QTL-seq Analysis of Seed Size Trait in Grape Provides New Molecular Insight on Seedlessness

Li Wang

Hebei Agricultural University

Songlin Zhang

Northwest Agriculture and Forestry University

Chen Jiao

Northwest Agriculture and Forestry University

Zhi Li

Northwest Agriculture and Forestry University

Chonghuai Liu

Chinese Academy of Agricultural Sciences & Peking Union Medical College Hospital of Skin Diseases and Institute of Dermatology

Xiping Wang (✉ wangxiping@nwsuaf.edu.cn)

Northwest Agriculture and Forestry University <https://orcid.org/0000-0002-8377-855X>

Research article

Keywords: grape, genome, seedless, SNP

Posted Date: March 24th, 2020

DOI: <https://doi.org/10.21203/rs.2.17413/v2>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background Seedlessness in grape (*Vitis vinifera*) is an important commercial trait for both the fresh and drying markets. However, despite numerous studies, the mechanisms and key genes regulating grape seedlessness are mostly unknown. **Results** In this study, we sequenced the genomes of the *V. vinifera* seeded cultivar 'Red Globe', the seedless cultivar 'Centennial', as well as the derived hybrids. Nonsynonymous SNPs were identified and analyzed with respect to published transcriptome data. All the DEGs containing nonsynonymous SNPs were further analyzed in terms of expression patterns, Gene Ontology and pathway enrichment. A potential QTL region associated with seed size was characterized based on SNP indices for both seedless and seeded progeny. Expression analysis of candidate genes during ovule development in multiple seeded and seedless grape cultivars further indicates their potential function in grape seed development. **Conclusion** In summary, DEGs containing nonsynonymous SNPs were mainly protein kinase, transcription factors, cytochrome P450 and other factors related to seed development, which were mainly involved in biological processes like hormone balance, seed coat and endosperm development, reproductive organ development, oxidation and reduction, senescence and cell death. Based on SNP-index and expression pattern analysis, three genes were further identified as potential seedlessness-related genes. Overall the data cast light on the differences of seed development between seeded and seedless progeny in perspective of both functional variants and expression pattern, which provides valuable candidates for future functional study.

Background

Seedlessness in grape is an important focus of research, since seedless fruit are preferred for fresh consumption and processing. Based on distinctions related to pollination and fruit development, seedlessness in grape can be classified as parthenocarpy, where fertilization is absent, or stenospermocarpy, characterized by normal fertilization with subsequent abortion [1,2]. Most commercial seedless grape cultivars are stenospermocarpic, and since this trait is heritable which can pass hereditary feature to the next generation and usually used as parent in breeding, such as the typical seedless cultivar 'Thompson Seedless'.

Over the past several decades, significant progress has been made in characterizing the genetic mechanisms underlying grape seedlessness. On one hand, research has been focused on mapping and identification of genetic loci influencing this trait, aiming to develop molecular markers to increase the speed and efficiency of seedless grape breeding. Seedlessness has been reported to be a quantitative trait mapping to at least seven loci [3-6], comprising one major-effect locus and several complementary recessive loci [4]. One Random Amplified Polymorphic DNA (RAPD) marker, *UBC-269500*, was found to be linked to a major seedlessness locus using bulked segregant analysis (BSA) [7]. Two RAPD markers were obtained and tightly linked to the dominant *I* gene, and the closest marker was further developed into a SCAR (Sequence Characterized Amplified Region) marker, designated *SCC8* [8,9]. In addition, another SCAR marker, *SCF27*, was identified [10]. Grape seedlessness was further studied using QTL (Quantitative Trait Locus) analyses, and a major QTL *VMC7F2* underlying seed fresh weight was developed [3]. On the other hand, an assortment of candidate genes associated with grape seed development has also been identified. A gene named *Chloroplast Chaperonin 21 (ch-Cpn21)* was found to be differentially expressed in the inflorescence of seedless and seeded grapes, and silencing of this gene promoted seed abortion in tobacco and tomato fruits [11].

In addition, an EF-hand calcium-binding protein designated VvCBP1 was reported to be related to seedlessness, and silencing of this gene in tomato reduced seed number in fruit [12]. And *VvAGL11*, an ortholog of the *AGAMOUS*-like *11* gene of *Arabidopsis*, was identified to be the major function candidate gene for seedlessness [13]. Moreover, the grape berry-specific basic helix-loop-helix transcription factor VvCEB1 was reported to affect cell size [14]. In addition, some MADS-box genes, which show strong homology to an *Arabidopsis* ovule identity gene, were found to be differentially expressed during ovule development in multiple seeded and seedless grape cultivars, indicating a potential role in ovule

development [15]. Moreover, nowadays omics analysis has cast light on mechanism related to grape seedlessness, evidence showed that parthenocarpy in the grapevine seedless somatic variant Corintobianco was caused by the absence of a mature macrogametophyte which was probably due to meiosis arrest [16]. A transcriptome analysis of ovule development in grape hybrids was used to develop a network map of possible mechanisms influencing seed size [17]. Recently, the major origin of seedless grapes was found to be associated with a missense mutation in the MADS-box gene *VviAGL 11* with two F1 mapping populations [18]. In general, despite all the progress made in molecular biology underlying grape seedlessness, the grape ovule identity gene and corresponding regulation mechanism was still not clear enough.

Recently, decreasing sequencing costs and increasing quality of output has facilitated the generation of large amounts of sequence data. Whole-genome resequencing combined with QTL analyses (QTL-seq) has enabled rapid identification of the genomic regions underlying specific traits [19]. For example, QTL-seq in rice successfully identified QTLs associated with seedling vigor and resistance to the fungal rice blast disease [20]. Likewise, sequencing of bulked-segregant populations has contributed to a number of scientific studies. For example, a study combining transcriptome sequencing (RNA-seq) with BSA developed single nucleotide polymorphisms (SNPs) in polyploidy species [21], whereas another identified gene associated with disease resistance against enteric septicemia in catfish [22]. In addition, deep genome sequencing of bulked-segregant populations identified a *MIXTA*-like R2R3 *MYB* gene influencing epidermal cell development and carotenoid pigmentation in *Mimulus lewisii* [23].

