPs are slightly enriched in *cis*-element responsive to stresses and that product of these genes were predicted to localize in ROS-enriched organelles, chloroplast and mitochondria, further confirm their functions. Taken together, this work proposes unique evidence on methionine oxidation in proteins and its possible role in regulating the plant's activities.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was funded by the National Foundation for Science and Technology Development (NAFOSTED) under the Grant no. 106-NN.02-2013.46 to Dung Tien Le. Stephanie K. Dalquist (skd@mit.edu) is acknowledged for correcting English usage of this paper. Dung Tien Le wishes to thank Vadim Gladyshev (Harvard Medical School) and Son Tran (RIKEN CSRS) for helpful discussions during the early phase of this study.

References

[1] E. R. Stadtman, "Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions," *Annual Review of Biochemistry*, vol. 62, pp. 797–821, 1993.

[2] E. M. Balog, L. E. Norton, R. A. Bloomquist et al., "Calmodulin oxidation and methionine to glutamine substitutions reveal methionine residues critical for functional interaction with ryanodine receptor-1," *The Journal of Biological Chemistry*, vol. 278, no. 18, pp. 15615–15621, 2003.

[3] N. J. Carruthers and P. M. Stemmer, "Methionine oxidation in the calmodulin-binding domain of calcineurin disrupts calmodulin binding and calcineurin activation," *Biochemistry*, vol. 47, no. 10, pp. 3085–3095, 2008.

[4] J. R. Erickson, M.-L. A. Joiner, X. Guan et al., "A dynamic pathway for calcium-independent activation of CaMKII by methionine oxidation," *Cell*, vol. 133, no. 3, pp. 462–474, 2008.

[5] B. Ezraty, R. Grimaud, M. El Hassouni, D. Moinier, and F. Barras, "Methionine sulfoxide reductases protect Ffh from oxidative damages in *Escherichia coli*," *The EMBO Journal*, vol. 23, no. 8, pp. 1868–1877, 2004.

[6] D. T. Le, M.-Y. Yoon, Y. T. Kim, and J.-D. Choi, "Roles of conserved methionine residues in tobacco acetolactate synthase," *Biochemical and Biophysical Research Communications*, vol. 306, no. 4, pp. 1075–1082, 2003.

[7] Z. Su, J. Limberis, R. L. Martin et al., "Functional consequences of methionine oxidation of hERG potassium channels," *Biochemical Pharmacology*, vol. 74, no. 5, pp. 702–711, 2007.

[8] H. Sun, J. Gao, D. A. Ferrington, H. Biesiada, T. D. Williams, and T. C. Squier, "Repair of oxidized calmodulin by methionine sulfoxide reductase restores ability to activate the plasma

membrane Ca-ATPase," *Biochemistry*, vol. 38, no. 1, pp. 105–112, 1999.

[9] W. Vogt, "Oxidation of methionyl residues in proteins: tools, targets, and reversal," *Free Radical Biology and Medicine*, vol. 18, no. 1, pp. 93–105, 1995.

[10] F. Takahashi, T. Mizoguchi, R. Yoshida, K. Ichimura, and K. Shinozaki, "Calmodulin-dependent activation of MAP kinase for ROS homeostasis in *Arabidopsis*," *Molecular Cell*, vol. 41, no. 6, pp. 649–660, 2011.

[11] J. Snijder, R. J. Rose, R. Raijmakers, and A. J. R. Heck, "Site-specific methionine oxidation in calmodulin affects structural integrity and interaction with Ca^{2+}/calmodulin-dependent protein kinase II," *Journal of Structural Biology*, vol. 174, no. 1, pp. 187–195, 2011.

[12] J. Gao, D. H. Yin, Y. Yao et al., "Loss of conformational stability in calmodulin upon methionine oxidation," *Biophysical Journal*, vol. 74, no. 3, pp. 1115–1134, 1998.

[13] T. Le Dung, X. Liang, D. E. Fomenko et al., "Analysis of methionine/selenomethionine oxidation and methionine sulfoxide reductase function using methionine-rich proteins and antibodies against their oxidized forms," *Biochemistry*, vol. 47, no. 25, pp. 6685–6694, 2008.

[14] X. Liang, A. Kaya, Y. Zhang, D. T. Le, D. Hua, and V. N. Gladyshev, "Characterization of methionine oxidation and methionine sulfoxide reduction using methionine-rich cysteine-free proteins," *BMC Biochemistry*, vol. 13, no. 1, article 21, 2012.

[15] D. B. Oien, T. Canello, R. Gabizon et al., "Detection of oxidized methionine in selected proteins, cellular extracts and blood serums by novel anti-methionine sulfoxide antibodies," *Archives of Biochemistry and Biophysics*, vol. 485, no. 1, pp. 35–40, 2009.

[16] N. B. Wehr and R. L. Levine, "Wanted and wanting: antibody against methionine sulfoxide," *Free Radical Biology and Medicine*, vol. 53, no. 6, pp. 1222–1225, 2012.

[17] L. Tarrago, S. Kieffer-Jaquinod, T. Lamant et al., "Affinity chromatography: a valuable strategy to isolate substrates of methionine sulfoxide reductases?" *Antioxidants and Redox Signaling*, vol. 16, no. 1, pp. 79–84, 2012.

[18] B. Ghesquière, V. Jonckheere, N. Colaert et al., "Redox proteomics of protein-bound methionine oxidation," *Molecular & Cellular Proteomics*, vol. 10, no. 5, Article ID M110.006866, 2011.

[19] S. Jacques, B. Ghesquière, P.-J. De Bock et al., "Protein methionine sulfoxide dynamics in *Arabidopsis thaliana* under oxidative stress," *Molecular and Cellular Proteomics*, vol. 14, no. 5, pp. 1217–1229, 2015.

[20] A. Koc, A. P. Gasch, J. C. Rutherford, H.-Y. Kim, and V. N. Gladyshev, "Methionine sulfoxide reductase regulation of yeast lifespan reveals reactive oxygen species-dependent and -independent components of aging," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 21, pp. 7999–8004, 2004.

[21] D. T. Le, B. C. Lee, S. M. Marino et al., "Functional analysis of free methionine-R-sulfoxide reductase from *Saccharomyces cerevisiae*," *Journal of Biological Chemistry*, vol. 284, no. 7, pp. 4354–4364, 2009.

[22] H. Ruan, X. D. Tang, M.-L. Chen et al., "High-quality life extension by the enzyme peptide methionine sulfoxide reductase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 5, pp. 2748–2753, 2002.

[23] J. Moskovitz, S. Bar-Noy, W. M. Williams, J. Requena, B. S. Berlett, and E. R. Stadtman, "Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 23, pp. 12920–12925, 2001.

[24] H. M. Romero, B. S. Berlett, P. J. Jensen, E. J. Pell, and M. Tien, "Investigations into the role of the plastidial peptide methionine sulfoxide reductase in response to oxidative stress in Arabidopsis," *Plant Physiology*, vol. 136, no. 3, pp. 3784–3794, 2004.

[25] S.-K. Oh, K.-H. Baek, E. S. Seong et al., "CaMsrB2, pepper methionine sulfoxide reductase B2, is a novel defense regulator against oxidative stress and pathogen attack," *Plant Physiology*, vol. 154, no. 1, pp. 245–261, 2010.

[26] C.-W. Li, S.-H. Lee, P.-S. Chieh, C.-S. Lin, Y.-C. Wang, and M.-T. Chan, "Arabidopsis root-abundant cytosolic methionine sulfoxide reductase B genes MsrB7 and MsrB8 are Involved in tolerance to oxidative stress," *Plant and Cell Physiology*, vol. 53, no. 10, pp. 1707–1719, 2012.

[27] E. Laugier, L. Tarrago, A. Courteille et al., "Involvement of thioredoxin y2 in the preservation of leaf methionine sulfoxide reductase capacity and growth under high light," *Plant, Cell and Environment*, vol. 36, no. 3, pp. 670–682, 2013.

[28] E. Laugier, L. Tarrago, C. Vieira Dos Santos, F. Eymery, M. Havaux, and P. Rey, "*Arabidopsis thaliana* plastidial methionine sulfoxide reductases B, MSRBs, account for most leaf peptide MSR activity and are essential for growth under environmental constraints through a role in the preservation of photosystem antennae," *The Plant Journal*, vol. 61, no. 2, pp. 271–282, 2010.

[29] S.-H. Lee, C.-W. Li, K. W. Koh et al., "MSRB7 reverses oxidation of GSTF2/3 to confer tolerance of *Arabidopsis thaliana* to oxidative stress," *Journal of Experimental Botany*, vol. 65, no. 17, pp. 5049–5062, 2014.

[30] T. Ichikawa, M. Nakazawa, M. Kawashima et al., "The FOX hunting system: an alternative gain-of-function gene hunting technique," *Plant Journal*, vol. 48, no. 6, pp. 974–985, 2006.

[31] A. Tiessen, P. Pérez-Rodríguez, and L. Delaye-Arredondo, "Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes," *BMC Research Notes*, vol. 5, article 85, 2012.

[32] J. Zhang, "Protein-length distributions for the three domains of life," *Trends in Genetics*, vol. 16, no. 3, pp. 107–109, 2000.

[33] O. Thimm, O. Bläsing, Y. Gibon et al., "MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes," *Plant Journal*, vol. 37, no. 6, pp. 914–939, 2004.

[34] R. Nishiyama, D. T. Le, Y. Watanabe et al., "Transcriptome analyses of a salt-tolerant cytokinin-deficient mutant reveal differential regulation of salt stress response by cytokinin deficiency," *PLoS ONE*, vol. 7, no. 2, Article ID e32124, 2012.

[35] R. Nishiyama, Y. Watanabe, M. A. Leyva-Gonzalez et al., "*Arabidopsis* AHP2, AHP3, and AHP5 histidine phosphotransfer proteins function as redundant negative regulators of drought stress response," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 12, pp. 4840–4845, 2013.

[36] D. T. Le, R. Nishiyama, Y. Watanabe et al., "Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis," *PLoS ONE*, vol. 7, no. 11, Article ID e49522, 2012.

[37] O. Emanuelsson, H. Nielsen, S. Brunak, and G. von Heijne, "Predicting subcellular localization of proteins based on their

N-terminal amino acid sequence," *Journal of Molecular Biology*, vol. 300, no. 4, pp. 1005–1016, 2000.

[38] P. Horton, K.-J. Park, T. Obayashi et al., "WoLF PSORT: protein localization predictor," *Nucleic Acids Research*, vol. 35, no. 2, pp. W585–W587, 2007.

