UNIVERSITÀ DEGLI STUDI DI TORINO

SCUOLA DI DOTTORATO IN SCIENZE DELLA NATURA E TECNOLOGIE INNOVATIVE

DOTTORATO IN SCIENZE AGRARIE, FORESTALI ED AGROALIMENTARI

CICLO: XXVII

Genetic investigation of seed development in grapevine

Chinedu Charles Nwafor

Docente guida: **Prof. Silvia Guidoni**

Coordinatore del Ciclo: **Prof. Dr. Aldo Ferrero**

ANNI

2012; 2013; 2014

Table of Contents

List of figures

List of tables

1. Summary

In a comprehensive attempt to understand the molecular and cellular processes driving seedlessness in grapevine, a seeded variety (wild-type) and its seedless somatic variant (mutant) were characterized at the morphological, genomic and transcriptomic levels in relation to berry development and seed content. The overall importance of clonal variability and the application of Next Generation Sequencing technology in highlighting the molecular events during seed formation within a developing berry have been clearly demonstrated.

In this thesis three hypothesis were formulated, tested and confirmed. First it was hypothesized that the seedless mutant has a gross morphology identical to the wild-type except for berry size. In testing this hypothesis quantitative and qualitative traits that relate to berry development and seed content were compared in the two clones. Here traits that were significantly different in the two lines are those related only to berry size and seed content. This evaluation was performed both in control conditions (self-pollination) and after anther/stigma removal which further allowed the investigation of a possible role for Parthenocarpy, Stenospermocarpy or other mechanisms in promoting the phenotype of the seedless somatic variant.

The second hypothesis states that the mutant is sterile or partly sterile hence cannot produce viable seeds. In order to verify this hypothesis pollen germination and viability assays were carried out in both clones. The tests confirmed pollen germination and vitality percentage of the mutant was significantly lower than that of the wild-type.

The third hypothesis concerned the existence of genomic/transcriptomic differences between the two lines and could be tested through the power of the Next generation Sequencing technology. In particular, we raised the following questions: are there somatic mutations that can allow the wild-type and mutant to be distinguished? What are the temporal and spatial changes that could occur in their respective transcriptomes?

Especially how does expression levels of key regulatory genes change before, during and after fertilization in the two clones? These key questions were addressed with the aid of Molecular marker analysis, Array based SNP genotyping and RNA-Seq approach. Using 58 microsatellites, the analyzed loci showed identical profile in the wild-type and the mutant. The 20K grapevine Illumina CHIP revealed 16333 identical SNP loci in the two clones, thus a further confirmation of the true identity of the seedless line. Conversely variant calling from

RNA-Seq enabled the identification of numerous somatic mutations at the whole-genome level in the two lines. A total of 71,557 SNPs and 37,121 INDELs were initially identified relative to the Pinot Noir reference sequence. Following filtering and selection based on putative functions relevant to the trait of interest, 142 candidate SNPs could be discovered out of which 120 were selected for Sanger sequencing. Thirty-one of them were true positives and mostly mapping to exonic regions, i.e. Coding SNPs.

At the same time, RNA-Seq allowed the creation of inventories of gene expression at successive stages of seed formation. i.e. stages E-L 15 (single flowers in compact groups), E-L 27 (young berries enlarging) and E-L 38 (berries harvest-ripe). Here the transcriptomes revealed by Illumina mRNA-Seq technology had approximately 98% of grapevine annotated transcripts and about 80% of them were commonly expressed in the two lines. Differential gene expression analysis revealed a total of 1075 differentially expressed genes (DE) in the pairwise comparison of developmental stages, which included DE genes specific to the wild-type background, DE genes specific to the mutant background and DE genes commonly shared in both backgrounds. The analysis of differential expression patterns and functional category enrichment of wild-type and mutant DE genes highlighted significant coordination and enrichment of pollen and ovule developmental pathways. The expression of some selected DE genes was further confirmed by real-time RT-PCR analysis.

To the best of our knowledge the work presented in this thesis represents the most comprehensive attempt to characterize the genetic bases of seed formation in grapevine. We have shown that a seeded wine grape and its seedless somatic variant are similar in several biological processes except for berry size and seed content. With a high throughput method we could identify an inventory of genes with altered expression in the mutant compared to the wild-type, which may be responsible for the seedless phenotype. The genes located within known genomic regions regulating seed content may be used for the development of molecular tools to assist table grape breeding. Therefore the data reported here have provided a rich genomic resource for practical use and functional characterization of the genes that potentially underpin seedlessness in grapevine.

2. Introduction

2.1. Seed formation during berry development in grapevine

Seed formation in grapevine represents a sexual reproduction process of which the first step is inflorescence initiation and its emergence, followed by flower and finally berry development. In brief, the establishment of sexual organs of grapevine, growing in temperate region, occurs over two successive growing seasons. In the first season uncommitted primordia, or Anlagen are initiated during the latent bud formation in late summer. Overwintered buds burst following the perception of spring in the preceding season and the formation of flowers takes place later during this period. Bud burst is preceded by the activation of all the structures in the latent bud (Figure 1A, Boss and Thomas 2002), especially the differentiation of inflorescences forming an organ primordium from which flower primordia may develop; these events mark the early stage of floral organ development. The successive development of floral organs is simultaneous in each flower of the inflorescence in the same primordium and follows an order of organ appearance that is similar to all angiosperms. A detailed review on floral organ development in grapevine has been reported by (Coombe 1973, Coombe 1976, Srinivasan and Mullins 1981, Mullins et al 1992, and Dokoozlian, 2000).

Most cultivated *Vitis vinifera* varieties have hermaphroditic (perfect) flowers, consisting of male (five stamens which are pollen-bearing organs of the flower) and female (pistil houses the stigma, style, and ovaries) organs. Each stamen is tipped with a pollen-producing anther and a filament or stalk. The stigma serves as a pollen receiver while the style is a short, slender column of tissue arising from the ovary to the stigma. The ovary contains four ovules with each ovule consisting of an embryo sac that houses a single egg. Located at the base of the flower are five odor glands. (Figure 1B and 1C, Coombe, 1992, Dokoozlian, 2000).

2.2. Fertilization

At anthesis, the calyptra (cap) is dislodged due to the growth of the stamen. Briefly, the calyptra is a cap shaped green structure produced by the fusion of petals during floral organ development, it encloses the reproductive organs and other tissues within the flower (Dokoozlian 2000). Following the detachment of the cap, the stamens and the pistil become exposed. At full bloom the anther splits open releasing their pollen grains. Multiple pollen grains may adhere to substances secreted on the stigma at the tip of the pistil which consists primarily of sugars, proteins, and mineral nutrients required for pollen tube development.

Grape varieties with hermaphroditic flowers are considered self-pollinating. With favourable environmental conditions the pollen grains germinate and form pollen tubes. The pollen tube grows down the pistil to the ovary and penetrate an ovule, where a male gamete (sperm) fuse with an egg to form the embryo. This association is termed fertilization and under normal field conditions, it typically occurs two to three days after pollination (Dokoozlian 2000, Williams 2000). Following fertilization is a period (fruit set) when the fertilized flower starts to develop a seed and grape berry which protects the seed (Figure 1D).

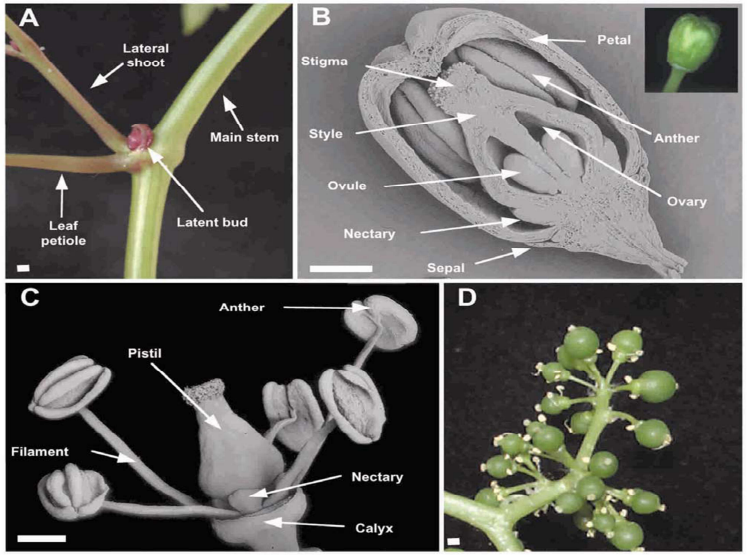


Figure 1. Grapevine sexual reproductive organs. **(A)** Bud burst **(B)** Cross section of a developing inflorescence with arrows indicating the male and female gametophytes. **(C)** A hermaphroditic (perfect) grape flower. **(D)** Fruit set.

2.3. Seed: embryo and endosperm

There is a paucity of information on the regulatory regime of pre- and post-embryonic development in grapevine. However earlier work on the reproductive anatomy of grapevine classified embryo formation as the Geum variation of the Asterad (Figure 2A and B). Similarly the order and pattern of cell division in the embryo was reported to be similar to other angiosperms (Pratt 1971, Mullins et al 1992). Following successful fertilization the zygote goes through a resting period of several weeks before cellular division. The pattern of zygotic

cellular division in grapevine has been reviewed by (Pratt 1971). The author reported endosperm cellular division to have occurred before cytokinesis during zygote development. Also the author classified the pattern of endosperm formation in the grapevine as helobial (free nuclear division) with the first cellular division of the primary endosperm nucleus, producing a transverse wall across the embryo sac, which in turn forms a small chalazal cell and a large micropylar cell (Figure 2B). The micropylar cell contains the endosperm nucleus, which further divides severely, without developing cell wall. Three to six free-nuclear divisions occur in the micropylar chamber before any wall formation occurs, in contrast to the chalazal cells where every division is accompanied by cell wall formation. In the mature seed the colour of the embryo ranges from toques orange to yellowish brown and varies in size and length. While the endosperm is whitish and irregular in shape, its cells contains nutrients which are absorbed as the embryo develops (Figure 2C).

Following fertilization intensive meristematic growth takes place in the inner and outer integument (Figure 2B). The rate of mitosis in the outer integument was reported to be maximum at 25 days after bloom and ceases by 45 days after anthesis. The outer integument thickens and elongates to form the beak. The middle layers of the outer integument in the basal half of the seed forms two projections on either side of the raphe which push the inner integument and nucellus inward. These projections are called seed folds or fossettes. The cells of the inner integument divide anticlinally to keep pace with the growth of the outer integument. The nucellus grows with the integuments by cell enlargement and division. For details on the physiological and anatomical features of endosperm and integument formation as well as development see (Pratt 1971, Mullins et al 1992).

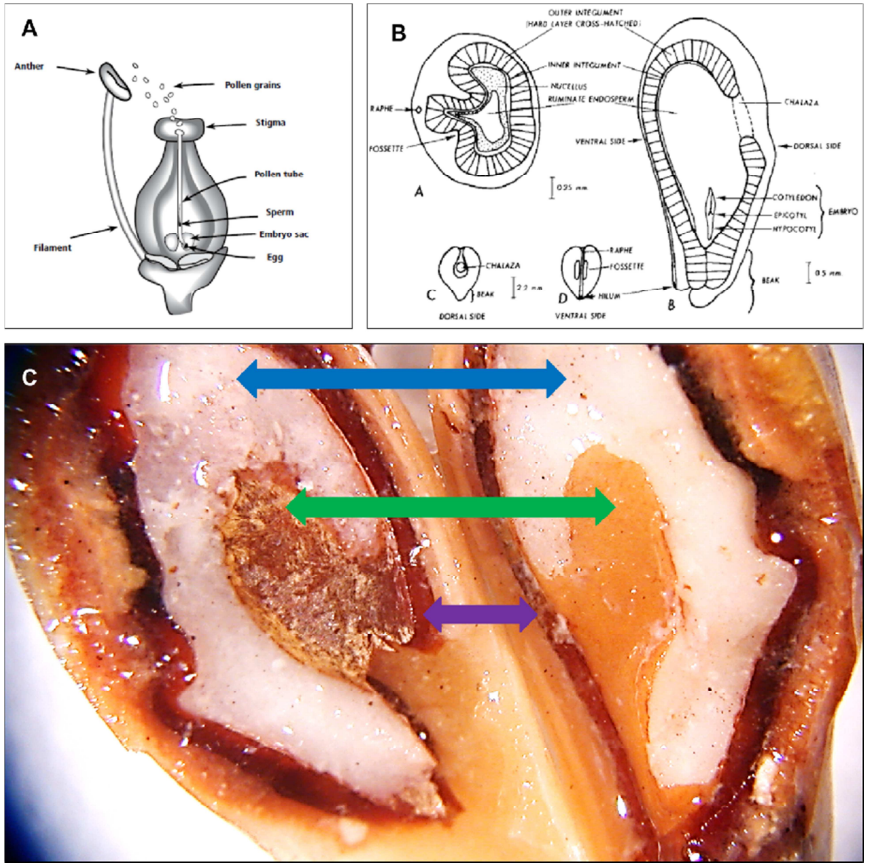


Figure 2. Seed development. (**A)** Animation of fertilization process. (**B)** Illustration of different stages of embryo development after pollination. (**C)** Dissection of a matured seed, 60 days post anthesis. Violet arrow indicates the seed coat, green arrow indicates embryo embedded in the endosperm and blue arrow indicates the endosperm. (**A)** and (**B)** were adapted from (Dokoozlian 2000 and Pratt 1971) respectively.

2.4. Seedlessness in grapevine

The origin of ancient seedless cultivars is unclear, however they probably arose due to single mutation that took place in one shoot of an otherwise normal vine. Seedless varieties of *Vitis vinifera* have been cultivated and prized for many years, mainly because they are preferred

for fresh and dry fruit consumption. Currently two main mechanisms, namely parthenocarpy and stenospermocarpy, have been described to be responsible for most seedless forms of grape (Ledbetter and Ramming 1989). Parthenocarpic conditions are usually referred to as fruit developmental process where true seedlessness occurs, i.e. fruit develops from the ovary in the absence of fertilization, yielding small berries that completely lack seeds. Examples are Corinto and its related cultivars that are used mostly for making raisins (Ledbetter and Ramming 1989, Cabezas et al 2006). Whereas in stenospermocarpy, pollination and fertilization take place normally, but seed development aborts at an early stage (2-4 weeks) after fertilization (Ledbetter and Ramming 1989, Mejia et al 2007). In this more prevalent mechanism, the pericarp (berry flesh) keeps growing but the embryo and/or endosperm arrests its development, resulting in the presence of seed traces and a reduced berry size at harvest (Doligez et al. 2002, Fanizza et al. 2005, Mejia et al. 2007).

Previous studies on seedless grapes have focused on the anatomical and morphological difference between seeded and seedless cultivars, showing that the gross morphology of the vines is mostly similar except for seed formation and berry size (Pearson 1933, Olmo 1934, Olmo 1937, Barritt 1970). For instance, in Black Corinth cultivars known to be parthenocarpic, all embryo sacs observed at anthesis were at various stages of degeneration (Olmo 1937). In many cases the entire egg apparatus was missing or appeared abnormal. In addition the ovules were very small, with only one layer of sclerenchymatous cells in the outer integument. For stenospermocarpic cultivars, double fertilization was shown to trigger fruit development (Ledbetter and Ramming 1989). Ovule development was reported to be abnormal and normal in some cases (occasionally berries within clusters contain lignified seeds). Also endosperm development was observed to precede embryo development prior to its degeneration. The endosperm degenerates from 20 to 25 days post anthesis depending on cultivar although embryos may remain viable as they are usually arrested after endosperm degeneration. Aborted embryos appear as small whitish seeds or seed traces in the ripe berry. Integument development in stenospermocarpic cultivars was also reported to be abnormal and independent of embryo sac development (Ledbetter and Ramming 1989). In most abnormal ovules the inner integument protrudes beyond the outer integument and probably towards the chalazal outside the ovule showing little or no sclerenchyma cells. Comparison of three stenospermocarpic cultivars from pre-bloom through to 25 days after anthesis showed that at ripening the size of seed traces is relative to the time of endosperm/embryo degeneration

(Barritt 1970, Ledbetter and Ramming 1989).

Numerous degree of seed development with continuous variation have been observed in grapevine, a summary of the various types of seeds are shown in **Table 1** as described by (Striem et al 1992).

Table 1. Seed types

Type	Description
Normal seed	Complete lignified seeds
Empty/soft seed	Well-developed seed but devoid of embryo, endosperm or nucleus
Large seed trace	Berries with late endosperm/embryo degeneration. Traces are visible and measure from 4 to 5.5 mm
Medium seed trace	Visible trace but smaller than large seed trace
Small seed trace	Very small trace measuring below 2 mm

2.5. Genetic basis of seedlessness in grapevine

Several models were earlier proposed for controlling inheritance of seedlessness, however the widely accepted model suggests that genetic inheritance of seedlessness is governed by the expression of three independent recessive genes under the control of a dominant regulator gene named *Seed Development Inhibitor (SDI)* (Bouquet et al 1996, Lahogue et al 1998, Adam-Blondon et al 2001, Doligez et al 2002). These studies were based on the analysis of a segregating population, i.e a progeny segregating for seedlessness obtained by crossing two partially seedless genotypes. To date several other studies adopting a similar approach have reported the existence of a number of QTLs controlling seedlessness in grapevine such as the QTL intervals located on linkage groups (LGs) 1, 2, 4, 5, 12, 14 15 and 18 (Doligez et al 2002, Fanizza et al 2005, Cabezas et al 2006, Mejía et al 2007, Costantini et al 2008, Doligez et al 2013). Costantini et al (2007) and Mejía et al (2011) proposed a MADS-box ovule identity gene *(VvAGL11)* to be the most probable candidate gene for the major QTL on LG18 controlling both berry weight and seed traits. It was also successfully tested for

usefulness in marker-assisted selection (Bergamini et al 2013). Recently Doligez et al (2013), reported the most stable QTL intervals for berry weight and seed traits containing many genes whose functions are possible related to seedlessness.

While the identification of candidate gene through co-localization with QTL has been useful in shedding light on the positional genes with functions relevant to seedlessness, it is probably not a comprehensive approach towards characterizing seedlessness in grapevine. Indeed, identifying the genomic regions that regulate seed content variation within a segregating population is still quite far away from understanding the underlying biological processes. Additionally, all the QTL studies performed till now focused on a single type of seedlessness and genetic background, as they used Sultanina (Also known as Thomson seedless) or its derived varieties as parents.

2.6. A comprehensive approach for understanding the molecular mechanism underlying seedlessness in grapevine

A logical approach to better understand the processes driving seedlessness would be to highlight the molecular events during seed formation within a developing berry in a seeded cultivar and its seedless mutant. Various studies in grapevine have reported the existence of somatic variations affecting several traits (Torregrosa et al 2011). therefore somatic variants for seed content may be a valuable material.

Useful tools to this purpose are those providing a holistic view of the genomic or more importantly the transcriptional landscape during seed development in the two lines as well as allowing their direct comparison. Recently, novel approaches enabled by Next Generation Sequencing technologies (NGS) are proving invaluable towards archiving this feat.

2.7. Next Generation Sequencing technologies (NGS)

NGS technologies have wide range of applications, nowadays more are being developed at a fast rate compared to five years ago when they were initially introduced. NGS techniques allow the sequencing of thousands of genomes from humans to plants through to microbes. This has opened entirely new areas of biological inquiry resulting in the ability of researchers to investigate biological questions that were not previously possible such as ancient genomes, human disease, ecological diversity. Detailed review of current and emerging NGS technologies can be found in (Wold et al 2008, Wang et al 2009, Ponting et al 2009, Caniato 2011).

2.7.1. Current applications of NGS include

I. Full genome (re-) sequencing or variant discovery by re-sequencing of targeted regions of interest among individuals (mapping of structural rearrangements, which may include copy number variation, deletions, insertions and chromosomal inversions
II. Transcriptome analysis (RNA-Seq, gene prediction and annotation, alternative splicing discovery).
III. Epigenetic (large scale analysis of DNA methylation).

3. Aims and objectives

Seedless grapes are of interest for both fresh fruit consumption and raisin production. This thesis is an attempt aimed at unravelling the genetic regulation of seed development in grapevine through an integrative approach. A seeded variety (wild-type) and a seedless somatic variant (mutant) are compared at the morphological, genomic and transcriptomic level at different developmental stages in order to understand the biological processes underlying the two distinct phenotypes. Once identified, the allelic differences determining phenotypic differences might be integrated in marker-assisted breeding programs with the potential to produce a new generation of seedless grapevine .

4. Phenotypic characterization of wild-type and mutant

4.1. The importance of cultivar identification in viticulture and enology

Grapevine (*Vitis vinifera* L) is one of the most important fruit crops in the world. In Italy, it is presently cultivated over an area of about 700,000 ha with a productivity of approximately six million Tonnes (FAO stat 2012). These grapes are mainly used for wine purposes. Viticulturists and winemakers are increasingly interested in exploring genetic diversity among grapevine varieties, mainly to rationalise, preserve and exploit genetic resources. Special attention is given to local and old materials, which are mostly forgotten (relic). This interest resulted in a strong emphasis on proper identification of grapevine cultivars (ensuring trueness to type) and of their genetic relationships sometimes providing evidence of intra-varietal variation (berry colour variants being the most frequent). Indeed the identification of grape varieties including their synonyms (different names for the same cultivar) and homonyms (same name for different cultivars), and how they relate with other grape cultivars is not only crucial for conservation and genetic improvement. It is equally important for wine production and marketing in several regions of the world. For example, proper identification and verification of synonyms has a practical significance with respect to countries where wine regulation is enforced by legislation e.g. Italy, a member state within the European Union, where the use of wine grapes for cultivation is strictly regulated: only registered and specifically authorized cultivars can be grown. In addition, the rules for the wine geographic appellations establish the grapes to be used (Schneider et al 2001).

4.2. PCR based methods used for accurate cultivar identification

PCR based molecular markers, i.e. Microsatellites also known as SSR (Simple Sequence Repeats), RAPDs (Random Amplified Polymorphic DNAs), AFLPs (Amplified Fragment Length Polymorphism) and SNPs (Single Nucleotide Polymorphism) present an alternative and objective means for grapevine cultivar identification independent of the phenotypic characteristics used in ampelography (Pellerone et al 2001). In particular, SSRs are widely adopted for genetic assessment of grapevine cultivars mainly because they are highly polymorphic and co- dominantly inherited (Maul et al 2012). Also SSR and more recently SNP data are easy to interpret and both assay are amenable to automation allowing the genotyping of several hundred cultivars in a single run thereby saving cost and time

(Emanuelli et al 2008).

4.3. Identification of a seedless somatic variant of Sangiovese in the Calabria grapevine germplasm

Calabria is one of the regions in the south of Italy with ancient viticultural practices and wine tradition. In this area several vineyards are small, holding ancient, local as well as widespread varieties, and represent a potential source of grapevine varietal diversity. In 2009, more than 250 grapevine accessions held in the private collection of the Librandi winery (Cirò Marina, KR, Italy) were characterized using SSR and ampelographic descriptors with the aim of evaluating the genetic diversity. The study identified several synonyms for the major cultivars and the homonyms were distinctly defined (Schneider et al 2009). Among the synonyms was an accession named Corinto Nero which had the same profile at 10 SSR loci as Sangiovese, a widespread wine cultivar in Italy. Surprisingly, gross morphology of the Corinto Nero vines was the same as that of Sangiovese, except for the reduced berry size and seedless berries (berries with rudimental seeds).

4.4. Is the Corinto Nero grown in Calabria the true Corinto Nero?

In today's world the grape cultivar Corinto Nero as it is known in Italy is alleged to have come from Korinthos (Corinth) in the North-east of Peloponnese Greece where it was first called Korintiaki and known as Corinthian in Greek (Robinson et al 2012). Korintiaki has many synonyms which differed from country to country and region to region within a particular country. For example in some parts of Australia and United States of America (USA) it is called Currant Grape while in other parts it is called Zante Currant, in fact, in California it is called Black Corinth. In France it is called Corinthe Noir as well as Raisin de Corinthe. In Italy it has three official synonyms namely Corinto Nero, Passerilla and Passula di Corinto.

In the DOC (Denominazione di origine controllata) wine Malvasia delle Lipari produced in the Eolie islands, off the northern coast of Sicily (not far from Calabria), 5% of Corinto Nero grapes are allowed. However the results of Schneider et al (2009) cast doubt on the trueness to type and origin of the so called "Corinto Nero" grown in the Eolie islands and Calabria region, if it was the true Korintiaki or perhaps a seedless form of Sangiovese mistaken for Korintiaki (the true Corinto).

Although the true origin of Korintiaki may not be in Greece as genetic and morphological studies have distinguished it from other Mediterranean wine grape varieties (Robinson et al 2012). Also Vargas et al (2007) reported that neither Corinthe Blanc from Greece nor Corinto

Noir from Italy were mutants of Korintiaki. Taken together, these information suggest that the Corinto Nero grown in the Eolie islands and Calabria may not be the true Korintiaki.

This chapter is aimed at confirming the trueness to type of Corinto Nero (hereafter mutant) identified in Calabria region by performing a comprehensive genotypic and phenotypic characterization of the mutant in comparison with a reference true to type Sangiovese cultivar (hereafter wild-type).

4.5. Main objectives are

I. To analyse the DNA profile of mutant and compare it to the profile a reference wild-type using fifty eight microsatellite loci.
II. To perform quantitative and qualitative characterization of the phenotypic differences observed between mutant and wild-type if any.
III. To investigate the physiological process possibly responsible for seedlessness phenotype of the mutant (parthenocarpy or stenospermocarpy).
IV. To determine whether the seedless phenotype of the mutant is heritable.
V. To test the viability of the mutant pollen.

4.6. Methods

4.6.1. Sample collection

The grapevine germplasm collection of Grinzane Cavour maintained by CNR-Istituto di Virologia Vegetale di Grugliasco (Torino, Italy), holds the same mutant accession identified in Calabria since it was vegetatively propagated, as well as the true reference Sangiovese.

For molecular marker analysis, young leaves were collected from wild-type and mutant.

For pollen germination and viability tests, wild-type and mutant pollens were obtained from inflorescence harvested on May 29th, 2014 when the plants were at flowering stage. Samples were kept in cooler bags with silica gel. Each genotype had two replicates, i.e. WT I, WT II, MT I, and MT II consisting of 1-2 opened flower clusters with a few flowers still closed.

4.6.2. Genomic DNA extraction and SSR genotyping of the wild-type and the mutant

Total genomic DNA was extracted from young immature leaves as described by Emanuelli et al (2013). Fifty eight SSR markers, spread across the nineteen chromosomes of grapevine genome, were used to genotype the wild-type and the mutant (Appendix 1). Of this set, twenty SSR markers were previously described by (Emanuelli et al 2013), thirty-two SSR

markers used by (Costantini et al 2008) and six SSR markers developed by (Mejía et al 2011).

PCR amplifications for multiplex panels were carried out in a final volume of 12.5 µl containing 10 ng of genomic DNA, 0.25 mM of each dNTP, 2 mM MgCl2, 1.5 U Taq DNA Polymerase (AmpliTaq Gold™, Applied Biosystems, Foster City, CA). The amplification protocol was as follows: 7 min at 95 ℃; 30 cycles of 45 sec at 95 ℃, 1 min at 54 ℃, 30 sec at 72 ℃; and 1 hour at 72 ℃. Primers failing to amplify at 54 ℃ were further tested in single panel at different annealing temperatures. PCR products (0.5 µl) were mixed with 9.3 µl of formamide and 0.2 µl of the GeneScan™ 500 ROX® Size Standard (Applied Biosystems) and 0.5 µl of this mix was subjected to capillary electrophoresis on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems) to separate DNA fragments. GeneMapper v3.5 (Applied Biosystems) was employed for the allele size estimation.

