Comparative analysis of the genetic variation of flowering time genes in *Vitis vinifera*

Dissertation

submitted by
Nadia Kamal

for the Degree of
Doctor of Natural Sciences (*Dr. rer. nat.*)
at the Faculty of Biology
Bielefeld University, Germany

Bielefeld, October 2017

Contents

1 Summary

Grapevine is one of the economically most valuable fruit crops in the world. In recent years climate change and global warming have led to significant strains on viticulture. Higher temperatures at earlier dates within the growing season advance the date of bud burst and flowering. Consequently, the predicted shift in véraison to earlier dates causes ripening to occur under higher temperatures which impairs wine quality. It is of high interest for viticulture to investigate the timing of flowering and elucidate its genetic architecture. Genetic variation between cultivars leads to differences in the expression of phenotypic traits such as flowering time. A deeper understanding of varietal differences in the phenology of flowering time is thus critical to select varieties that are well adapted for production under current and future climatic conditions.

The goal of this thesis was therefore to identify alleles of Flowering Time Control (FTC) candidate genes that might be involved in the timing of flowering. More than four hundred FTC candidate genes were identified in grapevine by using functional data from *A. thaliana* to exploit the grapevine genome for homologues. In addition, previously defined quantitative trait locus (QTL)-regions for FTC in grapevine were considered for the selection of candidate genes. From this selection, 72 genes were chosen for an amplicon sequencing approach with a focus on flowering time related QTL-regions. The amplicons subjected to sequencing were derived from the parental lines and 35 individuals of the F1-mapping population GF.GA-47-42 x 'Villard Blanc'. This population segregates for the trait flowering time. In order to identify alleles of amplicon sequenced genes, a bioinformatic workflow for the phasing of the alleles of each gene was established.

Alleles in genomic regions of 46 genes on 16 chromosomes with a length of up to 8.3 kb were distinguished. The inheritage of alleles of closely neighbored genes remains largely constant, which indicates the functionality of the established workflow. Moreover, analyzing inheritance patterns within an F1-population including the parental lines proved to be an effective method for the validation of allele phasing results. Functional molecular markers that are capable of distinguishing between up to four alleles and suitable for analyzing polymorphisms located in genomic regions were developed.

A genetic association study between alleles of FTC candidate genes and the timing of flowering was performed. Using a Wilcoxon Rank-Sum test, correlation between specific alleles of FTC candidate genes and flowering time phenotypes was investigated. This revealed a correlation within and outside of QTL-regions for flowering time on chromosome 1, 4, 14, 17, and 18. Among others, appearance of an allele of *Vitis vinifera WITH NO LYSIN KINASE 6 (VvWNK6)* inherited from early flowering GF.GA-47-42 was found to highly correlate with early flowering.

In addition to the above mentioned population, alleles of FTC genes were analyzed in other *Vitis* cultivars. They were investigated for common alleles of FTC candidate genes with a focus on alleles correlating with flowering time phenotypes. Common alleles of FTC candidate genes were found across different cultivars. However, common alleles were usually only shared between a few cultivars since grapevine exhibits a high genetic diversity. Within a diverse panel of eleven well-known grapevine cultivars, all late flowering cultivars were found to harbor an allele of *Vitis vinifera TOMATO MADS BOX GENE 6 (VvTM6)*.

In order to further analyze and confirm the role of FTC candidate genes, RNA-Seq experiments were performed. Differential gene expression was analyzed over consecutive developmental stages of buds and inflorescences of early flowering GF.GA-47-42 and late flowering 'Villard Blanc'. Both analyzed cultivars show similar expression patterns for most genes. Many of the analyzed FTC candidate genes showed differential expression between consecutive time points, indicating functional roles of these genes in the flowering control network. Moreover, most of these genes were found to be expressed in buds and inflorescences but not in leaves. This further supports their role in flowering initiation and floral development.

The findings of this study provide indications of the genetic factors controlling or influencing flowering time in grapevine. For future research the clarification of these factors can contribute to the development of molecular genetic markers that are capable of predicting flowering time phenotypes. This is of crucial importance in light of the need to breed new grapevine varieties adapted for production under changing climatic conditions.

2 Background

2.1 *Vitis vinifera*

Vitaceae is a pantropical family of fourteen genera and about 910 known species (Christenhusz *et al.* 2016) and covers mostly woody or herbaceous lianas (Mullins *et al.* 1992). The most common genus of the *Vitaceae* is *Vitis* which contains 79 accepted species[1]. *Vitis* species are mainly found in temperate zones of the Northern hemisphere and are almost equally distributed between the American and Asian continents. Only one *Vitis* species is native to Europe, *Vitis vinifera* L., which consists of two species, *Vitis vinifera* L. subspecies *sylvestris*, the wild grapevine, and *Vitis vinifera* L. subspecies *vinifera*, the cultivated grapevine (Boss *et al.* 2003; Mullins *et al.* 1992). The wild species *Vitis vinifera sylvestris* was domesticated by humans to obtain hermaphrodite plants producing many large, sweet berries (Duchêne 2016). As a result, *Vitis vinifera vinifera* (hereafter called *Vitis vinifera*) is hermaphroditic while *Vitis vinifera sylvestris* is dioecious.

Grapevine is globally one of the major fruit crops based on cultivated areas, economic value and various uses. The growing of grapes and the making of wine has been of great importance for Western civilizations since ancient times (Mullins *et al.* 1992; Wan *et al.* 2013). Grapevine cultivation and domestication seems to have its origin between the seventh and fourth millenia BC in the near East between the Black Sea and Iran (Terral *et al.* 2010). From there, cultivated forms were spread by humans throughout the middle and near East up to central Europe. Today, *Vitis vinifera* is the most cultivated grapevine species worldwide (Wan *et al.* 2013). All common known grape varieties, such as 'Chardonnay', 'Riesling', and 'Cabernet Sauvignon' are derived from it. The estimated surface area for grape growing was about 7.5 million ha in 2014, with the greatest portion in Europe. More than 74 million tons of grapes are produced annually[2]. Grapes are consumed fresh as table grapes and dried as raisins. They are used for the production of fresh grape juice, jelly, jam, ethanol, vinegar, grape seed oil, tartaric acid, and fertilizer. Moreover, they have several industrial purposes; such as the prodcution of anthocyanin pigments and antioxidants for food additives, cosmetics,

[1] http://www.theplantlist.org/
[2] http://faostat.fao.org

and pharmaceutical products (Carmona *et al.* 2008). However, the main usage of grapes is wine production. According to the Food and Agriculture Organization of the United Nations (FAO), in 2014 53% of the world's yield of grape was used for winemaking. *Vitis vinifera* is hence of crucial importance for the production of wine spirits that are consumed not only in Europe but in countries all over the world.

2.1.1 The genome of *Vitis vinifera*

Vitis vinifera is a diploid plant with 38 chromosomes (n = 19). It has a genome size of approximately 475–500 Mb (Jaillon *et al.* 2007) and about 32,000 genes (Vitulo *et al.* 2014). Outbreeding by means of wind and insect pollination is the norm. Therefore, all grapevine varieties are highly heterozygous. Inbreeding depression is severe; sterility often occurs from the first to third generation of selfing. It has been shown that the amount of sequence divergence between alleles is about 13% (Jaillon *et al.* 2007; Velasco *et al.* 2007). Nevertheless, a reference genome sequence is available since 2007 (Jaillon *et al.* 2007). This sequence is derived from a cultivated clone of *Vitis vinifera* 'Pinot Noir' (PN40024), which has been bred to about 93% of homozygosity by successive selfing. The *Vitis* genome sequence was first established from an 8x Sanger sequencing assembly (Jaillon *et al.* 2007) and was updated with a 12x assembly in 2010 when additional Sanger sequences were added to the sequences pool. The PN40024-12xv0 assembly, available on the website of the international Grape Genome Browser Genoscope[3], was improved and updated to the 12xv2 assembly by the French-Italian Public Consortium but the sequences of contigs and scaffolds remained unchanged compared to version 12xv0[4] (Adam-Blondon *et al.* 2011). Several different gene sets derived from different gene predictions are available at present. One gene set derived from PN40024-12xv0 is available since 2009 and can be found on the Genoscope website[5]. The gene prediction 12Xv1 is the result of the union of v0 and a gene prediction performed at the *Centro di Ricerca Interdipartimentale per le Biotecnologie Innovative* (CRIBI) in Padova, Italy (Forcato 2010) and available on the CRIBI website[6]. In 2014 an improved gene set was established on the 12xv0 assembly version at the CRIBI (Vitulo *et al.* 2014).

[3]http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/
[4]https://urgi.versailles.inra.fr/Species/Vitis/Data-Sequences
[5]http://www.genoscope.cns.fr/externe/Download/Projets/Projet_ML/data/12X/annotation/
[6]http://genomes.cribi.unipd.it

2.1.2 The reproductive developmental cycle of grapevine

In contrast to annual plants that flower once in their life cycle, *Vitis vinifera* is a polycarpic perennial plant and flowers multiple times during its lifetime. Patterns of seasonal flowering occur repeatedly and the plants cycle between periods of flowering and vegetative growth. The reproductive developmental cycle is completed over two consecutive growing seasons separated by a dormancy period (Carmona *et al.* 2008) (Fig. 1). In spring of the first season, meristematic protuberances, called uncommitted primordia, lateral meristems or *'Anlagen'* (from the German word for uncommitted primordia) are formed. Latent buds are formed in the axis of grapevine leaves and do not burst until the second year after formation. Both vegetative and reproductive tissues are formed simultaneously on the same shoot by the shoot apical meristem. Uncommitted primordia formed in latent buds develop into inflorescence primordia while uncommitted primordia formed on rapidly growing shoots usually develop into tendrils (Carmo Vasconcelos *et al.* 2009). Tendrils and inflorescences share a common ontogenic origin which determines important differences in the vegetative development of the plant as well as in its flowering transition. Inflorescence primordia enter dormancy in autumn of the first year when day length decreases. They then stay dormant throughout winter (Fennell *et al.* 2015). Bud growth is halted during bud dormancy, which is an adaptive strategy for grapevine plants to better tolerate winter conditions such as low temperatures and frost (Fennell *et al.* 2015). The second season is initiated with bud burst in the following spring when environmental conditions become permissive (Carmona *et al.* 2008; Lebon *et al.* 2008). Shortly before and during bud burst, flower initials are formed and flower parts become distinct. The conversion from inflorescence primordia to inflorescences resumes as shoot development begins in the spring (Carmo Vasconcelos *et al.* 2009). Inflorescence formation is regulated at the level of formation of uncommitted primordia and at the level of differentiation. Flowering occurs around May - June in the Northern hemisphere, after which berry formation and berry ripening take place.

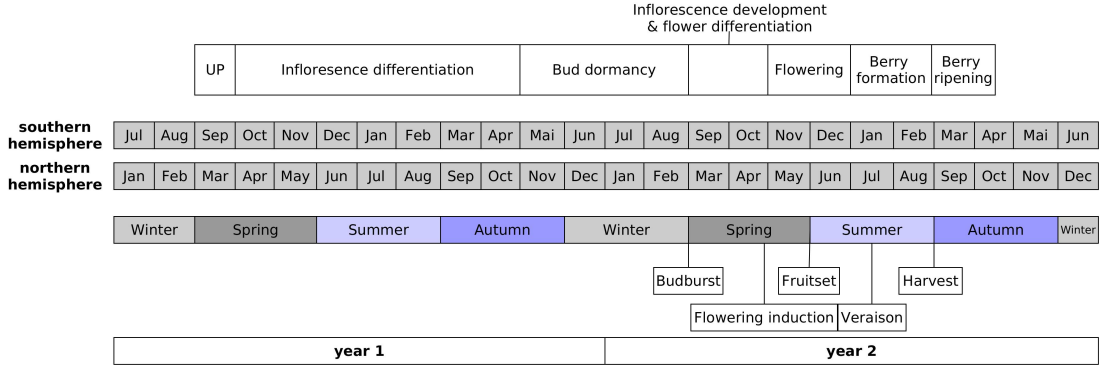

Figure 1: Illustration of the reproductive developmental cycle of grapevine showing the consecutive developmental stages of inflorescence formation and differentiation over the two growing seasons. UP: uncommitted primordia. (modified after Carmona *et al.* 2008; Fraga *et al.* 2013).

2.2 Flowering in plant species

In higher plants flowering represents the process of sexual reproduction, allows genetic recombination and hence the evolution of plants. Seeds and fruits are the major components of yield in crops. Thus the investigation of the processes influencing flowering in crop plants has been of great interest in agricultural research, especially in regards of climate change and global warming in the past decades.

Flowering is influenced by numerous external and internal factors. It needs to be strictly regulated in plants to guarantee for reproductive success by enabling the completion of seed development under favorable environmental conditions. The timing of flowering in plants depends on various environmental factors such as photoperiod, temperatures, vernalization, ambient CO_2 concentration, soil microbiota, and water and soil nutrient availability (Blackman 2017). However, the most important factors influencing the timing of flowering within a season are responses to temperature and photoperiod in addition to vernalization. Photoperiod is the duration a plant is exposed to light (Andrés *et al.* 2012). Plants with different photoperiodic patterns initiate flowering in different seasons since variations in day length are associated with the changing seasons. Temperature affects numerous living processes in plants, such as germination, flowering, and seed set. It regulates and controls plant development and can modify the timing of phenological cycles in plants significantly. Increased temperatures can cause phenological developmental processes such as flowering to occur earlier during

the growing season (Craufurd *et al.* 2009; Parker 2012). Vernalization describes the continuous exposure to low temperatures mostly during winter, which promotes flowering in many plant species. Moreover, internal factors including phytohormones and sugars help regulate a diverse range of developmental processes and influence the timing of flowering.

2.2.1 Flowering and flowering time control in grapevine

In grapevine flowering is not promoted by photoperiod and vernalization as in many other plant species (Carmona *et al.* 2008). The environmental stimuli that promote flowering in grapevine are short-term exposure to high temperature and high light intensities (Boss *et al.* 2002; Carmo Vasconcelos *et al.* 2009; Carmona *et al.* 2008, 2007; Mullins *et al.* 1992). The natural habitat of *Vitis vinifera* is forest. In order to reach full sunlight needed for flowering tendrils are used for climbing into the canopy to a height of 20-30 m (Boss *et al.* 2002). An increase in light and temperature are the environmental stimuli that a wild grapevine cane encounters when reaching the forest canopy. This promotes flowering induction and causes flower formation the following season. Under natural conditions grapevine plants need several years before producing the first flowers. Flowering induction occurs in latent buds during the summer of the first of the two consecutive growing seasons. However, flower initiation and development does not take place until the following spring when the second growing season begins (Carmona *et al.* 2007; Mullins *et al.* 1992). For optimum initiation of inflorescences it has been shown that a pulse of four to five hours of temperatures above 20°C with some variation among cultivars was sufficient (Carmo Vasconcelos *et al.* 2009).

Even though it has been reported that flowering is not affected by photoperiod in grapevine, grapevines are facultative long-day plants since inflorescence initiation and floral development are long-day facultative (Sreekantan *et al.* 2010). Long-day plants flower when the day length exceeds a critical value. There are no photoperiod requirements for flowering induction in grapevine but some cultivars produce higher numbers of inflorescence primordia per bud under long-day than under short-day photoperiods. It has been shown that American varieties such as *Vitis riparia* and *Vitis labrusca* enter dormancy at longer day lengths than *Vitis vinifera* (Sreekantan *et al.* 2010). Moreover, *Vitis vinifera* x *Vitis labrusca* vines grown in Delaware, USA form many more inflorescences when grown under long day conditions (Sreekantan *et al.* 2010). These indications suggest that photoperiod has an impact on the regulation of floral development in *Vitis*.

At the hormonal level, gibberellins and cytokinins are the principal regulators of

flowering in grapevine. Giberellic acid (GA) promotes the formation of uncommitted primordia, which develop into either inflorescences or tendrils. GAs are inhibitors of flowering in many fruit species but their role in grapevine varies with the stage of bud development. The initiation and development of lateral meristems is promoted by GAs as well as their development into tendrils. Inflorescence development on the other hand is suppressed by GAs. GA is a promoter of flowering at an early stage but acts as an inhibitor of flowering later on and promotes vegetative growth. Genetic evidence of the flowering inhibitory role of GAs in grapevine is provided by the phenotype of a grapevine dwarf mutant derived from the L1-meristematic layer of the champagne cultivar *Pinot Meunier*. This mutant is gibberellin-insensitive through a mutation in the gene *VvGAI*, the *A. thaliana GIBBERELLIC ACID INSENSITIV (GAI)* orthologue (Boss *et al.* 2002). The mutant plants are dwarfs with short internodes and tendril-primordia transformed to inflorescence-primordia. Cytokinins act as promoters of inflorescence development from lateral meristems. The variation of cytokinin- and GA-levels might determine the fate of lateral meristems. Subsequently, external stimuli, such as light and temperature, might regulate the differentiation of lateral meristems through their effects on hormone biosynthesis or response pathways (Carmona *et al.* 2007).

2.2.2 Impact of climate change on flowering time

In recent years the need to explore the implications of the changing climate on crop yields has become increasingly urgent. Climate change not only involves the changes in mean weather conditions, but also the changes to climate variability, seasonal weather patterns, and the frequency and magnitude of extreme weather events (Mosedale *et al.* 2015). In viticulture, seasonal variation contributes to enormous variation in yield and quality, in particular in cool-climate viticulture (Carmo Vasconcelos *et al.* 2009). Yields are not only affected by mean seasonal conditions but also vulnerable to the risk of damaging or unsuitable weather conditions at key phenological stages during the growing season, such as bud burst and flowering. Global warming leads to a tendency to higher temperatures at earlier dates during the growing season. This causes the processes of bud burst, flowering and véraison (the time when grape color changes) to occur earlier. The predicted shift in véraison to earlier dates causes ripening to occur under higher temperatures (Leeuwen *et al.* 2013). This affects grape composition and hence wine quality. Elevated temperatures impair the accumulation of anthocyanins in the berries which is unfavorable for wine quality (Duchêne 2016). In addition, with earlier ripening a greater proportion of fruit is also ripening in a shorter time window, leading to a compressed harvesting period. This results in a significant strain on harvesting resources and logistics as the time available to transport and process the fruit is compressed (Sweetman *et al.* 2014). Climate change can also

affect the number of flowers and thus the number of berries per plant. It has been demonstrated that the higher the temperatures around bud burst, the lower the number of flowers per inflorescence (Duchêne 2016). Furthermore, increased temperatures, lead to longer drought periods and water stress which also poses a strain to viticulture (Pavlousek 2011). Water deficit during floral initiation can lead to a decreased number of inflorescences and can have a negative influence on berry weight especially when applied after véraison (Duchêne 2016).

