

SCIENTIFIC REPORTS



OPEN

Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon

Wei Dong¹, Defeng Wu², Guoshen Li¹, Dewei Wu³ & Zicheng Wang¹

Dwarfism is one of the most valuable traits in watermelon breeding mainly because of its contribution to yield as well as the decreased labor required to cultivate and harvest smaller plants. However, the underlying genetic mechanism is unknown. In this study, a candidate dwarfism gene was identified by applying next-generation sequencing technology to analyze watermelon plants. We completed a whole-genome re-sequencing of two DNA bulks (dwarf pool and vine pool) generated from plants in an F₂ population. A genome-wide analysis of single nucleotide polymorphisms resulted in the detection of a genomic region harboring the candidate dwarfism gene *Cla010726*. The encoded protein was predicted to be a gibberellin 20-oxidase-like protein, which is a well-known “green revolution” protein in other crops. A quantitative real-time PCR investigation revealed that the *Cla010726* expression level was significantly lower in the dwarf plants than in the normal-sized plants. The SNP analysis resulted in two SNPs locating in the *Cla010726* gene promoter of *dsh* F₂ individuals. The results presented herein provide preliminary evidence that *Cla010726* is a possible dwarfism gene.

Watermelon (*Citrullus lanatus*) is among the top five most consumed fresh fruits worldwide, accounting for 7% of the global area devoted to vegetable production¹. Dwarfism is one of the most valuable traits in watermelon breeding because of its positive effect on yield as well as the associated decreased labor required for cultivating and harvesting the crop. In watermelon, two allelic genes, *dw-1* and *dw-1^s*, and two independent loci, *dw-2* and *dw-3*, have been reported to confer dwarfism^{2–5}. Additionally, the cucumber (*Cucumis sativus* L.) genes *cp*, *cp-2*, and *scp* have been identified as responsible for dwarfism-related plant architecture^{6–9}. Meanwhile, in tropical pumpkin (*Cucurbita moschata* Duch.) and squash (*Cucurbita pepo* L.), the dwarfism of the vines is regulated by the *Bu* gene^{10–13}. Although many studies have been conducted on dwarfism traits in cucurbitaceae plants, the responsible genes have not been cloned^{9,14,15}.

Dwarf plant mutations have been important for elucidating the regulatory molecular mechanisms underlying plant growth and development¹⁶. The main causes of dwarfism have been mutations in hormone biosynthesis or signal transduction pathway-related genes affecting the production of gibberellin (GA)^{17–19}, cytokinin²⁰, brassinosteroids^{21,22}, and other key hormones influencing plant growth and development. Additionally, abnormally developed plant cell membranes or walls can also lead to dwarfism in plants^{23,24}.

With the release of sequenced genomes, the combined application of bulked segregant analysis (BSA) and next-generation sequencing technology represents a new way to accelerate the identification of candidate genes controlling important agronomic traits in various crops^{25–27}. In 2013, a high-quality draft genome sequence of the Asian watermelon cultivar ‘97103’ ($2n = 2 \times = 22$) was produced. The draft sequence included 23,440 predicted protein-coding genes^{1,28,29}, and represented an important resource for plant researchers, particularly those interested in the genetic improvement of crops. The objective of this study was to identify the dwarfism gene in the *dsh* mutant watermelon line, which was derived from line ‘I911’ (Code I911; inbred hybrid selected in the seventh generation). Compared with the ‘I911’ plants, the *dsh* plants had short vines and stems, numerous branches and flowers, and small fruits. The ratio of long-vine to short-vine plants for the F₂ and BC₁ populations conformed to Mendel’s segregation ratios of 3:1 and 1:1, respectively³⁰. Thus, we concluded that the dwarfism trait is a qualitative characteristic (QC) controlled by a single gene. We re-sequenced the whole genome of two DNA bulks (i.e.,

¹School of Life Science, Henan University, Plant Genetics Laboratory, Kaifeng, Henan, 475004, People’s Republic of China. ²School of Physical Education, Henan University, Kaifeng, Henan, 475001, People’s Republic of China.

³Jiangsu Provincial Key Laboratory of Crop Genetics and Physiology, Yangzhou University, Yangzhou, Jiangsu, 225009, People’s Republic of China. Correspondence and requests for materials should be addressed to Z.W. (email: wzc@henu.edu.cn)

dwarf pool and vine pool) developed from plants in an F_2 population. A genome-wide analysis of single nucleotide polymorphisms (SNPs) enabled the detection of a genomic region harboring the dwarfism gene. Results from this study provide preliminary evidence that *Cla010726* is a possible candidate gene encoding the dwarfism trait.

