Integrative genomic and transcriptomic analyses of a bud sport mutant 'Jinzao Wuhe' with the phenotype of large berries in grapevines

Jianquan Huang[1†], Guan Zhang[2,3†], Yanhao Li[1,4], Mingjie Lyu[2], He Zhang[1], Na Zhang[1*], Rui Chen[2*]

[1] The Research Institute of Forestry and Pomology, Tianjin Academy of Agricultural Sciences, Tianjin, China

[2] Institute of Crop Germplasm and Biotechnology, Tianjin Academy of Agricultural Sciences, Tianjin, China

[3] College of Biotechnology and Food Science Food Science, Tianjin University of Commerce, Tianjin, China

[4] College of Horticulture and Gardening, Tianjin Agricultural University, Tianjin, China

[†] Jianquan Huang and Guan Zhang are co-first authors.

Corresponding Author:

Na Zhang[1*], Rui Chen[2*]

Tianjin Academy of Agricultural Sciences, Tianjin, China

Email address: Na Zhang (2371254838@qq.com) and Rui Chen (chenrui_aglab@126.com)

Abstract

Background: Bud sport mutation occurs frequently in fruit plants and acts as an important approach for grapevine improvement and breeding. 'Jinzao Wuhe' is a bud sport of the elite cultivar 'Himord Seedless' with obviously enlarged organs and berries. To date, the molecular mechanisms underlying berry enlargement caused by bud sport in grapevines remain unclear.

Methods: Whole genome resequencing (WGRS) was performed for two pairs of bud sports and their maternal plants with similar phenotype to identify SNPs, InDels and structural variations (SVs) as well as related genes. Furthermore, transcriptomic sequencing at different developmental stages and weighted gene co-expression network analysis (WGCNA) for 'Jinzao Wuhe' and its maternal plant 'Himord Seedless' were carried out to identify the differentially expressed genes (DEGs), which were subsequently analyzed for Gene Ontology (GO) and function annotation.

Results: In two pairs of enlarged berry bud sports, a total of 1,334 SNPs, 272 InDels and 74 SVs, corresponding to 1,022 target genes related to symbiotic microorganisms, cell death and other processes were identified. Meanwhile, 1,149 DEGs associated with cell wall modification, stress-response and cell killing might be responsible for the phenotypic variation were also determined. As a result, 42 DEGs between 'Himord Seedless' and 'Jinzao Wuhe' harboring

39 genetic variations were further investigated, including pectin esterase, cellulase A, cytochromes
40 P450, UDP-glycosyltransferase (UGT), zinc finger protein, auxin response factor (ARF), NAC
41 transcription factor (TF), protein kinase, etc. These candidate genes offered important clues for a
42 better understanding of developmental regulations of berry enlargement in grapevine.
43 **Conclusion:** Our results provide candidate genes and valuable information for dissecting the
44 underlying mechanisms of berry development and contribute to future improvement of grapevine
45 cultivars.
46 **Key words:** grapevine; bud sport; berry enlargement; genetic variation; WGRS; RNA-seq.
47

Introduction

49 Grape (*Vitis vinifera L.*) is one of the most favorite fruits worldwide with great economic and
50 nutritional value (Yang & Xiao, 2013). Different from routine improvement strategies in crops,
51 grapevine is highly heterozygous and exhibit serious genetic segregation, which hinders the
52 development process of excellent traits. So far, various breeding techniques had been applied to
53 improve the genetic background of grapes and achieve reliable and stable inheritance of
54 desirable traits (Yamada & Sato, 2016). Out of these approaches, bud sport is an important
55 strategy for grapevine improvement and breeding. Bud sport mutation is a phenomenon usually
56 happened in fruit plants in which stable somatic cells are mutated in the meristematic tissue of
57 the bud and exhibit novel traits due to endogenous and exogenous factors. In this process,
58 favorable phenotypes caused by mutation (e.g., shorter ripening cycle, larger berries, and
59 different skin color) will be selected by breeders that are significantly different from the maternal
60 plants with desired and superior characteristics (Foster & Aranzana, 2018). To date, numerous
61 bud sport mutants have been observed and successfully utilized in grapevine breeding, such as
62 early-ripening cultivars 'Fengzao', 'Tiangong Moyu', and 'Nantaihutezao' (Da-Long Guo &
63 Guo-Hai Zhang, 2015; Wei et al., 2020; Leng et al., 2021), as well as berry skin color mutants
64 'Pinot Grigio', 'Pinot Blanc' and 'Benitaka' (Azuma et al., 2009; Pan et al., 2012; Negri et al.,
65 2015). However, the phenotype with large berries induced by bud sport has not been reported
66 and the corresponding regulatory mechanisms remain poorly understood.
67 Berry size is a polygenic quantitative trait in grapevine. Similar to other fruit plants, grapevine
68 follows a double sigmoid curve corresponding to the fruit enlargement, véraison and maturity
69 stages. Few days after flowering, berry development occurs through cell division and expansion.
70 During the enlargement stage, the berry grows slowly and ends with the véraison stage, where it
71 softens and the color of the skin changes. In the véraison stage, berry growth restarts but
72 exclusively through cell enlargement. In the maturity stage, all the tissues that make up the
73 berries are formed. Previous studies in Arabidopsis (Krizek, 1999; Wang et al., 2016), apple
74 (Malladi & Hirst, 2010), tomato (Su et al., 2014) and peach (Scorzal et al., 1991) indicated that
75 cell division and cell expansion play crucial roles in controlling fruit size.
76 Several key genes and regulators had been demonstrated to be responsible for berry enlargement
77 which involved in cell wall modification, phytohormone production, water and sugar transport
78 processes. "Pectin Lyase (PL)", "wall-associated receptor kinase (WAK)" and "Pectin Esterase

79   (PE)" are related with the accumulation of pectin, that can directly bind pectin polymers to
80   reduce stiffness and stimulate cell wall development (Tucker et al., 2018). Ethylene was
81   previously thought to have limited effects on grape growth and development, that is why
82   grapevines were classified as non-climacteric berries (Chervin et al., 2004). Recent evidence
83   suggested that ethylene coupled with increased ACO transcription could produce a modest peak
84   during berry ripening and promote larger berry diameter, higher sugar accumulation (De La
85   Torre-Ruiz et al., 2016), and higher anthocyanin synthesis (El-kereamy et al., 2003). UGT, as
86   sugar donors, are primarily responsible for the formation of secondary metabolites such as
87   antibiotics, flavanols, and steroids through glycosylation modifications that affect plant growth
88   and development (Kiedrowski et al., 1992; Mitchell, Hall & Barber, 1994; Breton, Fournel-
89   Gigleux & Palcic, 2012). Cytochrome P450 (CYP) is a highly conserved enzyme family that
90   plays significant roles in plant development and growth by increasing cell size, responding to
91   biotic and abiotic stresses, and stimulating phytohormone signal transduction (Ma et al., 2015).
92   *PaCYP78A9* was inferred to be involved in the regulation of organ size in sweet cherry (Qi et al.,
93   2017). CYPs, such as *EOD3/CYP78A6* and *CYP78A9* in Arabidopsis (Fang et al., 2012;),
94   *CYP78A5* in tomato (Chakrabarti et al., 2013), *GmCYP78A10* in soybean (Wang et al., 2015),
95   *CYP78A13* (Xu et al., 2015) and *GW10* in rice (Zhan et al., 2021), have all been identified as
96   seed size regulators.
97   TFs are also important regulators in controlling berry size which are typically transcribed in
98   response to hormonal and environmental stimuli. NAC transcription factor has been proven to be
99   functional in fruit development and maturation (Migicovsky et al., 2021). For example, the
100   participation of *VvNAC26* determinates the berry ultimate size in grapevine (Tello et al., 2015).
101   Overexpression of *VvCEB1* and *VvSAUR41* genes in grapevine embryos increase fruit diameter
102   by stimulating cell expansion (Nicolas et al., 2013; Li et al., 2021). *VvANT1*, an ortholog of
103   Arabidopsis ANT, could regulate cell proliferation and organ expansion (Chialva et al., 2016). In
104   addition, Martins et al. have demonstrated that calcium-related genes have a significant
105   regulatory effect on the cell wall biogenesis by influencing the expression of pectin methyl
106   esterase (PME) and polygalacturonase (PG) (Martins et al., 2018).
107   In this study, 'Jinzao Wuhe' is a very rare bud mutation discovered and developed by our
108   laboratory due to its remarkably enlarged berry size. Herein, we carried out WGRS and genetic
109   variation analysis for 'Jinzao Wuhe' and its maternal cultivar to explore the candidate genes in
110   response to berry development, as well as the potential mechanisms underlying bud sport in
111   grapevines. Furthermore, transcriptome sequencing and WGCNA analyses were performed at
112   four developmental time-points, leading to the identification of causal genes and regulators.
113   These results would shed light on the molecular mechanisms of bud sport mutation and offer
114   useful information for future grapevine improvement.
115