A huge risk to viticulture is late spring frost as it causes significant crop loss when occurring after bud burst. In addition to the probability of frost after bud burst, the risk of frost damage is determined by the frequency and severity of frosts. The latter depends on the duration and intensity of subzero temperatures to which plants are exposed. However, grapevines are able to tolerate temperatures as cold as -15°C. The acclimation to cold temperatures is a normal aspect of their physiology. Nevertheless, cold or rainy weather around flowering can reduce the number of grape clusters formed and thus lead to a reduction in yield (Mosedale *et al.* 2015; Mullins *et al.* 1992). In respect to the weather conditions, yield can vary up to 32.5%, which makes grapevine by far the crop with the highest seasonal variation in yield (Boss *et al.* 2003; Lebon *et al.* 2008).

Vitis vinifera is a very adaptable species and can be grown in a wide range of environments, such as the deserts of Australia and California on the one hand and Eastern England on the other hand. Climatic variations within the grape growing zones lead to variation in the composition and flavor of grapes and wines. Wine grapes mostly grow in temperate climates between the 50^{th} and 30^{th} degree latitude on both hemispheres (Fig. 2) (Mullins *et al.* 1992). Temperature is the main factor defining the upper latitude limits of viticulture. Thus, despite the disadvantages the increase in mean seasonal temperatures has brought for viticulture, it has also led to improved growing conditions in many cool-climate viticulture regions (Mosedale *et al.* 2015).

The variation in flowering time between varieties grown under different environmental conditions is an adaptive response to pressure caused by the seasonal timing of climatic factors like frost, temperatures, snowmelt, and drought. Biotic interactions play a role as well. The abundances of competitors, herbivores, seed predators, and pollinators vary seasonally and geographically and their cycles are adapting to the changing climate (Blackman 2017).

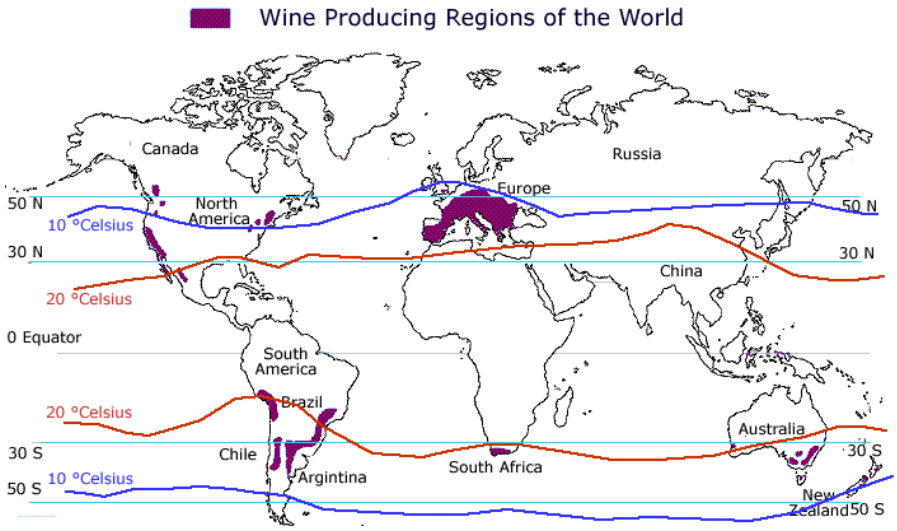

Figure 2: Main wine growing areas of the world between the 30th and the 50th degree latitude on both the Northern and Southern hemispheres. Figure from http://cgge.aag.org/GlobalEconomy1e/CaseStudy2_Chile%20Wine_Sep10/index.html

2.2.3 Flowering time control genes and pathways

The investigation of processes beneath the control of flowering and flowering time have become increasingly important due to climatic changes and the shift of flowering induction to earlier dates. Therefore, the genetics of natural variation in response to ambient temperature have been studied excessively. Via forward genetic approaches in *A. thaliana* and other model plants genetic pathways that promote flowering in response to environmental cues as well as many of the genes controlling flowering have been characterized (Andrés *et al.* 2012). More than 180 genes involved in the control of flowering time have been identified so far (Peng *et al.* 2015). Besides the identification of genes that regulate flowering, the determination of the specific genes that cause natural variation in flowering time is also required. Genes involved in natural variation of flowering time in *A. thaliana* include *CONSTANS (CO)*, *FLOWERING LOCUS C (FLC)*, *VERNALIZATION INSENSITIVE 3 (VIN3)*, *Phytochrome D (PHYD)*, *GAI*, and *FRIGIDA (FRI)* (Ehrenreich *et al.* 2009). In recent years, orthologues and paralogues of the genes identified in *A. thaliana* have been identified in various crop species such as rice, barley, poplar, maize, tomato, sunflower, and sugar beet. Nevertheless, several flowering control genes identified in crop species were not found in *A. thaliana*. The cereal flowering repressor, *VERNALIZATION2 (VRN2)* has no homologue in *A. thaliana* genome. *VRN2* inhibits long-day induction of

FT-like1 (FT1) prior to winter and is repressed during cold exposure. In rice, the *FLOWERING LOCUS T (FT)*-like gene *Hd3a* is repressed in long days and upregulated under short day conditions by genes including *Early heading date (Ehd1)* and *Oryza sativa MADS51 (OsMADS51)*, which both have no known homologues in the *A. thaliana* genome (Greenup *et al.* 2009). These findings demonstrate the rapidity with which the regulatory pathways controlling flowering responses to environmental cues have evolved, even among closely related species. The pathways that regulate the adaptation of flowering to environmental signals are not at all highly conserved among species. Mechanisms controlling flowering time show a high amount of variation between taxa indicating that the flowering time control network has the flexibility to evolve in many different ways (Andrés *et al.* 2012; Blackman 2017).

In grapevine, genes homologous to the *A. thaliana* floral development pathway genes or involved in photoperiod or vernalization responses were identified. Various studies have worked on the identification of grapevine homologues to *A. thaliana* flowering signal integrators, floral meristem identity genes, and flower organ identity genes (Boss *et al.* 2006; Calonje *et al.* 2004; Carmona *et al.* 2002; Fernandez *et al.* 2006; Joly *et al.* 2004; Sreekantan *et al.* 2006). *FRUITFUL-LIKE (VvFUL-L)* and *APETALA1 (VvAP1)* - *Vitis* homologues to the *A. thaliana FRUITFULL (FUL)* and *AP1* - are suggested to act on the specification of flower organ identity as their expression appeared in early developmental stages of lateral meristems and maintained in both inflorescence and tendril primordia (Carmona *et al.* 2007). The *Vitis LEAFY (LFY)* orthologue *Vitis vinifera FLORICAULA* (*VvFL*) participates in inflorescence- and flower-development. It is expressed in uncommitted primordia and later in inflorescence- and tendril-meristems. The expression of *VvFL* decreases in developing tendril meristems while increasing in proliferating inflorescence meristems. It reaches its highest level in the floral meristems after bud burst. This correlation might suggest that the expression of *VvFL* is required to reach a certain threshold in order to initiate reproductive development. In *A. thaliana* a threshold level of *LFY* expression has also been shown to promote flower initiation. It integrates flowering signals from different pathways and controls the expression of flower organ identity genes (Carmona *et al.* 2002). The expression of flower organ identity genes, like *LFY*, *AP1*, and *FUL* are regulated by flower signal integrators, such as *SUPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* and *FT* in *A. thaliana* (Carmona *et al.* 2008). Most of these genes are *MADS-box* genes, which are transcription factors involved in developmental processes and signal transduction in eukaryotes. They are defined by the presence of a conserved domain, the MADS-box, in the N-terminal region. The MADS-box domain is involved in DNA binding and dimerization with other MADS-box

proteins (Diaz-Riquelme *et al.* 2009). *MADS*-box genes in plants are called *MIKC*-genes and are divided into five classes (A-E) regarding their role in floral organogensis. *MIKC*-genes in *Vitis* were also found to be associated with specific floral developmental processes as well as flowering transition. For instance, SEPALLATA (SEP) genes are MIKC class-E floral homeotic genes involved in floral meristem determinacy and the specification of flower organ identity in *A. thaliana*. Their homologues in *V. vinifera* are expressed in flowers and fruits (Diaz-Riquelme *et al.* 2009). Moreover, *VvFLC1 and VvFLC2* are located in flowering time related QTL-regions in *Vitis* and highly expressed in buds (Diaz-Riquelme *et al.* 2009). In*A. thaliana, FLC* is a key regulator of flowering time.

A study of Achard *et al.* (2007) showed that activated ethylene signaling reduces gibberellic acid levels and leads to an increase in *GAI* accumulation in *A. thaliana*. *GAI* acts as a repressor of *LFY* and *SOC1* and thus represses flowering. Accumulation of GA increases the abundance of GA-biosynthesis gene transcripts, which consequently downregulate *GAI*. In the *Vitis vinifera GAI*-mutant (*VvGAI*), that forms inflorescences instead of tendrils (Boss *et al.* 2002), a correlation between inflorescence development and increased *VvFL* expression was observed. Gibberellins may repress inflorescence development, directly or indirectly by affecting *VvFL* expression. They promote the initiation of lateral meristems but inhibit their development as inflorescences in favor of tendril development. In *A. thaliana* gibberellins promote flowering through the expression of *LFY*. It is suggested that *A. thaliana* and grapevine share a gibberellin-mediated flowering pathway targeting *LFY/ VvFL*, although with opposite effects on its expression (Carmona *et al.* 2007).

Environmental cues modulate the timing of flowering through diverse transcriptional and posttranscriptional regulatory mechanisms, including DNA methylation, chromatin modification, small and long non-coding RNA activity, protein degradation, and protein transport (Blackman 2017). Moreover, temperature dependent splice variants of several genes, such as *FLOWERING LOCUS M (FLM)* and other *MIKC*-genes were found in *A. thaliana*. *Short Vegetative Phase (SVP)* was shown to have a temperature-dependent degradation pattern (Blackman 2017). In *Vitis SVP* is expressed in latent buds and in vegetative and reproductive organs similar to *A. thaliana* where *SVP* and *FLC* interact to negatively regulate the expression of the floral pathway integrators *SOC1* and *FT* (Lee *et al.* 2007). The seasonal separation between inflorescence and flower meristem formation in grapevine could require the participation of transcriptional repressors, such as members of the *FLC* and *SVP* subfamilies, to supress the development of flower meristems prior to the dormancy period (Diaz-Riquelme *et al.* 2009).

Key transcription factors that are responsive to environmental cues are responsible for the initiation of early stages of flowering in plants (Andrés *et al.* 2012). *MADS* transcription factor *VvMADS8* is most highly expressed in axillary buds in grapevine prior to flowering and hence seems to promote early flowering (Sreekantan *et al.* 2006). *VvMADS8* is a homologue to *A. thaliana SOC1* that is involved in flowering initiation (Sreekantan *et al.* 2006). In *Vitis* its expression is high during very early stages of inflorescence development, decreases in later stages of flower development and is not expressed in flowers and fruits at all. The expression of *VvFT* - the orthologue of the *A. thaliana FT* - is observed in infloresences and is associated with seasonal flowering induction in latent buds and the development of inflorescences, flowers, and fruits. *FT* expression is regulated by PHYTOCHROME INTERACTING FACTOR 4 (PIF4). The binding site of *PIF4* is more accessible under higher temperature through the release of a certain type of nucleosomes (H2A.Z) (Kumar *et al.* 2012). Overexpression of *VvFT* in *A. thaliana* is shown to accelerate flowering (Carmona *et al.* 2007). On the other hand, *PIF4* is repressed by DELLA proteins under cooler temperatures which delays flowering. DELLA proteins are transcriptional repressors and downregulate GA-response genes. The degradation of DELLA proteins is mediated by GA. Hence, GA influences *FT* expression in a temperature-dependent process revealing a possible mechanism where GA and temperature act together to regulate flowering (Kumar *et al.* 2012).

2.3 Analysis of the genetic variation of quantitative traits

An understanding of the genetic and environmental factors causing the variation of quantitative traits is of great importance in functional genomics. Various approaches have been established to gain information of the genetic variation behind complex trait. Connecting the genetic variation of genes and loci involved in a specific trait, such as flowering time, to the trait's phenotype allows a deeper understanding of genetic factors controlling and influencing the trait. This knowledge can subsequently be used for genomic selection and crop breeding for enhanced properties.

2.3.1 QTL-analyses

QTLs are genomic regions containing several genes that influence a certain quantitative trait. In the post-genome era, researchers have been working on

identifying QTL-regions responsible for various quantitative traits, such as flowering time (Samad *et al.* 2017). QTL-analyses are powerful for the identification of loci influencing traits of interest using the principle of genetic linkage (Scheben *et al.* 2017). It is essential to identify contributing candidate QTLs and genes influencing flowering time at a genetic level to accurately estimate phenological shifts in flowering time and to select crops better adapted to future environmental conditions. QTL-analyses aim to identify the genes within a QTL-region with an impact on the trait of interest and determine whether all genes influence the trait equally. QTLs are mapped by linkage disequilibrium with molecular markers that exhibit Mendelian segregation. If a QTL is linked to a marker locus, a difference in mean values of the quantitative trait among individuals with different genotypes will be observable at the marker locus (Mackay 2011). The closer the QTL and marker locus, the larger the difference in trait phenotypes between the marker genotypes, with the maximum difference when the marker genotypes coincide exactly with the QTL. A genetic map for a quantitative trait can be constructed when several QTLs for the trait are found in the genome.

The effect of a single QTL, and interactions between QTLs, can be efficiently studied by constructing near isogenic lines (NILs) that differ only at a single QTL-region. Segregating populations, in the order of thousands of individuals, derived from crossing such NILs, can be used to narrow down the position of the QTL to a small genomic region in which high confidence candidate genes can be found. Finally, the identity of a QTL is validated by genetic transformation and complementation tests. NILs that differ for alleles of a QTL can be used for high-resolution mapping and positional cloning of the candidate gene like shown in *A. thaliana* for the *early day-length insensitive (EDI)*-QTL (El-Assal *et al.* 2001).

Vitis species have been shown to have a high degree of variability with an average of 3.4 million SNPs (1/140 bp) per genotype (LePaslier *et al.* 2013). Due to this high heterozygosity, the severe inbreeding depression, and the long generation time, in *Vitis vinifera*, mapping strategies employing NILs, recombinant inbred lines (RILs) or multi parent RIL (MPRIL)/ multi-parent advanced generation inter-cross (MAGIC) populations cannot be used. A pseudo-testcross strategy (Grattapaglia *et al.* 1994) is usually applied for generating heterozygous F1-mapping populations with subsequent construction of genetic maps. This approach leads to two separate parental maps and a consensus map, respectively. The first filial generation is used for QTL-mapping unlike the norm in other crop or model species (Gascuel *et al.* 2017).

Within grapevine breeding programs QTLs for the timing of developmental stages

such as flowering have been identified (Duchêne *et al.* 2012; Fechter *et al.* 2014). Flowering and flowering time in grapevine are complex quantitative traits influenced by numerous genes of which many are localized in flowering time QTL-regions. Two independent QTLs responsible for variation in the phenological duration of flowering were identified. *VvFT* on chromosome 7 and the *CONSTANS*-like gene *VvCOL2* on chromosome 14 were the most reliable candidates identified from two of the underlying QTLs co-localizing for flowering time (Duchêne *et al.* 2012; Jagadish *et al.* 2016). *COL2* is known to be associated with the genetic variations of flowering time in *Medicago truncatula* and *Medicago sativa* (Jagadish *et al.* 2016). For flowering time, a first segregating locus was reported by Gökbayrak *et al.* (2006) but could not be resolved. Fechter *et al.* (2014) identified six QTLs on different chromosomes in a bi-parental pseudo F1-mapping population for the start of flowering.

2.3.2 Candidate genes studies

Candidate gene studies are focused on the selection of genes related to a trait and thus come with prior knowledge about gene function. The candidate gene approach begins with the selection of putative candidate genes based on their relevance in the mechanism of the investigated trait (Kwon *et al.* 2000). Several strategies for the identification of candidate genes can be applied. In the position dependent strategy, the identification of candidate genes is mainly based on the physical linkage information in QTL-regions previously identified by QTL-mapping analysis. Candidate genes are sought out from numerous genes harbored in the targeted QTL-region. Commonly used linkage analysis methods often result in QTLs containing hundreds of genes in the logarithmic odds ratio (LOD) interval supporting a QTL (Zhu *et al.* 2007). This makes the prioritization of positional candidate genes difficult. Moreover, quantitative traits are usually affected by multiple QTLs which further contributes to a high number of putative candidate genes. Another strategy for the identification of candidate genes is by applying comparative genomics. Candidate genes are selected based on their homology to genes whose role in affecting the phenotypic variation in the trait of interest has already been confirmed in other species. However, the phenotypic expression of the trait of interest can differ between species and the selected candidate genes can therefore have different roles in the investigated species (Zhu *et al.* 2007).