In this study, we exploited whole genome sequencing to identify genomic regions co-segregating with the seedless trait in bulked populations derived from a cross between a seedless and seeded cultivar. Variants were identified by genome sequencing and analyzed with respect to previously published ovule transcriptome data [17] in order to identify variants and genes potentially affecting seed size. The whole experiment design is shown in Figure 1. The results of our study will enhance the knowledge of the fundamental mechanisms of grape seed development and have potential to accelerate breeding efforts for seedless cultivars.

Results

Data Filtering and High Quality Clean Reads Mapped with Reference Genome

The base composition and quality value before and after filtering is shown in Additional file 1: Figure S1, Additional file 2: Table S1, and Additional file 3: Table S2. The Q20 value of all samples was >98.9% and the Q30 value 96%. The alignment rates to the reference genome for each sample, including the parental cultivars 'Red Globe' (R.G) and 'Centennial seedless' (C), and pooled seeded (S) and seedless (SL) progeny, were >95% (Table 1). The depth of sequencing coverage is 16X in 'C', 19X in 'R', 15X in 'SL', and 17X in 'S'.

Identification and Classification of Single Nucleotide Polymorphisms (SNPs) and Small Indels Based on Genome Re-sequencing

To catalog genetic diversity in grapevine and its potential influence on ovule development, we carried out whole genome re-sequencing of the parental cultivars ('Red Globe' and 'Centennial seedless'), and their pooled seeded and seedless progeny. A total of 6,379,614 SNPs and 745,433 small indels were identified. All of the identified variants were indexed based on their genomic position (Figure 2a). Most of the variants were found in the intergenic regions, followed by intronic, upstream, downstream and exonic regions. The least number of variants were found in the 5' UTR regions. A total of 222,652 variants were located within exons. These were further classified based on their potential effects on gene function (Figure 2b). And more attention was given to those variants predicted to cause amino acid changes in protein coding genes (frameshift, substitution, premature termination, and loss of termination). Four kinds of transition

and eight kinds of transversions were characterized, with C to G transversion and G to A transition the most common. The ratio of transition to transversion was 0.48, which is close to the ideal ratio of 0.5.

Gene Ontology (GO) and Pathway Analysis of Identified DEGs Containing Functional Variants

Functional variations were identified from genome data and the number of corresponding associated genes are shown in Table 2. In addition, a relative comprehensive statistical analysis was conducted in combination of both genome and transcriptome data, which used the same progeny population, and the number of functional variants containing genes identified as DEGs at least at one stage and at all three developmental stages were counted based on function classification of variants (Table 2). With respect to ovule development comparing the seeded and the seedless progeny at Stage 1, Stage 2 and Stage 3, a total of 2,425, 2,988 and 2,328 DEGs containing functional variants were identified, respectively (Additional file 4 Table S3). At each ovule developmental stage, the number of up-regulated DEGs (seedless progeny/seeded progeny) were found to be higher than down-regulated ones. Moreover, the number of DEGs containing functional variations were found to be highest at Stage 2, irrespective of whether the gene was up- or down-regulated.

To explore the biological processes and pathways associated with DEGs containing functional variations, we performed a GO and pathway analysis of the DEGs containing functional variants at three developmental stages. In general, results of both enriched GO and pathway were basically consistent and focused on aspects involved in seed development. A total of 162 GO terms were found to be significantly enriched at all three developmental stages (Figure 3a). The GO terms associated with up-regulated DEGs were mainly involved in biological processes like 'plant defense', 'regulation of reproductive development', 'regulation of hormone balance', 'seed coat development', 'endosperm and seed development', and 'cell death' (Figure 3b). On the other hand, GO terms associated with down-regulated DEGs were mainly related to 'lignin biosynthetic process', 'secondary metabolic process', 'oxidation reduction' and 'iron ion transport' (Figure 3b). A total of 124 pathways were found to be enriched at all three ovule developmental stages (Figure 4a), and these enriched pathways were mainly related to the regulation of hormone balance ('cytokinins degradation', 'cis-zeatin biosynthesis', 'jasmonic acid biosynthesis' and 'ethylene biosynthesis from methionine'), seed coat development ('cellulose biosynthesis' and 'suberin biosynthesis'), secondary metabolic biosynthesis ('flavonoid biosynthesis') and oxidation-reduction ('glutathione-mediated detoxification') (Figure 4b).

Functional and Expression Analysis of DEGs Containing Functional Variants

We focused further attention on functional variant-containing genes that were differentially expressed in the developing ovules, between seeded and seedless cultivars, at all three ovule developmental stages (Table 2). We estimated gene function based on their annotation, and focused on genes encoding potential disease resistance protein, protein kinase, transcription factor, cytochrome P450 and other factors affecting seed development.

Functional variants within genes encoding protein kinases were found to be nonsynonymous SNP, frameshift deletion, frameshift insertion, stopgain and stopless (Figure 5a). The expression patterns of these DEGs were almost similar at all three ovule developmental stages, i.e., mainly up-regulated in seedless progeny compared with seeded ones. We observed that many functional variant-containing DEGs encoded disease resistance proteins; these variants comprised nonsynonymous SNP, frameshift deletion, frameshift insertion and stopless (Figure 5b). Interestingly, the expression pattern of all the disease resistance related DEGs was almost same during three developmental stages, and most of the DEGs were up-regulated in the ovules of seedless progeny compared with seeded progeny. Many DEGs containing functional variants (nonsynonymous SNP, frameshift deletion, frameshift insertion, stopgain and stopless) were identified within cytochrome P450-like genes, and most of these DEGs were up-regulated in seedless progeny compared to seeded ones at all three developmental stages. It is noteworthy that, the expression pattern of these DEGs was almost the same during all developmental stages.

