

Article

# Genome-Wide Identification and Comparative Analysis of the 3-Hydroxy-3-methylglutaryl Coenzyme A Reductase (HMGR) Gene Family in *Gossypium*

Wei Liu <sup>1,†</sup>, Zhiqiang Zhang <sup>1,†</sup>, Wei Li <sup>2,†</sup>, Wei Zhu <sup>1</sup>, Zhongying Ren <sup>2</sup>, Zhenyu Wang <sup>2</sup>, Lingli Li <sup>1</sup>, Lin Jia <sup>1</sup>, Shuijin Zhu <sup>3,\*</sup> and Zongbin Ma <sup>1,\*</sup>

<sup>1</sup> Collaborative Innovation Center of Henan Grain Crops/Agronomy College, Henan Agricultural University, Zhengzhou 450002, China; lw6051@163.com (W.L.); 13598838901@163.com (Z.Z.); zhuwei\_2006z@126.com (W.Z.); ndlll@126.com (L.L.); cathylin2012@163.com (L.J.)

<sup>2</sup> State Key Laboratory of Cotton Biology/Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang 455000, China; lw8887@163.com (W.L.); renzhongyingcotton@163.com (Z.R.); wangzhenyu89@126.com (Z.W.)

<sup>3</sup> Department of Agronomy, Zhejiang University, Hangzhou 310058, China

\* Correspondence: shjzhu@zju.edu.cn (S.Z.); zongbinma@163.com (Z.M.); Tel.: +86-13067922851 (S.Z.); +86-13623716660 (Z.M.)

† These authors contributed equally to this work.

Received: 13 December 2017; Accepted: 21 January 2018; Published: 24 January 2018

**Abstract:** Terpenes are the largest and most diverse class of secondary metabolites in plants and play a very important role in plant adaptation to environment. 3-Hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) is a rate-limiting enzyme in the process of terpene biosynthesis in the cytosol. Previous study found the *HMGR* genes underwent gene expansion in *Gossypium raimondii*, but the characteristics and evolution of the *HMGR* gene family in *Gossypium* genus are unclear. In this study, genome-wide identification and comparative study of *HMGR* gene family were carried out in three *Gossypium* species with genome sequences, i.e., *G. raimondii*, *Gossypium arboreum*, and *Gossypium hirsutum*. In total, nine, nine and 18 *HMGR* genes were identified in *G. raimondii*, *G. arboreum*, and *G. hirsutum*, respectively. The results indicated that the *HMGR* genes underwent gene expansion and a unique gene cluster containing four *HMGR* genes was found in all the three *Gossypium* species. The phylogenetic analysis suggested that the expansion of *HMGR* genes had occurred in their common ancestor. There was a pseudogene that had a 10-bp deletion resulting in a frameshift mutation and could not be translated into functional proteins in *G. arboreum* and the A-subgenome of *G. hirsutum*. The expression profiles of the two pseudogenes showed that they had tissue-specific expression. Additionally, the expression pattern of the pseudogene in the A-subgenome of *G. hirsutum* was similar to its paralogous gene in the D-subgenome of *G. hirsutum*. Our results provide useful information for understanding cytosolic terpene biosynthesis in *Gossypium* species.

**Keywords:** *Gossypium*; HMGR; terpene biosynthesis; gene expansion; pseudogene

## 1. Introduction

Terpenes are a type of natural compound, which are widely distributed in nature and have diverse structures and functions [1,2]. Thousands of terpenes and derivatives are a good example of metabolic plasticity that is essential to survive in changing environments [3,4]. Additionally, many terpenes are specialized compounds that are rich sources of commercial products widely used as flavors, fragrances and pharmaceuticals by humans [5,6].

In plant cells, terpenes are synthesized by two independent pathways: the mevalonate pathway (MVA pathway) in the cytosol and the 2-C-methyl-D-erythritol 4-phosphate pathway (MEP pathway) in the plastid [7,8]. The 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) catalyzes the conversion of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) to mevalonate (MVA), which is considered to be a rate-limiting enzyme of the MVA pathway and plays a key role in the biosynthesis of plant cytosolic terpenes [9,10]. Currently, *HMGR* genes have been isolated and cloned from many species of plants, such as *Arabidopsis thaliana* [11,12], rice [13], wheat [14], cotton [15], melon [16], medicinal plants *Cymbopogon winterianus*, [17] and *Alisma orientale* [18], and so on. Many experiments have shown that *HMGR* holds an important control point in the MVA pathway and genetic manipulation of *HMGR* indeed increased terpenes content in plants. The *HMGR* gene of *Hevea brasiliensis* was introduced into tobacco by *Agrobacterium* transformation, and then the activity of HMGR in transgenic plants increased by 4–8 times and the total amount of sterols increased by six times [19]. Ginsenosides are glycosylated triterpenoids, and overexpression of the *HMGR* gene in ginseng could significantly increase the amount of ginsenosides [20]. Transgenic spike lavender plants expressing the *Arabidopsis HMGR* gene accumulated more essential oil constituents, which were composed of monoterpenes and sesquiterpenes [21]. Additionally, more and more evidence shows that HMGR is not only critical for normal plant development but also very important in adapting to changing environments. HMGR was negatively regulated by protein phosphatase 2A (PP2A) in *Arabidopsis* plants during development and in response to stress conditions [22]. In *Malus domestica*, the various putative *cis*-acting elements were present in the promoter of *MdHMGR1*, *MdHMGR2* and *MdHMGR4* to response to different hormones, and the expression patterns of *MdHMGR2* and *MdHMGR4* were significantly induced by ethephon (ETH), methyl jasmonate (MeJA), and salicylic acid (SA) [23,24]. In *Origanum vulgare* subsp. *gracile*, the expression of the *HMGR* gene was directly affected by the changing environmental condition and was enhanced under water stress conditions [25].

The *Gossypium* (cotton) genus contains 50 species, of which 45 diploid ( $2n = 2x = 26$ ) and five tetraploid ( $2n = 4x = 52$ ) species, and all diploid cotton species are divided into eight genomes, A, B, C, D, E, F, G and K [26]. At present, two diploid cottons, *Gossypium raimondii* (D<sub>5</sub>) [27,28] and *Gossypium arboreum* (A<sub>2</sub>) [29], and the tetraploid cotton, *Gossypium hirsutum* ((AD)<sub>1</sub>) [30,31] had completed whole genome sequencing. *G. raimondii* is a wild species belonging to the D-genome, and *G. arboreum* is a cultivar belonging to the A-genome [28,29]. They diverged from the same progenitor approximately 5–10 million years ago and *G. arboreum* underwent artificial domestication and selection [26,29]. Tetraploid cotton species are considered to be produced by interspecific hybridization between the African ancestor of an A-genome resembling *G. arboreum* and the American ancestor of a D-genome resembling *G. raimondii* approximately 1–2 million years ago [26,30]. *G. hirsutum*, as one of tetraploid species, is domesticated to provide the world's most natural textile fiber and become a major oilseed crop [30,31]. Additionally, *Gossypium* species serve as an ideal plant for studies of genome evolution and polyploidization [32,33]. *Gossypium* plants are also known to produce a specialized group of terpenes in the cytosol, including gossypol and related sesquiterpenoids, which could be used as phytoalexins in plant defense against pests and pathogens, as well as anticancer agents and male contraceptives in humans [34,35].

In previous studies, only a small portion of *HMGR* genes from cotton have been characterized, and our evolutionary analysis indicated that the *HMGR* genes underwent gene expansion and a unique gene cluster containing four *HMGR* genes was found in the diploid cotton, *G. raimondii* [36]. In order to explore the characteristics of the *HMGR* gene family in *Gossypium* genus, the *HMGR* gene family was identified in *G. raimondii*, *G. arboreum*, and *G. hirsutum* at a genome-wide level, and the phylogenetic relationship, chromosomal localization, gene structure, and protein motifs of *HMGR* genes in the three genomes were comprehensively analyzed in this study.

## 2. Results

### 2.1. Genome-Wide Identification of HMGR Genes in *Gossypium*

The candidate HMGR genes were identified from the *Gossypium* genome using the local blast program with the query sequences of *Arabidopsis* HMGR genes. The obtained sequences were submitted to the Pfam database to confirm the presence of conserved domains (PF00368). Then these sequences were further submitted to the Interpro database and validated to be the HMGR gene family membership (IPR004554). Finally, nine, nine and 18 putative HMGR genes were identified in *G. raimondii*, *G. arboreum*, and *G. hirsutum*, respectively (Table 1 and Supplementary Materials Table S3). The HMGR genes in *G. raimondii* were preferentially named GrHMGR1 to GrHMGR9, according to the published article [36]. Based on the orthologs in *G. raimondii*, the HMGR genes in *G. arboreum* were named GaHMGRs, with the same number as in *G. raimondii*. The HMGR genes in *G. hirsutum* were named GhHMGRs corresponding to the orthologs in *G. raimondii* and *G. arboreum*, and the D and A subgenomes were represented by suffixes D and A after each gene names, respectively. The genomic sequences with the upstream and downstream sequences of these gene loci were extracted and the coding sequences of these genes were re-predicted by the gene annotation tool FGENESH [37]. Then all the coding sequences were further manually verified by RT-PCR using gene-specific primers. As a result, 34 gene loci were confirmed to have the complete open reading frame (Supplementary Materials Table S4), while the *G. arboreum* GaHMGR1 and *G. hirsutum* GhHMGR1A were pseudogenes with a premature stop codon in their coding sequences. Compared with the initial annotation in genome database, the coding sequences of six HMGR genes (*GhHMGR3A*, *GhHMGR4A*, *GhHMGR8A*, *GhHMGR8D*, *GaHMGR2*, and *GaHMGR9*) were modified. Interestingly, the two gene loci (*GhHMGR3A* and *GhHMGR4A*) were originally found in one gene locus in the genome database of *G. hirsutum*, and their coding sequences were re-annotated subsequently. The coding sequences of *GhHMGR8A* and *GhHMGR8D* had a deletion according to the sequencing results. Additionally, the coding sequences of *GaHMGR2* and *GaHMGR9* had an insertion compared with their initial annotations. Overall, the results showed that *G. raimondii* and *G. arboreum* had the same the number of HMGR loci (nine loci) and that *G. hirsutum* had twice as many HMGR loci as the other two species (18 loci).