Materials & Methods

116   **Materials & Methods**
117   **Plant materials**

118   'Himord Seedless' is an early maturing cultivar with seedless and small berries, ripening before
119   the onset of the rainy season in Tianjin, China. 'Jinzao Wuhe' is a bud sport derived from
120   'Himord Seedless' discovered at Tianjin Academy of Agricultural Sciences (Tianjin, China)
121   orchard in 2014. The three-year-old vines 'Himord Seedless' and 'Thompson Seedless' and their
122   bud sports 'Jinzao Wuhe' and 'Dawuhebai' were planted in the grape garden of Tianjin
123   Academy of Agricultural Sciences, Tianjin, China, grown on Y- shaped trellises with 1.5m
124   (column) $\times$ 3m (row) interval. Management guidelines include techniques for irrigation,
125   fertilization, pruning, weeding, disease prevention, pest control and precautions during rainy
126   season. Compared with its maternal plant 'Himord Seedless', 'Jinzao Wuhe' exhibits larger
127   leaves, inflorescences, berries, thicker stems, roots, and greater vigorous growth compared to its
128   maternal plant. Especially, the berry size of 'Jinzao Wuhe' is almost twice as large as its
129   maternal plant, which is an excellent germplasm resource for grapevine breeding. In addition,
130   'Dawuhebai', another bud sport derived from table grape 'Thompson Seedless', also possesses
131   enlarged berries that would be helpful for investigating the regulatory mechanisms underlying
132   berry enlargement (Baiano & Terracone, 2012; Somkuwar et al., 2014).
133   The longitudinal diameter, transverse diameter, volume and weight of each ten berries for each
134   replicate were measured individually for every seven days started from the 10 days post-anthesis
135   (DPA). Three biological replicates were set for each treatment. For each grapevine cultivar, 2-
136   week-old leaves without pathogens and insect infestations were harvested for genomics DNA
137   isolation, library construction and WGRS. Inflorescences and berries of 'Himord Seedless' and
138   'Jinzao Wuhe' that were healthy and vigorous were sampled at the full-flowering stage (stage 1:
139   1 DPA), berry expansion stage (stage 2: 14 DPA), véraison stage (stage 3: 35 DPA) and maturity
140   stage (stage4: 56 DPA) for RNA extraction and transcriptome sequencing, respectively. Three
141   biological replicates were prepared for each sampling time point. These leaf, inflorescence and
142   berry samples were immediately snap-frozen by immersion in liquid nitrogen and stored at -80℃
143   until use.
144
145   **DNA and RNA extraction and sequencing**
146   Genomic DNA was extracted from grapevine leaves by using the modified protocol of cetyl
147   trimethyl ammonium bromide (CTAB). The concentration and purity of the extracted genomic
148   DNA was measured by NanoDrop-1000 spectrophotometer (NanoDrop Technologies, USA) and
149   the quality and integrity of isolated genomic DNA was assessed by 1% agarose gel
150   electrophoresis. Paired-end (PE) Deep sequencing libraries were prepared according to the
151   standard protocol (Illumina) with the Paired-end (PE) strategy. Finally, the qualified libraries
152   were sequenced using the Illumina HiSeq 2500 platform in Novogene Biotechnology Company
153   (Tianjin, China). Raw data was deposited in the NCBI SRA database under the BioProject
154   accession PRJNA795039 and PRJNA872374 (Table S1).
155   Total RNA was isolated from inflorescences and berries of 'Himord Seedless' and 'Jinzao
156   Wuhe' at 1, 14, 35, and 56 DPA using TRIzol reagent according to the manufacturer's
157   instructions. The quality of RNA was evaluated by NanoDrop-1000 spectrophotometer

158   (NanoDrop Technologies, USA), and RNA integrity was determined by 1% denaturing gel
159   electrophoresis. Three replicates at different stages were set for each sample. RNA-seq libraries
160   were then prepared with the Illumina TruSeq RNA Sample Prep Kit and sequenced by using the
161   Illumina HiSeq X Ten System in Novogene Biotechnology Company (Tianjin, China). All raw
162   RNA-seq data were deposited in the NCBI SRA database under BioProject accession
163   PRJNA795246 (Table S1).
164
165   **SNP, InDel and SV calling and annotation**
166   Illumina raw reads were filtered and trimmed for low quality bases and adapter sequences using
167   Fastp (default parameters), and then mapped onto the reference genome of *V. vinifera* Pinot Noir
168   PN40024 (12X, GCF_000003745.3) using BWA (version 0.7.17) (Li & Durbin, 2009) with
169   default parameters. Marking duplicated sequences caused by PCR reactions using the Genome
170   Analysis Toolkit (GATK4) (McKenna et al., 2010) with the MarkDuplicates module (default
171   parameters). Both SNPs and InDels were identified for each sample using the GATK4 pipeline
172   with the HaplotypeCaller module according to the recommended variation detection workflow
173   (gatk --java-options HaplotypeCaller -R $genome_file -ERC GVCF -I $bam -O $out). The
174   criteria for variant filtration were as follows: ⅰ) biallelic SNPs and InDels were retained; ⅱ)
175   sequencing depth of higher than 5 was selected; ⅲ) variant allele frequencies of less than 40%
176   were excluded; ⅳ) missing genotypes were deleted; ⅴ) InDels shorter than or equal to 50 bp
177   were retained.
178   Structural variations (SVs) were detected in all samples using LUMPY (version 0.2.13) (Layer et
179   al., 2014) and DELLY (version 0.8.7) (Rausch et al., 2012). Splitters and discordant reads were
180   extracted by using Lumpyexpress (lumpyexpress -B $sample.bam -S $sample.splitters.bam -D
181    $sample.discordants.bam -o lumpyout_single_$tag.vcf) for each individual sample. Bam files
182   were sorted and indexed using Samtools (version 1.12) (Li et al., 2009) and SVs were genotyped
183   using SVTyper software. For DELLY (version 0.8.7), SVs were presented with the
184   recommended instructions in BCF format (delly call -t ALL -g $genome $sample.bam -o
185    $sample.bcf), which were converted to VCF format using BCFtools (bcftools view $sample.bcf
186   > $sample.vcf) (Danecek & McCarthy, 2017). Moreover, SURVIVOR (SURVIVOR merge) was
187   used to merge SVs generated from both LUMPY and DELLY pipelines. As for filtration, SVs
188   with a length of 50 bp or more and the quality flag PASS, as well as the precise breakpoints,
189   were selected as the final dataset for further analysis.
190   The intersection of SNP, InDel, and SV datasets of the two pairs of bud mutation and their
191   maternal were selected separately. Then, the effect of all selected variants located in
192   chromosomes was annotated using Snpeff (java -jar snpEff.jar eff -csvStats $sample.csv -s
193    $sample.html -c snpEff.config -v grape $sample.vcf.gz >$sample_snpeff.gz) (Cingolani et al.,
194   2012), which classified them into HIGH, LOW, MODERATE, and MODIFIER categories
195   according to the predicted results for influencing gene function. Important and non-synonymous
196   SNPs, InDels and SVs with high impact were selected as target variants that might be
197   responsible for phenotype variation.

198
199 **Transcriptome analysis**
200 Tophat-Cufflinks pipeline was used to generate gene expression profiles and identify DEGs
201 (Trapnell, Pachter & Salzberg, 2009). First of all, raw reads were preprocessed using Fastp
202 (default parameters) (Chen et al., 2018) to remove adaptors and low quality bases. Tophat2
203 (default parameters) was carried out for genomic mapping. Gene expression levels were
204 estimated as fragments per kilobase of transcript per million (FPKM) by Cufflink (default
205 parameters) based on the alignments. DEGs between different developmental stages were
206 identified using Cuffdiff with regarding a false discovery rate (FDR) ≤ 0.05 and log2 Fold-
207 Change ≥ 1 as thresholds. Meanwhile, HiSAT2 (Kim, Langmead & Salzberg, 2015), feature
208 counts (Liao, Smyth & Shi, 2014) and DESeq2 (Love, Huber & Anders, 2014, p. 2) workflows
209 were also performed to achieve a more comprehensive result of DEGs analysis. The statistical
210 power calculated using RNASeqPower (Bioconductor - RNASeqPower) was 0.96. We visualized
211 the correlation among the three biological replicates for each sample and gene expression level
212 using the R package 'scatterplot3d' and 'pheatmap', respectively.
213
214 **WGCNA network analysis**
215 WGCNA was carried out in R program with the WGCNA package (Langfelder & Horvath,
216 2008). Following the WGCNA tutorial, the FPKM values of DEGs were chosen for further
217 investigation. The soft-thresholding power (β) was determined using the "pickSoftThreashold"
218 function. Hierarchical cluster tree displaying co-expression modules was constructed as follows:
219 minimodule size = 50, deep split = 2; merge cut height = 0.35. To identify the relationship
220 between co-expression modules and phenotypes, the module-trait relationships were
221 subsequently determined by Pearson correlation.
222

223 # Results
224 **Phenotypical comparison between 'Himord Seedless' and its bud sport mutant 'Jinzao**
225 **wuhe'**
226 Physiological traits were measured, where 'Jinzao Wuhe' showed a double sigmoid pattern curve
227 from fruiting (10 DPA) to maturity (59 DPA) with higher berry longitudinal diameter, transverse
228 diameter, volume and weight than 'Himord Seedless' (Fig. 1B, C, D, E; Table S2). Two
229 discriminable stages for berry expansion could be observed during berry growth. At the end of
230 maturity (59 DPA), 'Jinzao Wuhe' was about twice as large as the 'Himord Seedless' in terms of
231 volume and weight (Fig. 1A).
232
233 **SNP, InDel and SV calling**
234 To identify the genome-wide genetic variations between two maternal grapevines (WTs:
235 'Thompson Seedless' and 'Himord Seedless') and their bud mutants (MUTs: 'Dawuhebai' and
236 'Jinzao Wuhe'), WGRS was performed, producing a total of 305,400,258 raw reads (~ 45.8 Gb).
237 After removing the low-quality reads and bases, a total of 302,740,380 clean reads (~ 45.5 Gb)