[39] C.-S. Yu, Y.-C. Chen, C.-H. Lu, and J.-K. Hwang, "Prediction of protein subcellular localization," *Proteins*, vol. 64, no. 3, pp. 643–651, 2006.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 3219276, 6 pages
http://dx.doi.org/10.1155/2016/3219276

Research Article

SSR Mapping of QTLs Conferring Cold Tolerance in an Interspecific Cross of Tomato

Yang Liu, Tengxia Zhou, Haiyan Ge, Wen Pang, Lijie Gao, Li Ren, and Huoying Chen

School of Agriculture and Biology, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai 200240, China

Correspondence should be addressed to Huoying Chen; chhy@sjtu.edu.cn

Received 13 April 2016; Accepted 14 June 2016

Academic Editor: Wenwei Xiong

A population of 146 RILs (Recombinant Inbred Line) was derived from the cross between a cold-sensitive cultivated *Solanum lycopersicum* L. XF98-7 and a cold-tolerant wild *Solanum pimpinellifolium* LA2184. Relative germination ratio (RGR) and chilling index (CI) were used to evaluate the cold tolerance of the parental lines and RILs. It was found that the RGR and CI were significantly different between *S. lycopersicum* XF98-7 and *S. pimpinellifolium* LA2184 under cold treatment, indicating that wild species was more adapted to chilling temperature. The continuous and normal distribution of RGR and CI in RIL population suggested that the trait of cold tolerance was a typically quantitative trait controlled by multigenes. The molecular linkage map was constructed by using 120 simple-sequence repeat (SSR) markers, resulting in 15 linkage groups, with a total distance of 256.8 cM and average interval of 2.14 cM. Five QTLs controlling RGR and four QTLs for CI were detected with genetic contribution ranging from 0.95% to 19.55%. Thus, the nine QTLs will provide references for further fine position mapping for cold tolerance. The polymorphic markers could be used as a way of indirectly selecting the plant trait of interest and would promote developing new tomato variety by marker-assisted selection.

1. Introduction

Tomato (*Solanum lycopersicum* L.), one of the most widely cultivated vegetable crops, was originated from South America. Its rich nutrients and unique flavor promote health benefits to human. As a model plant, tomato also plays an important role in modern molecular biology science and genetic breeding research.

Tomato is a thermophilic vegetable and sensitive to low temperature. With the optimal growth temperature of 25°C, the germination and vegetative growth are inhibited when temperature is lower than 10°C and irreparable damage would occur under 6°C [1]. Thus, low temperature becomes one of the main restrictions for the yield improvement and geographic expansion in tomato. Breeding experts have been aiming at developing cultivated tomato with improved cold tolerance [2]. Thanks to the wild species that survive in suboptimal environment, they become the natural germplasm repository with abundant genetic diversity, allowing us to select the desirable or required characteristics, such as resistance to cold, drought, and diseases in tomato [3]. For example, wild species of *S. habrochaites* and *S. pimpinellifolium* growing in high-attitude regions often encounter chilling temperature at night, but they can survive and thrive during their growing seasons [4, 5]. Walker et al. found cold tolerance segregation in the BC_1F_2 population from the cross of cultivated tomato and *S. habrochaites* by investigating the chlorophyll fluorescence [6]. Foolad et al. also discovered the segregation in the BC_1S_1 population derived from cultivated tomato and *S. pimpinellifolium* by investigating seed germination under low temperature [7, 8].

The characteristic of cold tolerance in tomato is a quantitative trait, correlated with multigenes. Some quantitative trait loci (QTLs) for cold tolerance have been identified by using polymorphic molecular markers. The cold tolerance of a BC_1 population from the parental lines of *S. lycopersicum* NC84173 and *S. pimpinellifolium* LA722 was assessed, and three QTLs on chromosome 1 and two on chromosome 4 were detected by using 151 RFLP markers [7]. Using a BC_1 population between cultivated tomato T5 and wild accession LA1778, six QTLs were found to be linked to the shoot wilting after 2 hours of exposure to chilling temperature

(4°C) of the population's root [9]. The QTL mapping from BC_3S_1 population derived from *S. lycopersicum* 9706 and *S. pimpinellifolium* LA722 suggested that five QTLs were linked to seed germination and two related to vegetative growth for cold tolerance, with genetic contribution from 3.3% to 32.9% [10]. The root chilling experiment with a RILs population derived from cultivated tomato T5 and wild tomato LA1778 mapped a major QTL on chromosome 9 responsible for shoot turgor maintenance [11].

In this study, a RIL population, generated from the cross of *S. lycopersicum* XF98-7 and *S. pimpinellifolium* LA2184, was used to locate quantitative trait loci for cold tolerance by investigating the relative germination rate (RGR) and chilling index (CI) under low temperature conditions.

2. Materials and Methods

2.1. Plant Material. 146 RIL individuals (F_8) were developed from the cross *S. lycopersicum* XF98-7 (cultivated and cold-sensitive tomato) and *S. pimpinellifolium* LA2184 (wild and cold-tolerant tomato). *S. lycopersicum* XF98-7 and *S. pimpinellifolium* LA2184 were obtained from Shanghai Jiao Tong University in China and University of California in USA. The RIL population was constructed by Tomato Research Team from School of Agriculture and Biology of Shanghai Jiao Tong University.

When 4-5 true leaves appeared, leaf tissue from the RIL population and parental lines were collected for total genomic DNA extraction, which was used to identify molecular markers linked to the trait of interest.

2.2. The Evaluation of Cold Tolerance for Parental Lines and RIL Population. 100 random seeds of the parental lines and the RIL population were selected, sown on sterile petri plate that had wet filter papers on the bottom, and then placed into the dark incubators. The control samples were maintained under 25 ± 1°C, while the treated samples were kept at 11 ± 1°C. We used three biological replicates for RIL population and four for the parental lines. The amount of seed germination was recorded for 28 consecutive days after sample treatment, and the total seed germinating in the 28 days was available to calculate the germination ratio. The germination ratio (GR) and relative germination ratio (RGR) were used to estimate the cold tolerance of the individuals. A high RGR indicated a strong low temperature tolerance, which was calculated by the equation of RGR = (RG in cold treatment/RG in control treatment) × 100%.

After germination, 12 seedlings of the parental lines and RIL population were planted in 72-cell flats. At the stage of 4-5 true leaves, the plants were transferred to incubators with the light cycle of 12 h days/12 h nights and chilling temperature at 2 ± 1°C. After 48 hours, the chilling injuries of each individual by using chilling indexes (CI) were estimated for three replicates of RILs and four for parental lines. The injuries of the plants were visually scored based on the leaves and shoots wilting with 0 to 4 scale criterion. Score of 0 meant there were normal leaves and shoots on plants. Score of 1 meant the plant had only few flaccid leaf tips. Score of 2 represented that approximately 50% leaf tips were flaccid, but the main shoots were in normal condition. Score of 3 represented that more than 50% leaf tips were flaccid and shoots were wilted. Score of 4 described that the plant completely died. We used CI value to measure the cold tolerance by the equation of $CI = \sum Xa/(n\sum X) = (X_1a_1 + X_2a_2 + \cdots + X_na_n)/nT.X_n$ was the number of the plants of each injury score, and a_n stands for the score of the injury. T was the total number of investigated plants. A low CI indicated that the leaves and shoots were less injured and strongly tolerant for chilling temperature.

2.3. Date Analysis Procedures. The SPSS Statistics software version 20.0 (IBM Corp., Armonk, NY, USA) was used to analyze data variance, normal distribution, and correlation coefficient between variables.

2.4. DNA Extraction and SSR Reaction. The CTAB method of DNA extraction for 146 RILs and parental lines was used following the protocol of Murray and Thompson [12].

The sequence of 1183 primer pairs for SSR markers was obtained from the Sol Genomics Network (https://solgenomics.net/). The primers were synthesized by GenScript (Nanjing) Co., Ltd. The polymerase chain reaction (PCR) with a total volume of 20 μL contained 50 ng DNA template, 2.0 mM Mg^{2+}, 1x PCR buffer, 0.2 mM dNTPs mixture, 100 nM forward and reverse primers, and 1 U Taq DNA polymerase. The thermal program for PCR was an initial denaturation for 3 min at 94°C, denaturation for 30 s at 94°C, annealing for 45 s at 55–60°C, extension for 60 s at 72°C for 35 cycles, and an extra extension for 10 min at 72°C.

The PCR amplification products were separated by 5–8% vertical polyacrylamide gels and silver stained following the method of Zhang et al. [13].

2.5. Construction of Linkage Map and QTL Analysis. In this study, the molecular linkage map was constructed by JionMap 4.0 software with a LOD value 3.0 as the threshold. The map genetic distance (cM) was calculated in Kosambi function.

The location and effects of the QTLs were detected in composite interval mapping (CIM) method by using Windows QTL Cartographer version 2.5 software, and a minimum LOD score of 2.0 was set to indicate the presence of QTLs. The QTL nomenclature followed McCouch et al. [14].

3. Result

3.1. Chilling Tolerance of Parental Lines and RIL Population. The parental lines showed significant difference in RGR and CI in this study (Table 1). The RGR of *S. pimpinellifolium* LA2184 was significantly higher than that of *S. lycopersicum* XF98-7, and the CI of *S. pimpinellifolium* LA2184 was significantly lower than *S. lycopersicum* XF98-7's. The data indicated that *S. pimpinellifolium* LA2184 was much more cold tolerant than *S. lycopersicum* XF98-7, which was consistent with the previous assumption that wild species was more adapted to chilling temperature.

Continuous distribution was observed in RGR and CI trait among the RIL population (Figure 1). The Skewness and Kurtosis absolute values (0.327 and 0.902) of RGR trait in RIL were less than 1.0, so RGR followed approximately

Table 1: Phenotype analysis of parental lines.

Materials	Relative germination ratio (E ± SD)	Chilling index (E ± SD)
XF98-7	0.137 ± 0.025[a]	0.713 ± 0.052[a]
LA2184	0.626 ± 0.088[b]	0.345 ± 0.050[b]

Note: different small letters in each column mean significant difference at 0.05 level.

