Genome-wide identification and analysis of GDSL-type esterases/lipases in watermelon (Citrullus lanatus)

Runsheng Ren (runshengren@163.com)
Jiangsu Academy of Agricultural Sciences https://orcid.org/0000-0003-0942-6203
Xingping Yang
Jiangsu Academy of Agricultural Sciences
Jinhua Xu
Jiangsu Academy of Agricultural Sciences
Man Zhang
Jiangsu Academy of Agricultural Sciences
Guang Liu
Jiangsu Academy of Agricultural Sciences
Xiefeng Yao
Jiangsu Academy of Agricultural Sciences

Research article

Keywords: Citrullus lanatus, GDSL family, Lipase, Phylogenetic analysis, Genome-wide identification, Nitrogen

Posted Date: July 11th, 2019

DOI: https://doi.org/10.21203/rs.2.11240/v1

Abstract

Background The GDSL esterase and lipase families play important roles in abiotic stress, pathogen defense, seed development and lipid metabolism. Identifying the lipase activity of a putative GDSL lipase is necessary to determine its function. Systematic analysis of the GDSL gene family is still lacking in Citrullus lanatus. Results In this study, we identified 65 watermelon GDSL-type esterase/lipase genes and divided these genes into 6 clades based on phylogeny. The phylogenetic relationship of watermelon GDSL genes compared with Arabidopsis thaliana GDSL esterases/lipases was also determined, and these genes were divided into four groups related to morphological development, abiotic stress response, pathogen defense, and secondary metabolism. The chromosomal location of these genes revealed that they are distributed unevenly across all 11 watermelon chromosomes. Analysis of duplication events suggested that segmental duplication and tandem duplication were the major driving forces of GDSL family evolution. Synteny analysis indicated that GDSLs in watermelon were highly homologous to those in Arabidopsis thaliana, melon and cucumber. Transcriptome analyses showed the tissue-specific and common expression of the GDSL genes in leaf and root tissues and identified nitrogen-related genes under low nitrogen (N) stress compared with optimal N conditions. Conclusions Our results provide a basis for selecting candidate watermelon GDSL genes for further studies to determine the biological functions of the GDSL genes in watermelon.

Background

The family of GDSL lipases/esterases is a large conserved family and is widely present in all plants, animals, and microorganisms. Unlike the lipases/esterases with the GXSXG motif family, whose active serine site is located near the center of the conserved sequence, the catalytic triad in the GDSL-motif-like family was constituted by three highly conserved amino acid residues (i.e., serine, aspartic acid and histidine) and is located near the N- terminus [1-4].

Major advances have been reported that GDSL lipases/esterases are involved in various biological functions, such as development of seeds [5-9], deposition of epicuticular wax [10], biosynthesis of cutin [11, 12], hydration of pollen [13] and regulation of plant cell wall components [14]. Moreover, GDSL lipases are also involved in responses to various abiotic stresses, such as salt [15], drought [16], and freezing [17], as well as to biotic stresses, such as plant immune responses and pathogen defense [18-23].

Watermelon [*Citrullus lanatus* (Thunb) Mansfeld] of the *Cucurbitaceae* plant family is one of the most economically important vegetable crops in the world. Since the GDSL-motif-like family of lipases was first reported [3], considerable progress on GDSL esterases/lipases has been made in various plant species, including *Arabidopsis thaliana* [24], *Oryza sativa* [25, 26], *Brassica napus* [27], and *Rosaceae* species [28], but no GDSL-lipase members have been identified and functionally characterized in watermelon. The availability of whole-genome watermelon sequences offers an opportunity to search for GDSL-type lipase genes in watermelon. In this study, we performed a genome-wide analysis of GDSL-type lipase genes,

including genomic locations, chromosomal distributions, and evolutionary divergence. Additionally, transcriptome analysis also provided information on the identification of tissue-specific expression in response to low N stress. Taken together, these findings and analysis will provide a strong foundation for further studies on the roles of GDSL-type esterase/lipase genes in watermelon, and the comparative analysis between the GDSL-type esterase/lipase gene family from *Arabidopsis thaliana* and the other *Cucurbitaceae* crops will help to characterize the evolution of the GDSL-type esterase/lipase gene family species.

Results

Identification and characterization of the GDSL-type lipase genes in watermelon

Based on a hidden Markov model (HMM) search, a total of 65 GDSL-type esterase/lipase genes containing the GDSL domain were identified (Table 1) in the watermelon genome. Among these genes, the *ClCG09G018950* gene was identified as the smallest protein with 211 amino acids (aa), whereas the largest one was *ClCG02G019240* (1516 aa). The MW of the proteins ranged from 23.5 to 171.9 kDa, and the pI ranged from 4.99 (*ClCG07G011520*) to 40.8 (*ClCG02G015390*). The length of the watermelon GDSL coding sequence ranges between 633 and 4548 bp. The characteristics of all 65 watermelon GDSLs are listed in Tables 1 and S1.

Phylogenetic analysis of GDSL-type lipase genes

Based on the protein sequences, the phylogenetic analysis indicated that the 65 GDSL members were divided into six clades, corresponding to clades I, II, III, IV, V and VI (Fig. 1a). Among the 65 GDSL members, 16 belong to clade I, 4 to clade II, 4 to clade III, 6 to clade IV, 2 to clade V and 33 to clade VI. For comparative purposes, we further included comparatively well-characterized GDSL genes from model plant species *Arabidopsis thaliana* into a second phylogenetic tree, and the combined phyto tree could be divided into four groups (Fig. 1b). Group I (blue) of the combined phyto tree harbored almost half (32 genes) of the total watermelon GDSL genes, and most of the *Arabidopsis* GDSL genes (65 genes) grouped inside this group as well. Group II consists of 22 GDSL genes, including 8 from watermelon and 14 from *Arabidopsis*. Group III, containing 15 watermelon GDSL genes, clustered with 24 GDSL genes from *Arabidopsis*. Group IV contains the most GDSL genes, including 32 from watermelon and 67 from *Arabidopsis*. Most of the watermelon GDSL genes grouped the same as the first phylogenetic tree with only watermelon GDSL genes, but two genes grouped separately from their original clades; for example, *ClCG02G019240* and *ClCG02G019240* were grouped in group VI but were grouped originally in clade I and clade II, respectively. There are 3 phylogenetic subgroups in group I, designated I-a, I-b and I-c. The detailed subgroups in each group are shown in Fig. 1.

Gene structure and motif analysis of the watermelon GDSL gene family

The MEME results indicated that exon-intron organizations of all the identified GDSL genes were considerably diverse. As shown in Fig. 2b, all watermelon GDSL genes possessed three to thirty exons,

and fifty-eight (89.3%) of the family contained more than four exons. Seven genes (10.7%) had three exons. Genes with only one exon were not observed. The detailed genomic locations of GDSL genes are shown in Fig. 2b. The length of the motifs ranged from 15 amino acids to 34 amino acids. The details of the conserved motifs are shown in Fig. 2a.

Chromosomal distribution and gene duplication analysis

According to the physical locations of the GDSL genes, we constructed a map on the distribution of the GDSL genes on the 11 watermelon chromosomes. Fig. 3 shows that the 65 GDSL genes were unevenly distributed, and most of the GDSL genes (41/65) were concentrated on chromosomes 1, 2, 9, and 10. Chromosome 2 had the highest number of GDSL genes (14 genes, 22% of mapped genes), whereas chromosome 6 had the lowest number (2 genes, 3% of mapped genes).

Two types of genomic duplication (tandem duplication and segmental duplication) were observed for watermelon GDSL genes (Fig. 3 and 4). A total of eight genes identified as duplicated genes were clustered into seven duplication events (*ClCG02G007920/ ClCG08G001570*, *ClCG02G007920/ ClCG02G015390*, *ClCG02G015390/ ClCG08G001570*, *ClCG09G016490/ ClCG09G016520*, *ClCG07G013430/ ClCG07G013470*, *ClCG09G016490/ ClCG10G000920* and *ClCG10G000920/ ClCG09G016520*) in the watermelon genome and were randomly distributed on chromosomes 2 (2 genes), 7 (2 genes), 8 (1 gene), 9 (2 genes), and 10 (1 gene) (Fig. 3 and 4). Two single tandem duplication events (*ClCG07G013430/ ClCG07G013470* and *ClCG09G016490/ ClCG09G016520*) were positioned on chromosomes 7 and 9. In addition to the tandem duplication events, five duplication events involved in segmental duplications were also observed, showing collinearity among chromosomes 2, 8, 9 and 10. Interestingly, the segmental duplication genes on *ClCG02G007920* and *ClCG02G015390* were collinear with the gene *ClCG08G001570* on chromosome 8, and the tandem duplication genes on chromosome 9, *ClCG09G016490* and *ClCG09G016520* were collinear with the gene on chromosome *ClCG10G000920* (Fig. 3 and 4).

To better understand the evolutionary constraints of the duplicated watermelon GDSL family, the Ka/Ks ratios of the GDSL gene pairs were calculated. The results showed that 6 duplicated gene pairs had Ka/Ks < 1, with 4 of them being even less than 0.5, suggesting that these watermelon GDSL genes might have been subject to strong purifying selective pressure during evolution. The duplication dates for the 6 duplication events were estimated to have occurred approximately between 8 and 60 Mya (Fig. 3 and 4 and Table S2).

Evolutionary analysis of GDSL genes in watermelon and other species

To further infer the phylogenetic mechanisms of watermelon GDSL family genes, three comparative syntenic maps of watermelon with three representative species, including *Arabidopsis*, melon and cucumber, were constructed (Fig. 5). There are 35 orthologous GDSL gene pairs obtained between watermelon and *Arabidopsis*, 46 between watermelon and melon and 48 between watermelon and cucumber. Some *Arabidopsis* GDSL genes (19 genes) were found to be syntenic to the same two or three

watermelon GDSL genes, only two were syntenic for cucumber and none was syntenic for melon (Fig. 5 and Table S3).

