# The Major Origin of Seedless Grapes Is Associated with a Missense Mutation in the MADS-Box Gene *VviAGL11*[1]

**Carolina Royo,[a] Rafael Torres-Pérez,[a] Nuria Mauri,[a] Nieves Diestro,[a] José Antonio Cabezas,[b] Cécile Marchal,[c] Thierry Lacombe,[d] Javier Ibáñez,[a] Manuel Tornel,[e] Juan Carreño,[e] José Miguel Martínez-Zapater,[a] and Pablo Carbonell-Bejerano[a,2]**

[a]Instituto de Ciencias de la Vid y del Vino, Consejo Superior de Investigaciones Científicas-Universidad de La Rioja-Gobierno de La Rioja, 26007 Logrono, Spain
[b]Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria-Centro de Investigación Forestal, INIA-UPM, 28040 Madrid, Spain
[c]Institut National de la Recherche Agronomique, Centre de Ressources Biologiques de la Vigne, Domaine de Vassal, 34340 Marseillan-Plage, France
[d]AGAP, Universite Montpellier, CIRAD, Institut National de la Recherche Agronomique, Montpellier SupAgro, 34060 Montpellier, France
[e]Instituto Murciano de Investigación y Desarrollo Agrario y Alimentario, Sociedad Murciana de Investigación y Tecnología de Uva de Mesa, 30150 La Alberca, Spain

**ORCID IDs: 0000-0001-5067-882X (C.R.) (C.R.); 0000-0002-3696-4720 (R.T.); 0000-0002-5463-3802 (N.M.); 0000-0002-4728-6072 (J.A.C.); 0000-0001-9968-8228 (T.L.); 0000-0002-6286-5638 (J.I.); 0000-0002-5555-513X (J.C.) (J.C.); 0000-0001-7217-4454 (J.M.M.); 0000-0002-7266-9665 (P.C.)**

Seedlessness is greatly prized by consumers of fresh grapes. While stenospermocarpic seed abortion determined by the *SEED DEVELOPMENT INHIBITOR* (*SDI*) locus is the usual source of seedlessness in commercial grapevine (*Vitis vinifera*) cultivars, the underlying *sdi* mutation remains unknown. Here, we undertook an integrative approach to identify the causal mutation. Quantitative genetics and fine-mapping in two 'Crimson Seedless'-derived F1 mapping populations confirmed the major effect of the *SDI* locus and delimited the *sdi* mutation to a 323-kb region on chromosome 18. RNA-sequencing comparing seed traces of seedless and seeds of seeded F1 individuals identified processes triggered during *sdi*-determined seed abortion, including the activation of salicylic acid-dependent autoimmunity. The RNA-sequencing data set was investigated for candidate genes, and while no evidence for causal cis-acting regulatory mutations was detected, deleterious nucleotide changes in coding sequences of the seedless haplotype were predicted in two genes within the *sdi* fine-mapping interval. Targeted resequencing of the two genes in a collection of 124 grapevine cultivars showed that only the point variation causing the arginine-197-to-leucine substitution in the seed morphogenesis regulator gene *AGAMOUS-LIKE11* (*VviAGL11*) was fully linked with stenospermocarpy. The concurrent postzygotic variation identified for this missense polymorphism and seedlessness phenotype in seeded somatic variants of the original stenospermocarpic cultivar supports a causal effect. We postulate that seed abortion caused by this amino acid substitution in VviAGL11 is the major cause of seedlessness in cultivated grapevine. This information can be exploited to boost seedless grape breeding.

The reduction of seed content without altering fruit size is a major breeding goal in many fruit crops because it eases fruit eating, which increases attractiveness for consumers and improves the suitability of fruits for the food processing industry (Varoquaux et al., 2000). For centuries, seedlessness has been one of the most prized quality traits in grapevine (*Vitis vinifera*) berries intended for direct consumption either as fresh fruit or raisins (Ledbetter and Ramming, 1989). World demand for seedless grapes is rising, and vine growers are increasingly asking for seedless cultivars (FAO OIV, 2016). Understanding the genetic control of seedlessness, therefore, is paramount to boost the success of table grape breeding and fulfill such demands.

Although seeds are the sexual propagules of angiosperm plants, they are not required for the propagation of many woody crops such as grapevine that are vegetatively multiplied. This propagation method

[1]This work was partially supported by the European Union's Plant Genomics ERA-Net project GRASP and by the Spanish Ministry of Economy (MINECO) projects BIO2011-026229 and BIO2014-59324-R. This work also benefited from networking activities within the funded European Cooperation in Science and Technology (COST) Action FA1106 QualityFruit.

[2]Address correspondence to pablo.carbonell@icvv.es.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Pablo Carbonell-Bejerano (pablo.carbonell@icvv.es).

J.M.-Z. and P.C.-B. conceived the research; P.C.-B., J.M.-Z., and C.R designed experiments; C.R. and P.C.-B. performed experiments; C.R., P.C.-B., R.T.-P., N.M., N.D., and J.A.C. analyzed data; J.C., M.T., C.R., J.I., C.M., and T.L. provided materials and contributed to phenotyping tasks; P.C.-B., C.R., and J.M.M.-Z. wrote the article with input and comments from all the other authors.

www.plantphysiol.org/cgi/doi/10.1104/pp.18.00259

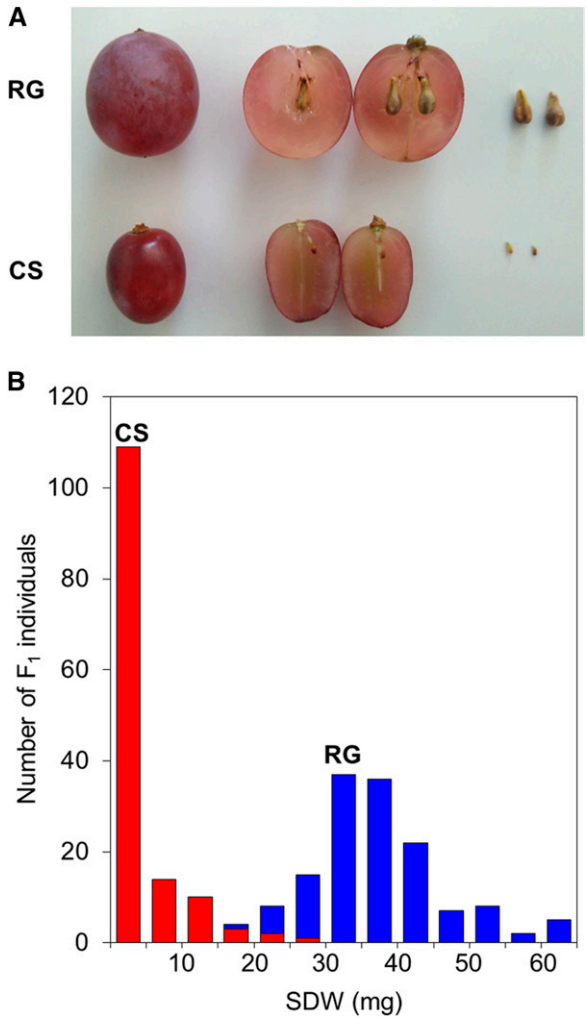



**Figure 1.** Stenospermocarpic phenotype and segregation in an RG×CS F1 population. A, Berries of RG and CS cross progenitors. Representative berries, seeds (RG), and stenospermocarpic seed traces (CS) at maturity are shown. B, Histogram showing the distribution of the SDW trait in RG×CS F1 progeny. SDW values correspond to the average of three seasons. The number of F1 individuals is denoted in different colors depending on the genotype in the *SDI* QTL-linked marker VMC7F2: *sdi*+ (198:200) and *sdi*– (200:200) individuals are represented in red and blue, respectively. RG and CS indicate the phenotypic class coincident with the phenotype of each progenitor.

enables the development of new cultivars from seedless somatic variants that appeared spontaneously during the history of grapevine domestication (This et al., 2006). Seedless grape variants can be classified into two major groups depending on the type of seedlessness (Stout, 1936; Pratt, 1971): (1) stenospermocarpy, in which fertilization and embryo development take place but seed development is prematurely aborted (Stout, 1936; Ledbetter and Ramming, 1989; Kovaleva et al., 1997); and (2) parthenocarpy, in which fruits develop in the absence of fertilization, yielding small berries that completely lack seeds, which has recently been related to impaired meiosis (Royo et al., 2016). Stenospermocarpy is widely used in the production of seedless table grape cultivars because berry size is less compromised, likely due to the presence of seminal rudiments or seed traces that promote fruit growth (Stout, 1936; Nitsch et al., 1960; Pratt, 1971).

A stable stenospermocarpy phenotype is shown by a few ancient oriental grapevine cultivars known as 'Kishmish' and derived varieties. They include the white-berried 'Kishmish', also known as 'Sultanina' or 'Thompson Seedless' (Dangi et al., 2001), which has been the major source of seedlessness in table grape breeding programs (Adam-Blondon et al., 2001; Ibáñez et al., 2009a, 2015). Therefore, studies of stenospermocarpy have focused on 'Sultanina' and 'Sultanina'-derived cultivars, which produce soft and often imperceptible seed traces generally lacking seed coat lignification. While the embryo usually remains viable, the degree of endosperm degeneration detected from 3 to 4 weeks after flowering (WAF) and the final size of seed traces are variable, depending on the genetic background (Pearson, 1932; Stout, 1936; Barritt, 1970; Pratt, 1971; Striem et al., 1992; Wang et al., 2015, 2016). Stenospermocarpy in 'Sultanina' has been associated with defects in the development of maternal seed coat tissues (Malabarba et al., 2017). Specifically, endotesta growth and lignification does not take place in seed traces. Abnormal development of the precursor inner ovule integument also has been reported in 'Sultanina' at earlier developmental stages (Pearson, 1932).

Concerning the genetic control of 'Sultanina'-derived seedlessness, different hypotheses were initially proposed depending on the approaches used to measure the trait and the genetic backgrounds analyzed (Bouquet and Danglot, 1996). However, a systematic analysis in several F1 cross progeny shows that a model involving three independent recessive loci regulated by a dominant locus could explain most segregations (Bouquet and Danglot, 1996). This dominant locus was later named *SEED DEVELOPMENT INHIBITOR* (*SDI*; Lahogue et al., 1998). Different quantitative genetic studies located the *SDI* quantitative trait locus (QTL) on linkage group (LG) 18, explaining up to 70% of the phenotypic variance in seed content parameters (Cabezas et al., 2006; Mejía et al., 2007, 2011; Costantini et al., 2008; Doligez et al., 2013).

Based on genetic linkage and putative homology, grapevine *AGAMOUS-LIKE11* (*VviAGL11* = *AGAMOUS-LIKE3*, *AG3* = *MINICHROMOSOME MAINTENANCE1, AGAMOUS, DEFICIENS, and SERUMRESPONSE FACTOR5, MADS5*) was proposed as the *SDI* candidate gene in the absence of information for other genes in the region (Costantini et al., 2008; Mejía et al., 2011). This assumption was done considering its homology to the MADS-box gene *AGL11*, also known as *SEEDSTICK* (*STK*; *At4g09960*), which controls ovule morphogenesis and seed coat differentiation in Arabidopsis (*Arabidopsis thaliana*; Pinyopich et al., 2003; Mizzotti et al., 2014). A role in seed morphogenesis also was shown for grapevine VviAGL11 homolog proteins (Malabarba et al., 2017). Molecular analyses

**Table I.** QTLs identified for SDW in the RG×CS F1 mapping population

QTLs detected in the CS map and in the consensus map (c) are indicated. QTLs were not detected in the RG map. LG, linkage group of the QTL; LOD, $Log_{10}$ of odds; variance, percentage of SDW variation explained by the QTL. Ranges of LOD and variance are shown if the QTL was detected in more than 1 year. cM, Centimorgan.

| QTL-Linked Marker | Map | LG | Year | Position | Confidence Interval | LOD | Variance |
|---|---|---|---|---|---|---|---|
| | | | | *cM* | | | % |
| VMC6F1-CS | CS | 2 | 07, 08, 09 | 15.6 | 0.0–36.6 | 1.4–4.7 | 0.7–1.6 |
| SNP1053_81-c | CS | 5 | 08 | 7.9 | 0.0–19.9 | 2.7 | 0.8 |
| VMC2H5-c | CS | 14 | 08 | 21.3 | 0.0–42.0 | 1.6 | 0.4 |
| VMC7F2-CS | CS | 18 | 07, 08, 09 | 66.1 | 64.7–66.1 | 67.4–109.7 | 71.0–83.0 |
| VVIB23-c | c | 2 | 07, 08, 09 | 23.2 | 16.2–31.2 | 3.0–9.9 | 1.6–2.3 |
| SNP1053_81-c | c | 5 | 08 | 9.3 | 0.0–36.3 | 4.4 | 1.0 |
| VMCNG1E1-CS | c | 14 | 08, 09 | 0.0 | 0.0–9.0 | 2.7–2.8 | 0.7–1.1 |
| VMC7F2-CS | c | 18 | 07, 08, 09 | 73.3 | 72.1–73.3 | 69.6–116.2 | 71.8–80.9 |

identified two nonsilent single-nucleotide polymorphisms (SNPs) in the seedless mutant haplotype (*sdi+*) of *VviAGL11* (Mejía et al., 2011; Malabarba et al., 2017). However, both amino acid substitutions were detected in homozygosity in the seeded wine grape 'Asyl Kara', a result that excluded them as functional dominant polymorphisms causing seedlessness (Mejía et al., 2011). Alternatively, the reduced expression of *VviAGL11* in fruits of seedless *sdi+* individuals compared with fruits or seeds of seeded individuals was related to the origin of seedlessness (Mejía et al., 2011; Ocarez and Mejía, 2016; Malabarba et al., 2017). Following this hypothesis, sequence polymorphisms observed in noncoding regions of the seedless allele of *VviAGL11* were proposed as putative mutations causing misexpression and seedlessness (Mejía et al., 2011; Di Genova et al., 2014). In a later effort to confirm the misexpression hypothesis, Ocarez and Mejía (2016) related several putative gene conversion events in the promoter region of *VviAGL11* with the reversion of stenospermocarpy in somatic variants of 'Sultanina' that produce regular seeds. However, those studies did not assess the question of whether the observed expression differences in *VviAGL11* are causes or consequences of the seedless syndrome. The misexpression hypothesis also has limitations in explaining the dominant nature of the *sdi* mutation. In addition, the fact that *VviAGL11* expression is decreased in seed tissues has not been proven consistently (Mejía et al., 2011; Ocarez and Mejía, 2016; Malabarba et al., 2017).

Here, we have reassessed the genetic and molecular origins of the 'Sultanina'-derived seedlessness in an independent unbiased study. We used large F1 crosses of table grape cultivars segregating for stenospermocarpy to delimit the location of the causal mutation through crossover mapping. To identify putative misexpression or coding mutations within the delimited interval, we followed a strategy that combined RNA-sequencing (RNA-seq) comparisons in F1 hybrids and targeted sequencing in a large collection of seeded and stenospermocarpic grapevine cultivars. The results clearly point to a single-nucleotide missense mutation in *VviAGL11* as the origin of the dominant seedless phenotype.

