SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

Received: 18 February 2019
Accepted: 21 May 2019
Published online: 13 June 2019

Transcriptome profiles of three Muscat table grape cultivars to dissect the mechanism of terpene biosynthesis

Lei Sun[1], Baoqing Zhu[2], Xuanyin Zhang[1], Guojun Zhang[1], Ailing Yan[1], Huiling Wang[1], Xiaoyue Wang[1] & Haiying Xu[3,4]

Vitis vinifera **is widely grown worldwide for making wine and for use as table grapes. Of the existing cultivars, some have a floral and fruity flavour, referred to as a Muscat flavour. It is well-documented that this flavour originates from a series of terpene compounds, but the mechanism of terpene content differences among the Muscat-type cultivars remains unclear. Transcript and terpene metabolite profiles were integrated to elucidate the molecular mechanism of this phenomenon. In this research, three genotypes with different aromatic strengths were investigated by RNA sequencing. A total of 27 fruit samples from three biological replicates were sequenced on Illumina HiSeq2000 at three stages, corresponding to the veraison; berries had intermediate Brix value and were harvest-ripe. After quality assessment and data clearance, a total of 254.18 Gb of data with more than 97% Q20 bases were obtained, approximately 9.41 Gb data were generated per sample. These results will provide a valuable dataset for the discovery of the mechanism of terpene biosynthesis.**

Background & Summary

The trait of aroma is one of the most important parameters for the quality of grapes and is the main concern when consumers buy grape products. For genetic improvement research and breeding, the biosynthesis mechanism of aromatic compounds and their regulation has attracted much attention. Terpenes are the typical aromatic compounds in Muscat grapes, and they belong to the second metabolites[1–4]; they have a low olfactory threshold and can be easily precepted by humans. The terpenes mainly exist in the pericarp and in the flesh of some cultivars[5], with their content being affected by the genotype[6,7], developmental stage[8,9], environment and management of the grape[10–13]. Terpenes have two forms: the free form, which directly leads to the aromatic flavour, and the glycoside bound form, in which the potential aromatic compounds transfer to the free form by hydrolysis[14–16].

Biologically, the biosynthesis of terpene compounds in plants are synthesized by two pathways, the methyl-erythritol-4-phosphate pathway (DXP/MEP) in the plastid and the mevalonate pathway (MVA) in the cytoplasm[17], with terpenes located in the mesocarp and pericarp[18]. Starting from pyruvic acid and 3-phosphate glyceraldehyde, by 1-deoxy-D-xylulose-5-phosphate synthase (DXS), which is the entrance enzyme in the MEP pathway, the two compounds were changed into 1-deoxy-D-xyulose-5-phosphate and, then, through six enzymatic reactions, were converted into geranyl-diphosphate (GPP). Geranyl-diphosphate was the substrate for all the terpenes. Then, by a series of terpene synthases, the GPP was synthesized into hemiterpenes (C5), monoterpenes (C10), sesquiterpenes (C15) or diterpenes (C20)[19–22].

The genetic mechanism of Muscat flavour in grapevines has been studied through quantitative trait loci analysis (QTL) in different F1 populations[23,24], and in selfing populations, it has been shown that VvDXS is a structural candidate gene for geraniol, nerol, and linalool concentrations in wine grapes[25]. Battilana reported that single nucleotide polymorphism (SNP) mutations in VvDXS are the main cause of the Muscat flavour. The substitution

[1]Beijing Academy of Forestry and Pomology Sciences, Beijing, 100093, China. [2]College of Biological Sciences and Technology, Beijing Forestry University, Beijing, 100083, China. [3]Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China), Ministry of Agriculture and Rural Affairs, Beijing, 100093, China. [4]Beijing Engineering Research Centre for Deciduous Fruit Trees, Beijing, 100093, China. Correspondence and requests for materials should be addressed to H.X. (email: xuhaiying@baafs.net.cn)