238   were retrieved, accounting for more than 99.34% of the raw data. The Q20 and Q30 Phred
239   Scores of all samples were greater than 93.57% and 88.93%, respectively (Table 1). The
240   genomic mapping rates for clean reads ranged from 91.91% to 95.87% (Table 1).
241   For 'Thompson Seedless', 'Dawuhebai', 'Himord Seedless' and 'Jinzao Wuhe', 3,763,122,
242   3,947,775, 5,512,478 and 5,499,289 SNPs were detected, respectively. As a result, MUTs are
243   richer in SNPs than WTs and the number of SNPs of 'Jinzao Wuhe' is higher than 'Dawuhebai'
244   (Fig. 2A). We also discovered 2,658,536, 3,663,339, 2,531,663 and 3,670,669 transitions (Ts)
245   and 1,289,239, 1,835,950, 1,231,459 and 1,841,809 transversions (Tv) in 'Thompson Seedless',
246   'Dawuhebai', 'Himord Seedless', and 'Jinzao Wuhe', respectively, with genome-wide transitions
247   to transversions ratio (Ts/Tv) were 2.06, 2.00, 2.06, and 1.99 (Table S3). SnpEff annotation
248   revealed that the SNPs were mainly located in the upstream and downstream of genes, followed
249   by intergenic regions, intronic regions, and exon regions, while they were scarce in splice site
250   regions, UTR3' and UTR5' regions. In total, the heterozygous rates in 'Himord Seedless' and
251   'Jinzao Wuhe' were apparently higher than those in 'Thompson Seedless' and 'Dawuhebai'
252   (Table S3).
253   'Thompson Seedless', 'Dawuhebai', 'Himord Seedless' and 'Jinzao Wuhe' have 698,381
254   (insertion: 382,040; deletion: 316,341), 752,785 (insertion: 341,594; deletion: 411,191),
255   1,113,041 (insertion: 516,558; deletion: 596,483) and 1,108,155 (insertion: 513,699; deletion:
256   594,456) InDels, respectively. It could be inferred that 'Himord Seedless' and 'Jinzao Wuhe'
257   have more distant genetic relationships with the reference genome (Table S4). Besides, the
258   length distribution of InDels showed that mononucleotide insertions or deletions accounted for
259   the majority of InDels (Fig. 2B). The frequency of InDels located in the upstream and
260   downstream of genes was higher, averaging at 29.30 % and 30.26 %. Approximately 15.00 %,
261   11.00 %, and 1.00 % of InDels harbored in intergenic, intronic, and exon regions, respectively.
262   And they were almost undetectable for the splice site, UTR3'and UTR5' regions (all less than
263   1.00 %). In addition, homozygous and heterozygous InDels in four grapevine samples were also
264   calculated, with the heterozygous rate of 79.32 % and 79.19 % in 'Himord Seedless' and 'Jinzao
265   Wuhe', respectively, which were higher than 68.02 % and 70.50 % in 'Thompson Seedless' and
266   'Dawuhebai', respectively (Table S4).
267   SVs play an important functional role in genetic and morphological divergence, mainly including
268   deletions (DEL), translocation (BND), inversions (INV), duplications (DUP) and insertions
269   (INS). Due to the limitations of Illumina sequencing platform that usually producing short reads
270   less than 200 bp, the results of DUP, INV, and INS were excluded for further investigation. Only
271   DEL and BND were remained for subsequent analysis. In 'Thompson Seedless', 'Dawuhebai',
272   'Himord Seedless' and 'Jinzao Wuhe', 18,717, 23,237, 31,807 and 29,715 SVs were identified,
273   respectively. The number of SVs in MUTs was markedly higher than in WTs. The distribution of
274   SVs of four grape lines showed that the proportion of deletions was more than 53.73%, whereas
275   other SVs accounted for less than 46.27%. (Fig. 2C). In summary, the distribution and
276   proportions of SVs among these four grapevine lines were highly similar to those of SNPs and
277   InDels (Table S5).

278

Identification of candidate genes harboring SNPs, InDels and SVs

279

280 To identify potential causal genes that are affected by DNA variations, comparative analysis
281 between WTs and their MUTs were performed firstly. For 'Thompson Seedless' and
282 'Dawuhebai', 867,546, 184,194 and 21,817 different SNPs, InDels and SVs, as well as
283 3,535,491, 633,486 and 10,066 common SNPs, InDels and SVs were identified. For 'Himord
284 Seedless' and 'Jinzao Wuhe', 706,821, 216,022 and 30,521 different SNPs, InDels and SVs, as
285 well as 5,152,473, 1,002,587 and 15,497 common SNPs, InDels and SVs were identified,
286 respectively (Fig. 3A, B, C). Then, we compared the genetic variants between WTs and MUTs
287 individually. Between 'Thompson Seedless' and 'Himord Seedless', 2,682,997 SNPs, 456,929
288 InDels and 8,250 SVs were regarded as common genetic variants. Between 'Dawuhebai' and
289 'Jinzao Wuhe', 2,775,884 common SNPs and 481,550 common InDels and 9,000 common SVs
290 were identified (Fig. 3A, B, C).
291 These common variants derived from two independent bud sport mutation events could offer
292 crucial clues for explaining the phenotypic variation in MUT berries that growing larger. Finally,
293 35,733 SNPs, 9,128 InDels and 1,166 SVs that were absent in WTs but co-existed in MUTs, as
294 well as 23,987 SNPs, 5,955 InDels and 763 SVs that were absent in MUTs but co-existed in
295 WTs were deemed as target variants that caused by bud sport mutation (Fig. 3A, B, C).
296 According to the results of snpEff analysis, 61 SNPs, 272 InDels and 74 SVs were predicted to
297 have HIGH impact, whereas 1,273 SNP entries (missense variants) to have MODERATE impact
298 (Table S6). These variants were distributed unevenly across 19 chromosomes, with SNPs and
299 InDels predominating on Chr18 and Chr12, and SVs predominating on Chr 18 and Chr5 (Fig.
300 3D).
301 As a result, a total of 1,022 unique genes were finally identified based on these determining
302 genetic variations, in which 891, 171 and 68 genes were associated with SNPs, InDels and SVs,
303 respectively (Table S6). GO enrichment analysis of these genes showed 20 representative GO
304 terms, including "biological process involved in symbiotic interaction", "biological process
305 involved in interaction with symbiont", "plant-type hypersensitive response", "programmed cell
306 wall", "glycoside metabolic process", "detection of biotic stimulus", "detection of external biotic
307 stimulus", "structural constituent of nuclear pore", "regulation of cell fate commitment" etc.
308 (Fig. 3E).

309

Identification of the DEGs between 'Himord Seedless' and 'Jinzao Wuhe'

310

311 In order to obtain evidence at the transcriptomic level, RNA-seq was carried out and DEGs were
312 investigated at four stages during berry development for 'Himord Seedless' and its mutant
313 'Jinzao Wuhe'. There was a good correlation among the three biological replicates for each
314 sample (Fig. 4A). According to the analysis results of Cuffdiff, 928 (309 up-regulated and 619
315 down-regulated), 841 (358 up-regulated and 483 down-regulated), 1,627 (846 up-regulated and
316 781 down-regulated) and 727 (548 up-regulated and 179 down-regulated) DEGs were identified
317 at stages 1-4, respectively (Fig. 4B, C, E, F; Table S7). After the redundancy removal, a total of

318   3,321 DEGs were identified that differentially expressed at least one time point (Fig. 4D). To
319   achieve a more reliable result, another pipeline for RNA-seq and DEG analysis was performed
320   using the same datasets. The results of DESeq2 generated 995 (346 up-regulated and 649 down-
321   regulated), 1,080 (425 up-regulated and 655 down-regulated), 2,055 (1125 up-regulated and 930
322   down-regulated) and 999 (733 up-regulated and 266 down-regulated) DEGs at stages 1-4,
323   respectively (Table S8). As a result, 2,984 DEGs were finally confirmed by coupling two
324   separate bioinformatic pipelines.
325   The GO enrichment analysis of theses DEGs identified 20 representative GO terms at four berry
326   development stages. The terms were related to cell wall modification, such as "plant-type cell
327   wall modification", "cell wall modification", "callose deposition in cell wall", "cell wall
328   thickening", "pectin esterase activity" in all stages, which could provide a reasonable explanation
329   between cell wall biogenesis and berry enlargement. Several GO terms were associated with
330   stress-response, such as "response to heat", "response to toxic substance", "response to
331   antibiotic", "response to reactive oxygen", "response to auxin", "response to fatty acid",
332   "response to toxic substance", "response to jasmonic acid" and "response to karrikin". For DEGs
333   in stage 1 and 3, significantly enriched GO terms were "cell killing", "regulation of cell killing",
334   "positive regulation of cell killing", "killing of cells of other organism" and "regulation of killing
335   of cells of other organism". At the stage 2, the GO terms "phenylpropanoid biosynthetic
336   process", "phenylpropanoid metabolic process" and "suberin biosynthetic process" were
337   significantly enriched. These detected pathways and related genes offered novel clues and
338   contributed to our better understanding of the mechanisms taking place behind the phenomenon
339   of bud sport mutation which leading to berry enlargement in 'Jinzao Wuhe'. (Fig. 4G, H, I, J).
340
341   **Identification of genes related to berry enlargement using WGCNA analysis**
342   WGCNA analysis was performed utilizing the FPKM values of 2,984 DEGs to investigate co-
343   expression patterns and determine the hub genes from DEGs. When the soft threshold β was set
344   to 2, the gene network initially corresponded to a scale-free network (Fig. S1). Based on their co-
345   expression patterns, all genes were further grouped into 18 modules (Fig. 5A). Berry phenotype
346   data was used for correlation analysis including transverse diameter, longitudinal diameter,
347   volume and weight. As a result, the 'blue' and 'brown' modules (p < 0.01), corresponding to 816
348   and 333 genes, were significantly clustered and linked to berry enlargement (Fig. 5B; Table S9).
349
350   **Integrated results of DEGs and WGCNA for bud sport grapevine 'Jinzao Wuhe'**
351   Based on the previous results from WGRS, RNA-seq and WGCNA, key genes were gradually
352   determined by means of DEGs, specified DNA variations and co-expression pattern with berry
353   phenotypic data. Finally, 42 key genes carrying high or moderate SNPs, InDels and SVs were
354   discovered, which closely related to cell wall modifications (*LOC100258861*, *LOC100258876*),
355   cytochromes P450 (*LOC100244431*, *LOC100253660*, *LOC100266067*, *LOC100254242*), UGT
356   (*LOC100242998, LOC100260618*), zinc finger protein (*LOC100260826*), RING-type E3
357   ubiquitin transferase (*LOC100250551*), ARFs (*LOC100265555*), NAC domain-containing

358   protein (*LOC100251801*), ABC transporter C family member (*LOC100259590*), LRR receptor-

359   like serine threonine-protein kinase (*LOC100243949, LOC100247557, LOC109124293*), protein

360   kinase (*LOC100853763, LOC100258962, LOC100267572, LOC100266874, LOC100258928,*

361   *LOC104880223, LOC100244602, LOC100246192*), mitochondrial transcription

362   (*LOC100242048, LOC100245527, LOC100853897, LOC100264034*) and other functions (Fig 6,

363   Table S10).

364

365   **Discussion**

366   In grapevine, very few studies have been focused on the berry enlargement through the

367   identifications of genetic variation, quantitative trait loci (QTL) and candidate genes (Doligez et

368   al., 2002; Mejía et al., 2007). So far, there are very limited materials available for exclusively

369   dissecting the regulatory mechanisms of berry development, especially in bud sport mutation.

370   Doligez et al. detected QTLs related to berry weight using F1 progenies from a cross between

371   two partially seedless genotypes. Claudia et al discovered a number of informative and

372   transferable SNP and InDel markers, as well as 68 candidate genes linked with berry size using

373   'Ruby' and 'Sultanina' cultivars (Muñoz-Espinoza et al., 2020). Wang et al underlined candidate

374   genes for berry related traits using 'Moldova' (*V. labruscana × V. vinifera*) and 'Ruiduxiangyu'

375   cultivars (Wang et al., 2022). These varieties are seedless with enlarged berries. In this study, the

376   seedless cultivars 'Jinzao Wuhe' and 'Dawuhebai' are ideal objects with enlarged berries due to

377   bud sport mutation that derived from 'Himord Seedless' and 'Thompson Seedless' cultivars.