Variants in a selected population are genotyped and their role in affecting the trait's phenotype can be confirmed by evaluating the effects of the causative gene variants in an association analysis. Variants are then assessed for potential functional consequences, such as affecting gene regulation or the gene's product. Such polymorphisms are usually evaluated for their role as future molecular

markers for the identification of traits, like resistances and diseases. Instead of relying on markers that are evenly spread throughout the genome regardless of their function or context in a specific gene, candidate gene studies focus on genes that are selected because of a priori hypotheses about their role in a trait. Furthermore, a candidate gene study is usually conducted in a population based sample of affected and unaffected individuals or individuals that have different phenotypes of a trait, such as early and late flowering plants. A candidate gene study therefore takes advantage of both the increased statistical efficiency of association analysis of complex traits and the biological understanding of the phenotype, tissues, genes, and proteins that are likely to be involved in the trait (Patnala *et al.* 2013; Tabor *et al.* 2002).

Since a large number of flowering time genes has been identified through molecular developmental genetics in *A. thaliana*, flowering time is a particularly attractive trait for candidate gene association studies. However, in contrast to model organisms like *A. thaliana*, in *Vitis* many of the genes and pathways underlying ecologically significant traits, such as flowering time, are still not completely known.

2.3.3 Next generation sequencing technologies and bioinformatic tools

To investigate the genetic variation of an organism with an available reference genome, such as *Vitis vinifera*, DNA-sequencing, RNA-sequencing, or epigenome sequencing can be conducted. DNA-sequencing technologies include whole genome, whole exome, and targeted sequencing. The genetic variation can be accessed by comparing the sequencing outcome to a genome reference sequence. Variations include, for instance, single nucleotide polymorphisms (SNPs), structural variations, and copy number variations, and can be analyzed by using various bioinformatic tools (Park *et al.* 2016). Studies on transcriptomes, that include the synthesis of complementary DNA from RNA, allows the investigation of differential gene expression, RNA splicing, gene fusion, and mutations (Park *et al.* 2016). The transition of gene expression studies from previous-hybridization based microarray methods to Next Generation Sequencing (NGS)-based methods in recent years, enabled the identification and quantification of transcripts without prior genomic information (Van Dijk *et al.* 2014).

Advances in throughput, powerful computational pipelines, and cost reduction have made sequencing of entire genomes increasingly feasible. However, depending on the task to be analyzed it might not be necessary to sequence entire genomes. A vast amount of approaches that addresses defined regions of the genome have

emerged (Van Dijk *et al.* 2014). A high level of targeting is achieved by amplicon sequencing, that is based on the polymerase chain reaction (PCR) amplification of previously specified genomic regions using sequence specific primers. Amplicon sequencing requires sequence information of the target locus, is characterized by high levels of sequencing depth, and can be done very effectively (Van Dijk *et al.* 2014). It is however not practical for large-scale projects (Lancia *et al.* 2001).

The rapid enhancement in NGS technologies has profoundly influenced crop genotyping. The identification of sequence based polymorphisms such as Single Sequence Repeats (SSRs) and SNPs made informative molecular genetic markers largely available and led to an acceleration in their identification process (Bhat *et al.* 2016). Genetic markers are a powerful tool in crop breeding programs that can be used to identify genotypes and connect them to phenotypes (Scheben *et al.* 2017).

In recent years massive amounts of genomic data have been produced. This requires solutions for data investigation, analysis, and management which in turn led to great demands on bioinformatics tools. The first step in the investigation of high-throughput genotyping data is the quality control and subsequent quality trimming of raw sequences. Low-quality base calls, PCR duplicates, and sequence contaminants can lead to genotyping errors and influence downstream analyses (Scheben *et al.* 2017). For most genomic analyses the alignment of sequencing reads to a reference sequence is required before variations can be investigated. For many major crops, such as wheat, maize, rice, tomato, sugar beet, grapevine, and many others genome reference sequences are already available. Most common sequencing platforms, such as Illumina-HiSeq and -MiSeq, generate short sequence reads that are aligned to a reference sequence in an alignment step. Numerous alignment programs have been developed since the emergence of high-throughput short-read sequencing, such as Bowtie/Bowtie2 (Langmead *et al.* 2012) and the more accurate BWA/BWA-MEM (Li 2013; Li *et al.* 2009a) that use the fast Burrows–Wheeler transform approach. The performance of different alignment algorithms can depend on the input data and the quality of the reference genome (Scheben *et al.* 2017).

For genotyping analyses a variant discovery process is necessary beforehand. Also for this task a vast amount of tools has been developed. SAMtools (Li *et al.* 2009b) and GATK (DePristo *et al.* 2011; McKenna *et al.* 2010; Van der Auwera *et al.* 2013) are some of the most popular and are based on a probabilistic approach. Results of variant detection in high-throughput sequencing data is still inconsistent between different variant callers which is due to differences in how variants are identified. While SAMtools considers each site individually and applies a hidden

Markov model to estimate sequencing errors, GATKs HaplotypeCaller assembles local haplotypes and uses Bayesian methods for handling errors (Scheben *et al.* 2017). Moreover, GATK offers increased accuracy by improving alignments locally before calling variants and allows the selection of ploidy levels other than haploid or diploid. However, it requires extensive formatting of the input before variant calling and is less flexible compared to other tools. Raw variants are usually filtered for read depth, read mapping quality, genotype quality, and other parameters and can be carried out by most variant callers (Scheben *et al.* 2017).

The storage and management of the increasing amounts of genomic data are highly challenging. Sequencing platforms generally output FASTQ-format files containing quality encoded sequencing reads. BAM-format files containing aligned reads are usually several times smaller. However, they do not contain quality values and hence FASTQ-files must be stored. Eventually, Variant Call Format (VCF)-files containing detected variants are comparably smaller. Recently cloud-based storage systems are emerging and might provide a solution for the storage of large-scale genomic data (Scheben *et al.* 2017).

2.3.4 Haplotype analyses

Within a mapping population, segregating for a specific quantitative trait such as flowering time, the extreme phenotypes are likely to be inherited by specific combinations of alleles at several loci. As a diploid organism, grapevine has two sets of each chromosome and hence two possible combinations of each gene. The full sequence of nucleotides in an individual chromosome is referred to as "haplotype" and is not directly observable from sequence data as it generally consists of unphased genotypes. If the two sequences of a gene (alleles) on the homologous chromosomes are different, the organism is heterozygous with respect to that gene, while it is homozygous if they are the same. In order to correlate genotypes or allele sequences with the phenotype of a trait, the two alleles of heterozygous genes need to be distinguished.

Although high-throughput DNA sequencing technologies and thus the generation of genotypic data has become increasingly comprehensive and inexpensive, sequencing technologies still suffer from producing haplotype phase ambiguous sequence reads. Determining the haplotype phase of an individual is computationally challenging and experimentally expensive; but haplotype phase information is crucial in various analyses, such as genetic association studies, the reconstruction of phylogenies and pedigrees, genomic imputation, linkage disequilibrium, and SNP tagging (Aguiar *et al.* 2012; Browning *et al.* 2011; Martin *et al.* 2016). In diploid organisms, like grapevine, genes are generally expressed from both alleles of a gene. Different

alleles can show different gene expression patterns and consequently varying forms of specific traits. The determination of these alleles is an important step in the declaration of corresponding traits. Among other approaches, haplotypic information can be obtained from DNA sequence fragments to reconstruct the two haplotypes of an individual. A sequence fragment that covers at least two variant sites in a genomic region can link these variants together and thus separate or phase them. When fragments are long enough to encompass multiple variant sites, and the sequencing coverage is sufficient enough to provide overlaps between fragments, the fragments can be assembled to reconstruct long haplotypes (Bansal *et al.* 2008).

3 Objectives

Genetic variation in genes and alleles influencing or controlling the timing of flowering leads to differences in the phenology of flowering time between cultivars. The induction of flowering in grapevine is mainly affected by hours of sunlight and high temperature. Therefore, climate change and global warming have led to major constraints on viticulture in recent years. A deeper understanding of interactions between genes or allele combinations and environmental factors is critical to select varieties that are adapted to current and future climate conditions. Therefore, the goals of this thesis were threefold:

1. In order to investigate the genetic variation of candidate genes involved in flowering time in grapevine and connect trait's genotypes and phenotypes, a genetic association study between alleles of previously defined FTC candidate genes and flowering time phenotypes was performed. After amplicon sequencing of selected genes in an F1-mapping population segregating for flowering time, the first goal was the phasing of alleles of sequenced genes and the establishment of a suitable bioinformatic workflow. Thereafter, alleles of FTC genes were to be correlated with the timing of flowering aiming to identify combinations of alleles and loci influencing the timing of flowering in *Vitis vinifera*.

2. The extent of parentage among grapevine cultivars is relatively high since sexual reproduction is the major driver of genetic diversity in cultivated grapevine. Different *Vitis* cultivars were analyzed for common alleles of FTC candidate genes with a focus on alleles correlating with flowering time phenotypes. The goal of this task was to examine whether alleles of FTC candidate genes, that were found to correlate with the timing of flowering in the above mentioned population, lead to a similar phenology of flowering time in other cultivars. This would give indications towards the confirmation of the role of these alleles in the control of flowering time.

3. The expression of FTC candidate genes in grapevine underlies seasonal patterns. In order to further investigate and confirm the role of these genes in the timing of flowering, RNA-Seq experiments were performed in early flowering GF.GA-47-42 and late flowering 'Villard Blanc'. Subsequently, differential gene expression was investigated over consecutive developmental

stages of buds- and inflorescences for both cultivars. In addition, genes differentially expressed between the cultivars at certain time points, exhibiting a shift in the expression pattern, or showing different expression patterns of a gene were sought. This was performed with the goal to identify genes controlling or influencing flowering timing .

4 Materials and Methods

4.1 Identification and annotation of flowering time control genes

Within this work, putative FTC genes were defined (Dr. Daniela Holtgräwe, Chair of Bioinformatics and Genome Research, Bielefeld University). For the identification and characterization of putative flowering time genes, functional data from *A. thaliana* and other well studied model plants was systematically exploited to identify FTC candidate genes in the *Vitis* genome. Therefore the 12xv0 version of the published PN40024 reference genome sequence (Jaillon *et al.* 2007; Velasco *et al.* 2007) and the gene prediction 12xv0 from Genoscope[1] were used. Since it has been shown in *A. thaliana* that QTL-regions responsible for natural phenotypic variation are enriched for candidate genes (Atwell *et al.* 2010) candidate genes in significant QTL-regions for FTC were identified based on sequence homology to the grapevine reference protein sequences. FTC related QTLs in *Vitis* were identified previously (Fechter *et al.* 2014). Genomic sequences of the identified FTC genes were later aligned to the reference assembly PN40024-12xv2. Using blast (Camacho *et al.* 2009) protein sequences of candidate genes from *A. thaliana* were compared against the *Vitis* protein sequences. For the functional annotation of candidate genes the method of reciprocal best hits (Ward *et al.* 2014) was applied. A reciprocal best hit (RBH)-pair consists of two sequences from different sets of sequences, whereas each displays the highest global score in the other data set. Genomic sequences of FTC candidate genes, whereas the longest sequence was used in case a gene displayed several transcripts, were compared against protein sequences of *Vitis vinifera* with blastx. Protein sequences of *Vitis* FTC candidate genes were compared against *A. thaliana* protein sequences with blastp. Using tblastn the hit showing the highest score was compared back against *Vitis vinifera* coding genes. When the original query was found to have the highest score, the resulting RBH-pair was considered.

Using a comprehensive list of *A. thaliana*, micro-RNAs that are involved in

[1] http://www.genoscope.cns.fr/externe/Download/Projets/Projet_ML/data/12X/annotation/

flowering time (Coupland et al., personal communication), homologues were identified in *Vitis vinifera* by comparing the miRNA-sequences against the *Vitis* genome using blast and selecting the best blast hit.

For the definition of gene names of genes that were not already described elsewhere, they were named according to the *Vitis*-RBH or the *A. thaliana* RBH in case no *Vitis* RBH was available. When no RBH was available, the *Vitis* best hit was used. *"Vv"* for *"Vitis vinifera"* was added in front of the gene name. When several genes had identical names, a small alphabetic character was appended to the gene name.

The list of FTC candidate genes is accessible under the following link: https://docs.cebitec.uni-bielefeld.de/index.php/s/kQLUhp5HAt8j4Qa. Other supplemental data can also be downloaded there (see appendix: Additional data).

4.2 Flowering time control candidate genes for amplicon sequencing

From the selection of previously determined FTC candidate genes, 72 genes were chosen for amplicons sequencing (appendix table A1, Fig. 3). Genes for amplicon sequencing were selected with a focus on previously defined QTL-regions for flowering time (Fechter *et al.* 2014) as well as random candidates on other chromosomes.

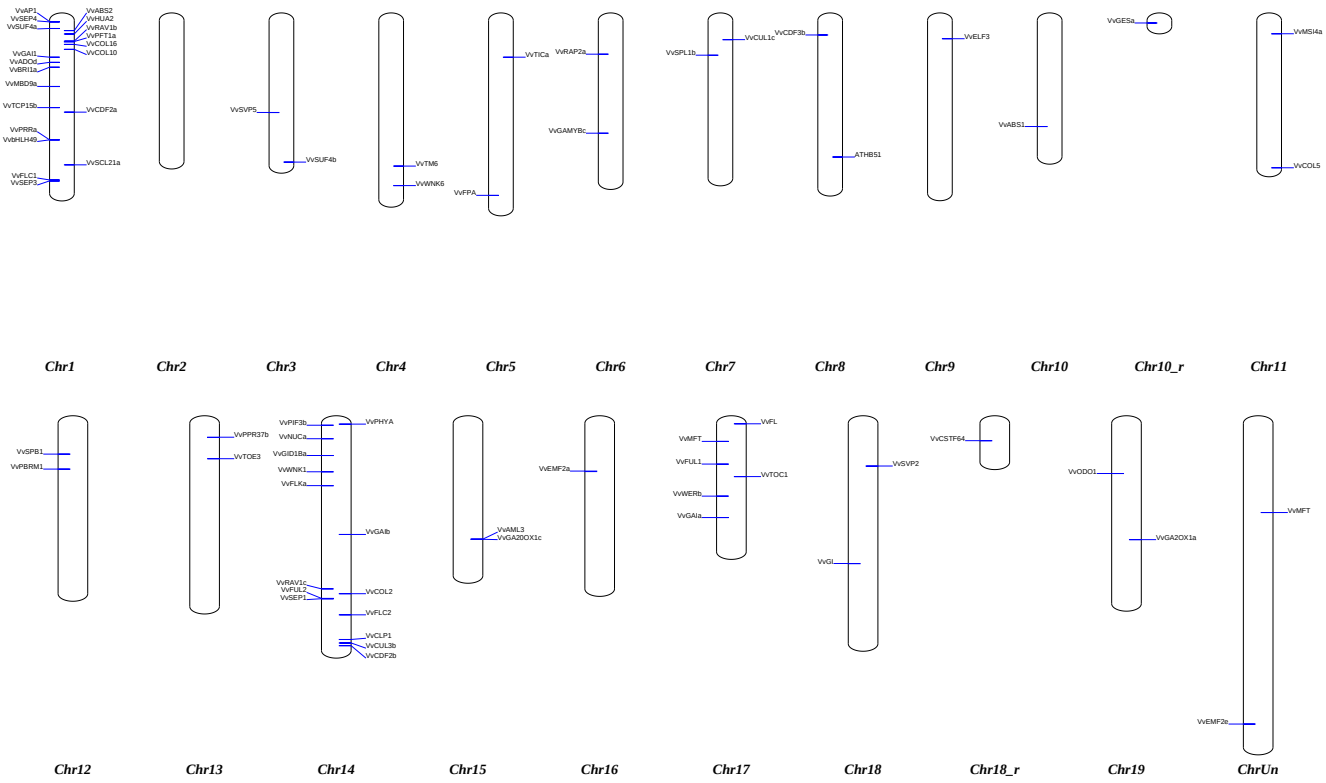

Figure 3: Physical chromosome map of *Vitis vinifera* showing FTC candidate genes selected for amplicon sequencing. Genes displayed on the left side of the chromosome are localized on the sense-strand while genes displayed on the right side are localized on the antisense-strand. Chr: Chromosome, r: random

4.3 Amplimer generation

The cDNA sequence of each candidate gene was compared against the grapevine reference sequence PN40024-12Xv0 with blast. The genomic DNA sequences including introns in addition to 1,000 bp from the 5'- and 3'-UTR-regions were extracted. Primers were designed for overlapping amplimers of up to 8 kb using the online-tool *Primer3*[2] (appendix table A2). Figure 4 shows an illustration of primer design for overlapping amplimers taking the gene *VvCOL16* as an example.

[2]http://biotools.umassmed.edu/bioapps/primer3_www.cgi

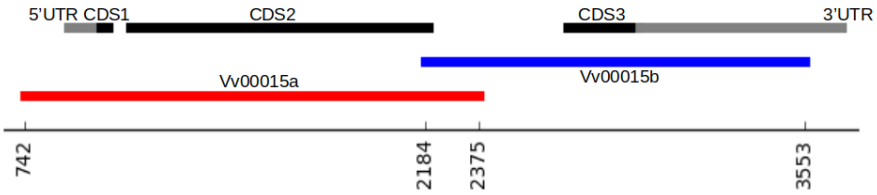

Figure 4: Illustration of amplimer and primer design through the example of *VvCOL16*. Vv00015a and Vv00015b are overlapping amplimers. CDS (coding sequence) is shown in gray and UTR-regions in black.