Methods

Plant materials and phenotyping for dwarfism. Two watermelon inbred lines, *dsh* and 'I911', were used as the parents to generate the F_1 and F_2 populations. The *dsh* watermelon plant (female parent) is a bush with a short vine, short internodes, thin stems, numerous branches, and small leaves, flowers, and fruits. The 139 F_2 individuals and 20 parent plants were grown and evaluated at the Henan University Genetics and Breeding Base in the spring of 2016.

Data generation. Genomic DNA was extracted from fresh leaves collected from the 2 parent plants as well as the 30 dwarf and 30 vine plants using the CTAB method³¹ for a subsequent QC-sequencing (QC-seq) analysis. The degradation and contamination of the extracted DNA were monitored on 1% agarose gels, while the DNA purity was checked using the NanoPhotometer[®] spectrophotometer (IMPLEN, CA, USA). The DNA concentration was measured using the Qubit[®] DNA Assay Kit and the Qubit[®] 2.0 Fluorometer (Life Technologies, CA, USA). For the QC-seq analysis, two DNA pools, namely the dwarf pool (D-pool) and vine pool (V-pool), were constructed by mixing an equal amount of DNA from the 30 dwarf and 30 vine F_2 plants collected in the autumn of 2016. A total of 1.5 μ g DNA per sample was used as the input material.

Sequencing libraries were generated using the TruSeq Nano DNA HT Sample Preparation Kit (Illumina, USA) following the manufacturer's recommendations. Separate index codes were added to attribute sequences to different samples. Briefly, DNA samples were sonicated to generate 350-bp fragments, which were then end-repaired, A-tailed, and ligated with the full-length adapter for Illumina sequencing by PCR amplification. Finally, the PCR products were purified using the AMPure XP system and the size distribution of the libraries was analyzed with the Agilent 2100 Bioanalyzer. The libraries were quantified by quantitative real-time (qRT)-PCR and then sequenced using the Illumina HiSeq 4000 platform to generate 150-bp paired-end reads, with an insert size around 350 bp.

Data analysis. To ensure reads were reliable and to prevent any artificial bias in the following analyses, the raw reads underwent a series of quality control procedures using in-house C scripts. Raw reads were removed based on the following criteria: (1) reads with $\geq 10\%$ unidentified nucleotides; (2) reads with $> 50\%$ bases having a phred quality score < 5 ; (3) reads with > 10 nucleotides aligned to the adapter, allowing $\leq 10\%$ mismatches; (4) putative duplicates generated by the PCR amplification during the library construction process. The BWA program was used to align the D-pool and V-pool clean reads against the reference genome sequence³². Alignment files were converted to BAM files using the SAMtools program³³. Additionally, potential PCR duplications were removed using the SAMtools command "rmdup". If multiple read pairs had identical external coordinates, only the pair with the highest mapping quality was retained.

Analyses of quality traits with SNP and InDel markers. All samples underwent variant calling using the Unified Genotyper function of the GATK program³⁴. The SNPs and InDels were filtered using the Variant Filtration parameter of GATK. ANNOVAR, which is an efficient software tool, was used to annotate the SNPs or InDels based on the GFF3 files for the reference genome³⁵. The homozygous SNPs/InDels between two parents were extracted from the vcf files. The read depth information for the homozygous SNPs/InDels in the D-pool and V-pool was obtained to calculate the SNP/InDel index²⁵. We used the dwarf genotype of the parent as the reference and for analyzing the read number for the D-pool and V-pool. We then calculated the ratio of the number of different reads to the total number of reads, which represented the SNP/InDel index of the base sites. We filtered out those points in which the SNP/InDel index in both pools was < 0.3 . Sliding window methods were used to determine the SNP/InDel index of the whole genome. The average of all SNP/InDel indices in each window was used as the SNP/InDel index for that window. We usually used a window size of 1 Mb and a step size of 10 kb as the default settings. The difference between the SNP/InDel index of two pools was calculated as Δ SNP/InDel index.