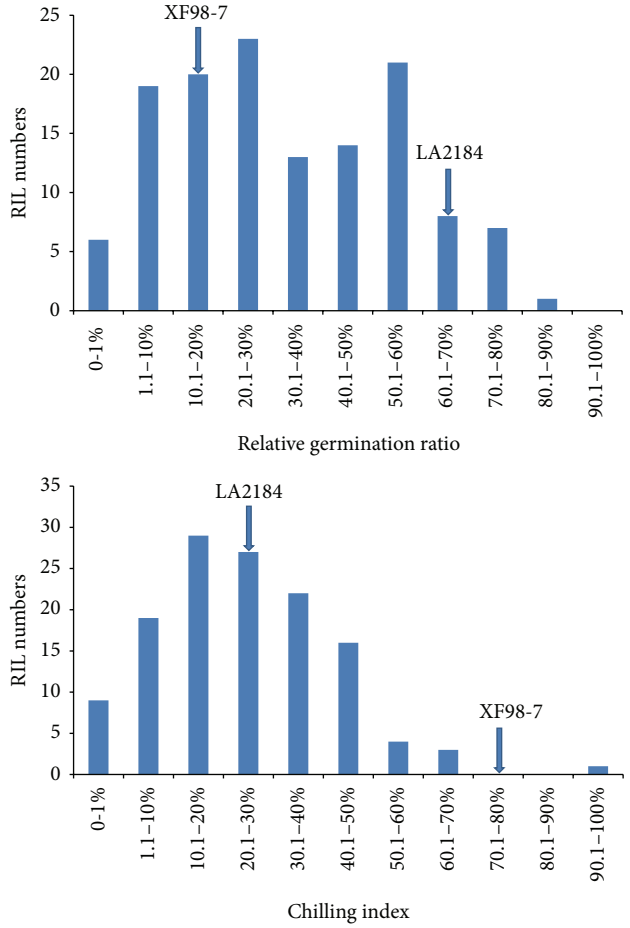


Figure 1: Frequency distribution of relative germination ratio and chilling index associated with cold tolerance in parental lines and RIL population.

normal distribution, and so was the CI trait (with Skewness = 0.382, and Kurtosis = 0.487). The data of the phenotypes also showed an obvious transgressive segregation, 23.3% of the RIL population expressed lower RGR than *S. lycopersicum* XF98-7, and 7.5% of RIL exhibited higher RGR than *S. pimpinellifolium* LA2184, and 44.52% demonstrated lower CI than *S. pimpinellifolium* LA2184.

The correlation analysis between RGR and CI showed that there was no significant correlation at a given significance level $P = 0.05$. The result indicated the cold tolerance of tomato during seed germination and vegetative growth was independent probably due to the different genetic background.

3.2. Genotype Analysis and Linkage Map Construction. A total of 1183 SSR primers were screened in parental lines and 142 primers (12.0%) were polymorphic. The informative SSR markers were continued to genotype RIL population. A molecular linkage map was built based on 120 SSR markers with a threshold value of 3.0 by using JoinMap 4.0 software (Figure 2). The molecular linkage map contained 15 linkage groups, with two groups on chromosomes 3, 4, and 6 and one group on other chromosomes. The total distance of the linkage map was 256.8 cM with an average interval distance of 2.14 cM. The range of genetic distance between two markers was from 0.18 cM to 8.30 cM.

3.3. Identification of QTL Associated with Cold Tolerance. The QTLs analysis was conducted by Windows QTL Cartographer version 2.5 software by using CIM method. Nine QTLs were mapped, linked to the cold tolerance of tomato, and distributed on chromosomes 1, 2, 3, 4, 9, and 12. The genetic contribution of single QTL ranged from 0.95% to 19.55% (Table 2).

There were five QTLs related to RGR, with two located on chromosome 1 and the other three on chromosomes 4, 9, and 12. The biggest and smallest genetic contributions were 19.55% and 5.95%. Except qRGI-4-1 with an additive effect 0.0763, all the positive genes were contributed from *S. lycopersicum* XF98-7. Four QTLs related to CI were observed on chromosomes 1, 2, 3, and 9, responsible for 28.91% of the total phenotypic variation. The positive genes of qCI-1-1 and qCI-3-1 were donated from *S. pimpinellifolium* LA2184, and the genes of qCI-2-1 and qCI-9-1 were from *S. lycopersicum* XF98-7.

4. Discussion

In spite of different evaluation criteria, we used the value of RGR and CI on a RIL population derived from the cross of *S. lycopersicum* XF98-7 and *S. pimpinellifolium* LA2184. There was a significant separation for the seed germination rate and CI on RILs [10, 15, 16]. We constructed a molecular linkage map from 120 SSR markers and also detected QTLs related to cold tolerance in tomato. The result revealed that the RGR and CI were significantly different between the parental lines, with a higher RGR and a lower CI for the cold-tolerant wild species of *S. pimpinellifolium* LA2184 than the cold-sensitive cultivated species of *S. lycopersicum* XF98-7. The continuous and normal distribution observed in RGR and CI among RIL population indicated that the trait of cold tolerance was a typically quantitative trait controlled by multigenes. It was surprising that there were few RILs with CI trait similar to *S. lycopersicum* XF98-7. We checked the fruits of RILs and found a partial separation, too. The fruit phenotype tend to be like

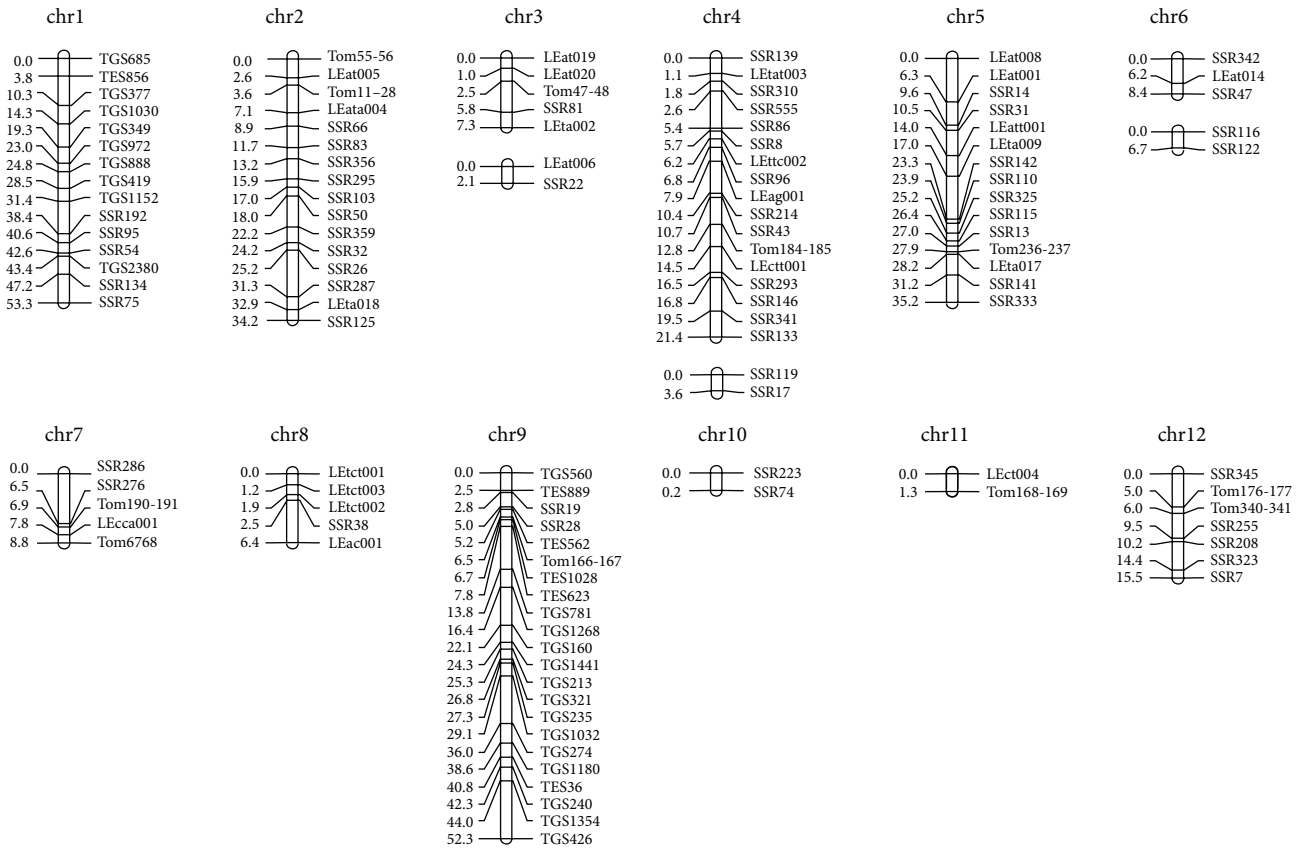


Figure 2: The genetic linkage map for SSR makers in RIL population.

Table 2: QTLs analysis for cold tolerance in RIL population using CIM method.

Trait	QTL	Chromosome	Position (cM)	Marker interval	LOD score	R^2 (%)	Additive effect
RGR	qRGI-1-1	1	40.4	SSR192 - SSR95*	5.45	19.55	−0.1379
	qRGI-1-2	1	47.2	TGS2380 - SSR134*	2.53	8.52	−0.0888
	qRGI-4-1	4	10.4	LEag001 - SSR214*	2.06	6.02	0.0763
	qRGI-9-1	9	7.8	TES623* - TGS781	2.12	5.95	−0.0617
	qRGI-12-1	12	8.0	Tom340-341* - SSR255	4.26	11.33	−0.1143
CI	qCI-1-1	1	9.8	TES856* - TGS377	3.25	0.95	0.0208
	qCI-2-1	2	18.0	SSR50* - SSR359	2.96	10.34	−0.0828
	qCI-3-1	3	0.0	LEat019* - LEat020	3.01	10.31	0.0776
	qCI-9-1	9	26.8	TGS321* - TGS235	2.35	7.31	−0.0652

*The closer molecular marker to the QTL position.

S. pimpinellifolium LA2184, and there were much more small fruits. We assumed it was because of the strong resistance and high ratio of fruit set of *S. pimpinellifolium* LA2184; more small fruit-plants were selected when the RIL population was built. There was no significant correlation between RGR and CI in RILs, which was consistent with the study done by Foolad [17].

Genetic markers have made the process of plant breeding more efficient and greatly promoted genetic research in plants, especially for QTL mapping to find genetic variation contributed to a trait we are interested in. Foolad et al. took advantage of polymorphic markers and found five QTLs associated with seed germination in low temperature on chromosomes 1 and 4 from the BC_1S_1 population derived from *S. lycopersicum* NC84173 and *S. pimpinellifolium* LA722 [7]. Three major QTLs on chromosomes 4, 8, and 9 and six minor QTLs on chromosomes 1, 7, 9, 11, and 12 were found to be linked to seed rapid germination in low temperature [18]. Another five QTLs from chromosomes 1, 2, 5, and 11 were also detected to be associated with cold tolerance [10]. In this study, five QTLs involved in RGR were identified on chromosomes 1, 4, 9, and 11. In summary, QTL conferring

cold tolerance on chromosome 1 was overlapped from the 4 studies above and our detection that QTLs on chromosomes 4, 9, and 11 were consistent with that in Foolad's studies [17]. The possibility of QTLs linked to cold tolerance located on chromosomes 1, 4, 9, and 11 provides references for fining position mapping and the genetic markers would promote developing new crop variety by marker-assisted selection.