Moreover, many DEGs identified as transcription factors (TF) were found to contain functional variants including nonsynonymous SNP, frameshift deletion, stopgain and stopless (Figure 6). These TFs were classified into 15 families, among which MYB and BHLH were families with the most number of DEGs having functional variants. Expression of DEGs associated with 10 TF families (WRKY, BHLH, SBP, AP2, Myb, HB, TCP, HSF, CCAAT, and GRAS) were found to be up-regulated at all three ovule developmental stages in seedless progeny compared to seeded progeny. In contrast, a group of genes was down-regulated at all three ovule developmental stages in seedless progeny compared to seeded ones; these included one gene from each family of NAC, GATA, bZIP, ERF and two genes from MADS-box family.

DEGs containing functional variants were also classified into 12 other functions associated with seed development, including heavy-metal-associated protein, cellulose-related protein, ankyrin repeat-containing protein, senescence-associated protein, hormone-related protein, calmodulin, methyltransferase, cell cycle, ABC transporter protein, thaumatin, aspartic proteinase and flavonoid synthase. Expression pattern of these DEGs is shown in Figure 7. Two kinds of functional variants were identified in heavy-metal-associated protein (nonsynonymous SNP and frameshift insertion) and methyltransferase (nonsynonymous SNP and frameshift deletion), while only one kind of function variant (nonsynonymous SNP) was identified in others. Expression of DEGs encoding cellulose-related protein, ankyrin repeat-containing protein, senescence-associated protein, calmodulin and cell cycle were up-regulated at all three ovule developmental stages in seedless progeny compared to seeded ones while expression of DEGs encoding several of the other functions were down-regulated. Except one DEG encoding methyltransferase, all other DEGs showed consistency in expression pattern at all three stages of ovule development.

Identification of seedless candidate genes based on SNP-index analysis of two extreme pools

For each progeny bulk, the frequency of each SNP variant (SNP-index) was calculated, and a delta SNP-index was determined (Additional file 5: Figure S2). A number of 3 peaks on chromosomes 11, 13, and 17, as well as 3 valleys on chromosomes 6, 7, and 8 under the threshold for delta-SNP index were identified. A total of 7,923 SNPs were identified within the peak regions, while 360 SNPs were identified in the valley regions. A total of 13 nonsynonymous SNPs were identified in the valley regions, including 1 variant on Chromosome 6 associated with 1 gene, 6 variants on Chromosome 7 associated with 4 genes, and 6 variants on Chromosome 8 associated with 4 genes (Additional file 6: Table S4). In the peak regions, a total of 38 functional variants were identified (Additional file 7: Table S5), among which 35 variants (34 nonsynonymous SNP and 1 stopgain) were located on Chromosome 13 and were associated with 18 genes. The other 3 nonsynonymous SNPs were found on Chromosome 17 and associated with 2 genes.

In general, based on the SNP-index analysis, a total of 33 genes were identified as potential preliminary candidates containing functional variant and some of them have more than one functional variant. We generated transcriptomic data using ovules from progeny of the same hybrid population as the genomic sample and examined the expression pattern of these genes during three different stages of ovule development. We defined DEGs in at least one of the three stages of ovule development with functional variant as the standard and nine candidate genes were further analyzed in this way and corresponding information can be found in Table 3 and Additional file 8: Figure S3. Among nine selected genes, *G2* was down-regulated during all ovule developmental stages in the seedless progeny. Whereas, the expression of the other eight candidate genes was up-regulated at least during one ovule developmental stage in seedless progeny compared to the seeded ones. It is notable that the arginine-197-to-leucine substitution on Chromosome 18 in *VviAGL11* [18] was also identified in our results, however, compared with other genes, this gene did not perform well in our SNP-index analysis, so not listed as a key candidate gene for further analysis. According to our transcriptomic analysis, *VviAGL11* is not identified as DEGs, and our results are consistent with the previous findings, so we have not considered this gene for further research. Motif analysis of the mutation region was carried out, and results showed that mutation in *G1* happened in 'ML domain', *G2* and *G3* in 'Low complexity region', and mutation in *G6* caused stopgain (Additional file 9: Table S6).

Expression analysis of candidate genes related to seedlessness

To further explore the potential roles of nine candidate genes in ovule development or abortion, their expression patterns were observed in different grapevine tissues (Figure 8) and in ovules at different development stages (Figure 9) from seeded and seedless cultivars. All candidate genes were expressed in reproductive tissues (flowers, tendrils and fruits). Both the *G1* and *G2* genes showed strong expression in reproductive tissues, compared with nutritional tissues, whereas the *G8* gene was expressed more strongly in nutritional tissues (Figure 8). The expression pattern of *G1*, *G8*, and *G9* genes were similar between seeded and seedless varieties, while *G5*, *G6* and *G7* genes were differentially expressed. For example, *G5* expression was relatively high in fruit of 'Red Globe' and leaves of 'Thompson Seedless', while *G6* and *G7* were relatively highly expressed in the root of 'Thompson Seedless'.

To further justify and validate our results, we analyzed expression of the nine candidate genes in two seeded ('Kyoho' and 'Red Globe') and two seedless cultivars ('Thompson Seedless' and 'Flame Seedless') (Figure 9). The expression pattern of most of the candidate genes was consistent with the transcriptome data, except that some genes showed differences at specific stages, likely related to cultivar difference (Figure 9). Six of the candidate genes exhibited a consistent expression pattern at all three ovule developmental stages between cultivars: four genes (*G4*, *G7*, *G8* and *G9*) showed higher expression in seedless cultivars compared to seeded ones, while two genes (*G2* and *G5*) showed the opposite results. In transcriptome data (Table 3), expression of *G5* gene was first up-regulated at stage 1 then down-regulated at stage 2 and 3 (seedless progeny/seeded progeny), while *G5* gene was down-regulated (seedless cultivars/seeded cultivars) at all three developmental stages in multiple cultivars. Despite *G5* gene showing some difference, the expression pattern of the remaining five candidate genes was consistent with transcriptome data, which indicates these candidate genes may be related to seed phenotype differences.

