

Genome-wide identification and expression profiling of SET DOMAIN GROUP family in *Dendrobium catenatum*

Dong-Hong Chen (✉ chendh212@163.com)

Zhejiang A&F University

Han-Lin Qiu

Zhejiang A and F University

Yong Huang

Hunan Agricultural University

Lei Zhang

Army Medical University

Jin-Ping Si

Zhejiang A and F University

Research article

Keywords: SDG, SET domain, histone lysine methylation, expression profiling, environmental stress

Posted Date: December 4th, 2019

DOI: <https://doi.org/10.21203/rs.2.16026/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on January 28th, 2020. See the published version at <https://doi.org/10.1186/s12870-020-2244-6>.

Abstract

Background: *Dendrobium catenatum* used as a precious Chinese herbal medicine is a kind of epiphytic orchid plant which grows on the trunks and cliffs and often faces up to diverse environmental stresses. SET DOMAIN GROUP (SDG) proteins act as histone lysine methyltransferases which are involved in pleiotropic developmental events and stress responses through modifying chromatin structure and regulating gene transcription, but their roles in *D. catenatum* are unknown. **Results:** In this study, we identified 44 SDG proteins from *D. catenatum* genome. Subsequently, comprehensive analyses related to gene structure, protein domain organization, and phylogenetic relationship were performed to evaluate these *D. catenatum* SDG (DcSDG) proteins, along with the well-investigated homologs from the model plants *Arabidopsis thaliana* and *Oryza sativa* as well as the newly characterized 42 SDG proteins from a closely related orchid plant *Phalaenopsis equestris*. We showed DcSDG proteins can be grouped into eight distinct classes (I~VII and M), mostly consistent with the previous description. Based on the catalytic substrates of the reported SDG members mainly in *Arabidopsis*, Class I (E(z)-Like) is predicted to account for the deposition of H3K27me_{2/3}, Class II (Ash-like) for H3K36me, Class III (Trx/ATX-like) for H3K4me_{2/3}, Class M (ATXR3/7) for H3K4me, Class IV (Su(var)-like) for H3K27me₁, Class V (Suv-like) for H3K9me, as well as class VI (S-ET) and class VII (RBCMT) for methylation of both histone and non-histone proteins. RNA-seq derived expression profiling showed that DcSDG proteins usually have wide but distinguished expression in different tissues and organs. Finally, environmental stresses examination showed the expression of DcASHR3, DcSUVR3, DcATXR4, DcATXR5b, and DcSDG49 is closely associated with drought-recovery treatment, the expression of DcSUVH5a, DcATXR5a and DcSUVR14a is significantly influenced by low temperature, and even 61% DcSDG genes are in response to heat shock. **Conclusions:** This study systematically identifies and classifies SDG genes in orchid plant *D. catenatum*, indicates their functional divergence during the evolution, and discovers their broad roles in the developmental programs and stress responses. These results provide constructive clues for further functional investigation and epigenetic mechanism dissection of SET-containing proteins in orchid.

Background

Spatiotemporal expression patterns of a set of genes associated with developmental and environmental stimuli are largely regulated by epigenetic modifications, which mainly consist of DNA methylation, non-coding RNAs, chromatin remodeling and histone modifications [1]. Histone modifications may occur post-translationally on many residues in histone tails and core regions. These modifications mainly include methylation, acylation, phosphorylation, ubiquitination, citrullination, hydroxylation, O-GlcNAcylation and ADP-ribosylation [2, 3], which constitute different combinations, that is, "histone code", and act in chromatin-templated processes [4, 5]. Among these covalent modifications, histone methylation displays complicated features as it not only occurs on distinct residues (lysine and arginine) and positions, but also involves different numbers (1~3) of methyl groups [6]. Histone lysine modification is catalyzed by histone lysine methyltransferases / SET DOMAIN GROUP (SDG) proteins which commonly possess an evolutionarily conserved SET domain as catalytic module [7], except for H3K79 methyltransferase DOT1,

which lacks a SET domain [8]. SDG-catalyzed histone methylation at specific lysine residues can cause similar or opposite effects on gene expression, such as those of H3K4 and H3K36 associated with gene activation, whereas those of H3K9, H3K27, H4K20 associated with gene repression [9].

Based on the sequence homology and phylogenetic reconstruction, plant SDG proteins can be categorized into seven distinct classes: class I, E(z) homologs (H3K27me writer); class II, ASH1 homologs (H3K36me writer); class III, Trx homologs and related proteins (H3K4me writer); class IV, proteins with a SET domain and a PHD domain; class V, Su(var) homologs and relatives (H3K9me writer); class VI, proteins with an interrupted SET domain (S-ET); class VII, RBCMT and other SET related proteins for targeting non-histone proteins [10, 11]. *SDG* genes have been discovered in bacteria, viruses, and eukaryotes [12, 13]. The presence of *SDG* genes in bacteria was initially considered a consequence of horizontal gene transfer from eukaryotic hosts [14, 15]. However, investigation on more released genomes of prokaryotic organisms including not only pathogens and symbionts, but also free-living bacteria and archaea suggests that *SDG* genes have undergone independent evolution in prokaryotes, which is unrelated to the evolution of eukaryotic SDGs, on the other hand, an ancient horizontal gene transfer occurred between bacteria and archaea [13, 16].

SDG family has currently been systematically identified and classified in the genomes of *Arabidopsis* (49 members) [10, 17], *Brassica rapa* (49) [11], *Vitis vinifera* (33) [18], *Populus trichocarpa* (59) [19], *Zea mays* (43) [20], *Oryza sativa* (43) [21], *Solanum lycopersicum* (52) [22], *Citrus sinensis* (47) [23], *Gossypium raimondii* (52) [24], C4 panicoid model *Setaria italica* (53) [25], and *Litchi chinensis* (48) [26]. However, the SDG family in orchid species, which constitutes an extremely evolutionary branch, remains elusive.

SDG proteins and related histone methylation marks are widely involved in diverse growth and developmental processes, such as seed dormancy, repression of vegetative-to-embryonic reversion, shoot branching, root system architecture, chloroplast development, flowering time, vernalization, floral organ development, ovule and anther development, embryo and endosperm development, plant senescence, carotenoid biosynthesis, and thigmomorphogenesis [27-32]. They are also implicated in the response to biotic and abiotic stresses. SDG8 is required for plant defense against necrotrophic fungal pathogens by regulating a subset of genes within jasmonic acid (JA) and/or ethylene signaling pathway [27] and for basal and R protein-mediated resistance to bacterial pathogens in *Arabidopsis* [33]. Loss-of-function mutant *sdg8* results in enhanced susceptibility to the fungal and bacterial pathogens.

ARABIDOPSIS TRITHORAX-LIKE PROTEIN1 (ATX1) as H3K4me3 writer orchestrates expression of defense response genes in antagonistic salicylic acid (SA)/JA signaling pathways by directly activating the expression of the SA/JA signaling mediator *WRKY70* gene through establishing H3K4me3 marks on its nucleosomes [34]. In addition, ATX1 is involved in drought stress response, and its disruption results in decreased tolerance to dehydration stress in *atx1* plants [35]. ATX1 modulates dehydration stress signaling in both abscisic acid (ABA)-dependent and -independent pathways. During ABA-dependent pathway, dehydration stress induces ATX1 binding to *NCED3* locus, which encodes the rate-limiting enzyme in ABA biosynthesis. Subsequently the deposition of H3K4me3 mark and recruitment of RNA polymerase II are increased, leading to enhanced *NCED3* expression and ABA production [36].

Dehydration stress causes dynamic and specific changes in global histone H3K4me1/2/3 patterns in *Arabidopsis*, especially H3K4me3 marks with broad distribution profiles on the nucleosomes of stress-induced genes [37]. Similarly, drought stress triggers massive changes in H3K4me3 enrichments on numerous loci (respectively including 3927/910 genes with increased/decreased depositions) in rice seedlings, showing positive correlation with their transcript changes in response to drought stress [38]. However, when *Arabidopsis* exposed to cold temperatures, H3K27me3 deposition gradually decreases in the chromatins of two cold-responsive genes, *COR15A* and *ATGOLS3* [39].

Dendrobium catenatum (also known as *Dendrobium officinale*) belongs to the Orchidaceae family, is a rare and precious Chinese medicinal herb. The stem of *D. catenatum* is the major medicinal part used for relieving upset stomach, promoting body fluid production, and nourishing Yin and antipyresis in traditional remedy and health care [40]. Furthermore, the plant stem contains bioactive extracts with anticancer, hepatoprotective, hypolipidemic, antifatigue, antioxidant, anticonstipation, hypoglycemic, gastric ulcer-protective, and antihypertensive effects, and immunoenhancement, as confirmed by modern pharmacology [41]. However, given its long-time extensive demand and over-exploitation, *D. catenatum* suffers from near extinction, and was once defined as an endangered medicinal plant in China [42]. In the past 20 years, *D. catenatum* has been successfully cultivated and became an important economic crop for health care. However, environmental stresses, such as drought, cold, and high temperature, extremely restrict its growth, resulting in heavy yield loss [43]. Hence, it is necessary to screen and identify the candidate genes conferring resistance to differential environmental stresses in *D. catenatum* molecular breeding.

To obtain a detailed understanding on the SDG family in medicinal orchid plants, we identified SDG members throughout the genome of *D. catenatum*, and subsequently performed comprehensive assessments on the phylogenetic relationship, gene structure, domain organization, gene expression profiling, and response to environmental stresses. Our results provide insights into the evolution and function of SDG genes in medicinal orchid plants.

Results

Identification of SDG Proteins in the *D. catenatum* Genome

To obtain all of the members of SDG proteins in *D. catenatum*, we performed BLASTP search using known *Arabidopsis* and rice SDG proteins as queries against *D. catenatum* genome (INSDC: JSDN00000000.2). First, we checked the SDG genes of *Arabidopsis* and rice in the Superfamily 1.75 database (<http://supfam.org/SUPERFAMILY/>), we discovered 49 genes in *Arabidopsis thaliana*, corresponding to those reported in literature (Additional file 1) [10, 17]. On the other hand, 46 genes were identified in *Oryza sativa* (Additional file 1), three (Os01g65730/OsSET44, Os01g74500/OsSET45, Os06g03676/OsSET46) more than the 43 reported genes [21]. Reciprocal BLAST was carried out to confirm that the hits from *D. catenatum* and its close relative *Phalaenopsis equestris* belong to the SDG

family. Finally, we obtained 44 SDG genes in *D. catenatum* (Table 1 and Fig. 1) and 42 in *P. equestris* (Additional file 2), and they were named after their Arabidopsis homologs.

To characterize and classify the SDG family in *D. catenatum*, we used SDG proteins in the dicot model plant *A. thaliana*, the monocot model plant *O. sativa*, and its close relative *P. equestris* as references for phylogenetic analysis. The results showed that the 181 SDG proteins from the above four species could be clustered into eight classes (I~VII and M), mostly corresponding to the classification criteria in Arabidopsis [10]. By contrast, ARABIDOPSIS TRITHORAX RELATED 3 (ATXR3) branch was previously classified into Class III [10], but it was separated from the other branches and near ATXR7-like proteins in Class IV in this study (Fig. 1). Considering their similar substrate specificities, we combined ATXR3 branch and the neighboring ATXR7-like proteins and categorized them under class M, as supported by the further phylogenetic analysis among Classes III, IV, and M (Fig. 4). Given the high limited reports on Class VI and VII, which feature potential functions for non-histone and histone methylation, we mainly focused on the roles of Classes I~V and M with well-investigated histone methylation specificity in this study.

Class I: E(z)-Like (H3K27me2/3)

Class I contains two E(z) homologs in each of monocot plants *D. catenatum*, *P. equestris* and rice, and three well-characterized homologs, namely, CURLY LEAF (CLF), SWINGER (SWN), and MEAEA (MEA), that represent three distinct clades in the dicot plant Arabidopsis (Fig. 2). The genes in this class contain 15~16 introns, which are extremely longer in the two orchid species compared with those of Arabidopsis and rice. This result suggests that overall intron length positively correlated with the corresponding genome size. A similar phenomenon related to intron/exon proportion was also observed in the members of the other classes as will be mentioned later. The three Arabidopsis E(z) proteins act as the catalytic subunits of the evolutionarily conserved Polycomb Repressive Complex 2 (PRC2), which is involved in the deposition of H3K27me3 repressive mark on the target gene locus [44]. CLF (dominant H3K27me3 writer) and SWN act redundantly in vegetative and reproductive development, whereas MEA functions exclusively in suppression of central cell proliferation and endosperm development [45-47]. Rice E(z) homologs SDG711 and SDG718 participate in mediating accurate photoperiod control of flowering time [48]. Clades I-1 (CLF-like) and I-2 (SWN-like) each contain one ortholog in the examined species, but Clade I-3 (MEA-like) is confined to Arabidopsis. Plant E(z)-like proteins generally harbor highly conserved domain organization at the C-terminal region, which includes a SANT domain, the cysteine rich CXC domain, and the signature SET domain, except for DcCLF and PeSWN, which lack the SANT domain, and DcSWN, which possesses an additional SANT domain at the N-terminus.

Class II: Ash-Like (H3K36me)

Class II can be further divided into five clades, each of which consists of a single member per plant species, except for two members of Clade II-1 in Arabidopsis and *D. catenatum* (Fig. 3). Clades II-1 to II-4 exist in all the examined species, but Clade II-5 is only found in rice and contains a single member, SDG707, with unknown function. Class II proteins generally share three conserved domains: an Associated with SET (AWS), SET, and PostSET domains [10, 49].

Clade II-1 (ASHH3-like) members are relatively shorter than their homologs in the four other clades. Arabidopsis ASHH3/SDG7 is required for proper timing in response to vernalization [50]. SDG7 lacks detectable HKMTase activity [51], but rice ortholog SDG724/LVP possesses H3K36 methylation activity. The loss of *SDG724* leads to a late flowering phenotype [52]. Notably, DcASHH3a/3b in *D. catenatum* lack AWS domain, different from PeASHH3 from its close relative *P. equestris* and ASHH3 orthologs in Arabidopsis and rice, and their functional divergence during speciation is interesting to investigate.

Clade II-2 (ASHR3-like) members are characterized by an additional PHD domain near the N-terminus, except for rice SDG736. ASHR3/SDG4 participates in regulating pollen tube growth and stamen development, and its overexpression leads to growth arrest, anther degeneration and male sterility [53, 54]. ASHR3 harbors catalytic activities on H3K36me1 and possible H3K36me2, which is involved in regulating cell division competence in the root meristem and quiescent center [55].

Clade II-3 (ASHH1-like) members display uniform protein length and highly conserved AWS-SET-PostSET domain combination at the N-terminus. Arabidopsis ASHH1/SDG26 localized in the nucleus can methylate oligonucleosomes in vitro, and ASHH1/SDG26 knockout leads to a late-flowering phenotype through decreasing H3K4me3 and H3K36me3 level at the *SOC1* locus [56, 57]. Similarly, the knockdown of rice ortholog *SDG708* causes a late-flowering phenotype and a genome-wide decrease in H3K36me1/2/3 levels during early growth stages [58]. Predictably, *D. catenatum* DcASHH1 harbors a similar function.