4.6.3. Evaluation of the phenotypic differences between wild-type and mutant: berry development and seed content

Wild-type and mutant phenotypes was reported for the first time by Schneider et al (2009). In the present study, quantitative and qualitative evaluation of berry development and seed content for wild-type and mutant was carried out in three successive growing seasons: 2012, 2013 and 2014. In each plant, if available, 3-5 representative clusters or bunches were randomly selected and measured for bunch length (BHL) and weight (BHW) in order to compute mean values. Next 25 berries were randomly taken from a mixture of 4-5 representative clusters, weighted (berry weight, BW) from which mean berry weight (MBW) was calculated. Subsequently 10 berry diameter readings were randomly taken and averaged for each cluster (MBD) .

Seeds and seed traces extraction was performed on 25 berries randomly sampled from a cluster mix. Counts were taken for the number of berries that contain seeds and seed traces in both clones, in order to compute total seed number (SN) and mean seed number per berry (MSN). Total seed fresh weight (TSFW) was measured and seed number was used to compute mean seed fresh weight (MSFW = TSFW/SN).

Qualitative assessment was performed based on classification of seed content, using a method previously described by Bergamini et al (2013). In brief, all analyzed clusters were divided in four classes, namely C1 for aborted and not evaluable seeds, C2 for aborted and rudimentary seeds, C3 for complete not lignified seeds, and C4 for lignified seeds.

The normality of each trait distribution was evaluated by the Kolmogorov-Smirnov test. Genotype effect was tested with analysis of variance and Kruskal-Wallis test ($p < 0.05$). Statistical analyses were performed with R packages.

4.6.4. Investigating the physiological process responsible for the seedlessness phenotype

In order to determine whether the mutant has a parthenocarpic or stenospermocarpic phenotype, emasculation and covering experiments were performed before anthesis in the two clones (Figure 3). When available 12 pre-capfall (fused petals) inflorescence were randomly selected, 8 inflorescences were manually decapped and emasculated. The emasculated inflorescences consisted of two groups: Group one (Emasculation plus stigma, EMS+ST) where only anthers were removed; group two (Emasculation minus stigma, EMS-ST) both anthers and stigma were carefully removed. Next the remaining 4 pre-capfall inflorescences were left un-emasculated (self-pollinated, SP). Finally both emasculated and un-emasculated groups were tagged for easy identification and covered with paper bags. The self-pollinated inflorescences were used as a control making three treatments with four replicates per clone. At harvest, quantitative and qualitative evaluation of berry and seed traits were performed as follows.

- **A.** Comparing the emasculated groups of mutant and wild-type.
- **B.** Comparing emasculated and un-emasculated groups of wild-type.
- **C.** Comparing emasculated and un-emasculated groups of the mutant.

Figure 3. Picture showing field experiment **(A)** Grapevine inflorescence before emasculation. **(B)** Emasculation process. **(C)** Tagging and covering of emasculated inflorescence. **(D)** Complete emasculation treatments imposed on experimental plants in the vineyard.

4.6.5. Heritability of the seedless phenotype

To test the heritability of the seedless trait of the mutant, control crossing experiments (Figure 4) were carried out between the mutant and two cultivars, Nebbiolo and Trebbiano toscano respectively. Briefly Nebbiolo is an early flowering cultivar while Trebbiano toscano is late flowering. Both varieties are highly productive. Prior to their respective anthesis (Mutant, Nebbiolo and Trebbiano toscano), if available at least 5 pre-capfall inflorescences were randomly selected from 4 plants of each cultivar. The inflorescence were manually decapped and emasculated. The emasculated inflorescences of mutant were manually pollinated with pollens obtained from Nebbiolo and vice versa. The same was done for Trebbiano toscano. The total number of crosses carried out are detailed in Table 2.

Table 2. Cross pollination

Cross pollination	Number of plants emasculated	Number of clusters emasculated
Mutant x Nebbiolo	4	4
Nebbiolo x Mutant	4	5
Mutant x Trebbiano toscano	4	4
Trebbiano toscano x Mutant	4	10

All manually pollinated inflorescences were tagged and covered with paper bags. At harvest quantitative and qualitative evaluation were performed for all crosses that set fruit, as described in **Section 4.6.3** above. All seeds obtained from the respective crosses were washed, disinfected and stored at 4^{o} C for three months prior to seed germination trials.

Figure 4. Schematic representation of cross pollination experiment. **(A)** Grapevine inflorescence before emasculation. **(B)** Emasculation process. **(C)** Emasculated inflorescence. **(D)** Nylon pollination brush. **(E)** Manual pollination. **(F)** Covering and tagging of pollinated inflorescence.

4.6.6. Extraction of pollen

Anthers from freshly opened flowers were gently removed by separating them from the petal and sepals. Next these anthers were then placed in Petri dishes in the desiccator (4 °C) in a refrigerator for three days before pollen collection.

For closed flowers the inflorescences were placed in water for three days to allow for

maturation. Unfortunately, the flowers did not open to release pollens as expected. As a result all individual flowers from this cohort were removed and kept them under controlled conditions for 20 h, exposed to a temperature of 29 ^{o} C, at 1 m from a 125 W lamp. Pollen grains were gathered from both opened and closed flowers separately, through a sieve. Finally all collected pollens were stored in a desiccator at (4 °C) in a refrigerator.

4.6.7. Germination test

Germination tests were performed independently for wild-type and mutant, as previously described by Carreno et al (2006). The solution used for pollen germination assay was composed of the following:

20% sucrose

100 mg / L boric acid

300 mg / L calcium nitrate.

Prior to germination test assay, the pollen grains were set to rehydrate (equilibration of pollen in humid air, Relative Humidity = 100%) at room temperature for at least an hour. Germination assay was carried out in a mass culture medium contained in 5-cm wide plastic Petri dishes at the recommended temperature of 25 ° C (approximately) for 24 hours. The quantity of pollen was proportional to the volume of the medium in each Petri dish (5 mg in 5 ml of germination medium). Finally slides were prepared and samples were analysed in replicates. Sprouted grains were counted in random fields by photomicrographs, those that are considered germinated are only granules where the length of the pollen tube appeared double compared to the granule.

4.6.8. Vitality test

Pollen viability test was performed individually for wild-type and mutant. A solution of 2,3,5-*triphenyl tetrazolium chloride* (TTC) was used for staining (few drops of 1% TTC (0.2 g. TTC and 12 g. sucrose) were dissolved in 20 ml distilled water). Replicates for each sample are prepared in the absence of light, TTC solution was dropped by Pasteur pipettes on microscope slides and pollen were re-shaken with a slim brush (one brush per plant type) covered with a coverslip. Next the microscope slides were placed in an incubator for one hour at temperature of 37 ℃. The staining for both clon es were analysed under the microscope. Out of a population identified by random fields, approximately 300 grains per slide were counted separately; granules that are viable appeared red and non-viable as (yellowish to

colorless). Also quantitative analysis of pollen morphology was performed for both clones by measuring the polar and equatorial diameter of 50 randomly selected pollen grains.

4.7. Results and discussion

4.7.1. Genotypic profile of wild-type and mutant based on molecular markers SSR

Using a total of 58 SSR molecular markers to genotyping the wild-type and mutant, means that previous genotyping assay performed on the two clones (Schneider et al 2009) has been extended by at least 4 folds. The markers were selected such that they are spread across the grapevine genome. Analysis of microsatellite results showed the two clones to have identical allele sizes at all the fifty-eight analyzed loci. See (Appendix 1).

4.7.2. Phenotypic characterization of wild type and mutant

The results of the first ampelographic characterization are described in (Schneider et al 2009). The authors reported that the two varieties shared all phenotypic characters except for the traits related to berry and seed size. Indeed the field observation carried out in this work for three successive years 2012, 2013 and 2014 showed the mutant had a gross morphology consistent with the wild-type except for the traits related to berry development and seed content, thereby confirming the earlier reports of Schneider and co-workers (Figures 5 and 6). Analysis of quantitative and qualitative data correlated to berry development and seed content for the growing seasons mentioned above produced very similar results therefore only data for one year (2014) are shown. Normality test indicates a departure from normal distribution of most traits (BHL, BHW, MBW, MBD, SN, TSFW, MSN and MSFW) under study, even after data sets were log-transformed; instead of a normal distribution most of the traits show a bimodal distribution, While only one trait (BHW) exhibited a normal distribution. Therefore both parametric and nonparametric test were employed to highlight significant difference between wild-type and mutant with respect to traits correlated to berry development and seed content, (Appendix 2).

4.7.2.1. Comparison between wild-type and mutant clusters

The results of comparison between wild-type and mutant self-pollinated experimental groups showed the mutant varied significantly ($p < 0.05$) from the wild-type in most of the traits under study (Appendix 2). For example the variation seen in bunch weight, mean berry weight, mean berry diameter, mean seed number, number of berries with and without seed (Figure

7). However there was no significant difference ($p > 0.05$) between the two clones in bunch length and mean seed fresh weight (Figure 7 and Appendix 2). With regards to the non-significant differences observed in the two clones, i.e. mean seed fresh weight is largely a less controlled parameter due to the fact that seeds may not be completely free of pulp after extraction from berry, which could lead to weight artefacts, perhaps a better augmentation to this parameter would be seed dry weight (SDM). In terms of bunch length, the non-significant difference observed further highlights the co-linearity of the two clones during inflorescence development.

Qualitative assessment of seed content revealed that the wild-type had two distinct berry sizes (Figure 5 and 6), all berries contained seeds and the two seed sizes can be visualized in (Figure 9). All wild-type seeds belonged to class C4 (Figure 6 and 9). Each berry contained a minimum of two seeds and a maximum of five seeds. Very few berries from mutant clusters had comparable size to those of the wild-type and contained seeds (at most two) belonging to class C4 (Figure 7 and 8). These berries were mostly located on the upper part of the cluster (Figure 5). Majority of the mutant belonged to classes C2 and C3 having small berry diameter ranging from (0.2-0.6cm), occasionally containing small seeds that are either greenish or whitish in colour (Figure 8).

Figure 5. Clusters from the two clones at harvest. **(A)** Wild-type cluster. **(B)** Mutant cluster.

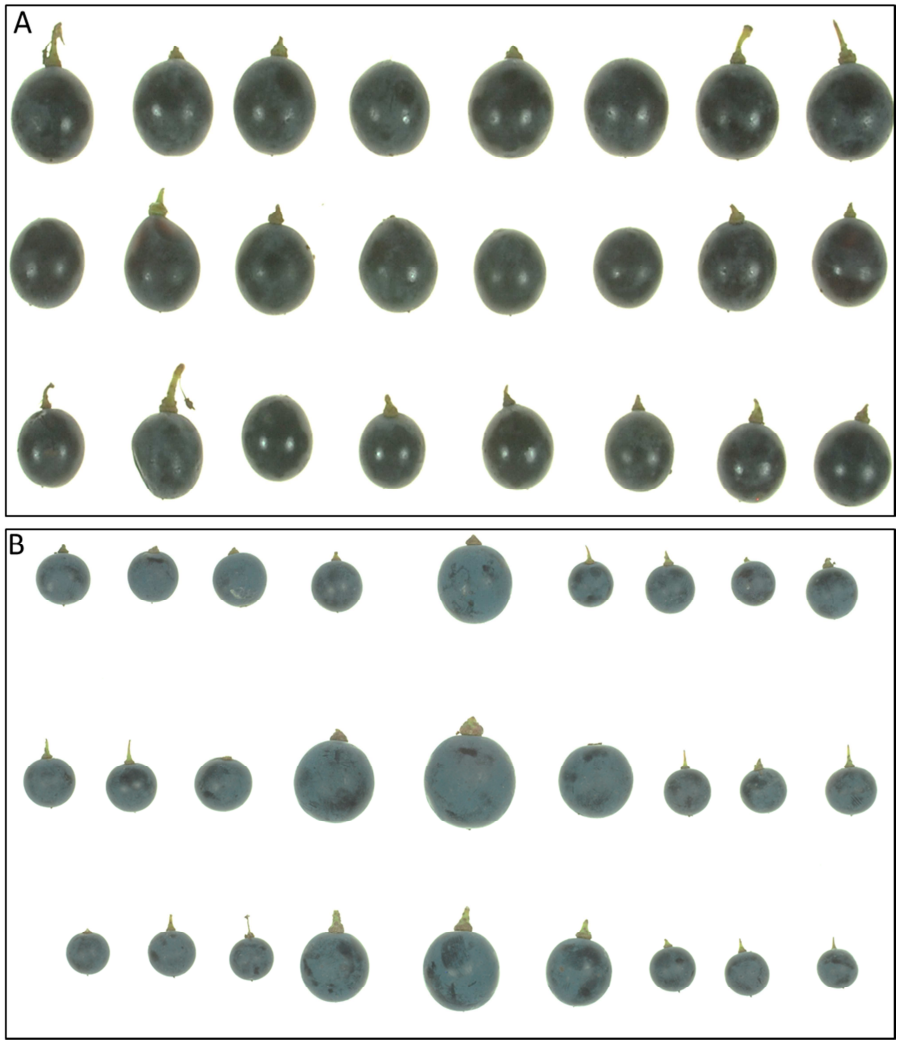


Figure 6. Berries from the two clones. **(A)** Wild-type berries. **(B)** Mutant berries.

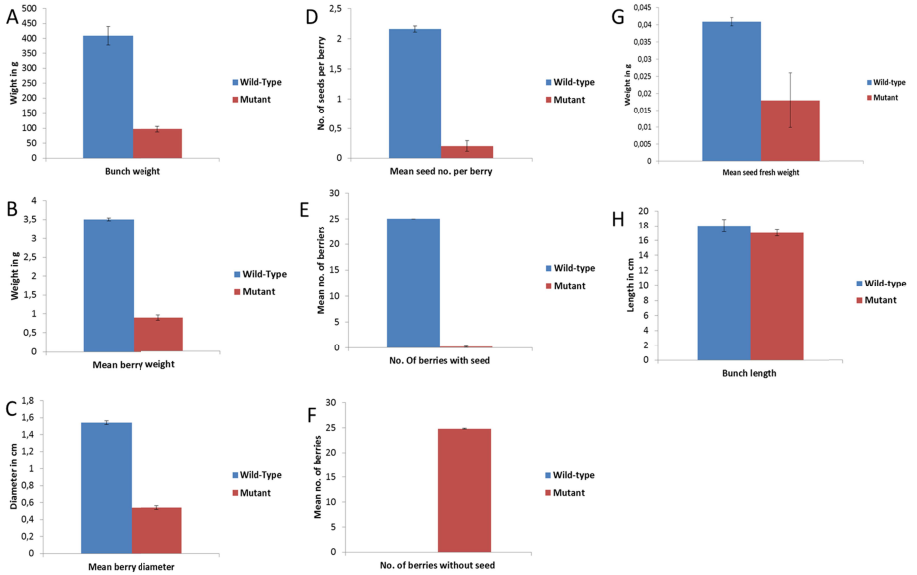


Figure 7. Comparison of berry- and seed- related traits in wild-type and mutant. **(A)** Mean bunch weight (BHW). **(B)** Mean berry weight (MBW). **(C)** Mean berry diameter (MBD). **(D)** Mean seed number per berry (MSN). **(E)** Mean number of berries with seed. **(F)** Mean number of berries without seed. **(G)** Mean seed fresh weight (MSFW) and **(H)** Bunch length (BHL).

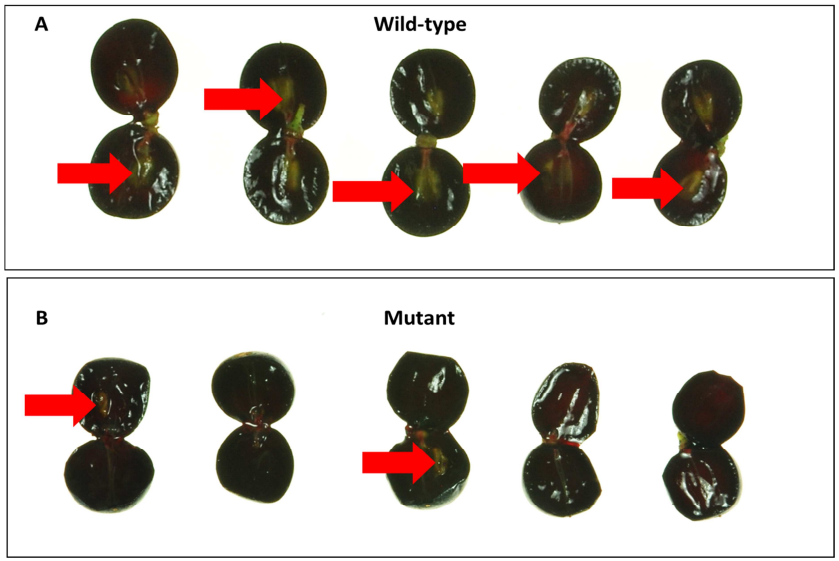


Figure 8. Dissected berries. **(A)** Wild-type berries **(B)** Mutant berries. Red arrows indicate the presence of seeds.

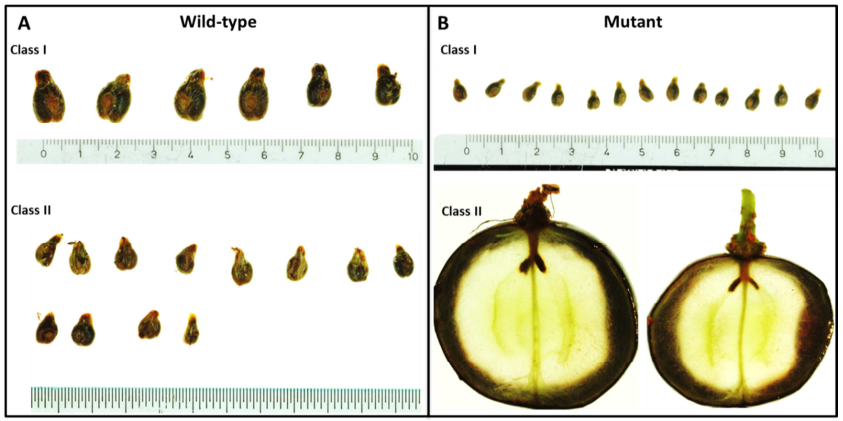


Figure 9. Qualitative analysis of seed. **(A)** Wild-type seeds from berries in **classes I** and **II**. **(B)** Mutant seeds from all the berries in **class I** and seed traces from randomly selected berries in **class II**.

4.7.3. Physiological process possibly responsible for the seedlessness phenotype

To investigating whether parthenocarpy or stenospermocarpy is responsible for the seedless phenotype of the mutant, three treatments were imposed on wild-type and mutant inflorescence respectively before anthesis. The experimental procedure included three groups (EMS+ST, EMS-ST and SP, see Figure 3 and Methods for details). Physical observation of all experimental groups 20 days post anthesis revealed that wild-type and mutant plants belonging to group EMS+ST were able to set fruits comparable to the SP groups after emasculation (Figure 10). Equally a few of the EMS-ST group had fruit set while majority of the inflorescences appeared dead.

Following these intriguing observation, all clusters within the experimental groups were monitored at intervals (every two weeks) from 20 days post anthesis through to harvest. During this monitoring period, it was observed, that as berry development progressed individual berries as well as clusters of SP group grew bigger in size compared to the EMS+ST and few surviving EMS-ST in both clones. At harvest clusters of EMS+ST group in both wild-type and mutant had reduced berry size, with the mutant EMS+ST groups having much smaller size than the wild-type EMS+ST group (Figure 11 and 12). Quantitative traits related to bunch and berry weight as well as seed content were analysed as described in the method. However the EMS-ST group was excluded from further analysis as very few samples were available.

4.7.3.1. Comparing the emasculated groups of the mutant and wild-type

The results shown in Figure 13 suggests that bunch weight and mean berry weight varied significantly between the experimental groups ($p < 0.05$, Appendix 3), while bunch length and mean berry diameter were not significantly different ($p > 0.05$), in addition to mean seed number, mean seed fresh weight, number of berries with and without seeds (Appendix 3).

4.7.3.2. Comparing emasculated and un-emasculated groups of wild-type

As shown in Figure 14, all the traits under study showed significant difference ($p < 0.05$, Appendix 4) between the experimental groups except MSFW. Here the observed significant difference was not surprising, because it was evident during field observation, where the emasculated groups exhibited reduced berry size compared to the self-pollinated group.

4.7.3.3. Comparing emasculated and un-emasculated groups of mutant

The results reported in Figure 15 showed that the mutant emasculated groups varied

significantly ($p < 0.05$, Appendix 5) from those of the self-pollinated group in bunch length, bunch and mean berry weight. While average berry diameter, mean seed number, mean seed fresh weight, number of berries with and without seeds were not significantly different ($p > 0.05$, Appendix 5).

4.7.4. Qualitative characterization of seed content

Figure 16 shows different types of seedlessness observed in mutant (self-pollinated and emasculated) and wild-type emasculated groups. Similarly characterization of qualitative seedless level gave support to the type of seedlessness observed here (See Appendix 6, Figure 1). For example the evaluation of most berries from the wild-type and mutant EMS+ST group, revealed they belonged to class C1. Although some berries from the wild-type EMS+ST group contained seeds that appeared to be lignified, when these seeds were weighed they had extremely low weight (0.4g). The mutant SP group belonged to classes C2 and C3, as most of the berries had noticeable seed traces and occasionally lignified seeds. While wild-type SP belonged to class C4 since nearly all the berries contained well lignified seeds (Figure 12).

Taken together these observations could suggest that stenospermocarpy may be responsible for the seedless phenotype observed in the mutant (SP groups) since pollination and fertilization had occurred. Furthermore the non-detectable seed trace (class C1) observed in wild-type and mutant EMS+ST groups could be attributed to parthenocarpy. However what is puzzling is how some of the berries could contain seed, if emasculation was done without errors or no form of pollination had occurred. Then it is worth investigating the source of the seed. There are reports in the literature about occurrence of cleistogamy and apomixis in grapevine hence it is worth carrying out further studies on these seeds at the genetic level n order to understand the origin of the embryo's.

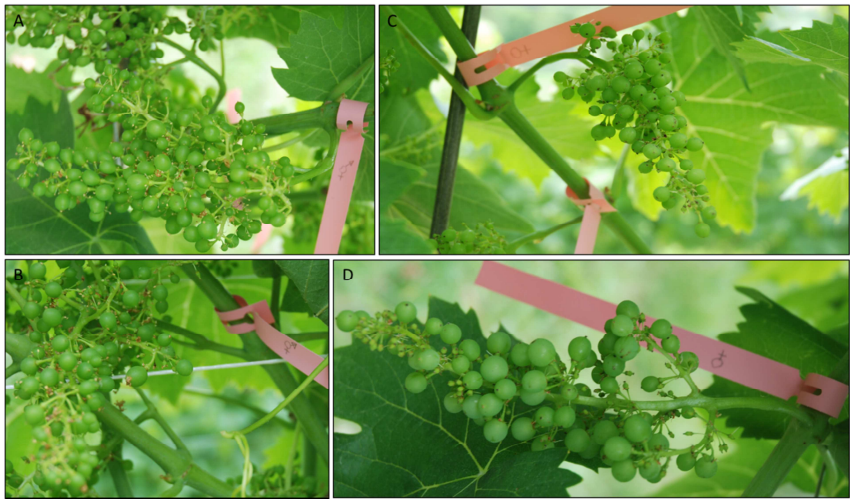


Figure 10. Physical observation of experimental groups 20 days post anthesis. **(A)** Wild-type self-pollination. **(B)** Mutant self-pollination. **(C)** Wild-type emasculated plus stigma. **(D)** Mutant emasculated plus stigma.

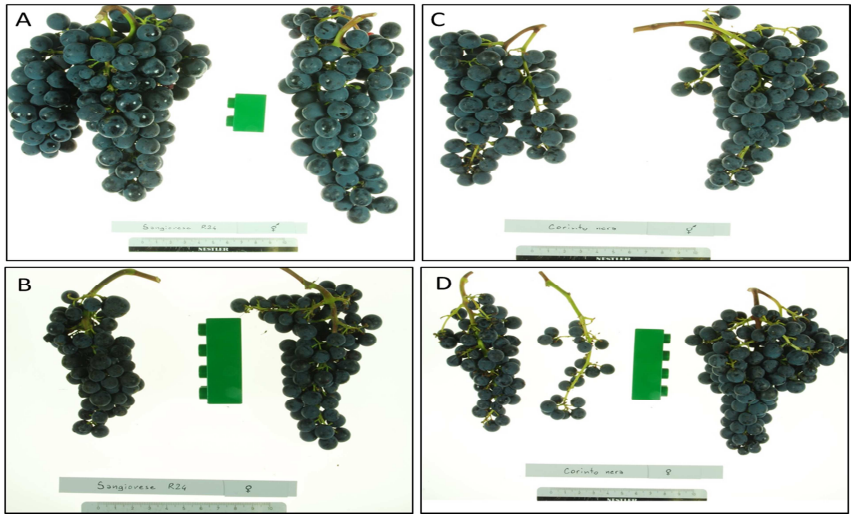


Figure 11. Clusters from two experimental groups. **(A)** Wild-type self-pollination. **(B)** Wild-type emasculated plus stigma. **(C)** Mutant self-pollination. **(D)** Mutant emasculated plus stigma.

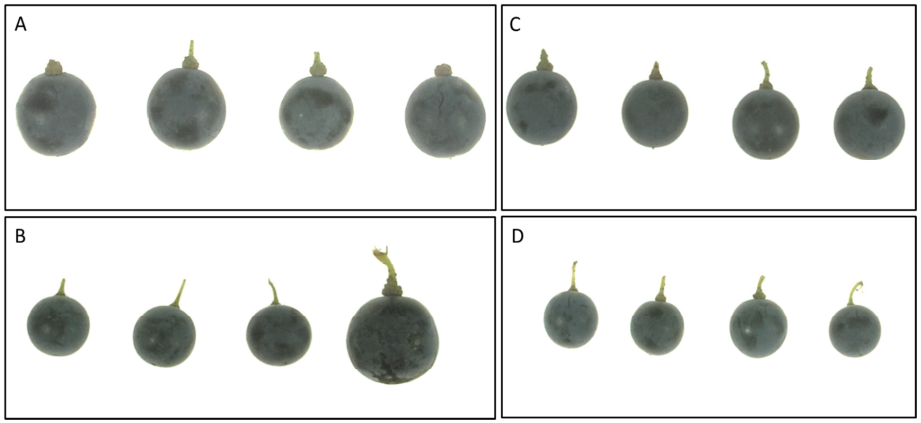


Figure 12. Berries from two experimental groups separated in two classes based on size. **(A)** Wild-type self-pollination. **(B)** Wild-type emasculated plus stigma. **(C)** Mutant self-pollination. **(D)** Mutant emasculated plus stigma.