378   Especially in 'Jinzao Wuhe', the berries were enlarged nearly twice bigger than its maternal

379   plant, making it an excellent breeding material for studying the regulatory mechanisms

380   underlying berry enlargement.

381   WGRS is a powerful strategy for genomic variation detection, which could provide direct

382   evidence for phenotypic alterations in plants mainly including SNPs and InDels (Horton et al.,

383   2012). Moreover, structural variations (SVs) could be also identified by using Illumina paired

384   end reads and offer clues to the molecular basis of bud sport mutations. In apples, for instance,

385   abundant SNPs and InDels were discovered in various bud mutants and other variants (Lee et al.,

386   2016; Wang et al., 2020). In grapevines, WGRS of 'Summer Black' and its bud mutant 'SBBM'

387   identified 635 genes carrying differential SNPs and InDels associated with the early-ripening

388   trait (Xu et al., 2016). 'Kyoho', 'Zana' and their bud mutants 'Fengzao' and '90-1' were

389   analyzed separately by Pei et al. to identify variations that could explain the altered phenotype

390   between bud sport mutations and their maternal plants in terms of early maturing trait(Pei et al.,

391   2021). Herein, WGRS of 'Himord Seedless' and 'Thompson Seedless' and their bud mutants

392   'Jinzao Wuhe' and 'Dawuhebai' resulted in the identification of 1,022 genes carrying differential

393   genetic variants. Combined with the GO enrichment results, in this case, symbiotic

394   microorganisms and plant-type hypersentive response might be the reason or consequence for

395   berry enlargement caused by bud mutation in grapevines (Marrano et al., 2018).

396   Transcriptome sequencing and WGCNA are also practical methods for characterizing and

397   quantifying the gene expression and exploring the potential genes and mechanisms of agronomic

398 traits. RNA-Seq of bud mutants 'Tiangong Moyu', 'Fengzao' and their maternal 'Summer
399 Black', 'Kyoho', respectively, at different stages of berry development led to uncovering
400 potential causal genes regulating ripening period (Wei et al., 2020). With the aid of examining
401 gene expression profiles, Theine et al. discovered regulatory genes that either delay or promote
402 berry development at the véraison stage of 'Pinot Noir' and 'Pinot Noir Precoce' grapes
403 (Adetunji et al., 2019). Muñoz-Espinoza et al. combined genetic variation detection and
404 transcriptomic approaches to find functional genes involved in berry size and prematurity and
405 develop molecular markers based on SNPs, InDels and SVs, demonstrating the practicability and
406 applicability of RNA-seq (Muñoz-Espinoza et al., 2020). Herein, we discovered 1,149 DEGs
407 associated with berry size based on WGCNA and GO functional categories between 'Jinzao
408 Wuhe' and 'Himord Seedless'. These results indicated that cell wall modification, secondary
409 metabolic processes, stress responses and cell death regulation might play significant roles
410 during berry enlargement (Sweetman et al., 2012).
411 The combination of high-through sequencing and QTL mapping is a practical strategy for
412 screening candidate genes for important agronomic traits in plants. In a very recent study, 772
413 candidate genes associated with berry size in table grapes have been identified by QTL approach
414 (Wang et al., 2022). We compared the results with 1,022 genes harboring SNPs, InDels and SVs
415 and 2,984 DEGs identified in this study by WGRS and RNA-seq approaches. As a result, 81
416 genes were found in the reported QTL regions including 'RING-type E3 ubiquitin transferase'
417 (*LOC100266002, LOC100250551*), 'Probable xyloglucan endotransglucosylase/hydrolase
418 protein 23' (*LOC100266747, LOC100251303, LOC100246166, LOC100263324,*
419 *LOC100241056)*, 'Expansin' *(LOC100261426, LOC100253046*), 'Auxin-induced protein
420 PCNT115' (*LOC100252909*), 'Auxin efflux carrier component' (*LOC100256460*), 'beta-1,3-
421 galactosyltransferase' (*LOC100246152*), 'Ethylene-responsive transcription factor'
422 (*LOC100259767*), 'LRR receptor-like serine/threonine-protein kinase' (*LOC100241613,*
423 *LOC100244275, LOC100244999, LOC100262226*), 'Heat shock protein' (*LOC100264870,*
424 *LOC100252916, LOC104879973*), 'Cytochrome P450' (*LOC100251316*), 'MATE efflux family
425 protein' (*LOC100242781, LOC100263305*) and several genes involved in 'Cell wall
426 modifications' (*LOC100267301, LOC100252982, LOC100242036, LOC100252799*) (Table
427 S11). These genes offered imported clues for understanding the mechanisms of berry
428 development which had been reported to be functionally related to berry size in grapes or other
429 plants (Xia et al., 2013; Doligez et al., 2013; de Jong et al., 2015; Muñoz-Espinoza et al., 2016,
430 2020).
431 Pectin and cellulose are key substances involved in the mechanical strength of the primary cell
432 wall and are important to the physical properties of plant cell. The degradation of pectin during
433 ripening seems to be responsible for tissue softening, as reported in several fruit plants including
434 kiwifruit (*Actinidia deliciosa*), lemon (*Citrus limon*), apple (*Malus pumila Mill*) and grapevines
435 (*Vitis vinifera L.*). Moreover, previous studies demonstrated that the cellulose-hemicellulose
436 networks and related genes could regulate cell wall modifications and cellulase degradation,
437 leading to decreased rigidity in grape berry during ripening (Fasoli et al., 2016). In this study, PE

438   and glycosyl hydrolase may catalyze the hydrolysis of pectin, resulting in soft berries during the
439   expansion stage of berry development. In rice, UGT has been demonstrated to regulate grain size
440   by modulating cell proliferation and expansion (Dong et al., 2020). The expression of UGT had
441   been shown to gradually decrease with fruit ripening, regulating the early stages of berry
442   development. In our results, the expression levels of two 'UDP-glycosyltransferase-related
443   genes' (*LOC100266067* and *LOC100254242*) significantly varied between MUT and WT. This
444   is consistent with the reported expression trend, suggesting their potential roles in regulating
445   berry size through cell proliferation or differentiation. In addition, 'arabinogalactan protein
446   (AGP)' is highly complex cell wall glycoprotein that plays a role in cell expansion and division
447   (Leszczuk et al., 2020). As a result, AGP (*LOC100252982*) was detected to have higher
448   expression in MUT than in WT, with specific accumulation in the second stage of berry
449   development, implying its regulatory role during berry expansion. However, these genes need to
450   be further investigated and functionally validation.
451   The CYP superfamily genes were also reported to positively associated with fruit size (Qi et al.,
452   2017). Four CYP genes (*LOC100244431*, *LOC100254242*, *LOC100258062*, *LOC100266067*)
453   were identified in this study, of which *LOC100258062* and *LOC100266067* were upregulated at
454   berry expansion stage in MUT, implying their functional roles in enlarging berry size in MUT.
455   Similar phenomenon was also discovered in Arabidopsis that overexpression of *AtCYP78A9*
456   resulting in enlarged fruit organs (Ito & Meyerowitz, 2000) In addition, auxin was regulated by
457   NAC, which thereby triggered the expression of ARFs (Tello et al., 2015; Li et al., 2022).
458   Ethylene coupled with increased ACO transcription could produce a modest peak during berry
459   ripening and promote larger berry diameter. Simultaneously, zinc finger protein(Zhao et al.,
460   2021), ABC transporter C family member , RING-type E3 ubiquitin transferase (Xia et al.,
461   2013), multi antimicrobial extrusion (MATE) (Suzuki et al., 2015) mitochondrial transcription,
462   leucine-rich repeat receptor-like protein kinase (Wang et al., 2019), F-box kelch-repeat protein
463   (Li et al., 2011) and protein kinase (Ye et al., 2017) resulted in the enlargement for berries. This
464   result is in agreement with genes related to fruit enlargement between pears, peaches, and
465   strawberries(Pei et al., 2020). However, limited results supported simultaneously by evidence at
466   genomic and transcriptomic levels suggested that bud sport mutation might be more complex
467   than previously assumed.
468
Conclusions
470   In conclusion, combined with WGRS, RNA-seq approaches and integrated bioinformatic
471   analysis, key genes including pectin esterase, cellulase A, cytochromes P450, UGT, zinc finger
472   protein, ARF, NAC TF, protein kinase were identified in 'Himord Seedless' and its bud sport
473   mutant 'Jinzao Wuhe'. These results could partially explain its phenotypic changes, especially
474   for berry enlargement in bud sport 'Jinzao Wuhe'. However, DEGs during berry development
475   and DNA variations between bud sport and its maternal cultivar could not agree well with each
476   other, suggesting the complexity of mechanisms underlying bud mutation. Our findings offer

477 new clues for a better understanding of bud sport in grapevines and provide useful information
478 for table grape improvement through molecular breeding.
479

Acknowledgements
481 We would like to thank Dr. Ming Li at Zhengzhou Fruit Research Institute, Chinese Academy of
482 Agricultural Sciences, for helping to provide materials of 'Thompson Seedless' and
483 'Dawuhebai'.
484

Funding
486 This work was supported by the China agriculture research system of MOF and MARA (CARS-
487 29) and the Innovation Research and Experiment Program for Youth Scholar, Tianjin Academy
488 of Agricultural Sciences (2021002).
489

References
491 Adetunji MO, Lamont SJ, Abasht B, Schmidt CJ. 2019. Variant analysis pipeline for accurate

492   detection of genomic variants from transcriptome sequencing data. *PloS One* 14:e0216838.

493   DOI: 10.1371/journal.pone.0216838.

494 Azuma A, Kobayashi S, Goto-Yamamoto N, Shiraishi M, Mitani N, Yakushiji H, Koshita Y. 2009.

495   Color recovery in berries of grape (Vitis vinifera L.) "Benitaka", a bud sport of "Italia", is

496   caused by a novel allele at the VvmybA1 locus. *Plant Science: An International Journal*

497   *of Experimental Plant Biology* 176:470–478. DOI: 10.1016/j.plantsci.2008.12.015.

498 Baiano A, Terracone C. 2012. Effects of bud load on quality of Beogradska besemena and

499   Thompson seedless table grapes and cultivar differentiation based on chemometrics of

500   analytical indices. *Journal of the Science of Food and Agriculture* 92:645–653. DOI:

501   10.1002/jsfa.4625.