Four QTLs conferring CI were detected on chromosomes 1, 2, 3, and 9, with the contribution of single QTL to phenotypic variation between 0.95% and 10.34%. There was neither correlation between the RGR and CI phenotype data in RIL population nor colocalization found for the two traits, which indicated that cold tolerance of tomato during seed germination and vegetative growth is controlled by different genes [19]. Vallejos and Tanksley found QTLs on chromosomes 6, 7, and 12 that affected the cold tolerance of tomato plants [20]. Liu et al. located two QTLs related to cold resistance in young plants on chromosomes 2 and 8, in which the genetic contributions were 3.3% and 4.9% [10]. Goodstal et al. identified a major QTL for shoot turgor maintenance when root was exposed to chilling temperature. Despite multiple QTLs conferring cold tolerance in vegetative growth, their locations were independent almost across the whole genome which lead to low genetic contribution to the variant phenotype [21]. And we can see that the results all above differ from each other; no QTLs locate in the same region. Except for the material difference in the studies, the density of the molecular linkage map also influenced the number of and the related QTLs. A single QTL with high effect value would be broken into several minor QTLs when the marker density of the linkage map increases and the population size gets larger [11].

The past studies showed that the cold tolerance of tomato was a complex quantitative trait controlled by multiple genes, and genes that control the low temperature resistant trait during different developmental stages were not the same. Though much effort had been made to explain the genetic background of tomato response to cold stress and some related QTLs were mapped, the chromosome and the position where the QTLs located were quite different and the repeatability of the results was low. We assume that no uniform phenotypic assessment criteria are a main obstruction for QTLs detection besides the inherent complexity of cold resistance trait. For example, to evaluate the cold tolerance during seed germination, the time when germination rate reaches 25% (T25), 50% (T50), and 75% (T75) was induced by Foolad et al. [7, 22], while Zhao et al. [23] set the inhibition growth of radicles that recovered from cold stress for 3 days as a criterion. Liu et al. [10] adopted germination index (GI) and relative germination index (RGI) to assess tomato cold tolerance. More different standards to evaluate the cold tolerance were used, such as chlorophyll fluorescence, the degree of shoot wilting, root ammonium uptake, the ratio of shoot fresh weight, and the leaf and shoot injuries [6, 9, 16, 24]. Thus, the detection results of QTLs conferring to cold tolerance in vegetative growth were greatly variable and low repeatable. To have a profound understanding of cold tolerance in tomato and accelerate the process of developing commercial cultivars tomato with improved cold tolerance, a uniform standard to evaluate the chilling injuries should be established immediately.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The research was financially supported by grant of Shanghai Science and Technology Committee (no. 13391901201), Shanghai Agricultural Committee (no. 5 2013), and National High Technology Research and Development Program of China (no. 2012AA100104).

References

[1] C. Geisenberg and K. Stewart, "Field crop management," in *The Tomato Crop: A Scientific Basis for Improvement*, J. Atherton and J. Rudich, Eds., pp. 511–557, Springer Science & Business Media, 1986.

[2] J. H. Venema, P. Linger, A. W. Van Heusden, P. R. Van Hasselt, and W. Brüggemann, "The inheritance of chilling tolerance in tomato (*Lycopersiconspp*.)," *Plant Biology*, vol. 7, no. 2, pp. 118–130, 2005.

[3] C. M. Rick, "Genetic variability in tomato species," *Plant Molecular Biology Reporter*, vol. 1, no. 2, pp. 81–87, 1983.

[4] B. D. Patterson, R. Paull, and R. M. Smillie, "Chilling resistance in *Lycopersicon hirsutum* Humb. & Bonpl., a wild tomato with a wild altitudinal distribution," *Functional Plant Biology*, vol. 5, no. 2, pp. 609–617, 1978.

[5] M. R. Foolad and G. Y. Lin, "Relationships between cold- and salt-tolerance during seed germination in tomato: germplasm evaluation," *Plant Breeding*, vol. 118, no. 1, pp. 45–48, 1999.

[6] M. A. Walker, D. M. Smith, K. P. Pauls, and B. D. McKersie, "A chlorophyll fluorescence screening test to evaluate chilling tolerance in tomato," *HortScience*, vol. 25, no. 3, pp. 334–339, 1990.

[7] M. R. Foolad, F. Q. Chen, and G. Y. Lin, "RFLP mapping of QTLs conferring cold tolerance during seed germination in an interspecific cross of tomato," *Molecular Breeding*, vol. 4, no. 6, pp. 519–529, 1998.

[8] M. R. Foolad, "Genetics of salt and cold tolerance in tomato: quantitative analysis and QTL mapping," *Plant Biotechnology*, vol. 16, no. 1, pp. 55–64, 1999.

[9] M. J. Truco, L. B. Randall, A. J. Bloom, and D. A. S. Clair, "Detection of QTLs associated with shoot wilting and root ammonium uptake under chilling temperatures in an interspecific backcross population from *Lycopersicon esculentum* × *L. hirsutum*," *Theoretical and Applied Genetics*, vol. 101, no. 7, pp. 1082–1092, 2000.

[10] B. Liu, Y. C. Du, X. X. Wang et al., "QTL analysis of cold tolerance from *Solanum pimpinellifolium* during seed germination and seedling stages using advanced backcross population," *Acta Horticulturae Sinica*, vol. 37, no. 7, pp. 1093–1101, 2010.

[11] E. M. Arms, A. J. Bloom, and D. A. St Clair, "High-resolution mapping of a major effect QTL from wild tomato *Solanum habrochaites* that influences water relations under root chilling," *Theoretical and Applied Genetics*, vol. 128, no. 9, pp. 1713–1724, 2015.

[12] 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.

[13] J. Zhang, W. Guo, and T. Zhang, "Molecular linkage map of allotetraploid cotton (*Gossypium hirsutum* L. × *Gossypium barbadense* L.) with a haploid population," *Theoretical and Applied Genetics*, vol. 105, no. 8, pp. 1166–1174, 2002.

[14] S. R. McCouch, Y. G. Cho, M. Yano et al., "Report on QTL nomenclature," *Rice Genetics Newsletter*, vol. 14, no. 11, pp. 11–13, 1997.

[15] B. Liu, *Genetic Introgression and QTL Analysis for Cold Tolerance in Wild Tomato (S. pimpinellifolium)*, Chinese Academy of Agricultural Sciences, 2009.

[16] Q. L. Dang, Y. U. Chao, and Z. L. Wang, "Studies on the chilling resistance and physiological index in processing tomatoes after treating seeds by ABA," *Journal of Shihezi University*, vol. 23, no. 3, pp. 349–351, 2005.

[17] M. R. Foolad, "Genome mapping and molecular breeding of tomato," *International Journal of Plant Genomics*, vol. 2007, Article ID 64358, 52 pages, 2007.

[18] M. R. Foolad, P. Subbiah, and L. Zhang, "Common QTL affect the rate of tomato seed germination under different stress and nonstress conditions," *International Journal of Plant Genomics*, vol. 2007, Article ID 97386, 10 pages, 2007.

[19] S. Doganlar, A. Frary, H.-M. Ku, and S. D. Tanksley, "Mapping quantitative trait loci in inbred backcross lines of *Lycopersicon pimpinellifolium* (LA1589)," *Genome*, vol. 45, no. 6, pp. 1189–1202, 2002.

[20] C. E. Vallejos and S. D. Tanksley, "Segregation of isozyme markers and cold tolerance in an interspecific backcross of tomato," *Theoretical and Applied Genetics*, vol. 66, no. 3-4, pp. 241–247, 1983.

[21] F. J. Goodstal, G. R. Kohler, L. B. Randall, A. J. Bloom, and D. A. S. Clair, "A major QTL introgressed from wild *Lycopersicon hirsutum* confers chilling tolerance to cultivated tomato (*Lycopersicon esculentum*)," *Theoretical and Applied Genetics*, vol. 111, no. 5, pp. 898–905, 2005.

[22] M. R. Foolad, G. Y. Lin, and F. Q. Chen, "Comparison of QTLs for seed germination under non-stress, cold stress and salt stress in tomato," *Plant Breeding*, vol. 118, no. 2, pp. 167–173, 1999.

[23] F. K. Zhao, X. H. Gao, J. H. Cheng, S. G. Fang, and J. H. Gu, "Analysis of chilling resistance of tomato cross and backcross progeny," *Journal of Changjiang Vegetable*, vol. 2, pp. 28–29, 2001.

[24] M. R. Foolad and G. Y. Lin, "Genetic analysis of cold tolerance during vegetative growth in tomato, *Lycopersicon esculentum* Mill," *Euphytica*, vol. 122, no. 1, pp. 105–111, 2001.

Hindawi Publishing Corporation
International Journal of Genomics
Volume 2016, Article ID 4235808, 9 pages
http://dx.doi.org/10.1155/2016/4235808

Research Article

Shotgun Quantitative Proteomic Analysis of Proteins Responding to Drought Stress in *Brassica rapa* L. (Inbred Line "Chiifu")

Soon-Wook Kwon,[1] Mijeong Kim,[2] Hijin Kim,[2] and Joohyun Lee[2]

[1]*Department of Plant Bioscience, Pusan National University, Milyang 627-706, Republic of Korea*
[2]*Department of Applied Bioscience, Konkuk University, Seoul 143-701, Republic of Korea*

Correspondence should be addressed to Joohyun Lee; joohyun00@gmail.com

Received 18 February 2016; Revised 13 May 2016; Accepted 18 May 2016

Academic Editor: Wenwei Xiong

Through a comparative shotgun quantitative proteomics analysis in *Brassica rapa* (inbred line Chiifu), total of 3,009 nonredundant proteins were identified with a false discovery rate of 0.01 in 3-week-old plants subjected to dehydration treatment for 0, 24, and 48 h, plants subjected to drought stress. Ribulose-bisphosphate carboxylases, chlorophyll a/b-binding protein, and light harvesting complex in photosystem II were highly abundant proteins in the leaves and accounted for 9%, 2%, and 4%, respectively, of the total identified proteins. Comparative analysis of the treatments enabled detection of 440 differentially expressed proteins during dehydration. The results of clustering analysis, gene ontology (GO) enrichment analysis, and analysis of composite expression profiles of functional categories for the differentially expressed proteins indicated that drought stress reduced the levels of proteins associated with photosynthesis and increased the levels of proteins involved in catabolic processes and stress responses. We observed enhanced expression of many proteins involved in osmotic stress responses and proteins with antioxidant activities. Based on previously reported molecular functions, we propose that the following five differentially expressed proteins could provide target genes for engineering drought resistance in plants: annexin, phospholipase D delta, sDNA-binding transcriptional regulator, auxin-responsive GH3 family protein, and TRAF-like family protein.