Clade II-4 (ASHH2-like) proteins are considerably longer than the others, and characterized by an additional CW domain near to the N-terminal triple domain combination. In Arabidopsis, ASHH2/SDG8, as the major H3K36 methyltransferase, specifically accounts for the di- and tri-methylation of H3K36 [57, 59]. *SDG8* deletion leads to pleiotropic phenotypes in vegetative and reproductive stage [60]. Consistently, the knockdown of rice ortholog *SDG725* causes wide-ranging defects, including dwarfism, shortened internodes, erect leaves and small seeds [32]. In the aspect of protein architecture, ASHH2 ortholog in *D. catenatum* or *P. equestris* is more similar to Arabidopsis SDG8 than rice SDG725.

Class III: Trx/ATX-Like (H3K4me2/3)

Class III consists of five members, which can be further divided into three clades in each examined plant species (Fig. 4). Class III proteins characterized tandem PHD domains in the middle region and SET-PostSET domain combination at the N-terminus. Moreover, several clades contain additional distinct

domains, such as PWWP domain specific to Clade III-1/2, and FYRN–FYRC domain combination specific to Clade III-1.

Clade III-1 (ATX1-like) contains two members in Arabidopsis, and one in each of the three other species. In Arabidopsis, ATX1 and ATX2 paralogs exhibit similar domain architectures [61], but have distinct expression patterns in most cases and influence the expressions of largely nonoverlapping gene sets [62]. For the shared targets, ATX1 and ATX2 account for the deposition of H3K4me3 and H3K4me2 marks, respectively [62]. Notably, *atx1* mutant displays a pleiotropic phenotype, whereas *atx2* mutant is wildtype-like [62, 63]. Different from Arabidopsis *atx1* with early-flowering phenotype [62], rice ortholog mutant *ostrx1/sdg723* exhibits late-flowering through decreased H3K4me3 levels at the central flowering time integrator *Ehd1*. OsTRX1 can rescue Arabidopsis *atx1* phenotype [64-66], suggesting that ATX1-like proteins demonstrate conserved biochemical and molecular functions during evolution. However, ATX1-like proteins produce specific phenotypes in distinct species due to the differences in developmental context. Thus, DcATX1 and PeATX1 in orchid may play important roles in flowering time control.

Clade III-2 (ATX3-like) includes three members in each tested species. Arabidopsis ATX3/4/5 are clustered together and separated from the monocot orthologs. The orthologs from *D. catenatum*, and *P. equestris* are consistently clustered together, concordant with their close relationship. In Arabidopsis, ATX3/4/5 exhibit a common evolutionary origin, and function redundantly in genome-wide H3K4me2/3 profiles. Furthermore, *atx3 atx4 atx5* triple mutant displays dwarfism and reduced fertility [67]. In rice, ATX3-like proteins SDG721 and SDG705 function redundantly in modulating H3K4 methylation levels. The loss of both genes results in semi-dwarfism, shortened panicle branching, and reduced cell length [68]. Considering the dwarf phenotype of ATX3-like mutants in Arabidopsis and rice, the homologs in *D. catenatum* and *P. equestris* might be involved in regulating plant architecture.

Clade III-3 shows specificity toward the examined monocots and contains one copy per species. *D. catenatum* DcATX3d and *P. equestris* PeATX3d are characterized by an additional Jas domain at the C-terminus, in contrast with the rice ortholog OsSET37/SDG732. Further survey of this clade will be helpful for understanding the evolution of SDG family in monocots.

Class M: ATXR3/7 (H3K4me)

Class M comprises of two clades, namely, Clade M-1 (ATXR7-like) and M-2 (ATXR3-like). Each clade contains one copy per plant species (Fig. 4). ATXR7-like proteins usually lack extra domains, except for PeATXR7 with a C-terminal GYF domain. Arabidopsis ATXR7/SDG25 acts as the writer of H3K4 monomethylation (H3K4me1), and the *sdg25* null mutant displays early flowering phenotype [59, 69]. ATXR3-like proteins also contain only one copy in each species, are characterized by the presence of DUF4339 domain in the middle region, except for OsSET27/SDG701. Arabidopsis ATXR3/SDG2 is the major H3K4me3 writer, whose depletion leads to pleiotropic development defects [28, 70, 71]. *D. catenatum* DcATXR3 and *P. equestris* PeATXR3 feature a more similar protein architecture to Arabidopsis

ATXR3/SDG2 than rice SDG701. This finding suggests that ATXR3-like proteins in orchid may retain their ancestral role, whereas rice ortholog may functionally diverge, as attributed to the loss of specific domain and partial sequence.

Class IV: Su(var)-like (H3K27me1)

Class IV can be divided into two clades, Clade IV-I (ATXR5-like) and IV-II (ATXR6-like), which encode shorter proteins characteristic of an N-terminal PHD domain in addition to the defined SET domain, except for PeATXR5 (Fig. 4). In Arabidopsis, ATXR5 and ATXR6 show largely overlapping functions. The depletion of both results in global H3K27me1 reduction, heterochromatin decondensation, and over-replication of heterochromatin [72, 73]. ATXR5/6 are involved in maintaining DNA replication [74] and repressing the expression of transposable element [75]. The overexpression of either ATXR5 or ATXR6 causes male sterility [76]. ATXR5 and ATXR6 probably perform separate roles because of ATXR5 with a dual localization in plastids and nucleus but ATXR6 solely in nucleus [76].

Class V: Suv-Like (H3K9me)

Class V contains 15 members in Arabidopsis, 14 in rice and *D. catenatum*, and 13 in *P. equestris*; These members can be further divided into two subclasses, SUVH and SUVR, which include Clades V-1 to V-3 and V-4 to V-6, respectively (Fig. 5). Class V proteins are usually characterized by PreSET–SET–PostSET or PreSET–SET domain combinations. SUVH proteins often contain another symbolic SET- and RING-ASSOCIATED (SRA) domain, whereas SUVR proteins in Clades V-4 and V-5 often include another WIYLD domain and tandem ZnF_C2H2 domains, respectively. SUVH genes usually lack introns, except for the members of SUVH4 branch and two members (PeSUVH45 and SDG727) of SUVH5 branch, whereas SUVR genes contain variable number of introns. In general, Class V members are responsible for methylation of histone H3 lysine 9 (H3K9me), in which H3K9 dimethylation (H3K9me2) is the critical mark for gene silencing and DNA methylation, and are involved in heterochromatin formation and reprogramming of gene expression [77].

SUVH subclass

In Clade V-1, the five members SUVH1/3/7/8/10 in Arabidopsis cluster together, and are distinct from the five homologs in rice and each of the two homologs in *D. catenatum* or *P. equestris*. This result indicates that duplication of these clade members occurred after divergence between dicots and monocots. However, the two orthologs in either *D. catenatum* or *P. equestris* respectively pair together, indicating that their gene duplication occurred before the split of *Dendrobium* and *Phalaenopsis*. Arabidopsis SUVH1/SDG32 performs a distinct anti-silencing function to promote the expression of genes targeted by

DNA methylation. The knockout of SUVH1/SDG32 causes no effect on H3K9me2 levels but reduces H3K4me3 levels [78]. Furthermore, SUVH1 binds to highly methylated genomic loci targeted by RNA-directed DNA methylation (RdDM) at the whole-genome level. However, SDG728, as one of SUVH1 homologs in rice, retains its classical function to mediate H3K9 methylation and participate in retrotransposon repression [79]. *D. catenatum* and *P. equestris* include two SUVH1-like proteins, far less than the five members in Arabidopsis and rice. Thus, the function of SUVH1 homologs and the evolutionary mechanisms in orchids require further investigation.

Clade V-2 comprises two members for each examined species, and these members lack PostSET domains, in contrast with those in clade V-1. In Arabidopsis, SUVH2 and SUVH9 as sister paralogs show overlapping functions in RdDM and heterochromatic gene silencing [80, 81]. SUVH2 displays a central role in heterochromatic gene silencing with strong dosage effects [82]. SUVH2 overexpression leads to ectopic heterochromatization accompanied with significant developmental defects, such as extreme dwarfism and curly cotyledon [80, 82]. SUVH2 and SUVH9 may feature inactive histone methyltransferase activity [83, 84]. However, the simultaneous absences of *SUVH2* and *SUVH9* lead to a marked decrease in H3K9me2 levels in the RdDM loci [81, 85]. SUVH2 and SUVH9 can bind to methylated DNA through their SRA domains and facilitate the recruitment of Pol V to RdDM loci via association with the chromatin-remodeling DDR complex and MORC complex [83, 85]. Considering the highly similar domain organization among SUVH2-like proteins in these examined species, their function should be evolutionarily conserved.

Clade V-3 proteins could be further divided into two branches (SUVH4 and SUVH5). SUVH4 branch possesses one member in each species, whereas SUVH5 branch contains two members in Arabidopsis and rice, and three members in *D. catenatum* or *P. equestris*. In Arabidopsis, SUVH4/KRYPTONITE (KYP), SUVH5, and SUVH6 as H3K9 methyltransferases are required to maintain DNA methylation, and act genetically downstream of H3K9 methylation [86-91]. SUVH4/KYP as the predominant H3K9 methyltransferase can mono- and di-methylate H3K9 [77]. SUVH5 and SUVH6 as sister paralogs in SUVH5 branch exhibit HKMTase activities with locus-specific features [77, 87, 88]. SUVH4/5/6 can bind 5-methyl-cytosine (5mC) in CG, CHH, and CHG sequence contexts via their SRA domains with different preferences [91, 92]. In rice, SUVH4-like protein SDG714 mediates H3K9 methylation, participating in DNA methylation, transposition of transposable elements, and genome stability [93]. Notably, PeSUVH4 protein in *P. equestris* is evidently short and lack SRA and PostSET domains, compared with SUVH4-like proteins in the three other species. The divergence of SUVH4 between orchid genus *Dendrobium* and *Phalaenopsis* is worthy of investigation.

SUVR subclass

In Clade V-4, there are 3 members in Arabidopsis and *P. equestris*, 1 in rice, and 4 in *D. catenatum*, respectively. In Arabidopsis, SUVR2 mediates transcriptional silencing in both RdDM-dependent and -independent manners [94]. SUVR4 participates in the epigenetic defense mechanism by introducing

H3K9me3 marks to repress potentially harmful transposon activity [95]. SUV4 specifically converts H3K9me1 into H3K9me3 at transposons and pseudogenes within the euchromatin [95, 96], but SUV1 and SUV2 show no detected histone methyltransferase activity *in vitro* [94, 97]. In this study, *D. catenatum* DcSUV4, *P. equestris* PeSUV4, and rice SDG712 were grouped together with Arabidopsis SUV4 but not with SUV1/2, implying that they possess ubiquitin binding and HKMTase activity, except for PeSUV4, which includes an obviously short sequence and lacks WIYLD and PreSET domains.

Clade V-5 contains one member in each tested species, and characterized by an additional tandem ZnF_C2H2 domain, except for SDG706. In Arabidopsis, SUV5 lacks the SRA domain but recognizes specific DNA sequences through its zinc finger motifs and establishes the heterochromatic state through H3K9me2 deposition in a DNA methylation-independent manner [98]. The knockout of SUV5 leads to delayed flowering, and no further enhanced phenotype occurs in the quintuple *suvr1 suvr2 suvr3 suvr4 suvr5* mutants [98, 99]. This finding suggests that SUV5 is a dominant developmental regulator in SUV subfamily.

Clade V-6 members exist in one copy in each species, and their encoding proteins are notably shorter than those of the other clades in this class. Arabidopsis SUV3 contains an additional AWS domain close to the SET-PostSET domain combination, and DcSUV3 contains an intact PreSET-SET-PostSET domain combination. However, SUV3 orthologs in rice and in *P. equestris* only contain a PreSET domain, suggesting that the genes in Clade V-6 may undergo less selective pressures and become increasingly divergent during evolution. The functions of the genes in this clade remain uncharacterized thus far.

Tissue and organ expression profiles of *DcSDG* genes

To investigate the potential roles of *DcSDGs* during growth and development in *D. catenatum*, we detected the expression profiles of *DcSDGs* by reanalyzing the RNA-seq data from different plant tissues and organs, including leaf, root, green root tip, white part of root, stem, flower bud, sepal, labellum (lip), pollinia, and gynostemium (column) [100].

On the basis of hierarchical clustering (Fig. 6 and Additional file 3), the expression patterns of *DcSDGs* could be divided into two groups, G1 and G2. G1 genes usually feature low expressions in most tissues and organs. However, several genes are highly expressed in specific tissues and organs, such as *DcATX3d* in flower bud, *DcASHR3* in root and flower bud, and *DcATXR5a*, *DcSUVH5a*, and *DcATX3a* in pollinia. G2 genes display diverse expression profiles in different tissues and organs. The majority of G2 group genes are highly expressed in most of the detected tissues and organs, whereas several show intermediate expressions in most tissues and organs, such as *DcASHH3a*, *DcASHH1*, *DcSUVH5b*, and *DcSDG47*. Furthermore, ~86% of genes (38/44) present intermediate and high expressions in flower buds, compared with ~77% (34) in root and gynostemium, ~57% (25) in leaf, ~45% (20) in stem, ~43% (19) in pollinia, ~55% (24) in sepal, and ~64% (28) in labellum. These findings suggest that *DcSDG* family plays essential roles in flower bud formation. Notably, *DcSDG51* in leaf, *DcSUVH2b* in pollinia, and

DcSUVH1a in root and stem display the most distinguished expressions, indicating their prominent functions in specific tissues.

The expression profiles of six duplicated *DcSDG* gene pairs were further compared (Additional file 4). In general, one copy showed higher expression levels than the others in all tissues, except for *DcSUVH5a/5b* pairs, suggesting that one paralog might performed a dominant function during plant growth and development. *DcASHH3a/3b*, *DcATX3b/3c*, *DcATXR5a/5b* and *DcSUVR14b/14c* exhibited similar expression patterns, whereas *DcSUVH5a/5b* pairs displayed differential expression profiles in the detected tissues and organs. These results indicate that distinct duplicated gene pairs might undergo different evolutionary pressures and diverge at varying periods.

Expression levels of *DcSDGs* in response to environmental stresses

D. catenatum is an epiphytic orchid plant that grows on trunks and cliffs and often experiences diverse environmental stresses, such as drought, cold, and high temperature. To detect the responses of *DcSDG* genes to drought stress, the expression profiles of *DcSDGs* were assessed by analyzing the RNA-seq data from the leaves under different drought treatments [101] (Fig. 7 and Additional file 5). In brief, the seedlings were irrigated on the 1st day, and no water was provided from the 2nd day to the 7th day, and recovered on the 8th day. Leaves were sampled at both 06:30 and 18:30 on the 2nd (DR5 and DR8), 7th (DR6 and DR10), and 9th (DR7 and DR15) days, respectively, and at 18:30 on the 8th day (DR11). The results showed that one-week of drought stress notably repressed the expressions of *DcCLF*, *DcASHR3*, *DcSUVR3*, and *DcSUVR14c*, but obviously induced the expression of *DcATXR5b*, *DcATXR4*, and *DcSDG49* when sampling at both dawn and dusk. Subsequently, rewatering restored the expression levels of *DcASHR3*, *DcSUVR3*, *DcATXR5b*, *DcATXR4*, and *DcSDG49*.