502 Breton C, Fournel-Gigleux S, Palcic MM. 2012. Recent structures, evolution and mechanisms of

503   glycosyltransferases. *Current Opinion in Structural Biology* 22:540–549. DOI:

504   10.1016/j.sbi.2012.06.007.

505 Chakrabarti M, Zhang N, Sauvage C, Muños S, Blanca J, Cañizares J, Diez MJ, Schneider R,

506   Mazourek M, McClead J, Causse M, van der Knaap E. 2013. A cytochrome P450 regulates

507   a domestication trait in cultivated tomato. *Proceedings of the National Academy of*

508   *Sciences of the United States of America* 110:17125–17130. DOI:

509   10.1073/pnas.1307313110.

510   Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.

511   *Bioinformatics (Oxford, England)* 34:i884–i890. DOI: 10.1093/bioinformatics/bty560.

512   Chervin C, El-Kereamy A, Roustan J-P, Latché A, Lamon J, Bouzayen M. 2004. Ethylene seems

513   required for the berry development and ripening in grape, a non-climacteric fruit. *Plant*

514   *Science* 167:1301–1305. DOI: 10.1016/j.plantsci.2004.06.026.

515   Chialva C, Eichler E, Grissi C, Muñoz C, Gomez-Talquenca S, Martínez-Zapater JM, Lijavetzky

516   D. 2016. Expression of grapevine AINTEGUMENTA-like genes is associated with

517   variation in ovary and berry size. *Plant Molecular Biology* 91:67–80. DOI:

518   10.1007/s11103-016-0443-1.

519   Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012.

520   A program for annotating and predicting the effects of single nucleotide polymorphisms,

521   SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*

522   6:80–92. DOI: 10.4161/fly.19695.

523   Da-Long Guo, Guo-Hai Zhang. 2015. A NEW EARLY-RIPENING GRAPE CULTIVAR -

524   "FENGZAO." *Acta Horticulturae*:153–156. DOI: 10.17660/ActaHortic.2015.1082.20.

525   Danecek P, McCarthy SA. 2017. BCFtools/csq: haplotype-aware variant consequences.

526   *Bioinformatics (Oxford, England)* 33:2037–2039. DOI: 10.1093/bioinformatics/btx100.

527   De La Torre-Ruiz N, Ruiz-Valdiviezo VM, Rincón-Molina CI, Rodríguez-Mendiola M, Arias-

528   Castro C, Gutiérrez-Miceli FA, Palomeque-Dominguez H, Rincón-Rosales R. 2016. Effect

529   of plant growth-promoting bacteria on the growth and fructan production of Agave

530   americana L. *Brazilian Journal of Microbiology: [publication of the Brazilian Society for*

531   *Microbiology]* 47:587–596. DOI: 10.1016/j.bjm.2016.04.010.

532 Doligez A, Bertrand Y, Farnos M, Grolier M, Romieu C, Esnault F, Dias S, Berger G, François P,

533   Pons T, Ortigosa P, Roux C, Houel C, Laucou V, Bacilieri R, Péros J-P, This P. 2013. New

534   stable QTLs for berry weight do not colocalize with QTLs for seed traits in cultivated

535   grapevine (Vitis vinifera L.). *BMC plant biology* 13:217. DOI: 10.1186/1471-2229-13-217.

536 Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith P, Edwards J, This P. 2002.

537   Genetic mapping of grapevine (Vitis vinifera L.) applied to the detection of QTLs for

538   seedlessness and berry weight. *TAG. Theoretical and applied genetics. Theoretische und*

539   *angewandte Genetik* 105:780–795. DOI: 10.1007/s00122-002-0951-z.

540 Dong N-Q, Sun Y, Guo T, Shi C-L, Zhang Y-M, Kan Y, Xiang Y-H, Zhang H, Yang Y-B, Li Y-

541   C, Zhao H-Y, Yu H-X, Lu Z-Q, Wang Y, Ye W-W, Shan J-X, Lin H-X. 2020. UDP-

542   glucosyltransferase regulates grain size and abiotic stress tolerance associated with

543   metabolic flux redirection in rice. *Nature Communications* 11:2629. DOI: 10.1038/s41467-

544   020-16403-5.

545 El-kereamy A, Chervin C, Roustan J-P, Cheynier V, Souquet J-M, Moutounet M, Raynal J, Ford

546   C, Latché A, Pech J-C, Bouzayen M. 2003. Exogenous ethylene stimulates the long-term

547   expression of genes related to anthocyanin biosynthesis in grape berries. *Physiologia*

548   *Plantarum* 119. DOI: 10.1034/j.1399-3054.2003.00165.x.

549 Fang W, Wang Z, Cui R, Li J, Li Y. 2012. Maternal control of seed size by EOD3/CYP78A6 in

550   Arabidopsis thaliana. *The Plant Journal: For Cell and Molecular Biology* 70:929–939.

551   DOI: 10.1111/j.1365-313X.2012.04907.x.

552   Fasoli M, Dell'Anna R, Dal Santo S, Balestrini R, Sanson A, Pezzotti M, Monti F, Zenoni S. 2016.

553   Pectins, Hemicelluloses and Celluloses Show Specific Dynamics in the Internal and

554   External Surfaces of Grape Berry Skin During Ripening. *Plant and Cell Physiology*

555   57:1332–1349. DOI: 10.1093/pcp/pcw080.

556   Foster TM, Aranzana MJ. 2018. Attention sports fans! The far-reaching contributions of bud sport

557   mutants to horticulture and plant biology. *Horticulture Research* 5:44. DOI:

558   10.1038/s41438-018-0062-x.

559   Horton MW, Hancock AM, Huang YS, Toomajian C, Atwell S, Auton A, Muliyati NW, Platt A,

560   Sperone FG, Vilhjálmsson BJ, Nordborg M, Borevitz JO, Bergelson J. 2012. Genome-wide

561   patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the

562   RegMap panel. *Nature Genetics* 44:212–216. DOI: 10.1038/ng.1042.

563   Ito T, Meyerowitz EM. 2000. Overexpression of a Gene Encoding a Cytochrome P450, CYP78A9,

564   Induces Large and Seedless Fruit in Arabidopsis. *The Plant Cell* 12:1541–1550. DOI:

565   10.1105/tpc.12.9.1541.

566   de Jong M, Wolters-Arts M, Schimmel BCJ, Stultiens CLM, de Groot PFM, Powers SJ, Tikunov

567   YM, Bovy AG, Mariani C, Vriezen WH, Rieu I. 2015. Solanum lycopersicum AUXIN

568   RESPONSE FACTOR 9 regulates cell division activity during early tomato fruit

569   development. *Journal of Experimental Botany* 66:3405–3416. DOI: 10.1093/jxb/erv152.

570   Kiedrowski S, Kawalleck P, Hahlbrock K, Somssich IE, Dangl JL. 1992. Rapid activation of a

571   novel plant defense gene is strictly dependent on the Arabidopsis RPM1 disease resistance

572   locus. *The EMBO journal* 11:4677–4684. DOI: 10.1002/j.1460-2075.1992.tb05572.x.

573   Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory

574   requirements. *Nature Methods* 12:357–360. DOI: 10.1038/nmeth.3317.

575 Krizek BA. 1999. Ectopic expression of AINTEGUMENTA in Arabidopsis plants results in

576   increased growth of floral organs. *Developmental Genetics* 25:224–236. DOI:

577   10.1002/(SICI)1520-6408(1999)25:3<224::AID-DVG5>3.0.CO;2-Y.

578 Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis.

579   *BMC bioinformatics* 9:559. DOI: 10.1186/1471-2105-9-559.

580 Layer RM, Chiang C, Quinlan AR, Hall IM. 2014. LUMPY: a probabilistic framework for

581   structural variant discovery. *Genome Biology* 15:R84. DOI: 10.1186/gb-2014-15-6-r84.

582 Lee HS, Kim GH, Kwon SI, Kim JH, Kwon YS, Choi C. 2016. Analysis of "Fuji" apple somatic

583   variants from next-generation sequencing. *Genetics and molecular research: GMR* 15.

584   DOI: 10.4238/gmr.15038185.

585 Leng F, Ye Y, Zhu X, Zhang Y, Zhang Z, Shi J, Shen N, Jia H, Wang L. 2021. Comparative

586   transcriptomic analysis between 'Summer Black' and its bud sport 'Nantaihutezao' during

587   developmental stages. *Planta* 253:23. DOI: 10.1007/s00425-020-03543-7.

588 Leszczuk A, Kalaitzis P, Blazakis KN, Zdunek A. 2020. The role of arabinogalactan proteins

589   (AGPs) in fruit ripening—a review. *Horticulture Research* 7:1–12. DOI: 10.1038/s41438-

590   020-00397-8.

591 Li M, Chen R, Gu H, Cheng D, Guo X, Shi C, Li L, Xu G, Gu S, Wu Z, Chen J. 2021. Grape Small

592   Auxin Upregulated RNA (SAUR) 041 Is a Candidate Regulator of Berry Size in Grape.

593   *International Journal of Molecular Sciences* 22:11818. DOI: 10.3390/ijms222111818.

594 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.

595   *Bioinformatics (Oxford, England)* 25:1754–1760. DOI: 10.1093/bioinformatics/btp324.

596 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,

597   1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map

598   format and SAMtools. *Bioinformatics (Oxford, England)* 25:2078–2079. DOI:

599   10.1093/bioinformatics/btp352.

600 Li M, Tang D, Wang K, Wu X, Lu L, Yu H, Gu M, Yan C, Cheng Z. 2011. Mutations in the F-

601   box gene LARGER PANICLE improve the panicle architecture and enhance the grain

602   yield in rice. *Plant Biotechnology Journal* 9:1002–1013. DOI: 10.1111/j.1467-

603   7652.2011.00610.x.

604 Li X, Zhang X, Shi T, Chen M, Jia C, Wang J, Hou Z, Han J, Bian S. 2022. Identification of ARF

605   family in blueberry and its potential involvement of fruit development and pH stress

606   response. *BMC genomics* 23:329. DOI: 10.1186/s12864-022-08556-y.

607 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for

608   assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* 30:923–

609   930. DOI: 10.1093/bioinformatics/btt656.