One salient feature of the plant HMGR proteins is that they usually have two transmembrane movements across the endoplasmic reticulum membrane, and their catalytic domains are exposed to the cytosol during the process of performing function [38,39]. The transmembrane domains of *Gossypium* HMGR proteins were predicted by the online prediction tool TMHMM Server v. 2.0 (Supplementary Materials Figures S1–S3). The results showed that all the 34 HMGR proteins had two transmembrane domains at the N-terminus, and it was deduced that, as with other plant HMGR genes, it was necessary for *Gossypium* HMGR proteins to anchor on the membrane by two transmembrane movements in the catalysis reaction.

### 2.2. Chromosomal Distribution and Phylogenetic Analysis of *Gossypium* HMGR Genes

In *G. raimondii*, the nine HMGR genes were found in five chromosomes: four on chromosome 5, two on chromosome 2, and one on chromosome 8, 12 and 13 each (Supplementary Materials Figure S4). Although the chromosome distribution was found to be little different from the published article [36], there was also an HMGR gene cluster containing four genes on *G. raimondii* chromosome 5 (*GrHMGR2*, *GrHMGR3*, *GrHMGR4* and *GrHMGR5*).

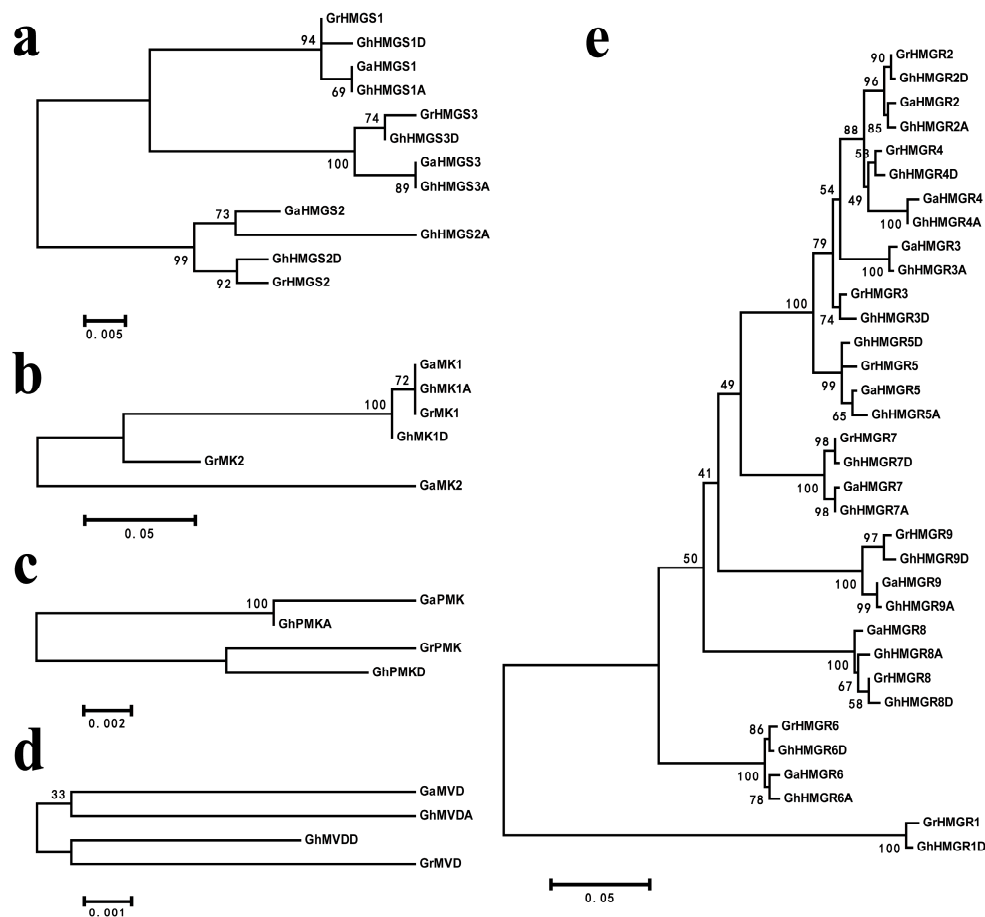
**Table 1.** The information of *HMGR* genes in *G. hirsutum*.

Gene Name	Gene Locus	Chromosome	Location	Strand	Protein Length	Mw(kDa) <sup>a</sup>	pI <sup>a</sup>
<i>GhHMGR1A</i> <sup>c</sup>	<i>HMGR</i> pseudogene	A01	39188679-39190582	+	–	–	–
<i>GhHMGR2A</i>	Gh_A03G1497	A03	95297747-95299778	–	582	62.33	6.24
<i>GhHMGR3A</i> <sup>b</sup>	Gh_A03G1496_1	A03	95253091-95255136	–	585	62.68	6.14
<i>GhHMGR4A</i> <sup>b</sup>	Gh_A03G1496_2	A03	95195182-95197211	–	585	62.67	6.00
<i>GhHMGR5A</i>	Gh_A03G1495	A03	95169197-95171221	–	585	62.71	6.25
<i>GhHMGR6A</i>	Gh_A01G2017	scaffold31_A01	38993-42152	+	585	62.82	6.20
<i>GhHMGR7A</i>	Gh_A12G0103	A12	1425978-1428020	–	585	62.62	6.24
<i>GhHMGR8A</i> <sup>b</sup>	Gh_A04G1424	scaffold1012_A04	184892-187043	–	581	62.26	6.69
<i>GhHMGR9A</i>	Gh_A13G0557	A13	13047920-13050876	–	628	67.60	6.26
<i>GhHMGR1D</i>	Gh_D01G1158	D01	25640487-25642401	+	560	60.40	6.53
<i>GhHMGR2D</i>	Gh_D02G1965	D02	63580366-63582399	–	582	62.33	6.43
<i>GhHMGR3D</i>	Gh_D02G1964	D02	63566380-63569291	–	585	62.54	6.00
<i>GhHMGR4D</i>	Gh_D02G1963	D02	63558389-63560423	–	583	62.40	6.17
<i>GhHMGR5D</i>	Gh_D02G1962	D02	63549091-63551149	–	585	62.78	6.25
<i>GhHMGR6D</i>	Gh_D01G0134	D01	984873-987923	+	585	62.79	5.83
<i>GhHMGR7D</i>	Gh_D12G0115	D12	1451036-1453088	–	585	62.44	6.43
<i>GhHMGR8D</i> <sup>b</sup>	Gh_D04G2012	scaffold3981_D04	28507-30658	+	581	62.26	6.49
<i>GhHMGR9D</i>	Gh_D13G0573	D13	7846186-7848822	+	628	67.46	6.50

<sup>a</sup> The theoretical Mw (molecular weight) and pI (isoelectric point) of the full-length protein are predicted by ProtParam tool (<http://web.expasy.org/protparam/>). <sup>b</sup> The coding sequences of genes are re-annotated. <sup>c</sup> *GhHMGR1A* is a pseudogene identified in this study.

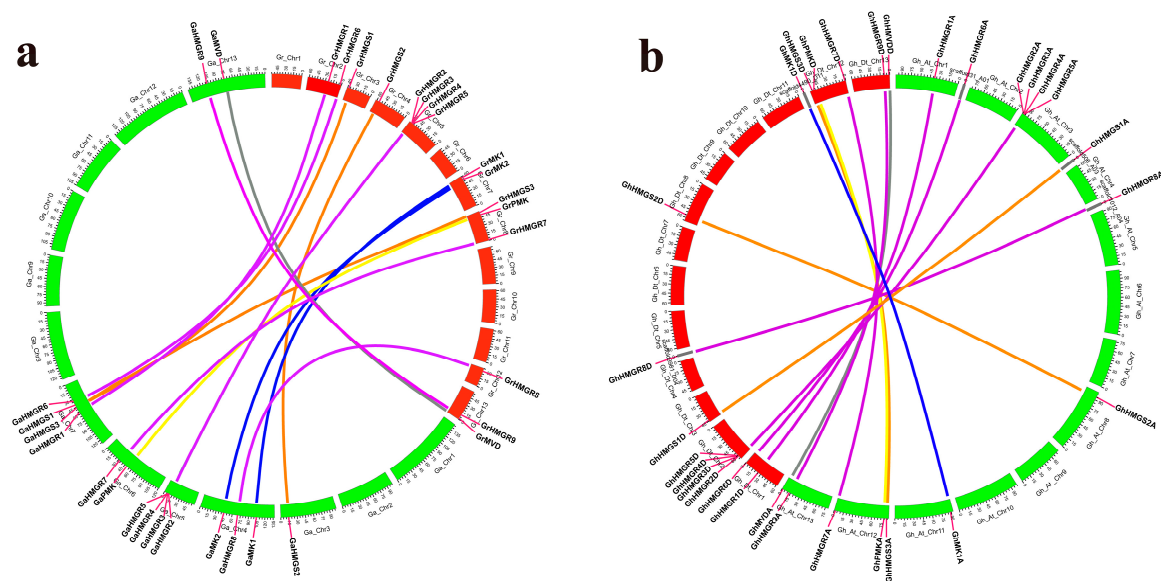
Likewise, in *G. arboreum*, the nine *HMGR* genes were found in five chromosomes (Supplementary Materials Figure S5). There was a single *HMGR* gene locus on chromosome 4, 6 and 13 each, and two *HMGR* gene loci on chromosome 7, but one of them was a pseudogene (*GaHMGR1*). Similarly, there was also an *HMGR* gene cluster on *G. arboreum* chromosome 5 (*GaHMGR2*, *GaHMGR3*, *GaHMGR4* and *GaHMGR5*). Simultaneously, the 18 *HMGR* genes were found in *G. hirsutum*: nine for each of the D-subgenome and the A-subgenome (Supplementary Materials Figure S6). There was a pseudogene named as *GhHMGR1A* in the A-subgenome. Additionally, both of the two subgenomes had an *HMGR* gene cluster that contained four closely adjacent genes (*GhHMGR2D*, *GhHMGR3D*, *GhHMGR4D* and *GhHMGR5D* in the *G. hirsutum* D-subgenome, and *GhHMGR2A*, *GhHMGR3A*, *GhHMGR4A* and *GhHMGR5A* in the *G. hirsutum* A-subgenome), which was consistent with the two diploid cottons, *G. raimondii* and *G. arboreum*.