To explore the possible roles of *SDG* proteins in response to cold stress, we evaluated the expression levels of *DcSDGs* through analyzing the raw RNA-seq reads from the leaves of *D. catenatum* seedlings treated at 20 °C (control) and 0 °C for 20 h, respectively [43] (Fig. 8 and Additional file 6). Data showed that 32% of *DcSDG* genes (14) showed transcription change in response to cold stress. For example, genes with upregulated expression consisted of *DcASHH1* (II), *DcATX3b* (III), *DcSUVH4* (V), *DcSUVH5a* (V), *DcSUVH5b* (V), *DcSDG45* (VII), and *DcSDG51* (VII), whereas genes with deregulated expressions included *DcATXR5a* (IV), *DcATXR5b* (IV), *DcSUVR14a* (V), *DcASHR1* (VI), *DcASHR2* (VI), *DcATXR2* (VI), and *DcSDG50* (VII). The expression levels of *DcSUVH5a*, *DcATXR5a*, *DcASHR2* and *DcSUVR14a* (fold change >2 or <0.5) are significantly influenced by cold.

To further understand the roles of the *DcSDG* proteins in response to high temperature (35°C) stress, the expression profiles of *SDG* genes in the leaves of *D. catenatum* seedlings were examined by quantitative *reverse-transcription*-polymerase chain reaction (RT-qPCR) (Fig. 9). The results show the diverse expression patterns of *DcSDG* genes during heat shock treatment. At 3 h after treatment (HAT), the number of upregulated genes (10) was slightly higher than that of downregulated genes (7). At 6 HAT,

more DcSDG genes were induced (15 upregulated genes versus 10 downregulated genes). At 12 HAT, the number of upregulated genes (27) was evidently higher than that of downregulated genes (3). Of the genes examined upon exposure to heat shock, three Class II genes (DcASHH3a/3b, DcATX3a), five Class V genes (DcSUVH2a/2b, DcSUR14b/14c, DcSUVR3), two Class VI genes (DcATXR1 and DcASHR1), two Class VII genes (DcSDG45/48), and one Class M gene DcATXR3 were distinguished from the corresponding control in at least at one time point ($P < 0.05$, Fig. 9).

Discussion

Characterization and classification of SDG proteins in *D. catenatum*

Over the last 20 years, SDG proteins in diverse plants, especially in the model plant *Arabidopsis* have been fully characterized [10]. A considerable amount of SDG proteins has been functionally investigated, providing the valuable reference information for precise protein classification at the genome-wide level. *D. catenatum* shows extensive application value in China and can be used in the food service industry, pharmaceutical, cosmetics, health products, and ornamental horticulture. The recent successful genome sequencing of *D. catenatum* associated with its unique developmental and living features promoted its use as a model orchid plant with great theoretical relevance [102]. In this study, 44 SDG proteins in *D. catenatum* was screened and identified from its genome sequence. Through the comprehensive evaluation of the sequence features, phylogenetic relationships, and substrate specificities based on previous classification criteria and published experimental data, this study divided the SDG proteins into eight classes: Class-I (H3K27me_{2/3}), Class-II (H3K36), Class-III (H3K4me_{2/3}), Class-M (H3K4me), Class-IV (H3K27me₁), Class-V (H3K9me), Class-VI/S-ET (undefined), Class-VII/ RBCMT (non-histone Rubisco).

To date, the understanding of Class VI and VII is highly limited. Class VI includes six clades (Additional file 7), whose members feature a long S-ET domain interrupted by a Zf-MYND domain and have not been functionally characterized in plants. In mammalian ASHR1 homologs, *SET And MYND Domain Containing 3* (SMYD3) may methylate histones H3K4 and H4K5, whereas SMYD2 can dimethylate H3K36 and repress gene transcription [103-105]. SMYD proteins also possess non-histone substrates, such as SMYD2 for p53 and estrogen receptor α and SMYD3 for VEGFR1 and MAP3K2, in the nucleus and cytoplasm [106-109]. Class VII consists of nine clades (Additional file 8), whose members are characterized by an additional Rubisco substrate-binding (Rubis-sub-bind) domain and conserved in plants and animals. In flowering plants, chloroplast-localized Rubisco large subunit N-methyltransferase (LSMT) performs the conserved and ancient functions to methylate the Rubisco small subunit and fructose-1,6-bisphosphate aldolase (FBA). Moreover, this enzyme evolved an extra novel role to catalyze the K14 trimethylation of Rubisco large subunit in Fabaceae, Cucurbitaceae, and Rosaceae [110, 111]. Based on the strict correlation between the presence of the His-Ala/Pro-Trp triad motif and the Rubisco methylation status [110], LSMTs in *D. catenatum* and the other three examined species only retained the ancestral function due to absence of this triad motif. In addition, pTAC14 associated with thylakoid is required for proper chloroplast development [112]. In animal, Class VII homolog SETD3 with additional Rubis-sub-bind domain exhibits a H3K36 methyltransferase activity [113]. Therefore, the members of

Class VI and VII in plants may be involved in the methylation of both histone and non-histone proteins; such assumption is beyond the previous expectation.

Evolution and function divergence of SDG proteins

Gene duplication, as a critical driving force, provides extra copy of genetic material to tolerate random mutations to survive the natural selection and create new species during evolution. Gene duplication, including tandem duplication and polyploidy (whole genome duplication, WGD), considerably contributes to the formation of gene families [24], such as the SDG family containing 44 members in *D. catenatum*. *D. catenatum* has undergone at least four rounds of WGD events, including two ancient WGDs before (ζ) and after (ϵ) the angiosperm split from gymnosperm, a τ WGD shared by most monocots and a recent WGD specific to the Orchidaceae lineage [100, 114]. Duplicate genes usually undergo three evolutionary fates, nonfunctionalization with silencing function, neofunctionalization with novel function, and subfunctionalization with partial function [115]. *D. catenatum* contains six pairs of duplicated DcSDG genes. In *DcSUVH2a/2b* pairs, *DcSUVH2a* was unexpressed in all examined tissues, implying that neofunctionalization might have occurred. *DcASHH3a/3b*, *DcATX3b/3c*, *DcATXR5a/5b* and *DcSUVR14b/14c* pairs displayed similar expression patterns, indicating that they might have experienced subfunctionalization. *DcSUVH5a/5b* pairs exhibited differential expression profiles in the different tissues and organs, indicating that novel expression specificity and function have developed.

Involvement of SDG proteins in plant development programs and environmental stress responses

SDG proteins associated with the corresponding histone methylations can output precise developmental instructions through diverse mechanisms. (1) Different SDG proteins show similar catalytic specificity, but act on distinct sets of genes. As the writers of H3K4 methylation, ATX3/4/5 have different targets from *ATXR3/SDG2* [67]. (2) Several SDG members act as the core subunits in specific conserved complexes. Class I E(z) homologs (CLF, SWN and MEA), together with p55-like protein MSI1; ESC-like protein FIE; and Su(z)12-like proteins FIS2, EMF2 and VRN2 constitute different PRC2 complexes, functioning in distinct developmental phases [116]. (3) Specific SDG proteins can interplay with other types of epigenetic regulators. SUVH4/5/6 interacts with the histone deacetylase HDA6 to silence a subset of transposons through histone H3K9 methylation and H3 deacetylation [117]. (4) Some SDG proteins can crosstalk with phytohormone signaling. Arabidopsis SDG8 and rice ortholog SDG725 mediate the deposition of H3K36me_{2/3} marks, and are involved in brassinosteroid (BR)-regulated gene expression through interacting with BR signal component BES1 and binding to BR biosynthesis or signaling genes *D11*, *BRI1*, and *BU1* [32, 118]. Further in-depth molecular dissection of SDG members in *D. catenatum* with specific developmental and living modes will enrich the action mechanisms of the SDG gene family.

Histone methylation established by SDG proteins are widely involved in responses to environmental stresses and pathogen challenges [27, 33, 35-38]. Studies have reported that drought stress can cause global changes of histone H3K4 methylation patterns in *Arabidopsis* and rice [37, 38]. Class III member ATX1 is implicated in drought stress response via both ABA-dependent and -independent pathways in *Arabidopsis* [35, 36]. Here we observed that the dynamic expression changes of DcASHR3 (Class II), DcSUVR3 (V), DcATXR5b (V), DcATXR4 (VI), and DcSDG49 (VII) were closely associated with water-drought-rewater treatment, indicating that methylations of H3K36 and H3K9 are also involved in drought response. H3K36me3 and H3K27me3 have been proven to play antagonistic roles in the cold-induced epigenetic switch at the *Arabidopsis FLC* locus [119]. In *D. catenatum*, we identified six significant cold-responsive *DcSDG* genes, including *DcASHH1* with putative H3K36 activity (Class II), DcATX3b/3d with putative H3K4 activity (Class III), DcATXR5a/5b with putative H3K27 activity (Class M), DcSUVH5a/5b and DcSUVR14a/3 with putative H3K9 activity (Class V) (Fig. 8). The results indicate diverse histone methylation marks with specific DcSDG proteins that perform certain roles during cold treatment. It will be intriguing to further investigate their direct targets by high throughput ChIP-seq method combined with excellent commercial antibodies against various histone methylation marks. Recently, Huang et al. [24] thoroughly identified cotton SDG genes and noted that the expressions of most of these genes decreased under high-temperature conditions. In this study, we observed 61% of *DcSDG* genes are in response to heat shock, but the number of upregulated genes was considerably higher than that of downregulated ones. This finding may be related to the particularly epiphytic lifestyle and crassulacean acid metabolism pathway in *D. catenatum*.

Conclusion

In this study, we identified 44 SDG proteins in *D. catenatum* and 42 in its close relative *P. equestris*, and replenished three other SDG members (Os01g65730/OsSET44, Os01g74500/OsSET45, and Os06g03676/OsSET46) into the previous rice SDG gene family (43 members). On the basis of the phylogenetic relationship and substrate specificity, these genes were divided into eight classes by using well-characterized *Arabidopsis* SDG members as references. In addition, we analyzed the expression profiles of *D. catenatum* SDG genes in different tissues and organs and their responses to diverse environmental stresses. Our findings provide comprehensive information on the classification and expression profiles on *D. catenatum* SDG genes, and will lay the foundation for the functional characterization of the SDG gene family in orchids.

Methods

Identification of SDG gene family in *D. catenatum* and *P. equestris*

All the SDG sequences including genomic DNAs, CDS and proteins in *Oryza sativa* and *Arabidopsis thaliana* were retrieved from the plant genomics resource Phytozome v12 (<http://phytozome.jgi.doe.gov/pz/portal.html>). Of these, the protein sequences were used as queries to search homologs of *D. catenatum* and *P. equestris* against NCBI database using BLASTp tool. All the hits

were further confirmed by the existence of the signature SET domain detected using on-line biosofts PROSITE (<http://prosite.expasy.org/>), SMART (<http://smart.embl-heidelberg.de/>) and Pfam (<http://pfam.xfam.org/search>).

Analysis of gene structure, domain architecture and phylogenetic relationship

The gene structure including the intron-exon distribution pattern was reconstructed by Gene Structure Display Server GSDS 2.0 (<http://gsds.cbi.pku.edu.cn/>) [120]. The domain organization was analyzed using SMART and Pfam databases. Phylogenetic analysis was performed using MEGA7 [121]. The full-length amino acid sequences of SDG proteins were used for constructing neighbor-joining (NJ) trees with the following settings: pairwise deletion option for gaps/missing data treatment; p-distance method for Substitution model; and bootstrap test of 1000 replicates for evaluation of internal branch reliability.

In silico expression profiling of *DcSDG* genes

For the tissue and organ expression profiling of *DcSDG* genes, The raw RNA-seq data of leaf (SRR4431601), root (SRX2938667), Green root tip (SRR4431599), white part of root (SRR4431598), stem (SRR4431600), flower bud (SRR4431603), sepal (SRR4431597), labellum (SRR4431602), pollinia (SRR5722145) and gynostemium (SRR4431596) in an individual of wild *D. catenatum* were downloaded from the NCBI Sequence Read Archive (SRA) provided by Zhang *et al* [100]. For drought stress and stress removal experiment in 8-month-old *D. catenatum* plants [101], irrigation was performed on the 1st day, omitted from the 2nd to the 7th day, and resumed on the 8th day, watering every two days at 15:30. Consequently, the raw RNA-seq reads were obtained from the leaves that were harvested at both 06:30 and 18:30 on the 2nd [DR5 (NCBI: SRR7223299) and DR8 (SRR7223300)], 7th [DR6 (SRR7223298) and DR10 (SRR7223296)], and 9th [DR7 (SRR7223301) and DR15 (SRR7223297)] days, respectively, and at 18:30 on the 8th day [DR11 (SRR7223295)]. For the expression analysis of *DcSDG* genes in response to cold stress, the raw RNA-seq reads of leaves under 20 °C control condition (SRR3210630, SRR3210635 and SRR3210636) and 0°C cold treatment for 20 h (SRR3210613, SRR3210621 and SRR3210626) were obtained from NCBI provided by Wu *et al* [43]. Reads of all the samples were aligned to the NCBI *Dendrobium* reference genome using HISAT package [122]. The mapped reads of each sample were assembled using StringTie [123]. Then, all transcriptomes from Samples were merged to reconstruct a comprehensive transcriptome using perl scripts. After the final transcriptome was generated, StringTie and edgeR was used to estimate the expression levels of all transcripts. StringTie was used to perform expression level for mRNAs by calculating FPKM. *DcSDG* genes were selected and differentially expressed genes were defined with log₂ (fold change) >1 or <-1 and with statistical significance (p value < 0.05) by R package. Heatmap was generated using TIGR MultiExperiment Viewer (MeV4.9) software [124].

Plant material and heat shock treatment

D. catenatum cultivar “Jingpin 1” (Breed NO. Zhe R-SV-DO-015-2014, Zhejiang Province, China) constructed by professor Jin-Ping Si as our lab PI was used in this study. *D. catenatum* was grown in greenhouse at 20 °C under a 12h light/12h dark regime. 1-year-old seedlings were treated at 35 °C heat shock for indicated time (3h, 6h and 12h) in a temperature-controlled incubator, compared with the mock plants at 20 °C. Then the leaves were harvested, snap frozen in liquid nitrogen and stored at -70 °C for further expression analysis.

Real-Time quantitative RT-PCR (RT-qPCR)

Total RNA was extracted from the leaves using TRIzol reagent (Invitrogen, USA) followed by RNase-free DNase I treatment. First strand cDNA synthesis was performed via PrimerScript RT Enzyme Mix I kit (TaKaRa, Japan), according to the manufacturer’s instructions. RT-qPCR reaction mixture (10 µl) was prepared according to the manual of SYBR® Premix Ex Taq™ II (Tli RNaseH Plus) kit (TaKaRa, Japan). Then the reaction was carried out on CFX96 Touch™ Real-Time PCR Detection System (BIO-RAD, USA) in three technical replicates for each biological triplicate using the primers listed in Additional file 9. The reaction condition was set as the following temperature profile: 94 °C for 3 m, 40 cycles of 94 °C for 20s, 60 °C for 20s, 72°C for 20s. The constitutive *DcACTIN* was used as the reference gene. The expression value of each gene tested was normalized with the internal reference gene, and the relative expression level was calculated with $2^{-\Delta\Delta CT}$ method [125].