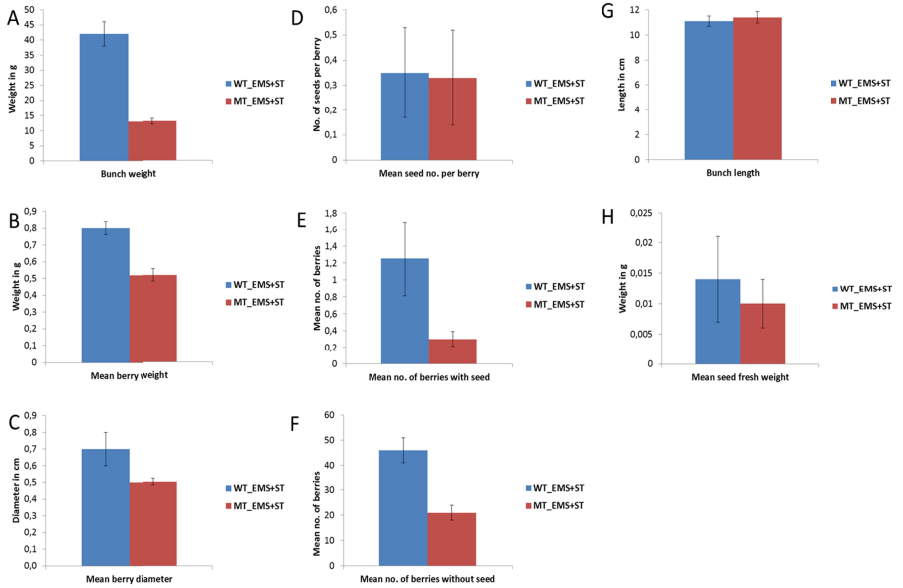


Figure 13. (A) Bunch weight (BHW). **(B)** Mean berry weight (MBW). **(C)** Mean berry diameter (MBD). **(D)** Mean seed number per berry (MSN). **(E)** Mean Number of berries with seed. **(F)** Mean Number of berries without seed. **(G)** Bunch length (BHL) and **(H)** Mean seed fresh weight (MSFW).

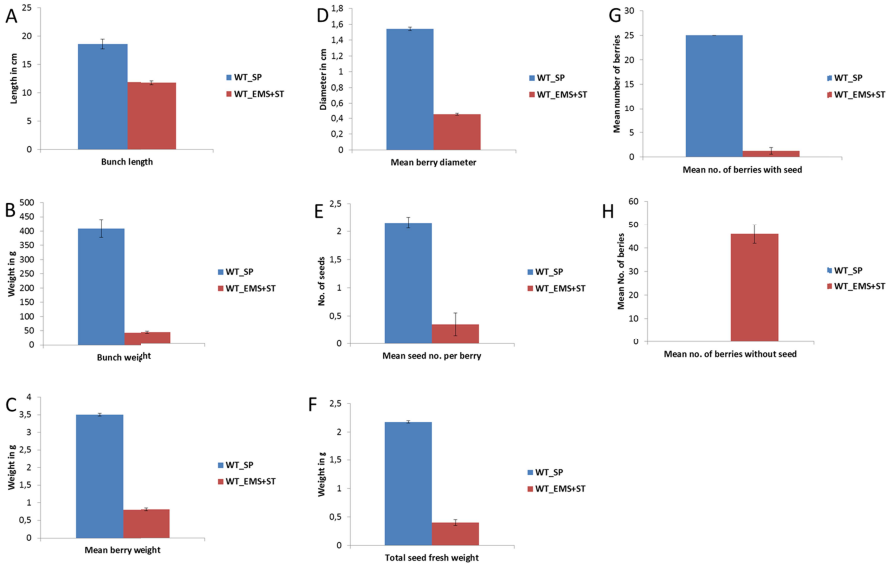


Figure 14. (A) Bunch length (BHL). **(B)** Bunch weight (BHW). **(C)** Mean berry weight (MBW). **(D)** Mean berry diameter (MBD). **(E)** Mean seed number per berry (MSN). **(F)** Total seed fresh weight (TSFW). **(G)** Mean number of berries with seed and **(H)** Mean number of berries without seed.

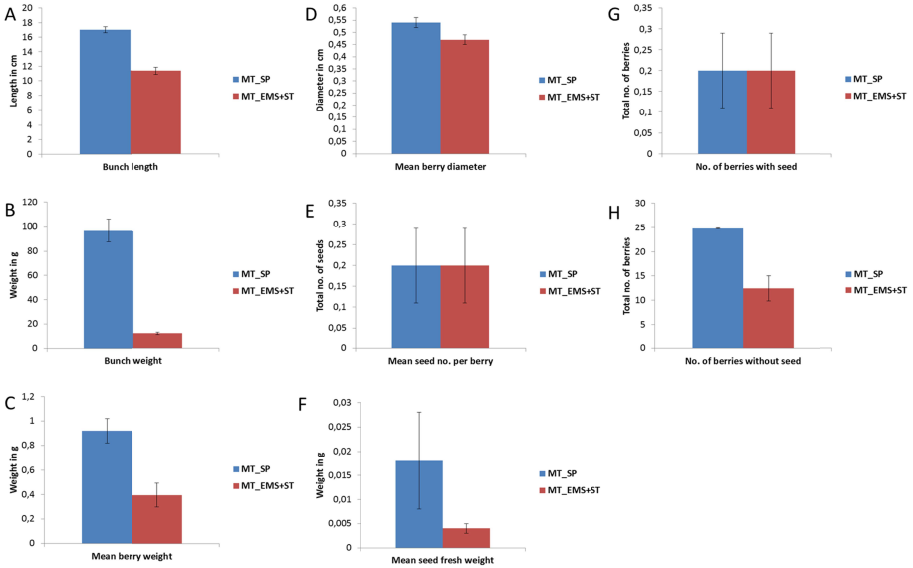


Figure 15. (A) Bunch length (BHL). **(B)** Bunch weight (BHW). **(C)** Mean berry weight (MBW). **(D)** Mean berry diameter (MBD). **(E)** Mean seed number per berry (MSN). **(F)** Mean seed fresh weight (MSFW), **(G)** Mean number of berries with seed and **(H)** Mean number of berries without seed.

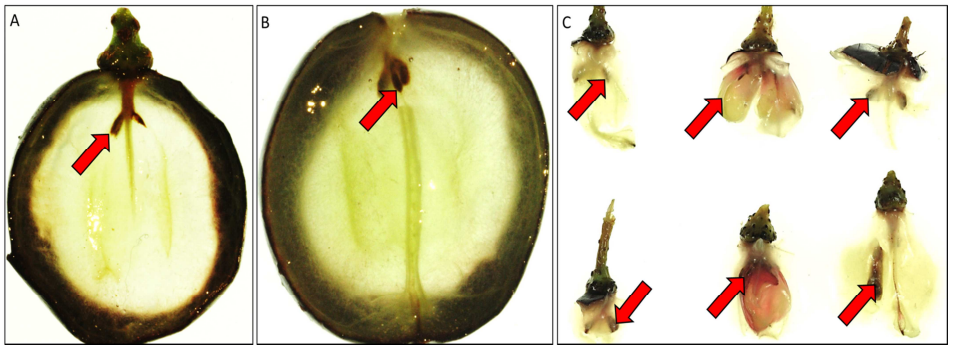


Figure 16. Various degree of seedlessness observed in the experimental groups**.** Red arrows indicate aborted or undeveloped seeds. **(A)** Wild-type emasculated group. **(B)** Mutant self-pollinated group. **(C)** Mutant emasculated group.

4.7.5. Phenotypic evaluation aimed at investigating the inheritance of the seedless phenotype

From 3 to 5 clusters each were observed in the crosses between Mutant x Nebbiolo and Trebbiano toscano x Mutant at harvest, indicating that the method described in Figure 3 was successful for pollination. For the Nebbiolo x Mutant cross, very few inflorescences were able to set fruit (two clusters and one later died due to infection). Therefore the cross between Mutant x Trebbiano toscano was excluded from the study due to lack of statistical power. Comparison of quantitative data for traits correlated to seedlessness among the crosses carried out was not possible due to the fact that only one cluster was observed for all the Nebbiolo x Mutant crosses. However seed count data showed that nearly all berries from the Nebbiolo x Mutant contained at most two seeds while majority of berries from cross Mutant x Nebbiolo had no seed. Furthermore qualitative data analysis showed that few berries from Mutant x Nebbiolo and Trebbiano toscano x Mutant contained lignified seeds and most of the berries without seed belonged to class C2. Finally all seeds obtained from the crosses failed to germinate after several germination trials. Taken together these results suggest that perhaps both male and female gametophyte of the mutant work in concert to promote seedlessness, how this is achieved in the mutant is still unknown, however a stated earlier our data suggests the mutant is a stenospermocarpic cultivar.

4.7.6. Pollen germination and viability test

To test the hypothesis of a non- functional pollen in the mutant, pollen germination and viability tests were performed, using pollen samples from both the wild-type and mutant plants. The experimental procedure consists of pollen extracted from both clones at two different phases. See methods for details of experimental design.

The result shown in Table 3 and Figure 17 A, suggest that pollens extracted from the opened flowers had a good germination capacity with respect to the wild-type (albeit with some differences between the replicates WTI and WTII). The pollen extracted subsequently showed very low (less than 4%) germination percentages. Furthermore the results indicated that none of the mutant pollen samples germinated, regardless of the replicate (MT I, MT II) and extraction phase (Table 3 and Figure 17 B).

Pollen viability test results are given in Table 4 and show the wild-type viability rate to be significantly higher than the mutant which had no reaction with the TTC in all stages and replicates under study (Figure 18 A and B).

Following these results, morphometric analysis of the pollen grains were carried out, comparing the pollen samples of the mutant to the wild-type, as well as data from literature (Table 5). When considering the range of morphometric measures reported in the literature (Bucher et al, 2004 for instance), *Vitis vinifera* pollen polar diameter average is about 22.8 µm and range from 22-25µm; while equatorial diameter mean is 23.7µm, ranging from 23-27µm.

Here about half of the mutant pollen grains measured (27 out of 50 for equatorial diameter; 26 out of 50 for polar diameter) exhibited values lower than 21µm for both equatorial and polar diameter. In contrast the measurements performed on the wild-type pollen samples were homogeneous and showed very low variability although slightly lower than the measurements given in the bibliography (Table 5).

Finally the morphometric result suggests mutant produced deformed pollen grains (Figure 19). This is a new finding, in addition to the other morphological differences found between the two clones as it is the first time morphometric data are studied in the two clones.

Table 3. Pollen germination test

Cultivar	Replicate	Pollen extraction	Total No. pollen observed	Total No germinated pollen	Average percentage of germinated pollen
Wild-type	I	Phase 1	1220	567	46,47
Wild-type	II	Phase 1	989	206	20,79
Wild-type	I	Phase 2	1220	0	0
Wild-type	II	Phase 2	989	0	0
Mutant	I	Phase 1	287	0	0
Mutant	II	Phase 1	415	0	0
Mutant	I	Phase 2	0	0	0
Mutant	II	Phase 2	0	0	0

Table 4. Pollen viability test

Cultivar	Replicate	Pollen extraction	Total No. pollen observed	TTC positive + Medium	Average percentage pollen positive TTC
Wild-type	I	Phase 1	387	146	37,7
Wild-type	II	Phase 1	401	195	48,6
Wild-type	I	Phase 2	392	39	9,9
Wild-type	II	Phase 2	403	46	11,4
Mutant	I	Phase 1	402	0	0
Mutant	II	Phase 1	406	0	0
Mutant	I	Phase 2	427	0	0
Mutant	II	Phase 2	412	0	0

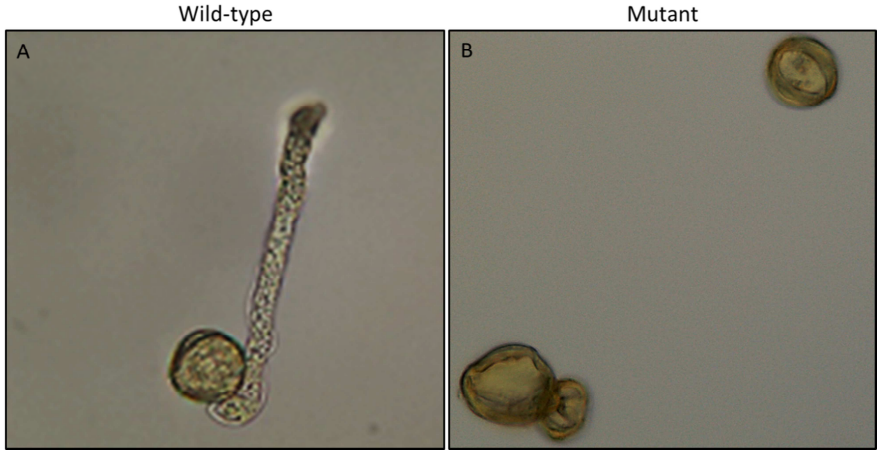


Figure 17. Pollen germination test. **(A)** Wild-type pollen tube growth. **(B)** Mutant pollen showing no germination.

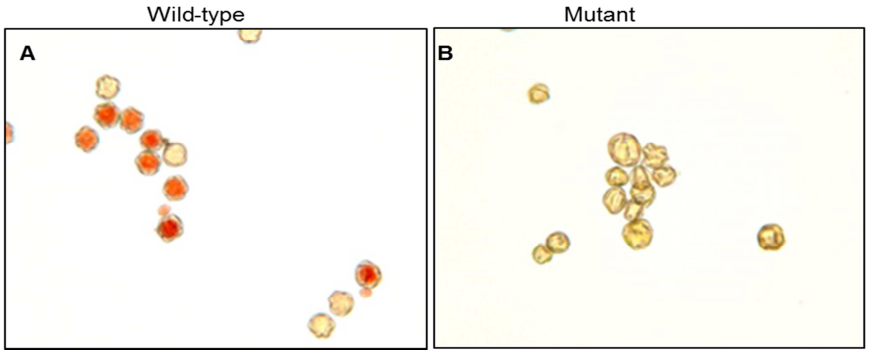


Figure 18. Pollen viability test. **(A)** Wild-type pollens showing positive staining with TTC (reddish purple color). **(B)** Mutant pollens showing no reaction with TTC (yellowish colour). Stainings for both clones were analyzed under the microscope.

Table 5. Morphometric measurement of mutant and wild-type pollens

	Equatorial diameter Mean-dev.st	Polar diameter mean-dev.st	Equatorial diameter min -max	Polar diameter min-max
CN I	22,96± 4,42	23,17±4,00	16,5-31,5	16,5-30,0
SG I	23,21±0,84	21,82±0,98	21,0-24,0	19,5-24,0

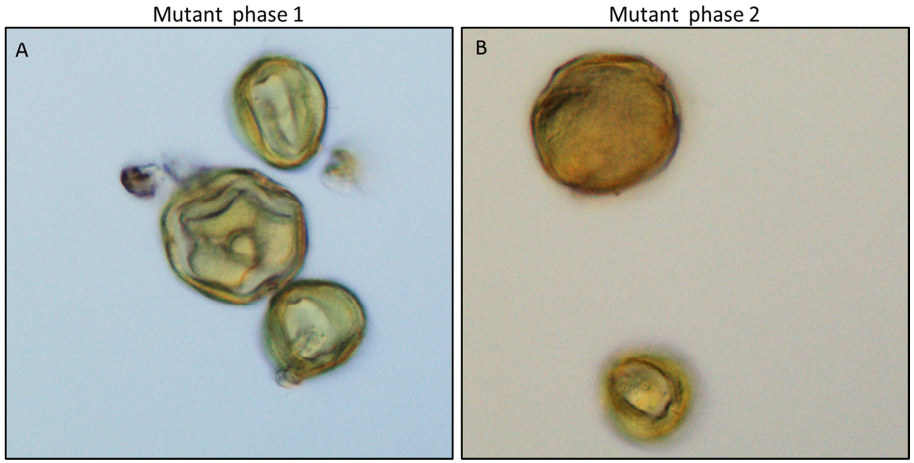


Figure 19. Morphometric measurement of mutant pollens revealed abnormal shape and size. **(A)** Pollen extracted in phase 1. **(B)** Pollen extracted in phase 2.

5. Transcriptomic and genomic variations between wild-type and mutant

The major events that take place in grapevine normal seed development, parthenocarpy and stenospermocarpy are shown schematically in (Appendix 6 Figure 2) and are described in detail by (Striem 1992, Varoquaux et al 2000).

In Arabidopsis, genetic studies have revealed several genes that participate in seed development like *SHOOT MERISTEMLESS (STM), CUP-SHAPED COTYLEDON (CUC1* and *CUC2), AINTEGUMENTA (ANT), SPATULA (SPT), AGAMOUS (AG) MADS box genes AG-SHATTERPROOF (SHP1* and *SHP2), SEEDSTICK (STK,* also known as *AGL11), NOZZLE/SPOROCYTELESS (NZZ/SPL), EMBRYO DEFECTIVE (EMB)* and *INO* (Skinner et al 2004, Jenik et al 2007, Devic 2008), including those that regulate endosperm formation such as *CRINKLY4* and *BET1* (Berger 1999, Huh et al 2008), embryo differentiation such as *EMBRYO-DEFECTIVE (EMB) and LEAFY COTYLEDON (LEC)* (Breuninger et al 2008, Braybrook and Harada 2008, Yin et al 2012), and seed coat development such as *APETALA 2 (AP2)* and *TRANSPARENT TESTA 16 (TT16)* (Dean et al 2011). Also, molecular studies with Arabidopsis, tomatoes, and other plants have revealed cis-regulatory elements of several genes active during seed development, mostly the transcription factors(TFs) that play a role in their regulation, i.e. *LEAFY COTYLEDON (LEC)* genes and *AGAMOUS like 15 (AGL15)* (Riechmann and Meyerowitz 1998, Le et al 2010, Ruan et al 2012). Nevertheless, in grapevine the identities of most regulators of seed development and their direct targets are largely unknown.

To date, very few studies have looked for genes possibly responsible for seedlessness by comparison of gene expression profiles in seeded and seedless grapes. For instance, differential expression analysis in seeded and seedless clones of cv Sultanina by (Hanania et al 2007, Hanania et al 200) allowed the identification of a chloroplast chaperonin (ch-Cpn21) resulting in seed abortion when silenced in tobacco and tomato, and of a ubiquitin extension protein (S27a) having a probable general role in the control of organ development in grapevine. Recently, differential expression analysis during ovule development in seeded and seedless cultivars identified grape metacaspase genes, consistent with a role of programmed cell death in stenospermocarpy (Zhang et al 2013).

To identify regulators and processes required for seed development that may be altered in the seedless phenotype, somatic variants are vital resources. At the same time an experimental

procedure that gives a broad view of the genomic and transcriptional landscape of both phenotypes in time and space is as important. In grapevine, somatic variation arises from mutation or epimutation events that first occur in a single cell belonging to a specific cell layer. Once at least one shoot apical meristem is colonized by the mutated cell in one or both cell layers, the mutation can be transmitted by bud propagation or eventually sexual reproduction (Torregrosa et al 2011). However, identification of somatic variants in grapevine is a time and labor intensive task, which requires genetic and phenotypic characterization of large germplasm collections (Schneider et al 2009). At the same time, the application of deep sequencing techniques to survey the total population of RNA within a tissue has made RNA-Seq a popular and comprehensive approach to deduce and quantify the transcriptome (Wang et al 2009). Its potential has been demonstrated in the de novo transcriptome characterization of *Vitis vinifera* cultivars (Zenoni et al 2012, Venturini et al 2013) and gene expression profile of grape berry during key developmental stages (Fasoli et al 2012, Sweetman et al 2012).
This chapter exploits the availability of a seedless somatic variant (also known as mutant, MT) described in Chapter two. This mutant is derived from Sangiovese (also known as wild-type, WT), a widespread seeded wine cultivar in Italy, see (Schneider et al 2009) for more detail. Here the overall aim is to highlight DNA sequence variation and transcriptional regulatory processes that may be altered in the mutant, bearing in mind that this mutant has a gross morphology of vines identical to the wild-type except for absence of seeds, reduced berry and bunch size at harvest. Therefore to understand the molecular mechanisms driving the seedless phenotype, Illumina mRNA-Seq technology was used to analyze the allelic variations as well as the transcriptional responses possibly related to seed development in the wild-type and the mutant.

5.1. Methods

5.1.1. Sample collection

Samples were collected from wild-type and mutant plants in the ger
mplasm collection of Grinzane Cavour maintained by CNR-Istituto di Virologia Vegetale di Grugliasco (Torino, Italy).
For array-based SNP genotyping and Sanger sequencing assay young leaves were gathered.
To create inventories of gene expression at successive stages of seed formation, three key time points along grape berry development were selected corresponding to stages E-L 15

(single flowers in compact groups), E-L 27 (young berries enlarging) and E-L 38 (berries harvest-ripe) of the modified E-L system described by (Coom.be 1995). Samples were collected for both clones in the following dates: 12th May, 10th June and 16th September 2010. When matched to the number of days from bloom (DFB) shown in (Appendix 6-Figures 2 and 3), these time points could be assigned to two main categories: “before” (E-L 15) and “after” (E-L 27 and 38) fertilization. A detailed description of how sampling dates were matched to DFB is reported in (Appendix 6). For each developmental stage two independent samples (biological replicates) were collected. A biological replicate was composed of the whole inflorescence for stage E-L 15 and of the whole bunch for stages E-L 27 and 38.

5.1.2. Array-based SNP genotyping: 20K grapevine Illumina CHIP

Genomic DNA was extracted with DNeasy Plant Mini Kits (Qiagen) from young leaves. DNA samples were quantified using Nanodrop 8000 (Thermo Scientific, Wilmington, DE) and quality was confirmed with gel electrophoresis. Next, 10ul of each DNA sample with a concentration of 100 ng/ul were sent to Illumina. For a detailed description of the experimental design of the 20k grapevine Illumina SNP CHIP see (link). Also detailed description of Infinium array chemistry can be found (here).
SNP calling was carried out with Illumina Genomestudio software. Briefly, genotypes are called for each sample by their signal intensity and allele frequency relative to canonical cluster positions for a given SNP marker, see link for further details.

5.1.3. Genotype filtering and polymorphism detection

A set of filtering criteria (quality thresholds) previously described by (Myles 2010) with slight modification were used to filter out inconsistent and bad quality genotypes. In short, SNPs with a call frequency of 0 were filtered out, we further required a minimum GenTrain score >0.6 and cluster separation >0.4. Finally an in house Perl script was used to carry out pairwise comparison of wild-type and mutant filtered genotype positions for polymorphism detection.

5.1.4. RNA extraction

For each sample total RNA extraction was performed from a lot of flowers/berries in triplicate (technical replicates), using the Spectrum™ Plant Total RNA kit (Sigma-Aldrich, St. Louis, MO) following the manufacturer’s protocol. RNA quality and quantity were determined using a Nanodrop 8000 (Thermo Scientific, Wilmington, DE) and a Bioanalyzer 2100 (Agilent, Santa

Clara, CA).

5.1.5. Library preparation and sequencing

For transcriptomic analysis a single biological replicate was used due to economic constraints. Total RNA from the three technical replicates of each sample were pooled for a total six pools representing each developmental stage for the two genotypes.

Libraries were prepared using the TruSeq SBS v5 protocol (Illumina, San Diego, CA). In particular, 10 µg of total RNA were used to isolate poly(A) mRNA after double purification of transcripts using poly(T) oligos attached with magnetic beads. Subsequent mRNA quality control was carried out on a Bioanalyzer 2100 (Agilent). Purified mRNA was fragmented using Zn-catalyzed hydrolysis and converted into double-stranded cDNA by random priming. Following end repair, single "A" base addition to 3'-end, indexed adapters were ligated and cDNA fragments of 200 ± 25 bp were purified. Purified cDNA was amplified by PCR and quality control was done by TOPO cloning and capillary sequencing. The cDNA libraries were quantified and diluted to 10 nM, after which they were multiplexed and sequenced with an Illumina HiSeq 2000 sequencer at Fasteris (Fasteris SA, Switzerland). A hundred-bp paired-end sequences were generated.

Image analysis, error estimation and base calling were carried out using Illumina Pipeline (version 1.4.5) to generate the sequence data. Indexed primers were used to identify the different reads from different samples in the sequence data. Some low-quality reads were removed using a custom algorithm. Illumina TruSeq adapter sequences were clipped and the remaining reads were considered suitable for further analysis after passing quality control at Fasteris.

5.1.6. cDNA sequence alignment and mapping to the reference genome

Short-read alignment and mapping of all the reads were carried on the 12x PN40024 genome assembly as well as 12x v1 transcript annotation (Vitulo et al 2014) using BWA (Burrows Wheeler Aligner) software (Li and Durbin 2010) with a maximum set of 2 mismatches in the first 32 bp sequences and a maximum of “n” mismatches in total (n from 2 to 9 depending on read length).

For polymorphism detection variants were called from reads mapped to the genome sequence using SAMtools pileup with default parameters (link). Since putative SNPs were called one library at a time, it was reasoned that the pileup file will contain every position in

the alignment where at least one base is a mismatch to the reference sequence, hence many will be false positives. Therefore filters were applied to remove those by following the set of criteria listed below,

(i) Minimum p-value for strand bias of 0.0001
(ii) Min p-value for end distance bias of 0.0001
(iii) Maximum read depth of 10000000
(iv) Minimum p-value for base quality bias of 1e-100
(v) Minimum RMS mapping quality for SNPs of 10
(vi) Minimum value for quality of 10
(vii) Minimum number of alternate bases of 2
(viii) Window size for filtering adjacent gaps of 10
(ix) Minimum p-value for map quality bias of 0
(x) SNP within Intron bp around a gap of 10 to be filtered
(xi) Minimum read depth of 5
(xii) Less than or equal to 0 for samples having a genotype mismatching
(xiii) The alternative base is observed in less than 2 reads in one of the directions
(xiv) The mutation affects a coding sequence but all the alternatives imply a amino-acid sequence identical to the reference amino-acid sequence. The output was reported in a VCF file format.

5.1.7. Variant call data analysis

5.1.7.1. Selection of putative SNPs related to the trait of interest

For a position to be considered a putative SNP or INDEL for the trait of interest in each library, the following approach was adopted.

A. It was required that the alternate base was supported by at least 3 reads and the frequency of the alternative alleles was ≥ 0.75 (since majority of the reads mapped to a single location) calculated on the total number of read pairs aligned on the region.
B. An ad hoc Perl script was written to take consensus positions that pass the initial filtering criteria in at least two libraries of wild-type and mutant respectively. From here INDELs were removed from further analysis.
C. Putative mutations from B above were annotated using the Variant Effect Predictor SNPeff program (Cingolani et al 2012).
D. An ad hoc Perl script was used to carry out a pairwise comparison between the wild-type

and mutant of all putative SNPs annotated as non-synonymous.

E. Putative SNP positions that are different in the two clones from D above were further selected based on function of the gene that harbour them and finally validated by Sanger sequencing.

5.1.7.2. Sanger sequencing

To validate putative SNPs found with the above method, polymerase chain reaction (PCR) amplification and Sanger sequencing were performed in the same panel of two clones used for RNASeq, for hundred and twenty gene fragments. DNA was extracted with DNeasy Plant Mini Kits (Qiagen), from young leaves. PCR and sequencing primers were designed based on the 12x PN40024 grapevine reference sequence using Primer 3 (Rozen and Skaletsky 2000, primers are available upon request). Chromatograms were trimmed, aligned and edited with MEGA software (Kumar et al 2004). Putative SNP loci and the genotypes of each individual identified by RNASeq-SNP calling were compared to the Sanger sequencing.

5.1.8. RNA-Seq raw read data

For transcriptomic analysis, the mapping results were processed with SAMtools, to extract for each transcript the number of mapped reads and determine, whether their mapping position is unique. Reads mapping to several positions on the reference sequence with the same “mapping quality” (i.e. number of mismatches and quality of the bases generating the mismatches) were attributed at random to one of them with a “0” mapping quality.

A Python script was developed to determine the distribution of mapped reads among genomic features for the wild-type and the mutant.

5.1.8.1. Gene expression analysis

Reads mapped to multiple locations and unmapped reads were excluded from gene expression analysis. Unique reads mapping to v1_mRNA annotated transcripts were summed for each gene model and normalized by million reads (RPM) because of read coverage bias towards 3' end of transcripts. A lower limit of detection for expression estimate was designated to be an RPM of 0.5 or, if the RPM value was less than 0.5, at least five uniquely mapped reads with identity > 98% over 100 bp, as previously described by (Sweetman et al 2012). The full raw expression dataset have been submitted to GEO under the accession number GSE58061 by Nwafor et al 2014.