## RESULTS

### QTL Mapping of Seed Content Variation

Since we used the bred cv 'Crimson Seedless' (CS) as the seedlessness donor (Ramming et al., 1995) in our study, we wanted to confirm the effect of the major *SDI* QTL in this genetic background. For this purpose, we used an F1 mapping population derived from a 'Red Globe' (RG) × CS cross hybridization. As an indicator of seed lignification that is useful to discriminate stenospermocarpy (Bouquet and Danglot, 1996), we analyzed the variation in seed dry weight per berry (SDW). Both progenitors showed extreme divergence in their seed content (Fig. 1). RG berries had an average of 3.2 ± 0.2 fully developed seeds with a mean SDW of 33.2 ± 9.3 mg, while CS berries had an average of 2.2 ± 0.7 seed traces with a mean SDW of 0.43 ± 0.2 mg (Supplemental Table S1). SDW distribution in F1 progeny was asymmetric and bimodal and did not fit a normal distribution in any of the analyzed years (Fig. 1; Supplemental Table S1). Values for SDW were highly correlated over the three seasons analyzed ($0.86 < r < 0.93$), in agreement with a high broad sense heritability (0.80–0.91) of the trait.

To map the QTLs responsible for SDW in RG×CS progeny, we built a total of three linkage maps corresponding to each progenitor as well as a consensus map for the cross. Linkage maps comprised both simple sequence repeat (SSR) and SNP markers, to a total of 191 markers in the case of RG, 227 for CS, and 290 for the consensus map. QTLs for SDW were analyzed in every linkage map and every season. The results confirmed the detection of a major QTL for seedlessness in LG 18 linked to the microsatellite marker VMC7F2 in both the CS and the consensus maps (Table I). This QTL was identified consistently in the 3 years with a very high LOD value and explained

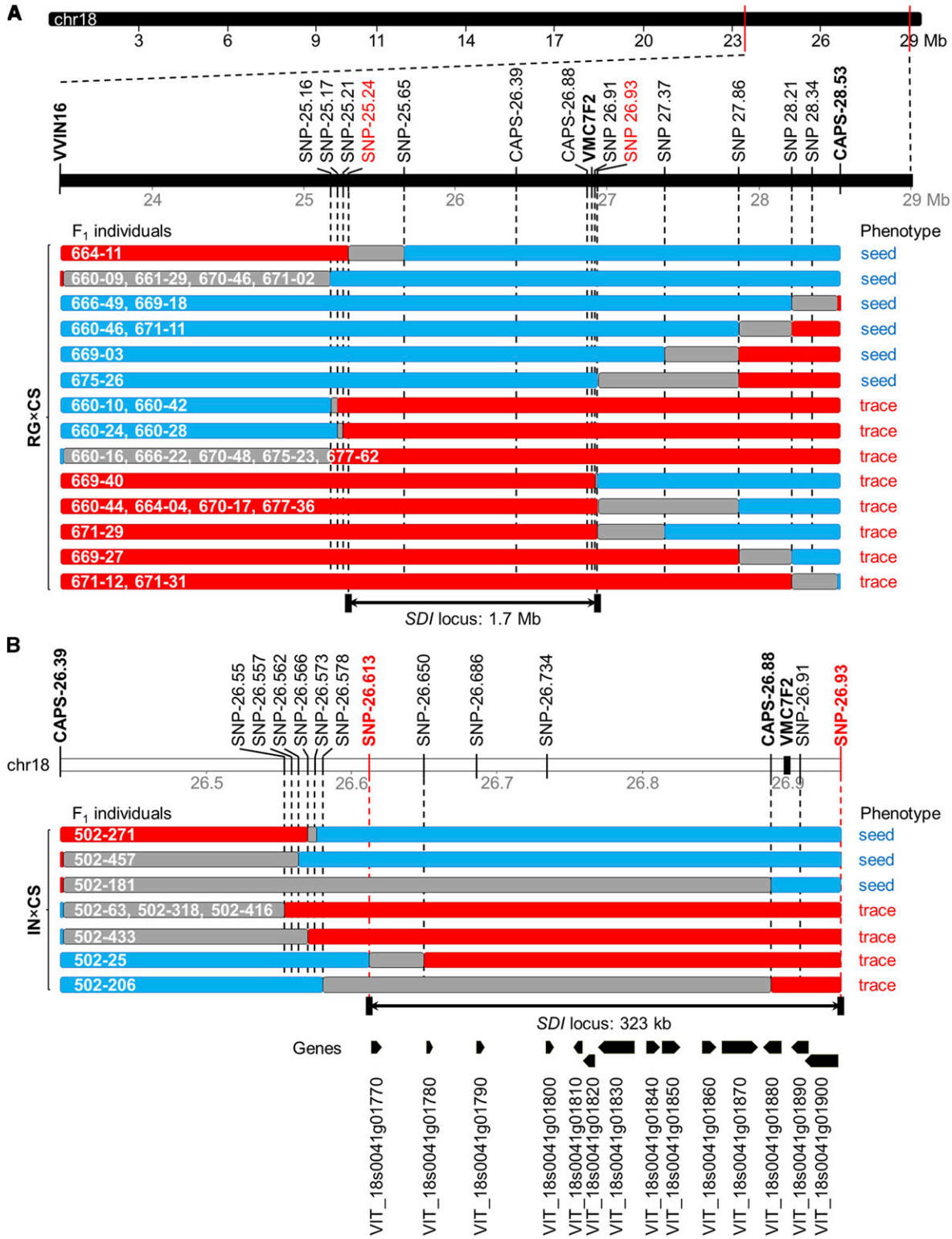
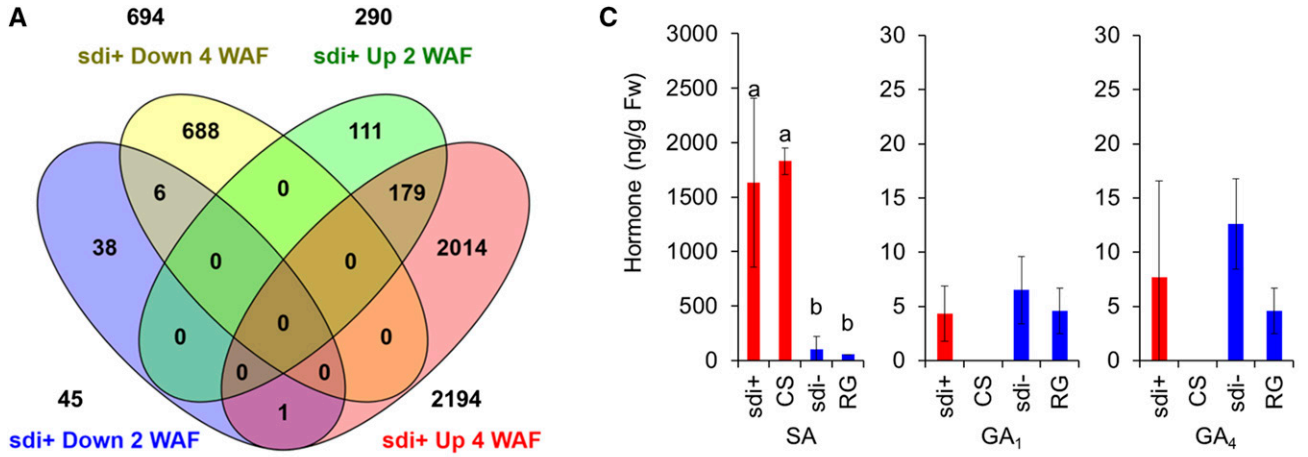



**Figure 2.** Fine-mapping of the *sdi* mutation in two seedless × seeded F1 populations. Schemes of crossover mapping for recombinants around the *SDI* locus in RG×CS (A) and 'Imperial Napoleon' (IN) × CS (B) F1 mapping populations are shown. Each class of recombinant F1 individuals is depicted in different lines, where red, blue, and gray colors denote chromosome fragments corresponding to the *sdi*+ seedless, *sdi*– seeded, and undetermined haplotypes, respectively. F1 individuals considered and their phenotypes (seed, seeded; trace, seedless) are indicated for each class on the left and right sides of the corresponding line, respectively. Markers studied in each F1 mapping population, the *sdi* fine-mapped interval, and the genes included on it according to distances in the PN40024 12X.0 reference genome and 12X V1 gene annotations are represented as well.




**B** Functional enrichment summary in DEGs between seedles (*sdi*+) and seeded (*sdi*-) individuals

| ***sdi*+ Up** | | ***sdi*+ Down** | |
|---|---|---|---|
| 2 WAF | 4 WAF | 2 WAF | 4 WAF |
| • Biotic stress response<br>• Protein kinase<br>• WRKY TFs<br>• SA signaling<br>• JA signaling<br>• Phenylpropanoid biosynthesis<br>• Stilbenoid biosynthesis<br>• Ascorbate and aldarate metabolism | • Biotic stress response<br>• NBS-LRR superfamily<br>• Protein kinase<br>• WRKY TFs<br>• Homeobox TFs<br>• NAC TFs<br>• Ethylene signaling<br>• SA signaling<br>• JA signaling<br>• Glycerolipid catabolism<br>• Tyrosine catabolism<br>• Auxin inactivation by conjugation<br>• Glutathione metabolism<br>• Photosynthesis<br>• Senescence | • Lignin metabolism<br>• Aromatic amino acid catabolism<br>• PLATZ TFs | • Cell growth<br>• Phenylpropanoid metabolism<br>• Monosaccharide metabolism<br>• Glycolysis/ Gluconeogenesis<br>• Nitrogen and sulfur metabolism<br>• Amino acid biosynthesis<br>• S1Fa-like TFs |

**Figure 3.** Transcriptome and phytohormone level comparisons between seedless (*sdi*+) and seeded (*sdi*–) RG×CS F1 individuals. A, Venn diagram comparison of DEGs identified in 2-WAF fruits (NimbleGen microarray) and 4-WAF aborting/developing seeds (RNA-seq). The number of up- and down-regulated DEGs in seedless (*sdi*+) versus seeded (*sdi*–) individuals in each experiment is shown. B, Summary of functional categories overrepresented in up- and down-regulated DEGs in each experiment. This table corresponds to a summary of the full list of functional enrichment results presented in Supplemental Table S4. JA, Jasmonic acid; TF, transcription factors. C, Levels of SA and bioactive GAs ($GA_1$ and $GA_4$) in RG×CS *sdi*+ and *sdi*– F1 individuals and in the cross progenitors, CS and RG. Hormones were measured in seed traces or in seeds at 4 WAF. Average and SD concentration values of five replicates for *sdi*+ and *sdi*– F1 individuals and two replicates for CS and RG are shown. Letters denote significant differences at $P < 0.01$ using Duncan's post hoc test. Fw, Fresh weight.

between 71% and 83% of the total variance for SDW, depending on the year and the linkage map. It colocalized with the *SDI* QTL detected previously in different genetic backgrounds related to stenospermocarpic seedlessness (Doligez et al., 2002, 2013; Cabezas et al., 2006; Mejía et al., 2007, 2011; Costantini et al., 2008). Another QTL detected in all the analyzed seasons for SDW was located on LG 2 (Table I). Rather than contributing to stenospermocarpy, the genotype in this QTL marker (VVIB23) correlates with the variation for seed number per berry (Supplemental Table S1) and likely colocalizes with the sex locus, as has been described previously (Costantini et al., 2008; Battilana et al., 2013; Doligez et al., 2013). Finally, two other minor QTLs for SDW were located on LGs 5 and 14, but they showed environmental interactions, since they were not detected in every season (Table I).

### Fine-Mapping of the *sdi* Mutation

A genetic fine-mapping strategy was developed to delimit the location of the mutation that underlies the

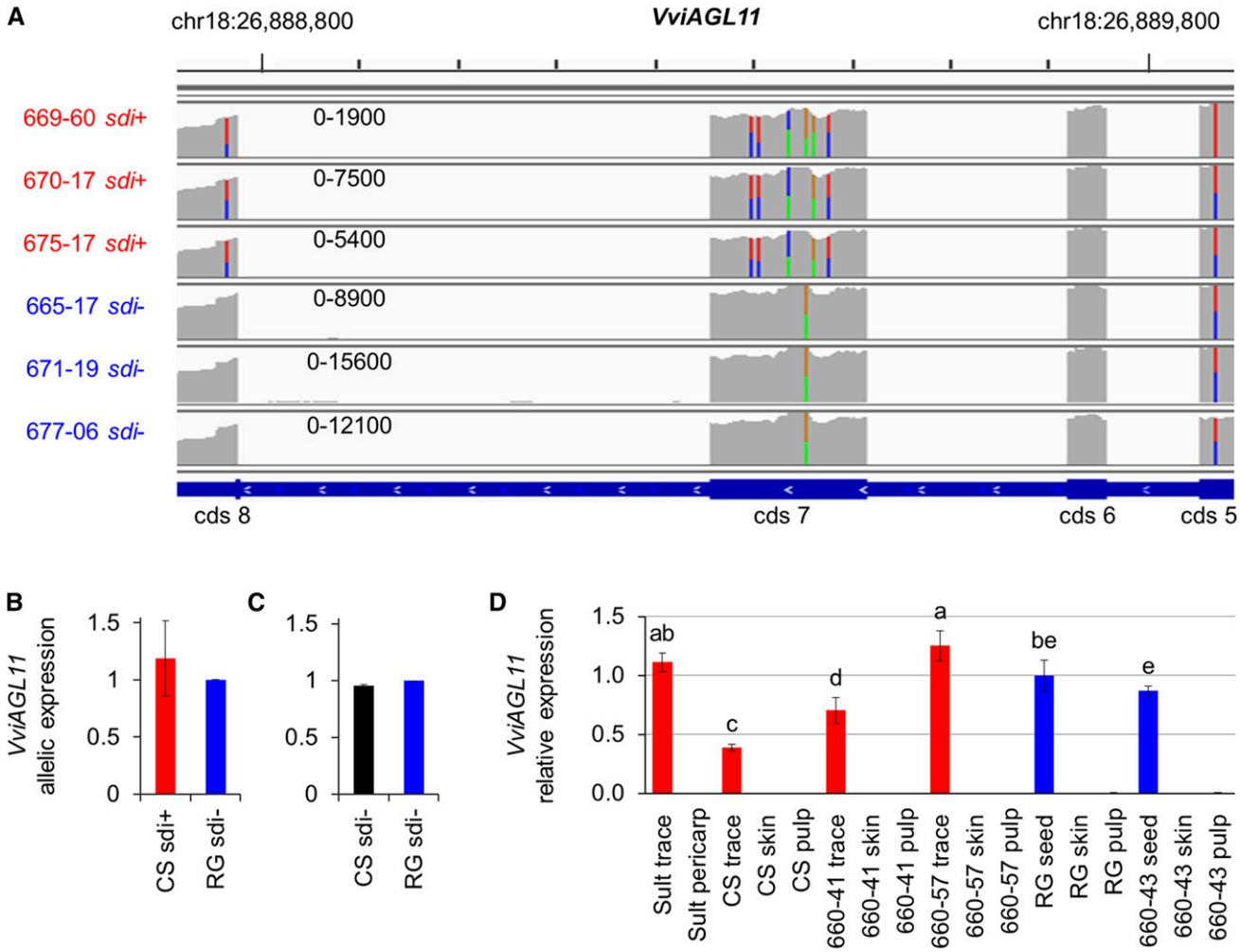



**Figure 4.** *VviAGL11* transcript levels. A, Representation of RNA-seq reads aligned in a region of *VviAGL11* with a high density of heterozygous SNPs. This image was obtained from Integrative Genomics Viewer. Read depth is shown separately for each studied seedless (*sdi*+) and seeded (*sdi*–) RG×CS F1 individual. The frequency of each allele at single-nucleotide variant (SNV) positions compared with the reference genome is denoted in colored bars, with red for T, green for A, blue for C, and orange for G nucleotide sequences. The read depth range depicted for each individual is indicated as well. A scheme of the portion of *VviAGL11* 12X V1 gene model annotated in this region of the PN40024 12X.0 reference genome is shown in blue at the bottom of the image, with coding sequence exons (cds) numbered from the start codon. B, Average *VviAGL11* allelic frequency in seedless RG×CS F1 individuals estimated from RNA-seq data. C, Average *VviAGL11* allelic frequency in seeded RG×CS F1 individuals estimated from RNA-seq data. In B and C, frequencies were normalized to the *sdi*– allele inherited from RG, for which expression levels are depicted in blue. Only positions with a heterozygous genotype in the three replicates (six and three SNPs in seedless and seeded individuals, respectively) were considered to estimate average frequencies and SD (represented by error bars) in the three seeded or seedless replicates. D, *VviAGL11* expression estimated by RT-qPCR. Expression in different pericarp and seed or seed trace (trace) tissues is shown for 'Sultanina' (Sult) at fruit set as well as for CS, RG, and RG×CS F1 individuals at 4 WAF. Expression levels were normalized relative to the *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* (*GPDH*) housekeeping gene and to RG seed tissue. Similar results were obtained when two other control genes were used for normalization (see "Materials and Methods"). Data represent means and SD of three technical replicates. Different letters denote significant differences between seed-related tissue samples with $P < 0.01$ using Duncan's post hoc test. Red and blue colors denote expression in stenospermocarpic and seeded individuals, respectively.