4.4 Plant material

4.4.1 The F1-mapping population GF.GA-47-42 x 'Villard Banc'

The pseudo F1-mapping population GF.GA-47-42 x 'Villard Blanc' was used for the analysis of flowering time. This population was crossed in 1989 using the breeding line GF.GA-47-42 and the cultivar 'Villard Blanc' and planted in the vineyards of the Institute for Grapevine Breeding Geilweilerhof in Siebeldingen, Southwestern Germany (49°13'05.0"N 8°02'45.0"E)[3] in 1996. It consists of 151 F1-individuals and segregates considerably for the trait flowering time. The maternal breeding line GF.GA-47-42 ('Bacchus weiss' x 'Seyval') is early flowering, while the paternal line 'Villard Blanc' (Seibel 6468 x 'Subereux') flowers very late compared to the median flowering time of the population (Fig. 5). This leads to differences in flowering time between the two lines of up to 17 days depending on the year. The parental lines of 'Villard Blanc' both flower late, while the parental lines of GF.GA-47-42 have an intermediate flowering time phenotype They flower earlier than 'Villard Blanc' but later than GF.GA-47-42 under similar environmental conditions. F1-individuals have numerical designations starting with "89-30-", followed by an individual number for every line. These identifiers are used within this work. The population GF.GA-47-42 x 'Villard Blanc' is described in more detail in Fechter *et al.* (2014) and Zyprian *et al.* (2006, 2016).

[3]http://www.jki.bund.de/de/startseite/institute/zuechtung-reben.html

(a) GF.GA-47-42

(b) 'Villard Blanc'

Figure 5: Inflorescences of GF.GA-47-42 and 'Villard Blanc' of the same date (June of the second year of the developmental cycle). GF.GA-47-42 has already reached full bloom since more than 50% of flowerhoods have fallen and anthers are visible. 'Villard Blanc' has not yet lost 50% of its flowerhoods and hence not reached full bloom.

4.4.1.1 Phenotyping for flowering time in the mapping population GF.GA-47-42 x 'Villard Blanc'

In cooperation with the working group of Prof. Reinhard Töpfer (JKI, Siebeldingen), phenotypic data of the mapping population GF.GA-47-42 x 'Villard Blanc' was collected. Phenotyping for flowering time (full bloom) was performed in nine years (1999 and 2009 - 2016) according to stage 65 of the BBCH-scale (50% of flower hoods fallen) (Lorenz *et al.* 1995). The date of flowering was then counted in days from the 1st of January of the respective year. The length of the flowering period varied between 10 (2016) and 17 (2012) days between the considered years. Flowering start in the population was between 147 (2011) and 168 (2016) days after January 1st. The end of the flowering period was between 157 (2011) and 183 (2013) days after January 1st (Fechter *et al.* 2014). The parental lines GF.GA-47-42 and 'Villard Blanc' are not the earliest or latest flowering individual of the population in every year. Table 1 shows the first and last day of the flowering period in days after January 1st in the years 1999 and 2009 - 2016 and the lengths of the flowering periods.

Table 1: Lengths of flowering periods in the mapping population GF.GA-47-42 x 'Villard Blanc' as well as the first and last day of the flowering period in days after January 1st in the analyzed years 1999 and 2009 - 2016.

Year	**Start of flowering period (days after January 1st)**	**End of flowering period (days after January 1st)**	**Length of flowering period (days)**
1999	165	178	14
2009	156	170	15
2010	151	180	19
2011	147	157	11
2012	153	169	17
2013	168	183	16
2014	150	161	12
2015	156	167	12
2016	168	177	10

Flowering dates were classified according to 6 stages for flowering time following OIV descriptor 302 (OIV 2009) (Zyprian *et al.* 2016) (1 = very early flowering; 2 = early flowering; 3 = medium early flowering; 4 = medium late flowering; 5 = late flowering; 6 = very late flowering). The flowering period of each year was divided into six even periods and given the descriptors 1 - 6. For the determination of the median of flowering time for each individual, the days of the flowering period of each year were numbered whereas the first day was numbered with one, the second day with two, etc. These numbers were then divided by the length of the flowering period. The resulting values were used to calculate the median (Anna Werner/ JKI, Siebeldingen).

From the 151 F1-individuals of the population, 35 were chosen for amplicon sequencing of candidate genes based on their flowering time phenotype. An attempt was made to select lines so as a normal distribution of flowering time phenotypes will be reached among them. For the selection of F1-individuals, the flowering classes 1 and 2, 3 and 4, and 5 and 6 were combined and termed 'early'-, 'intermediate'- and 'late'-flowering respectively. Approximately equal numbers of early, intermediate, and late flowering F1-individuals were chosen for amplicon sequencing. However, the selection of the lines was performed in 2013 and with the collection of further phenotypic data, flowering classes of some of the cultivars shifted. Table 2 shows the number of F1-indivuduals that was initially encountered per flowering time class according to phenotypic data from 1999 and 2009 - 2016.

Table 2: Number of F1-individuals per flowering time class according to data from the years 1999 and 2009 - 2016 and using days after January 1st.

Description of flowering time class	Flowering time class	Number of selected F1-individuals
very early flowering	1	1
early flowering	2	9
intermediate early flowering	3	6
intermediate late flowering	4	11
late flowering	5	7
very late flowering	6	1
total		35

4.4.2 Important grapevine cultivars and their flowering time phenotypes

In addition to the mapping population GF.GA-47-42 x 'Villard Blanc', eleven other important grapevine cultivars and several breeding lines were used for the analysis of FTC gene alleles within this work (Table 3). Phenotypic data of lines grown at the vineyards of Geilweilerhof were collected over seven years from 2009 - 2016 by members of the working group of Prof. Reinhard Töpfer (JKI, Siebeldingen). The data was collected at the Institute for Grapevine Breeding Geilweilerhof in Siebeldingen (49°13'05.0"N 8°02'45.0"E). The flowering time phenotypes of the cultivars are relative to each other; cultivars declared as early flowering, flower early compared to the other cultivars, etc. Especially, cultivars that do not exhibit the extreme phenotypes (either early or late flowering) are hard to distinguish. Nevertheless, the cultivars were allocated to 5 flowering time classes (early, intermediate-early, intermediate, intermediate-late, and late) (Ludger Hausmann/ JKI, Siebeldingen). See appendix table A3 for the studied lines, their flowering time phenotypes, and the phenotypes of the lines they originate from. Pedigree information was derived from the Vitis International Variety Catalogue (VIVC)[4].

[4]http://www.vivc.de

Table 3: Grapevine cultivars chosen for amplicon sequencing of FTC candidate genes.

Cultivar
'Börner'
'Chardonnay'
GF V.3125
GF-GA-47-42
GF.GA-52-42
'Grenache Noir'
'Pinot Noir'
'Pinot Noir précoce'
'Riesling'
'Silvaner'
'Solaris'
'Syrah'
'Villard Blanc'
35 F1-individuals of GF.GA-47-42 x 'Villard Blanc' (appendix table A4)

4.5 DNA isolation and amplimer amplification

The extraction of genomic DNA was performed from young leaves. The plant material was grounded under liquid nitrogen and subsequently used for DNA isolation according to the protocol *Purification of Total DNA from Plant Tissue (Maxi Protocol)* P. 28 of the DNeasy® Plant Maxi Kit (Qiagen). Amplicons were amplified by a long-distance touchdown PCR. Long distance PCRs are suitable for the amplification of long fragments while a touchdown protocol was applied in order to prevent the amplification of unspecific fragments and to increase the amplification rate of the PCR product. The annealing temperatures was increased by 1°C in every PCR cycle. The Q5® hot start high-fidelity DNA polymerase - a thermostable, hot start DNA-polymerase with a 3' → 5' exonuclease activity from NEB[5] - was used.

The PCR mix for amplicon-PCRs consisted of 6 µl of 10x reaction buffer, 0.6 µl dNTPs, 1.5 µl of each primer (10 mM), 0.3 µl Q5-Polymerase (5 U/µl), 17.1 µl H_2O (Merck), 3 µl DNA (5 ng/µl).

The following PCR program was used:

[5]http://www.neb-online.de

Temp.	Duration	PCR-step	
98°C	30 sec.	initial denaturation	
98°C 72°C – 57°C 72°C	10 sec. 30 sec. 5 min.	denaturation primer annealing elongation	15 cycles -1°C per cycle
98°C 58°C 72°C	10 sec. 30 sec. 5 min.	denaturation primer annealing elongation	25 cycles
72°C 10°C	2 min. unlimited	final elongation	

For gelelectrophoresis one percent agarose gels in 1x TAE-buffer were used (50x TAE-buffer: 242 g/l Tris, 57.1 ml/l acetic acid, 10 ml/l 0.5 M EDTA [pH 8.0]). 0,001% ethidiumbromid was added to the gel. 1x TAE was used for electrophoresis buffer and the 1 kb DNA-ladder (NEB) as a DNA size marker. 2 µl of loading buffer were added to 5 µl of the PCR samples and 6 µl were loaded on the gel.

For 100 ml of loading buffer 50 g 100% glycerin, 20 ml 0.5 M EDTA, 0.2 g xylencyanol FF 2.5%, 0.1 g bromphenolblue FF 2.5%, 0.2 g orange G, 0.1 g SDS, and 80 ml TAE were mixed and diluted 1:4 with 1x TAE and used in a concentration of 1:5.

4.6 Library preparation and amplicon sequencing

The deep sequencing of PCR products (amplicons) allows efficient variant identification and characterization in specific genomic regions. The sequencing of the amplicons was carried out at the Center for Biotechnology - CeBiTec of Bielefeld University by Prisca Viehöver. Amplicon sequencing was performed on a MiSeq in seven runs. All amplicons belonging to a respective F1-individual, parental line, or unrelated cultivar were pooled in equimolar amounts, fragmented and subsequently used for library preparation. The libraries were prepared as recommended by Illumina (TruSeq DNA Sample Preparation v2 Guide). Adaptors were ligated to the DNA fragments to allow paired-end sequencing. Adaptor-ligated fragments were size selected on a two percent low melt agarose gel. Fragments that carry adaptors on both ends were enriched by PCR. Final libraries were quantified using PicoGreen and quality checked by HS-Chips on a BioAnalyzer. Up to 20 libraries were pooled and sequenced on a MiSeq flowcell with PhiX spiked in. Libraries were sequenced on an Illumina MiSeq platform with either 2 x 250 (runs 1,3,4,5, and 7), 2 x 150 (run 2), or 2 x 300 (run 6) base pairs (bp) read length

and an average insert size of 500 bp. Diverse read lengths are due to the usage of multiple library preparation kits. After sequencing basecalling, demultiplexing and FASTQ-file generation was performed using a casava-based in house script.

4.7 Read processing and mapping

Adapter trimming of raw reads, quality filtering of reads with a window of four consecutive bases and with a quality value below 30 as well as the cropping of bases at the reads heads and tails with quality values below 30 was performed using the tool Trimmomatic (Bolger *et al.* 2014). Before and after trimming, the tool FastQC[6] was used to check the quality of the reads.

The processed reads were mapped to the grapevine reference sequence PN40024-12xv2 using the BWA-MEM-algorithm (Li *et al.* 2009a) which is suitable for long reads. Mapping was performed with default parameters and for each individual separately. Instead of the entire reference sequence, only the sequences of the genes were chosen for mapping in order to prevent false positive mapping results. SAM-format files were converted to BAM-format files and sorted using SAMtools (Li *et al.* 2009b). Readgroups were added and duplicated reads removed using Picard Tools[7].

4.8 Allele phasing

For the identification of the two alleles of the sequenced genes (phasing), a bioinformatic workflow including various tools from the genome analysis toolkit GATK (**VanderAuswera2013**; DePristo *et al.* 2011; McKenna *et al.* 2010) was established (Fig. 6).

4.8.1 Post-alignment quality enhancement of read alignments

After read alignment, the quality of the alignments was improved in two ways. Firstly, local realignments around InDels was performed using GATK's InDelRealigner to reduce the number of misalignments. Occasionally, the presence of insertions or deletions in individuals with respect to the reference genome leads to misalignments of reads to the reference, especially when InDels are covered at the start or end of a read. Such misalignments lead to many false positive SNPs. Secondly, base quality scores of reads in the aligned mapping files were

[6]http://www.bioinformatics.babraham.ac.uk/projects/fastqc
[7]http://broadinstitute.github.io/picard/

recalibrated using GATK's BaseRecalibrator in order to correct for variation in quality with machine cycle and sequence context. Thus, more accurate and more widely dispersed quality scores are provided.

4.8.2 Variant calling

Using GATK's HaplotypeCaller variants were called for each individual separately. The HaplotypeCaller determines which regions of the genome it needs to operate on, based on the presence of significant evidence for a variation. The program reassembles active regions using De Bruijn-like graphs and identifies possible haplotypes. Each haplotype is then realigned to the reference haplotype using the Smith-Waterman algorithm in order to identify variant sites. Using the PairHMM algorithm the tool performs a pairwise alignment of a read against each haplotype for every active region. These produced likelihoods of haplotypes are then marginalized to obtain the likelihoods of alleles for each putative variant site given the read data. For each putative variant site, the program calculates the likelihoods of each genotype per sample (i.e. an individual) given the read data observed for that sample. The most likely genotype is then assigned to the sample (McKenna *et al.* 2010).

Optimal parameters for variant calling using the HaplotypeCaller were decided on by considering the particularities of the experimental setup. A bias in allele frequency between the two alleles can lead to genotyping errors when the allele frequency drops below a certain threshold. Variants then fail to be detected. However, as variant calling was performed in a family of parents and their offspring, by the respect of Mendelian constraints, it was possible to distinguish between sequencing errors and true variants with very low allele frequencies due to a PCR bias. Hence, the ploidy was set to 12 for variant detection. The twelve artificial alleles were then expected to be present at approximately 8,3% each. This led to a detection of alleles with a frequency as low as 2%. The ploidy level was not set to values above 12, in favors of not introducing sequencing errors and other false positives as true variants. Moreover, variant calling under an elevated ploidy level is very time consuming and requires vast amounts of computing power, as the number of possible combinations to calculate increases with increasing ploidy.

Variant calling was performed in GVCF-mode for F1-individuals and the parental lines of the population GF.GA-47-42 x 'Villard Blanc'. GVCF stands for "genomic VCF" and differs from a regular VCF by having storing records for all sites. This serves the possibility to perform joint analysis in a cohort in subsequent steps. The records in a GVCF file include an accurate estimation of how confidently the sites are determined to be homozygous-reference or not. This estimation is

generated by the HaplotypeCaller's built-in reference model. Non-variant sites are grouped together into non-variant block records that represent intervals of sites for which the genotype quality (GQ) is within a certain range or band. The GQ ranges are defined in the GVCF file's header. The purpose of the blocks is to keep file size down. Otherwise, HaplotypeCaller was run with default parameters.

After variant calling, resulting variant files within the population GF.GA-47-42 x 'Villard Blanc' were merged using GATK's tool GenotypeGVCFs in order to apply further downstream steps on all samples together. At each position of the input GVCF-file, this tool combines all spanning records and output them to a new variant file.

Raw variants were hard-filtered according to GATK's "Best Practices" recommendations (DePristo *et al.* 2011; Van der Auwera *et al.* 2013). In addition, variants with read coverage depth and genotype quality below 20 were filtered out.

4.8.3 Physical phasing of candidate gene alleles

For the determination of allele specific sequences, physical phasing was performed using HapCUT (Bansal *et al.* 2008). The phasing of alleles was performed on the basis of sites polymorphic between the two alleles of a gene. As a first step, fragments were defined from the sequenced reads. A fragment is a mapped sequence read that has the non-polymorphic bases removed. After this, haplotype-informative reads that cover at least two heterozygous variants were extracted from the alignment file using the tool extractHairs from HapCUT and used for the assembly of haplotypes. The information of polymorphic sites was passed to HapCUT through a variants (VCF) file. Fragments containing two or more heterozygous variants contain valuable phase information as they link together variations of the same allele and define a potential phasing. Variants for which the individual is homozygous are not useful for assembly because a fragment containing either allele cannot uniquely identify which haplotype the allele was sampled from. These fragments were, thus, discarded. Furthermore, fragments containing none or one heterozygous variant are not useful for the assembly of haplotypes either and are also discarded (Aguiar *et al.* 2012; Bansal *et al.* 2008). HapCUT was then run with a maximum number of 600 iterations and the reference sequence was provided in order to extract reads covering both SNPs and InDels. HapCUT computes maximum cuts on a graph modeled from the fragment matrix to iteratively improve their phasing solution (Bansal *et al.* 2008) and initially delivers the information of variants that are linked together and thus phases alleles within one sample at heterozygous sites. Using various python scripts, intervals in which phasing was possible in individuals of the population GF.GA-47-42 x 'Villard Blanc' including the parents and F1-individuals

were determined and homozygous alternate variants were added to the variant files.

4.9 Determination of allele combinations in F1-individuals

Using GATK's FastaAlternateReferenceMaker FASTA-format files with alternative sequences were created for each individual within the genomic regions in which allele phasing was performed. A modified VCF-file with the result of allele phasing was passed to the tool. The created allele sequences of the progeny of the mapping population GF.GA-47-42 x 'Villard Blanc' were compared against the allele sequences of the parental lines using blastn (Camacho *et al.* 2009) with default parameters. Blast hits with at least 99% identity were tracked. Each of the progeny's alleles is supposed to be identical to one of the parental alleles. Thus, it was possible to determine inheritance patterns within the population for each gene.