610 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for

611   RNA-seq data with DESeq2. *Genome Biology* 15:550. DOI: 10.1186/s13059-014-0550-8.

612 Ma M, Wang Q, Li Z, Cheng H, Li Z, Liu X, Song W, Appels R, Zhao H. 2015. Expression of

613   TaCYP78A3, a gene encoding cytochrome P450 CYP78A3 protein in wheat (Triticum

614   aestivum L.), affects seed size. *The Plant Journal: For Cell and Molecular Biology*

615   83:312–325. DOI: 10.1111/tpj.12896.

616 Malladi A, Hirst PM. 2010. Increase in fruit size of a spontaneous mutant of "Gala" apple (Malus

617   x domestica Borkh.) is facilitated by altered cell production and enhanced cell size. *Journal*

618   *of Experimental Botany* 61:3003–3013. DOI: 10.1093/jxb/erq134.

619 Marrano A, Micheletti D, Lorenzi S, Neale D, Grando MS. 2018. Genomic signatures of different

620   adaptations to environmental stimuli between wild and cultivated Vitis vinifera L.

621   *Horticulture Research* 5:34. DOI: 10.1038/s41438-018-0041-2.

622   Martins V, Garcia A, Costa C, Sottomayor M, Gerós H. 2018. Calcium- and hormone-driven

623   regulation of secondary metabolism and cell wall enzymes in grape berry cells. *Journal of*

624   *Plant Physiology* 231:57–67. DOI: 10.1016/j.jplph.2018.08.011.

625   McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler

626   D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: A MapReduce

627   framework for analyzing next-generation DNA sequencing data. *Genome Research*

628   20:1297–1303. DOI: 10.1101/gr.107524.110.

629   Mejía N, Gebauer M, Muñoz L, Hewstone N, Muñoz C, Hinrichsen P. 2007. Identification of

630   QTLs for Seedlessness, Berry Size, and Ripening Date in a Seedless x Seedless Table

631   Grape Progeny. *American Journal of Enology and Viticulture* 58:499–507.

632   Migicovsky Z, Yeats TH, Watts S, Song J, Forney CF, Burgher-MacLellan K, Somers DJ, Gong

633   Y, Zhang Z, Vrebalov J, van Velzen R, Giovannoni JG, Rose JKC, Myles S. 2021. Apple

634   Ripening Is Controlled by a NAC Transcription Factor. *Frontiers in Genetics* 12:671300.

635   DOI: 10.3389/fgene.2021.671300.

636   Mitchell HJ, Hall JL, Barber MS. 1994. Elicitor-Induced Cinnamyl Alcohol Dehydrogenase

637   Activity in Lignifying Wheat (Triticum aestivum L.) Leaves. *Plant Physiology* 104:551–

638   556. DOI: 10.1104/pp.104.2.551.

639   Muñoz-Espinoza C, Di Genova A, Correa J, Silva R, Maass A, González-Agüero M, Orellana A,

640   Hinrichsen P. 2016. Transcriptome profiling of grapevine seedless segregants during berry

641   development reveals candidate genes associated with berry weight. *BMC plant biology*

642   16:104. DOI: 10.1186/s12870-016-0789-1.

643   Muñoz-Espinoza C, Di Genova A, Sánchez A, Correa J, Espinoza A, Meneses C, Maass A,

644   Orellana A, Hinrichsen P. 2020. Identification of SNPs and InDels associated with berry
645   size in table grapes integrating genetic and transcriptomic approaches. *BMC Plant Biology*
646   20:365. DOI: 10.1186/s12870-020-02564-4.

647 Negri AS, Prinsi B, Failla O, Scienza A, Espen L. 2015. Proteomic and metabolic traits of grape
648   exocarp to explain different anthocyanin concentrations of the cultivars. *Frontiers in Plant*
649   *Science* 6:603. DOI: 10.3389/fpls.2015.00603.

650 Nicolas P, Lecourieux D, Gomès E, Delrot S, Lecourieux F. 2013. The grape berry-specific basic
651   helix-loop-helix transcription factor VvCEB1 affects cell size. *Journal of Experimental*
652   *Botany* 64:991–1003. DOI: 10.1093/jxb/ers374.

653 Pan Q-H, Chen F, Zhu B-Q, Ma L-Y, Li L, Li J-M. 2012. Molecular cloning and expression of
654   gene encoding aromatic amino acid decarboxylase in "Vidal blanc" grape berries.
655   *Molecular Biology Reports* 39:4319–4325. DOI: 10.1007/s11033-011-1219-y.

656 Pei M-S, Cao S-H, Wu L, Wang G-M, Xie Z-H, Gu C, Zhang S-L. 2020. Comparative
657   transcriptome analyses of fruit development among pears, peaches, and strawberries
658   provide new insights into single sigmoid patterns. *BMC plant biology* 20:108. DOI:
659   10.1186/s12870-020-2317-6.

660 Pei M-S, Liu H-N, Wei T-L, Yu Y-H, Guo D-L. 2021. Detection and characterization of genome-
661   wide genetic variation associated with the early-ripening phenotype of grape mutants.
662   *Scientia Horticulturae* 285:110195. DOI: 10.1016/j.scienta.2021.110195.

663 Qi X, Liu C, Song L, Li Y, Li M. 2017. PaCYP78A9, a Cytochrome P450, Regulates Fruit Size in
664   Sweet Cherry (Prunus avium L.). *Frontiers in Plant Science* 8:2076. DOI:
665   10.3389/fpls.2017.02076.

666 Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: structural variant

667   discovery by integrated paired-end and split-read analysis. *Bioinformatics (Oxford,*

668   *England)* 28:i333–i339. DOI: 10.1093/bioinformatics/bts378.

669   Scorzal R, May LG, Purnell B, Upchurch B. 1991. Differences in Number and Area of Mesocarp

670   Cells between Small- and Large-fruited Peach Cultivars. *Journal of the American Society*

671   *for Horticultural Science* 116:861–864. DOI: 10.21273/JASHS.116.5.861.

672   Somkuwar RG, Bahetwar A, Khan I, Satisha J, Ramteke SD, Itroutwar P, Bhongale A, Oulkar D.

673   2014. Changes in growth, photosynthetic activities, biochemical parameters and amino

674   acid profile of Thompson Seedless grapes (Vitis vinifera L.). *Journal of Environmental*

675   *Biology* 35:1157–1163.

676   Su L, Bassa C, Audran C, Mila I, Cheniclet C, Chevalier C, Bouzayen M, Roustan J-P, Chervin

677   C. 2014. The auxin Sl-IAA17 transcriptional repressor controls fruit size via the regulation

678   of endoreduplication-related cell expansion. *Plant & Cell Physiology* 55:1969–1976. DOI:

679   10.1093/pcp/pcu124.

680   Suzuki M, Sato Y, Wu S, Kang B-H, McCarty DR. 2015. Conserved Functions of the MATE

681   Transporter BIG EMBRYO1 in Regulation of Lateral Organ Size and Initiation Rate. *The*

682   *Plant Cell* 27:2288–2300. DOI: 10.1105/tpc.15.00290.

683   Sweetman C, Wong DC, Ford CM, Drew DP. 2012. Transcriptome analysis at four developmental

684   stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and

685   coordinated gene expression. *BMC genomics* 13:691. DOI: 10.1186/1471-2164-13-691.

686   Tello J, Torres-Pérez R, Grimplet J, Carbonell-Bejerano P, Martínez-Zapater JM, Ibáñez J. 2015.

687   Polymorphisms and minihaplotypes in the VvNAC26 gene associate with berry size

688   variation in grapevine. *BMC plant biology* 15:253. DOI: 10.1186/s12870-015-0622-2.

689   Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq.

690   *Bioinformatics (Oxford, England)* 25:1105–1111. DOI: 10.1093/bioinformatics/btp120.

691 Tucker MR, Lou H, Aubert MK, Wilkinson LG, Little A, Houston K, Pinto SC, Shirley NJ. 2018.

692   Exploring the Role of Cell Wall-Related Genes and Polysaccharides during Plant

693   Development. *Plants (Basel, Switzerland)* 7:E42. DOI: 10.3390/plants7020042.

694 Wang X, Li Y, Zhang H, Sun G, Zhang W, Qiu L. 2015. Evolution and association analysis of

695   GmCYP78A10 gene with seed size/weight and pod number in soybean. *Molecular Biology*

696   *Reports* 42:489–496. DOI: 10.1007/s11033-014-3792-3.

697 Wang X, Shen F, Gao Y, Wang K, Chen R, Luo J, Yang L, Zhang X, Qiu C, Li W, Wu T, Xu X,

698   Wang Y, Cong P, Han Z, Zhang X. 2020. Application of genome-wide insertion/deletion

699   markers on genetic structure analysis and identity signature of Malus accessions. *BMC*

700   *Plant Biology* 20:540. DOI: 10.1186/s12870-020-02744-2.

701 Wang Q, Sun G, Ren X, Du B, Cheng Y, Wang Y, Li C, Sun D. 2019. Dissecting the Genetic

702   Basis of Grain Size and Weight in Barley (Hordeum vulgare L.) by QTL and Comparative

703   Genetic Analyses. *Frontiers in Plant Science* 10. DOI:

704   https://doi.org/10.3389/fpls.2019.00469

705 Wang Q, Xue X, Li Y, Dong Y, Zhang L, Zhou Q, Deng F, Ma Z, Qiao D, Hu C, Ren Y. 2016. A

706   maize ADP-ribosylation factor ZmArf2 increases organ and seed size by promoting cell

707   expansion in Arabidopsis. *Physiologia Plantarum* 156:97–107. DOI: 10.1111/ppl.12359.

708 Wang H, Yan A, Wang X, Zhang G, Liu Z, Xu H, Sun L. 2022. Identification of QTLs and

709   candidate genes controlling berry size in table grape by integrating QTL and transcriptomic

710   analysis. *Scientia Horticulturae* 305:111403. DOI: 10.1016/j.scienta.2022.111403.

711 Wei L, Cao Y, Cheng J, Xiang J, Shen B, Wu J. 2020. Comparative transcriptome analyses of a

712   table grape 'Summer Black' and its early-ripening mutant 'Tiangong Moyu' identify

713   candidate genes potentially involved in berry development and ripening. *Journal of Plant*

714   *Interactions* 15:213–222. DOI: 10.1080/17429145.2020.1760367.

715 Xia T, Li N, Dumenil J, Li J, Kamenski A, Bevan MW, Gao F, Li Y. 2013. The ubiquitin receptor

716   DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in

717   Arabidopsis. *The Plant Cell* 25:3347–3359. DOI: 10.1105/tpc.113.115063.

718 Xu F, Fang J, Ou S, Gao S, Zhang F, Du L, Xiao Y, Wang H, Sun X, Chu J, Wang G, Chu C.

719   2015. Variations in CYP78A13 coding region influence grain size and yield in rice. *Plant,*

720   *Cell & Environment* 38:800–811. DOI: 10.1111/pce.12452.