1. Introduction

Drought is a widespread environmental stress that is becoming increasingly problematic for agriculture due to the effects of climate change. Drought is caused by continuous shortages in water supply due to altered precipitation patterns in cropped areas [1, 2]. In general, drought stress causes 40% of global crop yield losses annually [3] and inhibits plant growth and development [4] by reducing root expansion, root development, leaf size, and seed development [5, 6]. Water deficit directly affects photosynthesis by altering the photosynthetic systems and reducing CO_2 availability [7]. Plant CO_2 assimilation is reduced by stomatal closure, damaged thylakoid membranes, and disrupted activity of enzymes involved in CO_2 fixation and adenosine triphosphate synthesis [5]. Drought stress affects ribulose-1,5-bisphosphate (RuBP) regeneration or reduces the levels of functional RuBP, which limits photosynthesis [8]. Secondary effects of drought stress include oxidative stress, which is toxic for aerobic metabolism [9].

Plants exhibit complex responses to drought stress, including changes in chloroplast metabolism and gene expression. Drought stress inhibits photosynthetic activity and causes an imbalance between light capture and light utilization [10]. Drought-mediated alterations in leaf photochemistry and photosynthetic electron transport generate potentially dangerous active oxygen species [11] and superoxide radicals [12]. Abscisic acid (ABA), a plant hormone that functions in plant growth and development, has two important roles in water stress, including regulating cellular water status to protect cell systems and inducing genes that express dehydration tolerant proteins [4]. Genetics and

breeding studies suggest that the pattern of dehydration tolerance inheritance may reflect that it is conferred by quantitative trait loci (QTL). Plant stress response is a dynamic process and the drought stress response is complex in plants, thus investigation of changes in gene expression in genomic level possibly to reveal the global snapshot of the response. The development of methods to monitor gene expression at the genomic level has enabled transcriptomic and proteomic analyses of plant cell responses under stress conditions. Proteins are key components of cellular structure. Shotgun proteomics analysis can provide direct functional information by exploring broad cellular expression patterns of proteins responding to environmental or extracellular stimuli [13].

Severe drought stress can be lethal for leaf vegetables such as *B. rapa*, which is widely produced in Asia. The whole genome of Chinese cabbage (inbred line Chiifu) was sequenced [14]. This genome sequencing database provided a reference and promoted transcriptomic and proteomic studies of other genomes. Yu et al. performed tag sequencing on *B. rapa* L. ssp. *pekinensis* and identified 1,092 drought-responsive genes; 37 genes were transcription factors, 28 were involved in signal transduction, and 61 were water-sensing and osmosensing responsive genes [15]. Microarray analysis of seedlings of *B. rapa* L. ssp. *pekinensis* subjected to 48 h of drought treatment determined that 738 genes (including 56 transcription factors) were differentially expressed in response to drought [16]. The identification of abiotic stress-responsive genes using DNA and RNA analytical tools is ongoing. However, proteomic analysis of *B. rapa* subjected to abiotic stress has not been sufficiently investigated.

Gel-LC/MSMS approach is a one shotgun proteomic approach where bottom up protein identification is performed from the protein mixture [17, 18]. In Gel-LC/MSMS, proteins are first separated using 1D SDS-PAGE or IPG-IEF and then digested into peptides from the divided gel pieces. These peptides are analyzed through mass spectrometry (MSMS) combined with high-pressure liquid chromatography (HPLC) [17]. Proteins detected by MS-MS are identified by comparison with protein databases [19, 20]. Shotgun proteomic analysis is usually performed for qualitative analysis, but spectral count normalization enables relative quantitative analysis [21]. The normalization of spectral counts is a label-free approach, which utilizes the mass signal strength of the sample's proteins and the spectral counts of the protein [22].

We investigated the effects of drought on *Brassica rapa* at the protein level using shotgun proteomic analysis. Nonredundant proteins were detected in *B. rapa* seedling extracts at 0, 24, and 48 h after the start of dehydration. The relative levels and patterns of proteins were compared to provide insights into changes in proteins in response to drought and to identify candidate proteins that could confer drought resistance to other plants.

2. Materials and Methods

2.1. Plant Material and Drought Stress Treatment. *Brassica rapa* ssp. pekinensis (inbred line Chiifu) seeds were germinated in water, and then one seedling was transplanted into soil in one pot (90 mm id × 90 mm). *Brassica rapa* plants were grown for 3 weeks in a growth chamber at 25°C (16 h day/8 h night, 40–70% relative humidity) with sufficient water supply until 1 day before drought treatment. Drought treatment was performed by removing the plastic pot from the plant root mass and exposing the soil to air [16]. This drought treatment proceeded for 24 and 48 h, at which times the whole plant except root tissue was harvested. Three-week-old plants without any drought treatment served as control (or 0 h drought treatment). Triplicate biological replicates were analyzed for all the three time points (0, 24, and 48 h) for statistical analysis.

2.2. Protein Extraction and Trichloroacetic Acid/Acetone Precipitation. The harvested plant tissue was ground in liquid nitrogen, and proteins were extracted from the ground tissue powder by adding extraction buffer (8 M urea, 5 mM dithiothreitol (DTT), 1% lithium dodecyl sulfate (LDS), and 100 mM Tris, pH 8.5). The suspension was incubated at room temperature for 30 min with vortexing, followed by centrifugation at 14,000 ×g for 15 min. The supernatant was recovered, and extracted proteins were precipitated overnight with 20% (v/v) trichloroacetic acid (TCA), washed several times with cold acetone to remove chlorophyll, and resolubilized in 8 M urea/Tris-HCl, pH 8.5. Sample protein concentrations were determined using the 2D-Protein Quant Kit (GE Healthcare, Piscataway, NJ, USA).

2.3. One-Dimensional LDS-PAGE and In-Gel Trypsin Digestion. Fifty micrograms of protein samples was prepared with NuPage® LDS Sample Buffer (Invitrogen, Carlsbad, CA, USA). Proteins were separated on a 4–12% NuPAGE Novex Bis-Tris gel (Invitrogen, Carlsbad, CA, USA) and stained with Coomassie Blue G-250 (Invitrogen). After the proteins were resolved on the gel, each sample gel lane was cut out and divided into seven equal-sized pieces, and proteins were in-gel digested with trypsin using the method of Shevchenko et al. [23].

2.4. LC MS/MS Analysis. A nanoflow HPLC instrument (Easy nLC, Thermo Fisher Scientific, Waltham, MA, USA) was coupled online to a Q Exactive Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA, USA). Analytical columns (12 cm, 75 μm inner diameter) were packed in-house with Alltima C18-AQ 5 μm resin. Reverse-phase chromatography was performed with a binary buffer system consisting of 0.1% formic acid (buffer A) and acetonitrile in 0.1% formic acid (buffer B). The sample was separated with a linear gradient of 3–60% buffer B at a flow rate of 250 nL/min. The total run time for the LC MS/MS was 110 min. MS data were acquired using a data-dependent top 8 method and dynamically choosing the most abundant precursor ions from the survey scan (300–2,000 Da) for higher-energy collisional dissociation (HCD) fragmentation. Dynamic exclusion duration was 60 s, and the precursor isolation window was performed with four. Survey scans were acquired at a resolution of 70,000 at m/z 200. The resolution for HCD spectra was set to 17,500 at m/z 200.

(a) (b) (c)

Figure 1: Response of *B. rapa* seedling during drought stress. (a) Control, (b) 24 h drought treatment, and (c) 48 h drought treatment.

2.5. Analysis of Proteomic Data. Proteome Discoverer (version 1.3) software (Thermo Fisher Scientific) was used for protein identification and spectral count acquisition for each identified protein. The fragmentation spectra were searched against the *Brassica rapa* (Brassica V 1.2) protein database with precursor and fragment mass tolerances set to 10 ppm and 0.8 Da, respectively, and with up to two missed cleavages. Cysteine carbamidomethylation was set as a fixed modification, and methionine oxidation was set as a variable modification for database searching. Both peptide and protein identifications were filtered at 1% false discovery rate which were evaluated through decoy database which was created by reversing all of the *B. rapa* protein sequences. In addition the proteins identified with only one spectral count (SC) were discarded.

2.6. Comparative Analysis of Relative Protein Abundances. The Proteome Discoverer output was exported to Microsoft Excel to calculate normalized spectral counts (NSpC) [24–26]. The NSpC for each protein k is given by

$$(\mathrm{NSpC})_k = \frac{(\mathrm{SpC}/L)_k}{\sum_{i=1}^{n} (\mathrm{SpC}/L)_i}, \quad (1)$$

where the total number of MS/MS spectra matching peptides from protein k (SpC) is divided by the protein's length (L) and then divided by SpC/L for all N proteins in the experiment.

2.7. Bioinformatics Analysis. Gene ontology (GO) annotations of *Brassica rapa* proteins were retrieved from BRAD *B. rapa* genome data V 1.2. GO enrichment analysis was performed in agriGO (http://bioinfo.cau.edu.cn/agriGO/) with customized parameters using the *B. rapa* whole genome as the background/reference. The clustering analyses were conducted with Genesis software [27] using centered correlation, and the average linkage procedure and tree were visualized with the same software. Composite expression profile analysis was performed by summing averages of NSpC for all proteins of a given functional category at each of the three time points of drought treatment.

3. Results and Discussion

3.1. Phenotypic Changes during Drought Stress. *Brassica rapa* 3-week-old plants were subjected to drought stress for 0, 24, and 48 h (Figure 1). All leaves were withered on plants that were dehydrated for 24 h; the lower leaves were severely damaged so that few of the lower leaves were folded. Some leaves displayed slight chlorosis at the leaf margin. Similar effects were observed in plants subjected to dehydration for 48 h; all leaves were severely withered and folded, and slight chlorosis at the leaf margin was observed, similarly to that of plants dehydrated for 24 h. These morphological changes suggest that the drought treatment was effective, and the analyzed proteomes for plants subjected to 0 (control), 24, and 48 h of drought stress represent plant responses to normal conditions, mild drought stress, and severe drought stress, respectively.