Expression analysis of candidate dwarfism genes by quantitative real-time PCR. We investigated the expression patterns of *Cla010721*, *Cla010725*, *Cla010726*, and *Cla010750* using qRT-PCR. The *dsh* and 'I911' tissue culture seedlings were grown for 10 and 30 days. Each collected sample represented one replicate. Total RNA was extracted from all samples using the Trizol reagent (Invitrogen, Carlsbad, CA, USA). The PrimeScript[™] RT reagent Kit with gDNA Eraser (TaKaRa, Dalian, China) was used to reverse transcribe cDNA from the extracted total RNA. The resulting cDNA samples were analyzed by qRT-PCR in a 20- μ l reaction volume containing 10 μ l SYBR Premix Ex Taq II (TaKaRa). The *CIYLS8* gene (encoding yellow-leaf-specific protein 8) was included as a control for normalizing gene expression data (Kong *et al.*, 2014). The qRT-PCR was completed using an ABI 7500 Fast Real-time PCR system. There were five biological repeats for *dsh* and 'I911'. Each sample was analyzed three times (i.e., technical replicates). The primers used to detect the transcripts of structural and regulatory genes are listed in Table 1.

Analysis of the SNP in the promoter of the candidate dwarfism gene. Genomic DNA was extracted from fresh leaves collected from the 30 dwarf and 30 vine plants using the CTAB method³¹. The degradation and contamination of the extracted DNA were monitored on 1% agarose gels, while the DNA purity was checked using the NanoPhotometer[®] spectrophotometer (IMPLEN, CA, USA). The fragments with SNP were amplified by PCR using the following primer pair: forward, 5'-TGTTGAAATTTGGTGACGAGGT-3', and

Primer name	Forward primer sequence	Reverse primer sequence
Cla010721	GAGCAACTGGGGATGGCGACAT	GGCAAGCACCGGCATGAGTA
Cla010725	GGCCGCCAACGTCTACATGCTT	CGCCAATCCAACGCAGAGT
Cla010726	CGACTTAGGGTTTACGGAAC	GCTCTCAAAATATCTCCCA
Cla010750	CATACTCATCCTTTATCACC	TATATGTTGCAGATCGCTTT
CIYLS8	AGAACGGCTTGTGGTCATTC	GAGGCCAACACTTCATCCAT

Table 1. Sequences of the quantitative real-time PCR primers

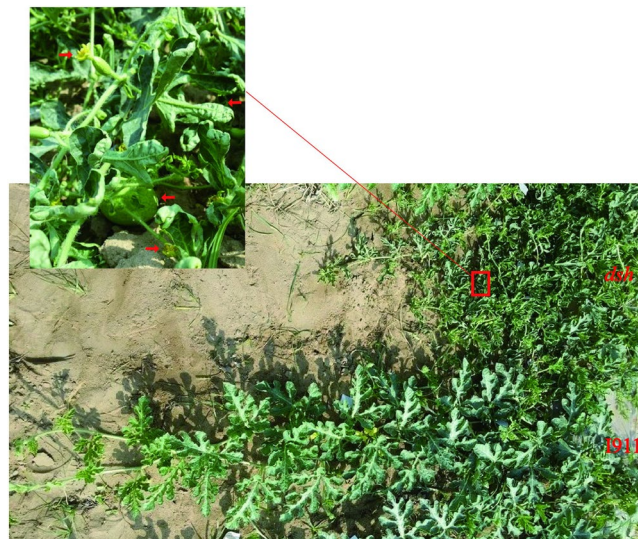


Figure 1. Comparison of the morphological indices between the 'dsh' mutant and the wild-type 'I911'.

reverse, 5'- TGAATTAACGTTTCGGGCAC -3' in a 20- μ l reaction volume containing 0.2 μ l Ex Taq (TaKaRa). And then the PCR products were sequenced.

Results

Morphology of dwarf watermelon plant. After years of screening and cultivating, the *dsh* dwarf watermelon plant was detected in an inbred watermelon line derived from 'I911' in July 2009. There were considerable morphological differences between *dsh* and 'I911' plants (Fig. 1). The *dsh* plants produced short vines and stems, many branches and flowers, and small fruit, which was unlike the 'I911' plants. Moreover, the leaf edges of *dsh* plants were curled. The results showed that F1 generation all developed long vines, whereas the ratio of long-vine to short-vine plants for the F₂ populations conformed to Mendel's segregation ratios of 3:1 (maximum χ^2 value as 0.54, $P > 0.05$).