Discussion

In this study, we combined whole genome sequencing with bulk segregation analysis of seeded and seedless progeny derived from the seeded maternal parent 'Red Globe' and the seedless paternal parent 'Centennial seedless' to identify genes with potential function in ovule/seed development. These genes were further assessed in relation to published ovule transcriptome data using the same plant progeny materials. SNP-index analysis combined with expression profiles resulted in the identification of potential candidate genes containing function variants. This adds to the knowledge of grape seedlessness and will provide new molecular resources for seedless grape breeding in the future.

Overall Differences of DEGs Containing Functional Variants between Seeded and Seedless Grape Progeny

DEGs containing functional variants were identified by combining data from genome sequencing and RNA-seq using the same grape progeny. Further functional analysis showed that overall differences of these DEGs between seeded and seedless grape progeny were mainly enriched in important biological processes and pathways involved in seed development, including reproductive organ development, hormone balance regulation, seed coat development, endosperm development, senescence and cell death (Figure 3 and 4). In addition, many DEGs containing functional variants were identified as protein kinases, TFs, TRs, cytochrome P450 and genes related to seed development. Protein kinases are key factors in plant biological processes such as energy metabolism and signal transduction. Protein kinases participate in hormonal signaling cascades in plants, and there have been increasing reports of protein kinases regulating plant reproductive tissue development. In *Petunia*, a Pollen-Expressed Receptor-like Kinase gene (*PRK1*) was found to play a role in signal transduction events during pollen development and pollination [24]. In *Arabidopsis thaliana*, a receptor-like protein kinase, HAESA, was reported to control floral organ abscission [25], and GSK3/Shaggy like kinase AtSK3-2 was found to play an important role in the control of cell elongation during flower development [26]. Based on the fact that plant protein kinases participate in the regulation of reproductive tissue

development, we infer that DEGs containing functional variants, identified as protein kinases, may be related to phenotypic differences of ovules between seeded and seedless progeny.

Other factors influencing seed development include cytochrome P450s, ankyrin repeat-containing proteins, calmodulins, ABC transporter proteins and so on. In plants, ankyrin repeat-containing proteins mediate protein-protein interactions to regulate signal transduction related to development and plant defense response. In lily, an ankyrin repeat-containing protein characterized as a ubiquitin ligase was found to be required for pollen germination and pollen tube growth [27]. Moreover, plant calcium ions can influence pollen tube growth, and the *VvCBP1* gene encoding an EF-hand calcium-binding protein was reported to be associated with fruit seedlessness of grapevine [12]. Other factors such as ABC transporter proteins have also been reported to influence plant reproductive tissue development. For example, an ATP-binding cassette transporter from rice, *OsABCG15*, was found to be required for anther development and pollen fertility [28]. As the DEGs containing functional variants represent various functional classes, we surmise that they may have a role in different biological processes and pathways involved in seed development or ovule abortion.

Some DEGs Containing Functional Variants were Identified as Candidates associated with Grape Seed Size by QTL-Seq Analysis

Based on the potential function of candidate genes, we further carried out gene expression analysis during ovule development among multiple cultivars. The results, in combination with both gene annotation and previous related functional studies, suggest that three genes in particular, *G2*, *G4* and *G8*, may be determinants of grape seedlessness.

In *Arabidopsis*, the *AUXIN RESPONSE FACTOR 2 (ARF2)* gene links auxin signalling, cell division, and the size of seeds and other organs [29]. As auxin regulators, ARFs can interact with AUX/IAA proteins to repress the content of auxin and regulate corresponding subsequent biological processes. The grapevine *VvCEB1* gene, which was reported to influence cell size, functions by influencing auxin metabolism and cell signaling [14]. Prior to fertilization, ovary growth is blocked by the repressive action of AUX/IAA–ARF complexes on auxin-responsive genes. After the pollination and fertilization, the level of auxin in ovules or ovary increases and derepresses ovary growth, which is necessary for fruit initiation [30]. ARFs function to guarantee the corresponding biological process proceeds in an orderly manner to ensure seed viability. For example, in *Arabidopsis* and tomato, mutation of *ARF8* and *SlARF7* both lead to seedlessness [30,31]. In our results, the *G2* gene, identified as encoding a putative auxin response factor, was differently expressed in ovules between seeded and seedless progeny, and this expression pattern was also consistent in ovules of multiple cultivars. In view of the reported function of ARFs, this implies that *G2* may be a key gene related to seedlessness.

Furthermore, the *G4* was annotated as a putative cytochrome P450 monooxygenase. This class of gene is known to participate in the balance regulation of GA and auxin. In rice, *ELONGATED UPPERMOST INTERNODE (EUI)* encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins to reduce their concentration [32]. The *Arabidopsis* cytochrome P450 monooxygenase gene *71A13* and *YUCCA* were reported to be involved in auxin synthesis [33,34]. These studies indicate that the influence of monooxygenases on plant growth and development is via regulating GA and auxin content. In the seedless vs. seeded progeny populations used for transcriptional profiling, the content of both GA and auxin was significantly higher in seedless progeny compared to seeded ones during ovule development [17]. Moreover, overexpression of *Arabidopsis CYP78A9*, which encodes cytochrome P450, resulted in large and seedless fruit [35]. These results suggest that the content of GA and auxin is critical for seed formation, and that *G4* is a key gene associated with seedlessness. Whereas, the *G8* gene was annotated as a putative Aluminum-activated Malate Transporter (ALMT). Previously, ALMT has been well studied for its important function in malic acid transport [36]. In apple, a natural mutation leading to truncation of the ALMT protein at the *Ma* locus was reported to be associated with low fruit acidity [37], but to our knowledge, ALMT has not been explored for its role in seed development. According to our results, *G8* is found within the genomic interval identified by QTL-seq analysis influencing seedlessness. Moreover, the expression of

the *G8* gene is significantly different during ovule developmental stages in seedless and seeded cultivars. These results implicate that, *G8* and *ALMT* genes function as key factors influencing seedlessness. Our results suggest *G8* is an excellent candidate for further functional studies.