The genes encoding other related enzymes in upstream or downstream of *HMGR* in the *MVA* pathway were also identified in *Gossypium*. As a result, there were three genes encoding 3-hydroxy-3-methylglutaryl coenzyme A synthase (*HMGS*), two genes encoding mevalonate kinase (*MK*), one gene encoding phosphomevalonate kinase (*PMK*) and one gene encoding mevalonate diphosphate decarboxylase (*MVD*) in each of *G. raimondii* and *G. arboreum*, and six *HMGS* genes, two *MK* genes, two *PMK* genes and two *MVD* genes in *G. hirsutum* (Supplementary Materials Table S5). Phylogenetic trees based on the protein sequences of these genes in the *MVA* pathway of *Gossypium* were constructed to investigate the evolutionary relationships (Figure 1).

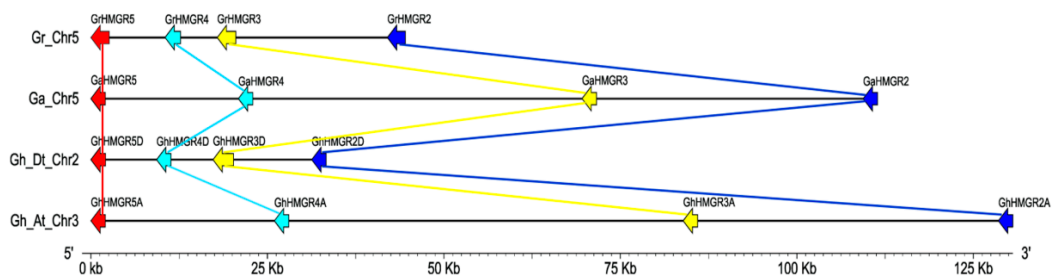


**Figure 1.** Phylogenetic relationship of *HMGS*, *MK*, *PMK*, *MVD* and *HMGR* proteins from *G. raimondii*, *G. arboreum* and *G. hirsutum*. (a) *HMGS*; (b) *MK*; (c) *PMK*; (d) *MVD*; (e) *HMGR*. Numbers at the nodes represent bootstrap support values (1000 replicates). The bars in (a–e) indicate 0.5%, 5%, 0.2%, 0.1% and 5% sequence divergence, respectively.

The position of each gene and the homologous gene pairs were displayed intuitively with Circos diagrams (Figure 2). We identified 16 pairs of orthologous genes between *G. raimondii* and *G. arboreum* and 15 pairs of paralogous genes between the D-subgenome and the A-subgenome of *G. hirsutum*. It showed that there were one-to-one relationships between homologous genes of the MVA pathway in the two diploid cottons or in the two subgenomes of the tetraploid cotton. For the three *HMGS* genes and two *MK* genes in *G. raimondii*, they had orthologous genes in *G. arboreum*. In *G. hirsutum*, the three *HMGS* genes and one *MK* gene in the D-subgenome had corresponding paralogs in the A-subgenome. However, the orthologous genes of *GaMK2* and *GrMK2* were not found in *G. hirsutum*, indicating that they might be lost during the formation of *G. hirsutum*. For the *PMK* genes, including *GaPMK*, *GrPMK*, *GhPMKA* and *GhPMKD*, they exhibited the corresponding homologous relationships in *G. raimondii* and *G. arboreum* or the two subgenomes of *G. hirsutum*. Similarly, two *MVD* genes as a paralogous gene pair in *G. hirsutum* also had a corresponding ortholog in *G. raimondii* and *G. arboreum*. Furthermore, we elaborately detected the homologous relationships of *HMGR* genes, two *HMGRs* on chromosome 2, four *HMGRs* in a gene cluster on chromosome 5, three *HMGRs* on chromosome 8, 12 and 13 of *G. raimondii* had one to one orthologous relationships with two *HMGRs* on chromosome 7, four *HMGRs* in a gene cluster on chromosome 5, three *HMGRs* on chromosome 4, 6 and 13 of *G. arboreum*. Additionally, five *HMGRs* on chromosome 1, scaffold31\_A01, scaffold1012\_A04, 12 and 13, and four *HMGRs* in a gene cluster on chromosome 3 of the A-subgenome had one to one corresponding paralogous relationships with five *HMGRs* on chromosome 1, scaffold3981\_D04, 12 and 13, and four *HMGRs* in a gene cluster on chromosome 2 of the D-subgenome of *G. hirsutum*. In general, it indicated that only the *HMGR* gene formed a gene cluster containing four genes and the gene cluster was present in all three cotton genomes (Figure 3).



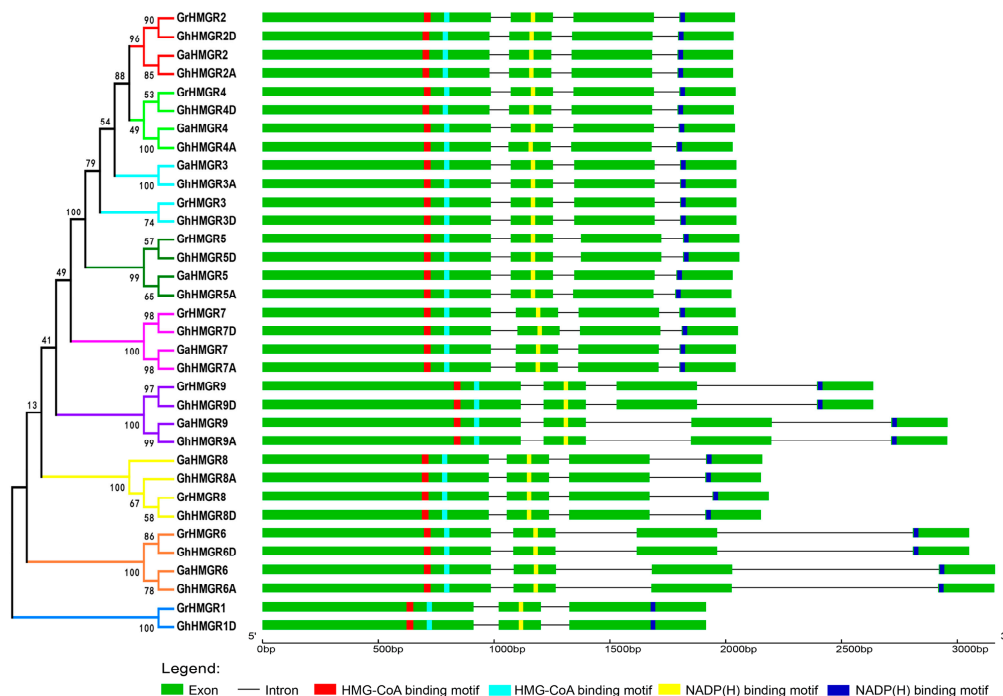
**Figure 2.** Locations and homologous relationships of the MVA pathway genes in *G. raimondii* and *G. arboreum*, and in the A-subgenome and D-subgenome of *G. hirsutum*. (a) Locations and orthologous relationships of the MVA pathway genes in *G. raimondii* and *G. arboreum*; (b) Locations and paralogous relationships of the MVA pathway genes in the D-subgenome and the A-subgenome of *G. hirsutum*. The chromosomes of *G. raimondii*, *G. arboreum*, *G. hirsutum* D-subgenome, and *G. hirsutum* A-subgenome are shown with different colors and labeled as Gr, Ga, Gh\_Dt and Gh\_At, respectively. The putative homologous gene pairs belonging to the *HMGS*, *HMGR*, *MK*, *PMK* and *MVD* gene families are connected by orange, purple, blue, yellow and grey lines, respectively. Several genes are located on the scaffolds that do not determine the exact locations and are placed next to the corresponding chromosomes.



**Figure 3.** The *HMGR* gene clusters in *G. raimondii*, *G. arboreum* and *G. hirsutum*. The putative homologous gene pairs are displayed by arrows of the same color and connected by lines of the same color. The direction of arrows indicates the direction of transcriptions.

### 2.3. Gene Structure and Conserved Protein Motifs of *Gossypium* *HMGR* Genes

The gene structure of *Gossypium* *HMGR* genes was determined (Figure 4). Except for *GrHMGR1* and *GhHMGR1D*, the majority of protein-coding *HMGR* genes in the three *Gossypium* species had the typical gene structure with three introns and four exons, which was the same as the gene structure of most *HMGR* genes in plants [36]. *GrHMGR1* and *GhHMGR1D* lacked the last intron, becoming the structure with two introns and three exons. Most of the *HMGR* genes had almost the same length of exons, only intron length varied greatly. The second and third introns of *GaHMGR6*, *GhHMGR6A*, *GrHMGR6* and *GhHMGR6D* were relatively long. *GaHMGR9*, *GhHMGR9A*, *GrHMGR9* and *GhHMGR9D* had also changed the gene structure. Although they had almost the same exon length as each other, their first exon was longer than other *HMGR* genes. Additionally, *GaHMGR9* and *GhHMGR9A* had a short insertion in the second intron compared with *GrHMGR9* and *GhHMGR9D*. The *HMGR* genes in the gene clusters had almost the same gene structure.