Abbreviations

ABA: abscisic acid; ATX1: *ARABIDOPSIS* TRITHORAX-LIKE PROTEIN1; ATXR3: *ARABIDOPSIS* TRITHORAX RELATED 3; BR: Brassinosteroid; ChIP-Seq: Chromatin immunoprecipitation-based sequencing; CLF: CURLY LEAF; E(Z): zeste; FBA: Fructose-1,6-bisphosphate aldolase; FLC: FLOWERING LOCUS C; HKMTase: Histone lysine methyltransferase; JA: Jasmonic acid; KYP: KRYPTONITE; MEA: MEAEA; NJ: Neighbor-joining; PRC2: Polycomb Repressive Complex 2; RBCMT: Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) methyltransferase; RdDM: RNA-directed DNA methylation; RT-qPCR: Real-time quantitative polymerase chain reaction; Rubis-sub-bind: Rubisco substrate-binding; SA: salicylic acid; SDG: SET DOMAIN GROUP; SET: Su(var), E(z), and Trithorax; SRA: SET- and RING-ASSOCIATED; SMYD3: *SET And MYND Domain Containing 3*; Su(var): Suppressor of variegation; SWN: SWINGER; Trx: Trithorax; WGD: Whole genome duplication

Declarations

Ethics approval and consent to participate

The *D. catenatum* used in this study is a commercial cultivar "Jingpin 1", which was cultivated by Prof Jinping Si (Zhejiang A&F University), was authorized by Zhejiang Province with Breed NO. Zhe R-SV-DO-015-2014. It does not require ethical approval.

Consent to publish

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its Additional files. The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by National Key R&D Program of China (2017YFC1702201) focusing on wild imitation cultivation and stress response mechanisms of *Dendrobium*, National Natural Science Foundation of China (31870310) focusing on epigenetic mechanism of *Dendrobium* epiphytic lifestyle, Scientific R&D Foundation for Talent Start-up Project of Zhejiang A&F University (2017FR035) focusing on the functional dissection of histone methylation in *Dendrobium*, and the State Key Laboratory of Subtropical Silviculture (ZY20180206) focusing on metabolic mechanism of environmental fitness in *Dendrobium*.

Authors' contributions

DHC and JPS planned and designed the research. HLQ performed the experiments. DHC, HLQ, YH and LZ analyzed the data. DHC and JPS wrote the article. All of the authors approved the manuscript.

Acknowledgements

We thank the manuscript editing provided by the professional editing company EnPapers (<http://www.enpapers.com/>).

Authors' Information

¹ State Key Laboratory of Subtropical Silviculture, SFGA Engineering Research Center for *Dendrobium catenatum* (*D. officinale*), Zhejiang A&F University, Lin'an, Hangzhou, Zhejiang 311300, China, ² Key Laboratory of Education Department of Hunan Province on Plant Genetics and Molecular Biology, Hunan Agricultural University, Changsha 410128, China

References

1. Goldberg AD, Allis CD, Bernstein E: **Epigenetics: a landscape takes shape.** *Cell* 2007, **128**(4):635-638.
2. Huang H, Sabari BR, Garcia BA, Allis CD, Zhao Y: **SnapShot: histone modifications.** *Cell* 2014, **159**(2):458-458 e451.
3. Lawrence M, Daujat S, Schneider R: **Lateral Thinking: How Histone Modifications Regulate Gene Expression.** *Trends in genetics : TIG* 2016, **32**(1):42-56.
4. Jenuwein T, Allis CD: **Translating the histone code.** *Science* 2001, **293**(5532):1074-1080.
5. Strahl BD, Allis CD: **The language of covalent histone modifications.** *Nature* 2000, **403**(6765):41-45.
6. Liu C, Lu F, Cui X, Cao X: **Histone methylation in higher plants.** *Annual review of plant biology* 2010, **61**:395-420.
7. Tschiersch B, Hofmann A, Krauss V, Dorn R, Korge G, Reuter G: **The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes.** *The EMBO journal* 1994, **13**(16):3822-3831.
8. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y: **Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain.** *Current biology : CB* 2002, **12**(12):1052-1058.
9. Zhang Y, Reinberg D: **Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails.** *Genes & development* 2001, **15**(18):2343-2360.
10. Ng DW, Wang T, Chandrasekharan MB, Aramayo R, Kertbundit S, Hall TC: **Plant SET domain-containing proteins: structure, function and regulation.** *Biochimica et biophysica acta* 2007, **1769**(5-6):316-329.
11. Huang Y, Liu C, Shen WH, Ruan Y: **Phylogenetic analysis and classification of the Brassica rapa SET-domain protein family.** *BMC plant biology* 2011, **11**:175.
12. Mujtaba S, Manzur KL, Gurnon JR, Kang M, Van Etten JL, Zhou MM: **Epigenetic transcriptional repression of cellular genes by a viral SET protein.** *Nature cell biology* 2008, **10**(9):1114-1122.
13. Alvarez-Venegas R, Sadler M, Tikhonov A, Avramova Z: **Origin of the bacterial SET domain genes: vertical or horizontal?** *Molecular biology and evolution* 2007, **24**(2):482-497.
14. Aravind L, Iyer LM: **Provenance of SET-domain histone methyltransferases through duplication of a simple structural unit.** *Cell cycle* 2003, **2**(4):369-376.
15. Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q *et al*: **Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.** *Science* 1998, **282**(5389):754-759.
16. Alvarez-Venegas R: **Bacterial SET domain proteins and their role in eukaryotic chromatin modification.** *Frontiers in genetics* 2014, **5**:65.

17. Pontvianne F, Blevins T, Pikaard CS: **Arabidopsis Histone Lysine Methyltransferases**. *Advances in botanical research* 2010, **53**:1-22.
18. Aquea F, Vega A, Timmermann T, Poupin MJ, Arce-Johnson P: **Genome-wide analysis of the SET DOMAIN GROUP family in grapevine**. *Plant cell reports* 2011, **30**(6):1087-1097.
19. Lei L, Zhou SL, Ma H, Zhang LS: **Expansion and diversification of the SET domain gene family following whole-genome duplications in Populus trichocarpa**. *BMC evolutionary biology* 2012, **12**:51.
20. Qian Y, Xi Y, Cheng B, Zhu S: **Genome-wide identification and expression profiling of DNA methyltransferase gene family in maize**. *Plant cell reports* 2014, **33**(10):1661-1672.
21. Lu Z, Huang X, Ouyang Y, Yao J: **Genome-wide identification, phylogenetic and co-expression analysis of OsSET gene family in rice**. *PloS one* 2013, **8**(6):e65426.
22. Aiese Cigliano R, Sanseverino W, Cremona G, Ercolano MR, Conicella C, Consiglio FM: **Genome-wide analysis of histone modifiers in tomato: gaining an insight into their developmental roles**. *BMC genomics* 2013, **14**:57.
23. Xu J, Xu H, Liu Y, Wang X, Xu Q, Deng X: **Genome-wide identification of sweet orange (Citrus sinensis) histone modification gene families and their expression analysis during the fruit development and fruit-blue mold infection process**. *Frontiers in plant science* 2015, **6**:607.
24. Huang Y, Mo Y, Chen P, Yuan X, Meng F, Zhu S, Liu Z: **Identification of SET Domain-Containing Proteins in Gossypium raimondii and Their Response to High Temperature Stress**. *Scientific reports* 2016, **6**:32729.
25. Yadav CB, Muthamilarasan M, Dangi A, Shweta S, Prasad M: **Comprehensive analysis of SET domain gene family in foxtail millet identifies the putative role of SiSET14 in abiotic stress tolerance**. *Scientific reports* 2016, **6**:32621.
26. Peng M, Ying P, Liu X, Li C, Xia R, Li J, Zhao M: **Genome-Wide Identification of Histone Modifiers and Their Expression Patterns during Fruit Abscission in Litchi**. *Frontiers in plant science* 2017, **8**:639.
27. Berr A, McCallum EJ, Alioua A, Heintz D, Heitz T, Shen WH: **Arabidopsis histone methyltransferase SET DOMAIN GROUP8 mediates induction of the jasmonate/ethylene pathway genes in plant defense response to necrotrophic fungi**. *Plant physiology* 2010, **154**(3):1403-1414.
28. Berr A, McCallum EJ, Menard R, Meyer D, Fuchs J, Dong A, Shen WH: **Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development**. *The Plant cell* 2010, **22**(10):3232-3248.
29. Cazzonelli CI, Cuttriss AJ, Cossetto SB, Pye W, Crisp P, Whelan J, Finnegan EJ, Turnbull C, Pogson BJ: **Regulation of carotenoid composition and shoot branching in Arabidopsis by a chromatin modifying histone methyltransferase, SDG8**. *The Plant cell* 2009, **21**(1):39-53.
30. Cazzonelli CI, Nisar N, Roberts AC, Murray KD, Borevitz JO, Pogson BJ: **A chromatin modifying enzyme, SDG8, is involved in morphological, gene expression, and epigenetic responses to mechanical stimulation**. *Frontiers in plant science* 2014, **5**:533.
31. Dong G, Ma DP, Li J: **The histone methyltransferase SDG8 regulates shoot branching in Arabidopsis**. *Biochemical and biophysical research communications* 2008, **373**(4):659-664.

32. Sui P, Jin J, Ye S, Mu C, Gao J, Feng H, Shen WH, Yu Y, Dong A: **H3K36 methylation is critical for brassinosteroid-regulated plant growth and development in rice.** *The Plant journal : for cell and molecular biology* 2012, **70**(2):340-347.
33. Palma K, Thorgrimsen S, Malinovsky FG, Fiil BK, Nielsen HB, Brodersen P, Hofius D, Petersen M, Mundy J: **Autoimmunity in Arabidopsis acd11 is mediated by epigenetic regulation of an immune receptor.** *PLoS pathogens* 2010, **6**(10):e1001137.
34. Alvarez-Venegas R, Abdallat AA, Guo M, Alfano JR, Avramova Z: **Epigenetic control of a transcription factor at the cross section of two antagonistic pathways.** *Epigenetics* 2007, **2**(2):106-113.
35. Ding Y, Lapko H, Ndamukong I, Xia Y, Al-Abdallat A, Lalithambika S, Sadler M, Saleh A, Fromm M, Riethoven JJ *et al*: **The Arabidopsis chromatin modifier ATX1, the myotubularin-like AtMTM and the response to drought.** *Plant signaling & behavior* 2009, **4**(11):1049-1058.
36. Ding Y, Avramova Z, Fromm M: **The Arabidopsis trithorax-like factor ATX1 functions in dehydration stress responses via ABA-dependent and ABA-independent pathways.** *The Plant journal : for cell and molecular biology* 2011, **66**(5):735-744.
37. van Dijk K, Ding Y, Malkaram S, Riethoven JJ, Liu R, Yang J, Laczko P, Chen H, Xia Y, Ladunga I *et al*: **Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in Arabidopsis thaliana.** *BMC plant biology* 2010, **10**:238.
38. Zong W, Zhong X, You J, Xiong L: **Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress.** *Plant molecular biology* 2013, **81**(1-2):175-188.
39. Kwon CS, Lee D, Choi G, Chung WI: **Histone occupancy-dependent and -independent removal of H3K27 trimethylation at cold-responsive genes in Arabidopsis.** *The Plant journal : for cell and molecular biology* 2009, **60**(1):112-121.
40. Yan L, Wang X, Liu H, Tian Y, Lian J, Yang R, Hao S, Wang X, Yang S, Li Q *et al*: **The Genome of Dendrobium officinale Illuminates the Biology of the Important Traditional Chinese Orchid Herb.** *Molecular plant* 2015, **8**(6):922-934.
41. Tang H, Zhao T, Sheng Y, Zheng T, Fu L, Zhang Y: **Dendrobium officinale Kimura et Migo: A Review on Its Ethnopharmacology, Phytochemistry, Pharmacology, and Industrialization.** *Evidence-based complementary and alternative medicine : eCAM* 2017, **2017**:7436259.
42. Ng TB, Liu J, Wong JH, Ye X, Wing Sze SC, Tong Y, Zhang KY: **Review of research on Dendrobium, a prized folk medicine.** *Applied microbiology and biotechnology* 2012, **93**(5):1795-1803.
43. Wu ZG, Jiang W, Chen SL, Mantri N, Tao ZM, Jiang CX: **Insights from the Cold Transcriptome and Metabolome of Dendrobium officinale: Global Reprogramming of Metabolic and Gene Regulation Networks during Cold Acclimation.** *Frontiers in plant science* 2016, **7**:1653.
44. Zheng B, Chen X: **Dynamics of histone H3 lysine 27 trimethylation in plant development.** *Current opinion in plant biology* 2011, **14**(2):123-129.
45. Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G: **A Polycomb-group gene regulates homeotic gene expression in Arabidopsis.** *Nature* 1997, **386**(6620):44-51.