Expression profiling of watermelon GDSL genes under low N stress

To predict the possible functions of watermelon GDSL family genes, we analyzed the expression of the GDSL genes in leaves and roots treated with 0.2 mM and 9 mM N, respectively. The results revealed that the GDSL genes had diverse expression patterns in the leaf and root. Among the 65 GDSL gene members, twenty-seven genes were expressed in both the leaf and the root tissue, and some members, including *ClCG09G001270*, *ClCG02G016030*, *ClCG10G009690*, *ClCG05G025850*, *ClCG01G003090*, *ClCG02G001050*, *ClCG07G004140*, *ClCG01G023580*, *ClCG02G019240* and *ClCG10G019120,* showed the highest transcription level (Fig. 6 and Table S4), implying that these GDSL genes might play important roles in leaf and root development. Conversely, twenty-three genes displayed very low or could not be detected in either of the two tissues, suggesting that these genes might not play roles in the leaf and root tissues, although they might be primarily expressed in other tissues of watermelon not tested or under some special conditions. Some GDSL genes exhibited tissue-specific expression. For instance, the genes *ClCG08G000570*, *ClCG09G000290*, *ClCG07G013470* and *ClCG10G005280* were only expressed in leaves, while *ClCG04G009930*, *ClCG04G009920* and *ClCG05G011430* were expressed specifically in roots (Fig. 6 and Table S4).

Under the treatment of low concentrations of N, the results showed that the expression of some GDSL genes was significantly induced/repressed compared to the optimal treatment of N. In the leaves, fourteen GDSL genes (*ClCG06G003270*, *ClCG10G013760*, *ClCG08G000570*, *ClCG09G000290*, *ClCG11G010220*, *ClCG09G001270*, *ClCG09G020870*, *ClCG02G016030*, *ClCG10G009690*, *ClCG01G024110*, *ClCG02G001070*, *ClCG03G007300*, *ClCG05G025850* and *ClCG02G001050*) were repressed by the low concentration of N treatment. Interestingly, the transcript levels of many GDSL genes, such as *ClCG07G014350*, *ClCG01G020480*, *ClCG01G023600*, *ClCG10G005260* and *ClCG01G023580*, were upregulated by the low concentration of N treatment. In the root, the expression levels of seven genes (*ClCG06G003270*, *ClCG10G013760*, *ClCG09G001270*, *ClCG02G016030*, *ClCG10G009690*, *ClCG02G001050* and *ClCG10G005260*) were downregulated, whereas two genes (*ClCG07G014350* and *ClCG07G004140*) were upregulated (Fig. 6 and Table S4). The overall expression data analysis suggested that GDSL genes showed diverse expression patterns and might play crucial roles in leaf and root development in watermelon.

Discussion

The GDSL lipase/esterase family has been demonstrated to play multiple functional roles in developmental processes and in responses to abiotic and biotic stresses in plants. In the present study, a comprehensive set of 65 GDSL family genes was identified, and these genes were divided into 6 clades. As a model plant, extensive efforts have been made to functionally characterize the genes of *A. thaliana*. To speculate the possible functions of the GDSL genes identified in this study, we additionally performed

a phylogenetic analysis together with the GDSL genes in the model plant species, *Arabidopsis*, which can provide useful information regarding the possible roles of GDSL genes in watermelon based on their similarities between syntenic genes. The combined phylogenetic tree showed that the GDSL genes from *Arabidopsis* and watermelon were grouped into four main groups (Fig. 1b), which is consistent with the results of earlier studies conducted in Arabidopsis [40]. Group 1 contains 10 *Arabidopsis* and 10 watermelon GDSL genes, and most of these genes have no known function, except *AT2G38180*, which was reported to be involved in ethylene (ET) defense signaling pathways [41]. Accumulating evidence indicates that the plant GDSL esterase/lipases are also involved in secondary metabolism in plants. According to the phylogenetic tree, some of the genes in group 2 were also related to secondary metabolism; in *Arabidopsis*, for example, *AT1G54790* (seed fatty acid reducer, *SFAR1*) was reported to be involved in fatty acid (FA) metabolism in *Arabidopsis* seeds [9]. It has been reported that *AT3G48460* (*SFAR4*) is a GDSL-type esterase involved in fatty acid metabolism by reducing the fatty acid content during post germination and seedling development in *Arabidopsis* [8]. *AT1G67830* (*AtGELP33*) was reported to be related to xyloglucan metabolism and cell wall composition [9, 14]. These findings indicate that several genes in clade 2 are involved in some stress responses. Eleven members of group 3 were reported to be involved in plant resistance/immunity responses, namely, *AT5G40990* (*GLIP1*), *AT1G53940* (*GLIP2*), *AT1G53990* (*GLIP3*), *AT3G14225* (*GLIP4*), *AT1G53920* (*GLIP5*), *AT1G7112*0 (*GLIP6*) and *AT1G54030*, *AT1*G54020, *AT1G54010*, *AT1G54000* and *AT3G14210* [21, 42-45]. Among these genes, *AT5G40990* (*GLIP1*) is reported to regulate plant immunity through regulation of ethylene signaling, and regulation is mediated by its activity to accumulate a systemic signal(s) in the phloem [18, 46, 47]. The gene expression of *AT5G40990* (*GLIP1*), as well as *AT1G53990* (*GLIP3*) and *AT3G14225* (*GLIP4*), was regulated by two pathogen-responsive MAPKs, MPK3 and MPK6 [48]. *AT1G53940* (*GLIP2*) plays a role in plant immune responses and pathogen defense and is involved in the resistance to *Erwinia carotovora* via negative regulation of auxin signaling [19]. *AT1G54030* could cause organizational defects in the endoplasmic reticulum (ER) and aberrant protein trafficking in the plant secretory pathway [42]. The gene *AT3G14210* (Epithiospecifier modifier 1, *ESM1*) has been reported to suppress nitrile formation, increase isothiocyanate production, and correlate with plant resistance against herbivores [35]. In addition, it has been reported that the genes *AT5G40990* and *AT3G14210* are also related to the biotic stress response [18, 19, 35, 46, 47]. Moreover, *AT3G14210* and *AT3G14220* are tandem neighbors of *AT3G14225* [35]. These findings suggested that the watermelon GDSL genes classified into group 3 might be involved in plant resistance or immunity. For the genes in group 4, Arabidopsis genes *AT3G11210*, *AT2G38180* and *AT5G45920* are homologs of watermelon genes *ClCG07G004140*, *ClCG10G022120* and *ClCG01G023570*, respectively. *AT3G04290* was first reported to play a role in salt tolerance and may also be involved in defense reactions against pathogens [15]. In 2017, the gene *AT3G04290* was retrieved by a yeast two-hybrid screen using VACUOLELESS GAMETOPHYTES (*VLG*, *AT2G17740*) as bait, which is essential for the development of female and male gametophytes in Arabidopsis [49]. *AT1G58430* (*SFAR2*), *AT2G42990* (*SFAR3*) and *At4g18970* (*SFAR5*) have been demonstrated to act downstream of the GA signaling pathway and are also involved in fatty acid degradation in Arabidopsis seeds [9]. Moreover, *AT1G58430* (*SFAR2*) and *AT2G42990* (*SFAR3*) are also involved in important functions in plant development, morphogenesis, and glucose stress tolerance [9]. *AT5G45670* (*LIP1*) has been reported to

be specifically expressed in the epidermis and highly induced by GA and repressed by DELLAs during seed imbibition [50]. *AT5G45670* (*LIP1*) functions as a negative factor through its L1 box present in the LIP1 element for seed germination [51]. *At4g30140* (*CDEF1*) has cutinase activity, being secreted from cells and directly degrading the polyester in the cuticle, and it is also involved in the penetration of the stigma by pollen tubes and facilitating the emergence of the lateral roots [52, 53]. It has been reported that overexpression of the *AT1G29670* gene enhances seed germination and seedling establishment, suggesting that the gene could be a promising target to achieve the features of increased germination and higher oil content in plant breeding [54]. It has been reported that the gene *AT1G29660* is involved in phloem-mediated long-distance signaling regulating responses to biotic and abiotic stress [55-57]. It has been reported that the genes *AT5G18430* and *AT5G33370* had approximately 70% identity and over 80% similarity of their amino acid sequences with LTL1, which functions as a GDSL-motif lipase and was associated with salt resistance [15, 49]. *AT1G75910* (*AtGELP42*) functions as an extracellular lipase to facilitate pollen hydration on the stigma in the early pollination stage of Arabidopsis [13]. It has been reported that the gene *AT1G75930* plays a role in efficient pollination [13] and that the gene *AT1G75930* (*EXL6*) is a target of a key transcription factor that coordinates pollen wall development and sporopollenin biosynthesis in Arabidopsis [58]. It has been reported that the gene *AT1G75930* plays a role in pollen exine formation and is essential for pollen development in Arabidopsis [27].