3.2. Identification of Total Proteins from Young Brassica rapa Plants. Shotgun proteomic analysis was used to identify 3,009 nonredundant proteins (Supplementary Table S1 (see Supplementary Material available online at http://dx.doi.org/10.1155/2016/4235808)) with a false discovery rate of 0.01 after 0, 24, and 48 h of drought stress with three replicates. Shotgun proteomics analysis inherently contains a level of analytical incompleteness; the number of identified proteins for each sample ranged from 1,446 to 1,819 (Supplementary Table S2) [28]. The distributions of molecular weights (MWs) and pI values (Figure 2) for the identified *B. rapa* proteomes were compared with those of proteins encoded by the *Brassica* genome. The MWs of the identified proteins ranged from 5.5 kDa (Bra001019, unknown protein) to 534.5 kDa (Bra039167, the auxin transport protein BIG). The overall MW distribution of the identified proteins was similar to that of the *B. rapa* genome; however, the proportion of

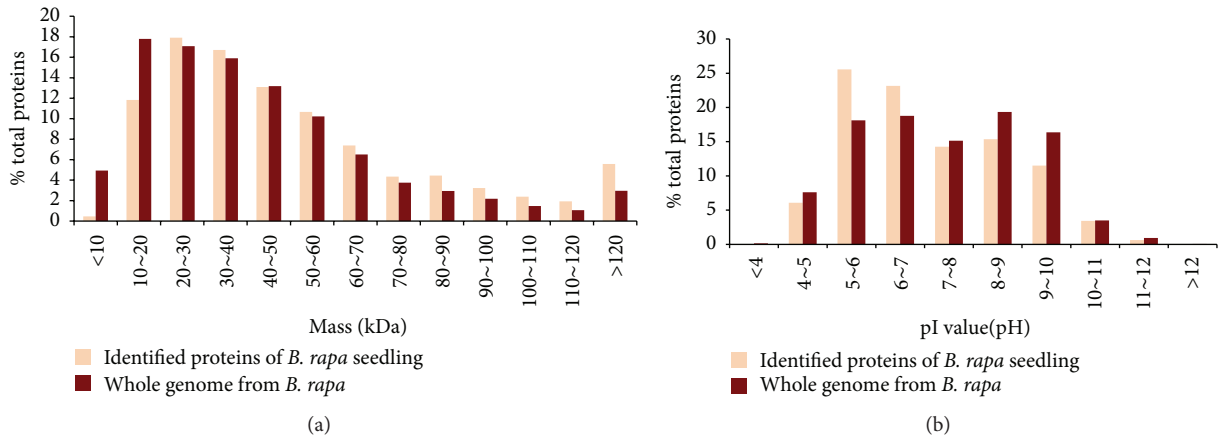


Figure 2: Distribution of molecular weight (a) and pI value (b) of proteins from *B. rapa* seedling (brown) and those of proteins encoded by the whole genome from *B. rapa* (orange).

proteins less than 20 kDa was quite low and the proportion of large MW proteins was relatively high. This is possibly due to the fact that the low MW proteins contained a small number of amino acids and produced a small number of trypsin-digested peptides. Therefore, the chance that these small peptides were detected by mass spectrometry was relatively low. For the protein pI values, the lowest pH was 4.27 (Bra002368, calmodulin 5) and the highest pH was 12.2 (Bra035628, ribosomal protein S30 family protein) for identified proteins. Proteins in the range of pH 7 to pH 10 were slightly less abundant in the *B. rapa* genome. This is possibly due to technical difficulties in solubilizing basic proteins using our protein extraction method. However, more than 45.2% of the identified proteins were basic proteins (pI higher than pH 7), and the overall distribution of pI values of identified proteins was similar to that of the genomic proteins. This suggested that the *B. rapa* proteome identified by our shotgun proteomic analysis was not biased. The distribution of the identified proteins was similar to *B. rapa*. This unbiased protein identification indicated that shotgun proteomic analysis was suitable for evaluating global protein expression patterns in *B. rapa*.

We performed relative quantification of the identified *B. rapa* proteins using spectral count normalization. The most abundant proteins were ribulose-bisphosphate carboxylases (RuBisCOs; Bra028181, Bra028087, Bra028406, Bra025431, Bra041116, and Bra034028), chlorophyll a/b-binding proteins (Bra010807 and Bra030182), and light harvesting complex in photosystem II (Bra039070, Bra037913, Bra013183, Bra029732, Bra000708, Bra028906, Bra004989, and Bra026099). RuBisCO is involved in carbon fixation and is the most abundant protein in nature. Chlorophyll a/b-binding proteins have a role in light harvesting in the thylakoid in cooperation with the light harvesting complex in photosystem II [29, 30]. These three proteins have the highest expression in plant leaves. The relative amounts of RuBisCO, chlorophyll a/b-binding protein, and light harvesting complex in photosystem II were 9%, 2%, and 4% of the total protein. RuBisCO accounts for 30–50% of total plant protein in green tissues [31, 32]. The relative proportion of RuBisCO in our analyses was 9% and we detected the chlorophyll a/b-binding protein and the light harvesting complex in photosystem II accounted for 6% of total proteins. This possibly reflects the leaf tissues analyzed.

3.3. Identification of Differentially Expressed Proteins during Drought Stress. A total of 3,009 proteins were not reproducibly detected in all Gel-LC/MSMS analyses even among the replicates for the same treatment due to the phenomenon of analytical incompleteness. The randomly detected proteins can be misjudged as differentially expressed proteins; therefore, we excluded randomly detected proteins from comparative analysis. All proteins that were detected in all treatments and replicates were subjected to comparison analysis. For the remaining proteins that were not detected in all Gel-LC/MSMS runs, proteins were included in the comparison analysis if they were detected in all the three replicates for a certain treatment. Using this criterion, 1,567 proteins were subjected to comparative analysis. The relative protein quantity was estimated with SCs (refer to Section 2). The SCs for these 1,567 proteins were globally normalized (NSpC) followed by logarithmic transformation (natural log(Ln) of NSpC) to conduct ANOVA (Supplementary Table S3). For the missing SC data, an arbitrary value of 0.1 was assigned to distinguish them with minimal SC of 2. The proteins identified with one SC were excluded in this study because the possible random identification even under 0.01 false positive cutoff. Thus the minimum difference of spectral count was 6 for the proteins were not detected one of the condition and judged as differentially expressed proteins. To evaluate the reproducibility of the 1,567 proteins, the coefficient of determination (R^2) between NSpCs for the biological replicates was estimated. The average R^2 was 0.93 and ranged from 0.88 to 0.95. This result suggested that the relative quantities of the 1,567 proteins were reproducible in the biological replicates. The comparative analysis detected 440 differentially expressed proteins (Supplementary Table S4) in 3-week-old plants during drought stress.

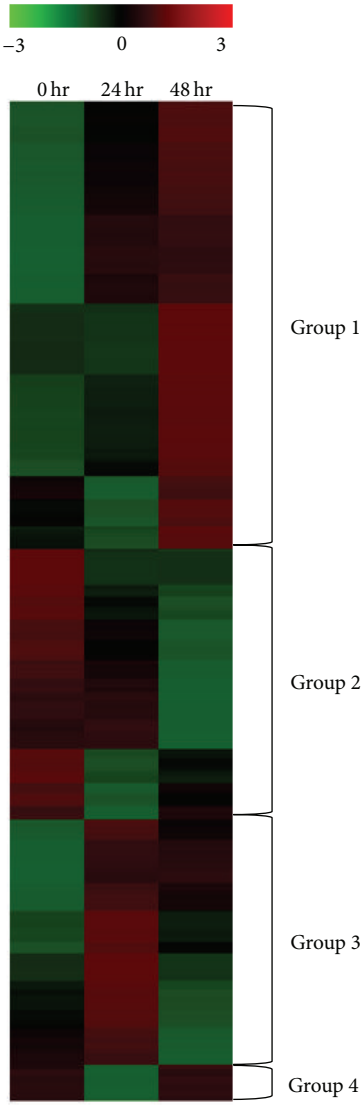


Figure 3: Clustering analysis for the 440 differentially expressed proteins.

3.4. Clustering Analysis. The clustering analysis was conducted with 440 differentially expressed proteins, which were arbitrarily grouped into four groups based on similar expression patterns (Figure 3, Supplementary Table S5). Proteins in Group 1 increased steadily and showed highest expression at 48 h of drought treatment. A total of 197 proteins belonged to Group 1. The GO enrichment analysis of Group 1 proteins revealed that 85 GO terms (51 biological processes, 13 molecular functions, and 21 cellular components) (Supplementary Table S6) were enriched, suggesting that numerous metabolic changes occurred in response to drought stress. Many GO terms associated with catabolic processes were enriched. This may represent altered metabolism in drought-stressed plant cells, in which nutrients were not properly provided through photosynthesis, and catabolism was increased as a counteraction. Some proteins in this group were previously reported to be involved in resistance to various stresses. Twelve of the enriched GO terms were responses to various stresses, suggesting that these protein levels increased to respond to the drought stress, although their expression did not prevent leaf wilting at 48 h. Group 1 included proteins associated with autoxidation such as glutathione S-transferase and peroxidase, proteins associated with molecular chaperonin such as heat shock protein and late embryogenesis abundant proteins, and proteins associated with defense to biotic stress such as chitinase and PR proteins. Proteins previously reported as candidate genes for plant stress responses in other plant species also were detected in Group 1. Three annexins (Bra008737, Bra034402, and Bra036764) were detected. Annexins belong to a multigene family of Ca^{2+}-dependent phospholipid- and cytoskeletal-binding proteins. Annexins are upregulated under various stress conditions. Overexpression of the mustard annexin (AnnBj1) in transgenic cotton plants conferred tolerance to various abiotic stresses such as sodium chloride, mannitol, polyethylene glycol, and hydrogen peroxide. In *Arabidopsis*, annexin had an important role in resistance to osmotic stress [33]. We observed that Bra008737 and Bra034402 levels increased rapidly during drought stress. A role for annexin has not been reported in *B. rapa*. The previously reported annexin functions associated with osmotic stress and drought tolerance in transgenic plants of other crop species suggest a possible role of annexin in drought stress tolerance in *B. rapa*. Group 1 also contains phospholipase D delta (Bra017730), which is involved in phospholipid metabolism responses to drought and salinity stress [34]. Cell membranes are major targets of environmental stresses, and lipids have crucial roles in preserving cell compartments under water stress conditions. In *Arabidopsis*, phospholipase D delta mRNA accumulated in response to dehydration and high salt stress [34, 35], suggesting a possible role of phospholipase D delta in drought tolerance through phospholipid metabolism.