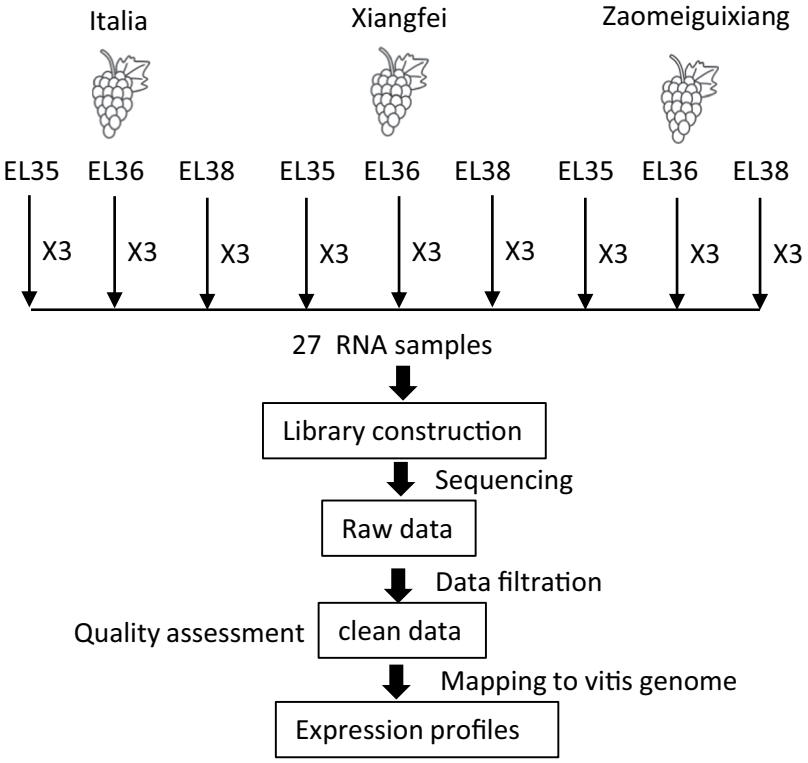

Fig. 1 Flowchart of the experimental design. Berry samples were collected at three developmental stages, and three biological replicates per sample were used for transcriptome sequencing. All raw reads were quality controlled and assessed. Then, the clean data were mapped to the *V. vinifera* reference genome (V2) by Hisat2. Gene expression levels were calculated with RSEM.

of a lysine with an asparagine at position 284 of the VvDXS amino acid sequence affects the monoterpene content of Muscat flavour and neutral cultivars[26].

In Muscat grapes, some cultivars have a very strong flavour, while others have moderate or light flavour. The terpene type and concentration varied among the cultivars. To date, terpene accumulation has been well-documented in some wine grapes. Terpene accumulation in developing Gewurztraminer grapes has been shown to be correlated with an increase in the transcript abundances of early terpenoid pathway enzymes[27]. Some transcription factors involved in controlling terpene biosynthesis have been predicted in the grapevine cultivar Muscat Blanc à Petits Grains through gene co-expression network analysis[28]. A Nudix hydrolase was also found to be a component of a terpene synthase-independent pathway, with cytochrome P450 hydroxylases, epoxide hydrolases and glucosyltransferases genes potentially involved in monoterpene metabolism[29]. However, there are few reports on the table grape[30].

In this study, we present the transcriptome analysis of three genotypes of table grapes. During berry development, 27 samples, in total, were sequenced on the Illumina HiSeq Platform. After quality assessment and data clearance, a total of 254.18 Gb high-quality base pairs with more than 97% Q20 bases were obtained, and an approximately 9.41 Gb per sample. In the aggregate, a total of 776 million reads were yielded, with an average of 31.66 million reads per sample. Furthermore, approximately 76.65% of the total reads were uniquely aligned to the grape genome (V2)[31]. These data will provide useful information for investigating terpene biosynthesis.

Methods

Overview of the experimental design. The berries of three genotypes were collected at three developmental stages. Approximately 300 grape berries were randomly collected for each replicate, with three replicates harvested for each stage. The experimental design and analysis pipeline are shown in Fig. 1.

Materials and methods. *Plant materials.* Three *V. vinifera* cultivars were used for transcript study. 'Xiangfei' was registered by our team and has a strong Muscat flavour and a green to golden skin colour, while 'Italia,' the famous mid-late season table grape cultivar that originated in Italy, has a moderate Muscat flavour. 'Zaomeiguixiang' has a purple-reddish colour and a strong Muscat flavour.