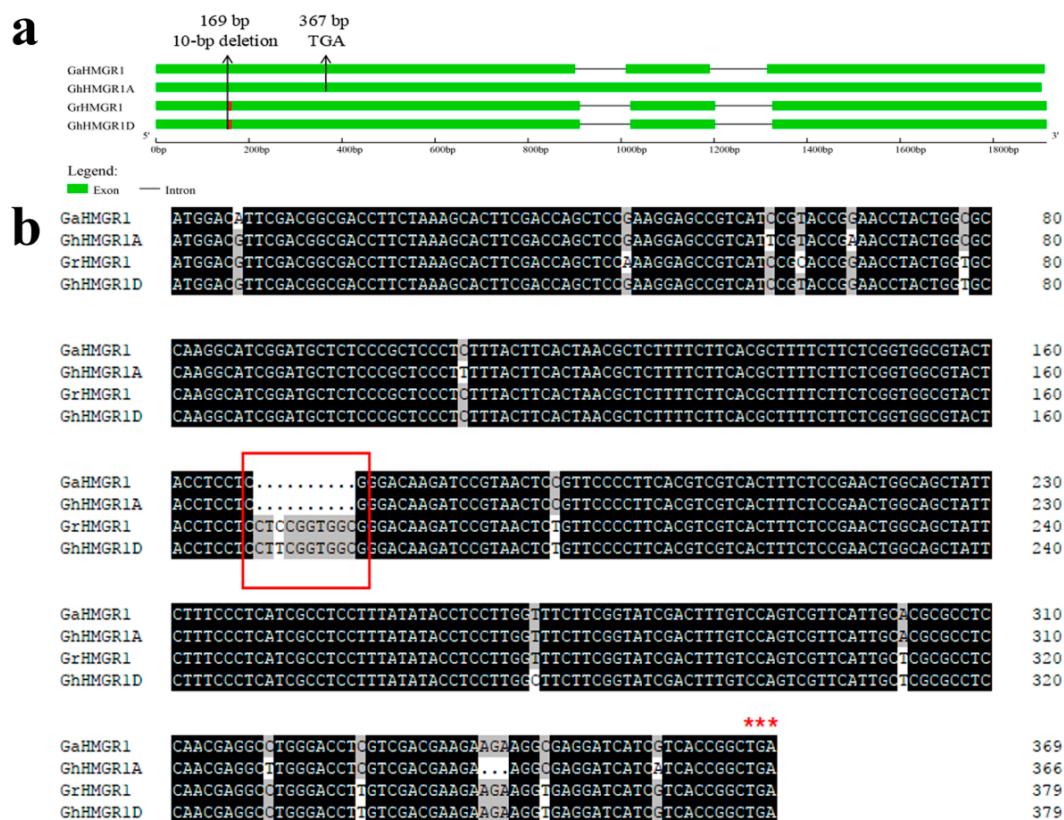


**Figure 4.** Phylogenetic relationship, gene structure, and conserved motifs of *HMGR* genes from *G. raimondii*, *G. arboreum* and *G. hirsutum*. Exons are represented by green boxes and introns by black lines. The two HMG-CoA binding motifs (EMPVGYVQIP and TTEGCLVA) and two NADP(H) binding motifs (DAMGMNM and GTVGGT) are represented by red, light blue, yellow and dark blue boxes, respectively.

In the catalytic domain of the HMGR proteins, there were four highly conserved motifs: two HMG-CoA binding motifs (EMPVGYVQIP and TTEGCLVA) and two NADP(H) binding motifs (DAMGMNM and GTVGGGT) [36,40,41]. All 34 *HMGR* proteins of the three *Gossypium* species had the four conserved motifs, in which the relative position of these motifs was also conserved and remained constant (Figure 4 and Supplementary Materials Figure S7). Specifically, the first HMG-CoA binding motif (EMPVGYVQIP) was separated from the second HMG-CoA binding motif (TTEGCLVA) by 19 amino acid residues, there was the first NADP(H) binding motif (DAMGMNM) after 88 amino acid residues and the second NADP(H) binding motif (GTVGGGT) was at the C-terminus, separated from the first NADP(H) binding motif by 142 amino acid residues. The sequences of *GaHMGR9*, *GhHMGR9A*, *GrHMGR9* and *GhHMGR9D* at the N-terminus before the first HMG-CoA binding motif were longer than that of other *HMGR* genes, whereas *GrHMGR1* and *GhHMGR1D* were shorter than other genes in this region.

#### 2.4. Identification of HMGR Pseudogenes

We found a special *HMGR* gene locus *GaHMGR1* in *G. arboreum*. The sequence alignment of *GaHMGR1* and its orthologous gene in *G. raimondii* *GrHMGR1* showed that their sequences were very similar, and there were only a few nucleotide insertion, deletion and substitution mutations. However, *GaHMGR1* lacked a 10-bp fragment in the first exon region compared with *GrHMGR1*, leading to a premature stop codon (TGA) mutation (Figure 5).



**Figure 5.** The *HMGR* pseudogenes in *G. arboreum* and *G. hirsutum*. (a) The gene structure of *GaHMGR1*, *GhHMGR1A*, *GrHMGR1* and *GhHMGR1D*. Exons are represented by green boxes and introns by black lines. The red boxes in the first exons of *GrHMGR1* and *GhHMGR1D* indicate the 10-bp deletion at the 169-bp position in *GaHMGR1* and *GhHMGR1A*; (b) The alignment of predicted coding sequence of *GaHMGR1* and *GhHMGR1A*, and corresponding sequence of *GrHMGR1* and *GhHMGR1D*. The red outlined box indicates the 10-bp deletion of *GaHMGR1* and *GhHMGR1A*. The three red stars indicate the premature stop codon (TGA) at the 367-bp position of *GaHMGR1* and *GhHMGR1A*.



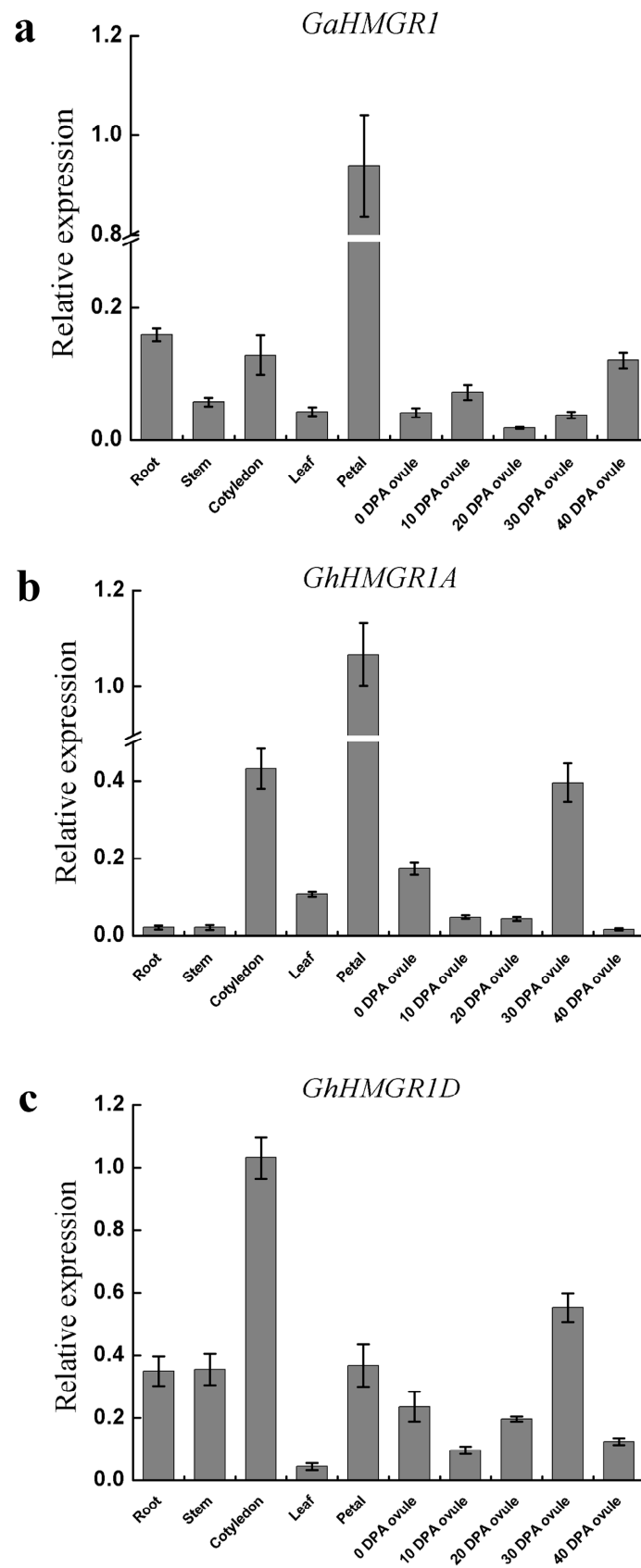
Through cloning and sequencing, it was found that *GaHMGR1* in *G. arboreum* could transcribe the full length RNA sequence similar to the *GrHMGR1*. However, the predicted protein sequence based on the RNA sequence showed that the translation was terminated prematurely due to the advance of the stop codon and the functional protein could not be translated.

Based on the genomic sequence of *GaHMGR1*, an *HMGR* pseudogene in the A-subgenome of *G. hirsutum* was identified by the Blast method and named *GhHMGR1A*. The sequence alignment revealed that its sequence was very similar to *GaHMGR1*. And both of them had the frameshift mutation and premature stop codon (TGA) due to the 10-bp deletion at the same position. These results indicated that *GaHMGR1* and *GhHMGR1A* were really orthologous genes. In particular, we cloned *GhHMGR1A* from *G. hirsutum* (TM-1) using the seedling cDNA as the template, and sequencing results showed that its transcripts contained the whole sequence belonging to exons and introns in its protein-coding counterpart. In order to eliminate the effects of DNA contamination and alternative splicing and further confirm the result, we cloned the pseudogene using the cDNAs of roots, stems, cotyledons, leaves, and petals of TM-1, and at least four clones for the pseudogene from each of the materials were picked randomly and sequenced. Finally, the results indicated that the transcripts were consistent in all materials. Then in order to study the distribution of the *HMGR* pseudogene in *Gossypium*, the pseudogene was identified in several *Gossypium* species, including wild species of the D-genome, wild species and cultivars of A-genome, and wild species, semi-domesticated species and cultivars of the AD-genome. We collected leaves of these species and used their cDNAs to clone the pseudogene with gene-specific primers, and then four clones of each material were picked randomly for sequencing. It was found that all the A-genome species and tetraploid species (AD-genome) we used had the *HMGR* pseudogene (Table 2). Therefore, it could be deduced that the pseudogene was derived from an ancient A-genome species and transferred to the A-subgenome of the tetraploid species during the *Gossypium* evolution.