The expression of 119 proteins in Group 2 steadily declined during drought treatment. GO enrichment analysis of Group 2 revealed enrichment of 90 GO terms (42 biological processes, 7 molecular functions, and 41 cellular components) (Supplementary Table S6). Among these terms, 34 were associated with photosynthesis, such as chloroplast organelle, photosynthesis dark and light reactions, and chlorophyll. The expression patterns of proteins in Group 2 primarily represent cell damage from drought stress. Drought stress reduces photosynthesis, and we observed that proteins associated with photosynthesis also were decreased. Proteins related to chlorophyll synthesis such as phytoene synthase (Bra006391), phytoene desaturase (Bra010751), and protochlorophyllide oxidoreductase (Bra026349) were downregulated during drought stress. Phytoene synthase catalysis generates two geranylgeranyl pyrophosphates. Phytoene synthase and protochlorophyllide oxidoreductase are related to carotenoid and ABA metabolism as well as chlorophyll [36–38]. Protochlorophyllide oxidoreductase catalyzes the phototransformation of protochlorophyllide to chlorophyllide [39]. Several ribosomal proteins were detected in Group 2, which also declined during drought stress. These reductions in proteins associated with photosynthesis and ribosomal proteins represent the cellular damage caused by drought stress. However, the auxin-responsive GH3 family protein (Bra006196) identified

in Group 2 may be involved in resistance to drought stress, because it is involved in drought stress signaling pathways [40]. GH3 disrupts the conjugation of the phytohormone auxin (IAA) and amino acids by degrading free IAA and negatively regulating auxin homeostasis [41, 42]. Transgenic rice plants overexpressing GH3 were dwarf and had larger stomatal apertures, which were susceptible to water loss [43]. Thus, a reduction of GH3 in *B. rapa* may function to minimize water loss.

The 108 proteins in Group 3 had the highest expression levels at 24 h after the start of dehydration. The expression levels of some proteins in Group 3 were lower at 24 h compared to those at 0 h. However, the levels of all proteins in this group were enhanced by drought stress, and the expression levels of many proteins were higher at 42 h than at 0 h. Thus, the response of Group 3 is similar to that of Group 1, and the GO enrichment analysis was similar to that of Group 1. GO enrichment analysis of Group 3 revealed that 110 GO terms were enriched (69 biological processes, 16 molecular functions, and 25 cellular components) (Supplementary Table S6). GO terms associated with responses to various stresses and catabolic processes were enriched. Several antioxidant proteins were detected, including glutathione S-transferase F3 (Bra000311), peroxidase superfamily protein (Bra013576), stromal ascorbate peroxidase (Bra037859), manganese superoxide dismutase 1 (Bra029879), and copper/zinc superoxide dismutase 2 (Bra034394). These results for Groups 1 and 3 suggest that drought stress causes oxidative stress. Group 3 contains two candidate genes for resistance to drought stress. The level of ssDNA-binding transcriptional regulator (Bra024809) increased at 24 h. Gene expression modulation is an important cellular response to external stimuli, and transcriptional regulators have crucial roles in controlling transcription from specific genes in response to specific stimuli [44]. The target gene of ssDNA-binding transcriptional regulator is not known. However, the early increase in response to drought stress may suggest a role in activating gene(s) involved in drought resistance. Increased tumor necrosis factor receptor-associated factor- (TRAF-) like family protein (Bra020541) also was detected. TRAF-like family proteins are known to function in proteasome-mediated regulation during various developmental processes [45]. A knockout mutant in *Arabidopsis* of TRAF-like family protein had greater susceptibility to drought stress, and the TRAF-like family protein seven in absentia 2 (SINA2) promotes drought tolerance in an ABA-dependent manner in *Arabidopsis* [46].

The levels of 16 proteins in Group 4 were reduced at 24 h and recovered at 48 h. Only one GO term was enriched, the term for translation. Ribosomal proteins, translational initiation factor, and release factor are in Group 4. The reduced ribosomal protein levels at 24 h resulted from drought stress-mediated damage, similar to that observed in Group 2. However, the recovered expression levels at 48 h for proteins in Groups 2 and 4 were unexpected.

3.5. Composite Expression Profile of Functional Categories during Drought Treatment. To characterize global protein expression patterns involved in specific processes, expression graphs were represented by summing NSpC for each protein in each functional category according to the functional catalog by Bevan et al. [47] (Figure 4). The results of GO enrichment analysis for each clustered group indicate that functional categories of chloroplast, catabolic process, response to abiotic stimulus, response to osmotic stress, and antioxidant activity were representative of the molecular changes that occurred during drought stress in *B. rapa*. The composite expression profiles of those five representative functional GO categories for all 440 differentially expressed proteins were analyzed. The results of these composite analyses were consistent with the results of the GO enrichment analysis for the individual groups. The composite expression patterns of proteins in chloroplast categories continuously declined, whereas those of the catabolic process category continuously increased. The responses of these two functional categories revealed the responses to drought stress-mediated cellular damage. For the response to abiotic stimulus, protein expression rapidly increased and remained high in response to drought stress treatment, although protein expression slightly declined at 48 h compared with that at 24 h. Proteins involved in response to osmotic stress and antioxidant activity were continuously expressed during drought stress. The temporal expression patterns of these proteins suggest that they could be evaluated as candidate proteins for drought tolerance.

4. Conclusion

We performed shotgun proteomic analysis and identified 3,009 nonredundant proteins in young *B. rapa* plants. RuBisCOs, chlorophyll a/b-binding protein, and light harvesting complex in photosystem II were abundantly expressed in *B. rapa* leaf tissues.

We compared the relative abundance of 1,567 reproducibly detected proteins and identified 440 proteins that were differentially expressed in response to drought stress. Our quantitative proteomics study using clustering analysis, GO enrichment analysis, and composite expression profiles of functional categories with the 440 differentially expressed proteins provides comprehensive molecular insights into protein level changes and modifications induced by drought stress. The representative responses to drought stress included a reduction in proteins associated with photosynthesis and an increase in proteins responding to various stresses. The expression of proteins involved in antioxidant activities increased during drought stress. These proteins likely have important roles in the removal of active oxygen species produced during drought stress. The *B. rapa* inbred line "Chiifu" is not drought tolerant and did not show a drought resistant morphology during 48 h of the experiment. However, we detected the induction of many proteins involved in abiotic stress responses, including osmotic stress, and proteins involved in antioxidant reactions. We propose that annexin, phospholipase D delta, ssDNA-binding transcriptional regulator, auxin-responsive GH3 family protein, and TRAF-like family protein are candidate genes for engineering drought tolerance or drought resistance in *B. rapa* based on their expression patterns under drought stress and previously reported molecular functions.

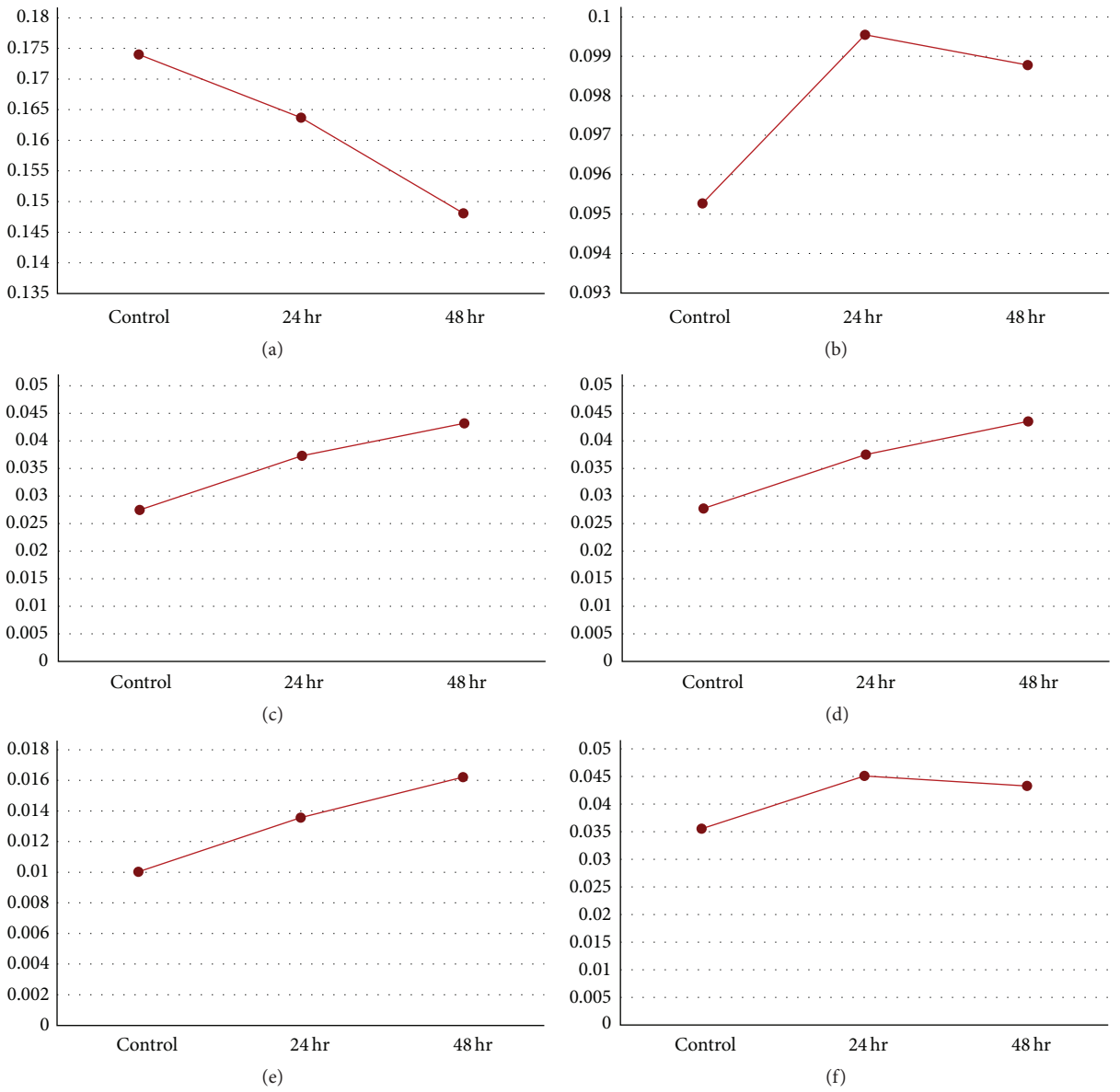


Figure 4: Composite protein expression patterns of gene function categories. (a) Chloroplast, (b) response to abiotic stimulus, (c) response to salt stress, (d) response to osmotic stress, (e) antioxidant activity, and (f) catabolic process.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This paper was supported by Konkuk University in 2013.

References

[1] D. Z. Habash, Z. Kehel, and M. Nachit, "Genomic approaches for designing durum wheat ready for climate change with a focus on drought," *Journal of Experimental Botany*, vol. 60, no. 10, pp. 2805–2815, 2009.

[2] R. R. Mir, M. Zaman-Allah, N. Sreenivasulu, R. Trethowan, and R. K. Varshney, "Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops," *Theoretical and Applied Genetics*, vol. 125, no. 4, pp. 625–645, 2012.