Sampling. The vines were grown in the experimental vineyard at the Beijing Academy of Forestry and Pomology Sciences in China (39°58′N and 116°13′E) under a plastic cover and were trained into a two-wire vertical trellis system with a 2.5-m row space and a 0.75 m plant space. In 2017, berry samples from three vines were harvested at the developmental stages corresponding to EL35, EL36, and EL38[32]. The berry begins to colour and soften at EL 35 (about 5% of the berries started to colour and soften), progresses to the complete veraison with an intermediate

Sample name	Total soluble solids	Titratable acidity(g/l)	pH
X-EL35-1	10.84	4.25	3.11
X-EL35-2	10.80	4.20	3.15
X-EL35-3	10.95	4.26	3.16
X-EL36-1	13.46	4.01	3.53
X-EL36-2	13.30	3.98	3.50
X-EL36-3	13.80	4.05	3.58
X-EL38-1	16.62	3.73	3.75
X-EL38-2	16.40	3.70	3.71
X-EL38-3	16.42	3.68	3.77
Y-EL35-1	5.18	5.15	3.07
Y-EL35-2	5.20	5.20	3.05
Y-EL35-3	5.18	5.22	3.01
Y-EL36-1	7.61	4.85	3.14
Y-EL36-2	7.45	4.80	3.18
Y-EL36-3	7.40	4.79	3.17
Y-EL38-1	14.80	4.51	3.47
Y-EL38-2	14.50	4.52	3.48
Y-EL38-3	14.57	4.48	3.45
Z-EL35-1	9.78	3.96	3.30
Z-EL35-2	9.70	3.95	3.32
Z-EL35-3	9.80	3.99	3.32
Z-EL36-1	12.90	3.48	3.75
Z-EL36-2	12.95	3.55	3.78
Z-EL36-3	12.88	3.52	3.71
Z-EL38-1	17.25	3.05	3.85
Z-EL38-2	17.20	2.96	3.80
Z-EL38-3	17.29	3.07	3.82

Table 1. Physiochemical parameters for each sample. X stands for cultivar Xiangfei, Y for cultivar Italia and Z for cultivar Zaomeiguixiang. EL35: the berry begins to colour and soften, EL36: complete of veraison with an intermediate Brix, EL38: berry reaches harvest ripeness.

Brix of EL 36, and reaches harvest ripeness at EL38. At each stage, three replicates were harvested; approximately 300 grape berries were randomly collected for each replicate.

Physiochemical parameters. Fifty berries of each replicate were pressed and centrifuged to determine total soluble solids (TSS), pH value and titratable acidity. TSS was measured by a digital refractometer (PAL-1, Atago, Tokyo, Japan). The pH value was measured by a pH meter (FiveGo F2-Standard, Mettler Toledo, Switzerland). Titratable acidity was analysed by titration with NaOH (0.1M) to the end point of pH 8.2 and expressed as tartaric acid equivalents in accordance with the National Standard of People's Republic of China (GB/T15038-2006, 2006). The other berries were then frozen in liquid nitrogen and stored at −80 °C.

RNA extraction and sequencing. The extraction of total RNA from the berries was carried out by a Plant RNA extraction kit (Aidlab Biotechnologies, Beijing, China). The quality of the RNA was verified by agarose gel electrophoresis, and the concentration was determined by the absorbance ratio (A260/A280, 1.8–2.0) on an Implen P330 nanophotometer (Implen GmbH, Munich, Germany).

The RNA-Seq libraries were constructed from 27 samples according to the methods of Wang[33]. The enriched mRNA was obtained by using oligo (dT) magnetic beads then fragmented by 94 °C for 5 min. cDNA was synthesized by Superscript®III Reverse Transcriptase, followed by purification, end repair and dA-tailing and was then ligated with the sequencing adaptor. Afterwards, PCR amplification was conducted by indexed primers. The constructed library was QC checked by Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR System and then sequenced by Illumina HiSeq2000 platform at BGI Life Tech Co., Ltd. (Shenzhen, China). Low quality reads (more than 20% of the base qualities are lower than 10), reads with adaptors and reads with unknown bases (N bases more than 5%) were filtered to get clean reads and were stored in FASTQ format. The clean reads were mapped onto the reference grapevine genome (V2) using Hisat2[34].