46. Grossniklaus U, Vielle-Calzada JP, Hoepfner MA, Gagliano WB: **Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis.** *Science* 1998, **280**(5362):446-450.
47. Kiyosue T, Ohad N, Yadegari R, Hannon M, Dinneny J, Wells D, Katz A, Margossian L, Harada JJ, Goldberg RB *et al.*: **Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis.** *Proceedings of the National Academy of Sciences of the United States of America* 1999, **96**(7):4186-4191.
48. Liu X, Zhou C, Zhao Y, Zhou S, Wang W, Zhou DX: **The rice enhancer of zeste [E(z)] genes SDG711 and SDG718 are respectively involved in long day and short day signaling to mediate the accurate photoperiod control of flowering time.** *Frontiers in plant science* 2014, **5**:591.
49. Springer NM, Napoli CA, Selinger DA, Pandey R, Cone KC, Chandler VL, Kaeppeler HF, Kaeppeler SM: **Comparative analysis of SET domain proteins in maize and Arabidopsis reveals multiple duplications preceding the divergence of monocots and dicots.** *Plant physiology* 2003, **132**(2):907-925.
50. Lee J, Yun JY, Zhao W, Shen WH, Amasino RM: **A methyltransferase required for proper timing of the vernalization response in Arabidopsis.** *Proceedings of the National Academy of Sciences of the United States of America* 2015, **112**(7):2269-2274.
51. Sahr T, Adam T, Fizames C, Maurel C, Santoni V: **O-carboxyl- and N-methyltransferases active on plant aquaporins.** *Plant & cell physiology* 2010, **51**(12):2092-2104.
52. Sun C, Fang J, Zhao T, Xu B, Zhang F, Liu L, Tang J, Zhang G, Deng X, Chen F *et al.*: **The histone methyltransferase SDG724 mediates H3K36me_{2/3} deposition at MADS50 and RFT1 and promotes flowering in rice.** *The Plant cell* 2012, **24**(8):3235-3247.
53. Cartagena JA, Matsunaga S, Seki M, Kurihara D, Yokoyama M, Shinozaki K, Fujimoto S, Azumi Y, Uchiyama S, Fukui K: **The Arabidopsis SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen.** *Developmental biology* 2008, **315**(2):355-368.
54. Thorstensen T, Grini PE, Mercy IS, Alm V, Erdal S, Aasland R, Aalen RB: **The Arabidopsis SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS).** *Plant molecular biology* 2008, **66**(1-2):47-59.
55. Kumpf R, Thorstensen T, Rahman MA, Heyman J, Nenseth HZ, Lammens T, Herrmann U, Swarup R, Veiseth SV, Emberland G *et al.*: **The ASH1-RELATED3 SET-domain protein controls cell division competence of the meristem and the quiescent center of the Arabidopsis primary root.** *Plant physiology* 2014, **166**(2):632-643.
56. Berr A, Shafiq S, Pinon V, Dong A, Shen WH: **The trxG family histone methyltransferase SET DOMAIN GROUP 26 promotes flowering via a distinctive genetic pathway.** *The Plant journal : for cell and molecular biology* 2015, **81**(2):316-328.
57. Xu L, Zhao Z, Dong A, Soubigou-Taconnat L, Renou JP, Steinmetz A, Shen WH: **Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering**

- time regulation and other processes in Arabidopsis thaliana.** *Molecular and cellular biology* 2008, **28**(4):1348-1360.
58. Liu B, Wei G, Shi J, Jin J, Shen T, Ni T, Shen WH, Yu Y, Dong A: **SET DOMAIN GROUP 708, a histone H3 lysine 36-specific methyltransferase, controls flowering time in rice (Oryza sativa).** *The New phytologist* 2016, **210**(2):577-588.
59. Lee S, Fu F, Xu S, Lee SY, Yun DJ, Mengiste T: **Global Regulation of Plant Immunity by Histone Lysine Methyl Transferases.** *The Plant cell* 2016, **28**(7):1640-1661.
60. Zhao Z, Yu Y, Meyer D, Wu C, Shen WH: **Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36.** *Nature cell biology* 2005, **7**(12):1256-1260.
61. Pien S, Fleury D, Mylne JS, Crevillen P, Inze D, Avramova Z, Dean C, Grossniklaus U: **ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation.** *The Plant cell* 2008, **20**(3):580-588.
62. Saleh A, Alvarez-Venegas R, Yilmaz M, Le O, Hou G, Sadler M, Al-Abdallat A, Xia Y, Lu G, Ladunga I *et al*: **The highly similar Arabidopsis homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions.** *The Plant cell* 2008, **20**(3):568-579.
63. Alvarez-Venegas R, Pien S, Sadler M, Witmer X, Grossniklaus U, Avramova Z: **ATX-1, an Arabidopsis homolog of trithorax, activates flower homeotic genes.** *Current biology : CB* 2003, **13**(8):627-637.
64. Choi SC, Lee S, Kim SR, Lee YS, Liu C, Cao X, An G: **Trithorax group protein Oryza sativa Trithorax1 controls flowering time in rice via interaction with early heading date3.** *Plant physiology* 2014, **164**(3):1326-1337.
65. Jiang P, Wang S, Jiang H, Cheng B, Wu K, Ding Y: **The COMPASS-Like Complex Promotes Flowering and Panicle Branching in Rice.** *Plant physiology* 2018, **176**(4):2761-2771.
66. Jiang P, Wang S, Zheng H, Li H, Zhang F, Su Y, Xu Z, Lin H, Qian Q, Ding Y: **SIP1 participates in regulation of flowering time in rice by recruiting OsTrx1 to Ehd1.** *The New phytologist* 2018.
67. Chen LQ, Luo JH, Cui ZH, Xue M, Wang L, Zhang XY, Pawlowski WP, He Y: **ATX3, ATX4, and ATX5 Encode Putative H3K4 Methyltransferases and Are Critical for Plant Development.** *Plant physiology* 2017, **174**(3):1795-1806.
68. Jiang P, Wang S, Ikram AU, Xu Z, Jiang H, Cheng B, Ding Y: **SDG721 and SDG705 are required for rice growth.** *Journal of integrative plant biology* 2018, **60**(7):530-535.
69. Berr A, Xu L, Gao J, Cognat V, Steinmetz A, Dong A, Shen WH: **SET DOMAIN GROUP25 encodes a histone methyltransferase and is involved in FLOWERING LOCUS C activation and repression of flowering.** *Plant physiology* 2009, **151**(3):1476-1485.
70. Yun JY, Tamada Y, Kang YE, Amasino RM: **Arabidopsis trithorax-related3/SET domain GROUP2 is required for the winter-annual habit of Arabidopsis thaliana.** *Plant & cell physiology* 2012, **53**(5):834-846.
71. Pinon V, Yao X, Dong A, Shen WH: **SDG2-Mediated H3K4me3 Is Crucial for Chromatin Condensation and Mitotic Division during Male Gametogenesis in Arabidopsis.** *Plant physiology* 2017, **174**(2):1205-1215.

72. Jacob Y, Feng S, LeBlanc CA, Bernatavichute YV, Stroud H, Cokus S, Johnson LM, Pellegrini M, Jacobsen SE, Michaels SD: **ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing.** *Nature structural & molecular biology* 2009, **16**(7):763-768.
73. Jacob Y, Stroud H, Leblanc C, Feng S, Zhuo L, Caro E, Hassel C, Gutierrez C, Michaels SD, Jacobsen SE: **Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases.** *Nature* 2010, **466**(7309):987-991.
74. Jacob Y, Bergamin E, Donoghue MT, Mongeon V, LeBlanc C, Voigt P, Underwood CJ, Brunzelle JS, Michaels SD, Reinberg D *et al.*: **Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication.** *Science* 2014, **343**(6176):1249-1253.
75. Ma Z, Castillo-Gonzalez C, Wang Z, Sun D, Hu X, Shen X, Potok ME, Zhang X: **Arabidopsis Serrate Coordinates Histone Methyltransferases ATXR5/6 and RNA Processing Factor RDR6 to Regulate Transposon Expression.** *Developmental cell* 2018, **45**(6):769-784 e766.
76. Raynaud C, Sozzani R, Glab N, Domenichini S, Perennes C, Cella R, Kondorosi E, Bergounioux C: **Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis.** *The Plant journal : for cell and molecular biology* 2006, **47**(3):395-407.
77. Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, Singh PB, Cheng X, Schubert I, Jenuwein T, Jacobsen SE: **Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana.** *Chromosoma* 2004, **112**(6):308-315.
78. Li S, Liu L, Li S, Gao L, Zhao Y, Kim YJ, Chen X: **SUVH1, a Su(var)3-9 family member, promotes the expression of genes targeted by DNA methylation.** *Nucleic acids research* 2016, **44**(2):608-620.
79. Qin FJ, Sun QW, Huang LM, Chen XS, Zhou DX: **Rice SUVH histone methyltransferase genes display specific functions in chromatin modification and retrotransposon repression.** *Molecular plant* 2010, **3**(4):773-782.
80. Fischer A, Hofmann I, Naumann K, Reuter G: **Heterochromatin proteins and the control of heterochromatic gene silencing in Arabidopsis.** *Journal of plant physiology* 2006, **163**(3):358-368.
81. Kuhlmann M, Mette MF: **Developmentally non-redundant SET domain proteins SUVH2 and SUVH9 are required for transcriptional gene silencing in Arabidopsis thaliana.** *Plant molecular biology* 2012, **79**(6):623-633.
82. Naumann K, Fischer A, Hofmann I, Krauss V, Phalke S, Irmeler K, Hause G, Aurich AC, Dorn R, Jenuwein T *et al.*: **Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis.** *The EMBO journal* 2005, **24**(7):1418-1429.
83. Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, Zhong X, Marson G, Pellegrini M, Segal DJ *et al.*: **SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation.** *Nature* 2014, **507**(7490):124-128.
84. Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE: **SRA-domain proteins required for DRM2-mediated de novo DNA methylation.** *PLoS genetics* 2008, **4**(11):e1000280.
85. Liu ZW, Shao CR, Zhang CJ, Zhou JX, Zhang SW, Li L, Chen S, Huang HW, Cai T, He XJ: **The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA**

- methylation loci.** *PLoS genetics* 2014, **10**(1):e1003948.
86. Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S, Vashisht AA, Wohlschlegel JA, Patel DJ, Jacobsen SE: **Mechanism of DNA methylation-directed histone methylation by KRYPTONITE.** *Molecular cell* 2014, **55**(3):495-504.
87. Ebbs ML, Bartee L, Bender J: **H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases.** *Molecular and cellular biology* 2005, **25**(23):10507-10515.
88. Ebbs ML, Bender J: **Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase.** *The Plant cell* 2006, **18**(5):1166-1176.
89. Jackson JP, Lindroth AM, Cao X, Jacobsen SE: **Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase.** *Nature* 2002, **416**(6880):556-560.
90. Malagnac F, Bartee L, Bender J: **An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation.** *The EMBO journal* 2002, **21**(24):6842-6852.
91. Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, Reinberg D, Patel DJ, Jacobsen SE: **A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo.** *Genes & development* 2011, **25**(2):137-152.
92. Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE: **The SRA methylcytosine-binding domain links DNA and histone methylation.** *Current biology : CB* 2007, **17**(4):379-384.
93. Ding Y, Wang X, Su L, Zhai J, Cao S, Zhang D, Liu C, Bi Y, Qian Q, Cheng Z *et al.* **SDG714, a histone H3K9 methyltransferase, is involved in Tos17 DNA methylation and transposition in rice.** *The Plant cell* 2007, **19**(1):9-22.
94. Han YF, Dou K, Ma ZY, Zhang SW, Huang HW, Li L, Cai T, Chen S, Zhu JK, He XJ: **SUVR2 is involved in transcriptional gene silencing by associating with SNF2-related chromatin-remodeling proteins in Arabidopsis.** *Cell research* 2014, **24**(12):1445-1465.
95. Veiseth SV, Rahman MA, Yap KL, Fischer A, Egge-Jacobsen W, Reuter G, Zhou MM, Aalen RB, Thorstensen T: **The SUVR4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in Arabidopsis.** *PLoS genetics* 2011, **7**(3):e1001325.
96. Rahman MA, Kristiansen PE, Veiseth SV, Andersen JT, Yap KL, Zhou MM, Sandlie I, Thorstensen T, Aalen RB: **The arabidopsis histone methyltransferase SUVR4 binds ubiquitin via a domain with a four-helix bundle structure.** *Biochemistry* 2014, **53**(13):2091-2100.
97. Thorstensen T, Fischer A, Sandvik SV, Johnsen SS, Grini PE, Reuter G, Aalen RB: **The Arabidopsis SUVR4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9.** *Nucleic acids research* 2006, **34**(19):5461-5470.
98. Caro E, Stroud H, Greenberg MV, Bernatavichute YV, Feng S, Groth M, Vashisht AA, Wohlschlegel J, Jacobsen SE: **The SET-domain protein SUVR5 mediates H3K9me2 deposition and silencing at stimulus response genes in a DNA methylation-independent manner.** *PLoS genetics* 2012, **8**(10):e1002995.

99. Krichevsky A, Gutgarts H, Kozlovsky SV, Tzfira T, Sutton A, Sternglanz R, Mandel G, Citovsky V: **C2H2 zinc finger-SET histone methyltransferase is a plant-specific chromatin modifier.** *Developmental biology* 2007, **303**(1):259-269.
100. Zhang GQ, Liu KW, Li Z, Lohaus R, Hsiao YY, Niu SC, Wang JY, Lin YC, Xu Q, Chen LJ *et al*: **The Apostasia genome and the evolution of orchids.** *Nature* 2017, **549**(7672):379-383.
101. Zou LH, Wan X, Deng H, Zheng BQ, Li BJ, Wang Y: **RNA-seq transcriptomic profiling of crassulacean acid metabolism pathway in Dendrobium catenatum.** *Scientific data* 2018, **5**:180252.
102. Zhang GQ, Xu Q, Bian C, Tsai WC, Yeh CM, Liu KW, Yoshida K, Zhang LS, Chang SB, Chen F *et al*: **The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution.** *Scientific reports* 2016, **6**:19029.
103. Brown MA, Sims RJ, 3rd, Gottlieb PD, Tucker PW: **Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex.** *Molecular cancer* 2006, **5**:26.
104. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y: **SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells.** *Nature cell biology* 2004, **6**(8):731-740.
105. Van Aller GS, Reynoird N, Barbash O, Huddleston M, Liu S, Zmoos AF, McDevitt P, Sinnamon R, Le B, Mas G *et al*: **Smyd3 regulates cancer cell phenotypes and catalyzes histone H4 lysine 5 methylation.** *Epigenetics* 2012, **7**(4):340-343.
106. Huang J, Perez-Burgos L, Placek BJ, Sengupta R, Richter M, Dorsey JA, Kubicek S, Opravil S, Jenuwein T, Berger SL: **Repression of p53 activity by Smyd2-mediated methylation.** *Nature* 2006, **444**(7119):629-632.
107. Kunizaki M, Hamamoto R, Silva FP, Yamaguchi K, Nagayasu T, Shibuya M, Nakamura Y, Furukawa Y: **The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by SMYD3.** *Cancer research* 2007, **67**(22):10759-10765.
108. Mazur PK, Reynoird N, Khatri P, Jansen PW, Wilkinson AW, Liu S, Barbash O, Van Aller GS, Huddleston M, Dhanak D *et al*: **SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer.** *Nature* 2014, **510**(7504):283-287.
109. Zhang X, Tanaka K, Yan J, Li J, Peng D, Jiang Y, Yang Z, Barton MC, Wen H, Shi X: **Regulation of estrogen receptor alpha by histone methyltransferase SMYD2-mediated protein methylation.** *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(43):17284-17289.
110. Ma S, Martin-Laffon J, Mininno M, Gigarel O, Brugiere S, Bastien O, Tardif M, Ravel S, Alban C: **Molecular Evolution of the Substrate Specificity of Chloroplastic Aldolases/Rubisco Lysine Methyltransferases in Plants.** *Molecular plant* 2016, **9**(4):569-581.
111. Ying Z, Mulligan RM, Janney N, Houtz RL: **Rubisco small and large subunit N-methyltransferases. Bi- and mono-functional methyltransferases that methylate the small and large subunits of Rubisco.** *The Journal of biological chemistry* 1999, **274**(51):36750-36756.