**Table 2.** Distribution of the *HMGR* pseudogene gene in *Gossypium*.

Species	Type	Genomic Group	<i>HMGR</i> Pseudogene
<i>G. raimondii</i>	wild species	D <sub>5</sub>	Non-existence
<i>G. herbaceum</i> race. <i>africanum</i>	wild species	A <sub>1</sub>	Existence
<i>G. herbaceum</i> cv. Jinta	cultivar	A <sub>1</sub>	Existence
<i>G. arboreum</i> cv. Shixiya1	cultivar	A <sub>2</sub>	Existence
<i>G. darwinii</i>	wild species	(AD) <sub>5</sub>	Existence
<i>G. mustelinum</i>	wild species	(AD) <sub>4</sub>	Existence
<i>G. hirsutum</i> race. <i>latifolium</i>	semi-domesticated species	(AD) <sub>1</sub>	Existence
<i>G. hirsutum</i> cv. TM-1	cultivar	(AD) <sub>1</sub>	Existence
<i>G. hirsutum</i> cv. CCR141	cultivar	(AD) <sub>1</sub>	Existence
<i>G. barbadense</i> cv. Xinhai21	cultivar	(AD) <sub>2</sub>	Existence

Expression patterns of the two *HMGR* pseudogenes (*GaHMGR1* and *GhHMGR1A*) and the homologous gene of *GhHMGR1A* in the D-subgenome of *G. hirsutum* (*GhHMGR1D*) were analyzed by qRT-PCR in roots, stems, cotyledons, leaves, petals, and ovules collected at 0 DPA, 10 DPA, 20 DPA, 30 DPA and 40 DPA of *G. arboreum* and *G. hirsutum* (Figure 6). The results showed that all the three genes displayed tissue-specific expression patterns. Both of the pseudogenes had the highest expression level in petals. In addition, *GaHMGR1* had relatively high expression in roots, cotyledons, ovules at 40 DPA, and had the lowest expression in ovules at 20 DPA. *GhHMGR1A* was highly expressed in cotyledons and ovules at 30 DPA, and lowly expressed in roots, stems and ovules at 40 DPA. Moreover, *GhHMGR1D* as a protein-coding *HMGR* gene, which also was the paralog of the pseudogene *GhHMGR1A* in *G. hirsutum*, was expressed at a high level in roots, stems, cotyledons, petals and ovules at 30 DPA. Although *GhHMGR1D* had the highest expression level in cotyledons, its expression pattern was roughly similar to that of the pseudogene *GhHMGR1A*.



**Figure 6.** Expression patterns of *GaHMGR1*, *GhHMGR1A* and *GhHMGR1D* in different tissues. (a) *GaHMGR1*; (b) *GhHMGR1A*; (c) *GhHMGR1D*.

### 3. Discussion

It has been suggested that *HMGR* is a multigene family in cotton and there are seven to nine members in the tetraploid cotton according to the Southern blot technique [15,42]. Our previous studies have shown that there were nine *HMGR* genes in *G. raimondii* and the number of *HMGR* genes was significantly expanded compared with other plants [36]. In this study, another version of *G. raimondii* genome data [27] was used to further confirm that nine *HMGR* gene loci were included in *G. raimondii*. Unexpectedly, the chromosome distribution was little different. In the published study, *GrHMGR6* was located on chromosome 7, but in this study it was on chromosome 2. In addition, the order of four genes in the *HMGR* gene cluster was exactly the opposite. The *G. raimondii* genome used in the previous study [36] was sequenced on the Illumina HiSeq 2000 platform at the BGI-Shenzhen and was assembled using the SOAPdenovo with a *K*-mer of 41 and SSPACE software [28], while the *G. raimondii* genome used in this study was sequenced on the Applied Biosystems 3730xl, Roche 454 XLR and Illumina Genome Analyzer (GA)IIx machines at the U.S. Department of Energy Joint Genome Institute and was assembled using the modified version of Arachne v.20071016 with specific parameters [27]. Considering that the two versions of *G. raimondii* genome data were independent sequenced and assembled, there might be numerous different assembling which might result in the difference of chromosomal distribution of *HMGR* genes. Furthermore, nine and 18 *HMGR* genes were identified in *G. arboreum* and *G. hirsutum*, respectively, using the recently published genome database [29,31]. Phylogenetic analysis showed that the nine *HMGR* genes in *G. arboreum* had one-to-one orthologous relationships with the nine *HMGR* genes in *G. raimondii*, which indicated that these *HMGR* genes were distributed to *G. raimondii* and *G. arboreum* with speciation from the common ancestor of them, and the number of *HMGR* genes has expanded in the ancestral species. The number of *HMGR* genes in *G. hirsutum* was just twice than that of *G. raimondii* and *G. arboreum*, indicating that all the *HMGR* genes were retained in the process of polyploidization.

Gene duplication, including segmental duplication and tandem duplication, has been recognized as the main mechanisms which contributed to expansion of gene families [43,44]. In *Glycine max* and *Populus trichocarpa*, the segmental duplication was the main reason for the expansion of *MYB* and *WRKY* gene families, but there were also some clusters resulting from the tandem duplication [45–48]. *HMGR* genes were generally distributed on chromosomes dispersedly in other plants [36]. However, there was a gene cluster containing four closely adjacent *HMGR* genes on the chromosome 5 of *G. raimondii* and *G. arboreum*. There was also a gene cluster on the chromosome 3 of the A-subgenome and chromosome 2 of the D-subgenome in *G. hirsutum*. The sequences of four *HMGR* genes in each gene cluster were very similar. It was speculated that the *HMGR* genes had undergone tandem duplication in the common ancestor of *G. raimondii* and *G. arboreum*, leading to the emergence of a *HMGR* gene cluster. Previous studies have shown that segmental duplication and tandem duplication play similar roles in the expansion of the *HMGR* gene family in *G. raimondii* [36]. In this study, it was found that there were nine corresponding *HMGR* gene loci in *G. arboreum*, which further indicated that this segmental duplication and tandem duplication had occurred in the common ancestor of *G. raimondii* and *G. arboreum*, resulting in the expansion of *HMGR* genes. Genomic evolution analysis showed that a whole genome duplication event was uniquely occurred for *Gossypium* after speciation from its closely related species, *Theobroma cacao* [29], which supported the inference that segmental duplication was one of the causes of *HMGR* gene expansion.

In this study, several related genes of the MVA pathway were identified in *G. raimondii*, *G. arboreum* and *G. hirsutum*. There were three *HMGS* genes, nine *HMGR* genes, two *MK* genes, one *PMK* gene, and one *MVD* gene in each of *G. raimondii* and *G. arboreum*, and six *HMGS* genes, 18 *HMGR* genes, two *MK* genes, two *PMK* genes, and two *MVD* genes in *G. hirsutum*. In the model plant *Arabidopsis*, there were one *HMGS* gene, two *HMGR* genes, one *MK* gene, one *PMK* gene and two *MVD* genes [49]. Compared with *Arabidopsis*, the number of *HMGS*, *HMGR* and *MK* gene loci was more in *G. raimondii* and *G. arboreum*, the number of *PMK* loci was the same as that in *Arabidopsis*, and the number of *MVD* loci was one less. In general, only the number of *HMGR* genes in *Gossypium* species was most

significantly expanded in the MVA pathway, and there was a unique gene cluster that might have resulted from tandem duplication. *Gossypium* species synthesize gossypol and related sesquiterpenoids uniquely by the MVA pathway in the cytosol and accumulate in roots and pigment glands of aerial tissues, to resist the invasion of pests and pathogens [35,50,51]. The proteins encoded by *HMGR* genes are a rate-limiting enzyme of the MVA pathway and are important regulatory sites for the biosynthesis of terpenes in the cytosol [9,10]. Therefore, it could be speculated that this increase in the number of *HMGR* genes in *Gossypium* species might be likely related to the biosynthesis of more terpenes including gossypol in the cytosol during their growth and development. In addition, this study found that *HMGR* gene expansion and a unique *HMGR* gene cluster were present in the three *Gossypium* species, and the four genes within the *HMGR* gene cluster had almost the same gene and protein structure, which indicated that the gene cluster was quite conserved in the evolutionary process.

After the number of genes was expanded, functional differentiation has three fates: pseudogenization, loss of gene function; neo-functionalization, access to new gene function; sub-functionalization, both of the two copies retain the function of ancestral gene [52]. *GrHMGR1* of *G. raimondii* and its orthologous gene in the D-subgenome of *G. hirsutum*, *GhHMGR1D*, were lacking the third intron compared with other *HMGR* genes. Moreover, its orthologous gene in *G. arboreum*, *GaHMGR1*, and the orthologous gene of *GaHMGR1* in the A-subgenome of *G. hirsutum*, *GhHMGR1A*, had a 10-bp deletion at the same position, resulting in a frameshift mutation, and could not be translated into functional proteins. These results suggested that one functional member after expansion of *HMGR* genes in the common ancestor of *G. raimondii* and *G. arboreum*, might be differentiated by losing the third intron, then it became a pseudogene by losing the 10-bp fragment in the first exon in the ancestor of *G. arboreum* after the speciation of *G. raimondii* and *G. arboreum*. The pseudogene was identified in all the A-genome and AD-genome species collected in this study. Thus, it suggested that the pseudogene might be transferred from wild species to cultivars of the A-genome during process of domestication. Then, during the tetraploid formation by interspecific hybridization between the A-genome and D-genome, the pseudogene was transferred from the A-genome to the A-subgenome.