In view of the importance of nitrogen (N) as the primary inorganic nutrient in plant growth and development, especially for crops requiring large quantities of fertilizers, such as watermelon, and the key roles of GDSL lipases in regulating plant growth and development, the expression patterns of GDSL genes in the leaf and root of watermelon under optimal nitrogen (ON) and low nitrogen (LN) conditions were investigated in this study based on the available transcriptome data published previously [39]. According to the analysis of gene expression profiling, the watermelon GDSL genes showed diverse expression patterns (Fig. 6). The transcriptome data showed that five GDSL genes (*CLCG04G009920*, *CLCG02G001070*, *CLCG04G009930*, *CLCG01G024110* and *CLCG11G010220*) are expressed only in leaf tissue, four GDSL genes (*CLCG05G006600*, *CLCG01G023460*, *CLCG04G009910* and *CLCG10G005280*) are expressed only in root tissue, and approximately 21 GDSL genes are highly expressed in both leaf and root tissue. Among these genes, the Arabidopsis homolog *AT2G23540* for *ClCG02G013150* was reported to be highly expressed in leaf and root tissues [59], which was also observed for the tomato homolog of *Solyc02g090210* [60]. The Arabidopsis homolog gene *AT3G04290* for the gene *CLCG11G010220* has similar expression patterns and is expressed mainly in leaf and flower tissues [61]. Functionally, studies have demonstrated that the GDSL lipase plays a role in salt tolerance [15] and is also involved in defense reactions against pathogens [15, 62]. It has been reported that the gene *AT3G04290* may also play a role in cell wall differentiation and plant growth in the Arabidopsis response to ionizing radiation [63]. In contrast, the Arabidopsis homolog gene *AT5G55050* for the gene *CLCG10G005280* was also mainly expressed in the root tissue and significantly ($P = 0.03$) induced by > 3-fold (normalized) after 6 h of exposure of plants to allelochemicals identified in buckwheat (fagomine, gallic acid, or rutin) in the aquaculture medium [64].

Many researchers have reported that low nitrogen stress has comprehensive impacts on genes involved in various biosynthetic, catabolic and regulatory processes and thus severely inhibits plant growth and development [65-67]. Previous transcriptome data revealed that GDSL genes were also related to low nitrogen stress. For example, under LN stress, the GDSL gene *GRMZM2G034958* was only detected in cobs, and *GRMZM2G046306* and *GRMZM2G015708* were only detected in florets [68], suggesting that these three GDSL genes have negative roles in nitrogen-related metabolic processes. In the present study, the expression profiles of genes from group 2 did not show a significant change in their expression fold under low N stress, both in the leaf and root tissues. However, many members of group 4 show differential expression under the low N stress treatment (Fig. 6), implying the possible role of the genes from group 4 in plant growth and development. Notably, 5 GDSL genes, *ClCG02G013150*, *ClCG11G010220*, *ClCG02G006480*, *ClCG01G023460* and *ClCG10G005280,* had a significant change in their expression fold in the leaf and/or root under LN, which suggested that the genes played important roles in responses to low nitrogen stress. The expression of the GDSL gene *ClCG02G013150* in leaves was downregulated by low N, and its homolog in Arabidopsis, *AT2G23540*, was also downregulated by the stress of 2,4,6-trichlorophenol (2,4,6-TCP) [69]. A functional study demonstrated that the Arabidopsis homolog *AT2G23540* plays an important role in cell expansion and cuticle deposition in response to stresses [70]. These findings suggested that the watermelon *ClCG02G013150* negatively regulates low nitrogen tolerance and thus has a potential value in watermelon stress-resistance improvement. The Arabidopsis homolog GDSL genes in other groups were also reported to be involved in low-nitrogen stress. For example, for the GDSL gene *AT1G54010* in group 3, the expression level was significantly upregulated in response to both short- and long-term N availability increases [71], suggesting that it plays a key role in relating to the regulatory network for plant N responses. The two Arabidopsis GDSL genes *AT1G28570* and *AT1G28600* in group 2 were first upregulated under the severe N-limiting condition and then downregulated after the long-term N availability increase [71]. According to the transcriptome data, the gene expression patterns provide valuable clues on the possible functions of the watermelon GDSL genes in relation to nitrogen use. In summary, our study will help to elucidate the basics of GDSL information and provide a solid basis for the further investigation of the biological functions of GDSL genes in watermelon.

Conclusions

In summary, the present study identified 65 GDSL-type esterase/lipase genes in *C. lanatus*. Their gene structure, chromosomal location and phylogenetic analyses were performed, which will provide basic information for the functional characterization of GDSL genes in watermelon. RNA-seq data revealed that tissue-specific and common expression of the GDSL genes in leaf and root tissues, suggesting that the GDSL genes had clear function differentiation watermelon. The expression profiling under low N and optimal N conditions showed that some GDSL genes were significantly upregulated or downregulated, indicating their important roles in nitrogen related growth and development of watermelon. Overall, these data are useful for the follow-up study of the functional characteristics of GDSL genes in watermelon.

Methods

Identification of the GDSL-type lipase gene family in watermelon and chromosomal distribution

The genomic data of watermelon (*C. lanatus*) were downloaded from CuGenDB (Version 2.0, ftp://cucurbitgenomics.org/pub/cucurbit/genome/watermelon/WCG/). HMMER searches were first carried out (1e-3 as E-value cut-off.) in watermelon protein sequences using the GDSL domain Hidden Markov Model (HMM) profile (PF00657) downloaded from Pfam (http://pfam. xfam.org/) with a default e value threshold of 0.1 [29], then the complete GDSL family genes in watermelon were identified (0.1 as e-value cut-off.) with the new watermelon-specific HMM file as query using the “hmmbuild” module by HMMER V3.0 program. The GDSL family genes were mapped to watermelon chromosomes based on their physical location information from the watermelon genome database using Circos [30]. Gene characteristics, including the length of the coding sequence (CDS), the protein molecular weight (MW), and the isoelectric point (pI), were calculated by ExPASy (http://www.expasy.org/). The subcellular localization was predicted using the CELLO v2.5 server (http://cello.life. nctu.edu.tw/).

Conserved motifs, gene structures and phylogenetic analysis

The Multiple Expectation Maximization for Motif Elicitation program (MEME, http://meme.nbcr.net/meme/intro.html) [31] was used to identify conserved motifs based on the protein sequences of the identified genes. The parameters employed in the analysis were set with the minimum motif width, 6; maximum motif width, 50; and maximum number of motifs, 10. The online program Gene Structure Display Server (GSDS: http://gsds.cbi.pku.edu.cn) [32] was used to display the exon-intron organization of the GDSL genes based on the data from the genome annotation file. Multiple sequence alignments of the watermelon GDSL protein sequences were performed using the ClustalW program [33]. Phylogenetic trees were constructed using the Molecular Evolutionary Genetics Analysis (MEGA 7.0) with the maximum-likelihood (ML) method, 1000 repetitions of bootstrap value and Poisson model [34].

Identification of duplicated GDSL genes and nonsynonymous/synonymous substitution (Ka/Ks) ratios of gene pairs in watermelon

Gene duplication events were determined on the basis of multiple sequence alignments using ClustalW with the following criteria: the shorter sequences cover > 75% of the longer sequence after alignment, and the similarity of aligned regions is > 75%. Gaps in the alignments were manually removed by Bioedit. The nonsynonymous (Ka) and synonymous (Ks) values of the duplicated GDSL gene pairs were calculated by the program KaKs_Calculator [35]. The Ks values were used to estimate the approximate date of the duplication time (T = Ks/ (2 × 6.5 × 10−9) × 10−6 Mya), and the Ka/Ks ratio was used to show the selection pressure for the duplicate gene pairs [36].

Analysis of syntenic relationships of GDSL genes between watermelon, A. thaliana and other major cucurbits crops

To understand the evolutionary relationships of the orthologous GDSL family genes between watermelon, *A. thaliana*, melon and cucumber genomes, the MCscan program [37] was employed to identify orthologous regions with default parameters. The genomic and annotation data of melon (version 3.6.1) and cucumber (version 3) were downloaded from the Cucurbit Genomics Database (CuGenDB) (http://cucurbitgenomics.org/), and those of *Arabidopsis* were downloaded from the *Arabidopsis thaliana* Plant Genome Database (AtPGD; http://plantgdb.org/AtGDB/). The identification of the GDSL family genes was performed following the same procedures described above. The synteny relationship of the orthologous GDSL genes obtained between watermelon and other selected species was visualized using Circos [30] and TBtools software [38].

Expression analysis of GDSL family genes using transcriptome data

To explore the expression profiles of GDSL family genes, one set of RNA-Seq data that included 12 samples was utilized to draw heat maps according to fragments per kilobase per million mapped reads (FPKM). The RNA-seq experiment measured the transcriptome response of leaves and roots in response to low (0.2 mM) and high (9 mM) concentrations of nitrogen (N) in watermelon. Three biological replicates were performed, and the RNA-seq was run using an Illumina HiSeq 2000 paired end sequencing platform. The RNA-seq data were downloaded in “fastq” format from the public database (https://www.ebi.ac.uk/), and the accession number of the study was PRJNA422970, which included 12 run accessions (SRR6389278-SRR6389289). Details about the transcriptome data and analysis of watermelon leaves and roots under low nitrogen were described and carried out in a previous study [39].

Declarations

Acknowledgments

All the authors are grateful for the raw data download from the public database. We sincerely thank the Department of Biological Sciences, 385 Serra Mall, Stanford University, for providing access to the *Arabidopsis thaliana* data. We sincerely thank the Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, Barcelona, Spain, for providing access to the *Cucumis melo* data. We sincerely thank the Key Laboratory of Horticultural Crops Genetic Improvement of Ministry of Agriculture, Sino-Dutch Joint Lab of Horticultural Genomics Technology, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China, for providing access to the *Cucumis sativus* data. We sincerely thank the Key Laboratory of Horticultural Plant Biology, Ministry of Education/College of Horticulture and Forestry Sciences, Huazhong Agricultural University, for providing access to the RNA-seq data of *C. lanatus*.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No. 31601777) and the Jiangsu Provincial Support Program for Breeding New Varieties (PZCZ201716).