112. Gao ZP, Yu QB, Zhao TT, Ma Q, Chen GX, Yang ZN: **A functional component of the transcriptionally active chromosome complex, Arabidopsis pTAC14, interacts with pTAC12/HEMERA and regulates plastid gene expression.** *Plant physiology* 2011, **157**(4):1733-1745.
113. Kim DW, Kim KB, Kim JY, Seo SB: **Characterization of a novel histone H3K36 methyltransferase setd3 in zebrafish.** *Bioscience, biotechnology, and biochemistry* 2011, **75**(2):289-294.
114. Clark JW, Donoghue PCJ: **Whole-Genome Duplication and Plant Macroevolution.** *Trends in plant science* 2018, **23**(10):933-945.
115. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**(5494):1151-1155.
116. Mozgova I, Kohler C, Hennig L: **Keeping the gate closed: functions of the polycomb repressive complex PRC2 in development.** *The Plant journal : for cell and molecular biology* 2015, **83**(1):121-132.
117. Yu CW, Tai R, Wang SC, Yang P, Luo M, Yang S, Cheng K, Wang WC, Cheng YS, Wu K: **HISTONE DEACETYLASE6 Acts in Concert with Histone Methyltransferases SUVH4, SUVH5, and SUVH6 to Regulate Transposon Silencing.** *The Plant cell* 2017, **29**(8):1970-1983.
118. Wang X, Chen J, Xie Z, Liu S, Nolan T, Ye H, Zhang M, Guo H, Schnable PS, Li Z *et al*: **Histone lysine methyltransferase SDG8 is involved in brassinosteroid-regulated gene expression in Arabidopsis thaliana.** *Molecular plant* 2014, **7**(8):1303-1315.
119. Yang H, Howard M, Dean C: **Antagonistic roles for H3K36me3 and H3K27me3 in the cold-induced epigenetic switch at Arabidopsis FLC.** *Current biology : CB* 2014, **24**(15):1793-1797.
120. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G: **GSDS 2.0: an upgraded gene feature visualization server.** *Bioinformatics* 2015, **31**(8):1296-1297.
121. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Molecular biology and evolution* 2016, **33**(7):1870-1874.
122. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory requirements.** *Nature methods* 2015, **12**(4):357-360.
123. Perteua M, Perteua GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: **StringTie enables improved reconstruction of a transcriptome from RNA-seq reads.** *Nature biotechnology* 2015, **33**(3):290-295.
124. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M *et al*: **TM4: a free, open-source system for microarray data management and analysis.** *BioTechniques* 2003, **34**(2):374-378.
125. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25**(4):402-408.

Tables

Table 1 Classification of SDG genes in *D. catenatum*.

No.	Name	Gene ID	Clade	exon number	Protein (aa)	Isoform
1	DcCLF	LOC110113338	I	16	918	2
2	DcSWN	LOC110098601	I	17	895	0
3	DcASHH1	LOC110107485	II	10	498	0
4	DcASHH2	LOC110097541	II	18	1,992	0
5	DcASHH3a	LOC110095009	II	9	377	0
6	DcASHH3b	LOC110105604	II	9	235	0
7	DcASHR3	LOC110103080	II	11	487	1
8	DcATX1	LOC110096723	III	25	1,081	0
9	DcATX3a	LOC110094707	III	22	980	0
10	DcATX3b	LOC110093543	III	24	1,060	0
11	DcATX3c	LOC110116505	III	26	942	4
12	DcATX3d	LOC110100284	III	19	2,038	0
13	DcATXR3	LOC110093028	M	21	2,360	0
14	DcATXR7	LOC110110427	M	24	1,304	6
15	DcATXR5a	LOC110100023	IV	7	413	3
16	DcATXR5b	LOC110102586	IV	8	373	3
17	DcATXR6	LOC110107042	IV	4	268	0
18	DcSUVH1a	LOC110091884	V	3	682	0
19	DcSUVH1b	LOC110100064	V	3	680	1
20	DcSUVH2a	LOC110104355	V	3	664	1
21	DcSUVH2b	LOC110111597	V	2	640	0
22	DcSUVH4	LOC110103468	V	20	752	4
23	DcSUVH45	LOC110096224	V	2	765	1
24	DcSUVH5a	LOC110109813	V	2	1,048	1
25	DcSUVH5b	LOC110101480	V	2	1,099	0
26	DcSUVR14a	LOC110104063	V	13	783	8
27	DcSUVR14b	LOC110097487	V	9	668	0
28	DcSUVR14c	LOC110097357	V	8	785	0
29	DcSUVR3	LOC110096832	V	2	337	0
30	DcSUVR4	LOC110109382	V	13	723	8
31	DcSUVR5	LOC110092659	V	14	1,658	2
32	DcASHR1	LOC110098638	VI	14	496	0
33	DcASHR2	LOC110112079	VI	3	385	2
34	DcATXR1	LOC110097453	VI	1	521	0
35	DcATXR2	LOC110107001	VI	15	494	0

36	DcATXR4	LOC110103199	VI	8	339	0
37	DcSDG42	LOC110104315	VI	14	765	4
38	DcSDG45	LOC110107875	VII	6	490	0
39	DcSDG46	LOC110111739	VII	6	494	1
40	DcSDG47	LOC110091896	VII	11	556	2
41	DcSDG48	LOC110099042	VII	12	499	0
42	DcSDG49	LOC110092290	VII	16	476	5
43	DcSDG50	LOC110113229	VII	8	482	2
44	DcSDG51	LOC110094849	VII	13	503	0

Additional File Legends

Additional file 1: SDG protein sequences in Arabidopsis and rice retrieved from Phytozome 12 database.

Additional file 2: Identification and classification of SDG genes in *P. equestris*. Sequences and information of *P. equestris* SDG genes and proteins came from NCBI database (<https://www.ncbi.nlm.nih.gov/>).

Additional file 3: Expression data of *DcSDG* genes from different tissues and organs in *D. catenatum*. The FPKM values of *DcSDG* genes in different tissues and organs was used for expression analysis in Fig. 6. Lf: leaf, Ro: root, Gr: green root tip, Wr: white part of root, St: stem, Fb: flower bud, Se: sepal, Lb: labellum, Po: pollinia, and Gs, gynostemium.

Additional file 4: Expression patterns of duplicated SDG gene pairs. The FPKM values of the duplicated *DcSDG* genes in different tissues and organs was used for comparison. Lf: leaf, Ro: root, Gr: green root tip, Wr: white part of root, St: stem, Fb: flower bud, Se: sepal, Lb: labellum, Po: pollinia, and Gs, gynostemium.

Additional file 5: Expression data of *DcSDG* genes from different drought treatments. The FPKM values of *DcSDG* genes in leaves under different drought treatments were used for expression analysis in **Fig. 7**. The seedlings were watered on the 1st day, dried from the 2nd to the 7th day, and re-watered on the 8th day. Leaves were collected at different times; DR5/ DR8, DR6/DR10, and DR7/DR15 indicate sampling at 06:30 and 18:30 on the 2nd, 7th, and 9th days, respectively, and DR11 indicates sampling at 18:30 on the 8th day.

Additional file 6: Expression data of *DcSDG* genes in the absence and presence of cold treatment. The FPKM values of *DcSDGs* genes in leaves under cold stress / 20 °C (control) for 20 h were used for expression analysis in **Fig. 8**.

Additional file 7: Domain organization and gene structure of the class-VI DcSDGs. The NJ tree was generated using MEGA7 with parameter settings as Fig. 1 based on full-length amino acid sequences of Class-VI SDGs in *D. catenatum*, *P. equestris*, Arabidopsis and rice. The number along the tree branch indicates bootstrap value. Different conserved protein domains are colored as indicated. Gene structures of SDGs in each species were indicated in distinct colors. The solid boxes represent exons and black lines represent introns.

Additional file 8: Domain organization and gene structure of the class-VII DcSDGs. The NJ tree was generated using MEGA7 with parameter settings as Fig. 1 based on full-length amino acid sequences of Class-VII SDGs in *D. catenatum*, *P. equestris*, Arabidopsis and rice. The number along the tree branch indicates bootstrap value. Different conserved protein domains are colored as indicated. Gene structures of SDGs in each species were indicated in distinct colors. The solid boxes represent exons and black lines represent introns.

Additional file 9: Primers used for expression analysis of heat shock treatment in this study.

Figures

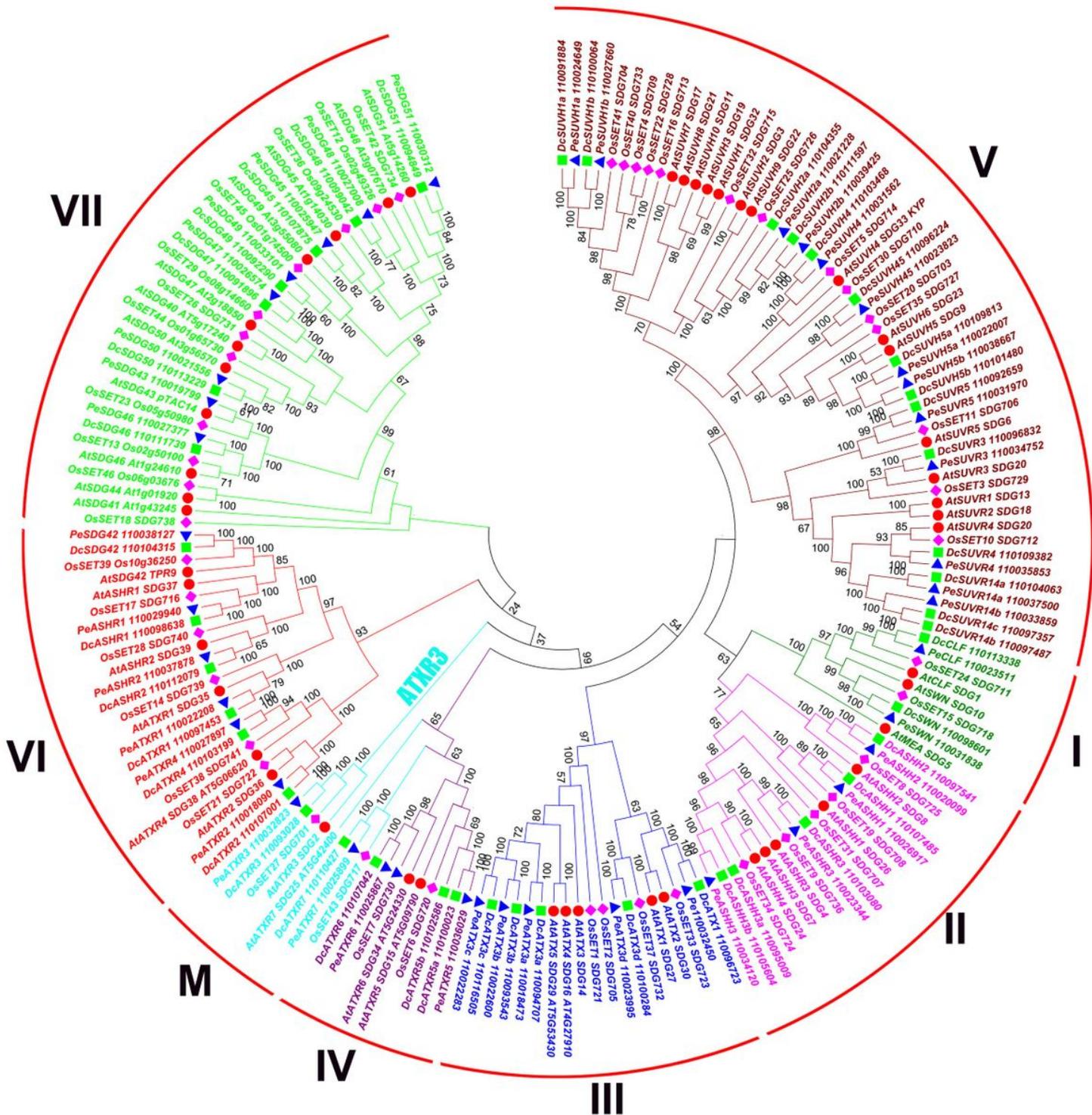


Figure 1

Phylogenetic analysis of SDGs in *D. catenatum*, *P. equestris*, *Arabidopsis* and rice. This tree includes 44 SET domain-containing proteins from *D. catenatum*, 42 from *P. equestris*, 49 from *A. thaliana* and 46 from *O. sativa*. The 181 SET domain-containing proteins were divided into eight classes combined with the further phylogenetic analysis in Fig. 4. SDGs amino acid sequences were aligned using Clustalx, and the phylogenetic tree was constructed using MEGA7 with the following settings: Tree Inference as Neighbor-Joining (NJ); Include Sites as Pairwise deletion option for total sequence analyses; Substitution

Model: p-distance; and Bootstrap test of 1000 replicates for internal branch reliability. Bootstrap values >50% are shown. Dc, *D. catenatum*; Pe, *P. equestris*; At, *A. thaliana*; Os, *O. sativa*.

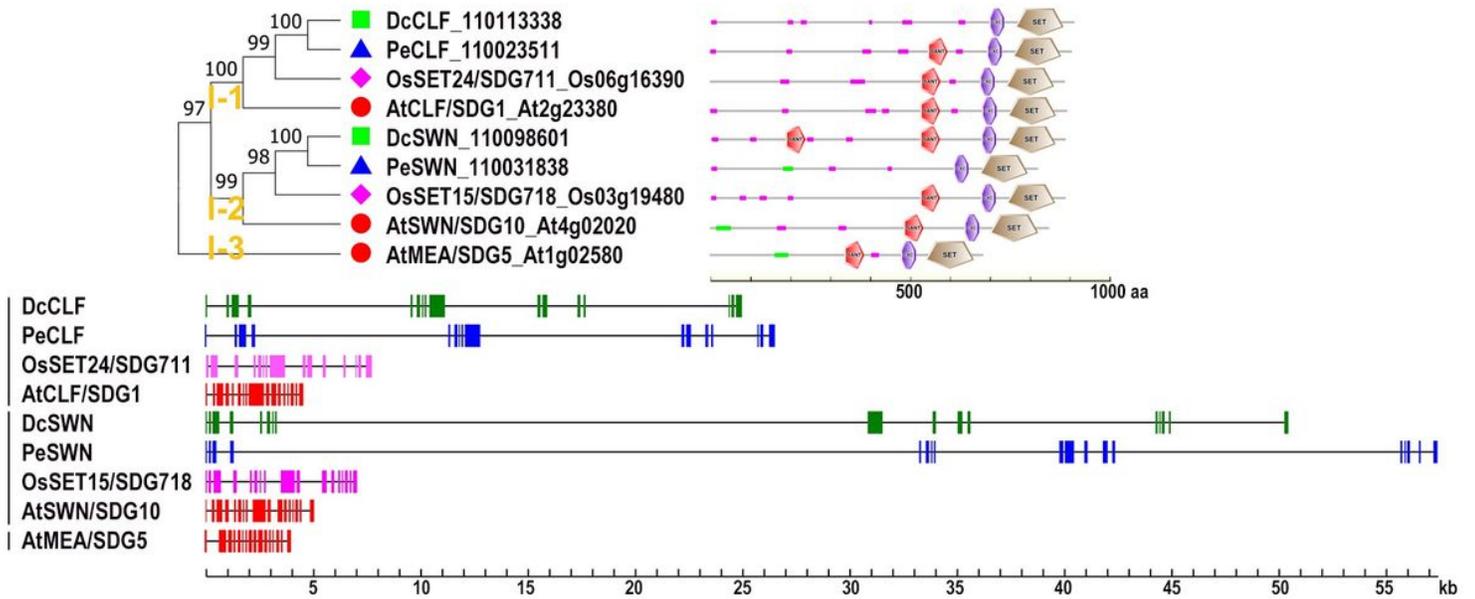


Figure 2

Domain organization and gene structure of the Class I DcSDGs. The NJ tree was generated using MEGA7 with parameter settings as Fig. 1 based on full-length amino acid sequences of Class I SDG proteins in *D. catenatum*, *P. equestris*, *Arabidopsis* and rice. The number along the tree branch indicates bootstrap value. Different conserved protein domains (SANT, CXC, and SET) are colored as indicated. Gene structures of SDGs in each species were indicated in distinct colors. The solid boxes represent exons and black lines represent introns.