*SDI* QTL. For every RG×CS F1 individual, the seedlessness phenotype de visu determined during at least 3 years was fully linked to the genotype in the VMC7F2 marker, with no recombinant individuals or false detection observed (Supplemental Table S1). Every F1 individual inheriting the *sdi*+ seedless allele (198 bp) or the *sdi*– seeded allele (200 bp) from the CS progenitor showed a seedless or seeded phenotype, respectively, confirming the dominant nature of the *sdi* mutation (Lahogue et al., 1998; Cabezas et al., 2006;

**Table II.** Nucleotide variation in the *sdi*+ haplotype generating protein sequence variation

SNVs specific to the *sdi*+ haplotype located within the *sdi* fine-mapping interval that were detected in the RNA-seq data set (Supplemental Table S5) with an impact on protein-coding sequence are presented. For each SNV, the position and gene annotation in the 12X V1 grapevine reference genome, the nucleotide change in the coding sequence of the *sdi*+ haplotype, the amino acid change, the PROVEAN score, and the predicted effect of the change in the function of the protein are indicated. PROVEAN score ≤ −2.5 and > −2.5 are considered deleterious and neutral, respectively (Choi et al., 2012). Predicted deleterious changes are shown in boldface.

| Genome Position | Gene Identifier | Functional Annotation | Coding Sequence Change | Amino Acid Change | Score | Predicted Effect |
|---|---|---|---|---|---|---|
| **chr18:26,859,228** | **VIT_18s0041g01870** | **VviPPAT2** | **451 C→T** | **Arg-151Cys** | **−6.44** | **Deleterious** |
| **chr18:26,871,891** | **VIT_18s0041g01870** | **VviPPAT2** | **584 A→T** | **Gln-195Leu** | **−4.27** | **Deleterious** |
| chr18:26,889,399 | VIT_18s0041g01880 | VviAGL11 | 628 A→G | Thr-210Ala | −0.21 | Neutral |
| **chr18:26,889,437** | **VIT_18s0041g01880** | **VviAGL11** | **590 G→T** | **Arg-197Leu** | **−5.40** | **Deleterious** |
| chr18:26,914,516 | VIT_18s0041g01890 | Unknown protein | 148 G→A | Asp-50Asn | 0.45 | Neutral |
| chr18:26,918,160 | VIT_18s0041g01900 | Translation elongation factor | 997 A→T | Thr-333Ser | 1.01 | Neutral |

Karaagac et al., 2012; Bergamini et al., 2013). Therefore, this F1 population was screened for meiotic recombinants by studying the VVIN16 SSR marker and a newly developed cleaved amplified polymorphic sequence (CAPS) marker (CAPS-28.53) mapping at either side of the VMC7F2 marker, respectively (Fig. 2). This interval comprises a 5.1-Mb physical distance in the 12X.0 grapevine reference genome assembly (Fig. 2), which is equivalent to 5.5 Mb in the recently published 12X.v2 assembly, as one small scaffold has been remapped between the VVIN16 and SNP-25.16 markers (Canaguier et al., 2017). A total of 29 recombinants were detected, 14 and 15 of them corresponding to recombination in upstream and downstream sequences relative to VMC7F2, respectively (Fig. 2). Additional inner markers were genotyped to map the meiotic recombination breakpoint in these individuals. In this manner, the *sdi* mutation was delimited to an interval of 1.69 Mb (chr18:25,246,979–chr18:26,936,376 positions in the 12X.0 reference genome) flanked by markers SNP-25.24 and SNP-26.93 (Fig. 2; Supplemental Table S2). Supporting its continuity in the grapevine reference genome, the delimited mapping interval remained intact in the reassembled 12X.v2 version (Canaguier et al., 2017).

To further delimit the location of the *sdi* mutation, additional recombinants within the interval delimited in RG×CS progeny were searched for in another seeded × seedless F1 mapping population derived from an IN×CS cross. Inheritance of seedlessness was fully linked to the presence of the *sdi*+ allele for the CAPS-26.88 marker in this mapping population, which colocalizes with *VviAGL11* (*VIT_18s0041g01880*) in close proximity to the VMC7F2 marker (Fig. 2). The IN×CS progeny were screened using CAPS-26.39 and CAPS-26.88 markers (Fig. 2; Supplemental Table S2), and nine recombinant F1 individuals were detected. Additional genotyping for crossover mapping delimited the *sdi* mutation to an interval of 323 kb (chr18:26,613,101–chr26,936,376) flanked by markers SNP-26.613 and SNP-26.93. Unfortunately, recombination breakpoints for individuals 502-181 and 502-206 could not be mapped more precisely to further restrict the interval length because material for these recombinants is no longer available. According to grapevine 12X V1 gene annotations (http://genomes.cribi.unipd.it/), the delimited *SDI* locus comprises 14 genes, although the two genes at each extreme were only partially included (Fig. 2). These 14 genes were considered as *SDI* QTL candidates in subsequent analyses.

### Transcriptome Analysis Reveals the Activation of Salicylic Acid-Dependent Defenses along with the Repression of Seed Morphogenesis during Stenospermocarpic Seed Abortion

Transcriptome comparisons were addressed to understand the processes involved in stenospermocarpic seed abortion and to search for candidate mutations underlying the *SDI* locus. To these aims, RNA-seq was used to compare seedless and seeded RG×CS F1 individuals. To minimize genetic background effects, different F1 individuals were analyzed as independent biological replicates. Individuals were selected to contrast in the allele of the *SDI* locus inherited from CS (*sdi*+ [seedless] or *sdi*− [seeded]; Supplemental Table S1). From pea-size fruits collected at 4 WAF, seeds or seed traces were extracted to specifically analyze gene expression in the affected organ. In this manner, we identified 2,888 differentially expressed genes (DEGs) with adjusted $P \leq 0.05$ in edgeR and 2-fold or greater change, most of which (76%) were up-regulated in seed traces of seedless F1 progeny (Fig. 3A; Supplemental Table S3). To understand the biological meaning underlying these differences, a functional enrichment analysis was carried out. Remarkably, several pathogen response-related functions were overrepresented among genes up-regulated in seed traces, which coincided with the activation of senescence/catabolism processes (Fig. 3B; Supplemental Table S4). WRKY, Homeobox, and NAC transcription factors may regulate these processes during seed abortion, as these transcription factor families were overrepresented in seed trace up-regulated genes as well. Likely connected with the lack of seed differentiation, genes up-regulated in seed traces also were overrepresented in photosynthesis-related categories. As expected, genes down-regulated in seed traces were overrepresented in functional categories related to seed morphogenesis, such as

phenylpropanoid metabolism (Fig. 3B; Supplemental Table S4). Collectively, these results are in line with general trends described by Wang et al. (2016) following a similar approach.

To assess for transcriptome differences at earlier stages, a similar comparison between *sdi*+ and *sdi*− RG×CS F1 individuals was carried out at the fruit set stage (2 WAF) using NimbleGen microarrays. Although we used whole setting fruits because the extraction of seed content was difficult at this stage, this strategy eased the analysis of a higher number of replicates. In this case, only 335 DEGs were detected (adjusted $P \leq 0.05$ in limma and 2-fold or greater change). Nevertheless, 55% of these DEGs coincided with 4-WAF DEGs, and seedless up-regulated genes predominated in this data set as well (86.6% of DEGs identified at 2 WAF), showing a consistency between the two experiments in spite of ontogeny, tissue, and interannual differences (Fig. 3A; Supplemental Table S3). Also in agreement with the 4-WAF results, genes down-regulated in seedless fruits at 2 WAF were overrepresented in seed differentiation-related functions such as lignin metabolism (Fig. 3B; Supplemental Table S4). Immune-like responses were activated in whole seedless fruits at 2 WAT, indicating that, at 4 WAT, these responses are not an artifact due to sample manipulation during the extraction of seed traces.

Immune-like responses activated in stenospermocarpic offspring comprised salicylic acid (SA) signaling pathway homologs such as *ENHANCED DISEASE SUSCEPTIBILITY1* (*EDS1*; *VIT_17s0000g07370*, *VIT_17s0000g07400*, *VIT_17s0000g07420*, and *VIT_17s0000g07560*), a gene annotated as putative *SYSTEMIC ACQUIRED RESISTANCE DEFICIENT1* (*SARD1*; *VIT_17s0000g03370*; National Center for Biotechnology Information [NCBI] annotated locus LOC100259493), and *PATHOGEN RESPONSE1* (*VIT_00s0207g00130*, *VIT_00s0207g00160*, *VIT_03s0088g00710*, *VIT_03s0088g00780*, and *VIT_03s0088g00810*; Supplemental Table S3). Considering that EDS1 and SARD1 activate SA production and defense responses in Arabidopsis (Zhang et al., 2010; Rietz et al., 2011), levels of SA were compared between seeds and seed traces at 4 WAF. A prominent 16-fold increase in SA levels was detected in seed traces when *sdi*+ and *sdi*− RG×CS F1 individuals were compared (Fig. 3C), suggesting a role of SA in *sdi*-mediated seed abortion. A greater difference (33-fold) was observed between the seedless progenitor CS, with small seed traces, and RG, with big seeds. When the active GAs $GA_1$ and $GA_4$ were measured, no significant difference was observed in F1 progeny (Fig. 3C), suggesting a minor contribution of this phytohormone to stenospermocarpy, at least in the studied developmental stage.

### Stenospermocarpy Does Not Associate with Misexpression Mutations

To identify candidate misexpression mutations, we inspected the transcriptome data set for the presence of DEGs within the *SDI* fine-mapping interval.

**Table III.** Genotypes of the two *VviPPAT2* candidate SNVs in a collection of 93 grapevine accessions

The numbers of seedless and seeded accessions identified for each genotype are specified. For each position, the allele encoding candidate deleterious substitutions is denoted in boldface. Accessions with genotypes that exclude these candidate SNVs as dominant *sdi* mutations are described in the footnotes. The full list of genotypes for each accession can be found in Supplemental Table S6. nd, No data were obtained.

| Genotype | VviPPAT2 Arg-151Cys | VviPPAT2 Gln-195Leu | Seedless Accessions | Seeded Accessions |
|---|---|---|---|---|
| 1 | **T**:**T** | **T**:**T** | 1 | 0 |
| 2 | **T**:**T** | A:**T** | 1 | 0 |
| 3 | C:**T** | A:**T** | 14 | 1[a] |
| 4 | C:**T** | A:A | 0 | 1[b] |
| 5 | C:C | A:A | 0 | 28 |
| 6 | **T**:T | nd | 0 | 1[c] |
| 7 | nd | **T**:T | 2 | 0 |
| 8 | C:**T** | nd | 0 | 7[d] |
| 9 | nd | A:**T** | 2 | 1[e] |
| 10 | C:C | nd | 0 | 1 |
| 11 | nd | A:A | 0 | 33 |

[a]'Verdil'. [b]'Pedro Ximenes'. [c]'Planta Nova'. [d]'Afus Ali', 'Alphonse Lavallee', 'Aramon', 'Aubun', 'Morio Muskat', 'Naparo', and 'Semillon'. [e]'Cornichon Blanc'.

Supporting the continuity of this interval in the grapevine reference genome (Jaillon et al., 2007), we found that it is almost fully covered by one contig (000279F) in the genome assembly that was independently produced in 'Cabernet-Sauvignon' (Chin et al., 2016). In addition, the same gene models are annotated in that interval of the reference genome, according to V1 and V2 gene predictions (Vitulo et al., 2014), and no additional missing gene was supported by our RNA-seq read alignments (PRJNA418130 in the NCBI Sequence Read Archive). While no DEG was detected in the RNA-seq assay, *VviAGL11* was the only one out of the 14 genes within the interval that was differentially expressed in the microarrays of whole berries at 2 WAF, showing a 3.4-fold repression in *sdi*+ fruits (Supplemental Table S3). When seed tissues were studied specifically in the RNA-seq experiment at 4 WAF, *VviAGL11* did not exceed the significance thresholds (adjusted $P = 0.23$ and 2-fold repression in *sdi*+ individuals). Regardless, any potential allelic imbalance in *VviAGL11* transcript levels was assessed from the RNA-seq data set. Considering the mean allelic frequency at heterozygous SNP positions (Fig. 4A), the expression between *sdi*+ and *sdi*− alleles was balanced in the seedless RG×CS F1 individuals analyzed (Fig. 4B). In fact, the seedless individual 669-60 showed ~50% more counts for the *sdi*+ allele (Fig. 4A), which indicates that no cis-acting regulatory mutation inhibiting the expression of *VviAGL11* was present in the *sdi*+ haplotype. A similarly balanced expression between the two *sdi*− *VviAGL11* alleles was observed in seeded F1 individuals (Fig. 4C). In agreement with RNA-seq results, the absence of allelic imbalance in *VviAGL11* was confirmed when allele-specific primers were used to compare the expression between the two alleles in