Authors' contributions

RSR, XPY and JXX conceived and supervised the research design. RSR, XFY, MZ and GL designed the research and analyzed the data. RRS, XPY and JXX drafted and modified the manuscript. All authors approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. Remington S, Franken S, Sussman J, Frolow F, Ollis D, Verschueren K, Cheah E, Cygler M, Dijkstra B, Harel M: **The alpha/beta hydrolase fold**. *Protein Eng* 1992, **5**:197-211.

2. Arpigny JL, JAEGER K-E: **Bacterial lipolytic enzymes: classification and properties**. *Biochemical journal* 1999, **343**(1):177-183.

3. Upton C, Buckley JT: **A new family of lipolytic enzymes?** *Trends in biochemical sciences* 1995, **20**(5):178-179.

4. Mølgaard A, Kauppinen S, Larsen S: **Rhamnogalacturonan acetylesterase elucidates the structure and function of a new family of hydrolases**. *Structure* 2000, **8**(4):373-383.

5. Clauß K, von Roepenack-Lahaye E, Böttcher C, Roth MR, Welti R, Erban A, Kopka J, Scheel D, Milkowski C, Strack D: **Overexpression of sinapine esterase BnSCE3 in oilseed rape seeds triggers global changes in seed metabolism**. *Plant physiology* 2011, **155**(3):1127-1145.

6. Barros M, Fleuri L, Macedo G: **Seed lipases: sources, applications and properties-a review**. *Brazilian Journal of Chemical Engineering* 2010, **27**(1):15-29.

7. Ejedegba B, Onyeneke E, Oviasogie P: **Characteristics of lipase isolated from coconut (Cocos nucifera linn) seed under different nutrient treatments**. *African Journal of Biotechnology* 2007, **6**(6).

8. Huang L-M, Lai C-P, Chen L-FO, Chan M-T, Shaw J-F: **Arabidopsis SFAR4 is a novel GDSL-type esterase involved in fatty acid degradation and glucose tolerance**. *Botanical studies* 2015, **56**(1):33.

9. Chen M, Du X, Zhu Y, Wang Z, Hua S, Li Z, Guo W, Zhang G, Peng J, Jiang L: **Seed Fatty Acid Reducer acts downstream of gibberellin signalling pathway to lower seed fatty acid storage in Arabidopsis**. *Plant, cell & environment* 2012, **35**(12):2155-2169.

10. Broun P, Poindexter P, Osborne E, Jiang C-Z, Riechmann JL: **WIN1, a transcriptional activator of epidermal wax accumulation in Arabidopsis**. *Proceedings of the National Academy of Sciences* 2004, **101**(13):4706-4711.

11. Kannangara R, Branigan C, Liu Y, Penfield T, Rao V, Mouille G, Höfte H, Pauly M, Riechmann JL, Broun P: **The transcription factor WIN1/SHN1 regulates cutin biosynthesis in Arabidopsis thaliana**. *The Plant Cell* 2007, **19**(4):1278-1294.

12. Takahashi K, Shimada T, Kondo M, Tamai A, Mori M, Nishimura M, Hara-Nishimura I: **Ectopic expression of an esterase, which is a candidate for the unidentified plant cutinase, causes cuticular defects in Arabidopsis thaliana**. *Plant and cell physiology* 2009, **51**(1):123-131.

13. Updegraff EP, Zhao F, Preuss D: **The extracellular lipase EXL4 is required for efficient hydration of Arabidopsis pollen**. *Sexual plant reproduction* 2009, **22**(3):197-204.

14. de la Torre F, Sampedro J, Zarra I, Revilla G: **AtFXG1, an Arabidopsis gene encoding α-L-fucosidase active against fucosylated xyloglucan oligosaccharides**. *Plant Physiology* 2002, **128**(1):247-255.

15. Naranjo MA, Forment J, RoldÁN M, Serrano R, Vicente O: **Overexpression of Arabidopsis thaliana LTL1, a salt-induced gene encoding a GDSL-motif lipase, increases salt tolerance in yeast and transgenic plants**. *Plant, cell & environment* 2006, **29**(10):1890-1900.

16. Katagiri T, Takahashi S, Shinozaki K: **Involvement of a novel Arabidopsis phospholipase D, AtPLDδ, in dehydration-inducible accumulation of phosphatidic acid in stress signalling**. *The Plant Journal* 2001, **26**(6):595-605.

17. Welti R, Li W, Li M, Sang Y, Biesiada H, Zhou H-E, Rajashekar C, Williams TD, Wang X: **Profiling membrane lipids in plant stress responses role of phospholipase Dα in freezing-induced lipid changes in Arabidopsis**. *Journal of Biological Chemistry* 2002, **277**(35):31994-32002.

18. Kwon SJ, Jin HC, Lee S, Nam MH, Chung JH, Kwon SI, Ryu CM, Park OK: **GDSL lipase-like 1 regulates systemic resistance associated with ethylene signaling in Arabidopsis**. *The Plant Journal* 2009, **58**(2):235-245.

19. Lee D, Kim B, Sj, Jin H, Park O: **Arabidopsis GDSL lipase 2 plays a role in pathogen defense via negative regulation of auxin signaling**. *Biochemical & Biophysical Research Communications* 2009, **379**(4):1038-1042.

20. La Camera S, Geoffroy P, Samaha H, Ndiaye A, Rahim G, Legrand M, Heitz T: **A pathogen-inducible patatin-like lipid acyl hydrolase facilitates fungal and bacterial host colonization in Arabidopsis**. *The Plant Journal* 2005, **44**(5):810-825.

21. Oh IS, Park AR, Bae MS, Kwon SJ, Kim YS, Lee JE, Kang NY, Lee S, Cheong H, Park OK: **Secretome analysis reveals an Arabidopsis lipase involved in defense against Alternaria brassicicola**. *The Plant Cell* 2005, **17**(10):2832-2847.

22. Jakab G, Manrique A, Zimmerli L, Métraux J-P, Mauch-Mani B: **Molecular characterization of a novel lipase-like pathogen-inducible gene family of Arabidopsis**. *Plant Physiology* 2003, **132**(4):2230-2239.

23. Hong JK, Choi HW, Hwang IS, Kim DS, Kim NH, Du SC, Kim YJ, Hwang BK: **Function of a novel GDSL-type pepper lipase gene, CaGLIP1, in disease susceptibility and abiotic stress tolerance**. *Planta* 2008, **227**(3):539-558.

24. Ling H: **Sequence analysis of GDSL lipase gene family in Arabidopsis thaliana**. *Pakistan Journal of Biological Sciences* 2008, **11**(5):763.

25. Jiang Y, Chen R, Dong J, Xu Z, Gao X: **Analysis of GDSL lipase (GLIP) family genes in rice (Oryza sativa)**. *Plant omics* 2012, **5**(4):351.

26. Chepyshko H, Lai C-P, Huang L-M, Liu J-H, Shaw J-F: **Multifunctionality and diversity of GDSL esterase/lipase gene family in rice (Oryza sativa L. japonica) genome: new insights from bioinformatics analysis**. *BMC genomics* 2012, **13**(1):309.

27. Dong X, Yi H, Han C-T, Nou I-S, Hur Y: **GDSL esterase/lipase genes in Brassica rapa L.: genome-wide identification and expression analysis**. *Molecular genetics and genomics* 2016, **291**(2):531-542.

28. Cao Y, Han Y, Meng D, Abdullah M, Yu J, Li D, Jin Q, Lin Y, Cai Y: **Expansion and evolutionary patterns of GDSL-type esterases/lipases in Rosaceae genomes**. *Functional & integrative genomics* 2018, **18**(6):673-684.

29. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching**. *Nucleic acids research* 2011, **39**(suppl_2):W29-W37.

30. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics**. *Genome research* 2009, **19**(9):1639-1645.

31. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching**. *Nucleic acids research* 2009, **37**(suppl_2):W202-W208.

32. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server**. *Bioinformatics* 2014, **31**(8):1296-1297.

33. Li K-B: **ClustalW-MPI: ClustalW analysis using distributed and parallel computing**. *Bioinformatics* 2003, **19**(12):1585-1586.

34. Tamura K, Stecher G, Peterson D, Kumar S, Mac O: **About MEGA**. 2011.

35. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J: **KaKs_Calculator: calculating Ka and Ks through model selection and model averaging**. *Genomics, proteomics & bioinformatics* 2006, **4**(4):259-263.

36. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**(5494):1151-1155.

37. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity**. *Nucleic acids research* 2012, **40**(7):e49-e49.

38. Chen C, Xia R, Chen H, He Y: **TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface**. *BioRxiv* 2018:289660.

39. Nawaz MA, Chen C, Shireen F, Zheng Z, Sohail H, Afzal M, Ali MA, Bie Z, Huang Y: **Genome-wide expression profiling of leaves and roots of watermelon in response to low nitrogen**. *BMC genomics* 2018, **19**(1):456.

40. Lai C-P, Huang L-M, Chen L-FO, Chan M-T, Shaw J-F: **Genome-wide analysis of GDSL-type esterases/lipases in Arabidopsis**. *Plant molecular biology* 2017, **95**(1-2):181-197.

41. Brodersen P, Petersen M, Bjørn Nielsen H, Zhu S, Newman MA, Shokat KM, Rietz S, Parker J, Mundy J: **Arabidopsis MAP kinase 4 regulates salicylic acid-and jasmonic acid/ethylene-dependent responses via EDS1 and PAD4**. *The Plant Journal* 2006, **47**(4):532-546.