QC-seq identified four candidate genes controlling dwarfism on chromosome 7. Illumina high-throughput sequencing resulted in 18,371,510,400 bp and 15,618,092,700 bp reads for the D-pool (45.94 \times average depth coverage or 99.68% coverage) and V-pool (39.62 \times average depth coverage or 99.65% coverage), respectively. These reads were aligned to the 355,247,419 bp reference watermelon genome and 352,235 SNPs were identified which included homozygous SNP and heterozygous SNP¹. We used the dwarf genotype of the parent as the reference to calculate the homozygous SNP index of 97,539 polymorphic markers between the two offspring. After screening, 97,186 polymorphic markers were obtained after filtration (Supplementary Table S1). An average SNP-index was computed in a 1 Mb window using a 10 kb step. The SNP-index graphs were generated for the D-pool (Fig. 2a) and V-pool (Fig. 2b) by plotting the average SNP-index against the position of each step in the genome assembly. By combining the information for the SNP-index in the D-pool and V-pool, the Δ (SNP-index) was calculated and plotted against the genome positions (Fig. 2c).

The Δ (SNP-index) value should be significantly different from 0 if a genomic region harbors a target gene. At the 95% significance level, 41 polymorphic marker loci were selected (Supplementary Table S2). At the 99% significance level, only one genomic region on chromosome 7 (27.66–30.61 Mb) had a Δ (SNP-index) value that was significantly different from 0 (Supplementary Table S3). The results of ANNOVAR's annotation indicated that there were four candidate watermelon genes responsible for the dwarfism phenotype in the 27.66–30.61 Mb region on chromosome 7. The four candidate watermelon genes were *Cla010721*, *Cla010725*, *Cla010726* and *Cla010750*. The SNP of *Cla010721* located in *exon* which was nonsynonymous mutation. And the others located in the promoter (Supplementary Table S4). But they all did not change the sequence of amino acids. It was