The expression of all identified transcripts were ranked by order of magnitude. In Brief, p-

values were computed to reflect the significance of the difference between two counts (n1 and n2 corresponding to any two library combination out of the six libraries) using a binominal model. The p-values were log-transformed in order to allow for greater numerical stability in comparing extreme values. Next all the p-values and the ratios of expression between the counts were considered to compute a ranking value for each transcript (Nwafor et al 2014, Appendix 8).

Raw uniquely mapped read counts for the wild-type and the mutant were independently subjected to differential expression (DE) analysis in a pairwise comparison between developmental stages (E-L 15 vs E-L 27, E-L 27 vs E-L 38 and E-L 15 vs E-L 38) using the software DESeq (Anders and Huber 2010) in R (parameters: false discovery rate (FDR) ≤ 5%, log2-fold change (FC) > 1). Next, DE genes were compared between the wild-type and the mutant. This strategy was preferred to the direct comparison of the two clones at each developmental stage in order to minimize the eventual differences due to asynchronous sampling.

An in-house R script was written to group DE genes with similar expression pattern based on the adjusted p-values. By indicating a significant up-regulation with "1", a significant down-regulation with "-1" and a non-significant difference with "0", the three comparisons between the developmental stages can be summarized with a triplet, e.g. “1, 0, 1”. This example indicates that there is a significant up-regulation going from the first to the second time point, no significant difference between the second and third time points, and a significant positive difference when comparing the first and last time points. Altogether, 27 different categories can be defined in this way, and 18 of these contain relevant patterns (for example the pattern “1, 1, -1” is impossible). These 18 groups are visualized in Figure 20. Each gene showing at least one significant difference between developmental stages was classified into one of these categories, for both the wild type and the mutant. The number of differentially expressed genes that fell to each pattern were compared between the wild-type and the mutant.

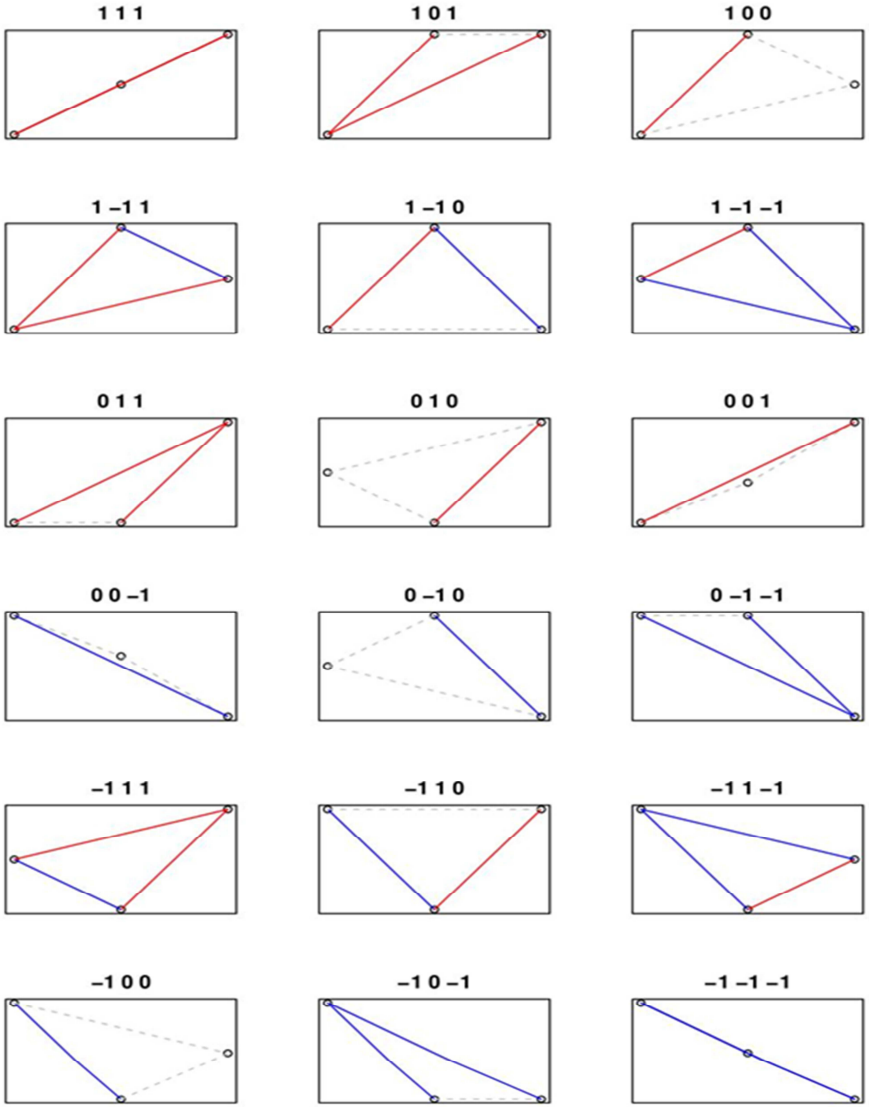


Figure 20. The eighteen relevant categories of triplets of significance.

5.1.8.2. Functional annotation and enrichment analysis

Wild-type and mutant genes were annotated against the v1 version of the 12x draft annotation of the grapevine genome using the CRIBI tools (link) combined with the grapevine molecular network VitisNet (Grimplet et al 2012). Next all DE genes for both genotypes were input into the AgriGO analysis tool (Du et al 2010). This allowed us to identify significantly enriched gene ontology (GO) terms in the whole set of DE genes or within each group when compared with GO terms in the complete Vitis vinifera genome. Using a hypergeometric test, a GO term was considered significantly enriched, if the FDR was < 0.05 and p-value < 0.01 when compared to all gene transcripts annotated in the reference genome (supported in AgriGO). Further, the REVIGO web server (Supek et al 2011) was used to summarize the processes represented in the lists of significantly enriched GO terms by removing redundant terms.

5.1.8.3. Selection of candidate genes

Candidate genes were chosen belonging to the three following groups:

I. Wild-type and mutant specific not DE genes, i.e. the transcripts which are expressed in the wild-type but not in the mutant and vice versa, with no significant differences between developmental stages. These genes were tested for GO annotation enrichment using AgriGO. Ultimately, genes were selected, if they fulfilled the following criteria: significant GO enrichment, RPM values above the lower limit of detection (0.5) and putative function relevant to seed development;

II. Wild-type and mutant specific DE genes, chosen based on their expression profile, fold change value, functional category enrichment, and putative function relevant to seed development. In addition, candidates were selected among DE genes with different expression profile or level of fold change in the two clones;

III. Candidate genes affecting seed content, previously identified in QTL analyses (Costantini et al 2008, Doligez et al 2013). These genes were compared with DE genes in the wild-type and the mutant, and the overlapping candidates were evaluated, based on their expression profile and the level of fold change.

5.1.8.4. Real-Time PCR validation of RNA-Seq data

Quantitative real-time PCR was carried out on cDNA obtained from both biological replicates described above, one of which was used for RNA-Seq. First-strand cDNA synthesis was performed with 1 µg of total RNA in triplicate using SuperScript™ III Reverse Transcriptase (Invitrogen, Carlsbad, CA) and oligo-dT according to manufacturer's protocol, after treatment

with DNase I (Invitrogen). The transcriptional profiles of 14 genes were analyzed. Sand and gadph (glyceraldehyde 3-phosphate dehydrogenase) were chosen as constitutive genes for normalization after evaluation of a set of five genes with the geNorm software (Vandesompele et al 2002). Their stable expression along development in the wild-type and the mutant was confirmed by RNA-Seq expression data. Details on gene IDs, gene annotations and primer sets are included in (Appendix 7-Table 1). Reactions were carried out with Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen) and specific primers using the LightCycler 480 (Roche Applied Science, Mannheim, Germany). The PCR conditions were: 95 ℃ for 5 min as initial step, followed by 50 cycles of 95 ℃ for 15 s, 68 ℃ for 30 s and 72 ℃ for 10 s. Finally, a post-PCR melting curve analysis was performed to verify the specificity of cDNA amplification. Each sample was examined in three technical replicates, and analyzed using the LightCycler 480 SV1.5.0 software (Roche Applied Science). REST 2009 software was used to calculate relative expression of each gene (Pfaffl et al 2002).

5.2. Results and discussion

5.2.1. Array-based SNP genotyping: 20K grapevine Illumina CHIP

Following visual inspection of clusters and filtering (see method in sections **5.1.2** and **5.1.3**), a total of 16563 SNPs displayed reliable cluster (Table 6). Analysis of pairwise comparison between the two clones revealed 16333 identical SNP loci. A total of 230 SNP loci were also identified with no call in either wild-type or mutant or in both.

Table 6. A pairwise comparison of SNP loci between the two clones

	Wild-type	**Mutant**	**Total**
SNP	20,000	20,000	20,000
Passing filters I (Illumina filters)	16,563	16,563	16,563
Identical SNP Passing filters II (Wild-type VS Mutant)	16,333	16,333	16,333

5.2.2. cDNA sequence alignment and mapping to the reference genome

Sequencing generated from 126 to 143 million and from 102 to 127 million 100-bp reads for the wild-type and the mutant, respectively (Appendix 7-Table 2). After pre-processing and quality control, the majority of reads from wild-type (≈ 79-81%) and mutant (≈ 70-81%) were successfully aligned to transcriptome (v1_mRNA version of the 12x draft annotation of the grapevine genome). Similar result was obtained for reads mapped to the genome assemble (wild-type ≈ 76-77% and mutant ≈ 75-76%), hence we report only reads aligned to the transcriptome (Appendix 7-Table 2).

For transcriptomic data analysis, a large fraction of mapped reads from each developmental stage for wild-type (≈ 87-89%) and mutant (≈ 85-87%) aligned to a single position. These uniquely mapped reads account on average for approximately 71% and 66% of the total number of sequenced reads for the wild-type and the mutant, respectively (Appendix 7-Table 2). Distribution of mapped reads among genomic features was similar for both reads mapped to the genome and transcriptome, therefore only reads mapped to the transcripts are shown. The results showed that a high proportion (49% for both the wild-type and the mutant) mapped to protein coding regions indicative of high coverage of actual transcribed sequences (Figure 21). The other reads mapped to splice junctions (27% and 26%), introns (14% and 16%) and untranslated regions (UTRs) (9% and 7%) for the wild-type and the mutant, respectively. The presence of intronic regions in RNA-Seq experiments is prevalent and has been attributed to various sources such as intron retention during splicing, DNA contamination during RNA-Seq preparation as well as alignment artefacts. Reads mapped to intronic regions in our data set are comparable to those obtained in similar experiments in grapevine (Zenoni et al 2010). Most of the intronic mapped reads in our data set show strand specificity, hence we infer they are mainly due to unspliced mRNA in our samples and others may be due to alignment artefacts.

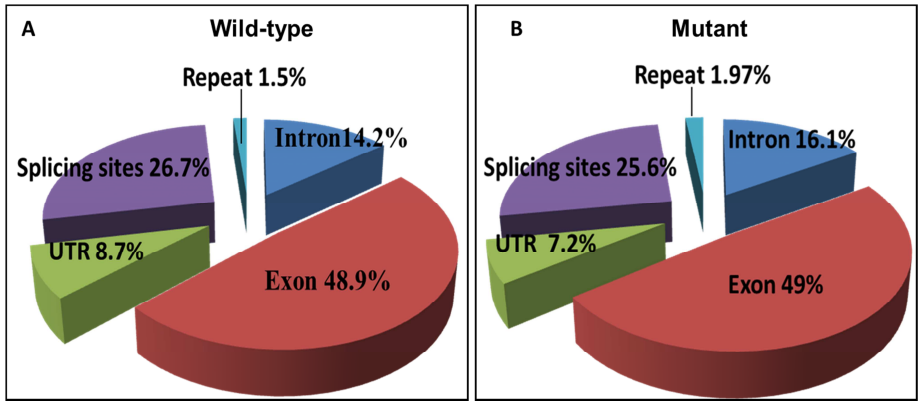


Figure 21. Distribution of mapped reads among genomic features. **(A)** Wild-type. **(B)** Mutant.

5.2.3. SNP detection in RNASeq variant call data

Summary of SNPs and INDELs predicted from the six libraries are shown in (Table 7). The predicted SNPs and INDELs are based on the reference sequence. A total of 71,557 SNPs and 37,121 INDELs satisfied the Initial filtering criteria described in Sections **5.1.6** and **5.1.7**. From this list, it was required for any position to be considered a candidate SNP, to be present in at least two libraries and to be different in wild-type compared to the mutant or vice versa (for instance, if such a SNP is homozygous it must be present in at least two libraries of either of the clones and heterozygous for one of the clones but not for both). This approach identified 1670 SNPs in at least two libraries. When combined with SNP selection based on putative functions relevant to the trait of interest, 142 candidate SNPs could be identified, from which 120 SNPs were selected for Sanger sequencing. Figures 22, 23, 24 and Tables 8 and 9 show various features used to characterize the variants called from RNA-Seq data. i.e. distribution of SNPs among genomic features, distribution of insertions and deletions length, coverage number of effects by impact and number of variants by functional class.

Table 7. Summary of variant call from RNA-Seq data

	SNP	INDEL
Total	373,407	206,050
Passing filtering criteria	71,557	37,121

5.2.4. Sanger validation of putative SNPs

To validate the putative SNPs identified by RNA-Seq, small DNA fragments (between 400 and 320 bp) from different genes in the same plants of wild-type and mutant used for RNA-Seq were sequenced.

A total of 120 putative loci were resequenced and 31 of them were true positives. Interestingly most of the true positives mapped to exonic regions and were located in genes that play significant role during berry development, while most of the false positives SNPs were those that appeared mainly in one library and occasionally in two libraries. Similarly individual inferred genotypes from RNA-Seq were check for concordance with Sanger method. Approximately fifty percent (50%) of the total inferred genotypes were in agreement with Sanger data (Appendix 7- Table 3 and Figure 1).

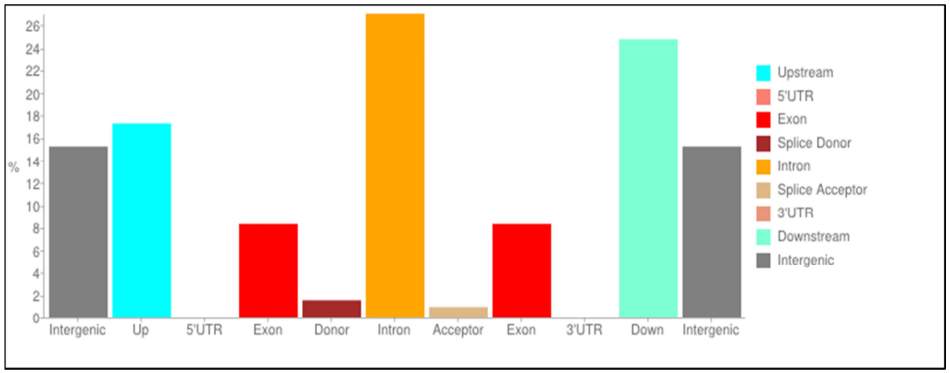


Figure 22. Distribution of variant effects by type and genomic region.

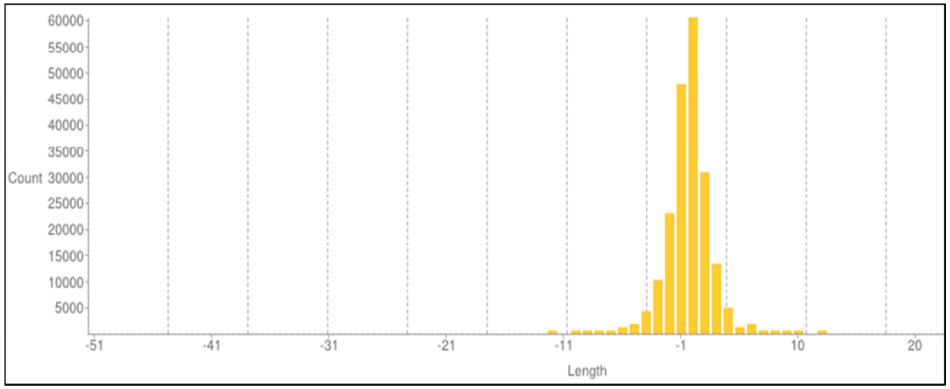


Figure 23. Insertions and deletions length.

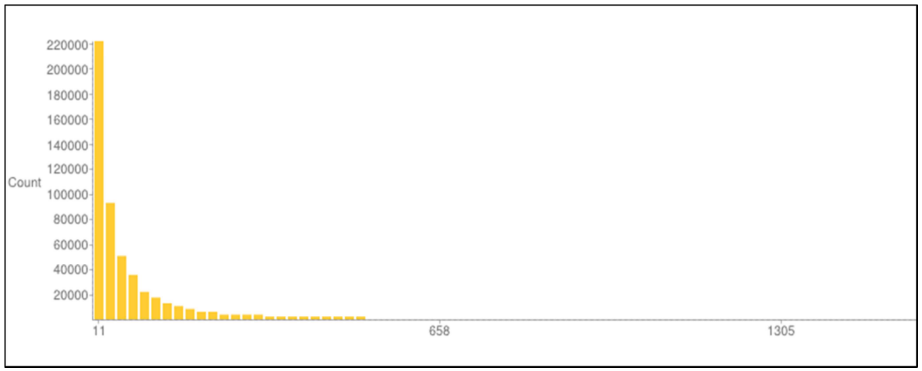


Figure 24. Variant coverage.

Table 8. Number of variant effects by impact

Type (alphabetical order)	Count	Percent
HIGH	38,926	3.4%
LOW	71,491	6%
MODERATE	46,861	4%
MODIFIER	1,006,336	86.5%

Table 9. Number of variant effects by functional class

Type (alphabetical order)	Count	Percent
MISSENSE	44,031	51.4%
NONSENSE	656	0.8%
SILENT	41,048	48%

5.2.5. Gene expression analysis

The digital, count-based nature of RNA-Seq provided a number of potential advantages for downstream data analysis and interpretation. For every gene detected in wild-type and mutant samples, uniquely mapped reads were used to generate raw expression counts and normalized expression values. The normalized expression values were calculated as RPM since it provides a useful way to assess overall expression levels between samples. Following the normalization of read counts, we analysed the most abundant transcripts within our samples by ranking them based on their p-value and ratio of expression. This in turn highlighted the top most highly expressed genes across all possible pairwise comparisons of the libraries (These data are reported in Nwafor et al 2014).

Overall the data-set identified approximately 98% of grapevine annotated transcripts (representing 27,495 genes) expressed throughout the three developmental stages under study. We detected a gene expression gradient from “before flowering” to “after flowering”, i.e. for wild-type E-L 15 (25,785 expressed genes) >E-L 27 (25,706 expressed genes) >E-L 38 (24,822 expressed genes) and for mutant E-L 15 (25,848 expressed genes) >E-L 27 (25,197 expressed genes) >E-L 38 (24,089 expressed genes) (Table 10).

To put these results into perspective, slightly more genes were expressed before fertilization in the mutant than in the wild-type and by far more genes were expressed after fertilization in the wild-type than in the mutant. In the wild-type and the mutant 23,640 and 23,072 genes were expressed in all three developmental stages, respectively (Figure 25). While it is not surprising the comparable number of genes shared by the three developmental stages in each clone, it is interesting to note that fewer genes were expressed specifically at each developmental stage: 586, 430 and 421 genes at stages E-L 15, E-L 27 and E-L 38 in the wild-type (Figure 25A) and 802, 337 and 351 genes at respective stages in the mutant (Figure 25B), which further highlights a reduction in gene expression in the mutant compared to the wild-type after fertilization. Thus we assessed what proportion of the expressed genes were common to both clones in the different stages and found that large number of expressed genes were shared among the wild-type and the mutant throughout development. In particular, 22,516 genes were commonly expressed in both clones in all three developmental stages (Table 11), 24,084 in the first two stages E-L 15 and E-L 27 (Figure 26A) and 22,790 in the last two stages E-L 27 and E-L 38 (Figure 26B). This was expected based on the phenotypic evaluation of the two clones that revealed similar berry development and ripening

(they were at the same developmental stage in the same date). Nevertheless, a fewer number of genes were exclusively expressed in a particular developmental stage and clone (Table 11), suggesting they could be responsible for the specificity of each clone. Finally, a total of 565 genes were not expressed at all (Table 11). This set of genes could be genotype specific and restricted to the grapevine clone PN40024 used for reference mapping.

The results of differential gene expression analysis of RNA-Seq data in the pairwise comparison between developmental stages are shown in Figure 27. In total 1075 genes were differentially expressed (DE) in both clones. With respect to the wild-type a total of 942 genes were found to be differentially expressed during development: 522 between stages E-L 15 and E-L 27, 354 between stages E-L 27 and E-L 38 and 393 between stages E-L 15 and E-L 38 (Figure 27A). For the mutant a total of 634 DE genes were identified: 458 between stages E-L 15 and E-L 27, 191 between stages E-L 27 and E-L 38 and 41 between stages E-L 15 and E-L 38 (Figure 27B). Analysis of data set overlap (Nwafor et al 2014) revealed that about 47% of the total DE genes (501/1075) were expressed in both the wild-type and the mutant (commonly shared expression), which supports the developmental alignment of the two clones. More strikingly, the percentage of DE genes specific to the wild-type with respect to all three developmental stages is 41% (441/1075), while for the mutant it is 12% (133/1075). We further evaluated the percentage of significantly up-regulated and down regulated genes in each pairwise comparison in both the wild-type and the mutant. On average approximately 67% of DE genes in the wild-type and 75% of DE genes in the mutant were down-regulated along development, while 33% and 25% of DE genes were induced in the wild-type and the mutant, respectively (Table 12). Taken together these results suggest that most of the expressed genes were active in different contexts along the grape berry developmental gradient (Table 10). However, significant quantitative changes occurred in individual gene expression level that corresponds to a particular stage or switch in development during seed formation. Here the mutant exhibited the strongest reduction in gene expression after fertilization (Table 12). It is tempting to speculate that it might be due to shut down in transcriptional processes resulting from incomplete fertilization or failure of embryo development. However, further work will be necessary to test this hypothesis.

Finally, we determined the expression pattern of all DE genes over the three developmental stages under investigation using the technique described in the methods. This approach revealed transcripts from a pool of DE genes that exhibit the same patterns of expression

over the three developmental stages. We present here 18 relevant groups (Figure 19). The wild-type and the mutant exhibited similar differential expression pattern except in groups 6, 10 and 18. Four main groups (3, 11, 12 and 16), accounted for about 67% of the DE genes along the three developmental stages of the wild-type. Similarly, groups 3, 11 and 16 accounted for 87% of DE genes in the mutant (Table 13). Additionally the analysis of expression pattern of all DE genes enabled us to identify relevant groups showing significant difference in the number of DE genes between the two clones, such as groups 2, 9, 10, 12 and 17 (Table 13).

5.2.6. Functional enrichment analysis

To assess the biological meaning of the wild-type and the mutant differential expression pattern, we examined representation of GO terms in the whole set of DE genes and within each of the eighteen groups. When considering the whole set of DE genes the most striking difference between the two clones was the wild-type specific enrichment in GO terms related to reproduction, such as anther wall tapetum development, cell division and microsporogenesis (see Nwafor et al 2014).

When considering the DE gene in each of the eighteen groups, for the wild-type we detected a number of significantly enriched GO terms in groups 3, 11, 12, 16 and 17, whereas in the mutant significantly enriched GO terms were found only in groups 11 and 16 (however, many of the GO terms in the wild-type group 17 were present in the mutant group 16) (These data are reported in Nwafor et al 2014). For example, we observed a specific significant enrichment of positively regulated (from stage E-L 15 to stage E-L 27) functional categories in the wild-type group 3, for which the genes were mainly related to cell wall modification. Here stage E-L 27 corresponded to "after fertilization", a phase of berry development mainly characterized with extensive cell division. Perhaps it is likely that these genes were highly active in the wild-type and may have played important role in cell wall re-assembly to encourage cell division during seed formation and embryo development.

5.2.7. Real-time PCR validation of RNA-Seq data

To confirm the results obtained by RNA-Seq, relative expression profiles of 14 genes were analysed by real-time PCR in the wild-type and the mutant. The tested genes encoded enzymes involved in cell wall metabolism, transcription factors from different families (MYB, MADS-346 box, PHD and AS2) and molecules playing a role in signalling, including hormone-

mediated signalling. For both clones and all genes, the real-time PCR results were consistent with the expression profiles determined from RNA-Seq data. Seven genes had similar expression profiles in the wild-type and the mutant, while the expression of the remaining 7 genes ranged from slightly different to completely opposite which suggests that some pathways may be altered in the seedless phenotype (Figure 28). In most cases biological replicates showed a consistent expression profile.

5.2.8. Selection of candidate genes

In this work gene expression analysis highlighted several genes with common and contrasting expression profiles in the two clones, which may contribute to trait variation (seed content, and the resulting berry size, are the only phenotypic differences between the two somatic variants). Therefore, in order to narrow down to specific genes whose expression and effect were altered in the seedless phenotype, we have applied the criteria described in Methods section **5.1.8.3**. This allowed us to select a number of candidate genes for the seedless phenotype, which are listed in Table 14 and described in detail in Chapter 4. Among them are genes required for fertility, cell growth and development, transcription factors and signalling molecules.

Table 10. Transcript abundance measurement at each developmental stage

	Wild-type			**Mutant**		
	E-L 15	**E-L 27**	**E-L 38**	**E-L 15**	**E-L 27**	**E-L 38**
RPM >200	873	873	890	888	869	861
RPM 10-200	11800	11748	9869	11681	11583	9412
RPM 0.5-10	7415	7583	7765	7596	7443	7555
RPM <0.5	5697	5502	6298	5683	5302	6261
Total detected	25785	25706	24822	25848	25197	24089

Table 11. Comparison of gene expression between the wild-type and the mutant

	Wild-type			Mutant		
Developmental stage	E-L 15	E-L 27	E-L 38	E-L 15	E-L 27	E-L 38
Genes expressed in all developmental stages in the two clones (common genes)	22516	22516	22516	22516	22516	22516
Exclusively uniquely expressed genes for each developmental stage	183	187	169	190	70	97
Non-detected expression for each developmental stage	1145	1224	2108	1082	1733	2841
Constitutively non-expressed genes in the two clones	565	565	565	565	565	565

Table 12. Evaluation of significantly up- and down-regulated genes in each pairwise comparison between developmental stages

	Wild-type						**Mutant**					
Pairwise comparison	E-L 27 vs E-L 15	Percentage	E-L 38 vs E-L 27	Percentage	E-L 38 vs E-L 15	Percentage	E-L 27 vs E-L 15	Percentage	E-L 38 vs E-L 27	Percentage	E-L 38 vs E-L 15	Percentage
Down-regulated genes	332	63.6	256	72.3	256	65.1	327	71.4	136	71.2	34	82.9
Up-regulated genes	190	36.4	98	27.7	137	34.9	131	28.6	55	28.8	7	17.1
Total	522		354		393		458		191		41	

Table 13. Number of genes in each group of differential expression patterns for the wild-type and the mutant

	Gene pattern	Number of genes (Wild-type)	Number of genes (Mutant)
Number of groups			
1	111	0	0
2	101	13	4
3	100	155	112
4	1-11	0	0
5	1-10	21	15
6	1-1-1	1	0
7	011	66	26
8	010	30	25
9	001	58	4
10	00-1	34	0
11	0-10	101	118
12	0-1-1	131	3
13	-111	0	0
14	-110	2	4
15	-11-1	0	0
16	-100	240	319
17	-101	88	4
18	-1-1-1	2	0

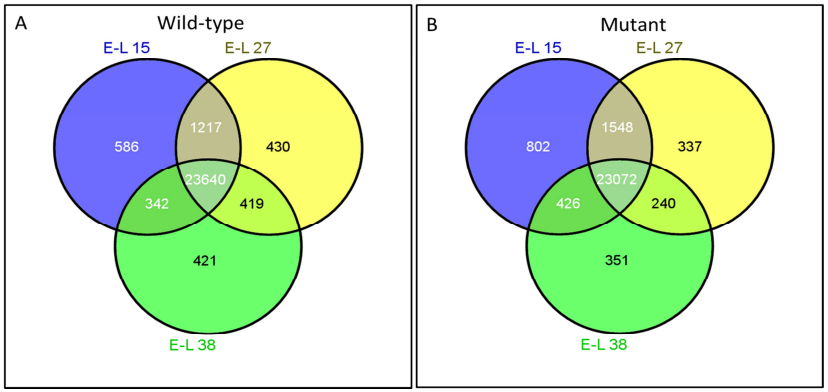


Figure 25. Gene expression overlap between the three key developmental stages in **(A)** wild-type and **(B)** mutant.