42. Lucia M, Giovanni S, Kentaro T, Chris H, Luciana R, Held MA, Federica B: **A missense mutation in the vacuolar protein GOLD36 causes organizational defects in the ER and aberrant protein trafficking in the plant secretory pathway**. *Plant Journal* 2010, **63**(6):901.

43. Hernández-Blanco C, Feng DX, Hu J, Sánchez-Vallet A, Deslandes L, Llorente F, Berrocal-Lobo M, Keller H, Barlet X, Sánchez-Rodríguez C: **Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance**. *The Plant Cell* 2007, **19**(3):890-903.

44. Rajjou L, Belghazi M, Huguet R, Robin C, Moreau A, Job C, Job D: **Proteomic investigation of the effect of salicylic acid on Arabidopsis seed germination and establishment of early defense mechanisms**. *Plant physiology* 2006, **141**(3):910-923.

45. Danon A, Miersch O, Felix G, op den Camp RG, Apel K: **Concurrent activation of cell death-regulating signaling pathways by singlet oxygen in Arabidopsis thaliana**. *The Plant Journal* 2005, **41**(1):68-80.

46. Kim HJ, Hong SH, Kim YW, Lee IH, Jun JH, Phee B-K, Rupak T, Jeong H, Lee Y, Hong BS: **Gene regulatory cascade of senescence-associated NAC transcription factors activated by ETHYLENE-INSENSITIVE2-mediated leaf senescence signalling in Arabidopsis**. *Journal of experimental botany* 2014, **65**(14):4023-4036.

47. Kim J-G, Stork W, Mudgett MB: **Xanthomonas type III effector XopD desumoylates tomato transcription factor SlERF4 to suppress ethylene responses and promote pathogen growth**. *Cell host & microbe* 2013, **13**(2):143-154.

48. Han X, Li S, Zhang M, Yang L, Liu Y, Xu J, Zhang S: **Regulation of GDSL Lipase Gene Expression by the MPK3/MPK6 Cascade and Its Downstream WRKY Transcription Factors in Arabidopsis Immunity**.

Molecular Plant-Microbe Interactions 2019:MPMI-06-18-0171-R.

49. D'ippólito S, Arias LA, Casalongué CA, Pagnussat GC, Fiol DF: **The DC 1-domain protein VACUOLELESS GAMETOPHYTES is essential for development of female and male gametophytes in Arabidopsis**. *The Plant Journal* 2017, **90**(2):261-275.

50. Castrillo G, Turck F, Leveugle M, Lecharny A, Carbonero P, Coupland G, Pazares J, Oñatesánchez L: **Speeding Cis - Trans Regulation Discovery by Phylogenomic Analyses Coupled with Screenings of an Arrayed Library of Arabidopsis Transcription Factors**. *Plos One* 2011, **6**(6):e21524.

51. Belén RC, Paloma RR, Raquel IF, Pilar C, Luis OAS: **Arabidopsis DELLA and two HD-ZIP transcription factors regulate GA signaling in the epidermis through the L1 box cis-element**. *Plant Cell* 2014, **26**(7):2905.

52. Kram BW, Bainbridge EA, Perera MADN, Carter C: **Identification, cloning and characterization of a GDSL lipase secreted into the nectar of Jacaranda mimosifolia**. *Plant Molecular Biology* 2008, **68**(1-2):173.

53. Kentaro T, Tomoo S, Maki K, Atsushi T, Masashi M, Mikio N, Ikuko HN: **Ectopic expression of an esterase, which is a candidate for the unidentified plant cutinase, causes cuticular defects in Arabidopsis thaliana**. *Plant & Cell Physiology* 2010, **51**(1):123-131.

54. Ding Z, Li S, An X, Liu X, Qin H, Wang D: **Transgenic expression of MYB15 confers enhanced sensitivity to abscisic acid and improved drought tolerance in Arabidopsis thaliana**. *Journal of Genetics & Genomics* 2009, **36**(1):17-29.

55. Akoh CC, Lee GC, Liaw YC, Huang TH, Shaw JF: **GDSL family of serine esterases/lipases**. *Progress in Lipid Research* 2004, **43**(6):534-552.

56. Huang D, Wu W, Abrams SR, Cutler AJ: **The relationship of drought-related gene expression in Arabidopsis thaliana to hormonal and environmental factors**. *Journal of Experimental Botany* 2008, **59**(11):2991-3007.

57. Breitenbach HH, Wenig M, Wittek F, Jordã L, Maldonado-Alconada AM, Sarioglu H, Colby T, Knappe C, Bichlmeier M, Pabst E: **Contrasting Roles of the Apoplastic Aspartyl Protease APOPLASTIC, ENHANCED DISEASE SUSCEPTIBILITY1-DEPENDENT1 and LEGUME LECTIN-LIKE PROTEIN1 in Arabidopsis Systemic Acquired Resistance**. *Plant Physiology* 2014, **165**(2):791-809.

58. Xu J, Zhang D: **ABORTED MICROSPORES Acts as a Master Regulator of Pollen Wall Formation in Arabidopsis**. *Plant Cell* 2014, **26**(4):1544.

59. Soler M, Serra O, Molinas M, Huguet G, Fluch S, Figueras M: **A genomic approach to suberin biosynthesis and cork differentiation**. *Plant Physiology* 2007, **144**(1):419-431.

60. Lashbrooke JG, Cohen H, Levy-Samocha D, Tzfadia O, Panizel I, Zeisler V, Massalha H, Stern A, Trainotti L, Schreiber L: **MYB107 and MYB9 Homologs Regulate Suberin Deposition in Angiosperms**. *Plant Cell* 2016, **28**(9):tpc.00490.02016.

61. Markus S, N Henriette U, Fran?Ois G, Monika D, Ray B, Detlef W, Lohmann JU: **Dissection of floral induction pathways using global expression analysis**. *Development* 2003, **130**(24):6001-6012.

62. Ascencio-Ibanez J, Sozzani R, Lee T, Chu T, Wolfinger R, Cella R, Hanley-Bowdoin L: **Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection**. *Plant Physiology* 2008, **148**(1):436-454.

63. Gicquel M, Taconnat L, Renou JP, Esnault MA, Cabello-Hurtado F: **Kinetic transcriptomic approach revealed metabolic pathways and genotoxic-related changes implied in the Arabidopsis response to ionising radiations**. *Plant Science* 2012, **195**(3):106-119.

64. Golisz A, Sugano M, Fujii Y: **Microarray expression profiling of Arabidopsis thaliana L. in response to allelochemicals identified in buckwheat**. *Journal of Experimental Botany* 2008, **59**(11):3099.

65. Hermans C, Hammond JPWhite PJ, Verbruggen N: **How do plants respond to nutrient shortage by biomass allocation?** *Trends in Plant Science* 2006, **11**(12):610-617.

66. Seebauer JR, Moose SP, Fabbri BJ, Crossland LD, Below FE: **Amino acid metabolism in maize earshoots. Implications for assimilate preconditioning and nitrogen signaling**. *Plant Physiology* 2004, **136**(4):4326-4334.

67. Jiman K, Turano FJ: **The putative glutamate receptor 1.1 (AtGLR1.1) functions as a regulator of carbon and nitrogen metabolism in Arabidopsis thaliana**. *Proc Natl Acad Sci U S A* 2003, **100**(11):6872-6877.

68. Pan X, Hasan MM, Li Y, Liao C, Zheng H, Liu R, Li X: **Asymmetric transcriptomic signatures between the cob and florets in the maize ear under optimal- and low-nitrogen conditions at silking, and functional characterization of amino acid transporters ZmAAP4 and ZmVAAT3**. *Journal of Experimental Botany* 2015, **66**(20):6149.

69. Li Z, Zhu B, Wang B, Gao J, Fu X, Yao Q: **Stress responses to trichlorophenol in Arabidopsis and integrative analysis of alteration in transcriptional profiling from microarray**. *Gene* 2015, **555**(2):159-168.

70. Cominelli E, Sala T, Calvi D, Gusmaroli G, Tonelli C: **Over-expression of the Arabidopsis AtMYB41 gene alters cell expansion and leaf surface permeability**. *The Plant Journal* 2008, **53**(1):53-64.

71. Bi Y, Ehirchiou D, Kilts TM, Inkson CA, Embree MC, Sonoyama W, Li L, Leet AI, Seo BM, Zhang L: **Identification of tendon stem/progenitor cells and the role of the extracellular matrix in their niche**. *Nature Medicine* 2007, **13**(10):1219-1227.