Data Records

The RNA-Seq clean data of the 27 samples were deposited at the NCBI Sequence Read Archive with accessions SRP184152[35]. The files of gene expression level were deposited in NCBI's Gene Expression Omnibus (GEO), and are accessible through GEO Series accession number GSE130386[36]. The information of the differentially expressed genes (DEGs) between samples were deposited in figshare[37].

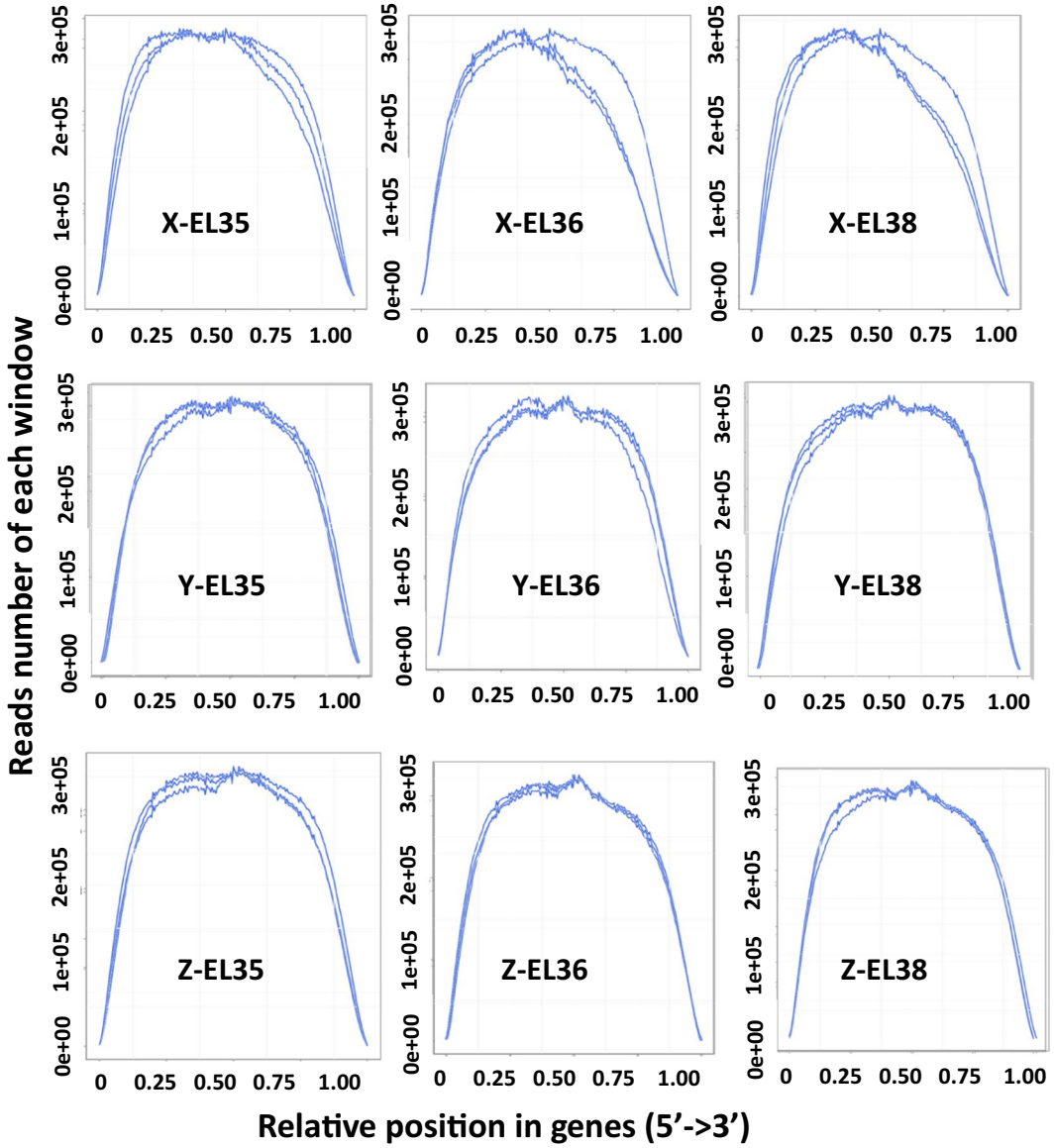

Fig. 2 Reads distribution on transcripts. The x-axis represents the position along transcripts, and the y-axis represents the number of reads.