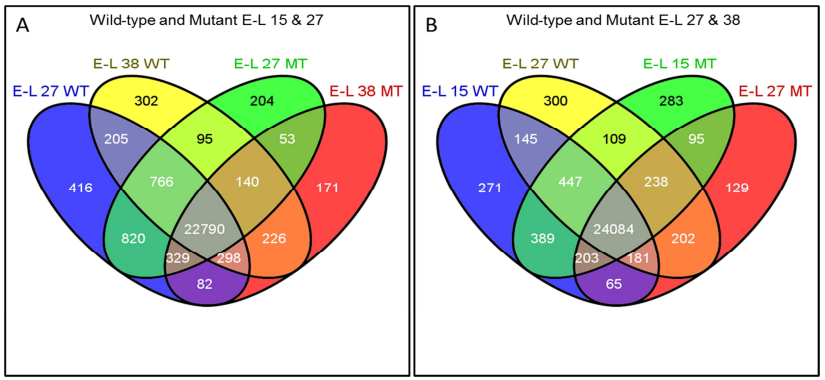


Figure 26. Gene overlap between the wild-type and the mutant in the first two and last two developmental stages. (A), Venn diagram showing shared and unique expressed genes between the wild-type and the mutant during the first two developmental stages E-L 15 and E-L 27. (B), Venn diagram showing shared and unique expressed genes between the wild-type and the mutant during the last two developmental stages E-L 27 and E-L 38.

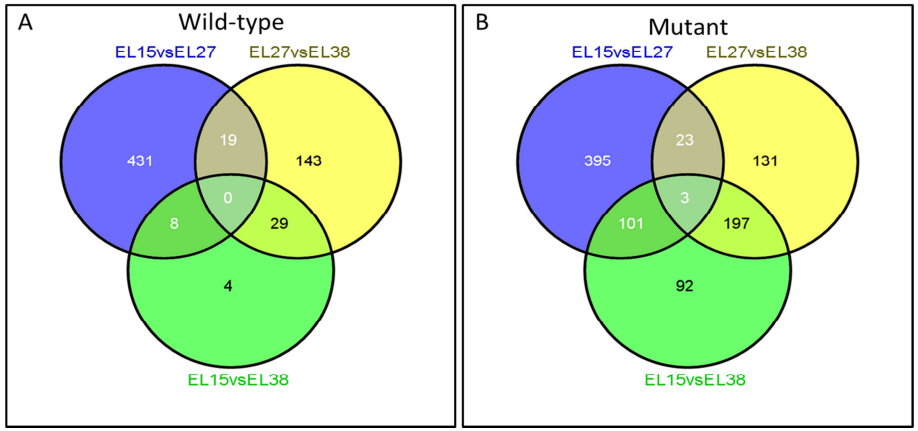


Figure 27. Comparison of differential gene expression in the pairwise comparison of developmental stages in wild-type and mutant plants. Venn diagrams indicate overlap of all differentially expressed genes obtained from each pairwise comparison between developmental stages (E-L 15 vs E-L 27, E-L 27 vs E-L 38 and E-L 15 vs E-L 38) in wild-type **(A)** and mutant **(B).**

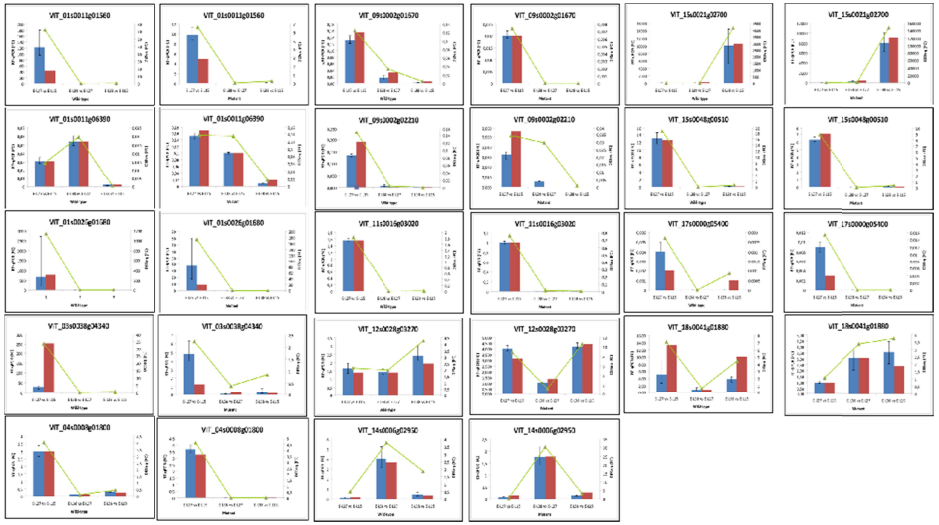


Figure 28. Quantitative real-time PCR validation of RNA-Seq data. Relative expression profile of 14 genes shows the expression fold change (FC) in the pairwise comparison between developmental stages for the wild-type and the mutant. Histograms represent expression fold changes as assessed by real-time PCR (by using REST), data are reported as means ± SE of three technical replicates (left axis). Green lines represent expression fold changes as assessed by RNA-Seq (by using DESeq, right axis). Blue column with error bar corresponds to the first biological replicate, while red column corresponds to the second biological replicate on which RNA sequencing was carried out.

Table 14. Candidate genes for seed content that have altered expression in the wild-type and the mutant. Abbreviations: nd= not detected in a pairwise comparison, inf= infinity (when the mean of one stage in a pairwise comparison is the denominator with value 0), sig = significant.

Gene ID	**Wild-type Gene Expression (RPM)**			**Mutant Gene Expression (RPM)**			**Wild-type Fold Change**			**Mutant Fold Change**			**Gene enrichment**	**Annotation**
	E-L 15	E-L 27	E-L 38	E-L 15	E-L 27	E-L 38	E-L 27 vs E-L 15	E-L 38 vs E-L 27	E-L 38 vs E-L 15	E-L 27 vs E-L 15	E-L 38 vs E-L 27	E-L 38 vs E-L 15		
Non-DE genes specific to the wild-type														
VIT_09s0002g01980	0.5	0.7	0.2	0	0	0								Myosin-like protein XIK
VIT_15s0048g01070	0.01	0.01	1.2	0	0	0								Vacuolar iron transporter 1
VIT_04s0044g01520	0	0.8	0	0	0	0								GA 20-oxidase 2
VIT_08s0058g01200	0	0.4	0	0	0	0							sig	Alpha-expansin 2
Non-DE genes specific to the mutant														
VIT_13s0106g00290	0	0	0	0.01	0	0.1								Histone deacetylase HDA14
VIT_03s0088g00900	0	0	0	0	0.1	0.01								Pathogenesis-related protein 1B
VIT_14s0006g00050	0	0	0	0.1	0	0								Transposase, IS4
Common genes differentially regulated in wild-type and mutant														
VIT_01s0026g01680	0.02	24.6	0	0.01	1.9	0	1,133	0	nd	172.5	nd	nd	sig	Pectate lyase
VIT_05s0020g04850	1.0	2.4	89.1	0.9	0.9	101.8	nd	nd	113.1	nd	153.3	nd	sig	H1flk
VIT_15s0021g02700	0.3	7.4	1,103	0	2.0	1,396	nd	nd	4,154	inf	nd	nd	sig	Beta-expansin (EXPB4)
VIT_15s0048g00510	9.1	170.5	4.4	8.8	74.9	2.4	18.8	nd	nd	8.7	nd	nd		Pectinesterase family
VIT_15s0021g02170	631.1	3.1	0.3	81.4	0.2	0.4	0.005	nd	0.0006	0.003	nd	nd		Chalcone and stilbene synthase

VIT_18s0089g00140	40.7	0.2	2.7	3.5	0	8.0	0.004	nd	nd	0	nd	nd		1,4-beta-mannan endohydrolase
VIT_19s0015g00960	150.2	0.7	0	59.3	0.1	0.1	0.004	nd	0	0.002	nd	nd		ABC transporter G member 4
VIT_18s0001g01760	969.1	8.1	0.04	854.3	3.2	1.0	0.008	nd	0.00005	0.004	nd	nd		PISTILLATA (PI) floral homeotic protein
VIT_18s0001g13460	107.1	13.8	0	84.4	7.9	0.1	0.1	nd	0	0.1	nd	nd		MADS-box AP3
Differentially expressed genes specific to the wild-type														
VIT_01s0011g06390	29.9	0.4	0.01	3.2	0.4	0.04	0.01	nd	0.0004	nd	nd	nd	sig	Male sterility 1
VIT_08s0007g07100	20.5	0.8	0.02	3.7	0.06	0.04	0.04	nd	nd	nd	nd	nd	sig	Male sterility 2
VIT_07s0005g05680	17.6	0.2	3.5	4.0	0.2	0.04	0.009	nd	nd	nd	nd	nd		Male sterility 5
VIT_07s0005g05720	29.3	1.6	2.3	14.5	2.6	1.2	0.06	nd	nd	nd	nd	nd		Male sterility 5
VIT_15s0107g00550	172.2	19.7	21.8	111.7	26.7	2.5	0.1	nd	nd	nd	nd	nd	sig	Male sterility 5
VIT_19s0014g03940	7.7	0.2	0.06	4.5	0.3	0.06	0.02	nd	nd	nd	nd	nd		Sporocyteless
VIT_12s0142g00040	49.6	2.6	0.07	12.5	0.7	0.08	0.05	nd	0.002	nd	nd	nd		Glycerol-3-phosphate acyltransferase 1
VIT_00s1404g00010	17.5	22.6	0.04	18.1	14.6	0.04	nd	0.002	nd	nd	nd	nd	sig	Calmodulin-binding
VIT_01s0026g01420	22.2	58.8	0.3	22.8	57.2	0.2	nd	0.007	nd	nd	nd	nd		Wall-associated kinase 4
VIT_06s0061g00730	406.0	887.8	0.3	423.5	1,138	0.2	nd	0.0004	0.0009	nd	nd	nd		Aquaporin GAMMA-TIP3/TIP1;3
VIT_18s0001g13200	31.7	143.9	0.3	38.7	110.9	0.5	nd	0.003	nd	nd	nd	nd		Cytokinin dehydrogenase 5 precursor
VIT_05s0094g00330	0.9	16.2	969.7	0.9	4.2	712.2	17.6	nd	1,332	nd	nd	nd		Chitinase, class IV
VIT_10s0003g03030	0.02	6.5	0.4	0.06	0.5	0.6	300.4	nd	nd	nd	nd	nd		Cation/hydrogen exchanger (CHX15)
VIT_01s0011g01560	0.2	14.2	0.07	0.4	2.4	0.07	72.6	nd	nd	nd	nd	nd		Transparent testa16
VIT_18s0001g03010	0	1.5	0	0	0.2	0	inf	nd	nd	nd	nd	nd		BZIP transcription factor
VIT_18s0041g01880	9.5	67.4	33.2	9.8	10.7	26.7	7.1	nd	nd	nd	nd	nd	sig	MADS-box protein SEEDSTICK

VIT_03s0038g04340	0.2	6.9	0.07	0.3	0.6	0.2	33.7	nd	nd	nd	nd	nd		Feronia receptor-like kinase
VIT_17s0000g08110	1.1	23.1	0.01	0.2	1.4	0	20.1	0.0006	nd	nd	nd	nd		Nodulin MtN3
VIT_17s0000g09000	0.8	0.4	36.3	0.5	0.009	0.04	nd	125.8	nd	nd	nd	nd		Oleosin OLE-2
VIT_07s0151g00640	12.4	5.2	453.2	17.9	3.0	1.1	nd	109.3	nd	nd	nd	nd		Globulin-1 S allele precursor
VIT_14s0128g00200	0.04	0.1	42.5	0.09	0	0	nd	540.1	1,242	nd	nd	nd		7S globulin precursor
VIT_13s0067g01250	0.06	0.1	26.0	0.02	0	0	nd	330.4	506.5	nd	nd	nd		Em protein GEA6 (EM6)
VIT_14s0108g00520	0	0.2	89.3	0.03	0.09	0.03	nd	637.6	inf	nd	nd	nd		Protease inhibitor/seed storage/lipid transfer protein (LTP)
VIT_16s0039g00220	0.4	0.2	25.2	0.3	0.1	0.7	nd	192.0	nd	nd	nd	nd		Aquaporin BETA-TIP
VIT_07s0005g05400	0	0.08	22.1	0.05	0	0	nd	360.2	inf	nd	nd	nd		Abscisic acid-insensitive protein 3 (ABI3)
VIT_19s0014g04130	0.7	0.2	23.6	0.3	0.4	6.4	nd	179.8	nd	nd	nd	nd		Serine/threonine-protein kinase receptor ARK3
VIT_18s0001g01570	0.2	0.5	265.1	0.3	0.07	0.3	nd	631.3	1,548	nd	nd	nd		Seed maturation protein PM31
VIT_14s0128g00340	0.1	0.08	17.0	0.1	0	0	nd	277.3	nd	nd	nd	nd		Seed maturation protein PM34
VIT_04s0008g01610	0	0.3	176.8	0	0	0	nd	776.9	inf	nd	nd	nd		Heat shock protein 17.6 kDa class II
Differentially expressed genes specific to the mutant														
VIT_14s0219g00270	20.3	2.5	0.08	21.6	0.5	0.08	nd	nd	nd	0.03	nd	nd		TEL1 (Terminal EAR1-like 1)
VIT_12s0059g00560	14.6	1.1	1.1	17.8	0.6	0.4	nd	nd	nd	0.04	nd	nd		Fimbrin 2
VIT_04s0008g04980	15.5	2.7	0.2	10.1	0.3	0.06	nd	nd	nd	0.03	nd	nd		Boron transporter-like protein 4
VIT_09s0002g01670	9.0	1.1	0.03	11.8	0.09	0	nd	nd	nd	0.007	nd	nd		Myb domain protein 26
VIT_09s0002g01370	13.2	1.3	0.2	14.0	0.3	0	nd	nd	nd	0.02	nd	nd		AP2 AINTEGUMENTA
VIT_14s0006g02950	25.4	12.6	37.4	53.1	5.4	123.3	nd	nd	nd	0.1	nd	nd		Lateral organ boundaries protein 41

VIT_15s0046g03080	3.4	0.9	0.02	2.9	0	0	nd	nd	nd	0	nd	nd		DTA2 (downstream target of AGL15 2)
VIT_12s0134g00240	17.7	36.7	21.5	9.7	77.1	22.7	nd	nd	nd	8.1	nd	nd		Avr9/Cf-9 rapidly elicited protein 20
VIT_12s0028g03270	14.2	30.5	48.0	7.3	62.5	55.7	nd	nd	nd	8.8	nd	nd		Ethylene-responsive transcription factor 9
VIT_16s0013g00950	1.8	5.1	2.1	1.1	22.1	0.4	nd	nd	nd	20.5	nd	nd		Ethylene-responsive transcription factor ERF105
VIT_16s0013g00990	2.4	5.0	0.7	1.2	18.6	0.4	nd	nd	nd	16.1	nd	nd		Ethylene-responsive transcription factor ERF105
VIT_16s0013g01050	3.1	6.3	1.1	1.8	23.3	0.3	nd	nd	nd	13.2	nd	nd		Ethylene-responsive transcription factor ERF105
VIT_16s0013g01120	3.2	13.2	63.7	3.1	29.5	49.1	nd	nd	nd	9.8	nd	nd		Ethylene-responsive transcription factor ERF105

6. General discussion and conclusions

The work described in this thesis is part of the on-going global effort to characterize cellular and molecular events that lead to seedless forms of grape. Seedlessness is an important economic trait for table grape market and understanding the genetic processes that underpin seedlessness is justified by this economic reason (Costenaro-da-Silva et al 2010).

This thesis proposed a new approach for unravelling the genetic processes that underpin seedlessness, see (Chapter 3). This approach was tested and validated in (Chapters 4 and 5). Basically the approach draws strength from two main sources: first the exploration of the phenotypic variation that exists in grapevine germplasm and secondly the application of Next Generation Sequencing (NGS) technology.

A mutant seedless form of Sangiovese, mistaken for another cultivar, was confirmed in (Chapter 4) proving that, for proper identification of cultivars, it is crucial not to rely only on traditional ampelographic data. In fact the combination of molecular (SSR) and morphological data provides a robust evidence for ascertaining a cultivar's true identity, an approach that is widely adopted not only for confirming trueness to type but also for kinship, pedigree and genetic diversity studies (Schneider et al 2001, Schneider et al 2009, Gasparro et al 2013, Emanuelli et al 2013).

Equally the availability of the somatic variant afforded the extraordinary opportunity to compare phenotypic and genotypic variation between a seeded and a seedless grapevine cultivar. Not only that, it provided a rare insight to the temporal and spatial changes in the transcriptomes of the two clones , with a special emphasis on the expression levels of key regulatory genes. This has never been done before exhaustively in grapevine. A lot of differentially expressed genes were identified including those involved in gametophyte development, cell cycling, transcription factors and signalling molecules. Hereafter major aspects of the thesis are discussed.

6.1. Investigating the physiological process responsible for seedlessness in a Sangiovese seedless somatic variant

The availability of this seedless somatic variant, see (Chapter 4), enabled the comparison of quantitative and qualitative traits that relate to berry development and seed content in the seedless somatic variant and its seeded type. Similarly it allowed the investigation of two main physiological processes known to cause seedlessness in grapevine (Parthenocarpy -like

in cv Corinto and Stenospermocarpy -like in cv Sultanina). In addition the heritability of the seedless trait coupled with the viability of the pollen of the seedless line could be tested, see (Chapter 4).

The results of quantitative and qualitative analysis of berry- and seed- related traits (Figures 5-9) and data from investigation of the two main physiological processes (Figures 10-16) suggest that the mutant is a stenospermocarpic cultivar. This could imply that stenospermocarpy is not restricted to Sultanina-derived cultivars because in most cases seed traces were detected in the mutant berries, and some normally developed seeds were observed, perhaps they were derived from non-aborted embryos. Whether the seedless phenotype of the mutant is heritable or not, remains to be confirmed as various trials and experiments to test this hypothesis failed. Here, embryo rescue experiments could be exploited in the future for the seeds obtained through crossing (Table 2). However pollen viability test suggests the mutant produced distorted and high variable sized pollen grains compared to the wild-type. Therefore it was concluded that the mutant pollen is not efficient perhaps defective and may be partly responsible for the seedless phenotype. However genetic validation of this hypothesis is required. For example pollen development marker genes (i.e. genes known to control morphometric patterning of pollen during development) could be compared in the two clones for variation in expression using several time points between before anthesis and at full bloom.

6.2. Berry development after emasculation in both wild-type and mutant

The development of berries after emasculation was first observed in 2012 and subsequently in 2013 and 2014 growing seasons. This phenomenon was indeed surprising, at first it was thought to be due to out-pollination mediated by insects or wind. Although *Forficula auricularia* were found in some flowers, however at the date of emasculation there was no pollen in the air since flowers were still closed. Also parthenocarpy and stenospermocarpy were suspected to be responsible for this behavior because most of the berries from the emasculated groups were seedless. But in grapevine, if parthenocarpy is "stimulative" (as usually thought to be), the development of berries in absence of pollen stimulation (not fertilization) would be impossible; if it is a "vegetative" parthenocarpy, it is perhaps possible, but this requires experimental prove or a confirmation from an expert grapevine physiologist. Nevertheless, normally developed seeds were observed, whereas in parthenocarpic berries seeds are

expected to be completely absent. Stenospermocarpy seems also unlikely because emasculation and inflorescence covering should have excluded any presence of pollen and consequently any fertilization event.

Alternative possibility are cleistogamy (Sampson et al 2001) and apomixis, (Koltunow et al 1995, Vielle-Calzada et al 1996, and Spillane et al 2001.).

Cleistogamy is an automatic self-pollination process, relating to flowers that does not open before fertilization, here self-pollination occurs in the bud early before anthesis. Evidence of this phenomenon in grapevine was reported by Sampson et al (2001).

In briefly, apomixis is said to be asexual reproduction through seed, in which meiosis precedes the formation of gametes, and double fertilization restores the somatic chromosome number. The resulting seed will have a genotype identical to that of its maternal parent. Two types of apomixis exist, namely: gametophytic and sporophytic. They both depend on the fate of the unreduced cells. If the unreduced cells give rise to a megagametophyte, then gametophytic apomixis occurs. If the unreduced cells give rise directly to an embryo, then sporophytic apomixis occurs (Spillane et al 2001, Vielle-Calzada et al 1996).

Elsewhere apomixis has been tested in grapevine. Chkhartishvili et al (2006) reported fruit-set was not observed for the emasculated flowers, suggesting it is not a characteristic phenomenon in grapevine. However the cultivars under study are different from those reported by Chkhartishvili et al (2006) and we assumed apomixis as a plausible cause of normal development of seeds after flower emasculation.

As convenient as it is to speculate on a possible role of apomixis considering all the scenario described above, care must be taken not to draw conclusions as there may be other unknown factors that could have possibly triggered normal seed development after emasculation and bunch covering. Maybe early cleistogamy, though this has been mentioned in grapevine before. One possible way to test the apomixis hypothesis, is to genetically characterize the embryo of the seeds from the berries that developed after emasculation, to determine if they originate from the maternal genome or not, however the major challenge here would be to accurately isolate the embryo without contamination with the testa and endosperm which are maternally derived.

6.3. Identification of genomic variations between wild-type and mutant

Recent advances in the Next Generation Sequencing technology (NGS) have changed the

way and manner genomes and transcriptomes are studied. For instance, the ability to use different templates (genomic DNA versus cDNA) allows for the study of diverse biological questions; considering the determination of mRNA sequences after conversion to cDNA (mRNA-Seq) in particular is proving invaluable for expression profiling and genome annotation. Many examples are cited in (Chapter 5), highlighting the potential of NGS application in grapevine.

To unravel the genotype/phenotype relationship between the wild-type and mutant we relied on effective identification of genomic variants. Here, two NGS techniques were applied (20K grapevine Illumina CHIP and RNA-Seq Variant call) to identify sequence variation that may be inherent in the two clones. The 20K grapevine Illumina CHIP is an array-based SNP genotyping method while RNA-Seq is a whole transcriptome sequencing approach.

Result of 20K grapevine Illumina CHIP experiment supported earlier result obtained from molecular marker analysis (SSR) confirming the mutant to be identical to the wild-type see (Chapter 5). This result is particularly significant in highlighting the true identity of the mutant. Although the SNPs used in the 20K grapevine Illumina CHIP were pre-determined and validated, the number of reliable SNPs (16,333) from this study is obviously too small to unequivocally say there are no sequence variations between the genomes of the two clones.

Variant calling from RNA-Seq is arguably a more cost effective means of identifying differences at a whole-genome level for somatic variants. However this approach is rife with transcriptome's intrinsic complexity, mainly due to splicing, which leads to the various technical difficulty during computational analysis. Nevertheless, variant calling from both wild-type and mutant transcriptomes (RNA-Seq libraries) identified several hundred thousands of SNPs and INDELs based on the reference genome. Because these SNPs and INDELs are reference-based many of them could be common in both clones including false positives. Here a direct comparison of all SNPs and INDELs found in both clones was a way of identifying the variants that differ in the two clones. Using the approach described in (Chapter 5) several putative SNPs were identified and confirmed by Sanger sequencing. In general these SNPs could be used to differentiate the two clones. Majority of the validated SNPs are " coding SNPs" (within protein-coding regions).

6.4. Evaluation of common and contrasting expression profiles of DE genes in wild-type and mutant: A candidate gene approach

6.4.1. Non-DE genes specific to the wild-type

Within this category very few genes met the RPM selection criteria, however many genes were significantly enriched and some of them had a putative functional role relevant to seed development. We selected four genes that play roles in cellular process, transport and signalling. Among cellular process genes, VIT_09s0002g01980 encodes the myosin-like protein XIK, which is involved in actin organization and biogenesis as well as actin-driven movement (Grimplet et al 2012). Among transporters, the gene VIT_15s0048g01070 encodes the vacuolar iron transporter 1 protein, implicated in iron transport and storage (Grimplet et al 2012). In seeds, iron has been demonstrated to be essential for Arabidopsis embryo development (Stacey et al 2008). Among the signalling genes are VIT_04s0044g01520 and VIT_08s0058g01200. VIT_04s0044g01520 encodes GA 20-oxidase 2, which is involved in gibberellic acid biosynthesis, whereas VIT_08s0058g01200 codes for the alpha-expansin 2 protein that participates in auxin-mediated signalling pathway as well as regulating cell growth (Grimplet et al 2012).

6.4.2. Non-DE genes specific to the mutant

All the genes that fell within this category did not meet the RPM selection criteria described in the Methods and did not have defined function when annotated; meaning that, many of them returned no hit upon functional annotation. Nevertheless, we noticed a few genes whose functional roles could be implicated in seed development. They included the histone deacetylase *HDA14* gene (VIT_13s0106g00290), involved in chromatin organization through protein acetylation and deacetylation, a gene (VIT_03s0088g00900) coding for a pathogenesis-related protein 1B implicated in jasmonate-mediated signalling as well as in plant-pathogen interaction and a transposase IS4 gene (VIT_14s0006g00050) that encodes a transposable element protein (Grimplet et al 2012).

6.4.3. Differential regulation of common transcriptional processes in the wild-type and the mutant

Significant number of expressed genes were common among wild-type and mutant growth stages, which suggests that the corresponding proteins may function in a common pathway to carry out a wide range of developmental processes. We reasoned that many of these shared genes will respond in both clones to the same signals that control the switch from one

developmental phase (before fertilization) to another (after fertilization), and will have similar pattern of expression. Indeed, differential expression analysis revealed 501 DE genes common to the wild-type and the mutant (47% of the total 1075 DE genes). Thirty-five of these genes showed different expression between the two clones along the time course. Among the 35 genes, six were significantly enriched and three of them had a functional annotation corresponding to seed development: pectate lyase, histone H1flk-like protein (H1flk), and beta-expansin (EXPB4). Pectate lyase is an enzyme involved in cell wall organization and biogenesis by catabolizing pectin. In tomato, two pectate lyases were found to be maximally expressed at the late stage of pollen development. It was suggested that the pollen expression of these genes might relate to a requirement for pectin degradation during pollen tube growth (Wing et al 1990). In the present study, the pectate lyase gene VIT_01s0026g01680 was up-regulated from stage E-L 15 to stage E-L 27 in both clones but the fold change was six times higher in the wild-type compared to the mutant. Based on its functional annotation, the H1flk-like gene VIT_05s0020g04850 plays a role in chromatin assembly. Its Arabidopsis homolog encodes a P-loop containing nucleoside triphosphate hydrolases superfamily protein that functions in ATP binding activity involved in cell killing (Wing et al 1990, TAIR). In the mutant background, this gene was specifically up-regulated from stage E-L 27 to stage E-L 38 while in the wild-type a significant differential expression with a lower fold change was 15 observed only between stages E-L 15 and E-L 38. The beta-expansin gene VIT_15s0021g02700 was not expressed at stage E-L 15 in the mutant. Differential expression analysis in the mutant showed specific up-regulation from stage E-L 15 to stage E-L 27, in contrast to a stable expression in the wild-type between the same stages. Based on its functional annotation, this gene encodes a protein involved in auxin-mediated signalling, which implies a late induction of auxin responsive genes in the mutant.