Tables

Table 1. Characteristic features of 65 GDSL family genes identified in this study

Gene ID	Genomic Location	Strand	CDS (bp)	MW (kDa)	pI	Subcellular localization[a]
ClCG01G001490	1: 1473675 - 1480964	+	1275	47.8	8.45	PM
ClCG01G003090	1: 3110977 - 3120817	-	2118	80.4	9.26	PM
ClCG01G020480	1: 34609016 - 34614446	+	720	26	5.75	PM
ClCG01G023460	1: 36795291 - 36798632	+	699	26	8.13	PM
ClCG01G023470	1: 36800506 - 36802831	-	1116	41.2	8.77	E, PM
ClCG01G023570	1: 36868238 - 36872409	-	738	27.2	5.32	Cy
ClCG01G023580	1: 36874226 - 36879304	+	2751	99.6	5.41	PM
ClCG01G023600	1: 36888308 - 36893954	-	3018	112.5	6.86	PM
ClCG01G023610	1: 36901385 - 36909586	+	2169	80	6.55	PM
ClCG01G024110	1: 37265796 - 37267519	-	1116	41.3	8.78	E
ClCG02G001050	2: 1239835 - 1248840	-	1851	68.9	8.87	Cy
ClCG02G001060	2: 1258206 - 1261413	-	810	30.4	8.98	V
ClCG02G001070	2: 1263200 - 1266670	-	1206	45.5	8.74	PM, Cy
ClCG02G006480	2: 7437752 - 7440765	-	1101	40	8.51	PM
ClCG02G007920	2: 9672552 - 9674003	-	1092	40	5.47	PM
ClCG02G013150	2: 26932465 - 26938456	+	1077	39.2	8.74	E, Ch
ClCG02G014720	2: 29082953 - 29084058	-	855	31.6	6.47	E
ClCG02G015300	2: 29646890 - 29653512	-	1686	62.3	5.56	E
ClCG02G015310	2: 29660907 - 29665911	+	2136	79	7.12	E, PM
ClCG02G015390	2: 29675925 - 29677555	-	1095	40.8	9.73	M
ClCG02G016030	2: 30439128 - 30454759	+	2547	93.3	7.13	PM
ClCG02G019240	2: 33967983 - 33996640	+	4548	171.9	6.14	Cy, N
ClCG02G021780	2: 36218462 - 36225276	-	2835	104.7	7.71	PM
ClCG02G024260	2: 38555646 - 38556876	+	1071	39.6	7.02	PM
ClCG03G004160	3: 4468561 - 4470874	-	1089	40	8.11	E, PM
ClCG03G007300	3: 8278114 - 8283697	+	984	35.8	5.32	E
ClCG03G007970	3: 9032733 - 9037005	-	741	28.2	8.98	PM
ClCG04G009620	4: 24542727 - 24545038	+	1056	38	6.06	E
ClCG04G009910	4: 24882560 - 24885373	+	1065	39.6	8.84	E, PM
ClCG04G009920	4: 24890537 - 24892016	-	1026	37.7	9.4	V
ClCG04G009930	4: 24899843 - 24908875	+	1998	73.1	9.6	PM
ClCG05G006600	5: 6617465 - 6619336	+	1086	39.8	8.44	PM
ClCG05G011430	5: 13615957 - 13619465	-	1056	38.7	8.52	E
ClCG05G025850	5: 37197191 - 37200146	-	1161	42.8	9.1	E
ClCG06G003270	6: 3931263 - 3938492	+	1029	37.7	8.72	E
ClCG06G016170	6: 29424797 - 29425976	-	870	31.8	5.94	E, PM
ClCG07G004140	7: 4846344 - 4853502	+	876	32.5	5.59	E, Ch
ClCG07G011520	7: 27629710 - 27631721	-	1185	44.1	4.99	PM
ClCG07G013430	7: 29879654 - 29880929	-	870	31.1	8.64	E
ClCG07G013470	7: 29900762 - 29902037	-	870	31.1	8.64	E
ClCG07G014350	7: 30778356 - 30784816	+	1926	72.3	8.73	Cy
ClCG08G000570	8: 1422069 - 1438536	+	2145	79	8.8	PM
ClCG08G001570	8: 3060313 - 3061594	+	1095	40.2	5.84	E
ClCG08G014050	8: 26905821 - 26908208	+	747	27.5	5.45	E, PM
ClCG09G000290	9: 294849 - 297525	-	774	28	9.04	E
ClCG09G001270	9: 1213294 - 1216949	-	1026	38.6	9.24	V
ClCG09G010300	9: 10108916 - 10113871	-	1029	37.1	9.1	E
ClCG09G011050	9: 10897447 - 10900549	-	921	34.1	5.16	PM, Ch

ClCG09G016490	9: 32381359 - 32382916	+	1125	41.9	5.93	E, PM, V
ClCG09G016520	9: 32422001 - 32423531	-	1107	41.4	5.66	PM
ClCG09G018940	9: 36192146 - 36197797	-	1146	42.2	9.35	E
ClCG09G018950	9: 36199038 - 36200370	-	633	23.5	7.07	E, Cy, N
ClCG09G020870	9: 37890860 - 37903778	+	1722	63.9	5.22	PM
ClCG10G000920	10: 975005 - 976544	+	1026	38.3	6.48	PM
ClCG10G005260	10: 6466496 - 6471342	-	1095	40.6	9.2	PM, V
ClCG10G005280	10: 6496224 - 6506461	+	969	35.7	8.83	E, PM
ClCG10G009690	10: 21982062 - 21987296	-	1071	39.8	7.18	E, Ch
ClCG10G011690	10: 25488464 - 25499754	+	2184	80.1	8.52	PM
ClCG10G013760	10: 28053668 - 28064476	+	2058	76.1	7.04	PM, PM
ClCG10G019120	10: 34045330 - 34052985	+	1116	40.8	9.04	E, PM, L
ClCG10G022120	10: 36599093 - 36601070	+	690	25.6	6.09	Cy, Ch
ClCG11G007470	11: 9677117 - 9680791	-	1071	39.6	9.23	E
ClCG11G008800	11: 12963403 - 12964798	+	1140	41.6	5.33	E, PM
ClCG11G010220	11: 18938812 - 18944953	-	1116	40.9	5.99	PM
ClCG11G010240	11: 19047305 - 19050868	-	1086	40.6	6.88	E

[a] Ch, Chloroplast; Cy, Cytoplasmic; E, Extracellular; L, Lysosomal; M, Mitochondrial; N, Nuclear; PM, plasma membrane; V, Vacuole

Additional File Legend

Table S1. Protein and CDS sequences of the 65 GDSL genes identified in this study.

Table S2. Segmental and tandem duplications of GDSL gene pairs in watermelon (*C. lanatus*) and inference of duplication time.

Table S3. Orthologous relationships of GDSL gene pairs between watermelon and three plant species of (ATH, *Arabidopsis thaliana*; MEL, Melon and CUM, Cucumber).

Table S4. RNA-seq data of 65 GDSL genes that were used in this study.

Figures

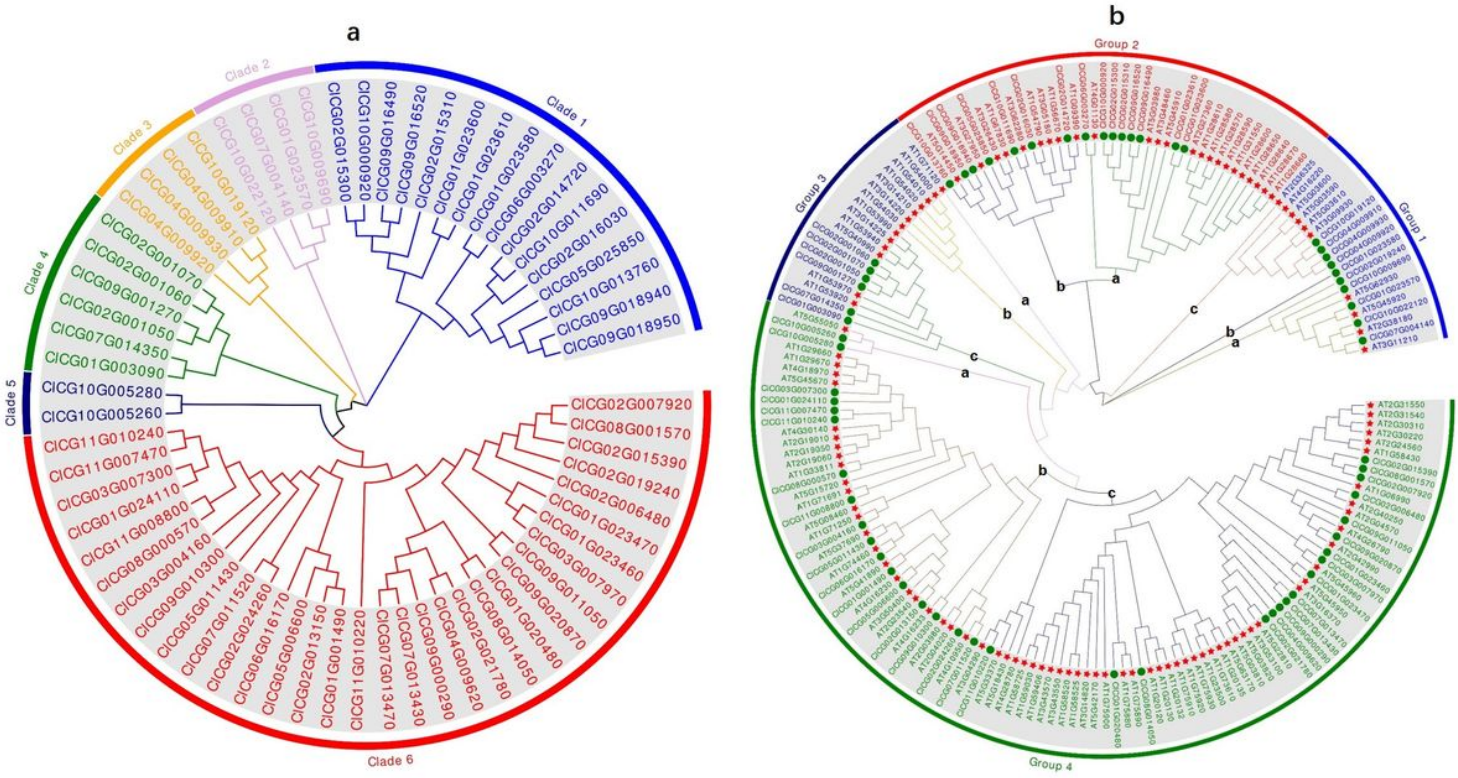

Figure 1

Phylogenetic analyses of GDSL proteins. a Phylogenetic classification of the watermelon GDSL proteins. b Phylogenetic relationships among GDSL proteins from C. lanatus and Arabidopsis thaliana. Genes on branch ends from watermelon and Arabidopsis are denoted by green solid circles and red stars, respectively. The different-colored arcs indicate different groups of GDSL proteins. The different-colored gene names indicate different groups (or clades). The subgroups (or subclades) were distinguished by different colored branches.