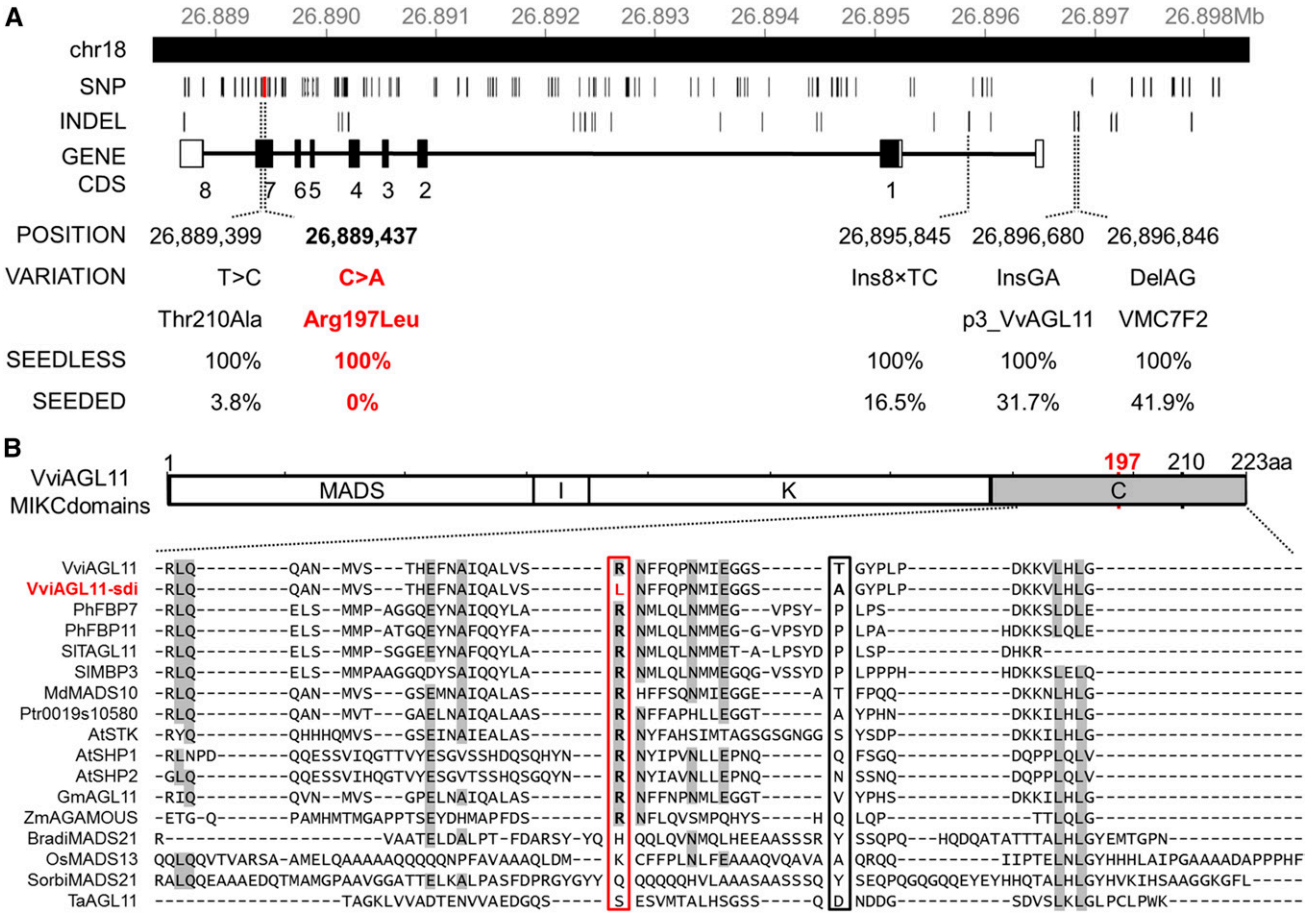



**Figure 5.** Sequence variation in *AGL11* and association with seedlessness. A, Nucleotide sequence variation in the *sdi*+ haplotype of *VviAGL11* and association of the *sdi*-linked variants with seedlessness in a collection of grapevine cultivars. *VviAGL11* (exons + introns) plus the 2-kb upstream sequence were targeted for Illumina paired-end sequencing in 105 seeded and five seedless accessions. The positions of SNPs and INDELs identified in the *sdi* haplotype compared with the PN40024 reference genome are depicted on a scheme of the gene (encoded in the minus strand), with coding sequences (CDS) numbered from the start codon of the gene. The frequency of the variant allele in stenospermocarpic seedless and seeded accessions is indicated for specific candidate polymorphisms selected from the full list of polymorphisms available in Supplemental Table S7. Genotype-phenotype correlation of 100% was identified only for the SNP causing the Arg-197Leu substitution in the *sdi* allele, which is highlighted in red. B, VviAGL11 protein domain model and alignment of the C-terminal domain from homologous AGL11 and related Arabidopsis AGAMOUS-lineage MIKC-type proteins. Grapevine proteins encoded by the seeded (VviAGL11) and the seedless (VviAGL11-sdi) alleles, AGL11 homologs from dicotyledonous (*Petunia hybrida*, FBP7 CAA57311.1 and FBP11 CAA57445.1; *Solanum lycopersicum*, SlAGL11 AY098736.2 and SlMBP3 XM_010324479.2; *Malus domestica*, MdMADS10 CAA04324.1; *Populus trichocarpa*, Ptr_0019s10580 U5FIJ5; Arabidopsis, AtSTK AT4G09960.3; and *Glycine max*, GmAGL11 B9MSS8) and monocotyledonous (*Zea mays*, ZmAGAMOUS NP_001105946.1; *Brachypodium distachyon*, BradiMADS21 G9BIK9; *Oryza sativa*, OsMADS13 Q2QW53; *Sorghum bicolor*, SorbiMADS21 C5XEN4; and *Triticum aestivum*, TaAGL11 ABF57916.1) species, as well as related AGAMOUS-lineage SHATTERPROOF Arabidopsis homologs (AtSHP1, AT3G58780.3 and AtSHP2, AT2G42830.2) were compared using the Conserved Domain Database tool available from NCBI (http://www.ncbi.nlm.nih.gov/cdd/). The alignment was performed with ClustalW in MEGA7 software (Kumar et al., 2016). The positions of the two amino acid substitutions detected in the VviAGL11-sdi allele are highlighted in boxes, and the seedlessness-associated Arg-197Leu substitution is shown in red.

'Sultanina' seed traces at fruit set stage by reverse transcription-quantitative PCR (RT-qPCR; *sdi*+:*sdi*− allelic expression ratio = 0.91 ± 0.1 [sd]). According to RNA-seq data, allelic imbalance also was absent in all other expressed genes located within the *SDI* interval (Supplemental Table S5; GSE107014 entry of the Gene Expression Omnibus database).

We also compared the absolute expression of *VviAGL11* in additional seedless and seeded accessions and different fruit tissues using RT-qPCR. This assay showed that, at fruit set and pea-size stages, the expression of *VviAGL11* was restricted to seed or seed traces and was not present in fruit pericarp (Fig. 4D). Remarkably, when *sdi*+ seed traces and *sdi*− developing seeds were

compared directly, *VviAGL11* transcript levels were similar or even higher in several seedless accessions. Altogether, the balanced allelic expression of *VviAGL11* in seedless *sdi*+/*sdi*− heterozygous accessions, along with the lack of correlation of *VviAGL11* absolute expression in seed organs and the seedlessness phenotype, reject the hypothesis of misexpression mutations in this gene as the origin of stenospermocarpy.

### Transcriptome Sequence Analysis Identifies *sdi* Candidate Coding Mutations

Once the presence of causal misexpression mutations was ruled out, we investigated the RNA-seq data set for candidate *sdi* mutations in coding sequences. Variant calling of RNA-seq data detected 68 variants specific to the *sdi*+ haplotype within the fine-mapping interval, and all of them were SNVs. While these SNVs affected the sequence of six genes within the interval, no high effect (loss of start or gain of stop codon) was predicted for any of them (Supplemental Table S5). Nonetheless, four of these genes collectively harbored six missense substitutions specific to the *sdi*+ haplotype: two located in the putative pantetheine-phosphate adenylyltransferase *VviPPAT2* (*VIT_18s0041g01870*) gene, two in *VviAGL11* (*VIT_18s0041g01880*), one in an unknown protein (*VIT_18s0041g01890*) gene, and one in a translation elongation factor (*VIT_18s0041g01900*) gene (Table II). An analysis using PROVEAN software predicted that three of these missense substitutions impact the biological functions of the affected proteins: Arg-151Cys (chr18:26,859,228) and Gln-195Leu (chr18:26,871,891) in *VviPPAT2* and Arg-197Leu (chr18:26,889,437) in its consecutive downstream gene *VviAGL11* (Table II). Thus, these three missense SNVs were tested as candidate *sdi* mutations in subsequent phenotype-association genotyping analyses.

### Candidate Missense Substitutions in *VviPPAT2* Do Not Associate with Seedlessness

*VviPPAT2* is homolog to a gene coding for an enzyme involved in coenzyme A biosynthesis and lipid storage in Arabidopsis seeds (Rubio et al., 2008). We named this gene in that manner because a *VviPPAT1* paralog is present in the grapevine reference genome (*VIT_04s0023g01990*). To assess for any possible connection between the two predicted deleterious amino acid substitutions in *VviPPAT2* and seedlessness determination, amplicons containing each SNV were sequenced in 20 and 73 *sdi*+ seedless and *sdi*− seeded cultivars, respectively (Table III; Supplemental Table S6). While both SNVs were validated by PCR in the CS progenitor, no genotype-phenotype association was confirmed. For the SNV at position chr18:26,859,228 (C:T), the T nucleotide allele (corresponding to the *sdi*+ haplotype of CS according to the RNA-seq data) was present in 10 seeded cultivars ('Afus Ali', 'Alphonse Lavallee', 'Aramon', 'Aubun', 'Morio Muskat', 'Naparo', 'Planta Nova', 'Pedro Ximenes', 'Semillon', and 'Verdil'); similarly, for the SNV at position chr18:26,871,891 (A:T), the T nucleotide allele of the *sdi*+ haplotype in CS was present in two seeded cultivars ('Cornichon Blanc' and 'Verdil'; Table III; Supplemental Table S6). These results do not support a role for candidate variants in *VviPPAT2* as major seedlessness-responsible dominant mutations.

### Stenospermocarpic Seedlessness Is Linked Specifically to One Missense Substitution in *VviAGL11*

Genotyping also was conducted to assess the role in stenospermocarpy for the candidate missense SNV detected in *VviAGL11*. Given that previous studies proposed several *sdi* candidate mutations in this gene (Mejía et al., 2011; Di Genova et al., 2014; Ocarez and Mejía, 2016; Malabarba et al., 2017), the whole gene plus the 2-kb upstream sequence were targeted for resequencing in a collection of 110 grapevine accessions using Illumina next-generation sequencing (NGS). Variant calling from NGS data identified 537 polymorphic sites in the 110 accessions, which included 448 SNPs and 89 insertions-deletions (INDELs; Supplemental Table S7). For 146 polymorphic sites (124 SNPs and 22 INDELs), the variant allele compared with the reference genome was shared in all five 'Sultanina'-derived seedless accessions studied (Fig. 5A). These variants were considered to be linked in the *sdi*+ haplotype because the analysis included two *sdi*+/*sdi*+ homozygous seedless 'Ruby Seedless' × 'Moscatuel' (RS×MO) F1 individuals obtained from a seedless × seedless cross. From all these positions, only the genotype for the SNP at position chr18:26,889,437 (A:C, according to the sequence in the plus strand) was fully associated with the seedlessness phenotype. Remarkably, this SNP causes the Arg-197Leu substitution with the predicted deleterious effect already detected in the RNA-seq variant calling (Table II). The A nucleotide allele was present and absent, respectively, in every stenospermocarpic and seeded accession (Fig. 5A; Supplemental Table S7). This amino acid substitution is located in the C-terminal domain characteristic of MIKC-type MADS-box genes (Fig. 5B). Although this domain is the most variable in MIKC genes (Kaufmann et al., 2005), the Arg residue at this position generally is conserved in AGL11 homologs of dicotyledonous species as well as in the related AGAMOUS-lineage SHATTERPROOF1 (SHP1) and SHP2 proteins (Fig. 5B; Pabón-Mora et al., 2014), which suggests that this residue could be functionally relevant.

Other candidate variants in *VviAGL11* were discarded in the resequencing analysis. For the SNP at position chr18:26,889,399 (C:T) corresponding to the second amino acid substitution in the *sdi*+ haplotype (Thr-210Ala), the C nucleotide detected in the *sdi*+ haplotype also was present in 'Aubun', 'Cornichon Blanc', and 'Verdil' seeded cultivars (Fig. 5A; Supplemental

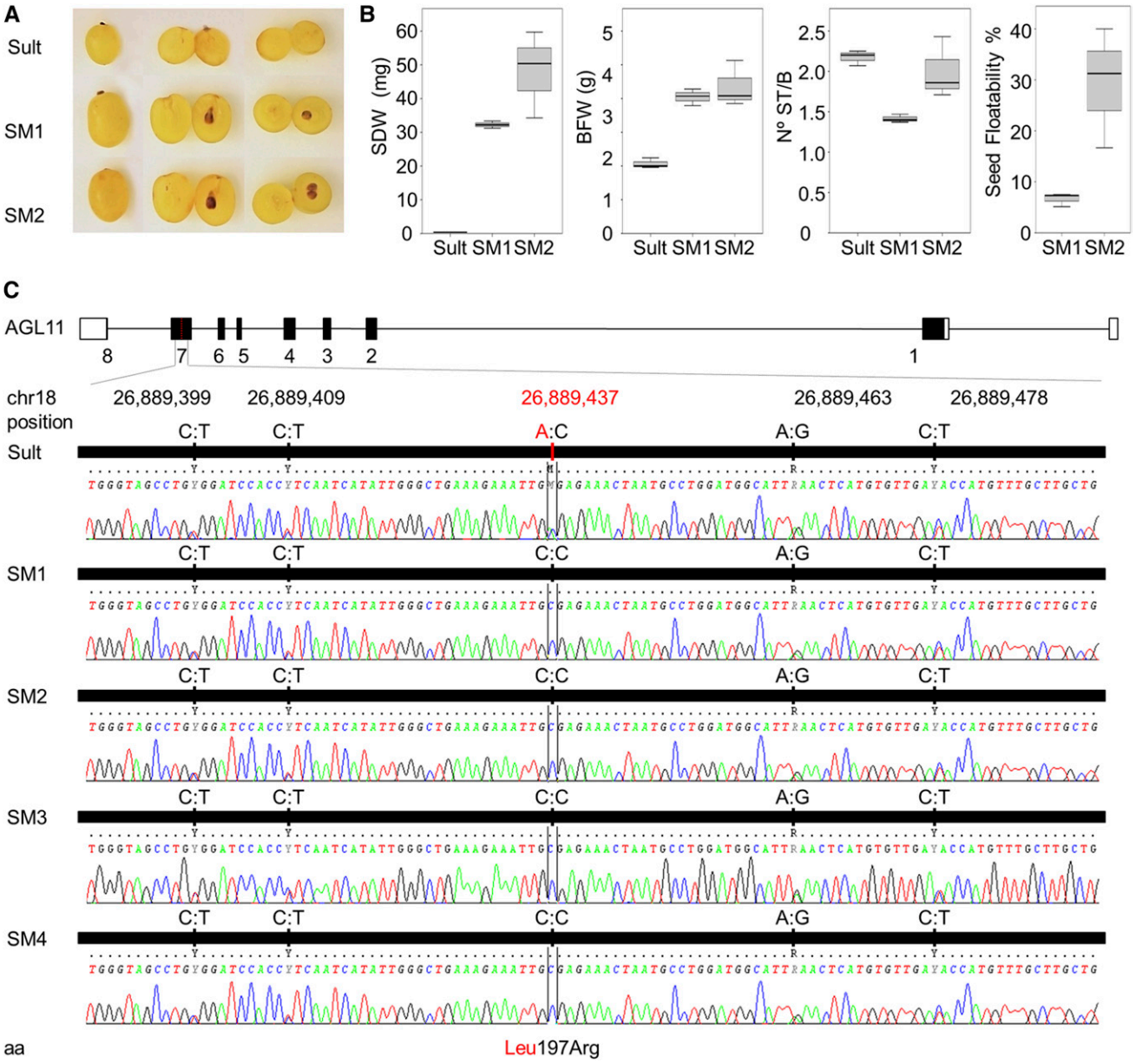



**Figure 6.** Concurrent somatic variation of the *sdi* mutation and seedlessness trait in 'Sultanina Monococco'. A, Images of representative fruits in stenospermocarpic 'Sultanina' (Sult) and in two seeded postzygotic variants of 'Sultanina' that are known as 'Sultanina Monococco' (SM1 and SM2). B, Phenotype characterization of 'Sultanina' somatic variants. SDW, berry fresh weight (BFW), number of seeds or traces per berry (N° ST/B), and seed floatability are presented for 'Sultanina', SM1, and SM2 accessions. C, Postzygotic variation in the SNVs determining the Arg-197Leu *sdi* missense substitution in four 'Sultanina Monococco' accessions (SM1–SM4). Data were obtained after capillary electrophoresis Sanger sequencing of a specific amplicon. aa, Amino acids.