As expected, 466 out of the 501 common DE genes shared the same group or expression profile in both the wild-type and the mutant. Functional annotation and GO term enrichment uncovered many biological processes, which included cell wall metabolism, cell cycling, primary and secondary metabolism, signalling and regulation of gene expression, water transport and abiotic stress responses. Within this set the following four genes are of interest. VIT_15s0048g00510 encodes a protein that belongs to the pectinesterase family, up-regulated from stage E-L 15 to stage E-L 27 with double fold change in the wild-type compared to the mutant. Functional annotation revealed the protein involvement in cell wall

modification through pectin degradation. In Arabidopsis, it has been shown that cell type-specific pectin degradation is required to separate microspores during pollen development (Rhee et al 1998). VIT_15s0021g02170, VIT_18s0089g00140 and VIT_19s0015g00960 showed a similar behavior: they were down-regulated from stage E-L 15 to stage E-L 27 in both clones, but much more expressed in the wild-type than in the mutant.
VIT_15s0021g02170 encodes chalcone and stilbene synthase. Its Arabidopsis homolog is involved in phenylpropanoid biosynthetic process and pollen exine formation (TAIR).
VIT_18s0089g00140 encodes 1,4-beta-mannan endohydrolase, which is implicated in fructose and mannose metabolic pathways (Grimplet et al 2012). Description of biological processes associated to its Arabidopsis homolog revealed a role in seed germination (TAIR).
The Arabidopsis homolog of VIT_19s0015g00960 is required for male fertility and pollen exine formation as it encodes an ATP-binding cassette transporter involved in tapetal cell and pollen development (TAIR). Finally, within this category we identified two genes already proposed to affect seed and/or berry development (Doligez et al 2013). They code for the PISTILLATA (PI) floral homeotic protein (VIT_18s0001g01760) and the MADS-box AP3 transcription factor (VIT_18s0001g13460). The latter co-localizes with the stable QTL for berry weight, seed number and fresh weight identified by (Doligez et al 2013).

6.4.4. Differentially expressed genes specific to the wild-type background

The 441 genes specifically modulated among the wild-type developmental stages represented 12 groups and included a range of functional categories. A large number (approximately 64%) of these genes were observed among nine groups, down-regulated from stage E-L 15 to stage E-L 27 and not differentially expressed from stage E-L 27 to stage E-L 38 or vice versa. The remaining 36% were observed in three groups, and were up-regulated in the same manner (Nwafor et al 2014).

6.4.4.1. Down-regulated genes specific to wild-type (from stage E-L15 to stage E-L 27)

Within this category we observed several interesting genes that showed significant enrichment of GO terms and very high negative fold change. They include five genes, three of which encode similar proteins: male sterility 1(MS1, VIT_01s0011g06390), male sterility 2 (MS2, VIT_08s0007g07100) and tetratricopeptide repeat domain male sterility MS5 (VIT_07s0005g05680, VIT_07s0005g05720 and VIT_15s0107g00550). The gene coding for MS1 protein belongs to the PHD family of transcription factors. The Arabidopsis MS1 gene

was described to be a sporophytic factor controlling anther and pollen development. It plays a critical role in the induction of pollen wall and pollen coat materials in the tapetum and, ultimately, the production of viable pollen. Indeed, mutants show a semi-sterile phenotype, as their pollen degenerates after microspore release. In addition their tapetum appears abnormally vacuolated (TAIR, Sanders et al 1999, Ito 2007, Yang et al 2007). The MS2 gene has an unclear function in Vitis vinifera, however its Arabidopsis best match was described as a fatty acid reductase gene, involved in oxidation-reduction process and pollen exine formation (Chen et al 2011). The function of the MS5 gene in Vitis vinifera is unknown, however in Arabidopsis it was suggested to be similar to POLLENNESS3 gene (Uniport). Mutants of this gene in Arabidopsis were shown to have defects in functional microspore production that lead to the degeneration of cells within the anther locules (Sanders et al 1999). One of the three MS5 gene predictions co-located with a minor QTL for mean seed fresh weight on chromosome 15 (Costantini et al 2008). The significant down-regulation of these genes from stage E-L 15 to stage E-L 27 in the wild-type implies that they were highly induced at stage E-L 15, where they exhibited maximum expression levels, perhaps to ensure viable and functional pollen development for complete fertilization. On the other hand, in the mutant, these genes were not differentially expressed. Further analysis of their RPM values in the mutant revealed very low level of expression at stage E-L 15, when compared to the wild-type. This observation might suggest abnormal pollen development in the mutant resulting in non-functional or partially sterile pollen. However, it needs to be tested and confirmed experimentally.
Within this category we found two additional genes with a putative role in ovule and pollen differentiation: SPOROCYTELESS (VIT_19s0014g03940) and glycerol-3-phosphate acyltransferase 1 (VIT_12s0142g00040). The SPOROCYTELESS gene of Arabidopsis was described to encode a transcription factor that is required for the initiation of both micro- and megagametogenesis and is expressed in the sporogenous tissue of the anther and the ovule. It is involved in establishing the prospective chalaza of the ovule, plays a central role in patterning both the proximal-distal and the adaxial-abaxial axes in the ovule and regulates the anther cell differentiation. Mutant is defective in the differentiation of primary sporogenous cells into microsporocytes, and does not properly form the anther wall (TAIR, Yang et al 1999, Liu et al 2009). The Arabidopsis homolog of glycerol-3-phosphate acyltransferase 1 gene was shown to be expressed in flower buds and siliques. Its protein is involved in metabolic

processes such as phosphatidylglycerol biosynthetic process, pollen sperm cell differentiation, and response to karrikin. Interestingly, the homozygous mutant plants are male sterile (TAIR, Li et al 2012).

6.4.4.2. Down-regulated genes specific to wild-type (from stage E-L 27 to stage E-L 38)

Within this category we observed about 30 genes with high negative fold change, the majority of which belong to the functional categories of cellular process and signalling. The most relevant for seed development appeared the genes encoding a calmodulin-binding protein (VIT_00s1404g00010), the wall-associated kinase 4 (WAK4, VIT_01s0026g01420), the aquaporin GAMMA-TIP3/TIP1;3 (VIT_06s0061g00730) and a precursor of cytokinin dehydrogenase (VIT_18s0001g13200). Indeed, in rice a calmodulin-binding protein was found to be essential to pollen development (Zhang et al 2012), the silencing of a member of the WAK family led to sterility due to anther indehiscence (Kanneganti and Gupta 2008), while the aquaporin GAMMA-TIP3/TIP1;3 in Arabidopsis was reported to be a pollen-specific water transporter contributing to male sterility in the double knockout mutant tip1;3/tip5;1(Wudick et al 2014), and cytokinins were demonstrated to regulate seed yield (Bartrina et al 2014).

6.4.4.3. Up-regulated genes specific to the wild-type (from stage E-L 15 to stage E-L 27)

Amongst this group we noticed a number of genes with high positive fold change value. Besides genes encoding proteins involved in cell wall organization and biogenesis, the most relevant for seed development were found in the categories: metabolism, transport, regulation overview and signalling. For instance, we identified a chitinase class IV gene (VIT_05s0094g00330), whose best Arabidopsis match was described to be expressed during somatic embryogenesis in nursing cells surrounding the embryos and additionally in mature pollen and growing pollen tubes until they enter the receptive synergid (TAIR, Passarinho et al 2001). Among transporters, a cation/hydrogen exchanger (VIT_10s0003g03030) showed its best match with an Arabidopsis protein involved in pollen tube growth (TAIR). Of particular interest were a set of genes encoding transcription factors and signalling molecules. Among the transcription factors were TRANSPARENT TESTA 16 (TT16 or AGL32, VIT_01s0011g01560), BZIP family protein (VIT_18s0001g03010) and the MADS-box protein SEEDSTICK (VIT_18s0041g01880). The TT16 gene encodes a MADS-box family

transcription factor (Grimplet et al 2012, TAIR). In Arabidopsis it was reported to determine the identity of the endothelial layer within the ovule, to play a maternal role in fertilization and seed development and to regulate proanthocyanidin biosynthesis and cell shape of the inner-most cell layer of the seed coat (TAIR, Nesi et al 2002). In canola (Brassica napus) it was further demonstrated that the tt16 deficiency affects pollen tube guidance, resulting in reduced fertility and negatively impacting embryo and seed development due to the altered expression of genes involved in gynoecium and embryo development, lipid metabolism, auxin transport, and signal transduction (Deng et al 2012). In addition, the TT16 gene was reported among the functional candidates potentially involved in seed and/or berry development that did not co-localize with QTLs detected for the same traits (Doligez et al 2013). The BZIP gene was previously described by (Lui et al 2014) to be expressed in pollen and other flower parts. Although the MADS-box protein SEEDSTICK gene did not show high positive fold change, it was significantly enriched in our data. In Arabidopsis and rice, this gene was described to encode a MADS-box transcription factor expressed in the carpel and ovules and to play a maternal role in fertilization and seed development. Mutants indeed exhibited reduced ovule fertilization and high seed abortion (TAIR, Favaro et al 2003, Mizzotti et al 2012, Dreni et al 2011). Interestingly, this gene was among those that co-localized with the stable QTLs for seed-related traits (Costantini et al 2008, Doligez et al 2013). The signalling molecules included FERONIA receptor-like kinase (VIT_03s0038g04340). In Arabidopsis , it was shown to mediate male-female interactions during pollen tube reception (Escobar-Restrepo et al 2007). Feronia mutant had impaired fertilization because pollen tube failed to arrest by continue growth inside the female gametophyte (Zou et al 2011). This study concluded that female control of pollen tube reception is based on a FERONIA-dependent signalling pathway. In our investigation, we observed low expression level (0.6 RPM) of FERONIA receptor-like kinase gene in the mutant, compared to higher expression (6.9 RPM) in the wild-type. Finally, within this category we identified a gene coding for a nodulin (VIT_17s0000g08110), which was up-regulated from stage E-L 15 to stage E-L 27 and down-regulated from stage E-L 27 to stage E-L 38. The Arabidopsis best match for this gene encodes a protein containing three domains, one of which is MtN3/saliva-related trans-membrane protein, and has function in sugar trans-membrane transporter activity (TAIR). In rice the genes Xa13/Os8N3/OsSWEET11 and Os11N3/OsSWEET14 encode proteins with two MtN3/saliva domains similar to that of Arabidopsis, and were identified to play important

role in regulating reproductive development through promotion of fertilization. These genes were reported to have a very high expression level in rice panicles and anthers compared to other tissues. Suppressed plants showed reduced fertility or were sterile due to blockage of microspore development at the unicellular pollen grain stage. This resulted in the gradual degeneration of the immature pollen suggesting the proteins are required for pollen development in rice. In addition knockout mutants showed reduced seed size and delayed growth (Yuan and Wang 2013). The significant up-regulation of the nodulin MtN3 gene from stage E-L 15 to stage E-L 27 in the wild-type compared to the mutant could imply an active role in promoting fertilization. In contrast, down-regulation of this gene from stage E-L 27 to stage E-L 38, which corresponds to a period of seed maturation (after fertilization), seems to support the notion that genes participating or promoting seed formation are tightly regulated.

6.4.4.4. Up-regulated genes specific to the wild-type (from stage E-L 27 to stage E-L 38)

Within this category we found a gene coding for oleosin OLE-2 protein (VIT_17s0000g09000), with a putative role in oil body organization and biogenesis as well as in reproduction and seed development.

Functional studies in Arabidopsis showed that the double mutant ole1/ole2 had irregular enlarged oil-containing structures throughout the seed cells which led to defects in germination or seed mortality (Shimada et al 2008). Three different genes encoded enzymes involved in primary metabolism, namely globulin-1 S allele precursor (GLB1, VIT_07s0151g00640), 7S globulin precursor (VIT_14s0128g00200) and Em protein GEA6 (EM6, VIT_13s0067g01250). Functional annotation revealed that the three genes participate in generation of metabolite precursors and serve as energy storage proteins. The maize GLB1 gene was found to be expressed throughout embryo development specifically in seed tissues (Belanger et al 1989). Similarly, 7S globulin precursor was described as a major storage protein in legume species (Kagawa et al 1987). In our study, the expression of the 7S globulin precursor gene was highest at wild-type stage E-L 38 while it was almost abolished in the mutant. This suggests that induction of these genes may be required to complete seed development. The best Arabidopsis match for the EM6 gene was described to be the Late Embryogenesis Abundant 6 gene, involved in response to abscisic acid, required for normal seed development, and regulating the timing of desiccation tolerance and the rate of water loss during seed maturation (TAIR, Gaubier et al 1993). Other interesting genes are those

involved in lipid and water transport, e.g . the genes coding for a protease inhibitor/seed storage/lipid transfer protein (VIT_14s0108g00520) and aquaporin BETA-TIP 20 (VIT_16s0039g00220).

Equally worth mentioning are two genes coding for signalling molecules, namely the abscisic acid-insensitive protein 3 ABI3 (VIT_07s0005g05400) and the serine/threonine-protein kinase receptor ARK3 (VIT_19s0014g04130). The expression of ABI3 gene was completely abolished in the mutant from stage E-L 27 to stage E-L 38. ABI3 is a putative seed-specific transcriptional activator acting as a central regulator in ABA signalling. In different species it was described to play a major role in seed maturation and to regulate the transition between embryo maturation and early seedling development (TAIR, Zeng et al 2013, Delmas et al 2013). In Arabidopsis the ARK3 gene was proposed to participate in recognition of pollen (TAIR, Pastuglia et al 2002). Four stress response genes were also present and specifically induced, including those coding for the seed maturation proteins PM31 (VIT_18s0001g01570) and PM34 (VIT_14s0128g00340). Finally, the gene prediction for the heat shock protein 17.6 kDa class II with a putative role in protein folding (VIT_04s0008g01610) was not expressed in the mutant in all three developmental stages.

6.4.5. Differentially expressed genes specific to the mutant background

The 133 DE genes, which were peculiar to the mutant, fell within 4 groups (3, 8, 11 and 17) and were all stage specifically induced. The majority of these genes (63%) were either down-regulated from stage E-L 15 to stage E-L 27 or from stage E-L 27 to stage E-L 38, whereas 37% of them were up-regulated in the same manner (Nwafor et al 2014). The genes related to seed development showed differential expression between stages E-L 15 and E-L 27.

6.4.5.1. Down-regulated genes specific to the mutant (from stage E-L 15 to stage E-L 27)

In this category we identified genes with high negative fold change encoding proteins with a role in cellular processes, transport and regulation of gene expression. Among the genes involved in cellular processes we selected Terminal EAR1-like 1 (TEL1, VIT_14s0219g00270) and Fimbrin 2 (VIT_12s0059g00560). The TEL1 gene encodes an RNA binding protein with a function in shoot development, conserved among land and vascular plants (TAIR, Vivancos et al 2012). The Arabidopsis best match of TEL1 is a member of the mei2-like gene family, which plays a role in meiosis. Specific multiple mutant combinations were reported to display sterility and a range of defects in meiotic chromosome behavior (Kaur et al 2006). The

Fimbrin 2 gene is involved in actin organization and biogenesis; its Arabidopsis homolog is FIMBRIN5, an actin bundling factor required for pollen germination and pollen tube growth (Wu et al 2010). The same function was reported in lily (Su et al 2012). We observed high expression of the TEL1 and Fimbrin 2 genes at stage E-L 15 in both clones, however as development progressed towards stage E-L 27 a significant repression of both genes in the mutant was evident in their very low RPM values as compared to a stable expression of these genes in the wild-type. In addition the Fimbrin 2 gene in grape fell within a stable QTL for mean seed fresh weight reported by (Doligez et al 2013). In the transport category we identified a gene encoding the boron transporter-like protein 4 (VIT_04s0008g04980). Previously, boron deficiency has been associated with the occurrence of parthenocarpic seedless grapes in some varieties of Vitis vinifera L (Pérez-Castro et al 2012). We also noticed a set of genes coding for transcription factors, which included the MYB domain protein 26 (MYB26, VIT_09s0002g01670), AP2 AINTEGUMENTA (VIT_09s0002g01370) and lateral organ boundaries protein 41 (LBD41, VIT_14s0006g02950). The Arabidopsis MYB26 protein was described to be involved in anther dehiscence, response to gibberellin stimulus and secondary cell wall biogenesis. Mutants for this gene produced fertile pollen but plants were sterile because anthers did not dehisce. When compared to wild type, no cellulosic secondary wall thickening was seen in the anther endothecium of the mutant (Yang et al 2007). The AP2 AINTEGUMENTA gene belongs to the AP2 (APETALA2)/EREBP (ethylene-responsive element binding protein) family of transcription factors, known to be key regulators of several developmental processes (Riechmann 1998). The Arabidopsis homolog was reported to have a role in ovule development among other functions. Mutants exhibited female-sterility as integuments did not develop and megasporogenesis was blocked at the tetrad stage (Elliott et al 1996). The LBD41 gene encodes a protein containing the conserved domain AS2/LOB. The Arabidopsis homolog of the LOB gene ASYMMETRIC LEAVES2 (AS2) was demonstrated to function in the repression of KNOX genes and in the specification of adaxial/abaxial organ polarity (Lin et al 2003). The maize ortholog was also reported to be required to prevent KNOX gene expression in lateral organs and, in addition, to promote the switch from proliferation to differentiation in the embryo sac. The failure to limit proliferation in mutant embryo sacs was shown to lead to a variety of structural defects, including the production of extra gametes and synergids. Moreover, the fertilization process was frequently abnormal, producing seeds with haploid embryos and embryos and endosperms derived from

fertilization by different pollen tubes (Evans 2007). Although the role of these regulatory genes in growth and development is well documented in model species, in Vitis vinifera L. their specific functions are not well characterized and can only be inferred. However, we observed a general pattern in the mutant, in which expression of these genes was almost abolished at stage E-L 27 when compared to their stable expression in the wild-type.
Finally, a gene DTA2 was observed (VIT_15s0046g03080, downstream target of AGL15). In Arabidopsis DTA2 was reported to encode an unknown protein with no significant similarity to any know n protein and to be expressed in developing seeds and in roots (Wang et al 2002). In our data, the DTA2 gene from the mutant was expressed at stage E-L 15, and the expression was abolished at stages E-L 27 and E-L 38 (in contrast to the stable expression in the wild-type).

6.4.5.2. Up-regulated genes specific to the mutant (from stage E-L 15 to stage E-L 27)

Within this category we selected six genes, one of which (VIT_12s0134g00240) encodes a signaLling molecule involved in stress response. This Avr9/Cf-9 rapidly elicited protein 20 was shown to function in the initial development of the defence response in tomato (Rowland et al 2005). The remaining five genes encode proteins involved in the ethylene-mediated signalling pathway. These are ethylene -responsive transcription factor 9 (ERF9, VIT_12s0028g03270) and ethylene-responsive transcription factor ERF105 (VIT_16s0013g00950, VIT_16s0013g00990, VIT_16s0013g01050 and VIT_16s0013g01120). The ERF9 gene was shown to take part in repressing the activation of pathogen related genes in Arabidopsis (Camehl et al 2010). The Arabidopsis homolog of ERF105 encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family that is involved in processes such as regulation of transcription, respiratory burst involved in defence responses, as well as responses to mechanical stimulus and wounding (TAIR, Camehl et al 2010, Libault et al 2007). We noticed that the expression levels of these genes were always higher at stage E-L 27 in the mutant compared to the wild-type.
It might be worthy of mention that a substantial proportion of our strongest candidate genes (that are the genes expressed specifically in either clone) were physically clustered in the vicinity of some previously identified QTLs (Costantini et al 2008, and Doligez et al 2013) mainly the loci on chromosomes 2 and 12 (Appendix 7-Table 4). While there may be no causal link between their expression and trait variation, they might provide a valuable starting point for developing DNA markers linked to the target trait, as discussed in (Jensen et al

2014).

7. List of reference

Adam-Blondon, A. F., Lahogue, F., Bouquet, A., Boursiquot, J. M., This, P., 2001. Usefulness of two SCAR markers for marker-assisted selection of seedless grapevine cultivars. Vitis - Journal of Grapevine Research, 40,147-155.

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data: Genome Biology 11, R106.

Barritt, B. H., 1969. Fruit set and ovule development in seedless grapes treated with growth retardants Alar and CCC. Ph.D. Thesis. Cornell Univ., Ithaca, NY.

Barritt, B. H., 1970. Ovule development in seeded and seedless grapes. Vitis - Journal of Grapevine Research 9, 7-14.

Bartrina, I., Otto, E., Strnad, M., Werner, T., Schmülling, T., 2011. Cytokinin regulates the activity of reproductive meristems, flower organ size, ovule formation, and thus seed yield in Arabidopsis thaliana. Plant Cell 23, 69-80.

Belanger, F. C., Kriz, A. L., 1989. Molecular characterization of the major maize embryo globulin encoded by the Glb1 gene. Plant Physiology 91, 636-643.

Bergamini, C., Cardone, M. F., Anaclerio, A., Perniola, R., Pichierri, A., Genghi, R., Alba, V., Forleo, L. R., Caputo, A. R., Montemurro, C., Blanco, A., Antonacci, D., 2013. Validation assay of p3_VvAGL11 marker in a wide range of genetic background for early selection of stenospermocarpy in *Vitis vinifera* L. Molecular Biotechnology 54, 1021-1030.

Berger, F., 1999. Endosperm development. Current Opinion in Plant Biology 2, 28-32.

Huh, J. H., Bauer, M. J., Hsieh, T. F., Fischer, R. L., 2008. Cellular programming of plant gene imprinting. Cell 132, 735-744.

Braybrook, S. A., Harada, J. J., 2008. LECs go crazy in embryo development. Trends Plant Science 139, 624-630.

Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., Laux, T., 2008. Differential expression of WOX genes mediates apical-basal axis formation in the Arabidopsis embryo. Dev Cell 14, 867-876.

Bica, D., 2007. Vitigni di Sicilia. pp:32

Boccacci, P., Torello Marinoni, D., Gambino, G., Botta, R., Schneider, A., 2005. Genetic characterization of endangered grape cultivars of Reggio emilia province. American journal of enology and viticulture 56, 411-416.

Boss, P. K., Thomas, M. R., 2002. Association of dwarfism and floral induction with a grape 'green revolution' mutation. Nature 416, 847-850.

Bouquet, A., Danglot, Y., 1996. Inheritance of seedlessness in grapevine *Vitis vinifera* L. Vitis - Journal of Grapevine Research 35, 35-42.

Cabezas, J. A., Cervera, M. T., Ruiz-Garcia, L., Carreño, J., Martinez-Zapater, J. M., 2006. A genetic analysis of seed and berry weight in grapevine. Genome 49, 1572-1585.

Camehl, I., Oelmüller, R., 2010. Do ethylene response factors-9 and -14 repress PR gene expression in the interaction between Piriformospora indica and Arabidopsis? Plant Signal Behaviour 5, 932-936.

Caniato, E., 2011. Development and Application of New Strategies for Genome Scaffolding and Gene Prediction applied to NGS data. Ph.D. Thesis. University of Padua, Italy.

Carreno, J., Oncina, R., Carreno, I., 2006. In vitro studies on pollen germination capability and preservation of different cultivars of Vitis vinifera L. ISHS Acta Horticulture 827: IX International Conference on Grape Genetics and Breeding.

Charlotte, P., 1971. Reproductive anatomy in cultivated grapes - a review. American Journal of Enology and Viticulture 22, 92-109.

Chen, W., Yu, X. H., Zhang, K., Shi, J., De Oliveira, S., Schreiber, L., Shanklin, J., Zhang, D., 2011. Male Sterile2 encodes a plastid-localized fatty acyl carrier protein reductase required for pollen exine development in Arabidopsis. Plant Physiology 157, 842-853.

Chkhartishvili, N., Vashakidze, L., Gurasashvili, V., And Maghradze, D., 2006. Type of pollination and indices of fruit set of some Georgian grapevine varieties. Vitis - Journal of Grapevine Research 45, 153–156.

Cipriani, G., Marrazzo, M. T., Di Gaspero, G., Pfeiffer, A., Morgante, M., Testolin, R., 2008. A set of microsatellite markers with long core repeat optimized for grape (Vitis spp.) genotyping. BMC Plant Biology 8, 127.

Cipriani, G., Spadotto, A., Jurman, I., Di Gaspero, G., Crespan, M., Meneghetti, S., Frare, E., Vignani, R., Cresti, M., Morgante, M., Pezzotti, M., Pe, E., Policriti, A., Testolin, R.,

2010. The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. Theoretical and Applied Genetics 121,1569-1585.
Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., Ruden, D. M., 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.
Coombe, B. G., 1973. The regulation and development of the grape berry. Acta Horticulture 34, 261-272.
Coombe, B. G., 1976. The development of fleshy fruits. Annual Review of Plant Biology 27, 507-28.
Coombe, B., Dry, PR., 1992. Viticulture Volume 2 Practices, Adelaide. Wine titles 376 pp.
Coombe, B. G., 1995. Growth stages of the grapevine: adoption of a system for identifying grapevine growth stages. Australian Journal of Grape and Wine Research 1, 100-110.
Costantini, L., Battilana, J., Lamaj, F., Fanizza, G., Grando, M. S., 2008. Berry and phenology-related traits in grapevine (*Vitis vinifera* L.): From quantitative trait loci to underlying genes. BMC Plant Biology 8, 38.
Costenaro-da-Silva, D., Passaiab, G., Henriquesa, J. A. P., Margisa, R., Pasqualia, G., Reversb, L. F., 2010. Identification and expression analysis of genes associated with the early berry development in the seedless grapevine (*Vitis vinifera* L.) cultivar Sultanine. Plant Science 179, 510–519.
Dean, G., Cao, Y., Xiang, D., Provart, N. J., Ramsay, L., Ahad, A., White, R., Selvaraj, G., Datla, R., Haughn, G., 2011. Analysis of gene expression patterns during seed coat development in Arabidopsis. Molecular Plant 4, 1074-1091.
Delmas, F., Sankaranarayanan, S., Deb, S., Widdup, E., Bournonville, C., Bollier, N., Northey, J. G., McCourt, P., Samuel, M. A., 2013. ABI3 controls embryo degreening through Mendel's I locus. Proceedings of the National Academy of Sciences 110, E3888-3894.
Deng, W., Chen, G., Peng, F., Truksa, M., Snyder, C. L., Weselake, R. J., 2012. Transparent Testa16 plays multiple roles in plant development and is involved in lipid synthesis and embryo development in canola. plant physiology 160, 978-989.
Devic, M., 2008. The importance of being essential: EMBRYO-DEFECTIVE genes in

Arabidopsis. Comptes Rendus Biologies 331, 726-736.
Dokoozlian, N. K., 2000. Grape Berry Growth and Development. in: Raisin Production Manual. University of California, Agricultural and Natural Resources Publication, Oakland, CA. 3393, 30-37.
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. Theoretical Applied Genetics 105, (5):780-795.
Doligez, A., Bertrand, Y., Farnos, M., Grolier, M., Romieu, C., Esnault, F., Dias, S., Berger, G., François, P., Pons, T., Ortigosa, P., Roux, C., Houel, C., Laucou, V., Bacilieri, R., Péros, J. P., This, P., 2013. New stable QTLs for berry weight do not colocalize with QTLs for seed traits in cultivated grapevine (*Vitis vinifera* L.). BMC Plant Biology 13, 217.
Du, Z., Zhou, X., Ling, Y., Zhang, Z., Su, Z., 2010. AgriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38, (Web Server issue):W64-W70.
Dreni, L., Pilatone, A., Yun, D., Erreni, S., Pajoro, A., Caporali, E., Zhang, D., Kater, M. M., 2011. Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. Plant Cell 23, 2850-2863.
Elliott, R. C., Betzner, A. S., Huttner, E., Oakes, M. P., Tucker, W. Q., Gerentes, D., Perez, P., Smyth, D. R., 1996. AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. Plant Cell 8, 155-168.
Emanuelli, F., Lorenzi, S., Grzeskowiak, L., Catalano, V., Stefanini, M., Troggio, M., Myles, S., Martinez-Zapater, J. M., Zyprian, E., Moreira, F. M., Grando, M. S., 2013. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. BMC Plant Biology 13, 39.
Escobar-Restrepo, J. M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W. C., Grossniklaus, U., 2007. The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. Science 317, 656-660.
Evans, M. M. S., 2007. The indeterminate gametophyte1 gene of maize encodes a LOB domain protein required for embryo sac and leaf development. Plant Cell 19, 46-62.
Fanizza, G., Lamaj, F., Costantini, L., Chaabane, R., Grando, M. S., 2005. QTL analysis for fruit yield components in table grapes (Vitis vinifera). Theoretical and Applied Genetics 111, 658-664.