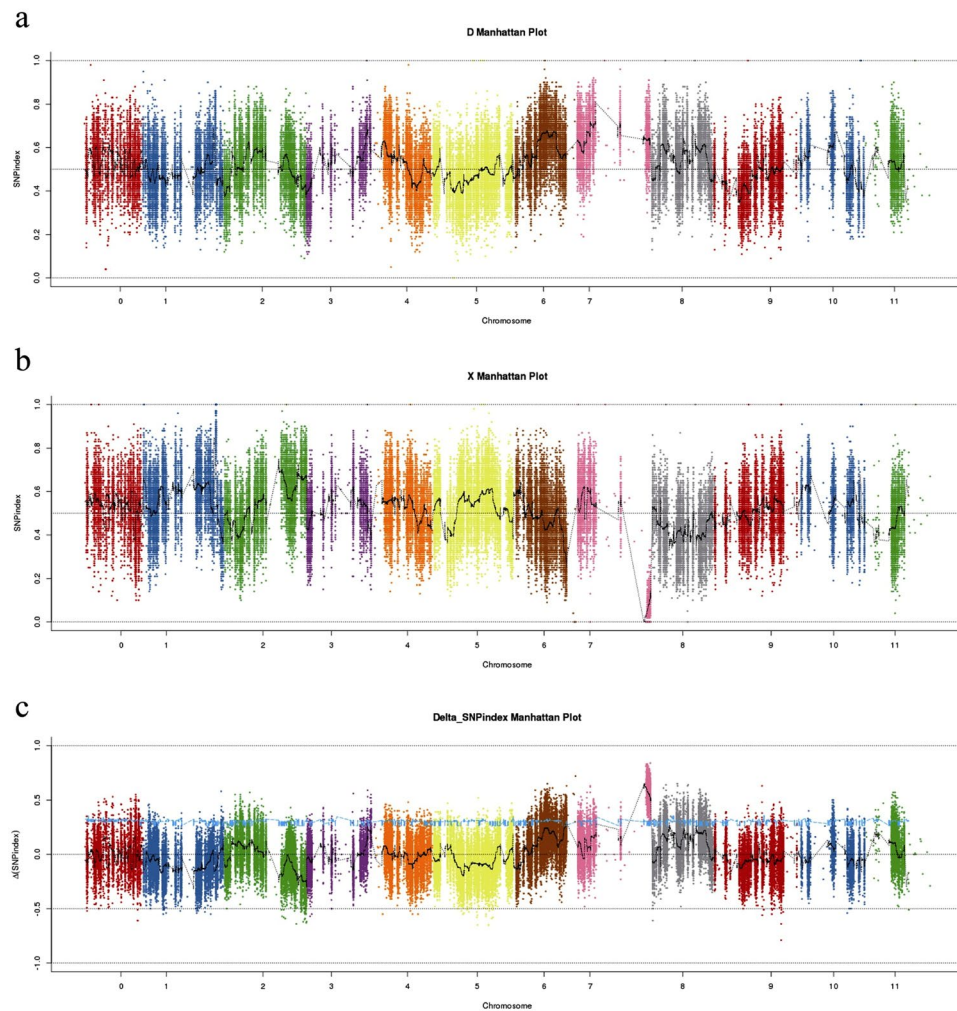


Figure 2. SNP-index graphs of the D-pool (a), V-pool (b), and Δ (SNPindex) (c) for the QC-seq analysis. The x-axis represents the position of seven chromosomes and the y-axis represents the SNP-index, which was calculated based on a 1 Mb window with a 10 kb step.

predicted that Cla010721 was an asparaginase-like protein, Cla010725 was a sugar transporter, Cla010726 was a GA20-oxidase-like protein and Cla010750 was a FAR1-related protein.

Identification of the dwarfism gene. We predicted the presence of four candidate dwarfism genes in a 27,800-kb region of watermelon chromosome 7. The highest Δ (SNP-index) value existed in the 1328 bp upstream region of the *Cla010726* gene (Fig. 3a, Supplementary Table S3). The *Cla010726* gene also appeared promising based on the gene annotation result. Cla010726 was predicted to be a GA20-oxidase-like protein that contains InterPro domain IPR005123 (oxoglutarate and iron-dependent oxygenase). The results of a BLAST alignment revealed that the identity between the *Cla010726* and the *C. sativus* GA20-oxidase 2-like gene was as high as 86%. Additionally, the sequence identity of the encoded proteins was 82.55%. Moreover, the *Arabidopsis thaliana* GA20ox family includes GA20-oxidase 1 and GA20-oxidase 2. The results of a protein BLAST alignment indicated that the sequence identity between Cla010726 and AtGA20ox1 was 30.50% and between Cla010726 and AtGA20ox2 was 29.60%. An important function of GA20ox in many plant species involves regulating GA concentrations. Thus, we proposed that *Cla010726* is a GA20ox homolog in watermelon and named this gene *ClaGA20ox* (*C. lanatus* gibberellin 20-oxidase). This gene represents the most likely candidate gene responsible for the dwarfism of watermelon plants.

We examined the four candidate watermelon genes *Cla010721*, *Cla010725*, *Cla010726* and *Cla010750* expression patterns in the two parental lines by qRT-PCR to assess whether the genes expression level influences the development of the dwarfism phenotype (Fig. 3a). The *Cla010726* expression level was considerably higher in 'I911' plants than in *dsh* plants ($P < 0.05$), further suggesting that *ClaGA20ox* may be responsible for the dwarfism in watermelon plants. We examined the SNP of *Cla010726* gene promoter in the F_2 population by sequencing. The SNP analysis resulted in two SNP locating in the *Cla010726* gene promoter of *dsh* F_2 individuals (Fig. 3b).