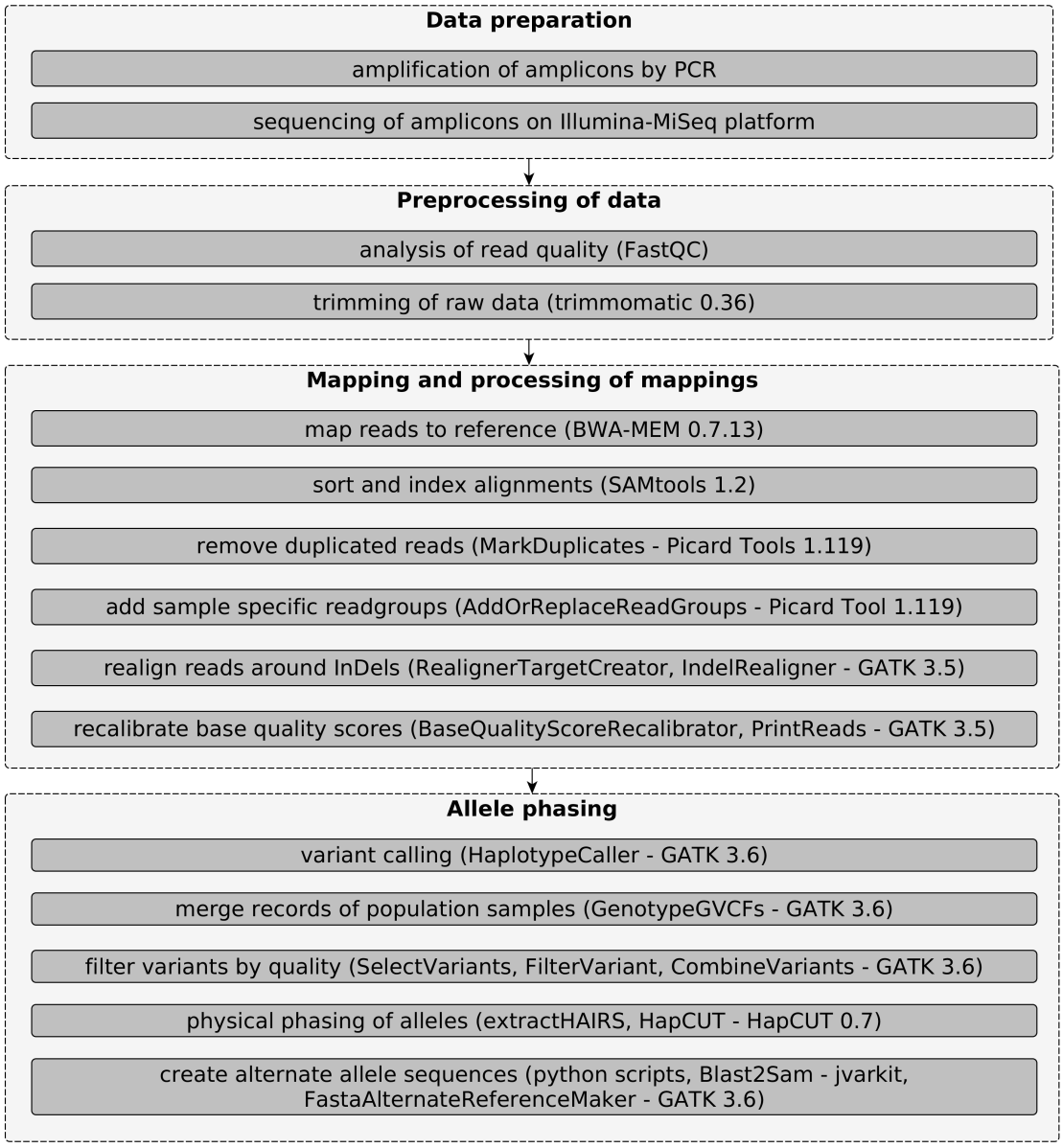

Figure 6: The established workflow for the phasing of alleles of amplicon sequenced genes.

4.10 Association analysis

In order to find alleles correlating with the phenotype of flowering time, a correlation analysis between the identified alleles of FTC candidate genes and early, intermediate or late phenotypes of flowering time was performed.

To test for the correlation of an allele and the flowering time phenotype, a Wilcoxon Rank-Sum test (Wilcoxon 1945) between a dichotomous variable (the presence or absence of an allele) and a continuous variable (flowering time) was carried out. For the continuous variable flowering time, the following data sets of phenotypic data were used:

1. days after January 1st for the years 1999 and 2009 - 2016,
2. accumulated temperature above 3 °C from November 1st of the previous year for the years 2011 - 2016 values, and
3. global radiation in KWh/m^2 from November 1st for the years 2011 - 2016.

The values for global radiation and accumulated temperature refer to the location of the vineyard in Siebeldingen, Germany (N49°21.675, E8°04.433) and were obtained from the DLR[8].

The null hypothesis assumed that the median of flowering time between groups of individuals carrying or not carrying a certain allele is equal. P-values for the correlation of 0.05 or below were considered to be significant. The null hypothesis was rejected and an association between an allele and the flowering time phenotype was found to exist when the p-value of correlation was equal or below 0.05.

4.11 Analysis of flowering time control candidate genes in important grapevine cultivars

The processing of data of the grapevine cultivars and breeding lines not beloning to the mapping population GF.GA-47-42 x 'Villard Blanc' was carried out in the same manner as within the population (Fig. 6), with the exception that variant calling was not performed in GVCF-but in default-mode since they do not belong to an F1-population. In order to compare allele sequences of unrelated individuals with those of the population GF.GA-47-42 x 'Villard Blanc', sequences were compared within the same intervals. Blast hits with an identity above 99% were considered to be identical.

4.12 Marker development

In cooperation with the Institute for Grapevine Breeding (JKI, Siebeldingen), carried out by Dr. Iris Ochßner, molecular markers that distribute between the maximum number of different alleles, were designed from the specified allele

[8]http://www.wetter.rlp.de

sequences for allele specific PCRs. Allele sequences of genes were scanned for InDel structures differing between the parental alleles. The designed markers were used to determine the allele distributions over all 151 heterozygous F1-individuals of the mapping population GF.GA-47-42 x 'Villard Blanc'. STS marker design and PCRs were carried out as described in Fechter *et al.* (2012).

4.13 Differential gene expression analysis

For the investigation of gene expression of FTC candidate genes, RNA-Seq experiments were performed. Plant material from buds and inflorescences from early flowering GF.GA-47-42 and late flowering 'Villard Blanc' was collected at several consecutive time points within the developmental cycle. The reproductive developmental cycle of grapevine is completed over two consecutive years beginning in the spring of the first year with bud development and ending in the summer of the second year with berry ripening and vintage (Fig. 1). Samples were collected over consecutive time points starting from latent winter buds until shortly before full bloom of the developmental cycle that was completed over the two consecutive years 2012 and 2013. Moreover, samples of consecutive time points of developing buds before dormancy in winter of the developmental cycle completed over 2013/2014 were collected. Table 4 shows an overview of the collected samples.

Samples (one replicate) were collected from four and three different vines of GF.GA-47-42 and 'Villard Blanc', respectively. The vines are located at the vineyards of Geilweilerhof, Germany (N49°21.675, E8°04.433). Sample collection was performed by Dr. Iris Ochßner and Anna Werner (JKI, Siebeldingen).

The phenological development of grapevine is divided into macro- and micro-stages in accordance with the BBCH-scale for the description of developmental stages of various cultivated plants (Hack *et al.* 1992).The BBCH-scheme divides the season into principal growth stages (macro stages) used to describe the developmental periods (coded 0-9), which is again each divided into secondary (micro) growth stages. Each of the stages is coded by a two digit code. However, macro stages 2, 3, and 4 are omitted from the BBCH-scale of grapevine since these stages are irrelevant in viticulture (Eichorn *et al.* 1977; Lorenz *et al.* 1995). Sample collection was performed according to BBCH-developmental stages insofar possible (Table 4).

Table 4: Samples collected from each of GF.GA-47-42 and 'Villard Blanc' for differential gene expression analysis, their developmental stage and the corresponding BBCH-stage.

Date of sample collection	**Developmental state**	**BBCH-stage**
Developmental cycle 1:		
20. December 2012	dormant buds	BBCH 0
8. March 2013	dormant buds	BBCH 0
22. March 2013	swelling buds	BBCH 0-5
12. April 2013	swelling buds	BBCH 5-9
26. April 2013	swelling buds/ first leaves	BBCH 11
3. May 2013	buds/ first leaves	BBCH 11-13
7. June 2013	inflorescences & leaves	BBCH 53
14. June 2013	inflorescences & leaves	BBCH 55
17. June 2013	inflorescences	BBCH 57
Developmental cycle 2:		
22. July 2013	buds & leaves	NA
2. August 2013	buds	NA
8. August 2013	buds & leaves	NA
16. August 2013	buds	NA
22. August 2013	buds & leaves	NA
5. September 2013	buds	NA
19. September 2013	leaves	NA

4.13.1 RNA extraction and sequencing

Total RNA was extracted from up to 100 mg of nitrogen grounded tissue using the Spectrum^{TM} Plant Total RNA kit (Sigma-Aldrich) according to the manufacturer's instructions for protocol B. After on-column DNase treatment with the DNase I Digest Set, the RNA was quantified. RNA-Seq (1 x 135 bp) was performed with TrueSeq technology on a HiSeq-1500 by using the Illumina RNA-Sequencing Kit complemented with reverse transcriptase according to the instructions of the manufacturer. One barcoded library was created for each of the time points and genotypes (Prisca Viehöver, CeBiTec, Bielefeld University).

4.13.2 RNA-Seq data processing

Using Trimmomatic (Bolger *et al.* 2014) quality filtering of reads with a window of four consecutive bases with a quality value below 30 was performed. Bases at the reads heads and tails with quality values below 30 were cropped. For

quality-check of reads before and after trimming the tool FastQC[9] was used. The reads were mapped to the grapevine reference sequence PN40024-12xv2 using tophat2 (Kim *et al.* 2013) which is capable of performing split read mapping and is thus intron-exon structure sensible. The intron size was set to 3,000, otherwise default parameters were used. Resulting BAM-format files were sorted and indexed using SAMtools (Li *et al.* 2009b).

With HTSeq (Anders *et al.* 2015) mapped reads were counted for each gene. Differential gene expression was analysed using the R-package DESeq2 (Love *et al.* 2014). As no biological replicates were available, an 'out-of-the-box' analysis was not possible. Therefore, the counts were modeled as a smooth function of time, and an interaction term of the condition with the smooth function was included. DESeq2's likelihood ratio test with a reduced design which does not include the interaction term was then applied. In order to test for cultivar specific effects a likelihood ratio test was performed, where the cultivar-specific differences over time were removed. This results in small p-values for genes that at one or more time point are differentially expressed between the cultivars. Genes that moved up or down over time in the same way in both cultivars are not given small p-values in this test.

[9]http://www.bioinformatics.babraham.ac.uk/projects/fastqc

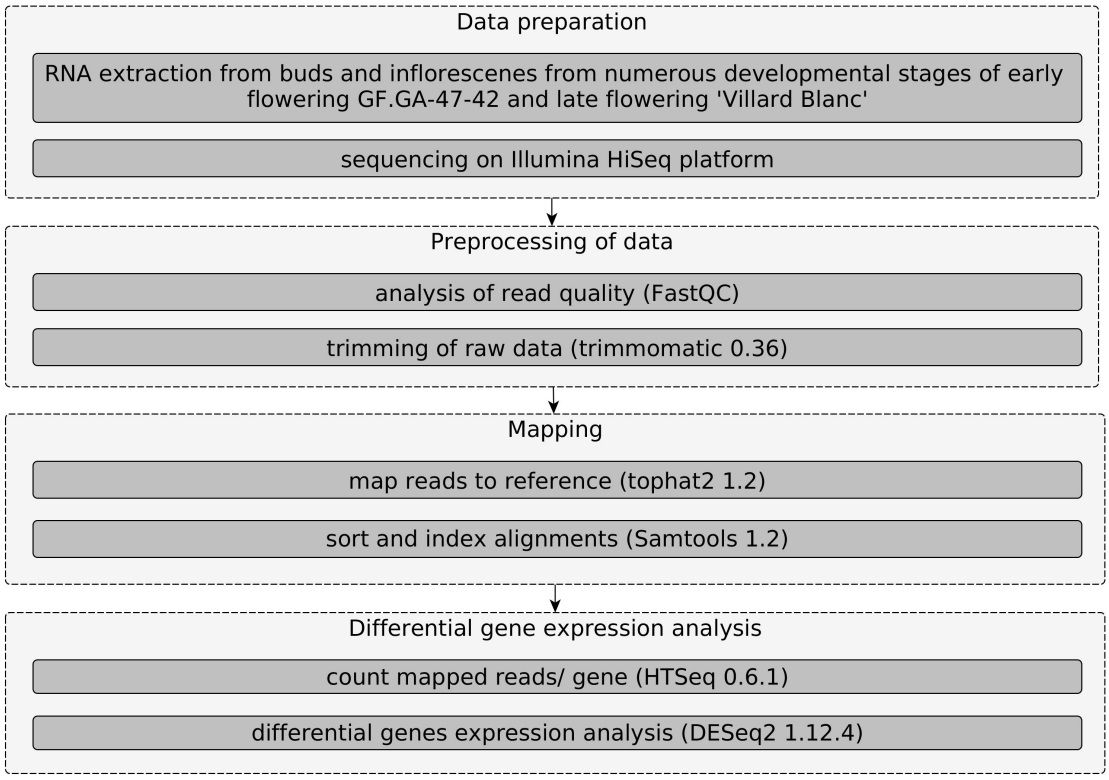

Figure 7: Workflow for the analysis of differential gene expression throughout consecutive developmental stages of buds, inflorescences, and leaves of early flowering GF.GA-47-42 and late flowering 'Villard Blanc'.

5 Results

5.1 Identification of flowering time control candidate genes in grapevine

More than four hundred candidate genes controlling or influencing flowering time in *Vitis* spread over all 19 chromosomes were identified. These genes were used as a basis for the analysis of the genetic variation of FTC genes in grapevine within this thesis. The highest number of candidate genes was found on chromosome 1 and 4 (34), followed by chromosome 14 (31) which posses previously identified QTL-regions for FTC. The chromosome with the smallest amount of candidate genes was chromosome 9. Ten genes were located on unordered contigs of chromosomes 7, 10, 13, and 18. Several *Vitis* genes had more than one (up to eight) orthologues in *A. thaliana*.

Candidate gene prediction was initially performed using the grapevine reference PN40024-12xv0. In order to determine the positions of the candidate genes in version 12xv2, the sequences were compared against PN40024-12xv2 by blast. From the selected genes, more than three hundred had an RBH in *Vitis* while only 24 genes had an RBH in the *A. thaliana* genome. Some genes had two or more hits such as *VvCOL10a*. In total, 57 genes had two hits, ten genes three hits, four genes four hits, and one gene with five and one with six hits were found. Ten genes were not found in the new assembly, among them *VvFT*. *VvFT* is, however, also not annotated in the 12xv0 assembly but only in the PN40024-8x assembly.

Many of the FTC candidate genes are transcription factors involved in flower development and morphogenesis such as members of the AP2/EREBP family (Riechmann *et al.* 1998) and homeodomain proteins (Gehring 1992). About eight MYB-transcription factors that participate in cell cycle control in many living taxa (Stracke *et al.* 2001) were among the identified FTC candidate genes in *Vitis*. More than 40 serum response factor (SRF) proteins, that play a role in cell cycle regulation, apoptosis, cell growth, and cell differentiation (Norman *et al.* 1988) were identified.

Several other protein families were among the FTC candidate genes, such as:

- TB1, CYC, and PCF (TCP) transcription factors that are involved in multiple developmental control pathways (Cubas *et al.* 1999) of which four were identified.

- bzip transcription factors (Jakoby *et al.* 2002) of which ten were among the FTC genes in *Vitis*.

- around a dozen GRAS and FRIGIDA proteins that are involved in flowering time and plant development. FRIGIDA proteins are required for the regulation of flowering time by upregulating *FLC* expression. Allelic variation at the FRIGIDA locus is an important determinant of natural variation in the timing of flowering (Johanson *et al.* 2000). The GRAS (GAI, RGA, SCR) family is a very important family of proteins involved in flowering in grapevine. GRAS proteins participate in GA signaling, which influences numerous aspects of plant growth and development (Pysh *et al.* 1999).

- Knotted1-like homeobox (KNOX) genes, that encode transcription factors required for meristem maintenance and patterning of organ initiation (Hake *et al.* 2004), of which seven were found.

- eight basic-helix-loop-helix (bHLH) transcription factors that participate in a diversity of development processes and cell activity (Murre *et al.* 1994).

- five DNA-binding-with-One-Finger (DOF) transcription factors that are involved in many fundamental processes in higher plants, such as seed development and responses to light and phytohormones as well as in seed maturation and germination (Bosu *et al.* 2008).

- sixteen SQUAMOSA PROMOTER BINDING PROTEIN (SBP)-domain proteins, that are transcriptional activators involved in a variety of processes such as flower and fruit development, plant architecture, GA signaling, and the control of early flower development (Chen *et al.* 2010).

- ten JmjC-domain proteins, that are protein hydroxylases catalysing a novel histone modification and belong to the cupin superfamily (Clissold *et al.* 2001).

- more than 25 2-oxoglutarate-dependent dioxygenase superfamily proteins (2OGDs) that are the second largest family of plant enzymes, involved in oxygenation/hydroylation reactions and were found to play a role in low temperature response in grapevine (Kim *et al.* 2016) were among the candidate genes.

- about five members of the cullin family were found, which are a family of hydrophobic proteins involved in many cellular processes by providing a scaffold for ubiquitin ligases (Bosu *et al.* 2008).
- and other identified domains, such as the Per-Arnt-Sim (PAS) domain which is a key structural motif involved in protein-protein interactions of the circadian clock and found in all kingdoms of life (Ponting *et al.* 1997) and the K-Homology (KH)-domain, that is the second most prevalent RNA binding motif in plants (García-Mayoral *et al.* 2007).

Eleven of the genes were found to be uncharacterized in *Vitis*.

5.2 Association analysis of flowering time control-candidate genes and alleles with the timing of flowering

In order to identify alleles of previously defined FTC candidate genes involved in the control of flowering time, a genetic association study between alleles of several of these genes and flowering time phenotypes was performed. A candidate gene approach including amplicon sequencing in the pseudo F1-mapping population GF.GA-47-42 x 'Villard Blanc' was established.

5.2.1 Primer design and amplimer amplification

In total, 72 FTC candidate genes were chosen for amplicon sequencing (Table A1), many of which are located in flowering related QTL-regions on chromosome 1, 14, and 17. Amplicon sequencing of candidate genes was performed in 48 different cultivars, 37 of which belong to the pseudo F1-mapping population GF.GA-47-42 x 'Villard Blanc' (appendix table 3).