Fasoli, M., Dal-Santo, S., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., Porceddu, A., Venturini, L., Bicego, M., Murino, V., Ferrarini, A., Delledonne, M., Pezzotti, M., 2012. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. Plant Cell 24, 3489-3505.

Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., Yanofsky, M. F., Kater, M. M., Colombo, L., 2003. MADS-box protein complexes control carpel and ovule development in Arabidopsis. Plant Cell 15, 2603-2611.

Food and agricultural organization statistics 2012. http://faostat3.fao.org/browse/Q/QC/E

Gasparro, M., Caputo, A. R., Bergamini, C., Crupi, P., Cardone, M. F., Perniola, R., Antonacci, D., 2013. Sangiovese and its offspring in southern Italy. Molecular Biotechnology 54, 581-9.

Gaubier, P., Raynal, M., Hull, G., Huestis, G. M., Grellet, F., Arenas, C., Pages, M., Delseny, M., 1993. Two different Em-like genes are expressed in Arabidopsis thaliana seeds during maturation. Molecular Genetics and Genomics 238, 409-418.

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

Hanania, U., Velcheva, M., Or, E., Flaishman, M., Sahar, N., Perl, A., 2007. Silencing of chaperonin 21, that was differentially expressed in inflorescence of seedless and seeded grapes, promoted seed abortion in tobacco and tomato fruits. Transgenic Res16, 515-525.

Hanania, U., Velcheva, M., Sahar, N., Flaishman, M., Or, E., Degani, O., Perl, A., 2009. The ubiquitin extension protein S27a is differentially expressed in developing flower organs of Thompson seedless versus Thompson seeded grape isogenic clones. Plant Cell Report 28, 1033-1042.

Ito, T., Nagata, N., Yoshiba, Y., Ohme-Takagi, M., Ma, H., Shinozaki, K., 2007. Arabidopsis MALE STERILITY1 encodes a PHD-type transcription factor and regulates pollen and tapetum development. Plant Cell 19, 3549-3562.

Robinson, J., Harding, J., Vouillamoz, J., 2012. A complete guide to 1,368 vine varieties, including their origins and flavours. In Wine and grapes, London Penguin Group Publishers.

Jenik, P. D., Gillmor, C. S., Lukowitz, W., 2007. Embryonic patterning in Arabidopsis

thaliana. Annual Review of Cell and Developmental Biology 23, 207-236.

Jensen, P. J., Fazio, G., Altman, N., Praul, C., McNellis, T. W., 2014. Mapping in an apple (Malus x domestica) F1 segregating population based on physical clustering of differentially expressed genes. BMC Genomics 15, 261.

Kagawa, H., Yamauchi, F., Hirano, H., 1987. Soybean basic 7S globulin represents a protein widely distributed in legume species. FEBS Letters 226, 145-149.

Kaur, J., Sebastian, J., Siddiqi, I., 2006. The Arabidopsis-mei2-lLike genes play a role in meiosis and vegetative growth in Arabidopsis. Plant Cell18, 545-559.

Koltunow, A., Bicknell, R . A., Chaudhury, A. M., 1995. Apomixis: Molecular strategies for the generation of genetically identical seeds without fertilization. Plant Physiology 108, 1345-1352.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinformatics 5, 150-63.

Lahogue, F., This, P., Bouquet, A., 1998. Identification of a co-dominant scar marker linked to the seedlessness character in grapevine. Theoretical Applied Genetics 97, 950-959.

Le, B. H, Cheng, C., Bui, A. Q., Wagmaister, J. A., Henry, K. F., Pelletier, J., Kwong, L., Belmonte, M., Kirkbride, R., Horvath, S., Drews, G. N., Fischer, R. L., Okamuro, J. K., Harada, J. J., Goldberg, R. B., 2010. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. PNAS, Proceedings of the National Academy of Sciences 107, 8063-8070.

Ledbetter, C. A., Ramming, D. W., 1989. Seedlessness in grapes. Horticultural Reviews 11, 159- 184.

Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler Transform. Bioinformatics 26, 589-595.

Li, XC., Zhu, J., Yang, J., Zhang, G. R., Xing, W. F., Zhang, S., Yang, Z. N., 2012. Glycerol-3-phosphate acyltransferase 6 (GPAT6) is important for tapetum development in Arabidopsis and plays multiple roles in plant fertility. Molecular Plant 5, 131-142.

Libault, M., Wan, J., Czechowski, T., Udvardi, M., Stacey, G., 2007. Identification of 118 Arabidopsis transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor. Molecular Plant-Microbe Interactions 20, 900-911.

Lin, W. C., Shuai, B., Springer, P. S., 2003. The Arabidopsis LATERAL ORGAN

BOUNDARIES-domain gene ASYMMETRIC LEAVES2 functions in the repression of KNOX gene expression and in adaxial-abaxial patterning. Plant Cell 15, 2241-2252.
Liu, X., Huang, J., Parameswaran, S., Ito T, Seubert B, Auer M, Rymaszewski A, Jia G, Owen HA, Zhao D. 2009. The SPOROCYTELESS/NOZZLE gene is involved in controlling stamen identity in Arabidopsis. Plant Physiology 151, 1401-1411.
Liu J., Chen N., Chen, F., Cai, B., Dal., Santo, S., Tornielli, G. B., Pezzotti, M., Cheng, Z. M., 2014. Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (Vitis vinifera). BMC Genomics 15, 281.
Maul, E., Sudharma, K.N., Kecke, S., Marx, G., Müller, C., Audeguin, L., Boselli, M., Boursiquot, J.M., Bucchetti, B., Cabello, F., Carraro, F., Crespan, M., De Andrés, M.T., Dias, J.E., Ekhvaia, J., Gaforio, L., Gardiman, M., Grando, M.S., Gyropoulos, D., Jandurova, O., Kiss, E.; Kontic, J., Kozma, P., Lacombe, T., Laucou, V., Legrand, D., Maghradze, D., Marinoni, D., Maletic, E., Moreira Maia, F., Muñoz, G., Nakhutsrishvili, G., Pijic, I., Peterlunger, E., Pitsoli, D., Pospisilova, D., Preiner, D., Raimondi, S., Regner, F., Savin, G., Savvides, S., Schneider, A., Sereno, C., Simon, S., Staraz, M., Zulini, L., Bacilieri, R., This, P. (2012). The European Vitis Database (www.eu-vitis.de): a technical innovation through an online uploading and interactive modification system. Vitis, 51 (2): 79-85.
Mejía, N., Gebauer, M., Muñoz, L., Hewstone, N., Muñoz, C., Hinrichsen, P., 2007. Identification of QTLs for seedlessness, berry size, and ripening date in a seedless × seedless progeny. American Journal of Enology and Viticulture 58, 499-507.
Mejía, N., Soto, B., Guerrero, M., Casanueva, X., Houel, C., Miccono, M. A., Ramos, R., Le Cunff, L., Boursiquot, J. M., Hinrichsen, P., Adam-Blondon, A. F., 2011. Molecular, genetic and transcriptional evidence for a role of VvAGL11 in stenospermocarpic seedlessness in grapevine. BMC Plant Biology 11, 57.
Miculan, M., 2014. Detection of somatic variants from next-generation sequencing data in grapevine bud sports. Ph.D. Thesis. University of Udine.
Mizzotti, C., Mendes, M. A., Caporali, E., Schnittger, A., Kater, M. M., Battaglia, R., Colombo, L., 2012. The MADS box genes SEEDSTICK and ARABIDOPSIS Bsister play a maternal role in fertilization and seed development. The Plant Journal 70, 409-420.
Mullins, M. G., Bouquet, A., Williams, L. E., 1992. Biology of the grapevine. In Biology of horticultural crops. Edited by Mullins MG, Cambridge: Cambridge University Press.

Myles, S., Chia, J. M., Hurwitz, B., Simon, C., Zhong, G. Y., Buckler, E., Ware, D., 2010. Rapid genomic characterization of the genus Vitis. PLoS One 5, e8219.
Nwafor, C. C., Gribaudo, I., Schneider, A., Wehrens, R., Grando, M. S., Costantini, L., 2014. Transcriptome analysis during berry development provides insights into co-regulated and altered gene expression between a seeded wine grape variety and its seedless somatic variant. BMC Genomics 15, 1030.
Nesi, N., Debeaujon, I., Jond, C., Stewart, A. J., Jenkins, G. I,, Caboche, M., Lepiniec, L., 2002. The TRANSPARENT TESTA16 locus encodes the ARABIDOPSIS BSISTER MADS domain protein and is required for proper development and pigmentation of the seed coat. Plant Cell 14, 2463-2479.
Passarinho, P.A., Van Hengel, A. J., Fransz, P. F., de Vries, S. C., 2001. Expression pattern of the Arabidopsis thaliana AtEP3/AtchitIV endochitinase gene. Planta 212, 556-567.
Pastuglia, M., Swarup, R., Rocher, A., Saindrenan, P., Roby, D., Dumas C, Cock J. M., 2002. Comparison of the expression patterns of two small gene families of S gene family receptor kinase genes during the defence response in Brassica oleracea and Arabidopsis thaliana. Gene 282, 215-225.
Pearson, H. M., 1933. Parthenocarpy and seed abortion in *Vitis vinifera* L. Proceedings of the American Society for Horticultural Science 34, 402-404.
Pellerone, F.I., Edwards, K. J., Thomas, M. R., 2001. Grapevine microsatellite repeats: Isolation, characterisation and genotyping of grape germplasm from Southern Italy. Vitis - Journal of Grapevine Research 40, 179-186.
Pérez-Castro, R., Kasai, K., Gainza-Cortés, F., Ruiz-Lara, S., Casaretto, J. A., Peña-Cortés, H., Tapia, J., Fujiwara, T., González, E., 2012. VvBOR1, the grapevine ortholog of AtBOR1, encodes an efflux boron transporter that is differentially expressed throughout reproductive development of Vitis vinifera L. Plant Cell Physiology 53, 485-494.
Pfaffl, M. W., Horgan, G. W., Dempfle, L., 2002. Relative Expression Software Tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30, e36.
Poupin, M. J., Matus, J. T., Leiva-Ampuero, A., Arce-Johnson, P., 2011. Flower development in grapevine: A tale of two seasons. In book: The Flowering Process and its Control in Plants: Gene Expression and Hormone Interaction, First Edition: Research

Signpost, Editors: M. Yaish 173-198.
Raimondi, S., Torello Marinoni, D., Schneider, A., 2006. Caratterizzazione Ampelografica E Genetica Di Vitigni Minori Del Basso Piemonte Oggetto Di Valorizzazione: Nuove Proposte Per I Viticoltori. Italus Hortus 32,154-157.
Rhee, S.Y., Somerville, C. R., 1998. Tetrad pollen formation in quartet mutants of Arabidopsis thaliana is associated with persistence of pectic polysaccharides of the pollen mother cell wall. The Plant Journal 15, 79-88.
Riechmann, J. L., Meyerowitz, E. M., 1998. The AP2/EREBP family of plant transcription factors. The Journal of Biological Chemistry 379, 633-646.
Rowland, O., Ludwig, A. A., Merrick, C. J., Baillieul, F., Tracy, F. E., Durrant, W. E., Fritz-Laylin, L., Nekrasov, V., Sjölander, K., Yoshioka, H., Jones, J. D., 2005. Functional analysis of Avr9/Cf-9 rapidly elicited genes identifies a protein kinase, ACIK1, that is essential for full Cf-9-dependent disease resistance in tomato. Plant Cell 17, 295-310.
Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology 132, 365–386.
Ruan, Y. L., Patrick, J. W., Bouzayen, M., Osorio, S., Fernie, A. R., 2012. Molecular regulation of seed and fruit set. Trends in Plant Science 17, 1360-1385.
Sampson, B., Noffsinger, S., Gupton, C., Magee, J., 2001. Pollination Biology of the Muscadine Grape. Horticultural Science 36:120–124.
Sanders, P. M., Anhthu, Q. B., Weterings, K., McIntire, K. N., Hsu, Y., Lee, P. Y., Troung, M. T., Beals, T. P., Goldberg, R. B., 1999. Anther developmental defects in Arabidopsis thaliana male-sterile mutants. Sex Plant Reproduction 11, 297-322.
Schlotterer, C., 2004. The evolution of molecular markers—just a matter of fashion? Nature Reviews Genetics 5, 63–69.
Schneider, A., Carra, A., Akkak, A., This, P., 2001. Verifying synonymies between grape cultivars from France and Northwestern Italy using molecular markers. Vitis - Journal of Grapevine Research 40, 197-203.
Schneider, A., Raimondi, S., Moreira, F. M., De Santis, D., Zappia, R., Marinoni, D., Librandi, N., Grando, M. S., 2009. Contribution to the identification of the main varieties of Calabria. Frutticoltura 71, 46-55.
Seed development flow-chart in grape berries [http://michaelstriem.com/files/Seed_Development_Flow-Chart_in_Grape_Berries.pdf].

Shimada, T. L., Shimada, T., Takahashi, H., Fukao, Y., Hara-Nishimura, I., 2008. A novel role for oleosins in freezing tolerance of oilseeds in Arabidopsis thaliana. Plant Journal 55, 798-809.
Skinner, D. J., Hill, T. A., Gasser, C. S., 2004. Regulation of ovule development. Plant Cell 16, S32-S45.
Spillane, C, A., Steimer, A., Grossniklaus, U., 2001. Apomixis in agriculture: the quest for clonal seeds. Sex Plant Reproduction. 14: 179-187.
Srinivasan, C., Mullins, M. G., 1981. Physiology of flowering in the grapevine. A review. American Journal of Enology and Viticulture 32, 47-63.
Stacey, M. G., Patel, A., McClain, W. E., Mathieu, M., Remley, M., Rogers, E. E., Gassmann, W., Blevins, D.G., Stacey, G., 2008. The Arabidopsis AtOPT3 protein functions in metal homeostasis and movement of iron to developing seeds. Plant Physiology 146, 589-601.
Striem, M. J., Spiegel-Roy, P., Baron, I., Sahar, N., 1992. The degrees of development of the seed-coat and endosperm as separate subtraits of stenospermocarpic seedlessness in grapes. Vitis - Journal of Grapevine Research 31, 149-155.
Su, H., Zhu, J., Cai, C., Pei, W., Wang, J., Dong, H., Ren, H., 2012. FIMBRIN1 is involved in lily pollen tube growth by stabilizing the actin fringe. Plant Cell 24, 4539-4554.
Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO summarizes and visualizes long lists of Gene Ontology terms. PloS ONE 6, e21800.
Sweetman, C., Wong, D. C., Ford, C. M., Drew, D. P., 2012. Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression. BMC Genomics 13, 691.
The Arabidopsis Information Resource (TAIR) [http://www.arabidopsis.org/].
This, P., Jung, A., Boccacci, P., Borrego, J., Botta, R., Costantini, L., Crespan, M., Dangl, G. S., Eisenheld, C., Ferreira-Monteiro, F., Grando, S., Ibáñez, J-, Lacombe, T., Laucou, V., Magalhães, R., Meredith, C. P., Milani, N., Peterlunger, E., Regner, F., Zulini, L., Maul, E., 2004. Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theoretical and Applied Genetics 109, 1448–1458.
Torregrosa, L., Fernandez, L., Bouquet, A., Boursiquot, J. M., Pelsy, F., Martínez-Zapater, J. M., 2011. Origins and consequences of somatic variation in grapevine. In Genetics, genomics, and breeding of grapes. Edited by Zapater JM, Blondon AM, Kole C.

New Hampshire: Science Publishers,68-92.
Uniprot [http://www.uniprot.org/].
Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3, 7
Vargas, A. M., Dolores Vélez, M., Teresa de Andrés, M., Laucou, V., Lacombe, T., Boursiquot, J. M., Borrego, J., Ibáñez, J., 2007. Corinto bianco: a seedless mutant of Pedro Ximenes. American Journal of Enology and Viticulture 58, 540-543.
Varoquaux, F., Blanvillain, R., Delseny, M., Gallois, P., 2000. Less is better: new approaches for seedless fruit production. Trends Biotechnology 18, 233-242.
Venturini, L., Ferrarini, A., Zenoni, S., Tornielli, G. B., Fasoli, M., Dal Santo, S., Minio, A., Buson, G., Tononi, P., Zago, E. D., Zamperin, G., Bellin, D., Pezzotti, M., Delledonne, M., 2013. De novo transcriptome characterization of Vitis vinifera cv. Corvina unveils varietal diversity. BMC Genomics 14,41.
Vielle-Calzada, J. P., Crane, C. F., and Stelly, D. M., 1996. Apomixis: the asexual revolution. Science 274, 1322-1323.
Vitulo, N., Forcato, C., Carpinelli, E. C., Telatin, A., Campagna, D., D'Angelo, M., Zimbello, R., Corso, M., Vannozzi, A., Bonghi, C., Lucchin, M., Valle, G., 2014. A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biolotechnology 14, 99.
Vivancos, J., Spinner, L., Mazubert, C., Charlot, F., Paquet, N., Thareau, V., Dron, M., Nogué, F., Charon, C., 2012. The function of the RNA-binding protein TEL1 in moss reveals ancient regulatory mechanisms of shoot development. Plant Molecular Biology 8, 323-336.
Wang, H., Tang, W., Zhu, C., Perry, S. E., 2002. A chromatin immunoprecipitation (ChIP) approach to isolate genes regulated by AGL15, a MADS domain protein that preferentially accumulates in embryos. Plant Journal 32, 831-843.
Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics10, 57-63.
Williams, L. E., 2000. Bud Development and Fruitfulness of Grapevines. in: Raisin Production Manual. University of California, Agricultural and Natural Resources Publication. Oakland, CA.3393,24-29.

Wing, R. A., Yamaguchi, J., Larabell, S. K., Ursin, V. M., McCormick, S., 1990. Molecular and genetic characterization of two pollen-expressed genes that have sequence similarity to pectate lyases of the plant pathogen Erwinia. Plant Molecular Biology 14, 17-28.
Wold, B., Myers, R. M., 2008. Sequence census methods for functional genomics. Nature Methods 5, 19-21.
Wu, Y., Yan, J., Zhang, R., Qu, X., Ren, S., Chen, N., Huang, S., 2010. Arabidopsis FIMBRIN5, an actin bundling factor, is required for pollen germination and pollen tube growth. Plant Cell 22, 3745-3763.
Wudick, M. M., Luu, D. T., Tournaire-Roux, C., Sakamoto, W., Maurel, C., Vegetative and sperm cell-specific aquaporins of Arabidopsis highlight the vacuolar equipment of pollen and contribute to plant reproduction. Plant Physiology 164, 1697-1706.
Yang, C., Vizcay-Barrena, G., Conner, K., Wilson, Z., 2007. MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. Plant Cell 19, 3530-3548.
Yang, C., Xu, Z., Song, J., Conner, K., Vizcay Barrena, G., Wilson, Z. A., 2007. Arabidopsis MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. Plant Cell 19, 534-548.
Yang, W. C., Ye, D., Xu, J., Sundaresan, V., 1999. The SPOROCYTELESS gene of Arabidopsis is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes and Development 13, 2108-2117.
Yin, T., Pan, G., Liu, H., Wu, J., Li, Y., Zhao, Z., Fu, T., Zhou, Y., 2012. The chloroplast ribosomal protein L21 gene is essential for plastid development and embryogenesis in Arabidopsis. Planta 235, 907-921.
Yuan, M., Wang, S., 2013. Rice MtN3/saliva/SWEET family genes and their homologs in cellular organisms. Molecular Plant- Oxford Journal 6, 665-674.
Zenoni, S., Ferrarini, A., Giacomelli, E., Xumerle, L., Fasoli, M., Malerba, G., Bellin, D., Pezzotti, M., Delledonne, M., 2010. Characterization of transcriptional complexity during berry development in Vitis vinifera Using RNA-Seq. Plant Physiology 152, 1787-1795.
Zeng, Y., Zhao, T., Kermode, A. R., 2013. A conifer ABI3-interacting protein plays important roles during key transitions of the plant life cycle. Plant Physiology 161, 179-195.
Zhang, Q., Li, Z., Yang, J., Li, S., Yang, D., Zhu, Y., 2012. A calmodulin-binding protein from rice is essential to pollen development. Journal of Plant Biology 1, 8-14.

Kanneganti, V., Gupta, A. K., 2008. Wall associated kinases from plants - an overview. Physiology and Molecular Biology of Plants 14, 109-118.
Zou, Y., Aggarwal, M., Zheng, W. G., Wu, H. M., Cheung, A. Y., 2011. Receptor-like kinases as surface regulators for RAC/ROP-mediated pollen tube growth and interaction with the pistil. AoB Plants. plr017.

8. Appendix

APPENDIX 1

Table 1: Genotypic characterization of wild-type and mutant.

Fifty eight SSR (simple sequence repeats) markers, spread across the nineteen chromosomes of grapevine genome, were used to genotype the wild-type and the mutant. Marker details and PCR conditions are described in methods. Symbols: * SSR markers commonly used to discriminate grapevine varieties, - indicates homozygous allele.

SSR Marker	Multiplex panel	Dye	Labeled primer concentration (µM)	Wild-type & Mutant Allele 1	Wild-type & Mutant Allele 2	Chromosome Location
VVS2*	1	6-FAM	0,2	130	-	Chr11
VVMD32*	1	6-FAM	0,6	253	257	Chr4
VVMD28*	1	NED	0,3	236	246	Chr3
VMC1B11	2	NED	0,3	167	-	Chr8
VVMD27*	2	HEX	0,6	167	183	Chr5
VVMD7*	2	6-FAM	0,2	240	263	Chr7
VrZAG62*	3	HEX	0,6	193	195	Chr7
VrZAG79*	3	6-FAM	0,3	243	259	Chr5
VVMD25*	4	6-FAM	0,3	243	251	Chr11
VVMD5*	4	HEX	0,6	225	236	Chr16
VVIQ52	5	NED	0,2	77	84	Chr9
VVMD24	5	6-FAM	0,2	207	213	Chr14
VVIN16	5	6-FAM	0,2	150	-	Chr18
VVIV37	6	NED	0,4	160	176	Chr10
VVIH54	6	6-FAM	0,2	166	174	Chr13
VMC4F8	7	6-FAM	0,2	112	124	Chr1
VVMD21	7	NED	0,4	244	250	Chr6
VVIN73	7	6-FAM	0,2	265	-	Chr17
VVIP31	8	NED	0,3	191	195	Chr19
VVIB01	8	6-FAM	0,3	289	291	Chr2
VMC7H3	9	HEX	0,08	120	130	Chr4
VVIN56	9	6-FAM	0,1	161	-	Chr7

VVIP77	10	NED	0,2	175	183	Chr4
VMC6D12	10	HEX	0,2	179	-	Chr9
VrZAG67	10	6-FAM	0,1	129	152	Chr10
VMC3E11.2	10	NED	0,4	89	97	Chr10
VVIB09	10	6-FAM	0,1	271	277	Chr17
VrZAG21	11	6-FAM	0,1	201	203	Chr4
VMC5G1.1	11	NED	0,2	123	-	Chr6
VMCNG1F1.1	12	HEX	0,2	148	158	Chr4
VVIP72	12	HEX	0,4	43	54	Chr6
VVIQ22	12	NED	0,2	94	-	Chr17
VrZAG83	13	6-FAM	0,4	190	194	Chr4
VMC3B7.2	18	HEX	0,4	102	104	Chr19
VMC8G6	13	HEX	0,2	158	-	Chr12
VMC5C5	14	NED	0,1	115	-	Chr6
VMC3D7	14	HEX	0,1	164	-	Chr10
VMC6C10	14	6-FAM	0,25	127	-	Chr14
VMC16F3	15	HEX	0,4	176	178	Chr7
VMC1A12	15	6-FAM	0,4	121	-	Chr7
VMC2H5	15	HEX	0,1	96	108	Chr14
VMC3B8	16	HEX	0,2	134	-	Chr12
VMC5A1	16	NED	0,1	167	169	Chr16
VMC4D9.2	17	6-FAM	0,2	227	229	Chr15
VMC7G5	17	6-FAM	0,1	166	184	Chr1
VMC5E9	18	NED	0,5	192	198	Chr19
VVIB63	18	6-FAM	0,1	142	-	Chr15
VVIM43	19	HEX	0,2	75	83	Chr6
VVIV16	19	HEX	0,1	104	-	Chr18
VVIB23	single panel	6-FAM	0,4	285	-	Chr2
VMC7F2	single panel	HEX	0,2	201	-	Chr18
VMC5G7	single panel	NED	0,1	200	216	Chr2

VVP18B19	single panel	6-FAM	0,4	140	147	Chr18
VVP18B32	single panel	HEX	0,2	266	276	Chr18
P3_VVAGL11	single panel	6-FAM	0,4	91	99	Chr18
VVP18B35	single panel	6-FAM	0,4	231	245	Chr18
VVP18B40	single panel	HEX	0,2	135	151	Chr18
VVP18B20	single panel	HEX	0,2	248	252	Chr18

APPENDIX 2

Phenotypic characterization of wild type and mutant

Comparison of experimental group: WT_SP vs MT_SP

A. Bunch length (BHL)

Anova summary

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety	1	3.02	3.025	0.281	0.61
Residuals	8	86.14	10.768		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.2727, df = 1, p-value = 0.6015

Experimental group	Wild-type	Mutant
Mean value	18.06	16.96

B. Bunch weight (BHW)

Anova summary

	Df	Sum sq	Mean Sq	F value	Pr(>F)
Variety	1	243048	243048	19.01	0.00241
Residuals	8	102290	12786		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.8182, df = 1, p-value = 0.009023

Experimental group	Wild-type	Mutant
Mean value	409	97

C. Berry weight (BW)

Anova summary

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety	1	10407	10407	231.6	3.45e-07
Residuals	8	360	45		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.8598, df = 1, p-value = 0.008816

Experimental group	Wild-type	Mutant
Mean value	87.5	23

D. Mean berry weight (MBW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.62039	1	64.867	0.03828
Residuals	0.66948	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 3.8723, df = 1, p-value = 0.04909

Experimental group	Wild-type	Mutant

Mean value	3.5	0.9

E. Mean berry diameter (MBD)

Anova summary

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Variety	1	2.500	2.500	192.3	7.07e-07
Residuals	8	0.104	0.013		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.9018, df = 1, p-value = 0.008611

Experimental group	Wild-type	Mutant
Mean value	1.54	0,54

F. Seed number (SN)

Anova summary

	Df	Sum Sq	F value	Pr(>F)
Variety	1	7236.1	3.512.670	6,78E-05
Residuals	8	164.8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 7.2581, df = 1, p-value = 0.007058

Experimental group	Wild-type	Mutant
Mean value	54	0,2

G. Mean seed number per berry (MSN)

Anova summary

	Df	Sum Sq	F value	Pr(>F)
Variety	1	96.040	72.319	2,81E-02
Residuals	8	10.624		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 7.2581, df = 1, p-value = 0.007058

Experimental group	Wild-type	Mutant
Mean value	2.16	0.2

H. Total seed fresh weight (TSFW)

Anova summary

	Df	Sum Sq	F value	Pr(>F)
Variety	1	110.460	1.722.978	1,08E-03
Residuals	8	0.5129		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 7.3052, df = 1, p-value = 0.006876

Experimental group	Wild-type	Mutant

Mean value	2.240	0.138

I. Mean seed fresh weight (MSFW)

Anova summary

	Df	Sum Sq	F value	Pr(>F)
Variety	1	0.0013094	15.832	0.2438
Residuals	8	0.0066167		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 2.6129, df = 1, p-value = 0.106

Experimental group	Wild-type	Mutant
Mean value	2.16	0.2

J. Number of berries with seeds

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	1537.6	1	15376	2,00E-14
Residuals	0.8	8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 8.3333, df = 1, p-value = 0.003892

Experimental group	Wild-type	Mutant

Mean value	25	0,2

K. Number of berries without seeds

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	1537.6	1	1,54E+04	2,00E-11
Residuals	0.8	8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 8.3333, df = 1, p-value = 0.003892

Experimental group	Wild-type	Mutant
Mean value	0	24.8

APPENDIX 3

Investigation of physiological process: a comparison of quantitative traits between wild-type and mutant experimental group one.