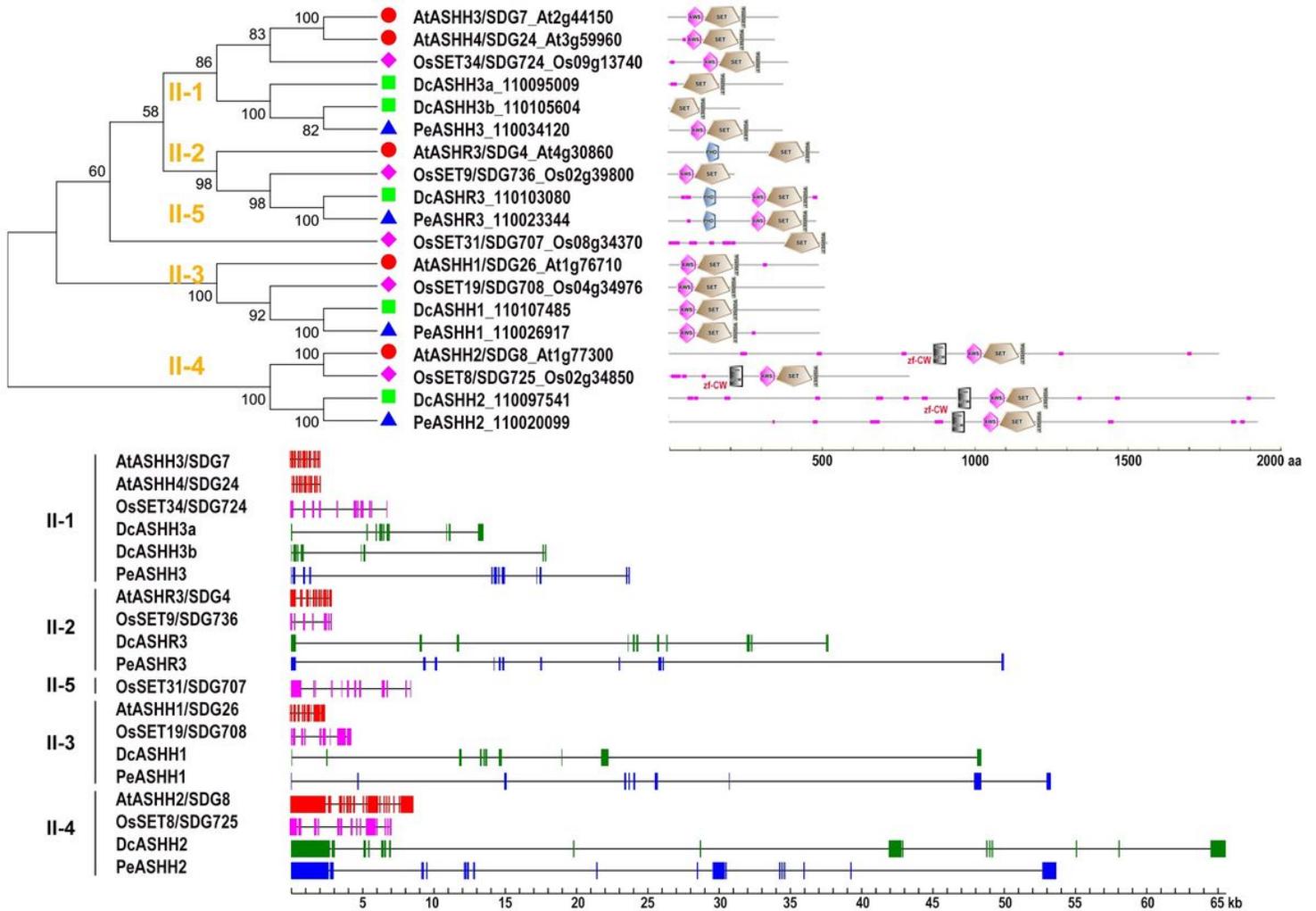


Figure 3

Domain organization and gene structure of the Class II SDGs. The NJ tree was generated using MEGA7 with parameter settings as Fig. 1 based on full-length amino acid sequences of Class II SDG proteins in *D. catenatum*, *P. equestris*, *Arabidopsis* and *rice*. The number along the tree branch indicates bootstrap value. Different conserved protein domains (AWS, PHD, zf-CW, SET and PostSET) are colored as indicated. Gene structures of SDGs in each species were indicated in distinct colors. The solid boxes represent exons and black lines represent introns.

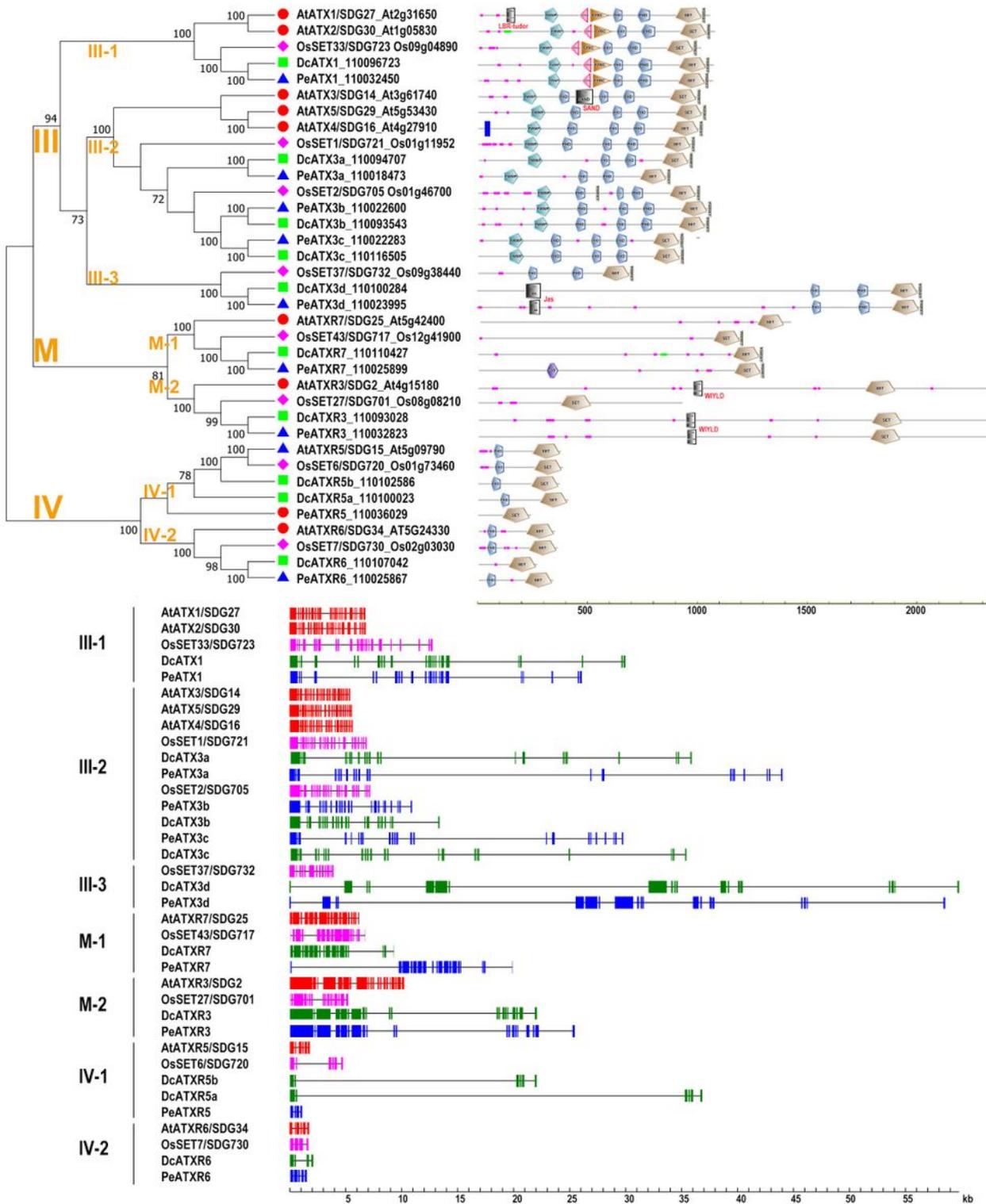


Figure 4

Domain organization and gene structure of the DcSDGs in Classes III, M and IV. The NJ tree was generated using MEGA7 with parameter settings as Fig. 1 based on full-length amino acid sequences of Class III/M/IV SDGs in *D. catenatum*, *P. equestris*, *Arabidopsis* and rice. The number along the tree branch indicates bootstrap value. Different conserved protein domains are colored as indicated. Gene structures

of SDGs in each species were indicated in distinct colors. The solid boxes represent exons and black lines represent introns.

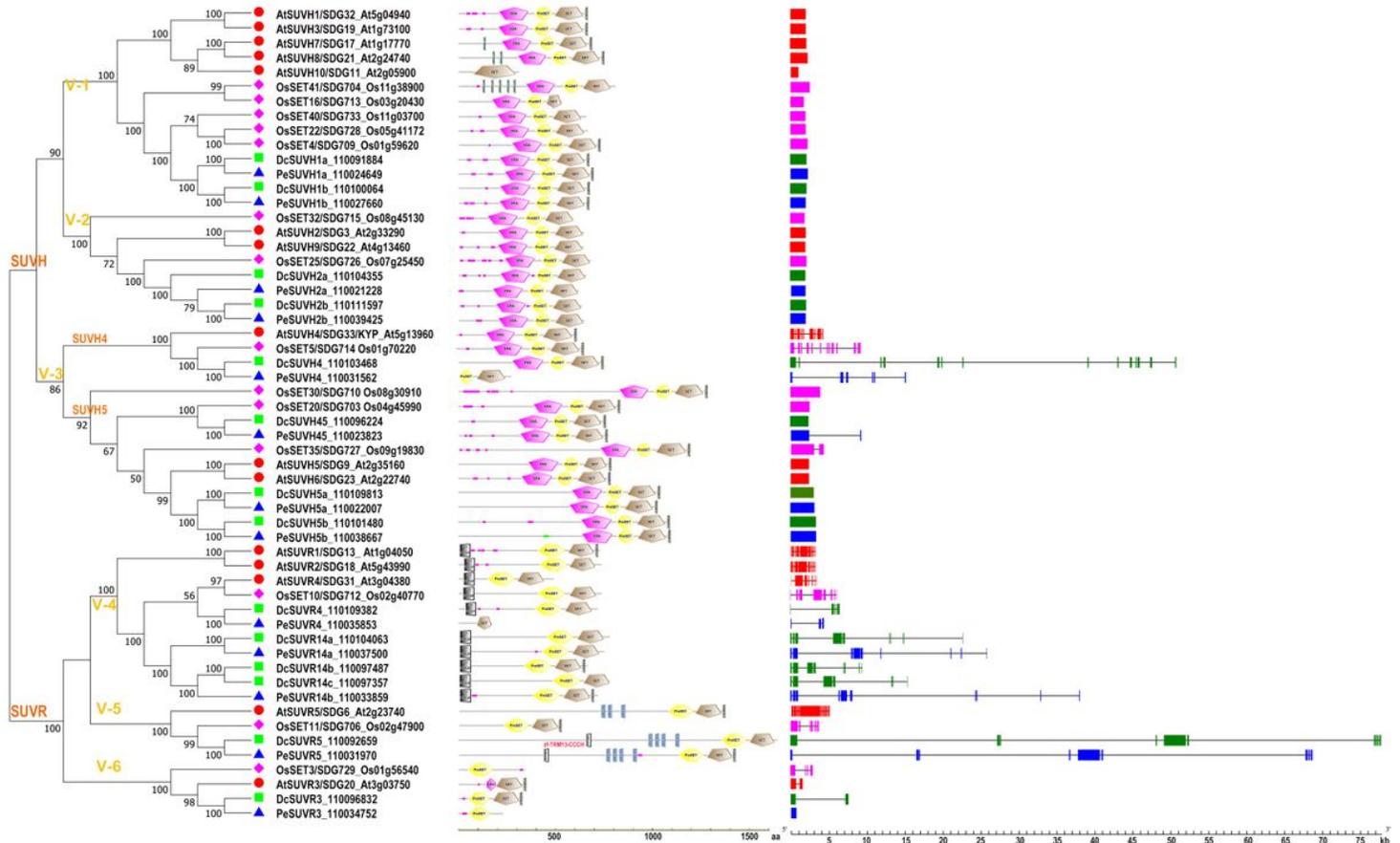


Figure 5

Domain organization and gene structure of the class V DcSDGs. The NJ tree was generated using MEGA7 with parameter settings as Fig. 1 based on full-length amino acid sequences of Class-V SDGs in *D. catenatum*, *P. equestris*, *Arabidopsis* and rice. The number along the tree branch indicates bootstrap value. Different conserved protein domains are colored as indicated. Gene structures of SDGs in each species were indicated in distinct colors. The solid boxes represent exons and black lines represent introns.