Table S7). Similarly, the poly-GA insertion candidate allele (8×TC according to the sequence in the plus strand) proposed in intron 1 (Di Genova et al., 2014) was detected in 17 seeded cultivars. In the putative promoter, the AG and GA *sdi*+ alleles of VMC7F2 and p3_VvAGL11 markers, respectively, were detected in 44 and 33 seeded cultivars (Fig. 5A; Supplemental Table S7), which does not support the role in seedlessness determination proposed by Mejía et al. (2011).

To validate the association of the Arg-197Leu substitution in VviAGL11 with stenospermocarpic seedlessness, 16 stenospermocarpic cultivars and one seeded cultivar that were not included in the targeted NGS were specifically genotyped following PCR and Sanger sequencing. The full association for position chr18:26,889,437 (A:C) was again confirmed in these accessions (Supplemental Table S6). This study included 'Asyl Kara', a seeded cultivar that, in our analysis, was

homozygous for the seeded allele (C:C) in that position, despite being described as A:A homozygous in a previous study (Mejía et al., 2011). 'Asyl Kara' was genotyped independently several times for this position, confirming the C:C genotype. The accession of 'Asyl Kara' used at the Instituto de Ciencias de la Vid y del Vino (ICVV) collection comes from the Vassal-Montpellier collection, and following microsatellite marker genotyping, we confirmed in the same DNA used for *VviAGL11* sequencing its true-to-typeness with the genetic profile indicated for 'Asyl Kara' in the Vitis International Variety Catalogue database (Supplemental Table S8; www.vivc.de). 'Asyl Kara' also was homozygous for the seeded allele (T:T) at position chr18:26,889,399 (Supplemental Table S6). Again, this genotype contrasts with that reported by Mejía et al. (2011), suggesting a likely sample mistaking by those authors. In summary, the putative deleterious Arg-197Leu substitution in *VviAGL11* is the only detected mutation within the *SDI* fine-mapping interval that is fully linked to the stenospermocarpy phenotype with no false detection in our comprehensive approach, which strongly suggests that it could be the *sdi* causal mutation.

### Concurrent Somatic Variation in the SNV Resulting in the *VviAGL11* Arg-197Leu Substitution and Seedlessness Trait Supports a Causal Effect

Somatic or postzygotic mutants occasionally appear during the characteristic vegetative propagation of grapevine cultivars (Torregrosa et al., 2011). In that manner, seeded somatic variants of the seedless cultivar 'Sultanina' have been reported previously and are known as 'Sultanina Monococco' (Adam-Blondon et al., 2001; Torregrosa et al., 2011; Ocarez and Mejía, 2016). We compared postzygotic variants of 'Sultanina' differing in their ability to develop seeds to evaluate the dominant effect of candidate *sdi* mutations.

We confirmed the seeded variant phenotype in two 'Sultanina Monococco' accessions (Fig. 6A). Mean SDW increased by more than 30 mg compared with a reference stenospermocarpic 'Sultanina' (Fig. 6B), reaching values comparable to those of seeded individuals in other genetic backgrounds (Fig. 1; Supplemental Table S1). Seeds produced by 'Sultanina Monococco' were generally filled, and most of them sank in water (Fig. 6B). Moreover, 59.9% germination success under regular soil conditions was obtained for 'Sultanina Monococco' (2085Mpt1 accession) self-cross filled seeds, confirming that they are functional. A similar seeded variant phenotype was confirmed visually in two other available accessions of 'Sultanina Monococco'. Genotyping of 13 microsatellite markers confirmed the identities of the four studied 'Sultanina Monococco' accessions as postzygotic variants of 'Sultanina' (Supplemental Table S8).

Amplicons containing the three candidate deleterious SNVs detected by RNA-seq in *VviAGL11* and *VviPPAT2* were sequenced in both seedless and seeded postzygotic variants of 'Sultanina' (Supplemental Table S6). Remarkably, the genotype obtained for the four 'Sultanina Monococco' accessions was identical to that in 'Sultanina', with the exception of the *VviAGL11* SNV at position chr18:26,889,437 (Fig. 6C; Supplemental Table S6), which, in fact, is the only polymorphism that was fully associated with seedlessness in a large collection of cultivars (Fig. 5A). At this position, the A:C genotype characteristic of 'Sultanina' and every other stenospermocarpic cultivar analyzed here was changed to C:C in all 'Sultanina Monococco' accessions. Altogether, given the unlikely casual coincidence in somatic variants, the concurrent postzygotic variation in both the *VviAGL11* missense SNV and seedlessness phenotype that we identified in 'Sultanina Monococco' is consistent genetic proof for the causal effect of this mutation. Accordingly, we postulate that the Arg-197Leu missense substitution in *VviAGL11* is the mutation responsible for stenospermocarpic seedlessness.

## DISCUSSION

### Multiple Lines of Evidence Converge on a *VviAGL11* Missense Mutation as the Cause of Stenospermocarpy

In this study, we carried out genome-wide approaches, which were devised independently of previous hypotheses, in an unbiased attempt to identify the mutation responsible for stenospermocarpic seedlessness in grapevine. Our data converge on a missense SNV resulting in an Arg-197Leu substitution in *VviAGL11* as the *sdi* mutation that is exploited worldwide for the production of commercial seedless grapes.

In the first place, crossover mapping in F1 recombinants allowed us to delimit the *sdi* mutation to a 323-kb segment (Fig. 2). These recombinants involve physical/genetic distance ratios between 166 and 555 kb/cM in this chromosomal region, depending on the mapping population (Fig. 2). These results are in line with average ratios between 198 and 390 kb/cM estimated in genetic maps produced in other grapevine crosses (Doligez et al., 2006; Houel et al., 2015; Teh et al., 2017); therefore, they do not support the recombination hotspot proposed by Mejía et al. (2011) in the vicinity of the *SDI* locus. Crossover mapping would be required to validate the smaller 92-kb confidence interval considered by those authors for the screening of the *SDI* gene.

In the second place, the RNA-seq, allele-specific RT-qPCR, and genotyping studies conducted here do not support the presence of cis-acting misexpression mutations within the *sdi* fine-mapping interval (Figs. 4 and 5; Supplemental Tables S3, S5, and S7). Specifically for *VviAGL11*, we show that, rather than misexpression, which was related by other authors to the cause of stenospermocarpy, the lower proportion of seed-related tissues in developing stenospermocarpic fruits is probably the origin of the differential expression that is detected for this gene when whole seeded and

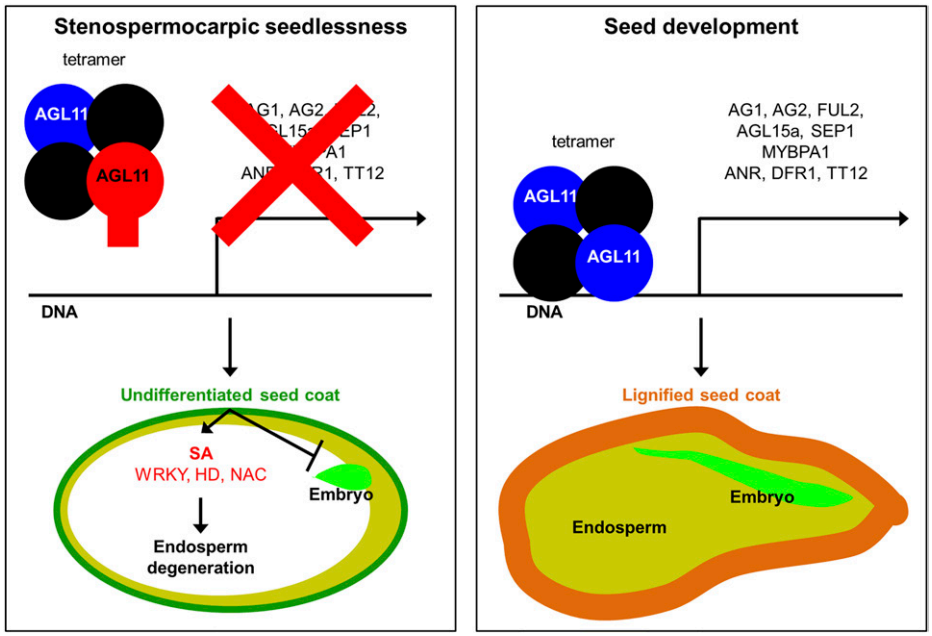



**Figure 7.** Model of grape seedlessness determined by the *sdi* locus. The image at left represents a heterozygous *sdi*+/*sdi*– individual in which the *sdi* Arg-197Leu substitution prevents the activation of gene expression by multiprotein complexes containing the mutant VviAGL11 protein (depicted in red). This lack of gene expression activation precludes seed coat differentiation, which triggers SA production along with the overexpression of WRKY, Homeobox domain (HD), and NAC transcription factors, and finally leads to endosperm degeneration and embryo developmental arrest in seed traces. The image at right depicts the activation of seed morphogenesis that takes place in seeded *sdi*–/*sdi*– individuals. Complexes containing the wild-type VviAGL11 protein (depicted in blue) directly or indirectly activate the expression of genes involved in seed coat sclerification, which enables endosperm maintenance and embryo development.

seedless fruits are compared (Mejía et al., 2011; Ocarez and Mejía, 2016; Malabarba et al., 2017). The relatively high expression of *VviAGL11* that we detected in seed traces of 'Sultanina' (Fig. 4D) also is contradictory to the absence of expression that Malabarba et al. (2017) reported in sections of developing flowers and fruits of this cultivar following in situ hybridization. In agreement with our results, some degree of *VviAGL11* expression has repeatedly been detected in developing flowers and fruits of *sdi*+ accessions using diverse approaches (Mejía et al., 2011; Wang et al., 2015, 2016; Ocarez and Mejía, 2016; Malabarba et al., 2017). Considering the fact that 80% of the probe used by Malabarba et al. (2017) for the in situ hybridization does not match with *VviAGL11* transcripts synthesized by 'Sultanina' but with a specific allele of 'Pinot Noir', deficient probe-transcript hybridization instead of a lack of expression would be a likely cause for the absence of signal in 'Sultanina' in that assay.

In line with the inconsistent difference in *VviAGL11* expression between seeds and seed traces (Fig. 4; Supplemental Table S3), candidate *sdi* polymorphisms that were proposed previously in regulatory regions of *VviAGL11* (Mejía et al., 2011; Di Genova et al., 2014; Ocarez and Mejía, 2016) can be specifically discarded according to the dominant nature of the *sdi* mutation and the presence in seeded accessions of the alleles linked in the *sdi*+ haplotype (Fig. 5A; Supplemental Table S7; Mejía et al., 2011; Karaagac et al., 2012; Bergamini et al., 2013). Importantly, irrespective of the presence of polymorphisms in regulatory regions of *VviAGL11*, we found a lack of allelic expression imbalance in seed traces of *sdi*+/*sdi*– heterozygous seedless individuals (Fig. 4), which rules out the presence of cis-acting misexpression mutations in the *sdi*+ haplotype of this gene. In the absence of allelic imbalance, mitotic recombination between the two alleles of 'Sultanina' could not recover *VviAGL11* expression, as was suggested previously for the origin of 'Sultanina Monococco' somatic variants (Ocarez and Mejía, 2016).

Finally, after RNA-seq screening of coding sequences within the *sdi* mapping interval (Table II), the SNV causing the Arg-197Leu substitution in *VviAGL11* was the only predicted deleterious variant in which a full linkage with the seedlessness phenotype was confirmed in a collection of grapevine cultivars (Fig. 5A). Although this missense variation was discarded previously according to its homozygous presence in the seeded 'Asyl Kara' (Mejía et al., 2011), in our study, this cultivar was homozygous for the Arg codon, similar to all other 106 seeded cultivars analyzed (Supplemental Tables S6 and S7). Remarkably, as consistent genetic proof, this polymorphism is the only candidate mutation that is not detected in 'Sultanina Monococco' seeded

somatic variants of the original seedless cv 'Sultanina' (Fig. 6). Given that unlikely coincidence in near-isogenic somatic variants, it can be assumed that somatic variations (either genetic or cellular) that altered the genotype of the missense SNV are the cause of the polymorphic seed development phenotype displayed by postzygotic variants of 'Sultanina'. This causality proof, together with all the other genetic and molecular evidence described above, led us to postulate that *VviAGL11* is the major dominant regulator gene and that the SNV resulting in the Arg-197Leu substitution is the monogenic *sdi* mutation that, two decades ago, were proposed for the main origin of seedlessness in grapevine (Bouquet and Danglot, 1996; Lahogue et al., 1998).

### Molecular Determination of Stenospermocarpy by the *sdi* Missense Mutation

In agreement with the absence of the *VviAGL11* missense mutation in seeded somatic variants of 'Sultanina' (Fig. 6C), *AGL11* homologs participate in the development of maternal seed tissues that is triggered after fertilization in other plant species (Colombo et al., 1997; Mizzotti et al., 2014; Malabarba et al., 2017). Resembling the phenotype of grape stenospermocarpy (Pearson, 1932; Barritt, 1970; Pratt, 1971; Striem et al., 1992; Malabarba et al., 2017), cosuppression of *FLORAL BINDING PROTEIN7* (*FBP7*) and *FBP11*, two *AGL11* homologs present in petunia (*Petunia hybrida*), caused the development of shrunken seeds that lack seed coat sclerification with a consequent degeneration of the endosperm, while the embryo frequently remained viable (Colombo et al., 1997). As reported for *FBP7* and *FBP11* transcripts in wild-type petunia or AGL11/STK protein in Arabidopsis (Colombo et al., 1997; Mizzotti et al., 2014), *VviAGL11* is transcribed specifically in the developing seed coat in 'Pinot Noir' seeded grape (Malabarba et al., 2017). These coincidences suggest that *AGL11* function is impaired in grapevines carrying the Arg-197Leu substitution.

Resembling the Arg-197Leu substitution identified in grapevine (Table II; Figs. 5 and 6), the variation of endocarp lignification in different accessions of oil palm (*Elaeis guineensis*) is associated with five independent missense substitutions in the MADS-box domain of *SHELL*, an oil palm homolog of *AGL11* (Singh et al., 2013; Ooi et al., 2016). The Arg-197Leu substitution in VviAGL11 is located in the C domain characteristic of MIKC-type MADS-domain proteins (Fig. 5B). This domain participates in the binding activity of MIKC tetramers to DNA and also may determine protein-protein interaction specificity (Honma and Goto, 2001; Melzer et al., 2009; van Dijk et al., 2010). According to this function, and considering also that VviAGL11-SEPALLATA heterodimers interact with additional VviAGL11 units (Mellway and Lund, 2013), the Arg-197Leu substitution might cause dominant effects by disrupting the action of VviAGL11-containing multiprotein complexes, even in the presence of the seeded allele in heterozygous individuals, as postulated for mutant alleles of SHELL (Singh et al., 2013, 2015). This mechanism would be compatible with the incomplete dominance of the *sdi* mutation that can be inferred in view of the more extreme seedless phenotype that often is displayed by *sdi*+/*sdi*+ homozygous individuals (Mejía et al., 2011; Ocarez and Mejía, 2016).