WT_EMS+ST vs MT_EMS+ST

A. Bunch length (BHL)

Anova summary

	Df	Sum Sq	F value	Pr(>F)
Variety	1	0.23	0.0662	0.8055
Residuals	6	21.21		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.0222, df = 1, p-value = 0.8815

Variety	WT_EMS+ST	MT_EMS+ST
Mean value	11.1	11.4

B. Bunch weight (BHW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	2072.39	1	142.264	0.009269
Residuals	874.03	6		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 5.3333, df = 1, p-value = 0.02092

Variety	WT_EMS+ST	MT_EMS+ST
Mean value	42	13

C. Berry weight (BW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	297.68	1	13.653	0.01015
Residuals	130.82	6		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 5.3976, df = 1, p-value = 0.02016

Variety	WT_EMS+ST	MT_EMS+ST
Mean value	20.0	10.2

D. Mean berry weight (MBW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.32962	1	65.417	0.04304
Residuals	0.30233	6		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 3.0361, df = 1, p-value = 0.08143

Variety	WT_EMS+ST	MT_EMS+ST
Mean value	20.0	10.2

E. Mean berry diameter (MBD)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.08533	1	0.4764	0.5158
Residuals	107.467	6		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.0267, df = 1, p-value = 0.8703

Experimental group	WT_EMS+ST	MT_EMS+ST
Mean value	.07	0.5

F. Seed number (SN)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	5.339	1	0.9953	0.3517
Residuals	37.550	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.1125, df = 1, p-value = 0.7373

Experimental group	WT_EMS+ST	MT_EMS+ST
Mean value	1.75	0.3

G. Mean seed number per berry (MSN)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.05	1	0.1542	0.7063
Residuals	2.27			

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.1125, df = 1, p-value = 0.7373

Experimental group	WT_EMS+ST	MT_EMS+ST
Mean value	0.35	0.33

H. Total seed fresh weight (TSFW)

Anova summary

	Df	Sum Sq	F value	Pr(>F)
Variety	1	0.03844	22.612	0.1711
Residuals	8	0.13600		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.8092, df = 1, p-value = 0.3684

Experimental group	WT_EMS+ST	MT_EMS+ST
Mean value	0.1	0.006

I. Mean seed fresh weight (MSFW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.0002351	1	0.5943	0.4660
Residuals	0.0027690	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.1125, df = 1, p-value = 0.7373

Experimental group	WT_EMS+ST	MT_EMS+ST
Mean value	0.014	0.010

J. Number of berries with seeds

Anova summary

	Df	Sum Sq	F value	Pr(>F)
Variety	1	3.6	1.44	0.2645
Residuals	8	20.0		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.8092, df = 1, p-value = 0.3684

Experimental group	WT_EMS+ST	MT_EMS+ST
Mean value	1.25	0.03

K. Number of berries without seeds

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	2471.6	1	123.057	0.009887
Residuals	1406.0	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 3.2311, df = 1, p-value = 0.07225

Experimental group	WT_EMS+ST	MT_EMS+ST
Mean value	46	21

APPENDIX 4

Investigation of physiological process: a comparison of quantitative traits between wild-type experimental groups one and two.

WT_SP vs WT_EMS+ST

A. Bunch length (BHL)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	85.70	1	79.791	0.0255988
Residuals	75.18	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6, df = 1, p-value = 0.01431

Experimental group	WT_SP	WT_EMS+ST
Mean value	18.6	11.9

B. Bunch weight (BHW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	299064	1	221.825	0.002183
Residuals	94374	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6, df = 1, p-value = 0.01431

Experimental group	WT_SP	WT_EMS+ST
Mean value	409	42

C. Berry weight (BW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	10149	1	546.284	6,66E-05
Residuals	130	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.0504, df = 1, p-value = 0.0139

Experimental group	WT_SP	WT_EMS+ST
Mean value	87.48	19.9

D. Mean berry weight (MBW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	162.384	1	546.284	6,66E-05
Residuals	0.2081	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.0504, df = 1, p-value = 0.0139

Experimental group	WT_SP	WT_EMS+ST

Mean value	3.5	0.8

E. Mean berry diameter (MBD)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	26.402	1	298.090	5,37E-04
Residuals	0.0620			

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.1538, df = 1, p-value = 0.01311

Experimental group	WT_SP	WT_EMS+ST
Mean value	1.54	0.45

F. Seed number (SN)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	6066.8	1	2.115.449	1,73E-03
Residuals	200.8	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.2069, df = 1, p-value = 0.01273

Experimental group	WT_SP	WT_EMS+ST

Mean value	54	1.75

G. Mean seed number per berry (MSN)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	72.802	1	294.167	0.0009832
Residuals	17.324			

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.2069, df = 1, p-value = 0.01273

Experimental group	WT_SP	WT_EMS+ST
Mean value	2.16	0.35

H. Total seed fresh weight (TSFW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	96.142	1	4.005.926	1,95E-04
Residuals	0.1680	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 6.4286, df = 1, p-value = 0.01123

Experimental group	WT_SP	WT_EMS+ST

Mean value	2.18	0.4

I. Mean seed fresh weight (MSFW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.00157241	1	42.569	0.07799
Residuals	0.00258564	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 1.5517, df = 1, p-value = 0.2129

Experimental group	WT_SP	WT_EMS+ST
Mean value	0.040	0.014

J. Number of berries with seeds

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	1253.47	1	4.679.630	1,14E-04
Residuals	18.75	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 7.5, df = 1, p-value = 0.00617

Experimental group	WT_SP	WT_EMS+ST

Mean value	25	1.25

K. Number of berries without seeds

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	4651.2	1	44.434	0.0002863
Residuals	732.8	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 7.2, df = 1, p-value = 0.00729

Experimental group	WT_SP	WT_EMS+ST
Mean value	0	46

APPENDIX 5

Investigation of physiological process: a comparison of quantitative traits between mutant experimental group one and two.

MT_SP vs MT_EMS+ST

A. Bunch length (BHL)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	57.27	1	16.912	0.00627
Residuals	20.32	6		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 5, df = 1, p-value = 0.02535

Experimental group	MT_SP	MT_EMS+ST
Mean value	17	11.4

B. Bunch weight (BHW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	13246.8	1	91.692	0.02316
Residuals	8668.2	6		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 5, df = 1, p-value = 0.02535

Experimental group	MT_SP	MT_EMS+ST
Mean value	96.8	12.8

C. Berry weight (BW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	302.10	1	64.451	0.04416
Residuals	281.24	6		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 3.8008, df = 1, p-value = 0.05123

Experimental group	MT_SP	MT_EMS+ST
Mean value	23	10.3

D. Mean berry weight (MBW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.62039	1	64.867	0.03828
Residuals	0.66948	7		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 3.8723, df = 1, p-value = 0.04909

Experimental group	MT_SP	MT_EMS+ST
Mean value	0.92	0.4

E. Mean berry diameter (MBD)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.01008	1	10.312	0.3490517
Residuals	0.05867	6		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.9205, df = 1, p-value = 0.3373

Experimental group	MT_SP	MT_EMS+ST
Mean value	0.54	0.46

F. Seed number (SN)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.0	1	0	1.00
Residuals	1.6	8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0, df = 1, p-value = 1

Experimental group	MT_SP	MT_EMS+ST
Mean value	0.2	0.2

G. Mean seed number per berry (MSN)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.0	1	0	1.00
Residuals	1.6	8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0, df = 1, p-value = 1

Experimental group	MT_SP	MT_EMS+ST
Mean value	0.25	0.25

H. Total seed weight (TSW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.04489	1	0.9421	0.3602
Residuals	0.38120	8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.0222, df = 1, p-value = 0.8815

Experimental group	MT_SP	MT_EMS+ST
Mean value	0.018	0.004

I. Mean seed fresh weight (MSFW)

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.00049	1	0.5765	0.4695
Residuals	0.00680	8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0.0222, df = 1, p-value = 0.8815

Experimental group	MT_SP	MT_EMS+ST
Mean value	0.018	0.004

J. No. Of berries with seeds

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	0.0	1	0	1
Residuals	1.6	8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 0, df = 1, p-value = 1

Experimental group	MT_SP	MT_EMS+ST
Mean value	0.2	0.2

K. No. Of berries without seeds

Anova summary

	Sum Sq	Df	F value	Pr(>F)
Variety	384.4	1	45.626	0.06518
Residuals	674.0	8		

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 2.6299, df = 1, p-value = 0.1049

Experimental group	MT_SP	MT_EMS+ST
Mean value	24.8	12.4

APPENDIX 6

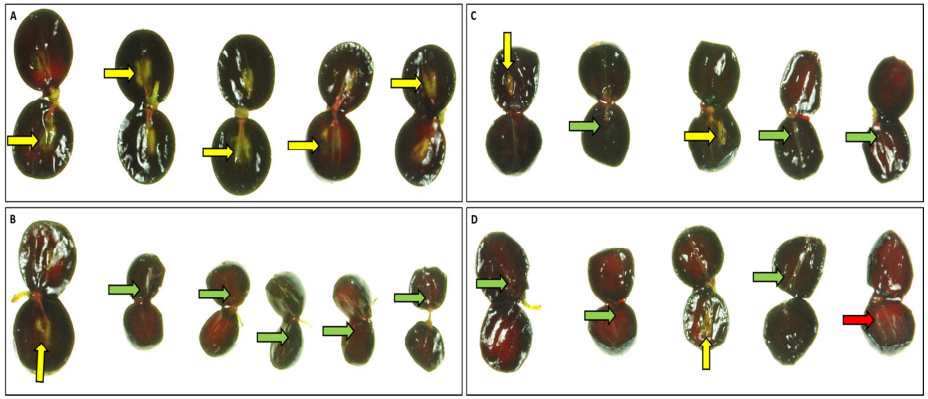


Figure 1. Sliced berries of all experimental groups. Yellow arrows indicates the presence of seeds, Red arrows indicate berries with seed traces, Green arrows indicates berries without seed trace. **(A)** Wild-type self-pollinated group. **(B)** Wild-type EMS+ST group, **(C)** Mutant self-pollinated group, **(D)** Mutant EMS+ST group.

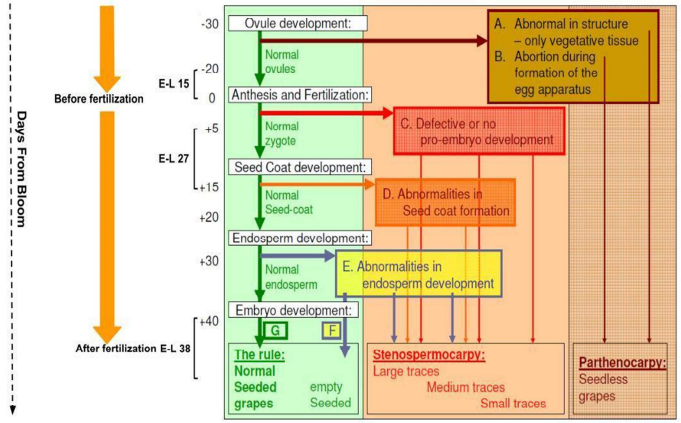


Figure 2. Diagram showing the major events that occur during seed development in grapevine. It was used as a guide for collecting RNA samples for RNA-Seq experiment (by matching sampling to days from bloom).

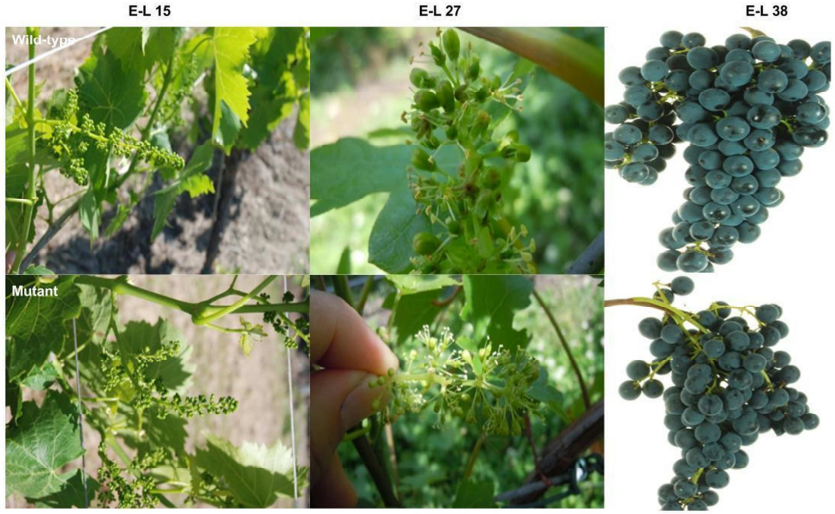


Figure 3. Picture of the materials collected from the two lines at each sampling date.

In 2009 anthesis was recorded for both lines between the last week of May and the first week of June (personal communication). With this prior knowledge of expected date of anthesis, RNA-Seq sampling for 2010 season was planned such that the initial sample collection will fall between 0 and 30 days before bloom in order to create an inventory of gene expression before fertilization (or flowering). For successive inventories of gene expression after fertilization, samplings were planned to take place between 0 and 15 days post anthesis and subsequently at harvest to cover the whole ripening process. Using the E-L system as a guide (see Schneider et al 2009), the first sampling was done on 12th May 2010 when 8 leaves were physically observed to be clearly separated and single flowers were in compact groups (corresponding to the stage E-L 15), and fell between 15 and 20 days from the expected date of full bloom. The same was done for the stage E-L 27, when the date of sampling (10th June 2010) was discounted from the actual date anthesis commenced, and it fell between 0 and 10 days post anthesis. The last sampling date (16th September 2010) corresponded to the stage E-L 38 (harvest), which was more than 40 days post anthesis. Note: anthesis in 2010 was observed between 31th May and 3rd June.

Appendix 7

Table 1. List of the genes analyzed in real-time PCR and primers used for their amplification.

The table reports the gene IDs, gene annotations and sequences of the primer sets used to analyze the transcriptional profile of 14 genes including the two constitutive genes used as reference for normalization.			
Gene ID (CRIBI)	**Gene annotation**	**Forward primer 5' -> 3'**	**Reverse primer 5' -> 3'**
VIT_01s0011g01560	Transparent testa 16	GTGATGGAGCAGTTCCCATT	TACTGGAGGGTGAGGTCCTG
VIT_01s0011g06390	MS1 (male sterility 1)	GGCAGCAAGGGTGATTGTTG	AGCTGCGTCGAACCAAGTAA
VIT_01s0026g01680	Pectate lyase	CAATACGAGCCCACATTGCG	TTCAGGTTCCTTCTCGTGCC
VIT_03s0038g04340	Feronia receptor-like kinase	TCTCCCATGGAAGGTCTGTC	AAGATCATCGACCCCCTTCT
VIT_04s0008g01800	Myb domain protein 7	TGCCGCTTTGGATCTTGACT	GCACGAGGACGTTTATAATGGA
VIT_06s0004g02820	Sand (reference gene)	CAATGTCGTCCGATTCGAGC	GATCTTGAAGGGAGTCGAGGG
VIT_09s0002g01670	Myb domain protein 26	ATTGAAACCAAGCCCATCAA	TGAGAGCCTGATGGGAGACT
VIT_09s0002g02210	Adhesion of calyx edges (ACE)	AGAGGGAGCCACATAGGGTT	TGCACTGAGCCACAGAAGAG
VIT_11s0016g03020	Pectinesterase family	ATTGGCACCTTCAATTCTGC	ATTCTAAATGCCACCGCTTG
VIT_12s0028g03270	Ethylene-responsive transcription factor 9	CAACGAAGTCCTCCCCTTCC	CAGCAGCGGAATTCACAACG
VIT_14s0006g02950	Lateral organ boundaries protein 41	AGCGGCTCTCTTTGGTTGAA	GAAGACAGGGTCGGATGGTG
VIT_15s0021g02700	Beta-expansin (EXPB4)	AGTCTTGGGGTGCCGTTTGG	GCCATCCCGCTGGAATGACA
VIT_15s0048g00510	Pectinesterase family	TCTCAAACATGGCTCAGCAC	GTGTTGCTGATGAGCTTGGA
VIT_17s0000g05400	Myb domain protein 35	GCCGAATGCAGATGGACAAC	TTCCTCCAGAAGGCTAGGGA
VIT_17s0000g10430	Glyceraldehyde 3-phosphate dehydrogenase (reference gene)	TTCTCGTTGAGGGCTATTCCA	CCACAGACTTCATCGGTGACA
VIT_18s0041g01880	MADS-box protein SEEDSTICK	AGGCTTCAGCAAGCAAACAT	CATTAAGCCGAGATGGAGGA

Table 2. Summary of read mapping to V1_mRNA version of 12X grapevine genome draft annotation.

	Wild-type			Mutant		
Developmental stage	E-L 15	E-L 27	E-L 38	E-L 15	E-L 27	E-L 38
Total No. of reads	128865364	125786280	143403274	127176972	102166350	115678764
No. of mapped reads	104054886	101638040	113217146	102918032	82120258	81346578
No. of reads mapped to single position	92582632	90292742	98800321	87999629	70331451	70888044
No. of reads mapped to multiple positions	11472254	11345298	14416825	14918403	11788807	10458534
Unmapped reads	24810478	24148240	30186128	24258940	20046092	34332186
Read coverage						
Percentage of total reads mapped	80.7	80.8	79.0	80.9	80.4	70.3
Percentage of total reads mapped to a single position	71.8	71.8	68.9	69.2	68.8	61.3
Percentage of mapped reads that align to a single position	89.0	88.8	87.3	85.5	85.6	87.1
Percentage of total reads mapped to multiple positions	8.9	9.0	10.1	11.7	11.5	9.0
Percentage of mapped reads that align to multiple positions	11.0	11.2	12.7	14.5	14.4	12.9
Percentage of total reads unmapped	19.2	19.2	21.0	19.1	19.6	29.7

Table 3. Selected individual genotypes inferred from RNA-Seq that matched Sanger sequencing.

	Reference	Alternate	Predicted genotype Mutant	Validated genotype Mutant	Predicted genotype Wild-type	Validated genotype Wild-type	Annotation
SNP A	T	G	TG	TG	TT	TT	Non-synonymous
SNP B	C	T	CT	CT	CC	CC	Non-synonymous
SNP C	G	T	GT	GT	GG	GG	Non-synonymous
SNP D	A	G	AG	AG	AA	AA	Non-synonymous
SNP E	G	T	GG	GG	GT	GT	Non-synonymous
SNP F	A	G	AG	AG	AA	AA	Non-synonymous
SNP G	C	A	CA	CA	CC	CC	Non-synonymous

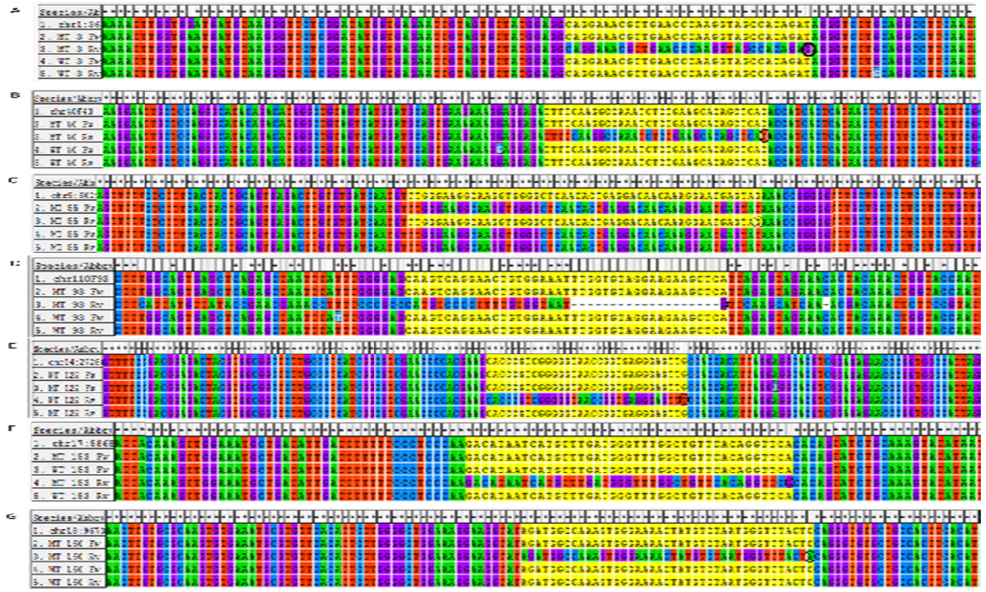


Figure 1. Selected putative SNPs confirmed through Sanger sequencing. Black circle indicates the base change.

Table 4: Proportion of RNA-Seq-derived candidate genes in the physical proximity of seed-related QTLs.

	Chr1	Chr2	Chr4	Chr5	Chr12	Chr14	Chr15	Chr18
10-Mbp window (Mbp)	0-9.5	0-9	15-25	0-9.8	0.5-10.5	11.8-21.8	6.6-16.6	22.7-32.7
No of candidate genes in the chromosome	75	63	84	86	97	86	72	143
No of candidate genes within the 10-Mbp window	32	41	32	32	59	21	34	35
Percentage of candidate genes in the 10-Mbp window	42,7	65,1	38,1	37,2	60,8	24,4	47,2	24,5

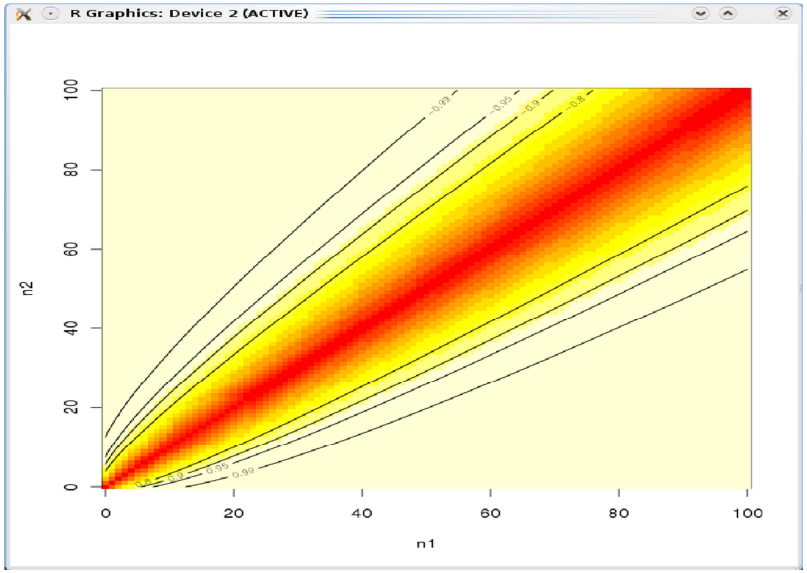
Appendix 8

Description of the procedure adopted to rank the transcripts by order of magnitude

P-values (or scores) were computed to reflect the significance of the difference between 2 counts (n1 and n2 corresponds to any two library combination out of the six libraries, independently of the genotype) using a binominal model. The model is described below. The p-values were log-transformed in order to allow for greater numerical stability in comparing extreme values. The sign of the p-value reflects the direction of the comparison (whether n1 is greater or lesser than n2). The smaller is absolute p-value, the more significant is the difference between the counts. Next we considered all the p-values and the ratios of expression between the counts to compute a ranking value for each transcript. Afterwards the ranking values were used to sort the transcripts and show on top the biggest differences in expressions between two of the libraries.

Model description

Assuming we sequenced N1 reads in sample1 (resp N2 reads in sample2), and n1 of those reads (resp. n2) are mapping into a given region of interest in the genome, we are interested in determining whether the expression in sample1 is significantly different from the expression in sample2. If we assume the events have the same probability of been observed in the two samples, n1 and n2 should follow a binomial distribution with the same probability of event p=(n1/N1 + n2/N2)/2. We can then estimate the probability of observing a count less than n1 or greater than n2 according to this model. Furthermore, we can put a sign on the probability to reflect the direction of the comparison. For example, a score of -0.9 may be interpreted as: «there is 90% chance that sample1 is under-expressed relatively to sample2»; and a score of +0.9 may be interpreted as «there is 90% chance that sample1 is over-expressed relatively to sample2». The picture below shows an overview of the score obtained when n1 and n2 are between 0 and 100, and N1, N2 are fixed to 1'000'000. We can for example see that under this model there is 95% probability that a count of n1=20 compared to a count of n2=40 is significantly different when there are 1'000'000 events in each sample. Note: We have observed that the model is not very well appropriate to compare large values. This issue may be related to a saturation effect.



Overview of the scores obtained with the binomial model when comparing two counts (n1, n2) between 0 and 100 with (N1,N2) fixed to 1'000'000.

9. Acknowledgement

I wish to express my gratitude to my supervisor Dr Laura Costantini for her helpful inputs and constructive criticism on this research project. Many thanks to Dr Stella Grando, Dr Anna Schneider and Dr Ivana Gribaudo for co-supervising this project. Stella provided me with all the support I needed to achieve my research objectives at the Fondazione Edmund Mach di San Michele all'Adige. Anna and Ivana have a special way of putting smile on my face any time I am in Torino. Their expert contribution to my research, assistance, support and encouragement kept me motivated throughout research project. Special thanks to Silvia Lorenzi for her patience and understanding during my learning of important research techniques used for the project, Silvia always makes complex techniques very simple to understand. Indeed I am thankful.

I am most grateful to the Fondazione Edmund Mach di San Michele all'Adige
for awarding me a PhD fellowship without which I would never have left Abuja for Trento to pursue the PhD programme.

I would further wish to thank my friends, Mridula, Francis, Dong, Lukas, Raymond and Paula for a good relationship we've had during the PhD programme. Personally I want to thank my family for their assistance, encouragement, patience and kind understanding during the PhD programme. Finally I return all glory to God almighty for showing me his mercy ever since I was born.

10. Dedication

This Thesis is dedicated to
Aina Ogbonna-Nwafor and Akaomachi Ogbonna-Nwafor