Conclusions

In summary, genome sequencing analysis in combination with transcriptome data was conducted using grape progeny derived from the seeded maternal parent 'Red Globe' and the seedless paternal parent 'Centennial seedless'. All DEGs containing functional variants were identified, so as to relate biological processes and pathways. Based on SNP index and expression analysis, a total of three genes in the QTL region were identified as candidates related to grape seedlessness, which provides valuable candidates for further functional research and adds our knowledge for seedless grape breeding.

Methods

Plant Material

A total of 114 F1 individuals were obtained from a cross between the seeded maternal parent 'Red Globe' (*V. vinifera*) and the seedless paternal parent 'Centennial seedless' (*V. vinifera*) in 2009, which is the same population used in an ovule transcriptome analysis [17]. All seedling plants were developed and maintained at the Zhengzhou Fruit Institute, Zhengzhou, China (113°42'N 34°43'E), and seed phenotypes were evaluated for three years. A total of 65 progeny with stable seed phenotype were obtained, including 31 seedless and 34 seeded. Of these, 19 seeded and 19 seedless individuals were randomly selected as experimental materials for genome sequencing. All the plant materials used and collected in this work comply with China's guidelines and legislation.

Sample Collection and Genomic DNA Library Construction, Sequencing and Data Analysis

For expression analysis, plant structures (young roots, stems, leaves, tendrils, flowers at the fully opening stage, and fruit at 30 days post-anthesis) were collected from 'Red Globe' and 'Thompson Seedless' seedlings. Ovules were dissected from two seeded ('Kyoho' and 'Red Globe') and two seedless ('Thompson Seedless' and 'Flame Seedless') cultivars at different developmental stages and weighed, as described previously [15]. All plants were identified and maintained in the grape germplasm resource orchard of Northwest Agriculture & Forestry University, Yangling, China (34°20'N 108°24'E). During ovule collection, ovule weight was measured along with development. Three landmark ovule developmental stages were defined based on seedless grape ovule weight change (initial stage, stage with the highest weight, and stage with the lowest weight) as done previously [17]. Samples were immediately frozen in liquid nitrogen and stored at -80 °C for subsequent gene expression analysis. And the plant materials used and collected in this work comply with China's guidelines and legislation.

For genomic sequencing, young leaves from 19 seedless or 19 seeded progeny were collected and pooled. DNA was extracted from each pooled leaf sample, using equivalent mass of sample from each pool. Sequencing libraries for the bulk and the parents were prepared using the Genomic DNA Sample Prep kit (Illumina, San Diego, CA) according to the manufacturer's protocol. The libraries were sequenced using an Illumina HiSeq 2500 system for 151 cycles, operating in the paired-end mode. The resulting raw Illumina reads were processed to remove duplicated read pairs (defined as those having identical bases in the first 100 bp of both left and right reads). Illumina adaptor sequences and low quality sequences were trimmed from the reads using Trimmomatic [38]. Read pairs with read < 40 bp were excluded from further analysis. Resulting paired-end reads were mapped to a grape reference genome (PN40024; 12X) [39] using BWA [40]. Only reads that mapped uniquely (i.e. having one single best hit) to the genome were used for SNP calling. SNPs and small indels (1-5 bp) were identified based on the alignment information in the pileup format generated by SAMtools [41].

DEGs containing nonsynonymous SNPs were identified using published transcriptome data (NCBI SRA accession SRP081137), previously generated from the same progeny populations [17]. GO and pathway enrichment analysis were performed using Plant MetGenMAP [42]. The raw sequencing data has been deposited in NCBI SRA under the accession number SRP174903.

QTL Analysis of Seed Size based on SNP Index

By aligning the sequence data of 'Seeded' and 'Seedless' DNA bulks to the reference sequence of one parent cultivar 'Red Globe', we counted the number (k) of short reads harboring SNPs that are different from the reference sequence. We defined the proportion of k in the total short reads (n) covering a particular genomic position ($=k/n$) as the SNP-index [20]. A SNP index was calculated for each SNP, and sliding window analysis was applied with a 2 Mb window size and 100 kb increment. The average SNP index within the window was used for sliding window plot drawing, and SNPs with a number less than 10 were excluded. To filter spurious SNPs called due to sequencing and/or alignment errors, SNP positions with SNP-index of <0.3 and read depth <7 were also excluded. A Δ (SNP-index), derived from SNP-index for each bulked pool, was further calculated [20]. The QTL region associated with seed size was then identified based on an arbitrary threshold of the top 0.5% of Δ (SNP-index) [43].

Quantitative RT-PCR Analysis

Quantitative real-time RT-PCR was carried out using a Bio-Rad iQ5 thermocycler (Bio-Rad, Hercules, CA, USA). For each sample, 1 μ g of total RNA was converted into cDNA using PrimeScript™ RTase and an oligo dT primer (TaKaRa Biotechnology, Dalian, China) and was subsequently diluted 6-fold with sterile water. The grapevine *Actin1* gene (GenBank Accession number AY680701) which was amplified with the primers F (5'-GATTCTGGTGATGGTGTGAGT-3') and R (5'-GACAATTTCCCGTTTCAGCAGT-3') was used as reference gene. Gene-specific primers can be found in Additional file 10: Table S7. For each reaction, 100 ng cDNA was used and carried out in triplicate with a reaction volume of 20 μ l containing 0.8 μ l each primer (1.0 μ M), 1.0 μ l of cDNA, 10 μ l of SYBR green (TaKaRa Bio Inc.), and 7.4 μ l sterile distilled water. The amplification parameters were 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 30 s. Melting-curve analyses were performed with initial incubation at 95°C for 15 s and then a constant increase from 60°C to 95°C. *ACTIN1* was used as a reference gene. Relative expression levels were analyzed using the IQ5 software and the normalized-expression method [44].