In spite of genetic proofs leading to the missense substitution in *VviAGL11* (Table II; Figs. 5 and 6) and all the other indirect evidence mentioned above, a heterologous assay in Arabidopsis was inconclusive about the role in seedlessness determination of the VviAGL11 protein encoded by the *sdi*+ haplotype (Malabarba et al., 2017). However, species-specific functions of AGL11 homologs are conceivable, since, contrasting with cosuppressed petunia lines and *sdi* stenospermocarpy, the Arabidopsis *stk* loss-of-function mutant overaccumulates polyphenols in the seed coat in the absence of endosperm degeneration or embryo arrest (Mizzotti et al., 2014). In fact, while 28 DEGs between *sdi*+ and *sdi*− RG×CS offspring were the closest grapevine homologs to any out of the 248 deregulated genes in the *stk* mutant (Mizzotti et al., 2014), 71% of these DEGs showed inverted expression response compared with the experiment in Arabidopsis (Supplemental Table S3). Genes involved in flavonoid accumulation were up-regulated in seeds of the *stk* mutant, including *DIHYDROFLAVONOL 4-REDUCTASE/TRANSPARENT TESTA3* (*DFR/TT3*), *BANYULS/ANTHOCYANIDIN REDUCTASE* (*BAN/ANR*), and *TT12*. In contrast, the grapevine homologs *VviDFR1* (*VIT_18s0001g12800*), *VviANR* (*VIT_00s0361g00040*), and *VviTT12* (*VIT_12s0028g01150*), respectively, together with the master regulator of proanthocyanidin accumulation in grape, *MYB PROANTHOCYANIDIN1* (*MYBPA1*, *VIT_15s0046g00170*; Bogs et al., 2007), were down-regulated in *sdi*+ seed traces (Supplemental Table S3), which, in turn, accumulate less polyphenols than *sdi*− grapevine seeds (Merin et al., 1983). In line with these inverted effects, a reversible role of AGL11/STK in the control of tissue lignification has been shown in Arabidopsis, depending on the presence of transcriptional cosuppressor partners (Balanzà et al., 2016). We thereby assume that functional studies in genetically engineered vines are required to ensure the presence of grapevine-specific protein partners. While functional studies of fruit traits take a long time in species with long juvenile phases, such as grapevine, we are conducting additional research in this direction to demonstrate the consequences of the Arg-197Leu substitution for VviAGL11 protein function and the determination of seed development abortion.

At this point, according to all the evidence described above, we suggest a working model in which the *sdi* missense substitution identified here disrupts the function of multimeric complexes containing VviAGL11 proteins (Fig. 7). This would prevent proper seed coat development and lignification (Malabarba et al., 2017), which is required for nutrient and signaling flux between maternal tissues and fertilization

products (Mizzotti et al., 2012; Figueiredo and Köhler, 2014). Apparently, this impaired flux in developing seeds of stenospermocarpic grapes leads to SA-dependent autoimmune responses (Fig. 3) and to endosperm degeneration and embryo development arrest (Pratt, 1971), which, at the transcriptome level, associate with the repression of secondary metabolism and MADS-box genes and the activation of photosynthesis-related genes (Fig. 3; Supplemental Tables S3 and S4). These molecular responses are very similar to those reported during SA-dependent seed abortion in incompatible Arabidopsis interspecific hybrids that, like *sdi*-determined stenospermocarpy, proceed with abnormal seed coat development (Burkart-Waco et al., 2013).

## CONCLUSION

Diverse approaches converge on the SNV causing the Arg-197Leu substitution in *VviAGL11* as the seedlessness-responsible mutation underlying the grapevine *SDI* locus. While molecular and expression assays rejected the hypothesis of cis-acting misexpression mutations, the deleterious effect predicted for this amino acid change, the full association between the variant allele and stenospermocarpy in a collection of grapevine cultivars, and the concurrent post-zygotic variation of both the missense polymorphism and the seedlessness phenotype in somatic variants of 'Sultanina' consistently point to this SNV as the causal mutation. This finding, together with previous discoveries in oil palm (Singh et al., 2013), show that variation in the amino acid sequence of AGL11 homologs has been selected in parallel during the domestication of distant monocot and dicot crops to control the proportion between pericarp and seed tissues of fruits as well as their level of sclerification. While seedlessness is the most relevant fruit feature for the greatly active table grape breeding field, our discovery enables the development of the most efficient marker-assisted selection to track the major locus controlling this trait. The simultaneous tracking of the *sdi* mutation and of minor *SDI*-independent loci controlling the degree of lignification and the size and number of seed traces will optimize the generation of new seedless cultivars in future table grape breeding programs. Remarkably, knowledge of the *sdi* mutation also can be exploited to introduce stenospermocarpy into other genetic backgrounds through controlled genome editing.

## MATERIALS AND METHODS

### Plant Material

Two table grape (*Vitis vinifera*) F1 mapping populations (RG×CS [$n$ = 292] and IN×CS [$n$ = 299]) were generated from controlled crosses carried out in 2003 and 2008, respectively. For these crosses, emasculated flowers of RG and IN seeded cultivars were pollinated with pollen collected from CS that, in both populations, was the donor of the *sdi* mutation. One plant for each F1 individual and three plants for each progenitor were grown in the same experimental field belonging to the Sociedad Murciana de Investigación y Tecnología de Uva de Mesa (ITUM) located in Blanca, Murcia, Spain. Plants were grafted onto 1,103 Paulsen rootstocks, conducted under parral trellis, ferti-irrigated, and cultivated under the same management practices.

For the study of sequence variation at selected loci, 124 grapevine cultivars and two RS×MO F1 individuals were used (Supplemental Tables S6 and S7). Samples of 120 cultivars were obtained from the Grapevine Germplasm Collection of the ICVV (ESP-217). At ICVV, 10 plants per cultivar were maintained under the same agronomical conditions in Finca La Grajera, located in Logrono, La Rioja, Spain. These plants were grafted in 2010 onto 110 Richter rootstocks, and their true-to-typeness was verified by genotyping SNP markers and comparing the results with the ICVV SNP database (Cabezas et al., 2011). RG and CS samples were obtained from ITUM, whereas RS×MO F1 individuals and MO were obtained and cultivated as described elsewhere (Carreño et al., 2015). Material for 'Chasselas Doré' was obtained from plants, grafted onto 110 Richter rootstocks in 2003, in the Vitis Germplasm Bank (ESP-080) of the Instituto Madrileño de Investigación y Desarrollo Rural, Agrario y Alimentario, located in Finca El Encín in Alcalá de Henares, Madrid, Spain. In addition to the previous 126 accessions, materials for four 'Sultanina Monococco' accessions (2085Mpt1, 2594Mtp2, 2594Mtp1, and 2777Mtp1; SM1–SM4, respectively) and for a second reference accession of 'Sultanina' (1566Mtp2), all of which were confirmed previously as clonal lines (Laucou et al., 2011), were obtained from the collection of the Institut National de la Recherche Agronomique Centre de Ressources Biologiques de la Vigne at Vassal-Montpellier (FRA-139), Marseillan-Plage, France. SM1 to SM4 accessions entered the Vassal-Montpellier repository in 1966, 1959, 1966, and 1975 from Cyprus, Turkey, Iran, and Bulgaria, respectively (https://bioweb.supagro.inra.fr/collections_vigne).

### Seedlessness Trait Assessment

For RG×CS F1 progeny and progenitors, seed content was evaluated at harvest during three years (2007–2009). For each individual and year, 20 fruits randomly collected from all the clusters produced by the plant were analyzed to estimate seed or seed trace number per berry and SDW. SDW was measured after 48 h of incubation of seeds and seed traces in an 80°C oven.

The seedlessness phenotype also was de visu registered as a qualitative trait for every RG×CS F1 individual in the same three seasons. For recombinants in the *SDI* mapping interval, de visu determination of seedlessness was conducted in 2012 and 2013 as well. To this end, seeded and seedless individuals were classified according to the presence of hard seeds with totally sclerified integuments or soft seed traces with unsclerified or partially sclerified integuments, respectively, as described previously (Costantini et al., 2008). Similar de visu classification of the seedlessness trait was carried out at least twice (between the 2013 and 2016 seasons) for the IN×CS F1 mapping population and for the 120 cultivars from the ICVV collection used for genotyping of candidate genes (Supplemental Tables S1, S6, and S7). At least 20 berries from two to three representative clusters from the same plant were inspected each time. Furthermore, seed content phenotypes in 'Sultanina' variant lines from the Vassal-Montpellier collection were measured quantitatively in the 2017 season following the same procedures described for RG×CS offspring.

### DNA Extraction

DNA was extracted from young leaves. For RG×CS progeny, DNA was extracted using the DNeasy Plant Mini Kit (Qiagen) and modified by adding 1% (w/v) PVP40 to the AP1 buffer. The BioSprint 96 DNA Plant Kit (Qiagen) was used for IN×CS progeny following the manufacturer's instruction. For all other accessions, the DNeasy Plant Mini Kit (Qiagen) was used without modifications.

### QTL Analysis

#### *Marker Genotyping*

RG, CS, and RG×CS F1 individuals were genotyped for a set of 223 SSR markers mapped previously on the grapevine genome (Doligez et al., 2002; Cabezas et al., 2006; Costantini et al., 2008), 178 of which showed segregation types suitable for genetic mapping. Primers were modified at the 3′ end with a fluorochrome (NED, 6-FAM, VIC, or PET) by Applied Biosystems. PCR amplifications were performed with Ecogen Taq polymerase. PCR products were analyzed with ABI Prism 3730 (Applied Biosystems) at the Unidad de Genómica-Campus Moncloa del Parque Científico de Madrid. Detection of amplified

fragments and allele assignment were carried out using GeneMapper version 3.7 (Applied Biosystems). Additionally, F1 individuals and progenitors were genotyped for 335 SNP markers described by Lijavetzky et al. (2007) and Cabezas et al. (2011), 112 of which showed segregation types suitable for genetic mapping and were analyzed as described previously (Cabezas et al., 2011).

### *Genetic Map Construction*

Genetic maps were built according to a two-way pseudo-testcross strategy (Grattapaglia and Sederoff, 1994) using JoinMap 3.0 (van Ooijen and Voorrips, 2001). Doubled Haploid and Cross Pollinators models were assumed for the construction of parental and consensus maps, respectively. The global segregation of each marker was tested for the expected Mendelian segregation ratio using a $\chi^2$ goodness-of-fit test ($P < 0.01$). Markers were grouped in the same LG according to the threshold LOD ≥ 5 and ordered by paired markers with LOD ≥ 3 and recombination frequency ≤ 0.35. These parameters were reduced to LOD ≥ 1 and recombination frequency ≤ 0.45 when required to adapt LGs to current information on the physical genome map. Mapping distances were calculated (in cM) using the Kosambi function (Kosambi, 1944).

### *QTL Detection*

QTL analyses were carried out on RG×CS mapping progeny independently for each of the three SDW evaluations (seasons 2007, 2008, and 2009) using the framework maps of each progenitor and the consensus framework map for the cross. QTL mapping was performed using MapQTL 4.0 (van Ooijen et al., 2002). Two parametric methods, Interval Mapping (Lander and Botstein, 1989; van Ooijen, 1992) and Multiple QTL Mapping (Jansen and Stam, 1994), were used. LOD score thresholds of 0.99 significance for each LG and 0.95 for the whole genome were established through 1,000 permutations for each trait (Churchill and Doerge, 1994). QTL position was estimated from the location of the maximum LOD value and a minimum confidence interval (minimum overlapping interval for one LOD interval independently detected for each year of analysis on each QTL and map).

## Fine-Mapping of the *SDI* QTL

RG×CS F1 individuals were screened for meiotic recombination around the *SDI* QTL according to the genotype at the microsatellite locus VVIN16 (Merdinoglu et al., 2005) and CAPS-28.53. This CAPS was designed for a *Bst*BI (New England Biolabs) restriction enzyme target specific of the *sdi+* haplotype in this progeny (Supplemental Table S2). Primer pairs used to obtain this and other target amplicons are described in Supplemental Table S9. PCR primers were designed with the NCBI Primer BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) using the grapevine 12X.0 reference genome assembly sequence as a template (https://urgi.versailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences). Similarly, recombinant individuals within the fine-mapping interval delimited in RG×CS F1 progeny were searched for in IN×CS F1 progeny by genotyping the CAPS-26.39 and CAPS-26.88 markers. These markers were designed to take advantage of the presence of *Blp*I and *Mse*I (New England Biolabs) restriction enzyme targets, respectively, that are specific to the *sdi+* haplotype in this F1 mapping population. Crossovers were mapped in recombinant F1 individuals by genotyping the SNP markers indicated in Supplemental Table S2, which were developed and genotyped through PCR amplification and Sanger capillary electrophoresis sequencing. In most cases, primers were designed to align in coding sequences according to grapevine 12X V1 gene annotations (http://genomes.cribi.unipd.it/). For PCR, KAPA2G Fast DNA Polymerase (KAPA Biosystems) was used for amplification from ~50 ng of gDNA. Amplification products were purified with ExoSAP-IT (USB Products Affymetrix) following the manufacturer's instructions and then sequenced by Sanger capillary electrophoresis using the same primers as in PCR. For every marker, the allele present in the *sdi+* haplotype was inferred according to the genotype observed in cross progenitors, F1 individuals with extreme seed/seedless phenotype, and the two RS×MO *sdi* homozygous F1 individuals (Supplemental Table S2).

## Transcriptome Analysis

### *RNA Extraction*

Total RNA was extracted from frozen tissue powder using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) with an added on-column DNase digestion step with the RNase-Free DNase Set (Qiagen).

### *Microarray Analysis*

Developing fruits at the fruit set stage (~2 WAF, flowering time May 20, 2010) were collected on June 3, 2010, around midday, from four seeded (660-43, 660-50, 669-45, and 671-19) and five seedless (660-40, 660-41, 660-57, 669-19, and 671-34) RG×CS F1 individuals and immediately frozen in liquid nitrogen. Individuals were selected according to contrasting genotypes in the VMC7F2 marker (Pellerone et al., 2001), which colocalizes with the *SDI* QTL (Table I), and to extreme phenotypes for SDW in the corresponding progeny distribution tail (Supplemental Table S1). Each individual was analyzed separately as an independent biological replicate. At least 12 developing fruits collected from two or more different clusters were ground for each RNA extraction. Analysis of RNA integrity as well as synthesis, labeling, and hybridization of cDNA to NimbleGen microarray 090818 Vitus exp HX12 (NimbleGen-Roche) and robust multiarray average normalization were performed as indicated elsewhere (Carbonell-Bejerano et al., 2014). Linear models for microarray data (limma) were run in Babelomics (Medina et al., 2010) to search for DEGs between seedless and seeded individuals. DEGs were identified considering a Benjamini-Hochberg adjusted $P \leq 0.05$ and 2-fold or greater change as significance thresholds.