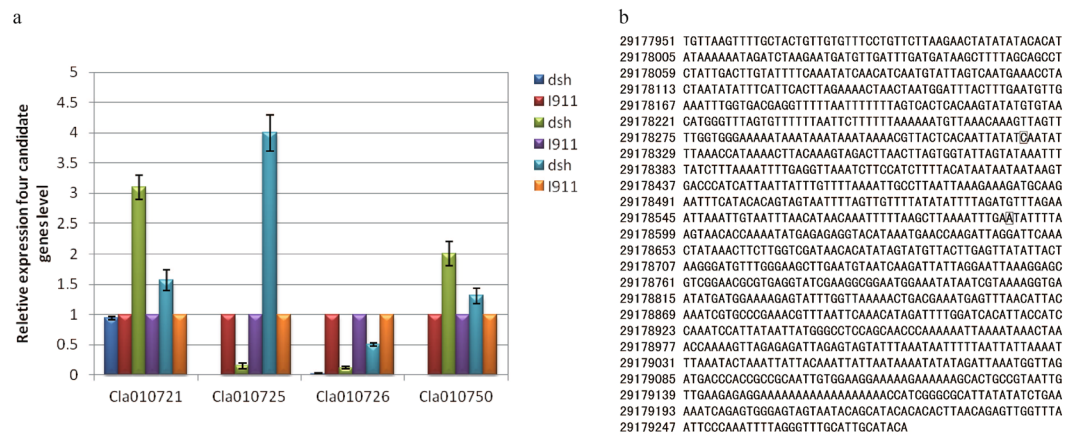


Figure 3. Expression of watermelon four candidate genes and nucleotide sequence of *Cla010726* gene promoter. **(a)** Relative expression levels of four candidate genes in *dsh* and ‘I911’ plants. Data are presented as the mean of three independent measurements. Error bars represent the standard deviation of the mean values. **(b)** Nucleotide sequence of the 1329 bp upstream region of *Cla010726*. The SNP site was in bold.

Discussion

Dwarfism is an important trait in watermelon breeding. In this study, *dsh* was a dwarf mutant that had been identified in an inbred watermelon line derived from ‘I911’. Advantages of this dwarf mutant include its high growth efficiency, low soil fertility requirements, and the fact it can be grown at a high planting density. This relatively new germplasm resource should be further developed in the future. However, the genetic mechanism underlying the dwarfism phenotype remains unknown. We completed a QC-seq analysis to characterize the dwarfism in watermelon plants using an F₂ mapping population. This method combined a high-throughput whole-genome re-sequencing with a bulked-segregant analysis, which represents a quicker and more efficient method of identifying a target gene. Functional orthologs of this gene with mutations have been selected as so-called “green revolution genes” in rice and barley.

Our analyses identified *Cla010726* on watermelon chromosome 7 as a potential gene responsible for the observed dwarfism. This gene is a homolog of *GA20ox* genes found in many plant species. Thus, we designated this gene *ClaGA20ox*. A previous study revealed that *GA20ox* is a key oxidase enzyme that contributes to GA biosynthesis by catalyzing the conversion of GA12 and GA53 to GA9 and GA20, respectively, via a three-step oxidation at C-20 of the GA skeleton¹⁹. Five copies of *GA20ox* genes have been detected in *A. thaliana*³⁶. Mutations to this gene have different effects on overall plant growth. Specifically, the *ga20ox1* line exhibits a semi-dwarf phenotype, whereas *ga20ox2* plants are only slightly smaller than wild-type plants¹⁹. Additionally, *AtGA20ox1* is an ortholog of the rice *SD1* (semi-dwarf 1) gene and barley *sdw1/denso* green revolution genes³⁷. Four *GA20ox*-like genes have been identified in the rice genome¹⁸. *OsGA20ox2* (or *SD1*) is a well-known gene that has been studied in green revolution rice varieties^{38,39}. It is one of the most important genes deployed in modern rice breeding programs^{9,40}. The *sdw1/denso* gene has been one of the most successful semi-dwarfing genes used in barley breeding worldwide^{41–43}. Furthermore, one of the *HvGA20ox2* genes was identified as a candidate gene for *sdw1/denso*, which is an ortholog of the rice *sd1* gene^{44,45}. The first *GA20ox* gene was isolated from pumpkin (*Cucurbita maxima* L.)⁴⁶. The *GA20ox* gene associated with a dwarf vine was also anchored in pumpkin (*C. maxima* D.)¹⁵ according to a high-density genetic map. Among the known *GA20*-oxidases, only that from developing pumpkin seeds has been shown to produce biologically inactive GA as the major product⁴⁷. Transgenic lettuce carrying the pumpkin *GA20ox* exhibited a dwarfism phenotype in the T₂ generation plants⁴⁸. Therefore, it is reasonable to postulate that *ClaGA20ox* is a viable candidate gene responsible for dwarfism in watermelon plants. However, further evidence is needed to functionally validate this. Accordingly, we are currently generating *ClaGA20ox* overexpressing and knock-out mutant lines for a subsequent examination of gene expression and function. Characterizing the mechanism underlying the dwarfism of *dsh* plants will likely be relevant for future molecular breeding efforts.