Primer pairs for 87 amplimers covering 72 genes were designed. Eventually, it was possible to amplify all 87 designed amplimers. Nine genes were covered by two amplimers, two genes by three amplimers, one gene by four amplimers, and the rest by one amplimer.

Some amplicons failed PCR amplification. About 15% of amplicons could not be amplified or had a sequencing depth below 20 (appendix tables A5 and A6). Ten genes could hardly be amplified in F1-individuals of the mapping population GF.GA-47-42 x 'Villard Blanc', many of which are located on chromosome 14. Individuals of the mapping population GF.GA-47-42 x 'Villard Blanc', for which

the amplification of these genes succeeded, namely the F1-individuals 89-30-257, 89-30-349, 89-30-198, 89-30-575, 89-30-328, 89-30-578, 89-30-632 as well as GF.GA-47-42 and 'Villard Blanc', were amplified and sequenced prior to and in a different library than the rest of the population. Therefore, the absence of the sequence information of the other lines might be due to technical issues. The amplification of amplimers in 'Pinot Noir' and 'Pinot Noir précoce' was performed for only 34 genes.

5.2.2 Amplicon sequencing and read processing

Raw reads were trimmed to improve sequence quality. Figure 8 shows the average quality of reads before and after trimming. Between 11.5 and 35.6% (20.2% on average [standard deviation (SD): 5.5%]) of reads were dropped through trimming.

Amplicon sequencing was performed in seven different runs (appendix table A7). Samples in all runs passed the sequence quality score criterion meaning that all sequences had universally satisfying quality values. All samples passed the per base N-content-report after trimming. An N is substituted for a conventional base when the sequencer is unable to make a base call. GC-content, Kmer-content, and the amount of overrepresented sequences remained noisy even after trimming, especially in runs 1 - 4. This is however common when sequencing low diversity libraries where only a limited number of genes is sequenced.

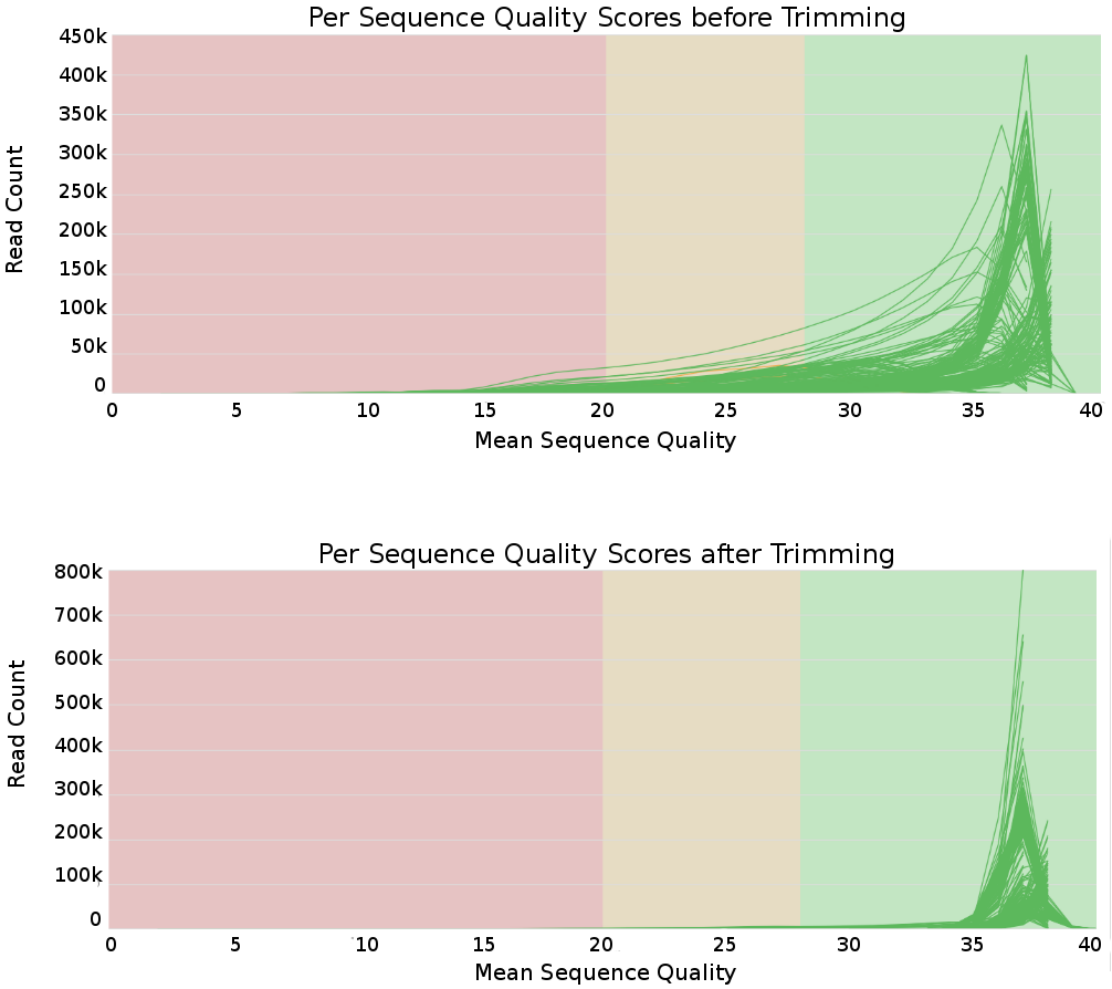

Figure 8: Mean sequence quality (phred score) of sequencing reads from all seven runs before and after quality trimming of reads. Quality analysis was performed with FastQC, image generated with MultiQC (Ewels *et al.* 2016) and modified. k: thousand.

5.2.3 Mapping and variant calling

5.2.3.1 Mapping

The amount of mapped reads was distributed between 45.7 and 86.5% across samples. On average 71.1% of reads were mapped (SD: 10.6%) (appendix table A8). After mapping of trimmed reads, the mappings were processed by several downstream steps. Typically, PCR duplicates arise in next generation sequencing. During library amplification prior to sequencing, multiple copies of each DNA molecule are created by PCR. When two copies of the same DNA molecule get

onto different beads or different primer lawns in a flowcell, PCR-duplicates occur. Optical duplicates, on the other hand, are single clusters that are called as two by the real time analysis software. While calling variations in heterozygous samples, duplicated reads are troublesome as they can introduce a bias towards one of the alleles when spanning variations. Therefore, duplicates were removed. The amount of duplicated reads among mapped reads was distributed between 6.2 and 43.9% across samples with an average of 20.4% (SD: 11.1%). Unpaired reads were also removed from the mapping files to prevent a bias in variant calling. Between 0.15 and 0.72% of reads were unpaired (average: 0.44%, SD: 0.16%).

The read depth (or depth of coverage) of mappings was distributed between 1 and 2,187 across samples. The average read depth was 286 (SD: 276). Figure 9 shows the average read depth for each of the genes in the analyzed lines. The F1-individuals 89-30-508 and 89-30-257 and the breeding line GF.GA-52-42 shows very high read depth values. Very high depth of coverage in read mapping correlates with a high sequencing output and is likely due to a high concentration of template DNA molecules. Several values are missing since amplicons were not amplified (See section 5.2.1). However, most samples had a mapping coverage between 100 and 300 (Figure 10).

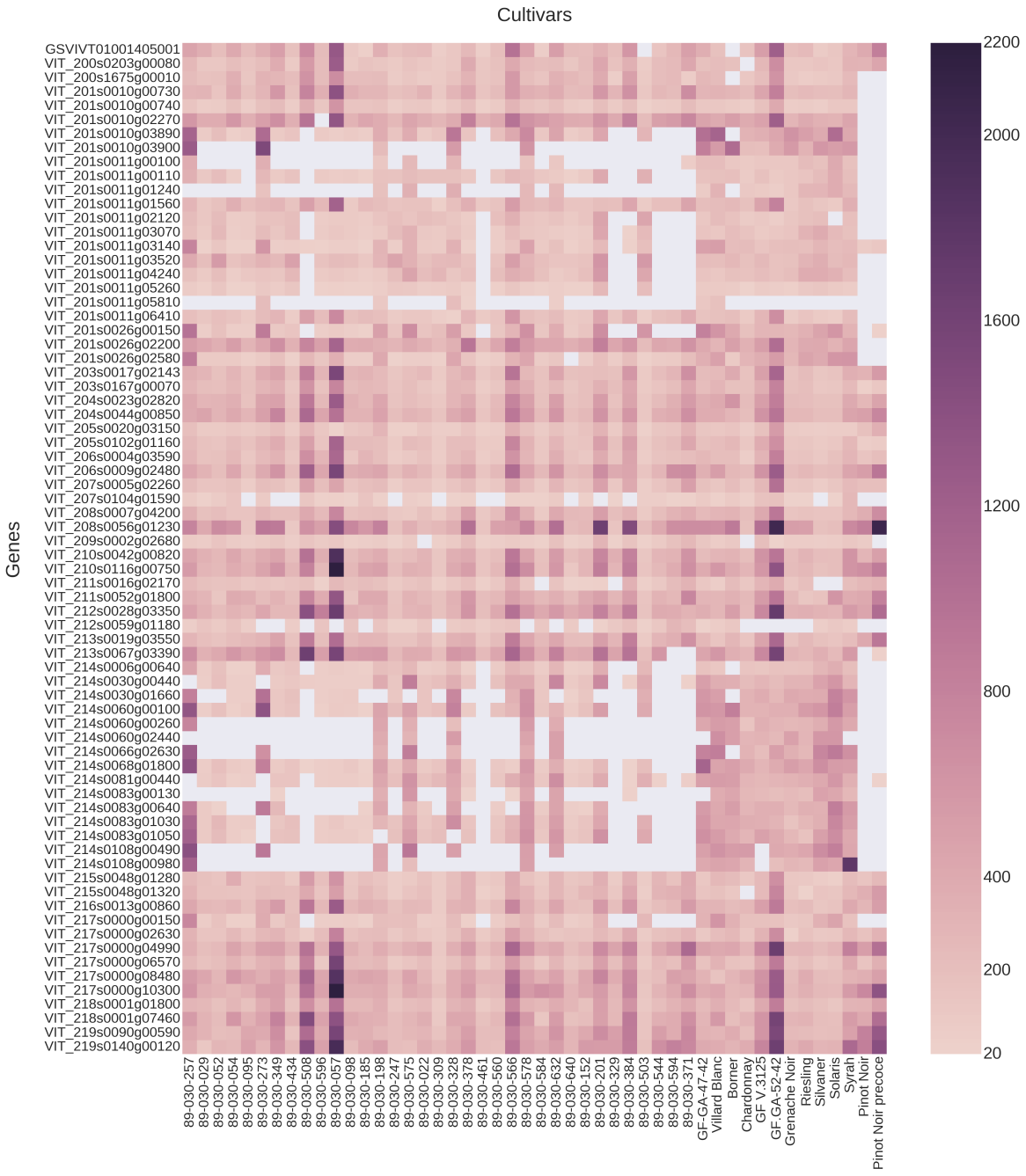

Figure 9: Heatmap of the average read depth of sequenced genes for each of the amplicon sequenced cultivars. Grey areas have average depth of coverage values below 20. Scale on the right side: depth of coverage.

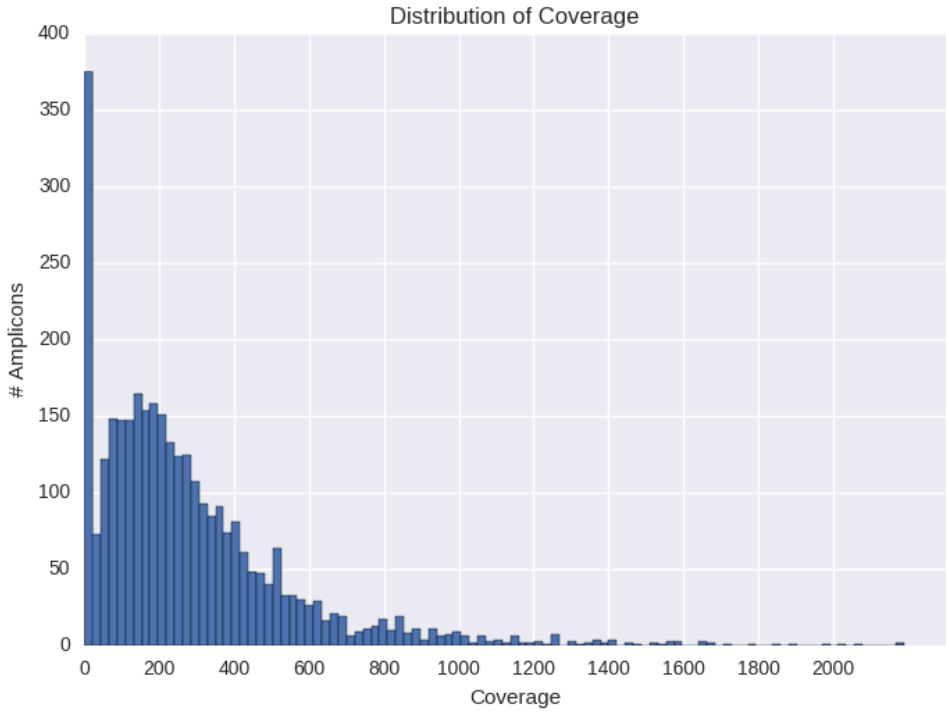

Figure 10: Histogram presenting depth of coverage values for the sequenced amplicons showing the number of amplicons with different depth of coverage values in read mappings. Most amplicons had a read depth around 100 - 300 in read mappings. Binsize = 50.

5.2.3.2 Variant Calling

To obtain information of polymorphic sites, variations in the amplicon sequenced data were called. Variants were called as described in section 4.8.2. After filtering, the density of variations was distributed between 0.79/100 bp ('Chardonnay') and 2/100 bp ('Börner'). SNPs were found with a density ranging between 0.74 and 1.87/100 bp, while InDels had a density between 0.05 and 0.13/100 bp. 'Börner' was found to have the highest amount of variations of all analyzed lines, while 'Chardonnay' had the smallest, followed by 'Silvaner', GF V.3125, 'Syrah', 'Pinot Noir précoce', 'Riesling', and 'Pinot Noir' (Fig. 11).

The density of homozygous variants was found to be ranging between 0.24/100 bp and 0.91/100 bp, while the density of heterozygous variants ranged between 0.54 and 1.31/100 bp (Fig. 11). The average length of sequenced amplicons was ranged between 166,274 bp and 469,519 bp across individuals with an average of 340,631 bp.

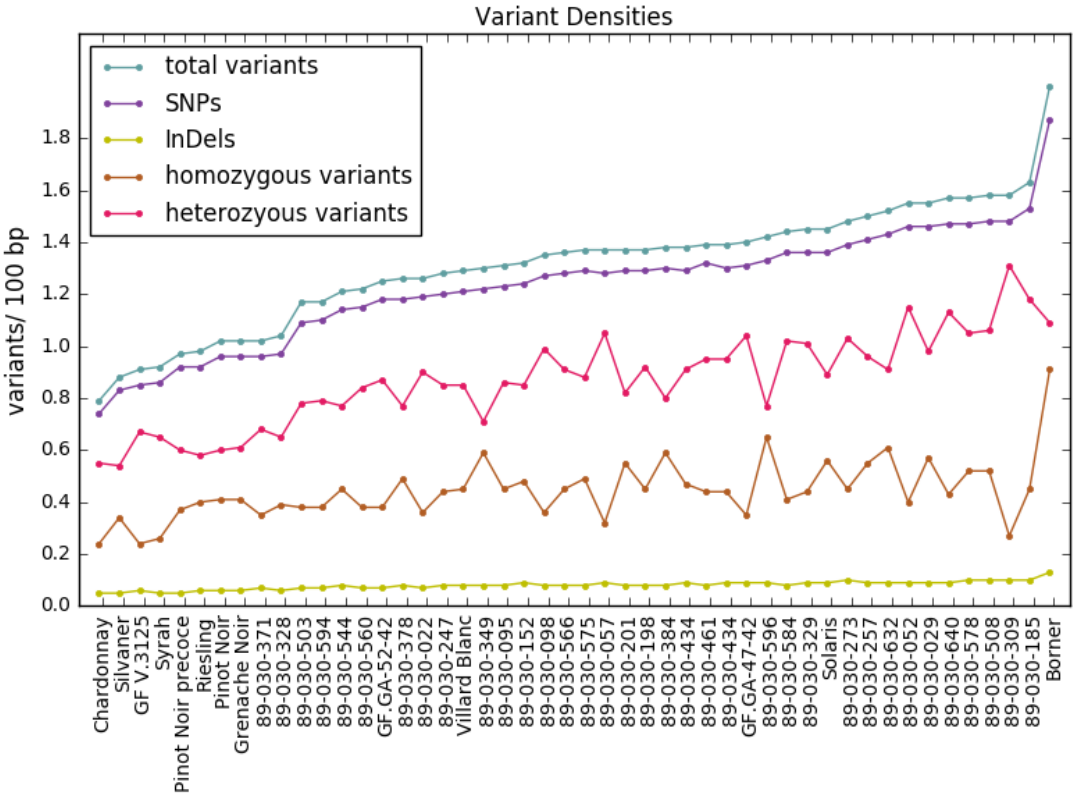

Figure 11: Average variant density in 100 bp in all analyzed lines in sequenced FTC candidate genes. Lines with the identifiers 89-030-XXX are F1-individuals of the population GF.GA-47-42 x 'Villard Blanc'.

Table 5: Density of variants (SNPs & InDels) and each of SNPs and InDels in 100 bp in the amplicon sequenced cultivars and the length of amplicon-sequenced sequences. D: density, homs: homozygous-, hets: heterozygous variants.