Technical Validation

Quality control. The physiochemical parameter of the samples was shown in Table 1. A total of 27 RNA samples were prepared and sequenced, with the sequencing depth ranging between 22.48 and 33.08 million reads; the Q20 values for the clean reads were above 97%, and the average mapping ratio was 84.72% (Online-only Table 1).

Analysis of RNA-Seq data. After novel transcript detection, novel coding transcripts were merged with reference transcripts to get a complete reference. Clean reads were mapped to the transcript by using Bowtie2[38]. Gene expression levels were calculated with RSEM[39]. The distribution of reads based on the detection of read coverage skewness showed good fragmentation randomness (Fig. 2). The differentially expressed genes (DEGs) between samples were identified by the R package, DESeq2[40]. The DEGs with a $|\log2ratio| \geq 1$ and a false discovery rate probability ≤ 0.001 were considered statistically significant. The statistical analyses of DEG are shown in Fig. 3.

Usage Notes

The RNA-Seq fastq.gz files were deposited at Gene Expression Omnibus and can be downloaded using the fastq-dump tool of the SRA Toolkit (https://www.ncbi.nlm.nih.gov). The V2 reference genome of *V. vinifera*, the annotated file, could be retrieved at (http://genomes.cribi.unipd.it/grape/).

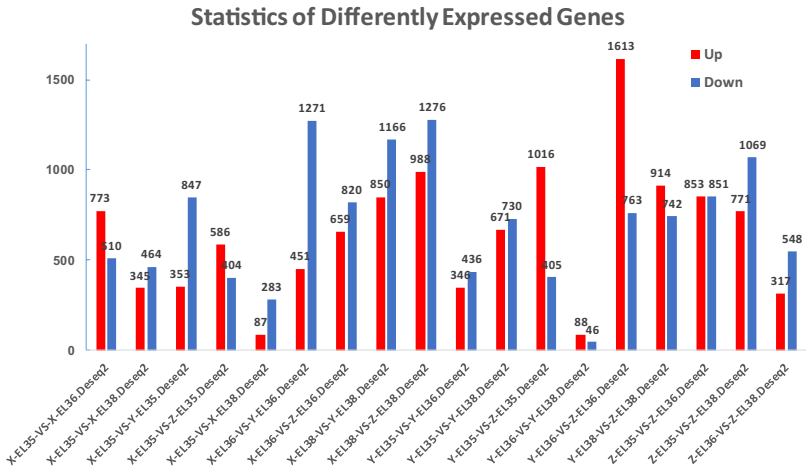

Fig. 3 Statistics of differently expressed genes. The X-axis represents the comparison method between groups and the y-axis represents DEG numbers. The red colour represents upregulated DEGs, and the blue colour represents downregulated DEGs.

Code Availability

SOAPnuke: https://github.com/BGI-flexlab/SOAPnuke. Version: v1.5.2. Parameters: -l 5 -q 0.51 -n 0.55 -i -Q 2–seqType 1.

HISAT2: http://www.ccb.jhu.edu/software/hisat. Version:v2.0.4.Parameters:–phred64–sensitive–no-discordant–no-mixed -I 1 -X 1000.

Bowtie2: http://bowtie-bio.sourceforge.net/Bowtie2. Version: v2.2.5. Parameters: -q–phred64–sensitive–dpad 0–gbar 99999999–mp 1,1–np 1–score-min L,0, −0.1 -I 1 -X 1000–no-mixed–no-discordant -p 1 -k 200.

RSEM: http://deweylab.biostat.wisc.edu/RSEM. Version: v1.2.12. Parameters: default.