Previous study has found an *HMGR* pseudogene named *ψhmg5* in *G. hirsutum* and its transcript was detected in cotton embryos [53]. However, because of the lack of genome data of other *Gossypium* species, they were not sure whether the pseudogene arose before or after the polyploidization. Through sequence alignment, we found that the pseudogene identified in the study was the same as the one in the previous study [53]. However, more deeply, we provided a possible mechanism for the formation of the pseudogene by comparative genome analysis. After the expansion of *HMGR* gene in the progenitor of *Gossypium* species, one functional gene might become a new gene with significantly different gene structure attributed to selective excision of the third intron, and led to a pseudogene through a 10-bp deletion in the first exon after a series of evolutionary processes in the A-genome, then transferred to the A-subgenome with polyploidization. In addition, transcripts of the identified pseudogene in *G. hirsutum* contained the whole sequence belonging to exons and introns in its protein-coding counterpart and could be detected in all materials we had collected using the qRT-PCR method.

Pseudogenes do not code for protein, so they have long been labeled as “junk” DNA [54]. However, recent results demonstrated that some pseudogenes could influence their parent genes through their transcripts, including negative regulation and positive regulation [55,56]. For example, the transcripts of pseudogene can produce endogenous siRNAs and then silence the expression of parent gene by RNA interference. In soybean, the inhibition of seed coat pigmentation induced by the *I* gene results from posttranscriptional gene silencing (PTGS) of chalcone synthase (*CHS*) genes and leads to a uniform yellow color of mature harvested seeds [57]. *GmIRCHS* (*Glycine max* inverted-repeat *CHS* pseudogene) was identified as a candidate for the *I* gene [58]. The siRNAs derived from *GmIRCHS* cleaved the mRNA of all *CHS* genes to inhibit their function, and occurred in the seed coat, specifically [59]. Additionally, the pseudogene transcripts can positively regulate its homologous gene by competitively binding miRNAs. For example, *PTEN* is a tumor suppressor

gene and maintaining *PTEN* protein levels can inhibit tumorigenesis. Its pseudogene *PTENP1* is highly similar to the homologous coding gene *PTEN* at the 3' untranslated region (UTR), which can bind miRNAs and reduce the cell concentration of miRNAs, leading to *PTEN* escape from miRNAs repression regulation [60]. In our study, the *HMGR* pseudogenes could still be detected after a long-term evolution, and the pseudogene *GhHMGR1A* showed tissue-specific expression and had the similar expression pattern with its paralogous gene *GhHMGR1D* in the other subgenome. Therefore, it suggested that the pseudogenes might have a potential role in regulation of other *HMGR* genes. Additionally, both of the pseudogenes had the highest expression level in petals. Many plants, such as snapdragon [61], *Hedychium coronarium* [62], and kiwifruit (*Actinidia deliciosa*) [63], can synthesize and emit lots of terpenes in their petals. It could be speculated that the high expression level of the two pseudogenes in petals might be related to the large demand for the precursors for terpene biosynthesis in petals of *G. arboreum* and *G. hirsutum*. Of course, the hypothesis requires further experimental evidence.

## 4. Materials and Methods

### 4.1. Identification of Genes in the MVA Pathway in *Gossypium*

The genome data of *G. raimondii* [27], *G. arboreum* [29] and *G. hirsutum* [31] were downloaded from the CottonGen database (<https://www.cottongen.org/>). Then, the local blast database was established for these genome data. The protein sequences of *Arabidopsis thaliana* genes in the MVA pathway were collected from the TAIR database (<http://www.arabidopsis.org>) [49,64]. BlastP and tBlastN programs were performed against the *Gossypium* genome local databases using the *Arabidopsis* protein sequences as queries with default parameters. All candidates were verified using the Pfam database [65] and InterPro database [66] to identify the members of gene family in the MVA pathway.

### 4.2. Sequence Alignment and Phylogenetic Tree Construction

Multiple sequence alignments were generated using ClustalX software (Version 2.1, Conway Institute UCD, Dublin, Ireland) [67] for the protein sequences with default parameters. Based on the result of multiple sequence alignment, phylogenetic trees were generated using the maximum likelihood method in MEGA software (Version 5.2, BioDesign Institute, Tempe, AZ, USA) [68], using the bootstrap method to assess the reliability with 1000 replicates.

### 4.3. Chromosomal Mapping, Protein Motif, and Gene Structure Analysis

The physical location data of identified genes were retrieved from the *Gossypium* genomes. MapInspect software (<http://mapinspect.software.informer.com/>) [69] and Circos software (Version 0.67, [www.circos.ca](http://www.circos.ca)) [70] were used to generate chromosomal distribution images for these genes in *G. raimondii*, *G. arboreum* and *G. hirsutum* according to their starting positions on chromosomes. The gene exon/intron structure was drawn using the Gene Structure Display Server (GSDS, <http://gsds.cbi.pku.edu.cn/>) online tool [71] by comparing the coding sequence (CDS) of each gene with its genomic sequence.

### 4.4. Cotton Plant Growth and Sample Collection

The seeds of *G. arboreum* cv. Shixiya 1, *G. hirsutum* cv. TM-1 and other species used in this study were supplied by Institute of Cotton Research, Chinese Academy of Agricultural Sciences (CAAS, Anyang, China). Whole seedlings, roots, stems, cotyledons and leaves were collected from two-week-old seedlings grown in a greenhouse. Petals were collected from plants on the day of flowering, and ovules were collected at 0, 10, 20, 30 and 40 days post anthesis (DPA). All samples were quick-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

#### 4.5. RNA Isolation and cDNA Synthesis

Total RNA was isolated from each sample using the RNA Extraction Kit (TIANGEN, Beijing, China). The RNA concentration was measured using a NanoDrop2000 microvolume spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and the integrity of RNA was analyzed on 1% agarose gels. One microgram of total RNA was used for first strand cDNA synthesis using PrimeScript™ 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China).

#### 4.6. Reverse Transcription PCR (RT-PCR) and Quantitative Real-Time RT-PCR (qRT-PCR)

The gene-specific primers were designed based on the nucleotide sequences by Oligo software (Version 7.60, Molecular Biology Insights, Cascade, CO, USA) and synthesized by Suzhou GENEWIZ (Supplementary Materials Tables S1 and S2). The RT-PCR was carried out as follows: 94 °C for 5 min; followed by 35 cycles of 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 1 min 30 s; then 72 °C for 10 min. The amplified fragments were purified with the MiniBEST Agarose Gel DNA Extraction Kit (TaKaRa, Dalian, China), cloned into the pMD18-Tvector (TaKaRa, Dalian, China) and verified by sequencing. The qRT-PCR was performed using a LightCycler480 system (Roche, Basel, Switzerland) with SYBR® Premix Ex Taq™ (TaKaRa, Dalian, China) and the cotton *UBQ7* gene was used as an internal control. The amplification parameters were as follows: stage 1: 95 °C, 5 min; stage 2: 40 cycles of 95 °C for 10 s, 60 °C for 10 s, 72 °C for 10 s; stage 3: extension at 72 °C for 10 min. Three biological replicates were used for each sample and the results were analyzed using the  $2^{-\Delta\Delta CT}$  method [72].

### 5. Conclusions

We performed a genome-wide identification of the *HMGR* gene family in *Gossypium* and analyzed their structure, conserved motif, and evolution. The results revealed that the *HMGR* genes were obviously expanded in the common ancestor of *Gossypium* mainly by segmental duplication and tandem duplication, and a gene cluster containing four closely adjacent genes was highly conserved during evolution. There was a pseudogene in *G. arboreum* and the A-subgenome of *G. hirsutum*, and they displayed tissue-specific expression patterns. This study is the first to characterize the *HMGR* gene family in *Gossypium* species and lays an important foundation for further study of cytosolic terpenes biosynthesis in cotton.

**Supplementary Materials:** The following are available online. Table S1. Primers for reverse transcription PCR. Table S2. Primers for quantitative real-time PCR. Table S3. The information of *HMGR* genes in *G. raimondii* and *G. arboreum*. Table S4. The coding sequences of *HMGR* genes in *G. raimondii*, *G. arboreum* and *G. hirsutum*. Table S5. The information of *HMGS*, *MK*, *PMK* and *MVD* genes in *G. raimondii*, *G. arboreum* and *G. hirsutum*. Figure S1. Predicted transmembrane domain for *G. raimondii* *HMGR* proteins. Figure S2. Predicted transmembrane domain for *G. arboreum* *HMGR* proteins. Figure S3. Predicted transmembrane domain for *G. hirsutum* *HMGR* proteins. Figure S4. Chromosomal distributions of *HMGR* genes in *G. raimondii*. Figure S5. Chromosomal distributions of *HMGR* genes in *G. arboreum*. Figure S6. Chromosomal distributions of *HMGR* genes in *G. hirsutum*. Figure S7. Phylogenetic relationship and distribution of conserved motifs in *HMGR* proteins from *G. raimondii*, *G. arboreum* and *G. hirsutum*.

**Acknowledgments:** This work was supported by the Special Fund for Henan Agriculture Research System (Grant No. S2013-07-G04) and the State Key Laboratory of Cotton Biology Open Fund (Grant No. CB2016A06).

**Author Contributions:** S.Z. and Z.M. conceived and designed the research. W. Liu, W.Z. and L.J. performed the experiments. W. Liu, Z.Z., W. Li and L.L. analyzed the data. W. Li, Z.R. and Z.W. prepared the figures. W. Liu and Z.Z. wrote the manuscript. S.Z. and Z.M. revised the manuscript. All authors read and approved the final manuscript.