List Of Abbreviations

Absciscic acid, ABA; *AGAMOUS*, AG; Aluminum-activated Malate Transporter, ALMT; *AUXIN RESPONSE FACTOR 2*, ARF2; Bulk Segregant Analysis, BSA; *Chloroplast Chaperonin 21*, *ch-Cpn21*; *ELONGATED UPPERMOST INTERNODE*, EUI; Gene Ontology, GO; Pollen-Expressed Receptor-like Kinase, PRK; Quantitative Trait Locus, QTL; Sequence Characterized Amplified Region, SCAR; Single Nucleotide Polymorphism, SNP; *S Receptor Kinase*, SRK; transcription factor, TF; transcription regulators, TR

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The raw sequencing data has been deposited in NCBI SRA under the accession numberSRP174903. The datasets supporting the conclusions of this article are included within the article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Joint Funds of the National Natural Science Foundation of China (U1603234), the Program for Innovative Research Team of Grape Germplasm Resources and Breeding (2013KCT-25), Chinese Universities Scientific Fund (Z109021571 and 2452019170) and the Strat-up Funds of Agricultural University of Hebei to Li Wang (3002013). These funding bodies had no role in the design of this study, collection, analysis, and interpretation of data or in writing the manuscript.

Authors' contributions

X.W. and L.W. designed the research; X.W. supervised the experiments; L.W. performed most of the experiments; S.Z. and Z.L. provided technical assistance; L.W. and C.J. analyzed the data; C.L. provided the experimental materials; L.W. wrote the article with contributions of all the authors. All authors read and approved the final manuscript.

Acknowledgements

We thank Prof. Steve van Nocker from Michigan State University for editing this manuscript. The authors thank all editors and reviewers for their comments on this manuscript.

References

1. Ledbetter CA, Ramming DW. Seedlessness in Grapes. In: Horticultural Reviews. John Wiley & Sons, Inc.; 1989; 159-184.
2. Ledbetter CA, Burgos L. Inheritance of Stenospermocarpic Seedlessness in *Vitis vinifera* L. J HERED. 1994; 85(2):157-160.
3. Cabezas JA, Cervera MT, Ruiz-García L, Carreño J, Martínez-Zapater JM. A genetic analysis of seed and berry weight in grapevine. Genome. 2006; 49(12):1572-1585.
4. Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith C, et al. Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. THEOR APPL GENET. 2002; 105 (5): 780-795.
5. Striem M, Spiegel-Roy P, Baron I, SaharN. The degrees of development of the seed-coat and the endosperm as separate subtraits ofstenospermocarpic seedlessness in grapesl. VITIS. 1992; 31: 149-155.
6. Striem MJ, Ben-HayyimG, Spiegel-Roy P. Identifying molecular genetic markers associated with seedlessness in grape. J AM SOC HORTIC SCI. 1996;121(5): 758-763.
7. Wang Y, Lamikanra O. Analysis of sequencing the RAPD marker linked to seedless genes in grapes. Acta Universitatis Agriculturae Boreali-occidentalis. 1996; 25(4):1-5.
8. Lahogue F, This P, Bouquet A. Identification of a codominant scar marker linked to the seedlessness character in grapevine. THEOR APPL GENET. 1998; 97(5):950-959.
9. Adam-Blondon AF, Lahogue F, Bouquet A, Boursiquot JM, This P. Usefulness of two SCAR markers for marker-assisted selection of seedless grapevine cultivars. VITIS. 2001; 40(3):147-156.