References

1. Bohlmann, J. & Keeling, C. I. Terpenoid biomaterials. *Plant J.* **54**, 656–669 (2008).
2. Magnard, J. L. *et al.* Plant volatiles. Biosynthesis of monoterpene scent compounds in roses. *Science.* **349**, 81–83 (2015).
3. Fenoll, J., Martniez, M. D. A., Hellin, P. & Flores, P. Changes of free and glycosidically bound monoterpenes and aromatic alcohols in Muscatel and Ruby Seedless table grapes during development. *J. Inter. Des Sciences. De La Vigne Et Du Vin* **46**, 41–50 (2012).
4. Croteau, R. Biosynthesis and catabolism of monoterpenoids. *Chem. Rev.* **87**, 929–954 (1987).
5. Luan, F., Mosandl, A., Munch, A. & Wust, M. Metabolism of geraniol in grape berry mesocarp of *Vitis vinifera* L. cv. Scheurebe: demonstration of stereoselective reduction, E/Z-isomerization, oxidation and glycosylation. *Phytochemistry.* **66**, 295–303 (2005).
6. Fenoll, J., Manso, A., Hellin, P., Ruiz, L. & Flores, P. Changes in the aromatic composition of the *Vitis vinifera* grape Muscat Hamburg during ripening. *Food Chem.* **114**, 420–428 (2009).
7. Liu, B. *et al.* The free and enzyme-released volatile compounds of distinctive *Vitis amurensis* var. Zuoshanyi grapes in China. *Eur. Food Res. Technol.* **240**, 985–997 (2015).
8. Kalua, C. M. & Boss, P. K. Evolution of volatile compounds during the development of Cabernet Sauvignon grapes (*Vitis vinifera L.*). *J. Agric. Food Chem.* **57**, 3818–3830 (2009).
9. Kalua, C. M. & Boss, P. K. Comparison of major volatile compounds from Reisling and Cabernet Sauvignon grapes (*Vitis vinifera L.*) from fruitset to harvest. *Aus. J. Grape and Wine Res* **16**, 337–348 (2010).
10. Bureau, S. M., Razungles, A. J. & Baumes, R. L. The aroma of Muscat of Frontignac grapes: effect of the light environment of vine or bunch on volatiles and glycoconjugates. *J. Sci. Food Agric* **80**, 2012–2020 (2000).
11. Wang, Y. *et al.* Effects of cluster thinning on vine photosynthesis, berry ripeness and flavonoid composition of Cabernet Sauvignon. *Food Chem.* **248**, 101–110 (2018).
12. Xu, X. Q. *et al.* Differences in volatile profiles of Cabernet Sauvignon grapes grown in two distinct regions of China and their responses to weather conditions. *Plant Physiol. Biochem.* **89**, 123–133 (2015).
13. Koundouras, S., Marinos, V., Gkoulioti, A., Kotseridis, Y. & Van, L. C. Influence of vineyard location and vine water status on fruit maturation of non-irrigated cv. Agiorgitiko (*Vitis vinifera L.*). effects on wine phenolic and aroma components. *J. Agric. Food Chem.* **54**, 5077–5086 (2006).
14. Wilson, B., Strauss, C. R. & Williams, P. J. The distribution of free and glycosidically-bound monoterpenes among skin, juice, and pulp fractions of some white grape varieties. *Amer. J. Enol.Viticult.* **37**, 107–111 (1986).
15. Hjelmeland, A. K. & Ebeler, S. E. Glycosidically bound volatile aroma compounds in grapes and wine: a review. *Amer. J. Enol. Viticult.* **66**, 1–11 (2015).
16. Voirin, S. G., Baumes, R. L., Bitteur, S. M., Gunata, Z. Y. & Bayonove, C. L. Novel monoterpene disaccharide glycosides of *vitis vinifera* grapes. *J. Agric. Food Chem.* **38**, 1373–1378 (1990).
17. Dubey, V. S., Bhalla, R. & Luthra, R. An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *J. Biosci.* **28**, 637–646 (2003).
18. Luan, F. & Wust, M. Differential incorporation of 1-deoxy-Dxylulose into (3S)-linalool and geraniol in grape berry exocarp and mesocarp. *Phytochemistry* **60**, 451–459 (2002).
19. Dudareva, N., Klempien, A., Muhlemann, J. K. & Kaplan, I. Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol.* **198**, 16–32 (2013).
20. Schwab, W., Davidovich-Rikanati, R. & Lewinsohn, E. Biosynthesis of plant-derived flavor compounds. *The Plant J* **54**, 712–732 (2008).
21. Withers, S. T. & Keasling, J. D. Biosynthesis and engineering of isoprenoid small molecules. *Appl. Microbiol. Biotechnol.* **73**, 980–990 (2007).