[3] J. S. Boyer, "Plant productivity and environment," *Science*, vol. 218, no. 4571, pp. 443–448, 1982.

[4] J.-K. Zhu, "Salt and drought stress signal transduction in plants," *Annual Review of Plant Biology*, vol. 53, pp. 247–273, 2002.

[5] M. Farooq, A. Wahid, N. Kobayashi, D. Fujita, and S. M. A. Basra, "Plant drought stress: effects, mechanisms and management," in *Sustainable Agriculture*, pp. 153–188, Springer, Amsterdam, The Netherlands, 2009.

[6] E. A. Bray, "Plant responses to water deficit," *Trends in Plant Science*, vol. 2, no. 2, pp. 48–54, 1997.

[7] M. M. Chaves, "Effects of water deficits on carbon assimilation," *Journal of Experimental Botany*, vol. 42, no. 1, pp. 1–16, 1991.

[8] C. Gimenez, V. J. Mitchell, and D. W. Lawlor, "Regulation of photosynthetic rate of two sunflower hybrids under water stress," *Plant Physiology*, vol. 98, no. 2, pp. 516–524, 1992.

[9] R. Mittler, "Oxidative stress, antioxidants and stress tolerance," *Trends in Plant Science*, vol. 7, no. 9, pp. 405–410, 2002.

[10] C. H. Foyer and G. Noctor, "Tansley review no. 112," *New Phytologist*, vol. 146, no. 3, pp. 359–388, 2000.

[11] D. Peltzer, E. Dreyer, and A. Polle, "Differential temperature dependencies of antioxidative enzymes in two contrasting species: *Fagus sylvatica* and *Coleus blumei*," *Plant Physiology and Biochemistry*, vol. 40, no. 2, pp. 141–150, 2002.

[12] A. R. Reddy, K. V. Chaitanya, and M. Vivekanandan, "Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants," *Journal of Plant Physiology*, vol. 161, no. 11, pp. 1189–1202, 2004.

[13] R. Rakwal and G. K. Agrawal, "Rice proteomics: Current status and future perspective," *Electrophoresis*, vol. 24, no. 19-20, pp. 3378–3389, 2003.

[14] J.-H. Mun, B.-S. Park, R. Schmidt, I. Bancroft, I. BancroftR, and I. Bancrofts, "Sequencing the gene space of brassica rapa," in *Genetics and Genomics of the Brassicaceae*, pp. 413–435, Springer, New York, NY, USA, 2011.

[15] S. Yu, F. Zhang, Y. Yu, D. Zhang, X. Zhao, and W. Wang, "Transcriptome profiling of dehydration stress in the Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*) by tag sequencing," *Plant Molecular Biology Reporter*, vol. 30, no. 1, pp. 17–28, 2012.

[16] S.-C. Lee, M.-H. Lim, J. A. Kim et al., "Transcriptome analysis in Brassica rapa under the abiotic stresses using Brassica 24K oligo microarray," *Molecules and Cells*, vol. 26, no. 6, pp. 595–605, 2008.

[17] J. Lee, W. M. Garrett, and B. Cooper, "Shotgun proteomic analysis of *Arabidopsis thaliana* leaves," *Journal of Separation Science*, vol. 30, no. 14, pp. 2225–2230, 2007.

[18] Y. Zhang, B. R. Fonslow, B. Shan, M.-C. Baek, and J. R. Yates III, "Protein analysis by shotgun/bottom-up proteomics," *Chemical Reviews*, vol. 113, no. 4, pp. 2343–2394, 2013.

[19] D. A. Wolters, M. P. Washburn, and J. R. Yates III, "An automated multidimensional protein identification technology for shotgun proteomics," *Analytical Chemistry*, vol. 73, no. 23, pp. 5683–5690, 2001.

[20] H. Liu, R. G. Sadygov, and J. R. Yates, "A model for random sampling and estimation of relative protein abundance in shotgun proteomics," *Analytical Chemistry*, vol. 76, no. 14, pp. 4193–4201, 2004.

[21] B. Zybailov, M. K. Coleman, L. Florens, and M. P. Washburn, "Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling," *Analytical Chemistry*, vol. 77, no. 19, pp. 6218–6224, 2005.

[22] W. Zhu, J. W. Smith, and C.-M. Huang, "Mass spectrometry-based label-free quantitative proteomics," *Journal of Biomedicine and Biotechnology*, vol. 2010, Article ID 840518, 6 pages, 2010.

[23] A. Shevchenko, H. Tomas, J. Havliš, J. V. Olsen, and M. Mann, "In-gel digestion for mass spectrometric characterization of proteins and proteomes," *Nature Protocols*, vol. 1, no. 6, pp. 2856–2860, 2007.

[24] A. C. Paoletti, T. J. Parmely, C. Tomomori-Sato et al., "Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 50, pp. 18928–18933, 2006.

[25] B. Zybailov, A. L. Mosley, M. E. Sardiu, M. K. Coleman, L. Florens, and M. P. Washburn, "Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*," *Journal of Proteome Research*, vol. 5, no. 9, pp. 2339–2347, 2006.

[26] N. M. Griffin, J. Yu, F. Long et al., "Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis," *Nature Biotechnology*, vol. 28, no. 1, pp. 83–89, 2010.

[27] A. Sturn, J. Quackenbush, and Z. Trajanoski, "Genesis: cluster analysis of microarray data," *Bioinformatics*, vol. 18, no. 1, pp. 207–208, 2002.

[28] M. R. Wilkins, R. D. Appel, J. E. Van Eyk et al., "Guidelines for the next 10 years of proteomics," *Proteomics*, vol. 6, no. 1, pp. 4–8, 2006.

[29] H. Paulsen, "Chlorophyll a/b-binding proteins," *Photochemistry and Photobiology*, vol. 62, no. 3, pp. 367–382, 1995.

[30] H. Paulsen, B. Finkenzeller, and N. Kühlein, "Pigments induce folding of light-harvesting chlorophyll *a*/*b*-binding protein," *European Journal of Biochemistry*, vol. 215, no. 3, pp. 809–816, 1993.

[31] I. Widjaja, K. Naumann, U. Roth et al., "Combining subproteome enrichment and Rubisco depletion enables identification of low abundance proteins differentially regulated during plant defense," *Proteomics*, vol. 9, no. 1, pp. 138–147, 2009.

[32] H. B. Krishnan and S. S. Natarajan, "A rapid method for depletion of Rubisco from soybean (*Glycine max*) leaf for proteomic analysis of lower abundance proteins," *Phytochemistry*, vol. 70, no. 17-18, pp. 1958–1964, 2009.

[33] S. Lee, E. J. Lee, E. J. Yang et al., "Proteomic of identification of annexins, calcium-dependent membrane binding proteins that mediate osmotic stress and abscisic acid signal transduction in arabidopsis," *Plant Cell*, vol. 16, no. 6, pp. 1378–1391, 2004.

[34] T. Katagiri, S. Takahashi, and K. Shinozaki, "Involvement of a novel Arabidopsis phospholipase D, AtPLDδ, in dehydration-inducible accumulation of phosphatidic acid in stress signalling," *Plant Journal*, vol. 26, no. 6, pp. 595–605, 2001.

[35] A. Gigon, A.-R. Matos, D. Laffray, Y. Zuily-Fodil, and A.-T. Pham-Thi, "Effect of drought stress on lipid metabolism in the leaves of *Arabidopsis thaliana* (Ecotype Columbia)," *Annals of Botany*, vol. 94, no. 3, pp. 345–351, 2004.

[36] L. O. Lindgren, K. G. Stålberg, and A.-S. Höglund, "Seed-specific overexpression of an endogenous arabidopsis phytoene synthase gene results in delayed germination and increased levels of carotenoids, chlorophyll, and abscisic acid," *Plant Physiology*, vol. 132, no. 2, pp. 779–785, 2003.

[37] P. K. Burkhardt, P. Beyer, J. Wünn et al., "Transgenic rice (*Oryza sativa*) endosperm expressing daffodil (*Narcissus pseudonarcissus*) phytoene synthase accumulates phytoene, a key intermediate of provitamin A biosynthesis," *The Plant Journal*, vol. 11, no. 5, pp. 1071–1078, 1997.

[38] X. Ye and P. Beyer, "Engineering the provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm," *Science*, vol. 287, no. 5451, pp. 303–305, 2000.

[39] M. Ryberg and K. Dehesh, "Localization of NADPH-protochlorophyllide oxidoreductase in dark-grown wheat (Triticum aestivum) by immuno-electron microscopy before and after transformation of the prolamellar bodies," *Physiologia Plantarum*, vol. 66, no. 4, pp. 616–624, 1986.

[40] M. Seki, M. Narusaka, J. Ishida et al., "Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray," *The Plant Journal*, vol. 31, no. 3, pp. 279–292, 2002.

[41] R. Kumar, P. Agarwal, A. K. Tyagi, and A. K. Sharma, "Genome-wide investigation and expression analysis suggest diverse roles of auxin-responsive *GH3* genes during development and response to different stimuli in tomato (*Solanum lycopersicum*)," *Molecular Genetics and Genomics*, vol. 287, no. 3, pp. 221–235, 2012.

[42] P. E. Staswick, B. Serban, M. Rowe et al., "Characterization of an arabidopsis enzyme family that conjugates amino acids to indole-3-acetic acid," *Plant Cell*, vol. 17, no. 2, pp. 616–627, 2005.

[43] H. Du, N. Wu, J. Fu et al., "A GH3 family member, OsGH3-2, modulates auxin and abscisic acid levels and differentially affects drought and cold tolerance in rice," *Journal of Experimental Botany*, vol. 63, no. 18, pp. 6467–6480, 2012.

[44] D. Desveaux, J. Allard, N. Brisson, and J. Sygusch, "A new family of plant transcription factors displays a novel ssDNA-binding surface," *Nature Structural Biology*, vol. 9, no. 7, pp. 512–517, 2002.

[45] M. Wang, Y. Jin, J. Fu et al., "Genome-wide analysis of SINA family in plants and their phylogenetic relationships," *DNA Sequence*, vol. 19, no. 3, pp. 206–216, 2008.

[46] Y. Bao, C. Wang, C. Jiang et al., "The tumor necrosis factor receptor-associated factor (TRAF)-like family protein SEVEN IN ABSENTIA 2 (SINA2) promotes drought tolerance in an ABA-dependent manner in Arabidopsis," *New Phytologist*, vol. 202, no. 1, pp. 174–187, 2014.

[47] M. Bevan, I. Bancroft, E. Bent et al., "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana," *Nature*, vol. 391, no. 6666, pp. 485–488, 1998.