References

- Guo, S. *et al.* The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet.* **45**, 51–58 (2013).
- Mohr, H. C. A mutant leaf form in watermelon. *Proc Assn Southern Agr Workers.* **50**, 129–130 (1953).
- Mohr, H. C. & Sandhu, M. S. Inheritance and morphological traits of a double recessive dwarf in watermelon, *Citrullus lanatus* (Thunb.) Mansf. *J Amer Soc Hort Sci.* **100**, 135–137 (1975).
- Liu, P. B. W. & Loy, J. B. Inheritance and morphology of two dwarf mutants in watermelon. *J Amer Soc Hort Sci.* **97**, 745–748 (1972).
- Huang, H. *et al.* Inheritance of male-sterility and dwarfism in watermelon [*Citrullus lanatus* (Thunb.) Matsum. and Nakai]. *Sci Horti.* **74**, 175–181 (1998).
- Robinson, R. W. & Mishanec, W. A new dwarf cucumber. *Veg Imp Newsl.* **7**, 23 (1965).
- Kauffman, C. S. & Lower, R. L. Inheritance of an extreme dwarf plant type in the cucumber. *J Am Soc Hortic Sci.* **101**, 150–151 (1976).
- Niemirowicz-Szczytt, K., Rucinska, M. & Korzeniewsia, A. An induced mutation in cucumber: super compact. *Cucurbit Genet Coop Rep.* **19**, 1–3 (1996).