Cultivar	**Length**	**D-variants**	**D-SNPs**	**D-InDels**	**D-homs**	**D-hets**
'Chardonnay'	443,271	0.79	0.74	0.05	0.24	0.55
'Silvaner'	448,794	0.88	0.83	0.05	0.34	0.54
GF V.3125	430,490	0.91	0.85	0.06	0.24	0.67
'Syrah'	459089	0.92	0.86	0.05	0.26	0.65
'Pinot Noir précoce'	173,789	0.97	0.92	0.05	0.37	0.60
'Riesling'	449,853	0.98	0.92	0.06	0.40	0.58
'Pinot Noir'	166,274	1.02	0.96	0.06	0.41	0.60
89-030-371	448,870	1.02	0.96	0.06	0.41	0.61
'Grenache Noir'	469,519	1.02	0.96	0.07	0.35	0.68
89-030-328	424,662	1.04	0.97	0.06	0.39	0.65
89-030-503	453,568	1.17	1.09	0.07	0.38	0.78
89-030-594	323,695	1.17	1.10	0.07	0.38	0.79
89-030-544	431,526	1.21	1.14	0.08	0.45	0.77
89-030-560	30,4842	1.22	1.15	0.07	0.38	0.84
GF.GA-52-42	433,558	1.25	1.18	0.07	0.38	0.87
89-030-378	315,039	1.26	1.18	0.08	0.49	0.77
89-030-022	391,063	1.26	1.19	0.07	0.36	0.90
89-030-247	328,479	1.28	1.20	0.08	0.44	0.85
'Villard Blanc'	460,834	1.29	1.21	0.08	0.45	0.85
89-030-349	316,522	1.30	1.22	0.08	0.59	0.71
89-030-095	376,766	1.31	1.23	0.08	0.45	0.86
89-030-152	322,664	1.32	1.24	0.09	0.48	0.85
89-030-098	317,866	1.35	1.27	0.08	0.36	0.99
89-030-566	311938	1.36	1.28	0.08	0.45	0.91
89-030-057	315,320	1.37	1.29	0.08	0.49	0.88
89-030-575	313,911	1.37	1.28	0.09	0.32	1.05
89-030-201	324,487	1.37	1.29	0.08	0.55	0.82
89-030-198	435,019	1.37	1.29	0.08	0.45	0.92
89-030-434	328,587	1.38	1.30	0.08	0.59	0.80
89-030-384	205,329	1.38	1.29	0.09	0.47	0.91
89-030-434	321,215	1.39	1.32	0.08	0.44	0.95
GF.GA-47-42	456,115	1.39	1.30	0.09	0.44	0.95
89-030-461	322,914	1.40	1.31	0.09	0.35	1.04
89-030-596	218,271	1.42	1.33	0.09	0.65	0.77
89-030-584	212,301	1.44	1.36	0.08	0.41	1.02
89-030-329	322,869	1.45	1.36	0.09	0.44	1.01
'Solaris'	448,167	1.45	1.36	0.09	0.56	0.89
89-030-273	313,237	1.48	1.39	0.10	0.45	1.03
89-030-257	321,132	1.50	1.41	0.09	0.55	0.96
89-030-632	206,119	1.52	1.43	0.09	0.61	0.91
89-030-052	326,916	1.55	1.46	0.09	0.40	1.15
89-030-029	311,674	1.55	1.46	0.09	0.57	0.98
89-030-640	209,116	1.57	1.47	0.09	0.43	1.13
89-030-578	200,292	1.57	1.47	0.10	0.52	1.05
89-030-508	318,250	1.58	1.48	0.10	0.52	1.06
89-030-309	320,252	1.58	1.48	0.10	0.27	1.31
89-030-185	212,397	1.63	1.53	0.10	0.45	1.18
'Börner'	383,444	2	1.87	0.13	0.91	1.09

5.2.4 Allele Phasing

The specificity of a trait influenced by a certain gene is often correlated with the combination of both alleles of the gene. In order to link certain alleles of the sequenced fragments to flowering time phenotypes, the two alleles of heterozygous genes had to be reconstructed using a mix of sequenced fragments from the two alleles. The outcome of variant calling was used for the attempt to reconstruct the alleles of each gene in every analyzed individual.

After defining intervals, alternative allele sequences were created. For genomic regions from 46 genes spread over 16 chromosomes, alleles were phased (Table 6). In 23 cases four different alleles were identified, three alleles in 18 cases, and two alleles in four cases. In one case only one allele was found meaning that all individuals of the population were homozygous for the respective locus. Phased regions of six genes were found to be homozygous in GF.GA-47-42 and three in 'Villard Blanc'. For the gene *VvSPL1b* (VIT_207s0005g0226) all F1-individuals were found to carry the L2-allele from 'Villard Blanc' and none was carrying the L1-allele. The lengths of the phased intervals was between 204 and 8,285 bp (Table 7).

Table 6: Number of genes with each one, two, three, or four alleles within the phased regions.

# Allele sequences	**# Genes**
four	23
three	18
two	4
one	1
total	46

Table 7: Genes for which allele sequences were distinguished within specified intervals, positions, and lengths of amplimers and phased regions in bp. For three genes (VIT_214s0081g00440, VIT_218s0001g01800, and VIT_218s0001g07460) several non-overlapping amplimers were designed that are listed separately.

Gene-ID	Amplimer length	Amplimer position	Interval length	Interval position
VIT_201s0011g00100	13,573	Chr1:149474..163047	3,854	Chr1:15091..154444
VIT_201s0011g00110	4,913	Chr1:193,108..198,021	4,438	Chr1:202,719..207,156
VIT_201s0011g01560	2,863	Chr1:1,353,206..1,356,069	2,825	Chr1:1,353,217..1,356,041
VIT_201s0011g02120	12,388	Chr1:1,796,990..1,809,378	5,958	Chr1:1,797,008..1,802,965
VIT_201s0011g03070	3,097	Chr1:2,750,739..2,753,836	2,993	Chr1:2,750,769..2,753,761
VIT_201s0011g03140	6,523	Chr1:2,832,803..2,839,326	5,080	Chr1:2,834,226..2,839,305
VIT_201s0011g04240	6,590	Chr1:3,848,721..3,855,311	3,581	Chr1:3,850,488..3,854,068
VIT_201s0011g05260	4,572	Chr1:4,894,044..4,898,616	1,481	Chr1:4,897,038..4,898,518
VIT_201s0011g06410	3,812	Chr1:6,230,731..6,226,919	1,186	Chr1:6,227,517..6,228,702
VIT_201s0026g00150	5,166	Chr1:8,786,973..8,792,139	3,194	Chr1:8,788,809..8,792,002
VIT_201s0026g02200	926	Chr1:11,611,459..11,610,533	367	Chr1:11,610,653..11,611,019
VIT_201s0026g02580	3,515	Chr1:12,212,051..12,215,566	4,937	Chr1:12,211,396..12,216,332
VIT_201s0010g00740	3,234	Chr1:15,907,173..15,903,939	3,080	Chr1:15,904,052..15,907,131
VIT_201s0010g02270	1,309	Chr1:19,231,131..19,232,440	1,233	Chr1:19,231,160..19,232,392
VIT_201s0010g03890	5,900	Chr1:21,257,213..21,263,113	2,607	Chr1:21,260,482..21,263,088
VIT_204s0023g02820	3,100	Chr4:19,394,945..19,398,045	1,600	Chr4:19,395,984..19,397,583
VIT_204s0044g00850	3,883	Chr4:21,994,322..21,998,205	491	Chr4:21,997,042..21,997,532
VIT_205s0102g01160	7,965	Chr5:23,296,885..23,304,850	2,591	Chr5:23,301,727..23,304,317
VIT_206s0004g03590	4,817	Chr6:4,494,149..4,498,966	2,943	Chr6:4,497,842..4,494,900
VIT_206s0009g02480	5,093	Chr6:15,015,975..15,021,068	4,958	Chr6:15,021,045..15,016,088
VIT_207s0005g02260	6,046	Chr7:4,629,108..4,635,154	1,865	Chr7:4,632,064..4,630,200
VIT_207s0104g01590	6,149	Chr7:2,560,922..2,567,071	895	Chr7:2,564,788..2,565,682
VIT_208s0007g04200	1,583	Chr8:18,187,348..18,188,931	204	Chr8:18,187,745..18,187,948
VIT_208s0056g01230	2,851	Chr8:1,939,911..1,937,060	1,469	Chr8:1,939,867..1,938,399
VIT_210s0116g00750	4,501	Chr10:344,218..339,717	1,012	Chr10:340,848..339,837
VIT_211s0052g01800	1,631	Chr11:19,620,833..19,622,464	1,520	Chr11:19,620,851..19,622,370
VIT_212s0028g03350	2,005	Chr12:4,110,705..4,108,700	1,626	Chr12:4,110,662..4,109,037
VIT_213s0019g03550	4,327	Chr13:4,719,797..4,724,124	3,963	Chr13:4,720,125..4,724,087
VIT_213s0067g03390	3,949	Chr13:1,862,156..1,866,105	2,338	Chr13:1,862,179..1,864,516
VIT_214s0006g00640	3,379	Chr14:14,805,916..14,809,295	3,265	Chr14:14,805,995..14,809,259
VIT_214s0030g00440	5,218	Chr14:4,293,121..4,298,339	1,043	Chr14:4,297,214..4,298,256
VIT_214s0068g01800	5,768	Chr14:25,525,842..25,531,610	3,662	Chr14:25,526,128..25,529,789
VIT_214s0081g00440	5,563 6,267	Chr14:8,304,570..8,310,133 Chr14:8,310,401..8,316,668	8,285	Chr14:8,304,600..8,312,884
VIT_214s0083g01030	6,044	Chr14:22,050,144..22,056,188	5,425	Chr14:23,324,870..23,319,446
VIT_214s0083g01050	6,100	Chr14:23,362,282..23,368,382	2,653	Chr14:23,373,019..23,370,367
VIT_215s0048g01280	6,434	Chr15:15,432,380..15,438,814	4,979	Chr15:15,433,754..15,438,732
VIT_216s0013g00860	13,078	Chr16:6,416,753..6,429,831	6,437	Chr16:6,416,816..6,423,252
VIT_217s0000g02630	2,311	Chr17:2,414,026..2,416,337	2,154	Chr17:2,416,209..2,414,056
VIT_217s0000g06570	7,955	Chr17:7,101,034..7,108,989	7,843	Chr17:7,101,119..7,108,961
VIT_217s0000g08480	1,733	Chr17:9,703,139..9,704,872	1,582	Chr17:9,704,843..9,703,262
VIT_217s0000g10300	2,165	Chr17:12,556,385..12,558,550	2,028	Chr17:12,558,521..12,556,494
VIT_218s0001g01800	4,385 5,340 5,041	Chr18:2,328,257..2,332,642 Chr18:2,317,315..2,322,655 Chr18:2,307,895..2,312,936	3,361	Chr18:2,320,702..2,317,342
VIT_218s0001g07460	7,672 4,401 5,312	Chr18:5694328..5702000 Chr18:5701976..5706377 Chr18:5706358..5711670	2,847	Chr18:5703037..5705883
VIT_200s0203g00080	1,988	ChrUn:11,888,203..11,890,191	588	ChrUn:11,888,548..11,889,135
VIT_200s1675g00010	3,994	Chr2:35343..39337	1,429	Chr2:37,862..39,290

In order to distinguish between two alleles, an appropriate amount of variants is needed within the span of a read-pair. The phase of the heterozygous genotypes can be determined since each read or pair of reads is obtained from a single haplotype. Read lengths after trimming were distributed between 80 and 300 bp with an average insert size of 500 bp. When variants were located farther apart than the maximum length that can be spanned by a read pair, they could not be linked together and were useless for phasing. Blocks upstream and downstream of regions lacking heterozygous variations could thus not be linked together. In such cases, the longest interval was chosen for further analyses (Fig. 12).

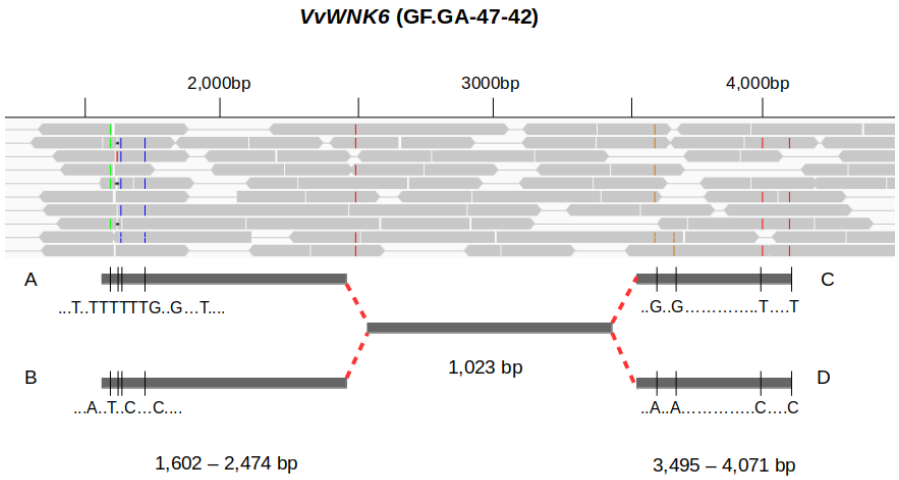

Figure 12: Illustration of allele phasing at the example of *VvWNK6* (VIT_204s0044g00850) in GF.GA-47-42. The phased blocks from 1,621 to 2,474 and 3,495 to 4,709 can not be linked together due to a predominantly homozygous region in between. It is hence uncertain whether the haplotype blocks A and C stem from the same haplotype while B and D stem from the other or vice versa.

HapCUTs output files contain blocks in which the phasing of alleles was performed at heterozygous sites. It is possible that such blocks differ in length between individuals of the mapping population due to the presence or absence of heterozygous sites. Homozygous variants are not useful for phasing but in order to compare the allele sequences of individuals of the mapping population, the heterozygous blocks were elongated by homozygous variants in both directions provided the homozygous variants were phased. Afterwards, the shortest interval of GF.GA-47-42 and 'Villard Blanc' was chosen. Within this interval alternative allele sequences for the parental lines and F1-individuals were created. When the region in which allele phasing was performed in an F1-individual was shorter than the interval,

the alternative allele sequence was created for the shorter interval. For example, *VvWNK6* amplified in GF.GA-47-42 had four phased intervals: 994-1,063 bp, 1,602-2,474 bp, 3,495-4,071 bp, and 4,615-4,709 bp, while 'Villard Blanc' had two: 1,621-4,071 bp and 4,679-4,709 bp (Fig. 13). Hence the longest overlapping phased interval was from 1,621 to 2,474 bp . In this special case however, the interval from 3,581 to 4,071 was chosen for further analysis since it was found to be harboring a SNP leading to a premature stop codon.

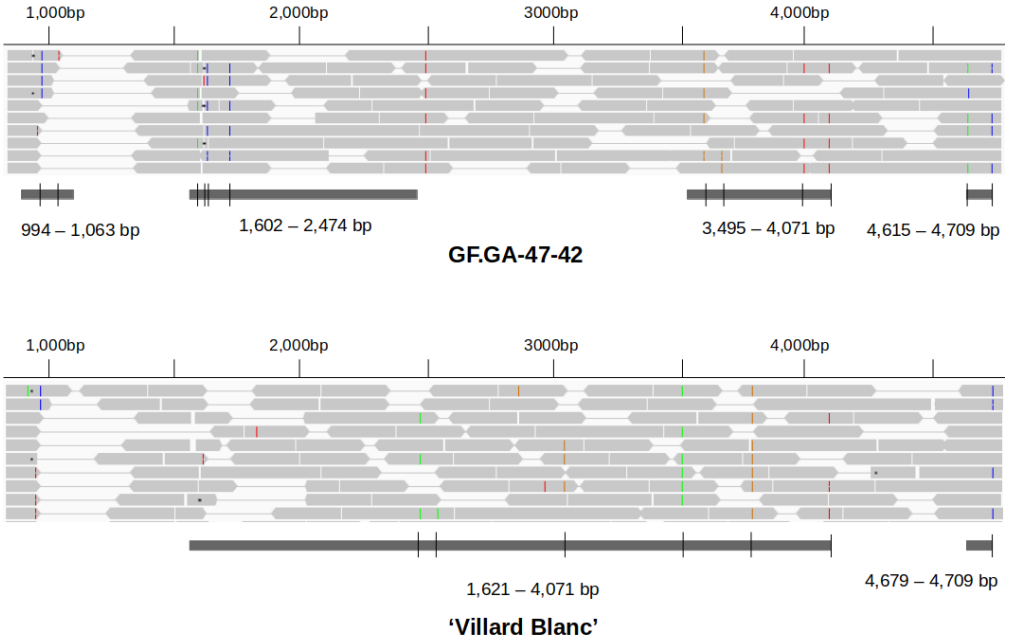

Figure 13: Illustration of phased intervals of both GF.GA-47-42 and 'Villard Blanc' in *VvWNK6*. The longest overlapping phased interval is from 1,621 to 2,474 bp.

An example of a gene for which four alleles were detected is *VvPFT1a* (VIT_201s0011g03140). The phased interval was located between 2,834,226 and 2,839,305 bp on chrosmosome 1 and had a length of 5,080 bp. Figure 14 shows a section of the gene in the parental lines and two of the F1-individuals that demonstrates the presence of four allele.