721 Xu Y, Gao Z, Tao J, Jiang W, Zhang S, Wang Q, Qu S. 2016. Genome-Wide Detection of SNP

722   and SV Variations to Reveal Early Ripening-Related Genes in Grape. *PLOS ONE*

723   11:e0147749. DOI: 10.1371/journal.pone.0147749.

724 Yamada M, Sato A. 2016. Advances in table grape breeding in Japan. *Breeding Science* 66:34–45.

725   DOI: 10.1270/jsbbs.66.34.

726 Yang J, Xiao Y-Y. 2013. Grape phytochemicals and associated health benefits. *Critical Reviews*

727   *in Food Science and Nutrition* 53:1202–1225. DOI: 10.1080/10408398.2012.692408.

728 Ye Y, Ding Y, Jiang Q, Wang F, Sun J, Zhu C. 2017. The role of receptor-like protein kinases

729   (RLKs) in abiotic stress response in plants. *Plant Cell Reports* 36:235–242. DOI:

730   10.1007/s00299-016-2084-x.

731 Zhan P, Wei X, Xiao Z, Wang X, Ma S, Lin S, Li F, Bu S, Liu Z, Zhu H, Liu G, Zhang G, Wang

732   S. 2021. GW10, a member of P450 subfamily regulates grain size and grain number in rice.

733   *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik* 134:3941–

734   3950. DOI: 10.1007/s00122-021-03939-3.

735 Zhao F, Zhang J, Weng L, Li M, Wang Q, Xiao H. 2021. Fruit size control by a zinc finger protein

736   regulating pericarp cell size in tomato. *Molecular Horticulture* 1:6. DOI: 10.1186/s43897-

737   021-00009-6.

738

The statistics of whole-genome resequencing data

1   **Table 1** The statistics of whole-genome resequencing data

Sample	Raw reads	Clean reads	Q20	Q30	Mapping rates
Thompson Seedless	54,003,664	53,434,316	95.52%	88.93%	93.57%
Dawuhebai	75,630,160	75,128,802	93.57%	90.50%	95.87%
Himord Seedless	91,982,400	91,164,032	96.30%	90.52%	91.91%
Jinzao Wuhe	83,784,034	83,013,230	93.70%	90.19%	95.39%

Figure 1

The comparison of 'Himord Seedless' and its bud sport 'Jinzao wuhe'.

Comparison of longitudinal diameter (B), transverse diameter (C), volume (D) and weight (E) of 'Himord Seedless' (black line) and 'Jinzao wuhe' (red line) at different developmental stages. Standard deviations (SDs) were calculated with 30 biological replicates for each timepoint. The error bars represented 1 SD.
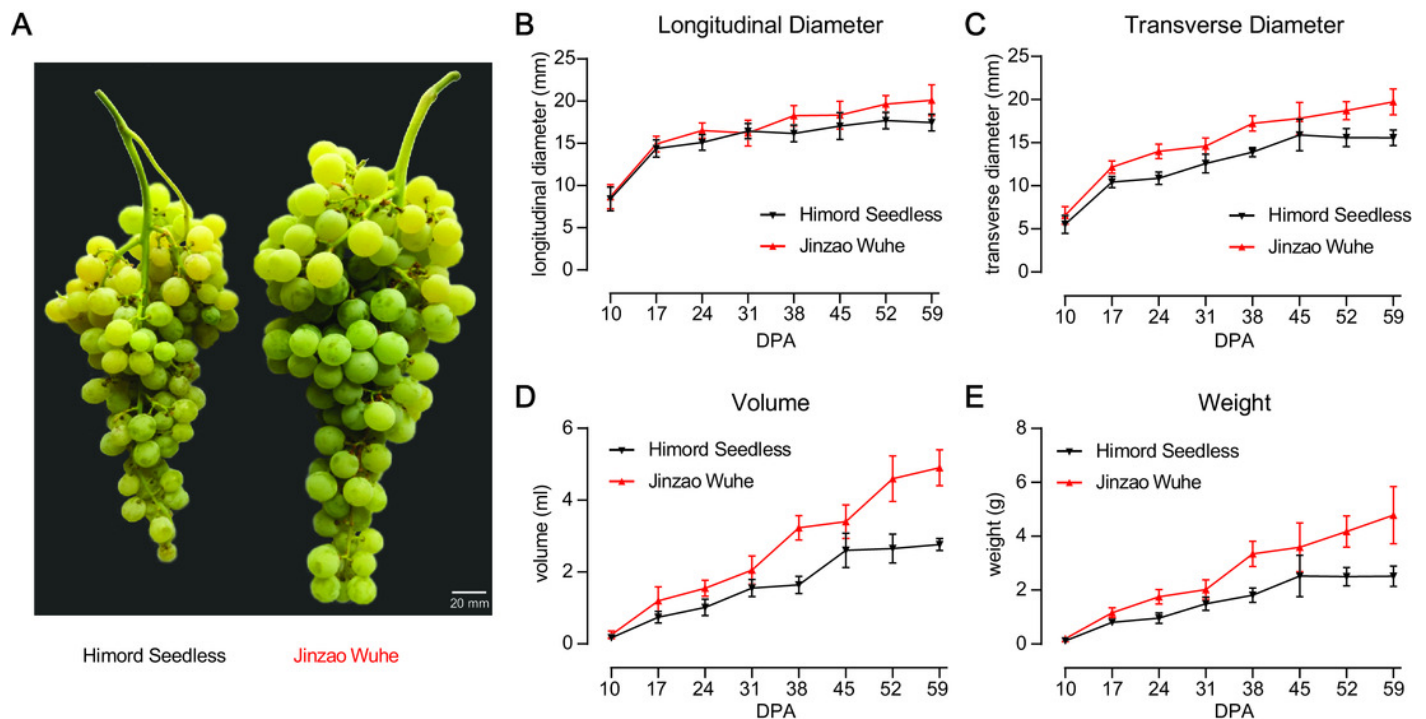
Figure 2

Annotation and distribution of SNPs, InDels and SVs of 'Thompson Seedless' and its bud sports 'Dawuhebai', 'Himord Seedless' and its bud sports 'Jinzao Wuhe'
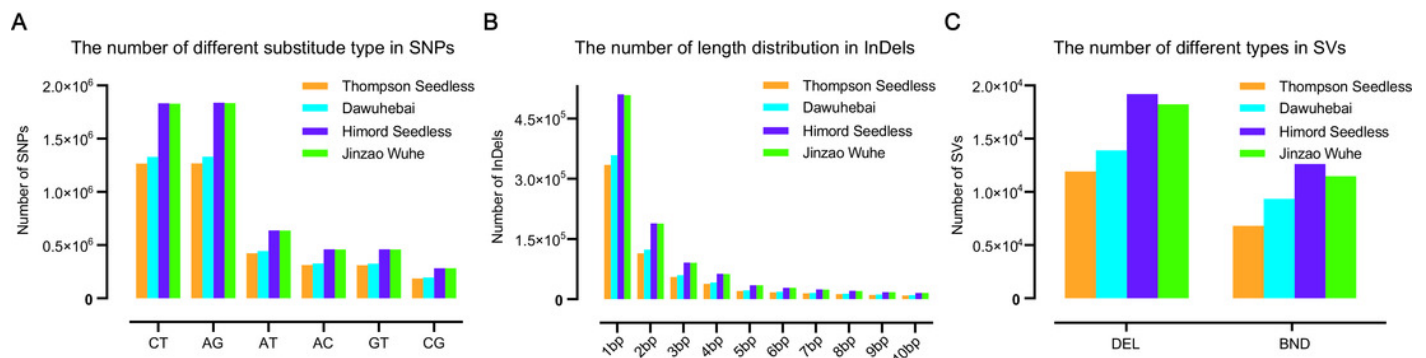
Figure 3

The statistics of SNPs, InDels and SVs and corresponding genes.

(A), (B) and (C) Venn diagrams represent the numbers of unique and shared SNPs, InDels and SVs, respectively. (a) 'Thompson Seedless'; (b) 'Dawuhebai'; (c) 'Himord Seedless'; (d) 'Jinzao Wuhe'. (D) The circos plots of SNPs, InDels and SVs distribution on the nineteen *V.vivfera* chromosomes. (a). Chromosome; (b). GC content; (c). Gene density; (d). SNP density; (e). InDel density; (f). SV density. (E) The GO annotation of potential causal genes harbored with determining SNPs, InDels and SVs
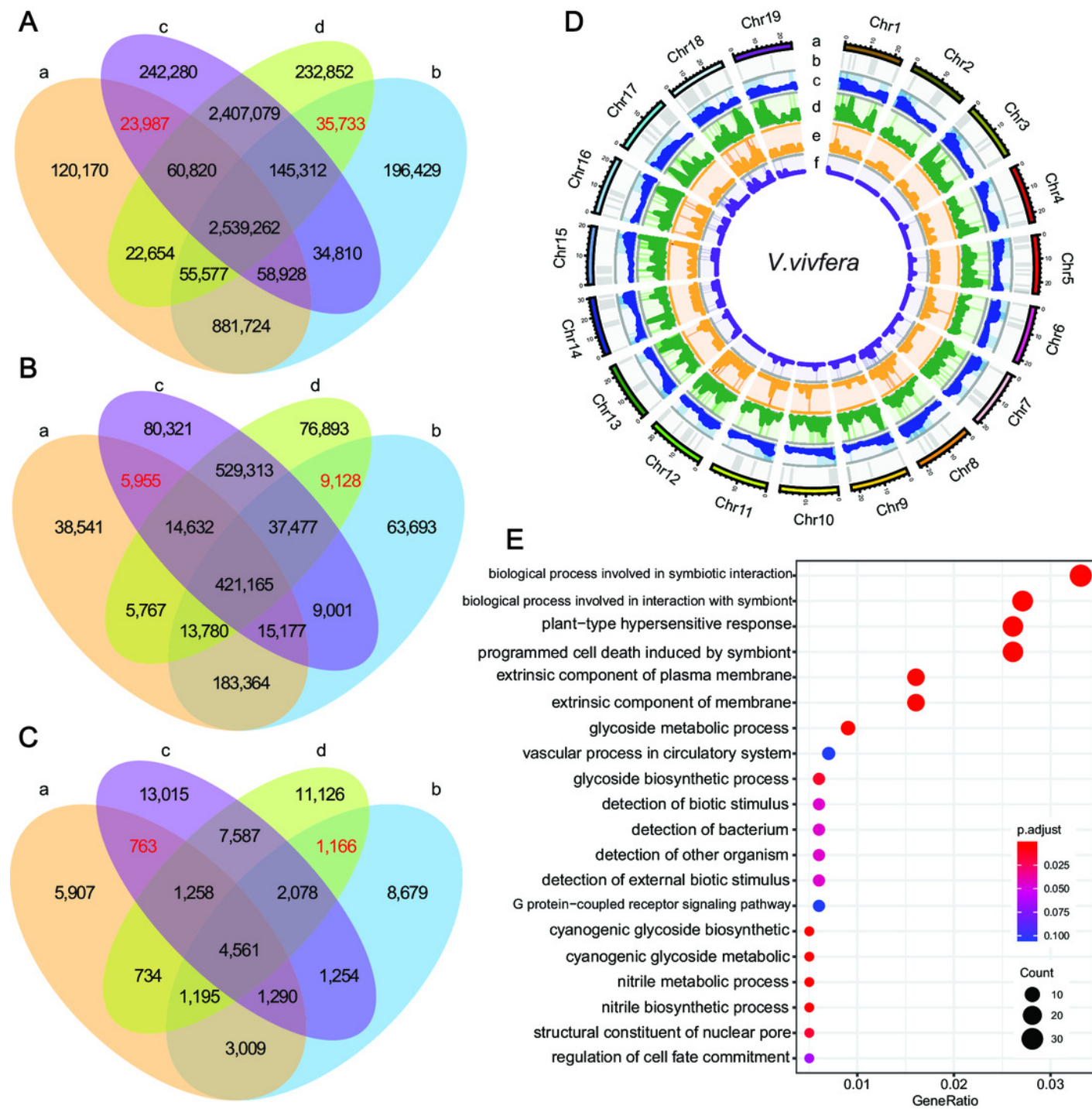
Figure 4