**Conflicts of Interest:** The authors declare that they have no competing interests.

### References

1. Withers, S.T.; Keasling, J.D. Biosynthesis and engineering of isoprenoid small molecules. *Appl. Microbiol. Biotechnol.* **2007**, *73*, 980–990. [[CrossRef](#)] [[PubMed](#)]

2. Lange, B.M.; Rujan, T.; Martin, W.; Croteau, R. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13172–13177. [[CrossRef](#)] [[PubMed](#)]
3. Nicotra, A.B.; Atkin, O.K.; Bonser, S.P.; Davidson, A.M.; Finnegan, E.J.; Mathesius, U.; Poot, P.; Purugganan, M.D.; Richards, C.L.; Valladares, F. Plant phenotypic plasticity in a changing climate. *Trends Plant Sci.* **2010**, *15*, 684–692. [[CrossRef](#)] [[PubMed](#)]
4. Bouvier, F.; Rahier, A.; Camara, B. Biogenesis, molecular regulation and function of plant isoprenoids. *Prog. Lipid Res.* **2005**, *44*, 357–429. [[CrossRef](#)] [[PubMed](#)]
5. Bohlmann, J.; Keeling, C.I. Terpenoid biomaterials. *Plant J.* **2008**, *54*, 656–669. [[CrossRef](#)] [[PubMed](#)]
6. Singh, B.; Sharma, R.A. Plant terpenes: Defense responses, phylogenetic analysis, regulation and clinical applications. *3 Biotech* **2015**, *5*, 129–151. [[CrossRef](#)] [[PubMed](#)]
7. Bick, J.A.; Lange, B.M. Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: Unidirectional transport of intermediates across the chloroplast envelope membrane. *Arch. Biochem. Biophys.* **2003**, *415*, 146–154. [[CrossRef](#)]
8. Laule, O.; Furholz, A.; Chang, H.S.; Zhu, T.; Wang, X.; Heifetz, P.B.; Grisse, W.; Lange, M. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6866–6871. [[CrossRef](#)] [[PubMed](#)]
9. Rodríguez-Concepción, M. Early Steps in Isoprenoid Biosynthesis: Multilevel Regulation of the Supply of Common Precursors in Plant Cells. *Phytochem. Rev.* **2006**, *5*, 1–15. [[CrossRef](#)]
10. Hemmerlin, A.; Harwood, J.L.; Bach, T.J. A *raison d'être* for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog. Lipid Res.* **2012**, *51*, 95–148. [[CrossRef](#)] [[PubMed](#)]
11. Learned, R.M.; Fink, G.R. 3-Hydroxy-3-methylglutaryl-coenzyme A reductase from *Arabidopsis thaliana* is structurally distinct from the yeast and animal enzymes. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 2779–2783. [[CrossRef](#)] [[PubMed](#)]
12. Enjuto, M.; Balcells, L.; Campos, N.; Caelles, C.; Arro, M.; Boronat, A. *Arabidopsis thaliana* contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 927–931. [[CrossRef](#)] [[PubMed](#)]
13. Ha, S.H.; Lee, S.W.; Kim, Y.M.; Hwang, Y.S. Molecular characterization of *Hmg2* gene encoding a 3-hydroxy-methylglutaryl-CoA reductase in rice. *Mol. Cells* **2001**, *11*, 295–302. [[PubMed](#)]
14. Aoyagi, K.; Beyou, A.; Moon, K.; Fang, L.; Ulrich, T. Isolation and characterization of cDNAs encoding wheat 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Plant Physiol.* **1993**, *102*, 623–628. [[CrossRef](#)] [[PubMed](#)]
15. Loguercio, L.L.; Scott, H.C.; Trolinder, N.L.; Wilkins, T.A. Hmg-coA reductase gene family in cotton (*Gossypium hirsutum* L.): Unique structural features and differential expression of *hmg2* potentially associated with synthesis of specific isoprenoids in developing embryos. *Plant Cell Physiol.* **1999**, *40*, 750–761. [[CrossRef](#)] [[PubMed](#)]
16. Kobayashi, T.; Kato-Emori, S.; Tomita, K.; Ezura, H. Detection of 3-hydroxy-3-methylglutaryl-coenzyme A reductase protein Cm-HMGR during fruit development in melon (*Cucumis melo* L.). *Theor. Appl. Genet.* **2002**, *104*, 779–785. [[PubMed](#)]
17. Devi, K.; Patar, L.; Modi, M.K.; Sen, P. An Insight into Structure, Function, and Expression Analysis of 3-Hydroxy-3-Methylglutaryl-CoA Reductase of *Cymbopogon winterianus*. *Bioinform. Biol. Insights* **2017**, *11*, 1–11. [[CrossRef](#)] [[PubMed](#)]
18. Gu, W.; Geng, C.; Xue, W.; Wu, Q.; Chao, J.; Xu, F.; Sun, H.; Jiang, L.; Han, Y.; Zhang, S. Characterization and function of the 3-hydroxy-3-methylglutaryl-CoA reductase gene in *Alisma orientale* (Sam.) Juz. and its relationship with protostane triterpene production. *Plant Physiol. Biochem.* **2015**, *97*, 378–389. [[CrossRef](#)] [[PubMed](#)]
19. Schaller, H.; Grausem, B.; Benveniste, P.; Chye, M.L.; Tan, C.T.; Song, Y.H.; Chua, N.H. Expression of the *Hevea brasiliensis* (H.B.K.) Mull. Arg. 3-Hydroxy-3-Methylglutaryl-Coenzyme a Reductase 1 in Tobacco Results in Sterol Overproduction. *Plant Physiol.* **1995**, *109*, 761–770. [[CrossRef](#)] [[PubMed](#)]
20. Kim, Y.J.; Lee, O.R.; Oh, J.Y.; Jang, M.G.; Yang, D.C. Functional analysis of 3-hydroxy-3-methylglutaryl coenzyme a reductase encoding genes in triterpene saponin-producing ginseng. *Plant Physiol.* **2014**, *165*, 373–387. [[CrossRef](#)] [[PubMed](#)]
21. Munoz-Bertomeu, J.; Sales, E.; Ros, R.; Arrillaga, I.; Segura, J. Up-regulation of an N-terminal truncated 3-hydroxy-3-methylglutaryl CoA reductase enhances production of essential oils and sterols in transgenic *Lavandula latifolia*. *Plant Biotechnol. J.* **2007**, *5*, 746–758. [[CrossRef](#)] [[PubMed](#)]

22. Leivar, P.; Antolin-Llovera, M.; Ferrero, S.; Closa, M.; Arro, M.; Ferrer, A.; Boronat, A.; Campos, N. Multilevel control of *Arabidopsis* 3-hydroxy-3-methylglutaryl coenzyme A reductase by protein phosphatase 2A. *Plant Cell* **2011**, *23*, 1494–1511. [[CrossRef](#)] [[PubMed](#)]
23. Lv, D.M.; Zhang, T.T.; Deng, S.; Zhang, Y.H. Functional analysis of the *Malus domestica* *MdHMGR2* gene promoter in transgenic *Arabidopsis thaliana*. *Biol. Plant.* **2016**, *60*, 667–676. [[CrossRef](#)]
24. Lv, D.; Zhang, Y. Isolation and functional analysis of apple *MdHMGR1* and *MdHMGR4* gene promoters in transgenic *Arabidopsis thaliana*. *Plant Cell Tissue Organ Cult.* **2017**, *129*, 133–143. [[CrossRef](#)]
25. Morshedloo, M.R.; Craker, L.E.; Salami, A.; Nazeri, V.; Sang, H.; Maggi, F. Effect of prolonged water stress on essential oil content, compositions and gene expression patterns of mono- and sesquiterpene synthesis in two oregano (*Origanum vulgare* L.) subspecies. *Plant Physiol. Biochem.* **2017**, *111*, 119–128. [[CrossRef](#)] [[PubMed](#)]
26. Wendel, J.F.; Cronn, R.C. Polyploidy and the evolutionary history of cotton. *Adv. Agron.* **2003**, *78*, 139–186.
27. Paterson, A.H.; Wendel, J.F.; Gundlach, H.; Guo, H.; Jenkins, J.; Jin, D.; Llewellyn, D.; Showmaker, K.C.; Shu, S.; Udall, J. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **2012**, *492*, 423–427. [[CrossRef](#)] [[PubMed](#)]
28. Wang, K.; Wang, Z.; Li, F.; Ye, W.; Wang, J.; Song, G.; Yue, Z.; Cong, L.; Shang, H.; Zhu, S. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* **2012**, *44*, 1098–1103. [[CrossRef](#)] [[PubMed](#)]
29. Li, F.; Fan, G.; Wang, K.; Sun, F.; Yuan, Y.; Song, G.; Li, Q.; Ma, Z.; Lu, C.; Zou, C. Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat. Genet.* **2014**, *46*, 567–572. [[CrossRef](#)] [[PubMed](#)]
30. Li, F.; Fan, G.; Lu, C.; Xiao, G.; Zou, C.; Kohel, R.J.; Ma, Z.; Shang, H.; Ma, X.; Wu, J. Genome sequence of cultivated Upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. *Nat. Biotechnol.* **2015**, *33*, 524–530. [[CrossRef](#)] [[PubMed](#)]
31. Zhang, T.; Hu, Y.; Jiang, W.; Fang, L.; Guan, X.; Chen, J.; Zhang, J.; Saski, C.A.; Scheffler, B.E.; Stelly, D.M. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **2015**, *33*, 531–537. [[CrossRef](#)] [[PubMed](#)]
32. Chen, Z.J.; Scheffler, B.E.; Dennis, E.; Triplett, B.A.; Zhang, T.; Guo, W.; Chen, X.; Stelly, D.M.; Rabinowicz, P.D.; Town, C.D. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* **2007**, *145*, 1303–1310. [[CrossRef](#)] [[PubMed](#)]
33. Ashraf, J.; Dongyun, Z.; Qiaolian, W.; Malik, W.; Youping, Z.; Abid, M.A.; Hailiang, C.; Qiuhong, Y.; Guoli, S. Recent Insights in Cotton Functional Genomics: Progress and Future Perspectives. *Plant Biotechnol. J.* **2017**. [[CrossRef](#)] [[PubMed](#)]
34. Tian, X.; Ruan, J.; Huang, J.; Fang, X.; Mao, Y.; Wang, L.; Chen, X.; Yang, C. Gossypol: Phytoalexin of cotton. *Sci. China Life Sci.* **2016**, *59*, 122–129. [[CrossRef](#)] [[PubMed](#)]
35. Cai, Y.; Xie, Y.; Liu, J. Glandless seed and glanded plant research in cotton. A review. *Agron. Sustain. Dev.* **2010**, *30*, 181–190. [[CrossRef](#)]
36. Li, W.; Liu, W.; Wei, H.; He, Q.; Chen, J.; Zhang, B.; Zhu, S. Species-specific expansion and molecular evolution of the 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) gene family in plants. *PLoS ONE* **2014**, *9*, e94172. [[CrossRef](#)] [[PubMed](#)]
37. Yao, H.; Guo, L.; Fu, Y.; Borsuk, L.A.; Wen, T.J.; Skibbe, D.S.; Cui, X.; Scheffler, B.E.; Cao, J.; Emrich, S.J. Evaluation of five ab initio gene prediction programs for the discovery of maize genes. *Plant Mol. Biol.* **2005**, *57*, 445–460. [[CrossRef](#)] [[PubMed](#)]
38. Campos, N.; Boronat, A. Targeting and topology in the membrane of plant 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Plant Cell* **1995**, *7*, 2163–2174. [[CrossRef](#)] [[PubMed](#)]
39. Leivar, P.; Gonzalez, V.M.; Castel, S.; Trelease, R.N.; Lopez-Iglesias, C.; Arro, M.; Boronat, A.; Campos, N.; Ferrer, A.; Fernandez-Busquets, X. Subcellular localization of *Arabidopsis* 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Plant Physiol.* **2005**, *137*, 57–69. [[CrossRef](#)] [[PubMed](#)]
40. Istvan, E.S.; Deisenhofer, J. The structure of the catalytic portion of human HMG-CoA reductase. *Biochim. Biophys. Acta* **2000**, *1529*, 9–18. [[CrossRef](#)]
41. Darabi, M.; Izadi-Darbandi, A.; Masoudi-Nejad, A.; Naghavi, M.R.; Nemat-Zadeh, G. Bioinformatics study of the 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) gene in Gramineae. *Mol. Biol. Rep.* **2012**, *39*, 8925–8935. [[CrossRef](#)] [[PubMed](#)]