### *RNA-Seq and Differential Gene Expression Analysis*

RNA was obtained from seed traces of three seedless (669-60, 670-17, and 675-17) and seeds from three seeded (665-17, 671-19, and 677-06) RG×CS F1 individuals. These individuals were selected according to contrasting genotypes for the *SDI* QTL marker VMC7F2 (Supplemental Table S1). Selected seeded individuals belonged to the higher tail of the distribution for SDW phenotype in the progeny, whereas seedless individuals without extremely low SDW were selected to ease the extraction of seed traces. Berries at the pea-size (10–12 mm) developmental stage were collected on June 12, 2012 (~4 WAF, flowering time May 18, 2012), around midday, and immediately frozen in liquid nitrogen. Berries were allowed to thaw briefly in the laboratory and opened by cutting with a scalpel, and then seed content was recovered rapidly with tweezers and refrozen in nitrogen. Each individual was analyzed separately as an independent biological replicate. Seed content from 12 to 50 fruits collected from two or more different clusters was used for each RNA extraction. RNA-seq was performed at the Centre for Genomic Regulation in Barcelona, Spain. The six corresponding cDNA libraries were prepared using the Illumina TruSeq Stranded mRNA Sample Prep kit starting from 1 µg of total RNA, as described previously (Royo et al., 2016). A mean fragment size of 302 bp was obtained. Library sequencing was performed on an Illumina HiSeq 2000 using v4 chemistry (flow cells and sequencing reagents). Paired-end strand-specific reads of 125 nucleotides were produced. Gapped alignment of reads to the PN40024 12X.0 grapevine reference genome assembly (http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/) was carried out using TopHat2 version 2.0.13 (Kim et al., 2013). TopHat2 was run under default parameters with the exception of gap length ≤ 8, mismatch ≤ 8, and edit distance ≤ 8, which were allowed in 125-nucleotide reads. These values were adjusted to compensate for the relatively high genetic distance between the table grape genotypes studied here and the reference genome of the wine grape-related PN40024 reference genome. As postfiltering, only uniquely mapped single-copy reads with quality ≥ 20, aligned in the same chromosome and paired in the expected orientation, were kept for further analysis. After filtering, an average of more than 15 million reads per replicate was considered for subsequent analysis (Supplemental Table S10). The htseq-count tool (version 0.5.4p5) from HTSeq (Anders et al., 2015) was used to estimate unambiguous read count for each 12X V1 annotated transcript. Normalization following the Trimmed Mean of M-values method (Robinson and Oshlack, 2010), as well as seed trace versus seed DEGs search (adjusted Benjamini-Hochberg $P \leq 0.05$ and 2-fold or greater change), were performed in edgeR version 2.2.6 (Robinson et al., 2010). Finally, reads per kilobase of exon per million fragments mapped was calculated using edgeR, and low-expressed transcripts were filtered out if average reads per kilobase of exon per million fragments mapped was less than one in both seed traces and seeds.

### *Functional Analysis of DEGs*

Lists of DEGs identified from microarray and RNA-seq analyses were compared using Venny (http://bioinfogp.cnb.csic.es/tools/venny/). Gene lists were analyzed further for functional enrichment compared with the whole set of transcripts predicted in the 12X V1 annotation of the grapevine reference

genome following a grapevine-specific functional classification (Grimplet et al., 2012). The analysis was carried out in FatiGO as described elsewhere (Carbonell-Bejerano et al., 2014).

### *Search of RNA-Seq Data for Candidate Allelic Imbalance*

Aligned reads within the *sdi* fine-mapped interval were inspected visually for allele-specific expression with the Integrative Genomics Viewer (IGV) software (Thorvaldsdóttir et al., 2013). For *VviAGL11*, the allelic ratio in each sample was estimated by comparing the frequency of each allele at heterozygous SNP positions in filtered Binary Alignment Map files. A mean value was estimated for all the SNPs detected along the gene according to the parameters indicated in the following paragraph. The *sdi*+ haplotype was inferred from the comparison of two seedless RG×CS F1 individuals that shared the same allele inherited from RG for the *SDI* locus with the three seeded individuals analyzed by RNA-seq (Supplemental Table S1).

### *Search of RNA-Seq Data for Candidate Sequence Variation*

To detect sequence variation specific to the *sdi*+ haplotype in expressed transcripts, RNA-seq alignments used for differential expression analysis also were analyzed for the presence of SNPs and INDELs, basically following a similar pipeline to that described previously (Royo et al., 2016) but adjusted to boost variant calling in both low-expressed and allelic-imbalanced genes. For variant calling, three tools implemented in SAMtools package version 1.5 (Li et al., 2009; Li, 2011) were used to compare each sample with the PN40024 reference genome: samtools mpileup for genotype probability estimation, bcftools for variant calling, and finally, strand bias and baseQ bias filters were applied using varfilter. Subsequently, using ad hoc Bash shell and Perl scripts (Royo et al., 2016), we selected polymorphisms within the *SDI* fine-mapping interval specific to the three seedless individuals according to the following filters: average depth of contrasting alleles (variant or reference) per F1 individual ≥ 5 counts, frequency of variant allele in *sdi*+ individuals ≥ 20%, frequency of variant allele in seeded individuals < 2.5%, and frequency of spurious alleles (alleles other than the reference and the first variant) < 2.5% in both samples. In this case, reads corresponding to each transcription strand direction were computed independently. The effect of detected candidate *sdi* mutations considering grapevine 12X V1 gene annotations was estimated using SnpEff version 2.0.3 (Cingolani et al., 2012), whereas the effect of *sdi* candidate amino acid substitutions on protein function was predicted using PROVEAN (Choi et al., 2012).

## Hormone Analysis

To compare phytohormone levels between developing seeds and seed traces, five seedless (665-16, 669-60, 670-17, 675-17, and 660-57) and five seeded (660-43, 660-50, 665-17, 671-19, and 677-06) RG×CS F1 individuals were analyzed (Supplemental Table S1). Additionally, two independent biological replicates of RG and CS, each from a different plant, were analyzed as well. Fruits at the pea-size stage (~4 WAF) were collected simultaneously with RNA-seq samples on June 12, 2012, and immediately frozen in liquid nitrogen. For each replicate, seed content from 20 or more fruits was extracted. SA and active GA ($GA_1$ and $GA_4$) levels were measured at the Plant Hormone Quantification Service of the Institute for Plant Molecular and Cell Biology in Valencia, Spain. In summary, 200 mg fresh weight of seeds or seed traces was extracted for each replicate, as indicated above for RNA-seq analysis. Frozen ground tissue was then analyzed using an ultra-performance liquid chromatography-tandem mass spectrometry system (Q-Exactive; Thermo Fisher Scientific). The statistical treatment of results was carried out using SPSS software (version 24.0 for Windows; IBM).

## RT-qPCR

RNA was obtained from berry skin, flesh, pericarp, seeds, or seed traces. Pea-size fruits collected from the ITUM experimental field on June 26, 2011, were used for three RG×CS F1 individuals (660-41, 660-43, and 660-57) and RG and CS progenitor accessions. 'Sultanina' samples from the ESP217-5186 accession were collected at fruit set (June 6, 2011) in the ICVV collection. At least eight fruits were used for each RNA extraction. For reverse transcription from total RNA (1 µg), SuperScript III First Strand (Invitrogen) and oligo(dT) were used following the manufacturer's instructions. Transcript levels were determined by RT-qPCR using a 7500 Real-Time PCR System (Applied Biosystems) and SYBR Green PCR Master Mix (Applied Biosystems). Reactions were performed in a final volume of 20 µL with 5 µL of a 1:10 dilution of cDNA. Gene-specific primers (Supplemental Table S9) were designed using the Oligo Explorer 1.2 software (Gene Link) and the gene sequences from the grapevine 12X.0 genome assembly (Jaillon et al., 2007) as design templates. For allele-specific RT-qPCR, the 3′ end of forward and reverse primers coincided with SNP positions in a way that each primer was specific for one allele. No-template controls were included for each primer pair, and each PCR was performed in triplicate. Amplification data were analyzed using the 7500 SDS software 1.3 (Applied Biosystems). Relative transcript levels were calculated after normalization to grapevine *UBIQUITIN* (*VIT_16s0098g01190*), *ELONGATION FACTOR1-α* (*VIT_06s0004g03220*), and *GPDH* (*VIT_17s0000g10430*) using the ΔΔCt method. For allele-specific RT-qPCR, two independent biological replicates from different stenospermocarpic 'Sultanina' plants were analyzed, and gDNA from the same accession was used for allelic ratio normalization assuming a balanced ratio in the genome. Graphical representations of relative transcript levels with SD values were created using Microsoft Excel software. The statistical treatment of results was carried out using SPSS software (version 24.0 for Windows; IBM).

## Genotyping of Candidate Mutations in *VviPPAT2*

For the validation of candidate SNVs detected from RNA-seq data, the gene *VviPPAT2* (*VIT_18s0041g01870*) was partially sequenced in a collection of 93 grapevine accessions (Supplemental Table S6). Primer pairs PPAT2-R151C and PPAT2-Q195L were used for PCR and Sanger sequencing to this end (Supplemental Table S9).

## Targeted Resequencing of *VviAGL11*

The gene *VviAGL11* (*VIT_18s0041g01880*) plus a 2-kb upstream sequence (9,849 bp in total, corresponding to positions chr18:26,888,672–chr26,898,521 in the PN40024 12X.0 grapevine reference genome) were sequenced in a collection of 110 accessions (Supplemental Table S7) by BGI, as described by Tello et al. (2016). Briefly, Agilent SureSelect was used to target the *VviAGL11* locus that was then sequenced in 90-nucleotide paired-end reads using Illumina HiSeq 2000. Sequencing data were analyzed as described by Tello et al. (2015): reads were aligned to the PN40024 12X.0 reference genome using Bowtie 2 (Langmead and Salzberg, 2012). Then, the variant calling tool implemented in SAMtools was used to detect SNPs and small INDELs by comparing the sequence of aligned reads in each of the 110 accessions with the 12X.0 PN40024 reference genome. Polymorphisms were initially filtered by means of ad hoc Perl scripts, as described previously (Tello et al., 2015). Significant polymorphisms present and absent in every seedless and seeded accession, respectively, were selected as candidate *sdi* mutations.

Validation by PCR and amplicon sequencing of the *sdi* candidate mutation identified in *VviAGL11* was carried out using the primer pair CAPS-26.88 (Supplemental Table S9) on gDNA of 15 additional seedless cultivars as well as in the seeded 'Asyl Kara' and in two 'Sultanina Monococco' accessions (Supplemental Table S6). Genotyping of 13 microsatellite markers in the same DNA sample used to genotype *sdi* candidate variants by PCR and Sanger sequencing was carried out to verify the identity of 'Asyl Kara' and 'Sultanina' seedless and seeded accessions (Supplemental Table S8). Microsatellite marker genotyping was performed as described elsewhere (Ibáñez et al., 2009b).

## Accession Numbers

Microarray hybridization and RNA-seq data have been deposited in the NCBI Gene Expression Omnibus database under accession numbers GSE106668 and GSE107014, respectively. Binary Alignment Map files for read alignments of *VviAGL11*-targeted NGS data were deposited under BioProject identifier PRJNA418130 (SRP124845) of the NCBI Sequence Read Archive.

## Supplemental Data

The following supplemental materials are available.

**Supplemental Table S1.** Phenotypes and genotypes of offspring in F1 mapping populations.

**Supplemental Table S2.** Genotyping for fine-mapping in seeded × seedless F1 mapping populations.

**Supplemental Table S3.** List of DEGs detected between seedless and seeded RG×CS progeny.

**Supplemental Table S4.** Functional enrichment analysis of DEGs detected between seedless and seeded RG×CS progeny.

**Supplemental Table S5.** List of SNVs linked to the *sdi* haplotype detected from RNA-seq data.

**Supplemental Table S6.** Genotypes of candidate missense substitutions in *VviPPAT2* and *VviAGL11* and seedlessness phenotype in a collection of grapevine cultivars.

**Supplemental Table S7.** Polymorphisms in *VviAGL11* detected by NGS and seedlessness phenotype in a collection of grapevine cultivars.

**Supplemental Table S8.** Microsatellite genotyping identification for 'Sultanina' and 'Asyl Kara' accessions.

**Supplemental Table S9.** Primer pairs used for PCR amplification and sequencing.

**Supplemental Table S10.** RNA-seq library size and normalization factors.

## ACKNOWLEDGMENTS

We thank J.L. Pérez-Sotés and the Servicio de Investigación Vitivinícola (Gobierno de La Rioja) at the ICVV collection and Rosa M. Arnau and María J. Candel at IMIDA and ITUM for plant maintenance and technical support in phenotyping tasks. We are very grateful to Nachi Montemayor, Rufino Aguirrezabal, Miguel Angulo, and Silvia Hernáiz for technical assistance. We also thank the Genomics Service of the CNB-Consejo Superior de Investigaciones Científicas for running microarray hybridizations and Dr. Iván Carreño for sharing information regarding RS×MO F1 progeny.

Received March 30, 2018; accepted May 23, 2018; published May 31, 2018.

## LITERATURE CITED

**Adam-Blondon AF, Lahogue-Esnault F, Bouquet A, Boursiquot JM, This P** (2001) Usefulness of two SCAR markers for marker-assisted selection of seedless grapevine cultivars. Vitis **40**: 147–155

**Anders S, Pyl PT, Huber W** (2015) HTSeq: a Python framework to work with high-throughput sequencing data. Bioinformatics **31**: 166–169

**Balanzà V, Roig-Villanova I, Di Marzo M, Masiero S, Colombo L** (2016) Seed abscission and fruit dehiscence required for seed dispersal rely on similar genetic networks. Development **143**: 3372–3381

**Barritt BH** (1970) Ovule development in seeded and seedless grapes. Vitis **9**: 7–14

**Battilana J, Lorenzi S, Moreira FM, Moreno-Sanz P, Failla O, Emanuelli F, Grando MS** (2013) Linkage mapping and molecular diversity at the flower sex locus in wild and cultivated grapevine reveal a prominent SSR haplotype in hermaphrodite plants. Mol Biotechnol **54**: 1031–1037

**Bergamini C, Cardone MF, Anaclerio A, Perniola R, Pichierri A, Genghi R, Alba V, Forleo LR, Caputo AR, Montemurro C,** (2013) Validation assay of p3_VvAGL11 marker in a wide range of genetic background for early selection of stenospermocarpy in *Vitis vinifera* L. Mol Biotechnol **54**: 1021–1030

**Bogs J, Jaffé FW, Takos AM, Walker AR, Robinson SP** (2007) The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. Plant Physiol **143**: 1347–1361

**Bouquet A, Danglot Y** (1996) Inheritance of seedlessness in grapevine, *Vitis vinifera* L. Vitis **35**: 35–42

**Burkart-Waco D, Ngo K, Dilkes B, Josefsson C, Comai L** (2013) Early disruption of maternal-zygotic interaction and activation of defense-like responses in *Arabidopsis* interspecific crosses. Plant Cell **25**: 2037–2055

**Cabezas JA, Cervera MT, Ruiz-García L, Carreño J, Martínez-Zapater JM** (2006) A genetic analysis of seed and berry weight in grapevine. Genome **49**: 1572–1585

**Cabezas JA, Ibáñez J, Lijavetzky D, Vélez D, Bravo G, Rodríguez V, Carreño I, Jermakow AM, Carreño J, Ruiz-García L,** (2011) A 48 SNP set for grapevine cultivar identification. BMC Plant Biol **11**: 153

**Canaguier A, Grimplet J, Di Gaspero G, Scalabrin S, Duchêne E, Choisne N, Mohellibi N, Guichard C, Rombauts S, Le Clainche I,** (2017) A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). Genom Data **14**: 56–62

**Carbonell-Bejerano P, Rodríguez V, Royo C, Hernáiz S, Moro-González LC, Torres-Viñals M, Martínez-Zapater JM** (2014) Circadian oscillatory transcriptional programs in grapevine ripening fruits. BMC Plant Biol **14**: 78