10. Mejia N, Hinrichsen P. A new, highly assertive SCAR marker potentially useful to assist selection for seedlessness in table grape breeding. In: VIII International Conference on Grape Genetics and Breeding 603: 2002; 2002: 559-564.
11. Hanania U, Velcheva M, Or E, Flaishman M, Sahar N, Perl A: Silencing of *chaperonin 21*, that was differentially expressed in inflorescence of seedless and seeded grapes, promoted seed abortion in tobacco and tomato fruits. TRANSGENIC RES. 2007; 16(4):515-525.
12. Wang S, Yu Y, Zhang C, Xu W, Wang Y. Molecular cloning and characterization of a novel gene encoding an EF-hand calcium-binding protein related to fruit seedlessness of grapevine. MSCI HORTIC-AMSTERDAM. 2011; 130(4):708-714.
13. Mejia N, Soto B, Guerrero M, Casanueva X, Houel C, Miccono Mde L, et al. Molecular, genetic and transcriptional evidence for a role of *VvAGL 11* in stenospermocarpic seedlessness in grapevine. BMC Plant Biol. 2011; 11: 57.
14. Nicolas P, Lecourieux D, Gomes E, Delrot S, Lecourieux F. The grape berry-specific basic helix-loop-helix transcription factor VvCEB1 affects cell size. J Exp Bot. 2013; 64(4):991-1003.
15. Wang L, Yin X, Cheng C, Wang H, Guo R, Xu X, et al. Evolutionary and expression analysis of a MADS-box gene superfamily involved in ovule development of seeded and seedless grapevines. Mol Genet Genomics. 2015; 290(3):825-846.
16. Royo C, Carbonell-Bejerano P, Torres-Perez R, Nebish A, Martinez O, Rey M, et al. Developmental, transcriptome, and genetic alterations associated with parthenocarpy in the grapevine seedless somatic variant *Corinto bianco*. J Exp Bot. 2015.
17. Wang L, Hu X, Jiao C, Li Z, Fei Z, Yan X, et al. Transcriptome analyses of seed development in grape hybrids reveals a possible mechanism influencing seed size. BMC genomics. 2016; 17(1):898.
18. Royo C, Torres-Pérez R, Mauri N, Diestro N, Cabezas JA, Marchal C, et al. The Major Origin of Seedless Grapes Is Associated with a Missense Mutation in the MADS-Box Gene *VviAGL 11*. PLANT PHYSIOL. 2018; 177(3):1234-1253.
19. Mackay I, Caligari P. Efficiencies of F and Backcross Generations for Bulk Segregant Analysis Using Dominant Markers. CROP SCI. 2000; 40(3):626-630.
20. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, et al. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. PLANT J. 2013; 74(1):174-183.
21. Trick M, Adamski NM, Mugford SG, Jiang CC, Febrer M, Uauy C. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biol. 2012; 12:14.
22. Wang R, Sun L, Bao L, Zhang J, Jiang Y, Yao J, et al. Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. BMC GENOMICS. 2013; 14:929.
23. Yuan YW, Sagawa JM, Di Stilio VS, Bradshaw HD. Bulk Segregant Analysis of an Induced Floral Mutant Identifies a MIXTA-like R2R3 MYB Controlling Nectar Guide Formation in *Mimulus lewisii*. GENETICS. 2013; 194(2):523-528.
24. Mu J-H, Lee H-S, Kao T. Characterization of a pollen-expressed receptor-like kinase gene of *Petunia inflata* and the activity of its encoded kinase. Plant Cell. 1994; 6(5):709-721.
25. Jinn TL, Stone JM, Walker JC. HAESA, an *Arabidopsis* leucine-rich repeat receptor kinase, controls floral organ abscission. GENE DEV. 2000; 14(1):108-117.
26. Claisse G, Charrier B, Kreis M. The *Arabidopsis thaliana* GSK3/Shaggy like kinase AtSK3-2 modulates floral cell expansion. PLANT MOL BIOL. 2007; 64(1-2):113-124.
27. Jian H, Feng C, Cecilia DC, Antonella A, Mouhua S, Shuai Y, et al. An ankyrin repeat-containing protein, characterized as a ubiquitin ligase, is closely associated with membrane-enclosed organelles and required for pollen germination and pollen tube growth in lily. PLANT PHYSIOL. 2006; 140(4):1374-1383.

28. Niu BX, He FR, He M, Ren D, Chen LT, Liu YG. The ATP-binding Cassette Transporter OsABCG15 is Required for Anther Development and Pollen Fertility in Rice. J INTEGR PLANT BIOL. 2013; 55(8):710-720.
29. Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ. The *AUXIN RESPONSE FACTOR 2* gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. DEVELOPMENT. 2006; 133(2); 251-261.
30. Pandolfini T, Molesini B, Spena A. Molecular dissection of the role of auxin in fruit initiation. TRENDS PLANT SCI. 2007; 12(8):327-329.
31. De Jong M, Wolters-Arts M, Feron R, Mariani C, Vriezen WH. The *Solanum lycopersicum* auxin response factor 7 (*SlARF7*) regulates auxin signaling during tomato fruit set and development. PLANT J. 2009; 57(1):160-170.
32. Zhu Y, Nomura T, Xu Y, Zhang Y, Peng Y, Mao B, et al. *ELONGATED UPPERMOST INTERNODE* encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice. PLANT CELL. 2006; 18(2):442-456.
33. Nafisi M, Goregaoker S, Botanga CJ, Glawischnig E, Olsen CE, Halkier BA, et al. *Arabidopsis* cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. PLANT CELL. 2007; 19(6):2039-2052.
34. Zhao Y, Christensen SK, Fankhauser C, Cashman JR, Cohen JD, Weigel D, et al. A Role for Flavin Monooxygenase-Like Enzymes in Auxin Biosynthesis. Science. 2001; 291(5502):306-309.
35. Ito T, Meyerowitz EM. Overexpression of a Gene Encoding a Cytochrome P450, *CYP78A9*, Induces Large and Seedless Fruit in *Arabidopsis*. PLANT CELL. 2000; 12(9):1541.
36. Meyer S, Scholz-Starke J, De Angeli A, Kovermann P, Burla B, Gambale F, et al. Malate transport by the vacuolar AtALMT6 channel in guard cells is subject to multiple regulation. PLANT J. 2011; 67(2):247-257.
37. Yang B, Dougherty L, Li M, Fazio G, Cheng L. A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the locus is associated with low fruit acidity in apple. MOL GENET GENOMIC. 2012; 287(8):663-678.
38. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England) 2014; 30(15):2114-2120.
39. Jaillon O, Aury J, Noel B, Policriti A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 2007; 449(7161):463 - 467.
40. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England) 2009; 25(14):1754-1760.
41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) 2009; 25(16):2078-2079.
42. Joung JG, Corbett AM, Fellman SM, Tieman DM, Klee HJ, Giovannoni JJ, et al. Plant MetGenMAP: an integrative analysis system for plant systems biology. Plant Physiol. 2009; 151(4):1758-1768.
43. Hill JT, Demarest BL, Bisgrove BW, Gorski B, Su Y-C, Yost HJ. MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. GENOME RES. 2013; 23(4):687-697.
44. Guo C, Guo R, Xu X, Gao M, Li X, Song J, et al. Evolution and expression analysis of the grape (*Vitis vinifera* L.) WRKY gene family. J Exp Bot. 2014; 65(6):1513-1528.