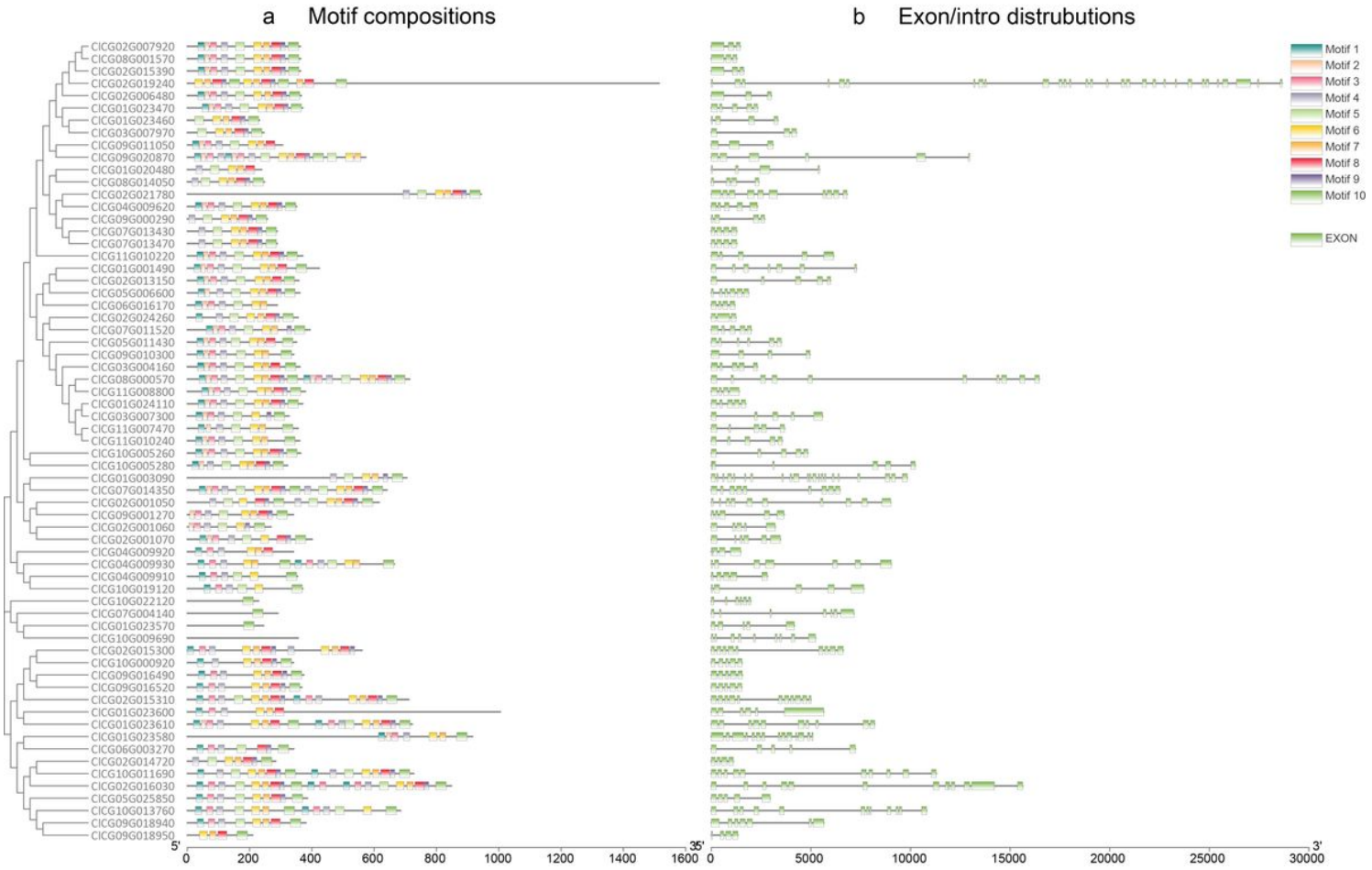

Figure 2

Conserved motif compositions and gene structure of watermelon GDSL proteins. a The motifs, numbers 1-10, are displayed in different colored boxes. The length of protein can be estimated using the scale at the bottom. b Exon/intron organizations of 65 GDSL genes. Solid green boxes and black lines indicate exons and introns, respectively. Their lengths are indicated in base pairs, and the scale is shown at the bottom.

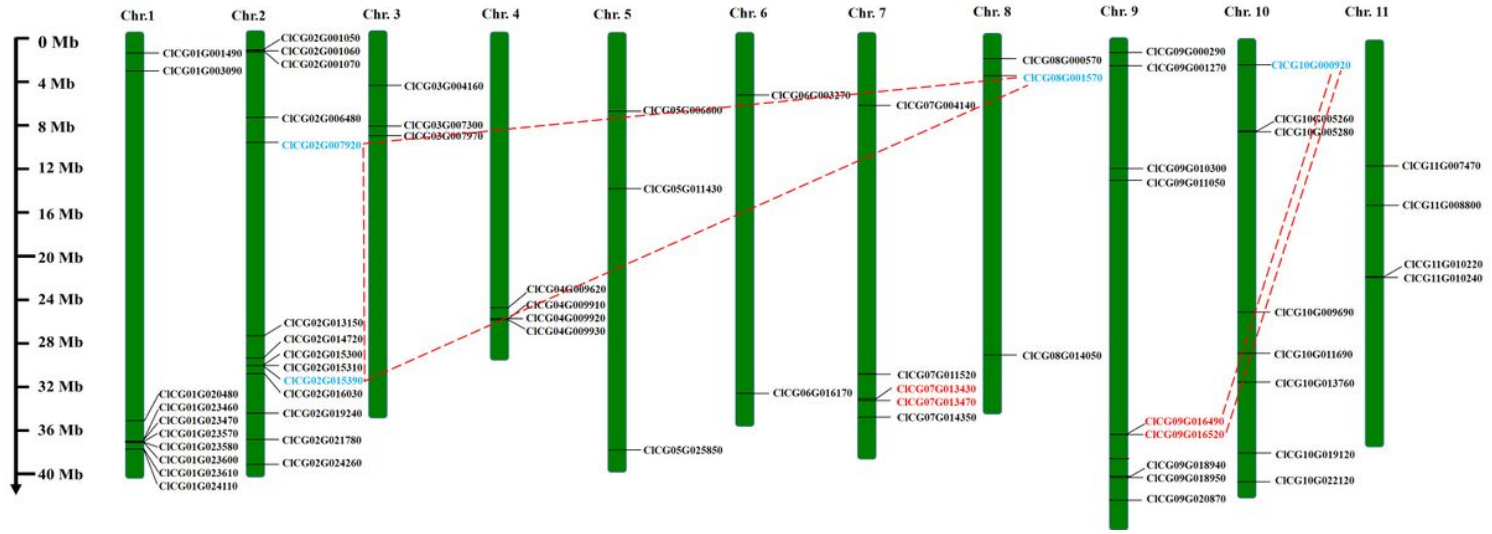

Figure 3

Genomic distributions of 65 GDSL genes on 11 watermelon chromosomes. Tandemly duplicated genes are colored in red. Segmentally duplicated genes are colored in blue and connected by red dotted lines. The scale bar on the left is shown in megabases (Mb).