**Carreño I, Cabezas JA, Martínez-Mora C, Arroyo-García R, Cenis JL, Martínez-Zapater JM, Carreño J, Ruiz-García L** (2015) Quantitative genetic analysis of berry firmness in table grape (*Vitis vinifera* L.). Tree Genet Genomes **11**: 818

**Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A,** (2016) Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods **13**: 1050–1054

**Choi Y, Sims GE, Murphy S, Miller JR, Chan AP** (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS ONE **7**: e46688

**Churchill GA, Doerge RW** (1994) Empirical threshold values for quantitative trait mapping. Genetics **138**: 963–971

**Cingolani P, Platts A, Wang L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM** (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin) **6**: 80–92

**Colombo L, Franken J, Van der Krol AR, Wittich PE, Dons HJ, Angenent GC** (1997) Downregulation of ovule-specific MADS box genes from petunia results in maternally controlled defects in seed development. Plant Cell **9**: 703–715

**Costantini L, Battilana J, Lamaj F, Fanizza G, Grando MS** (2008) Berry and phenology-related traits in grapevine (*Vitis vinifera* L.): from quantitative trait loci to underlying genes. BMC Plant Biol **8**: 38

**Dangi GS, Mendum ML, Prins BH, Walker MA, Meredith CP, Simon CJ** (2001) Simple sequence repeat analysis of a clonally propagated species: a tool for managing a grape germplasm collection. Genome **44**: 432–438

**Di Genova A, Almeida AM, Muñoz-Espinoza C, Vizoso P, Travisany D, Moraga C, Pinto M, Hinrichsen P, Orellana A, Maass A** (2014) Whole genome comparison between table and wine grapes reveals a comprehensive catalog of structural variants. BMC Plant Biol **14**: 7

**Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith P, Edwards J, This P** (2002) Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. Theor Appl Genet **105**: 780–795

**Doligez A, Adam-Blondon AF, Cipriani G, Di Gaspero G, Laucou V, Merdinoglu D, Meredith CP, Riaz S, Roux C, This P** (2006) An integrated SSR map of grapevine based on five mapping populations. Theor Appl Genet **113**: 369–382

**Doligez A, Bertrand Y, Farnos M, Grolier M, Romieu C, Esnault F, Dias S, Berger G, François P, Pons T,** (2013) New stable QTLs for berry weight do not colocalize with QTLs for seed traits in cultivated grapevine (*Vitis vinifera* L.). BMC Plant Biol **13**: 217

**FAO OIV** (2016) FAO-OIV Focus 2016 Table and Dried Grapes: Non-Alcoholic Products of the Vitivinicultural Sector Intended for Human Consumption. Food and Agriculture Organization of the United Nations, Rome

**Figueiredo DD, Köhler C** (2014) Signalling events regulating seed coat development. Biochem Soc Trans **42**: 358–363 10.1042/BST2013022124646244

**Grattapaglia D, Sederoff R** (1994) Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers. Genetics **137**: 1121–1137

**Grimplet J, Van Hemert J, Carbonell-Bejerano P, Díaz-Riquelme J, Dickerson J, Fennell A, Pezzotti M, Martínez-Zapater JM** (2012) Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. BMC Res Notes **5**: 213

**Honma T, Goto K** (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature **409**: 525–529

**Houel C, Chatbanyong R, Doligez A, Rienth M, Foria S, Luchaire N, Roux C, Adivèze A, Lopez G, Farnos M,** (2015) Identification of stable QTLs for

vegetative and reproductive traits in the microvine (*Vitis vinifera* L.) using the 18 K Infinium chip. BMC Plant Biol **15**: 205

**Ibáñez J, Vargas AM, Palancar M, Borrego J, de Andrés MT** (2009a) Genetic relationships among table-grape varieties. Am J Enol Vitic **60**: 35–42

**Ibáñez J, Vélez MD, de Andrés MT, Borrego J** (2009b) Molecular markers for establishing distinctness in vegetatively propagated crops: a case study in grapevine. Theor Appl Genet **119**: 1213–1222

**Ibáñez J, Carreño J, Yuste J, Martínez-Zapater JM** (2015) Grapevine breeding and clonal selection programmes in Spain. *In* **AG Reynolds**, ed, Grapevine Breeding Programs for the Wine Industry. Woodhead Publishing, Oxford, pp 183–209

**Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C,** (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**: 463–467

**Jansen RC, Stam P** (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics **136**: 1447–1455

**Karaagac E, Vargas AM, de Andres MT, Carreno I, Ibanez J, Carreno J, Martinez-Zapater JM, Cabezas JA** (2012) Marker assisted selection for seedlessness in table grape breeding. Tree Genet Genomes **8**: 1003–1015

**Kaufmann K, Melzer R, Theissen G** (2005) MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. Gene **347**: 183–198

**Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL** (2013) TopHat**2**: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14**: R36

**Kosambi DD** (1944) The estimation of map distances from recombination values. Ann Eugen **12**: 172–175

**Kovaleva LV, Smirnova NK, Milyaeva EL** (1997) Seedlessness: structure and metabolic activity of Vitis vinifera L female gametophyte (cv Kishmish chernyi). Russ J Plant Physiol **44**: 368–373

**Kumar S, Stecher G, Tamura K** (2016) MEGA**7**: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol **33**: 1870–1874

**Lahogue F, This P, Bouquet A** (1998) Identification of a codominant scar marker linked to the seedlessness character in grapevine. Theor Appl Genet **97**: 950–959

**Lander ES, Botstein D** (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**: 185–199

**Langmead B, Salzberg SL** (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods **9**: 357–359

**Laucou V, Lacombe T, Dechesne F, Siret R, Bruno JP, Dessup M, Dessup T, Ortigosa P, Parra P, Roux C,** (2011) High throughput analysis of grape genetic diversity as a tool for germplasm collection management. Theor Appl Genet **122**: 1233–1245

**Ledbetter CA, Ramming DW** (1989) Seedlessness in grapes. Hortic Rev **11**: 159–184

**Li H** (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **27**: 2987–2993

**Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R** (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078–2079

**Lijavetzky D, Cabezas JA, Ibáñez A, Rodríguez V, Martínez-Zapater JM** (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics **8**: 424

**Malabarba J, Buffon V, Mariath JEA, Gaeta ML, Dornelas MC, Margis-Pinheiro M, Pasquali G, Revers LF** (2017) The MADS-box gene Agamous-like 11 is essential for seed morphogenesis in grapevine. J Exp Bot **68**: 1493–1506

**Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, Conesa A, Tárraga J, Pascual-Montano A, Nogales-Cadenas R, Santoyo J,** (2010) Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Res **38**: W210–W213

**Mejía N, Gebauer M, Munoz L, Hewstone N, Munoz C, Hinrichsen P** (2007) Identification of QTLs for seedlessness, berry size, and ripening date in a seedless × seedless table grape progeny. Am J Enol Vitic **58**: 499–507

**Mejía N, Soto B, Guerrero M, Casanueva X, Houel C, Miccono MdeL, Ramos R, Le Cunff L, Boursiquot JM, Hinrichsen P,** (2011) Molecular, genetic and transcriptional evidence for a role of VvAGL11 in stenospermocarpic seedlessness in grapevine. BMC Plant Biol **11**: 57

**Mellway RD, Lund ST** (2013) Interaction analysis of grapevine MIKC(c)-type MADS transcription factors and heterologous expression of putative véraison regulators in tomato. J Plant Physiol **170**: 1424–1433

**Melzer R, Verelst W, Theissen G** (2009) The class E floral homeotic protein SEPALLATA3 is sufficient to loop DNA in 'floral quartet'-like complexes in vitro. Nucleic Acids Res **37**: 144–157

**Merdinoglu D, Butterlin G, Bevilacqua L, Chiquet V, Adam-Blondon AF, Decroocq S** (2005) Development and characterization of a large set of microsatellite markers in grapevine (*Vitis vinifera* L.) suitable for multiplex PCR. Mol Breed **15**: 349–366

**Merin U, Rosenthal I, Lavi U** (1983) A chemical method for the assessment of grapes according to their seed content. Vitis **22**: 306–310

**Mizzotti C, Mendes MA, Caporali E, Schnittger A, Kater MM, Battaglia R, Colombo L** (2012) The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. Plant J **70**: 409–420

**Mizzotti C, Ezquer I, Paolo D, Rueda-Romero P, Guerra RF, Battaglia R, Rogachev I, Aharoni A, Kater MM, Caporali E,** (2014) SEEDSTICK is a master regulator of development and metabolism in the Arabidopsis seed coat. PLoS Genet **10**: e1004856

**Nitsch JP, Pratt C, Nitsch C, Shaulis NJ** (1960) Natural growth substances in Concord and Concord Seedless grapes in relation to berry development. Am J Bot **47**: 566–576

**Ocarez N, Mejía N** (2016) Suppression of the D-class MADS-box AGL11 gene triggers seedlessness in fleshy fruits. Plant Cell Rep **35**: 239–254

**Ooi LC, Low ET, Abdullah MO, Nookiah R, Ting NC, Nagappan J, Manaf MA, Chan KL, Halim MA, Azizi N,** (2016) Non-tenera contamination and the economic impact of SHELL genetic testing in the Malaysian independent oil palm industry. Front Plant Sci **7**: 771

**Pabón-Mora N, Wong GK, Ambrose BA** (2014) Evolution of fruit development genes in flowering plants. Front Plant Sci **5**: 300

**Pearson HM** (1932) Parthenocarpy and seedlessness in Vitis vinifera. Science **76**: 594

**Pellerone FI, Edwards KJ, Thomas MR** (2001) Grapevine microsatellite repeats: isolation, characterisation and use for genotyping of grape germplasm from southern Italy. Vitis **40**: 179–186

**Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, Yanofsky MF** (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature **424**: 85–88

**Pratt C** (1971) Reproductive anatomy in cultivated grapes: a review. Am J Enol Vitic **22**: 92–109

**Ramming DW, Tarailo R, Badr SA** (1995) 'Crimson Seedless': a new late-maturing, red seedless grape. HortScience **30**: 1473–1474

**Rietz S, Stamm A, Malonek S, Wagner S, Becker D, Medina-Escobar N, Vlot AC, Feys BJ, Niefind K, Parker JE** (2011) Different roles of Enhanced Disease Susceptibility1 (EDS1) bound to and dissociated from Phytoalexin Deficient4 (PAD4) in Arabidopsis immunity. New Phytol **191**: 107–119

**Robinson MD, Oshlack A** (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol **11**: R25

**Robinson MD, McCarthy DJ, Smyth GK** (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–140

**Royo C, Carbonell-Bejerano P, Torres-Pérez R, Nebish A, Martínez Ó, Rey M, Aroutiounian R, Ibáñez J, Martínez-Zapater JM** (2016) Developmental, transcriptome, and genetic alterations associated with parthenocarpy in the grapevine seedless somatic variant Corinto bianco. J Exp Bot **67**: 259–273

**Rubio S, Whitehead L, Larson TR, Graham IA, Rodriguez PL** (2008) The coenzyme A biosynthetic enzyme phosphopantetheine adenylyltransferase plays a crucial role in plant growth, salt/osmotic stress resistance, and seed lipid storage. Plant Physiol **148**: 546–556

**Singh R, Low ET, Ooi LC, Ong-Abdullah M, Ting NC, Nagappan J, Nookiah R, Amiruddin MD, Rosli R, Manaf MA,** (2013) The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. Nature **500**: 340–344

**Singh R, Ti LLE, Li LOC, Abdullah MO, Nookiah R, Sambanthamurthi R, Ordway J, Lakey ND, Smith SW, Martienssen R** (Jan. 22, 2015) Expression of SEP-like Genes for Identifying and Controlling Palm Plant Shell Phenotypes. International Patent Application No. PCT/US2014/047468

**Stout AB** (1936) Seedlessness in Grapes. New York State Agricultural Experiment Station Technical Bulletin 238. New York State Agricultural Experiment Station, Geneva, NY

**Striem MJ, Spiegel-Roy P, Baron I, Sahar N** (1992) The degrees of development of the seed-coat and the endosperm as separate subtraits of stenospermocarpic seedlessness in grapes. Vitis **31**: 149–155

**Teh SL, Fresnedo-Ramírez J, Clark MD, Gadoury DM, Sun Q, Cadle-Davidson L, Luby JJ** (2017) Genetic dissection of powdery mildew resistance in

interspecific half-sib grapevine families using SNP-based maps. Mol Breed **37**: 1

**Tello J, Torres-Pérez R, Grimplet J, Carbonell-Bejerano P, Martínez-Zapater JM, Ibáñez J** (2015) Polymorphisms and minihaplotypes in the VvNAC26 gene associate with berry size variation in grapevine. BMC Plant Biol **15**: 253

**Tello J, Torres-Pérez R, Grimplet J, Ibáñez J** (2016) Association analysis of grapevine bunch traits using a comprehensive approach. Theor Appl Genet **129**: 227–242

**This P, Lacombe T, Thomas MR** (2006) Historical origins and genetic diversity of wine grapes. Trends Genet **22**: 511–519

**Thorvaldsdóttir H, Robinson JT, Mesirov JP** (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform **14**: 178–192

**Torregrosa L, Fernandez L, Bouquet A, Boursiquot JM, Pelsy F, Martínez-Zapater JM** (2011) Origins and consequences of somatic variation in grapevine. *In* **AF Adam-Blondon, JM Martinez-Zapater, C Kole**, eds, Genetics, Genomics, and Breeding of Grapes. CRC Press, Boca Raton, FL , pp 68–92

**van Dijk AD, Morabito G, Fiers M, van Ham RC, Angenent GC, Immink RG** (2010) Sequence motifs in MADS transcription factors responsible for specificity and diversification of protein-protein interaction. PLOS Comput Biol **6**: e1001017

**van Ooijen JW** (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet **84**: 803–811

**van Ooijen JW, Voorrips RE** (2001) JoinMap Version 3.**0**: Software for the Calculation of Genetic Linkage Maps. Plant Research International, Wageningen, The Netherlands

**van Ooijen JW, Boer MP, Jansen RC, Maliepaard C** (2002) MapQTL 4.0, Software for the Calculation of QTL Positions on Genetic Maps. Plant Research International, Wageningen, The Netherlands

**Varoquaux F, Blanvillain R, Delseny M, Gallois P** (2000) Less is better: new approaches for seedless fruit production. Trends Biotechnol **18**: 233–242

**Vitulo N, Forcato C, Carpinelli EC, Telatin A, Campagna D, D'Angelo M, Zimbello R, Corso M, Vannozzi A, Bonghi C,** (2014) 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

**Wang L, Yin X, Cheng C, Wang H, Guo R, Xu X, Zhao J, Zheng Y, Wang X** (2015) Evolutionary and expression analysis of a MADS-box gene superfamily involved in ovule development of seeded and seedless grapevines. Mol Genet Genomics **290**: 825–846

**Wang L, Hu X, Jiao C, Li Z, Fei Z, Yan X, Liu C, Wang Y, Wang X** (2016) Transcriptome analyses of seed development in grape hybrids reveals a possible mechanism influencing seed size. BMC Genomics **17**: 898

**Zhang Y, Xu S, Ding P, Wang D, Cheng YT, He J, Gao M, Xu F, Li Y, Zhu Z,** (2010) Control of salicylic acid synthesis and systemic acquired resistance by two members of a plant-specific family of transcription factors. Proc Natl Acad Sci USA **107**: 18220–18225