Tables

Table 1. Information of high quality clean reads mapped with reference genome

Sample	All Reads	Single Mapped Reads	Paired Mapped Reads	Unmapped Reads	Align Ratio(%)
C	57160754	136895	27239591	2544677	95.55
R	68761692	179104	32691359	3199870	95.35
SL	50978504	192890	24333081	2119452	95.84
S	56263426	224997	26903173	2232083	96.03

Table 2. Function classification of genes containing functional variants

Function classification	Associated genes	Associated DEGs at least at one stage	Associated DEGs at all three stages
frameshift deletion	454	124	35
frameshift insertion	309	91	24
nonsynonymous SNP	11381	3790	1147
stopgain	deletion	10	3
	insertion	13	3
	SNP	298	68
stopless	deletion	6	2
	insertion	15	6
	SNP	214	52

Table 3. Identified candidate genes based on expression analysis from transcriptome

Gene ID	Gene name	Chr	Position	Ref	Mut	Description	Ratio(Seedless/Seeded)		
							Stage 1	Stage 2	Stage 3
GSVIVG01036048001	<i>G1</i>	chr6	21418457	C	G	Phosphatidylglycerol/phosphatidylinositol transfer protein	1	2.19	0.89
GSVIVG01011008001	<i>G2</i>	chr7	2257286	G	A	Putative auxin response factor 2	0.2	0.08	0.56
GSVIVG01011007001	<i>G3</i>	chr7	2266373	C	T	Serine protease	2.37	1.48	6
			2266664	C	T				
GSVIVG01011006001	<i>G4</i>	chr7	2278664	A	C	Monooxygenase, putative	5.99	3.15	6.07
GSVIVG01000977001	<i>G5</i>	chr13	11040377	A	T	Histidine kinase-like ATPase domain-containing protein	2.75	0.12	0.83
GSVIVG01000993001	<i>G6</i>	chr13	11222492	G	A	Tir-nbs-lrr resistance protein	188	3.51	33.3
			11223120	T	C				
			11223228	C	A				
			11223476	C	T				
GSVIVG01000994001	<i>G7</i>	chr13	11229443	G	A	Aluminum-activated malate transporter	27.97	2.56	6.88
GSVIVG01000995001	<i>G8</i>	chr13	11240814	G	A	Putative aluminum-activated malate transporter	10.62	2.88	5.59
GSVIVG01001005001	<i>G9</i>	chr13	11336347	T	C	Protein trichome birefringence-like 36	5.83	6.6	4.09
			11337398	G	C				

Additional Files

Additional file 1: Figure S1. Base composition and quality value distribution after data filtering. Line A, T, C and G indicate the percent of corresponding base at each position, respectively. Mean refers to the average quality of all bases at each position, while Q20 and Q30 indicate the percent of bases with quality value no less than 20 and 30, respectively.

Additional file 2: Table S1. Base information before and after filtering. 'C' represents 'Centennial seedless', 'R' 'Red Globe', 'S' 'Seeded progeny', and 'SL' 'seedless progeny'.

Additional file 3: Table S2. Reads information before and after filtering. 'C' represents 'Centennial seedless', 'R' 'Red Globe', 'S' 'Seeded progeny', and 'SL' 'seedless progeny'.

Additional file 4: Table S3. Number of DEGs containing functional variants.

Additional file 5: Figure S2. Distribution of SNP-index. The x-axis indicates the chromosomal position, each spot represents a SNP. Average values of SNP-index or delta SNP were indicated by the black lines. (a). SNP-index plot of 'Seeded' bulk; (b). SNP-index plot of 'Seedless' bulk; (c). Plot of delta SNP-index. The red line indicate the threshold line.

Additional file 6:Table S4. Functional variants in the valley.

Additional file 7: Table S5. Functional variants in the peak.

Additional file 8Figure S3. Distribution of candidate genes based on expression analysis from transcriptome data.

Additional file 9: Table S6.Motif analysis of mutant sequences.

Additional file 10: Table. S7. The primer sequences used for expression analysis.

Figures

Figure 1

Overview of the experimental design used in this study. Scale bars are 1 cm.

Figure 2

Information about identified genomic variants. (a). Position of variants in relation to genes. (b). Functional consequences of variants within exons. (c). Transitions and transversions analysis of genomic SNP

Figure 3

Analysis of GO terms associated with DEGs containing functional variants. (a). Venn analysis of GO terms at three developmental stages. (b). Enrichment of identified GO terms. The red and green color indicate enriched GO terms associated with up- and down-regulated DEGs, respectively

Figure 4

Pathways enrichment of DEGs containing functional variants. (a). Venn analysis of pathways at three developmental stages. (b). Enrichment of identified pathways

Figure 5

Expression profile of DEGs containing functional variants at three developmental stages. (a). Expression profile of DEGs encoding protein kinases. (b). Expression profile of DEGs encoding disease resistant proteins. (c). Expression profile of DEGs encoding cytochrome P450s.

Figure 6

Expression profile of functional variant-DEGs encoding transcription factors at three developmental stages.

Figure 7

Expression profile of functional variant DEGs affecting seed development at three developmental stages.

Figure 8

Real-time quantitative PCR expression levels of candidate genes containing functional variants in different tissues. Mean values and SDs were obtained from three biological and three technical replicates. 'R.G' stands for 'Red Globe' and 'T.S' 'Thompson Seedless'.

Figure 9

Real-time quantitative PCR expression levels of candidate genes containing functional variants during ovule development of seeded and seedless cultivars. Mean values and SDs were obtained from three biological and three technical replicates. 'K' stands for 'Kyoho', 'R.G' 'Red Globe', 'F' 'Flame Seedless' and 'T.S' 'Thompson Seedless'.

Supplementary Files

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

- [Additionalfile3SupplementaryTable2.xls](#)
- [Additionalfile8SupplementaryFigure3.tif](#)
- [Additionalfile5SupplementaryFigure2.tif](#)
- [Additionalfile2SupplementaryTable1.xls](#)
- [Additionalfile4SupplementaryTable3.xls](#)
- [Additionalfile1SupplementaryFigure1.tif](#)
- [Additionalfile9SupplementaryTable6.xls](#)
- [Additionalfile10SupplementaryTable7.xls](#)

- [Additionalfile7SupplementaryTable5.xls](#)
- [Additionalfile6SupplementaryTable4.xls](#)