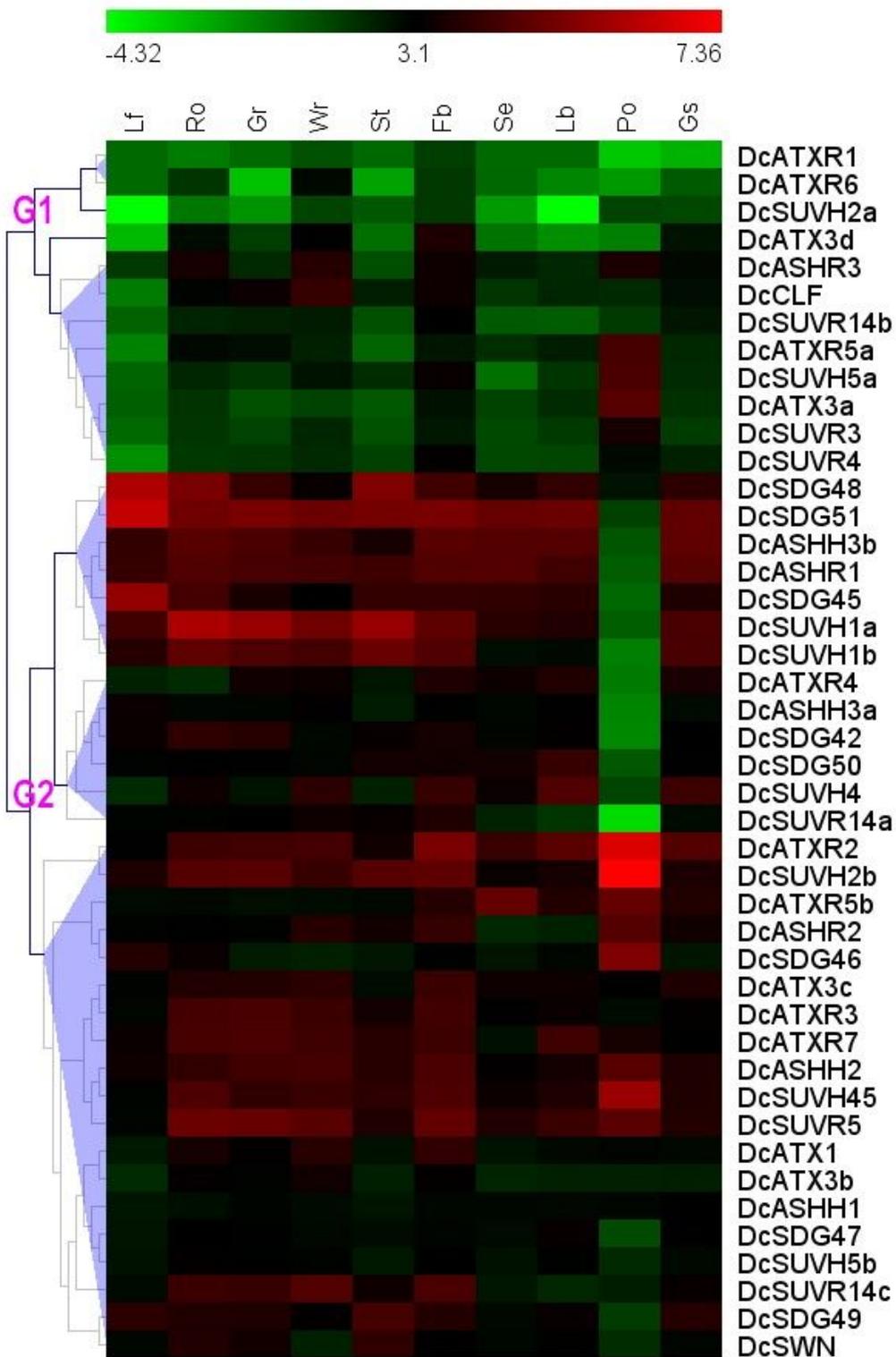


Figure 6

Tissue and organ expression profiles of DcSDGs across different tissues and organs. Hierarchical clustering of expression profiles of DcSDGs across different tissues and organs. Heat map generated using the software MultiExperiment Viewer (MeV) shows relative organ expression profiles of DcSDG genes in *D. catenatum*. Color scale at the top represents log₂ expression values, green and red represent low and high levels of transcript abundances, respectively. Lf: leaf, Ro: root, Gr: green root tip, Wr: white

part of root, St: stem, Fb: flower bud, Se: sepal, Lb: labellum (lip), Po: pollinia, and Gs, gynostemium (column).

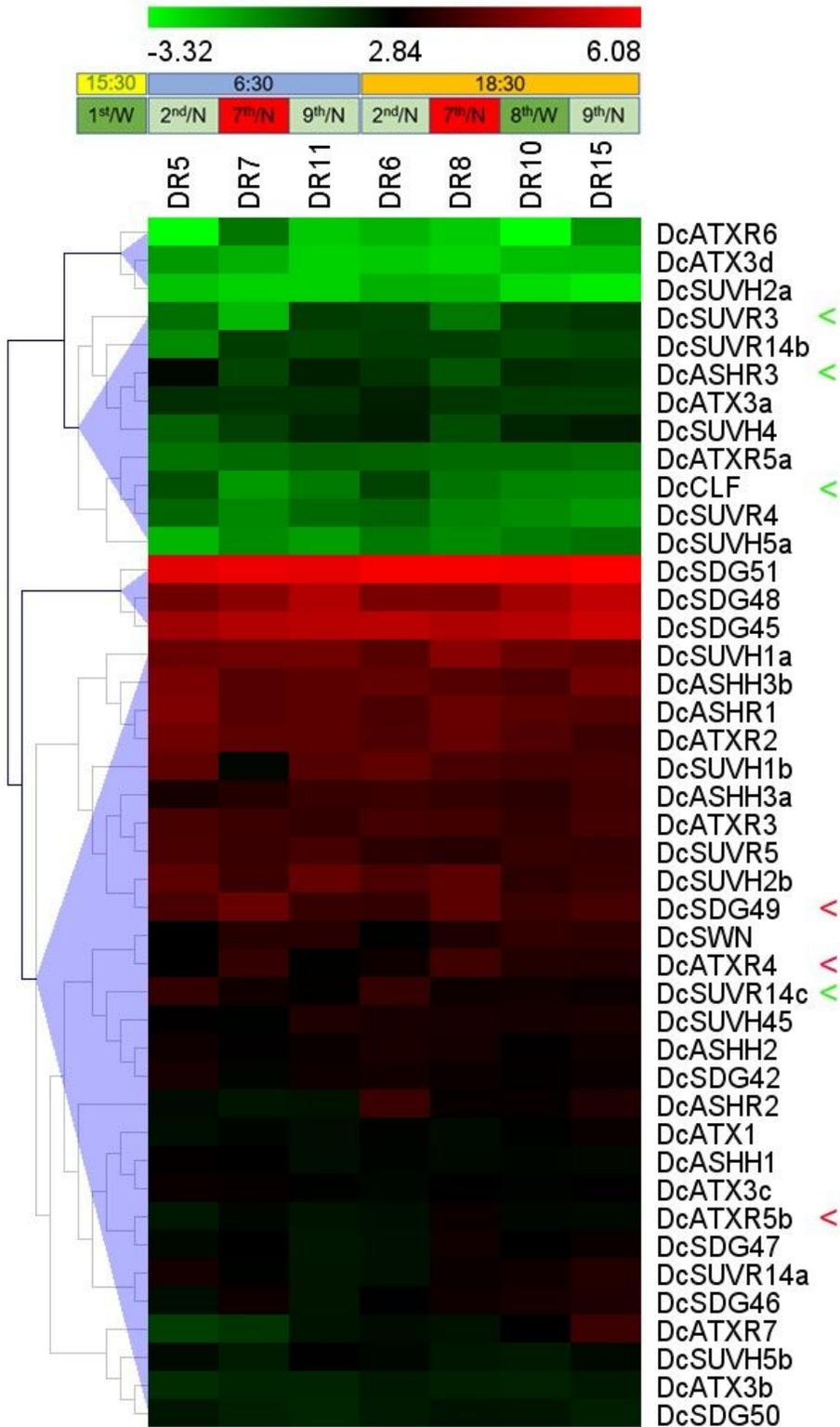


Figure 7

Expression of DcSDGs in response to drought stress. Heat map showing expression pattern of DcSDG genes in leaves under different drought treatments. The seedlings were watered on the 1st day, dried from the 2nd to the 7th day, and re-watered on the 8th day. Leaves were collected at different times; DR5/ DR8,

DR6/DR10, and DR7/DR15 indicate sampling at 06:30 and 18:30 on the 2nd, 7th, and 9th days, respectively, and DR11 indicates sampling at 18:30 on the 8th day. The Y-axis represents the value of the relative expression level [$\log_2(\text{FPKM} + 1)$].

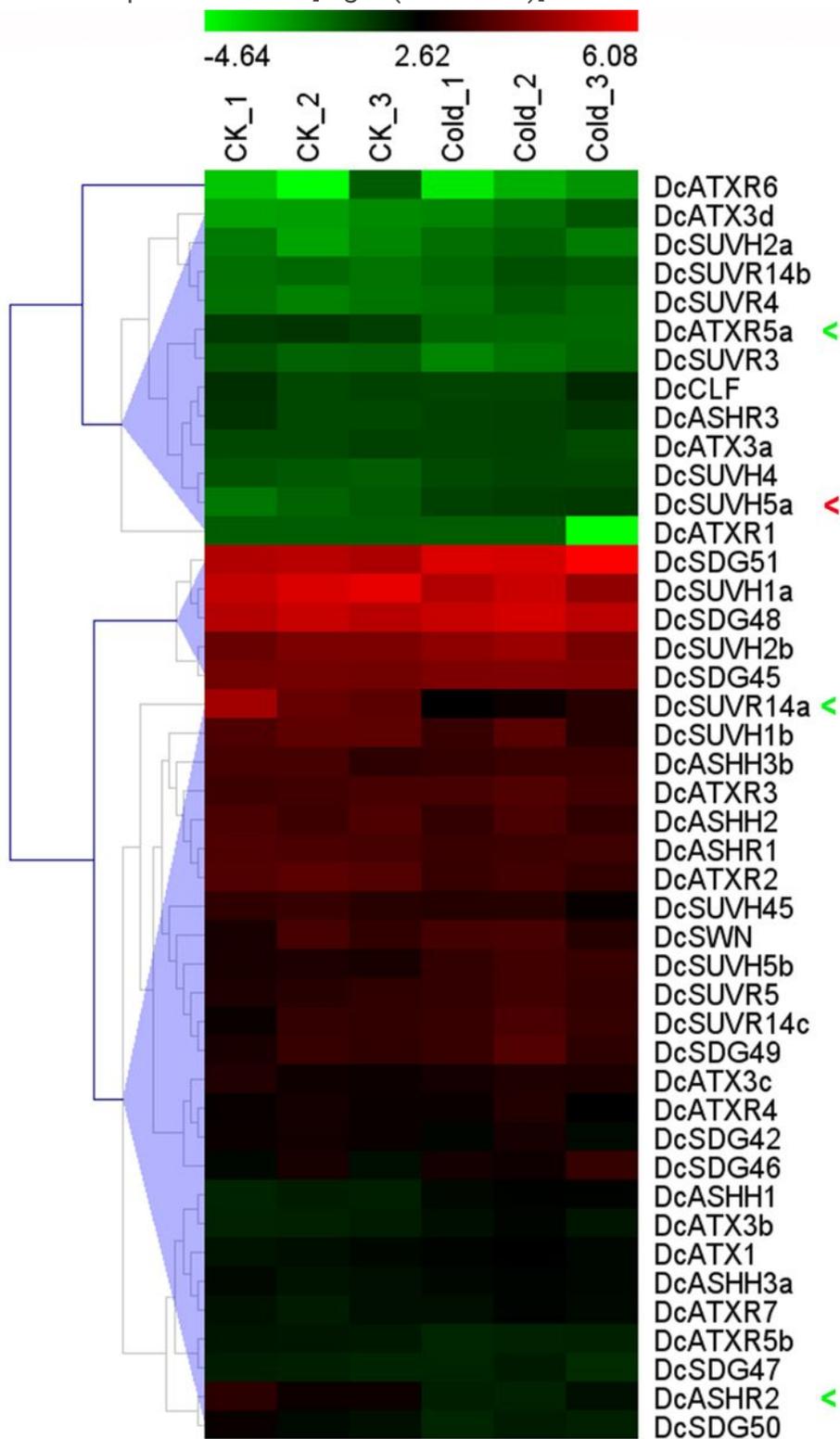


Figure 8

Expression of DcSDGs in response to cold stress. Heat map showing expression pattern of DcSDG genes in leaves under cold stress for 20 h. The Y-axis represents the value of the relative expression level [\log_2

(FPKM + 1)].

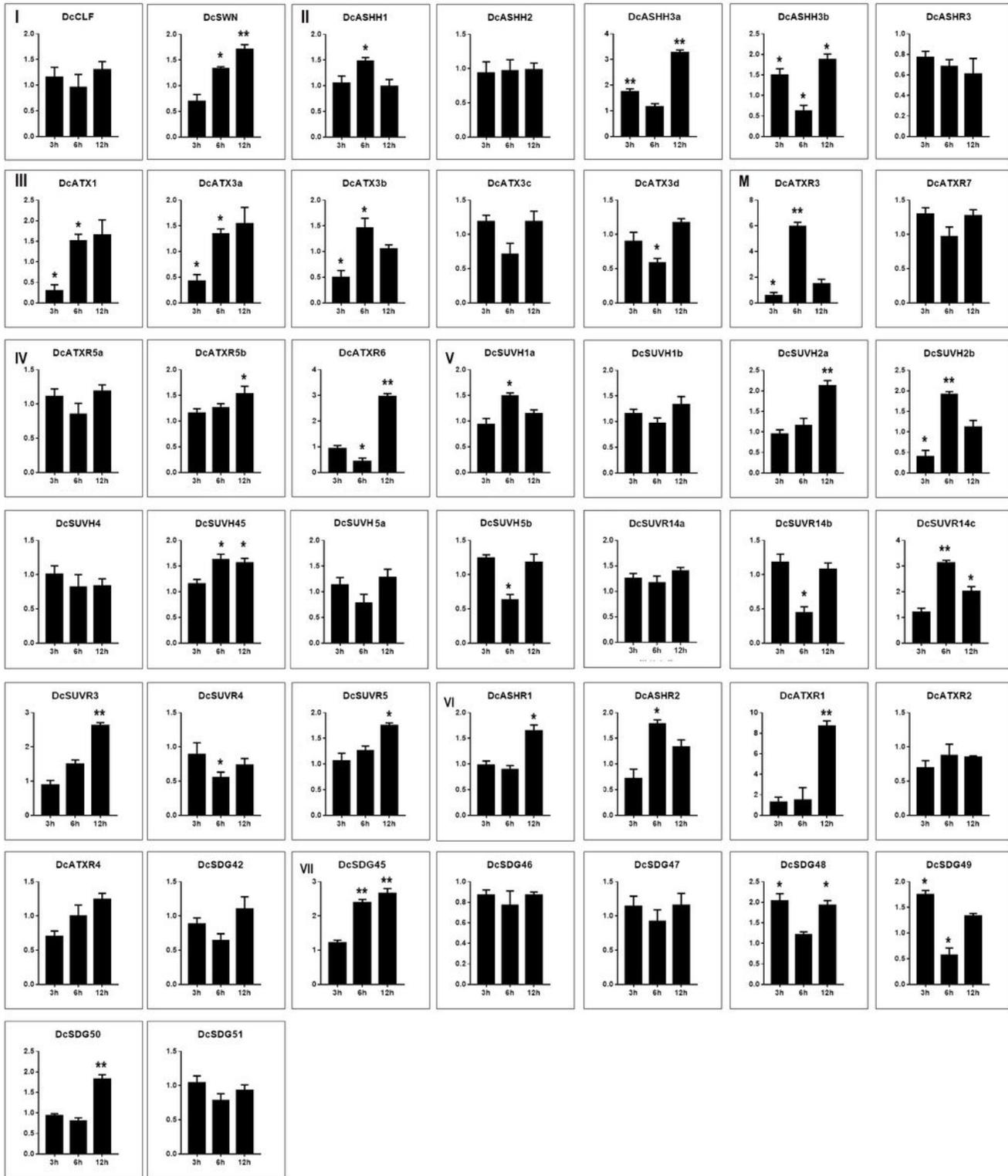


Figure 9

Expression of DcSDGs in response to high temperature stress through RT-qPCR assay. The actin gene of *D. catenatum* was used as an internal reference. The data are representative of three independent experiments. The error bars indicate SD and the asterisk shows the corresponding gene between the heat shock (35°C) and the control (20°C) significantly affected by Student's t test (*, P<0.05; **, P<0.01).

Supplementary Files

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

- [S6DcSDGcoldtreatmentdata.xlsx](#)
- [S1SDGproteinsinAtandOs.docx](#)
- [S2PeSDGs.docx](#)
- [S5Drought.xlsx](#)
- [S4expressionofduplicatedgenepairs.tif](#)
- [S9Primers.xlsx](#)
- [S7ClassVI.tif](#)
- [S8ClassVII.tif](#)
- [S3DcSDGorganexpressiondata.xlsx](#)