22. Degenhardt, J., Kollner, T. G. & Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **70**, 1621–1637 (2009).
23. Doligez, A., Audiot, E., Baumes, R. & This, P. QTLs for muscat flavour and monoterpenic odorant content in grapevine (*Vitis vinifera L.*). *Mol. Breeding* **18**, 109–125 (2006).
24. Battilana, J. *et al.* The 1-deoxy-D: -xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theor. Appl. Genet.* **118**, 653–669 (2009).
25. Duchene, E. *et al*. A grapevine (*Vitis vinifera L.*) deoxy-d-xylulose synthase gene colocates with a major quantitative trait loci for terpenol content. *Theor. Appl. Genet.* **118**, 541–552 (2009).
26. Battilana, J. *et al*. Functional effect of grapevine 1-deoxy-D-xylulose 5-phosphate synthase substitution K284N on Muscat flavour formation. *J. Exp. Bot.* **62**, 5497–5508 (2011).
27. Martin, D. M., Chiang, A., Lund, S. T. & Bohlmann, J. Biosynthesis of wine aroma: transcript profiles of hydroxymethylbutenyl diphosphate reductase, geranyl diphosphate synthase, and linalool/nerolidol synthase parallel monoterpenol glycoside accumulation in Gewürztraminer grapes. *Planta.* **236**, 919–929 (2012).
28. Wen, Y. Q. *et al*. Using the combined analysis of transcripts and metabolites to propose key genes for differential terpene accumulation across two regions. *BMC Plant Biol.* **15**, 240 (2015).
29. Costantini, L. *et al*. Drawing Links from Transcriptome to Metabolites: The Evolution of Aroma in the Ripening Berry of Moscato Bianco. (*Vitis vinifera L.*). *Front. Plant. Sci.* **8**, 780 (2017).
30. Wu, Y. *et al*. Aroma characterization based on aromatic series analysis in table grapes. *Sci. Rep.* **6**, 31116 (2016).
31. Vitulo, N. *et al*. A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biol.* **14**, 99 (2014).
32. Coombe, B. G. Adoption of a system for identifying grapevine growth stages. *Aust. J. Grape Wine Res.* **1**, 100–110 (1995).
33. Wang, L., Si, Y., Dedow, L. K., Shao, Y., Liu, P. & Brutnell, T. P. A low-cost library construction protocol and data analysis pipeline for Illumina-based strand specific multiplex RNA-seq. *PLoS One.* **6**, e26426 (2011).
34. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast-spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
35. NCBI Sequence Read Archive. http://identifiers.org/ncbi/insdc.sra:SRP184152 (2019).
36. Sun, L. & Zhu, B. Transcriptome profiles of three Muscat table grape cultivars at three developmental stage. *Gene Expression Omnibus*, http://identifiers.org/geo:GSE130386 (2019).
37. Sun, L. *et al*. Transcriptome profiles of three Muscat table grape cultivars to dissect the mechanism of terpene biosynthesis. *figshare*, https://doi.org/10.6084/m9.figshare.c.4378256.v1 (2019).
38. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
39. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
40. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgements

This work was supported by the Science and Technology Innovation Ability Construction Projects of Beijing Academy of Agricultural and Forestry Sciences (KJCX20180411), Earmarked Fund for China Agriculture Research System (CARS-29) and Beijing Municipal Natural Science Foundation (6192017).

Author Contributions

L.S. designed the experiments and wrote the manuscript. B.Q.Z. analysed the data. X.Y.Z. collected the samples and extracted RNA. G.J.Z., A.L.Y., H.L.W. and X.Y.W. reviewed the manuscript. H.Y.X. designed the experiments, reviewed the manuscript and supervised the study.

Additional Information

Competing Interests: The authors declare no competing interests.

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