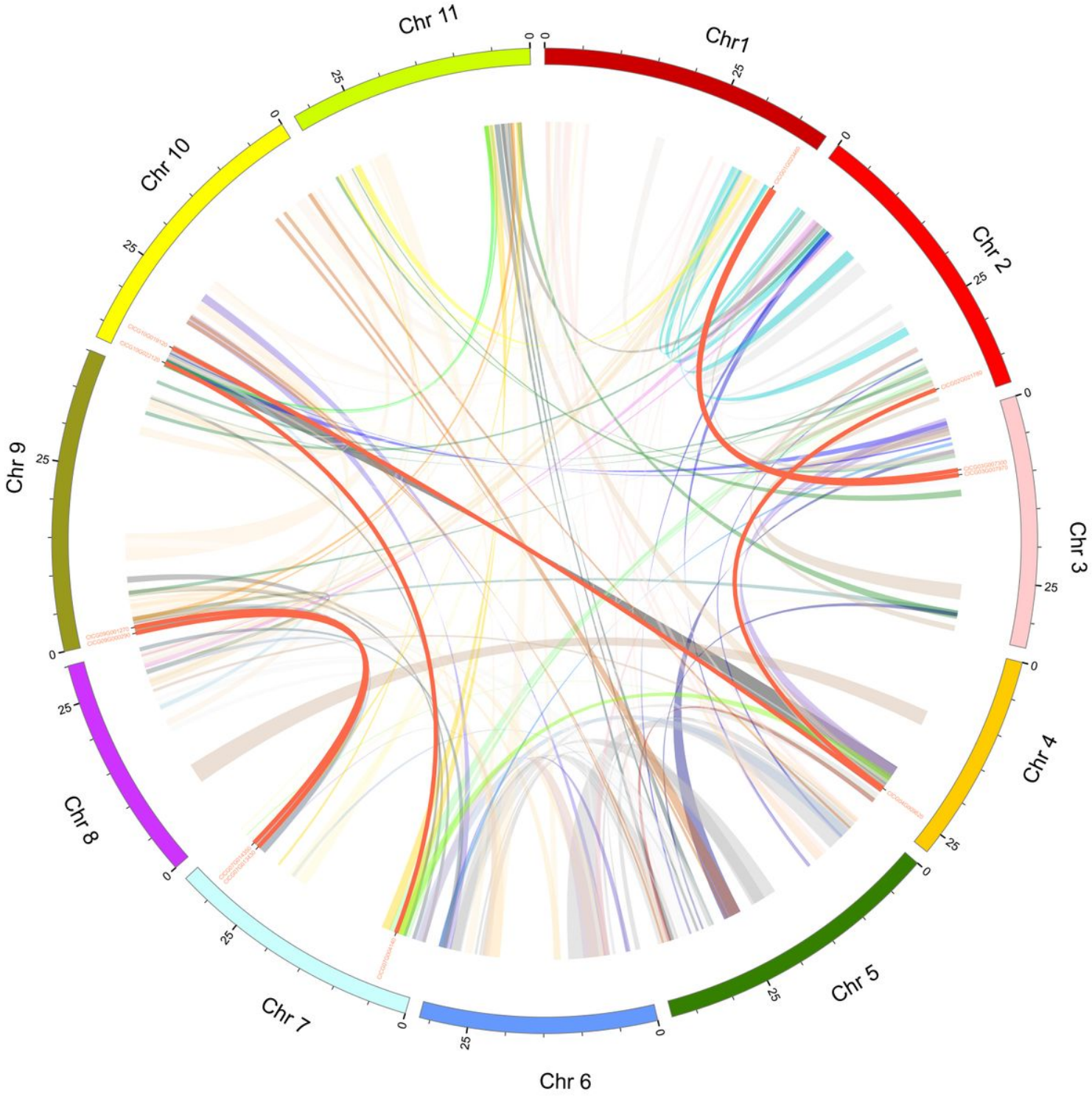

Figure 4

Duplication events of GDSL genes, including tandemly and segmentally duplicated genes in watermelon. The seven GDSL gene pairs are represented in bold red lines. The different color lines indicate all the tandemly and segmentally duplicated genes in watermelon. The picture was drawn with Circos.

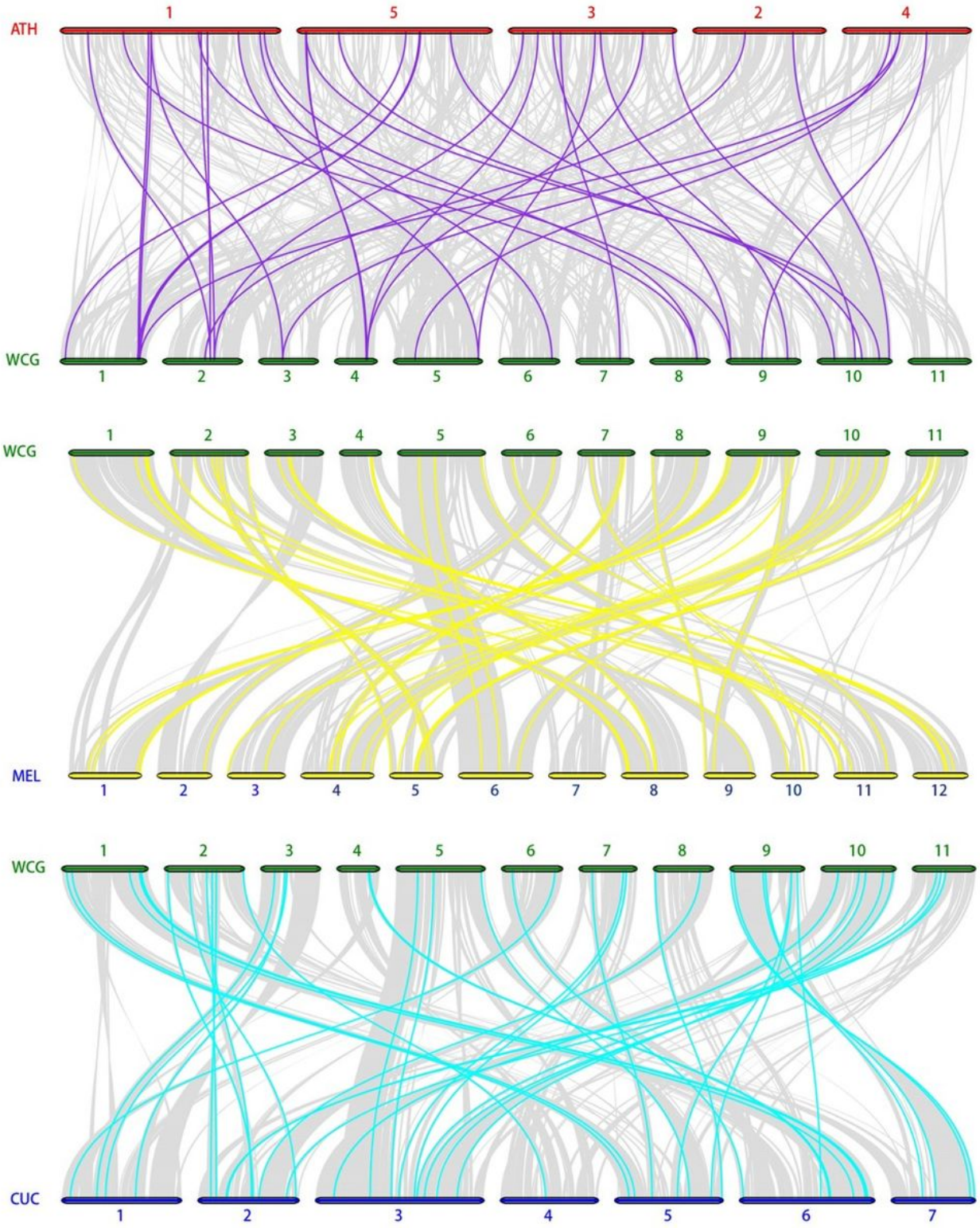

Figure 5

Syntenic relationships of GDSL genes between watermelon and three plant species (ATH, Arabidopsis thaliana; MEL, Melon and CUM, Cucumber). The colored lines in each figure represent the corresponding

syntenic GDSL gene pairs, and the gray lines in the background represent the syntenic blocks in watermelon and other plant species.

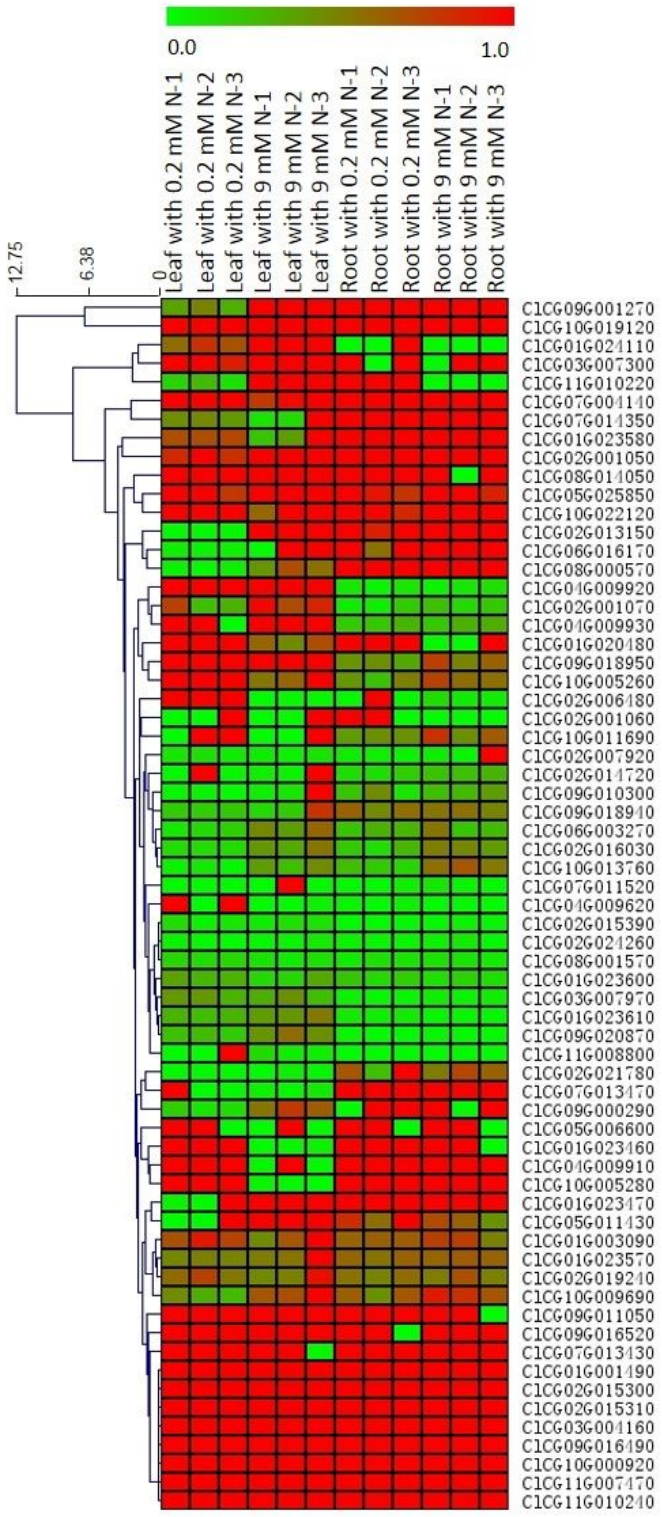

Figure 6

Expression profiles of watermelon GDSL genes. a The watermelon GDSL genes were clustered according to their expression profiles in leaves and roots under treatment with 0.2 mM and 9 mM N. The color scale represents the fold change in the gene expression value compared with the control. The lower expression

of genes was shown with the green shades, and higher expression of genes was shown using the red shades. The bar on the top represents relative expression values

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalTables.xlsx