Analyses and statistics of DEGs and GO enrichment.

Analyses and statistics of DEGs and GO enrichment. (A) Principal component analysis (PCA) of transcriptome data. (B), (C), (E), (F) Volcano plots for expressed genes at different developmental stages. Red dots represented the upregulated genes while blue dots represented the down-regulated genes." (D) Venn diagram of DEGs at four different stages. (G), (H), (I) and (J) GO enrichment results of DEGs at four development stages. WT: 'Himord Seedless'; MUT: 'Jinzao Wuhe'. Stage 1-4: full-flowering stage, fruit enlargement stage (14 days post-anthesis), véraison stage (35 days post-anthesis) and maturity stage (56 days post-anthesis). The 'Himord Seedless' in the stage 1-4 were the control group and 'Jinzao Wuhe' in the stage 1-4 were the experiment group.
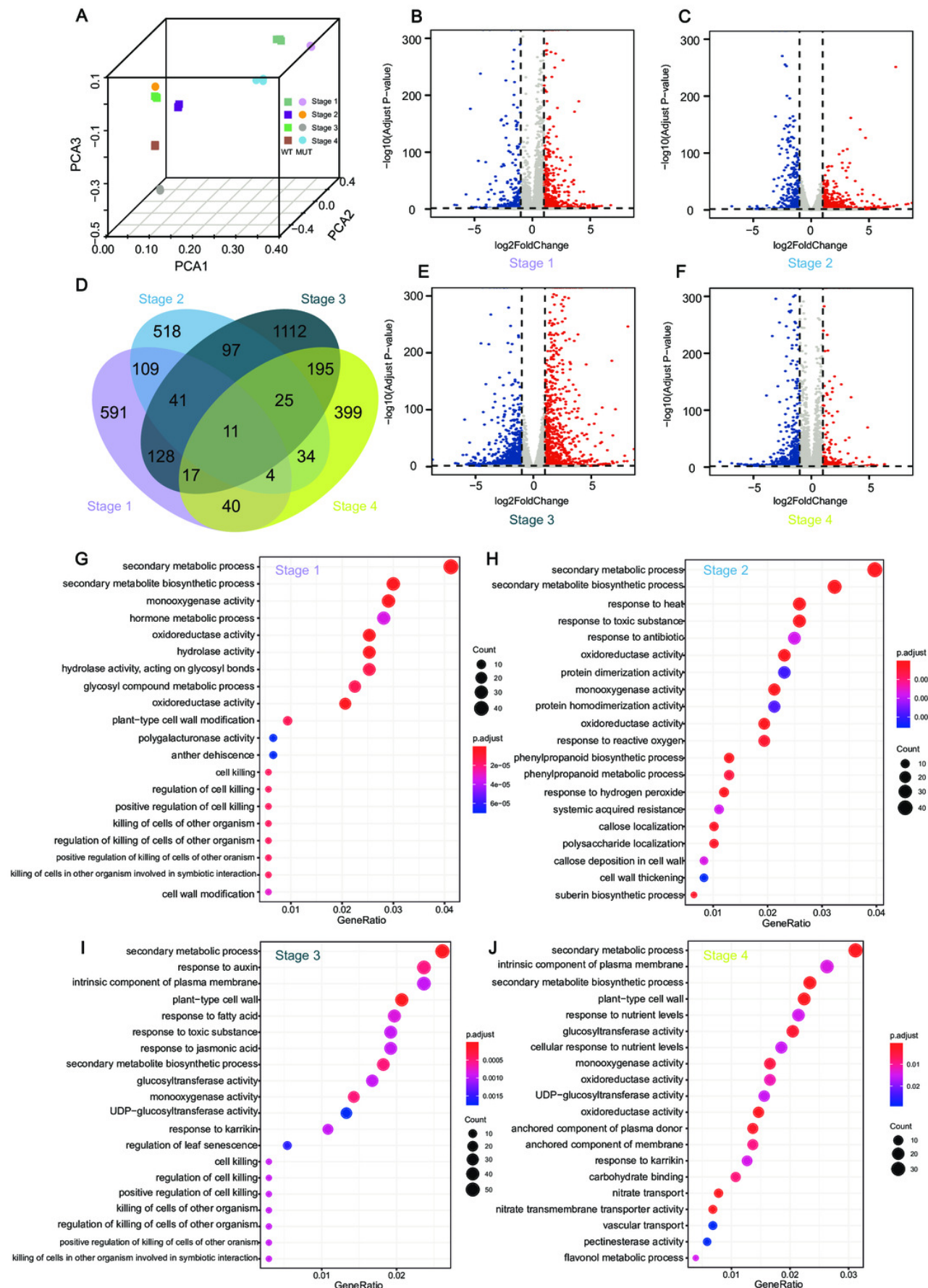
Figure 5

Weighted gene co-expression network analysis of berry development transcriptomes.

(A) Dendrogram plot of co-expression modules. This plot shows 18 modules and different colors represent different modules. (B) Module-trait relationship and corresponding p values. LD: longitudinal diameter; TD: transverse diameter
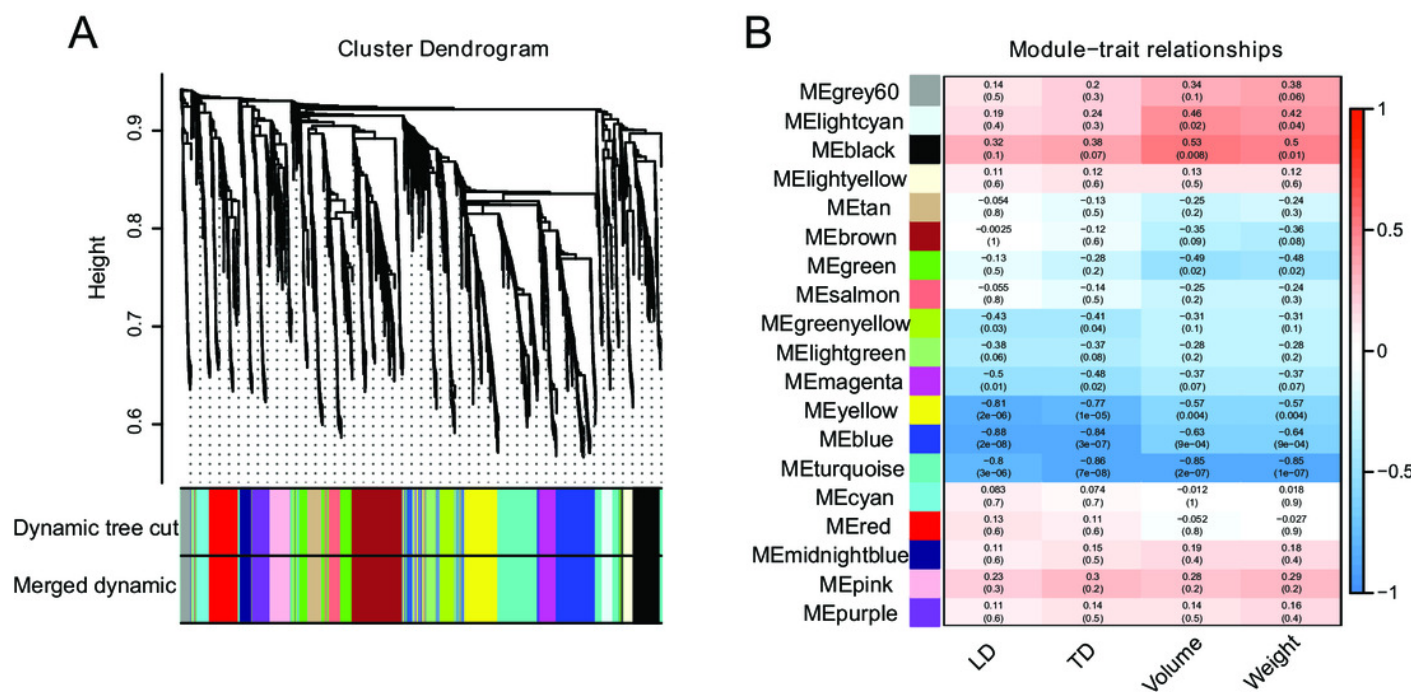
Figure 6

Analyses and statistics of DEGs carrying specified SNPs, InDels and SVs.

(A) The intersection between DEGs and genes containing SNPs, InDels and SVs. (B)

Expression profiles of DEGs harboring specified SNPs, InDels and SV