42. Joost, O.; Bianchini, G.; Bell, A.A.; Benedict, C.R.; Magill, C.W. Differential induction of 3-hydroxy-3-methylglutaryl CoA reductase in two cotton species following inoculation with *Verticillium*. *Mol. Plant Microbe Interact.* **1995**, *8*, 880–885. [[CrossRef](#)] [[PubMed](#)]
43. Hurles, M. Gene duplication: The genomic trade in spare parts. *PLoS Biol.* **2004**, *2*, e206. [[CrossRef](#)] [[PubMed](#)]
44. Freeling, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **2009**, *60*, 433–453. [[CrossRef](#)] [[PubMed](#)]
45. Du, H.; Yang, S.S.; Liang, Z.; Feng, B.R.; Liu, L.; Huang, Y.B.; Tang, Y.X. Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol.* **2012**, *12*, 106. [[CrossRef](#)] [[PubMed](#)]
46. He, H.; Dong, Q.; Shao, Y.; Jiang, H.; Zhu, S.; Cheng, B.; Xiang, Y. Genome-wide survey and characterization of the WRKY gene family in *Populus trichocarpa*. *Plant Cell Rep.* **2012**, *31*, 1199–1217. [[CrossRef](#)] [[PubMed](#)]
47. Yin, G.; Xu, H.; Xiao, S.; Qin, Y.; Li, Y.; Yan, Y.; Hu, Y. The large soybean (*Glycine max*) WRKY TF family expanded by segmental duplication events and subsequent divergent selection among subgroups. *BMC Plant Biol.* **2013**, *13*, 148. [[CrossRef](#)] [[PubMed](#)]
48. Chai, G.; Wang, Z.; Tang, X.; Yu, L.; Qi, G.; Wang, D.; Yan, X.; Kong, Y.; Zhou, G. R2R3-MYB gene pairs in *Populus*: Evolution and contribution to secondary wall formation and flowering time. *J. Exp. Bot.* **2014**, *65*, 4255–4269. [[CrossRef](#)] [[PubMed](#)]
49. Tholl, D.; Lee, S. Terpene Specialized Metabolism in *Arabidopsis thaliana*. In *Arabidopsis Book*; The American Society of Plant Biologists: Rockville, MD, USA, 2011; Volume 9, p. e0143.
50. Zhou, M.; Zhang, C.; Wu, Y.; Tang, Y. Metabolic engineering of gossypol in cotton. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 6159–6165. [[CrossRef](#)] [[PubMed](#)]
51. Fryxell, P.A. A Redefinition of the Tribe Gossypieae. *Bot. Gaz.* **1968**, *129*, 296–308. [[CrossRef](#)]
52. Lynch, M.; Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **2000**, *290*, 1151–1155. [[CrossRef](#)] [[PubMed](#)]
53. Loguercio, L.L.; Wilkins, T.A. Structural analysis of a hmg-coA-reductase pseudogene: Insights into evolutionary processes affecting the hmgr gene family in allotetraploid cotton (*Gossypium hirsutum* L.). *Curr. Genet.* **1998**, *34*, 241–249. [[CrossRef](#)] [[PubMed](#)]
54. Andersson, J.O.; Andersson, S.G. Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes. *Mol. Biol. Evol.* **2001**, *18*, 829–839. [[CrossRef](#)] [[PubMed](#)]
55. Pink, R.C.; Wicks, K.; Caley, D.P.; Punch, E.K.; Jacobs, L.; Francisco Carter, D.R. Pseudogenes: Pseudo-functional or key regulators in health and disease? *RNA* **2011**, *17*, 792–798. [[CrossRef](#)] [[PubMed](#)]
56. Xiao, J.; Sekhwal, M.K.; Li, P.; Ragupathy, R.; Cloutier, S.; Wang, X.; You, F.M. Pseudogenes and Their Genome-Wide Prediction in Plants. *Int. J. Mol. Sci.* **2016**, *17*, 1991. [[CrossRef](#)] [[PubMed](#)]
57. Senda, M.; Masuta, C.; Ohnishi, S.; Goto, K.; Kasai, A.; Sano, T.; Hong, J.S.; MacFarlane, S. Patterning of virus-infected *Glycine max* seed coat is associated with suppression of endogenous silencing of chalcone synthase genes. *Plant Cell* **2004**, *16*, 807–818. [[CrossRef](#)] [[PubMed](#)]
58. Kasai, A.; Kasai, K.; Yumoto, S.; Senda, M. Structural features of *GmIRCHS*, candidate of the I gene inhibiting seed coat pigmentation in soybean: Implications for inducing endogenous RNA silencing of chalcone synthase genes. *Plant Mol. Biol.* **2007**, *64*, 467–479. [[CrossRef](#)] [[PubMed](#)]
59. Tuteja, J.H.; Zabala, G.; Varala, K.; Hudson, M.; Vodkin, L.O. Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in *glycine max* seed coats. *Plant Cell* **2009**, *21*, 3063–3077. [[CrossRef](#)] [[PubMed](#)]
60. Polisenio, L.; Salmena, L.; Zhang, J.; Carver, B.; Haveman, W.J.; Pandolfi, P.P. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **2010**, *465*, 1033–1038. [[CrossRef](#)] [[PubMed](#)]
61. Nagegowda, D.A.; Gutensohn, M.; Wilkerson, C.G.; Dudareva, N. Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *Plant J.* **2008**, *55*, 224–239. [[CrossRef](#)] [[PubMed](#)]
62. Yue, Y.; Yu, R.; Fan, Y. Transcriptome profiling provides new insights into the formation of floral scent in *Hedychium coronarium*. *BMC Genom.* **2015**, *16*, 470. [[CrossRef](#)] [[PubMed](#)]
63. Nieuwenhuizen, N.J.; Wang, M.Y.; Matich, A.J.; Green, S.A.; Chen, X.; Yauk, Y.K.; Beuning, L.L.; Nagegowda, D.A.; Dudareva, N.; Atkinson, R.G. Two terpene synthases are responsible for the major sesquiterpenes emitted from the flowers of kiwifruit (*Actinidia deliciosa*). *J. Exp. Bot.* **2009**, *60*, 3203–3219. [[CrossRef](#)] [[PubMed](#)]

64. Vranova, E.; Coman, D.; Gruissem, W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* **2013**, *64*, 665–700. [[CrossRef](#)] [[PubMed](#)]
65. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2016**, *44*, D279–D285. [[CrossRef](#)] [[PubMed](#)]
66. Finn, R.D.; Attwood, T.K.; Babbitt, P.C.; Bateman, A.; Bork, P.; Bridge, A.J.; Chang, H.Y.; Dosztanyi, Z.; El-Gebali, S.; Fraser, M. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **2017**, *45*, D190–D199. [[CrossRef](#)] [[PubMed](#)]
67. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)] [[PubMed](#)]
68. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [[CrossRef](#)] [[PubMed](#)]
69. Liu, W.; Li, W.; He, Q.; Daud, M.K.; Chen, J.; Zhu, S. Characterization of 19 Genes Encoding Membrane-Bound Fatty Acid Desaturases and their Expression Profiles in *Gossypium raimondii* Under Low Temperature. *PLoS ONE* **2015**, *10*, e0123281. [[CrossRef](#)] [[PubMed](#)]
70. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [[CrossRef](#)] [[PubMed](#)]
71. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [[CrossRef](#)] [[PubMed](#)]
72. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)) Method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]

**Sample Availability:** Samples of the *Gossypium* species are available from the authors.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).