9. Li, J. *et al.* Mutation of rice BC12/GDD1, which encodes a kinesin-like protein that binds to a GA biosynthesis gene promoter, leads to dwarfism with impaired cell elongation. *Plant Cell*. **23**, 628–640 (2011).
10. Shiffriss, O. Developmental reversal of dominance in Cucurbita pepo. *Proc Am Soc Hort Sci*. **50**, 330–346 (1947).
11. Denna, D. W. & Munger, H. M. Morphology of the bush and vine habits and the allelism of the bush genes in Cucurbita maxima and C. pepo squash. *Proc Am Soc Hort Sci*. **82**, 370–377 (1963).
12. Robinson, R. W. *et al.* Genes of the Cucurbitaceae. *HortScience*. **11**, 554–568 (1976).
13. Paris, H. S. & Brown, R. N. The genes of pumpkin and squash. *HortScience*. **40**, 1620–1630 (2005).
14. Hwang, J. *et al.* Fine genetic mapping of a locus controlling short internode length in melon (*Cucumis melo* L.). *Mol Breeding*. **34**, 949–961 (2014).
15. Zhang, G. *et al.* A high-density genetic map for anchoring genome sequences and identifying QTLs associated with dwarf vine in pumpkin (*Cucurbita maxima* Duch.). *BMC Genomics*. **16**, 1101 (2015).
16. Tanabe, S. *et al.* A novel cytochrome P450 is implicated in brassinosteroid biosynthesis via the characterization of a rice dwarf mutant, dwarf11, with reduced seed length. *Plant Cell*. **17**, 776–790 (2005).
17. Boss, P. K. & Thomas, M. R. Association of dwarfism and floral induction with a grape ‘green revolution’ mutation. *Nature*. **416**, 847–850 (2002).
18. Sakamoto, T. *et al.* An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol*. **134**, 1642–1653 (2004).
19. Rieu, I. *et al.* The gibberellin biosynthetic genes AtGA20ox1 and AtGA20ox2 act, partially redundantly, to promote growth and development throughout the Arabidopsis life cycle. *Plant J*. **53**, 488–504 (2008).
20. Holst, K., Schmulling, T. & Werner, T. Enhanced cytokinin degradation in leaf primordia of transgenic Arabidopsis plants reduces leaf size and shoot organ primordia formation. *J Plant Physiol*. **168**, 1328–1334 (2011).
21. Mori, M. *et al.* Isolation and characterization of a rice dwarf mutant with a defect in brassinosteroid biosynthesis. *Plant Physiol*. **130**, 1152–1161 (2002).
22. Schwessinger, B. *et al.* Phosphorylation-dependent differential regulation of plant growth cell death, and innate immunity by the regulatory receptor-like kinase BAK1. *PLoS Genet*. **7**, e1002046 (2011).
23. Duan, Y. *et al.* Dwarf and deformed flower 1, encoding an F-box protein, is critical for vegetative and floral development in rice (*Oryza sativa* L.). *Plant J*. **72**, 829–842 (2012).
24. Rosa, M. *et al.* The Maize MID-COMPLEMENTING ACTIVITY Homolog CELL NUMBER REGULATORY13/NARROW ODD DWARF Coordinates Organ Growth and Tissue Patterning. *Plant Cell*. **29**, 474–490 (2017).
25. Takagi, H. *et al.* QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J*. **74**, 174–183 (2013).
26. Lu, Y. *et al.* Genetic mapping of a putative Agropyron cristatum-derived powdery mildew resistance gene by a combination of bulked segregant analysis and single nucleotide polymorphism array. *Mol Breeding*. **35**, 96 (2015).
27. Tiwari, S. *et al.* Mapping QTLs for salt tolerance in rice (*Oryza sativa* L.) by bulked segregant analysis of recombinant inbred lines using 50K SNP chip. *PLoS One*. **11**, e0153610 (2016).
28. Ren, Y. *et al.* An integrated genetic map based on four mapping populations and quantitative trait loci associated with economically important traits in watermelon (*Citrullus lanatus*). *BMC Plant Biol*. **14**, 33 (2014).
29. Ren, Y. *et al.* A High Resolution Genetic Map Anchoring Scaffolds of the Sequenced Watermelon Genome. *PLoS One*. **7**, e29453 (2012).
30. Li, Y. *et al.* Genetic Analysis of a Dwarf Vine and Small Fruit Watermelon Mutant. *Horticultural Plant J*. **2**, 224–228 (2016).
31. Murray, M. G. & Thompson, W. F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*. **8**, 4321 (1980).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–1760 (2009).
33. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
34. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. **20**, 1297–1303 (2010).
35. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*. **38**, e164 (2010).
36. Hedden, P. *et al.* Gibberellin biosynthesis in plants and fungi: a case of convergent evolution? *J Plant Growth Regul*. **20**, 319–331 (2002).
37. Barboza, L. *et al.* Arabidopsis semidwarfs evolved from independent mutations in GA20ox1, ortholog to green revolution dwarf alleles in rice and barley. *Proc Natl Acad Sci USA*. **110**, 15818–15823 (2013).
38. Sasaki, A. *et al.* Green revolution: A mutant gibberellin-synthesis gene in rice. *Nature*. **416**, 701–702 (2002).
39. Spielmeier, W., Ellis, M. H. & Chandler, P. M. Semidwarf (sd-1), “green revolution” rice, contains a defective gibberellin 20-oxidase gene. *Proc Natl Acad Sci USA*. **99**, 9043–9048 (2002).
40. Luo, Y. *et al.* Marker-assisted breeding of Indonesia local rice variety Siputeh for semi-dwarf phenotype, good grain quality and disease resistance to bacterial blight. *Rice*. **7**, 33 (2014).
41. Hellewell, K. B., Rasmusson, D. C. & Gallo-Meagher, M. Enhancing yield of semidwarf barley. *Crop Sci*. **40**, 352–358 (2000).
42. Ellis, R. P. *et al.* Phenotype/genotype associations for yield and salt tolerance in a barley mapping population segregating for two dwarfing genes. *J Exp Bot*. **53**, 1163–1176 (2002).
43. Von Korff, M. *et al.* AB-QTL analysis in spring barley: II. Detection of favourable exotic alleles for agronomic traits introgressed from wild barley (*H. vulgare* ssp. spontaneum). *Theor Appl Genet*. **112**, 1221–1231 (2006).
44. Jia, Q. J. *et al.* GA-20 oxidase as a candidate for the semidwarf gene sdw1/denso in barley. *Funct Integr Genomics*. **9**, 255–262 (2009).
45. Jia, Q. *et al.* Expression level of a gibberellin 20-oxidase gene is associated with multiple agronomic and quality traits in barley. *Theor Appl Genet*. **122**, 1451–1460 (2011).
46. Lange, T., Hedden, P. & Graebe, J. E. Expression cloning of a gibberellin 20-oxidase, a multifunctional enzyme involved in gibberellin biosynthesis. *Proc Natl Acad Sci USA*. **91**, 8552–8556 (1994).
47. Lange, T. Molecular biology of gibberellin synthesis. *Planta*. **204**, 409–419 (1998).
48. Niki, T. *et al.* Production of Dwarf Lettuce by Overexpressing a Pumpkin Gibberellin 20-Oxidase Gene. *Plant Physiol*. **126**, 965–972 (2001).

Acknowledgements

This work was supported by funding from the Foundation of Henan Science and Technology (Award Number:172102110231).

Author Contributions

W.D. and D.F.W. contributed equally to this work. W.D., D.F.W., G.L., D.W.W. and Z.W. designed the experiments and wrote the manuscript. W.D. and G.L. grew the F₁ and F₂ populations. W.D. and D.W.W. did the QC-seq analysis. W.D. and D.F.W. conducted almost all of the molecular analyses.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-21293-1>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018