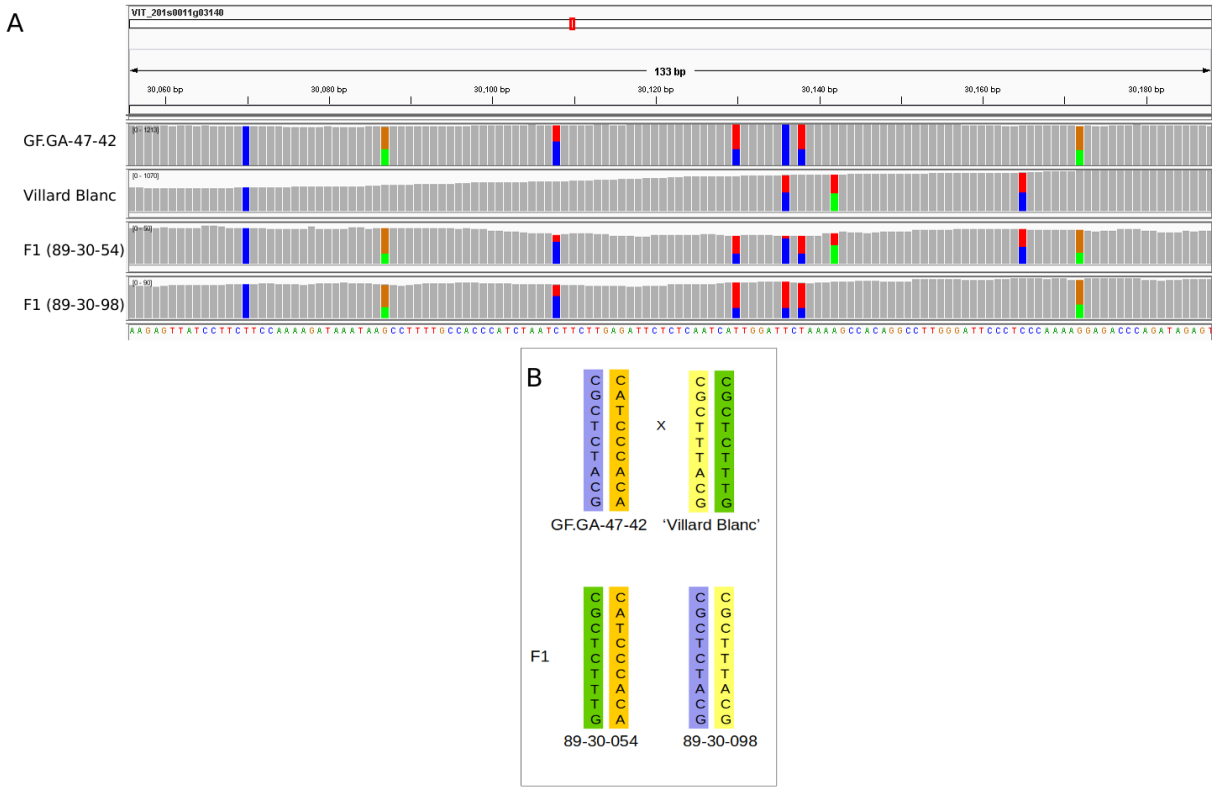

Figure 14: Illustration of the presence of four alleles for the gene VIT_201s0011g03140 (*VvPFT1a*) by the example of a section within the phased region of the gene. **A**: The upper two panels show the parental lines GF.GA-47-42 and 'Villard Blanc' and the lower panels show F1-individuals. GF.GA-47-42 is heteroyzgous for the sequences C-G-C-T-C-T-A-C-G and C-A-T-C-C-C-A-C-A, and 'Villard Blanc' has the sequences C-G-C-T-T-T-A-C-G and C-G-C-T-C-T-T-T-G. Both F1-individuals are heterozygous (C-G-C-T-C-T-T-T-G/ C-A-T-C-C-C-A-C-A and C-G-C-T-T-T-A-C-G/ C-A-T-C-C-C-A-C-A). Non-variant sites over all shown samples are ignored. The reference sequence is shown below. Image from IGV (Robinson *et al.* 2011; Thorvaldsdóttir *et al.* 2013). **B**: inheritage pattern of alleles within the shown region among the parental lines and two of the F1-individuals of the population GF.GA-47-42 x 'Villard Blanc'.

5.2.4.1 Nomenclature system

As a diploid organism grapevine has two sets of each chromosome and thus two alleles of each gene. Since the two alleles can be heterozygous, it is possible to detect four different alleles in an F1-population but it is also possible to detect less alleles. Depending on which alleles of the two parental lines are alike, this leads to a high number of different allele combinations in the population. Therefore, a nomenclature system for the description of alleles of genes within the population GF.GA-47-42 x 'Villard Blanc' was established (Table 8). Alleles inherited from GF.GA-47-42 were described as "E" referring to "early". "E1" and "E2" refer to both alleles in case they are heteroyzgous. "L" refers to late and originates from late flowering 'Villard Blanc'. Without analyzing the parental lines of GF.GA-47-42 and 'Villard Blanc', it is not possible to determine which of the GF.GA-47-42-alleles originates from which of its parents 'Bacchus' and 'Seyval'. Hence, E1- or E2-loci might originate from 'Bacchus' in some cases and from 'Seyval' in other cases. The same applies for 'Villard Blanc'.

When one cultivar was homozygous the alleles were both termed "0" ("E0/E0" for GF.GA-47-42 and "L0/L0" for 'Villard Blanc'). When all alleles were identical they were termed "N0". When GF.GA-47-42 and 'Villard Blanc' shared an allele, it was termed "N", which means that either L2 and E1 or E2 and L1 are alike. "N1" means that E1 and L1 are alike, while "N2" means that E2 and L2 are alike. "Na" means that E1, E2, and L1 are alike. "Nb" means that E1, E2, and L2 are alike. "Nc" means that E1, L1, and L2 are alike. "Nc" means that E2, L1, and L2 are alike (Table 8). Descriptions for allele combinations that distinguish between which of the two alleles of a parental line is identical to the two alleles of the other line (as in Na/Na x Na/L2), was implemented in order to be able to track patterns of allele combinations throughout QTL-regions and closely neighbored genes (Fig. 15).

Table 8: Nomenclature system for the description of allele combinations based on the population GF.GA-47-42 x 'Villard Blanc'. "E", as in "E1", "E2" and "E0", refers to early and originates from early flowering GF.GA-47-42, while "L", as in "L1", "L2" and "L0" refers to late and originates from late flowering 'Villard Blanc'. "N" means that both GF.GA-47-42 and 'Villard Blanc' share one or more alleles.

# Alleles	Similar Alleles	Alternative Description	Allele Names	Combinations in Offspring
4 alleles	none	/	E1 E2 L1 L2	E1/L1 E1/L2 E2/L1 E2/L1
3 alleles	E1 = E2	E0	E0 L1 L2	E0/L1 E0/L2
	L1 = L2	L0	E1 E2 L0	E1/L0 E2/L0
	E1 = L1	N1	E2 L2 N1	E2/N1 N1/L2 E2/L2
	E2 = L2	N2	E1 N2 L1	E1/N2 N2/L1 E1/L1
	E2 = L1	N	E2 N L1	E2/N N/L1 E2/L1
	E1 = L2	N	E1 N L2	E1/N N/L2 E1/L2
2 alleles	E1 = E2 = L1	Na	Na L2	Na/L2 Na/Na
	E1 = E2 = L2	Nb	Nb L1	Nb/L1 Nb/Nb
	E1 = L1 = L2	Nc	E2 Nc	E2/Nc Nc/Nc
	E2 = L1 = L2	Nd	E1 Nd	E1/Nd Nd/Nd
	E1 = E2, L1 = L2	E0 L0	E0 L0	E0/L0
	E1 = L1, E2 = L2	N1 N2	N1 N2	N1/N1 N2/N2 N1/N2
1 allele	E1 = E2 = L1 = L2	N0	N0	N0/N0

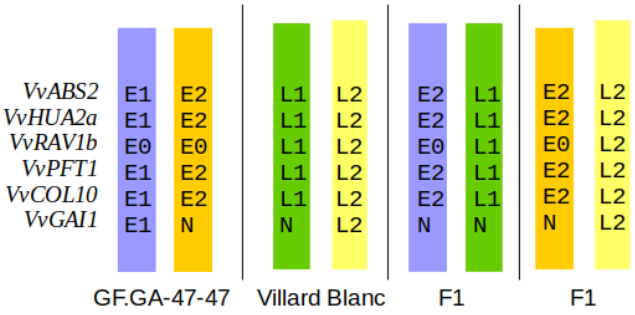

Figure 15: Illustration of the inheritage of closely neighbored genes by the example of genes within the FTC related QTL-region on chromosome 1. Each of the F1-individuals inherited a locus from one parental line. Loci containing identical alleles are shown in the same color.

5.2.4.2 Distribution of alleles of amplicon sequenced FTC candidate genes in F1-individuals

Aside from recombination events, a parent-offspring pair must share one haplotype for each chromosome and thus one identical by descent allele for every gene. Hence, Mendelian constraints were applied to validate the obtained allele specific sequences. Table 9 shows the segregation patterns of genes for which phasing was successfully performed within specified intervals (see appendix table A9 for the segregation patterns of all sequenced F1-individuals).

Table 9: Allele combinations of GF.GA-47-42 and 'Villard Blanc' in amplicon sequenced FTC candidate genes for which alleles were phased in the mapping population GF.GA-47-42 x 'Villard Blanc'.

Gene-ID	**GF.GA-47-42**	**'Villard Blanc'**
VIT_201s0011g00100	E1/E2	L1/L2
VIT_201s0011g00110	E1/N2	L1/N2
VIT_201s0011g01560	E1/E2	L1/L2
VIT_201s0011g02120	E1/E2	L1/L2
VIT_201s0011g03070	E0/E0	L1/L2
VIT_201s0011g03140	E1/E2	L1/L2
VIT_201s0011g04240	E1/E2	L1/L2
VIT_201s0011g05260	E1/N	N/L2
VIT_201s0011g06410	E1/E2	L0/L0
VIT_201s0026g00150	E1/E2	L1/L2
VIT_201s0026g02200	N1/E2	N1/L2
VIT_201s0026g02580	E1/E2	L1/L2
VIT_201s0010g00740	E1/E2	L1/L2
VIT_201s0010g02270	E1/E2	L1/L2
VIT_201s0010g03890	N1/E2	N1/L2
VIT_204s0023g02820	E1/N2	L1/N2
VIT_204s0044g00850	E1/N2	L1/N2
VIT_205s0102g01160	E1/E2	L1/L2
VIT_206s0004g03590	Na/Na	Na/L2
VIT_206s0009g02480	E1/E2	L1/L2
VIT_207s0104g01590	N1/N2	N1/N2
VIT_207s0005g02260	E1/E2	L1/L2
VIT_208s0056g01230	Nb/Nb	L1/Nb
VIT_208s0007g04200	Na/Na	Na/L2
VIT_210s0116g00750	E0/E0	L1/L2
VIT_211s0052g01800	E1/E2	L1/L2
VIT_212s0028g03350	E1/E2	L1/L2
VIT_213s0067g03390	E1/E2	L1/L2
VIT_213s0019g03550	E1/E2	L1/L2
VIT_214s0030g00440	E1/E2	L1/L2
VIT_214s0081g00440	E1/E2	L1/L2
VIT_214s0006g00640	E1/E2	L1/L2
VIT_214s0083g01030	E1/E2	L0/L0
VIT_214s0083g01050	N1/E2	N1/L2
VIT_214s0068g01800	E1/E2	L1/L2
VIT_215s0048g01280	E0/E0	L1/L2
VIT_216s0013g00860	N1/E2	N1/L2
VIT_217s0000g00150	N0/N0	N0/N0
VIT_217s0000g02630	E1/E2	L1/L2
VIT_217s0000g06570	N1/E2	N1/L2
VIT_217s0000g08480	N1/E2	N1/L2
VIT_217s0000g10300	E1/E2	L1/L2
VIT_218s0001g07460	E1/E2	L1/L2
VIT_218s0001g01800	E1/N2	L1/N2
VIT_200s0203g00080	E1/N2	L1/N2
VIT_200s1675g00010	N1/E2	N1/L2

5.2.4.3 Allele frequency bias

Amplification by PCR during sample preparation can introduce a bias as not all sequences are amplified in equal measure (Schirmer *et al.* 2015). This bias can lead to genotyping errors when variants are not detected. For diploid organisms GATK's HaplotypeCaller assumes alleles to be present at approximately 50% each. The ploidy is taken into account for the determination of genotypes. Variants thus fail to be detected when a variant is present in significantly less than 50% of reads covering the variant site. However, within the population GF.GA-47-42 x 'Villard Blanc' variant calling was performed within a family of parents and their offspring. Hence, by respect of Mendelian constraints, it was possible to distinguish sequencing errors from true variants with very low allele frequencies due to a PCR bias. Figure 16 shows an example of a strong biased allele frequency. In this case one parent is heterozygous for A/T at position 445, the other parent is homozygous for A and their offspring is heterozygous for A/T. 28 reads carrying the A-allele were amplified while 463 reads were amplified for the other allele. This resulted in an allele bias of 6%:94%. An allele frequency of 6% is below a threshold for variant detection under the default ploidy of two for diploid organisms. In other cases even lower allele frequencies were observed. A higher ploidy, lowers the threshold for variant detection. For example, when the ploidy is set to two, alleles that are approximately present at 50% each are detected. For ploidy 12, alleles are expected to be present at approximately 8,3% each. While setting the ploidy to 12, variants with an allele frequency as low as 2% can be detected. Variants with a lower allele frequency than 2% were hence not detected. For the genes *Vitis vinifera SCARECROW LIKE 21a (VvSCL21a), Vitis vinifera Squamosa Promoter-binding protein 1 (VvSPB1), VvSEP1*, and *Vitis vinifera MOTHER OF FT (VvMFT)* (VIT_201s0010g02270, VIT_212s0028g03350, VIT_214s0083g01050, and VIT_217s0000g02630 respectively) genomic regions were phased after calling variants with a ploidy value of 12.

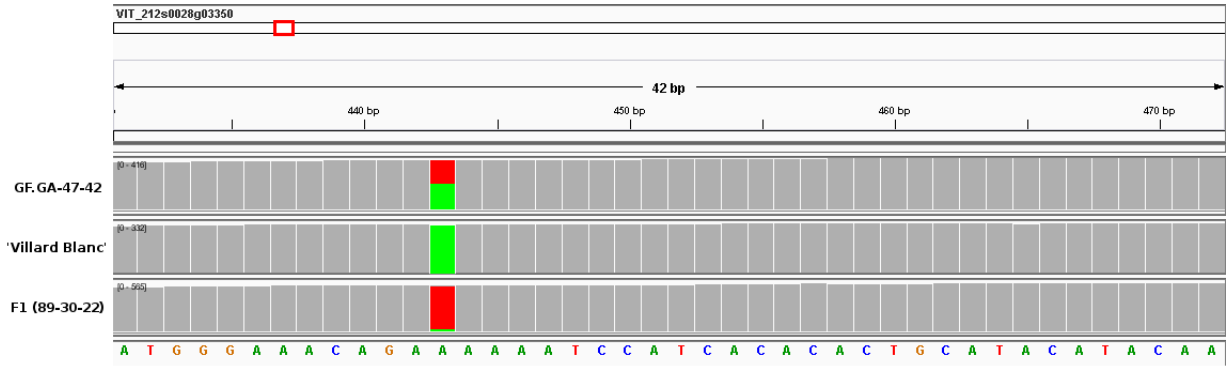

Figure 16: Example of a strong bias in allele frequency for the gene VIT_212s0028g03350 (*VvSPB1*). The upper two panels show the parental lines GF.GA-47-42 and 'Villard Blanc' and the lowest panel shows an F1-individual. Allele frequency of GF.GA-47-42: 48%/52% A/T, 'Villard Blanc': 100% A, offspring: 6%/94% A/T. The reference sequence is shown below. Image from IGV (Robinson *et al.* 2011; Thorvaldsdóttir *et al.* 2013).

Within the population GF.GA-47-42 x 'Villard Blanc' every detected variant was validated according to Mendelian inheritance patterns. Even though amplified from the same allele, the frequency of variants can vary as the amount of reads covering a variant differs from one variant to the next. When dealing with extremely biased allele frequencies, this can lead to some variants being detected while others remain undetected. When the frequency of a variant is as low as $\tilde{2}$ - 3%, another variant originating from the same allele, can have an even lower allele frequency and hence remain undetected. Setting the ploidy to high values can however introduce background noise. Non-unique primers can also introduce errors during variant detection due to the amplification of unspecific PCR-products. About 19% of sites were found to not fulfill possible inheritance patterns within a family of parents and F1-individuals in diploid organisms and were filtered out.

In some cases more than four alleles were found within the population GF.GA-47-42 x 'Villard Blanc' or alleles were found, that were not present in the parental lines. For these genes, allele phasing was performed with and without setting the ploidy to 12, and with and without filtering out sites that do not fulfill Mendelian constraints. When less than twelve of the F1-individuals were affected, these were excluded. In seventeen cases none of the approaches led to a satisfying result. Moreover, ten genes could only be amplified in very few individuals (Fig. 9). The only gene, that was amplified in only a small number of individuals (seven) and could still be phased was *VvAP1* (VIT_201s0011g00100). Overall, for 26 of the

selected FTC candidate genes allele phasing could not be performed.

5.2.4.4 Allele drop out

Another issue that emerged during variant detection was allele drop out, which also leads to genotyping errors. Allele drop out means the preferential amplification of one of the two alleles present at a heterozygous locus, causing an artificial increase in homozygosity when a heterozygous mutations falls on a primer binding region (Zucca *et al.* 2016). This leads to genotyping errors due to the missing allele. Since working with an F1-population and by applying Mendelian constraints, it was possible to identify cases where allele drop out occurred. The missing allele within the population GF.GA-47-42 x 'Villard Blanc' was determined but the sequence of the missing allele remained unknown. Figure 17 shows an example of allele drop out. One parent is homozygous for the sequence A-C-G-T-A, the other is heteroyzgous and has the sequences A-T-A-C-G and C-C-G-T-A. Both F1-individuals are homozygous for the sequences A-T-A-C-G and C-C-G-T-A. The offsprings have inherited the detected sequences from 'Villard Blanc'. An F1-individual of a given family inherits one haplotype or allele from each of the parents. The detected GF.GA-47-42-allele is not identical to one of the 'Villard Blanc'-alleles. Hence, an allelic sequence is missing. The missing sequence is common between both shown F1-individuals and GF.GA-47-42. In such cases it was possible to determine which allele was missing within the population GF.GA-47-42 x 'Villard Blanc', but the sequence of the missing allele remained unknown. Figure 17:B shows an illustration of how a missing allele is added. Table 10 shows the genes that were affected from allele drop out.

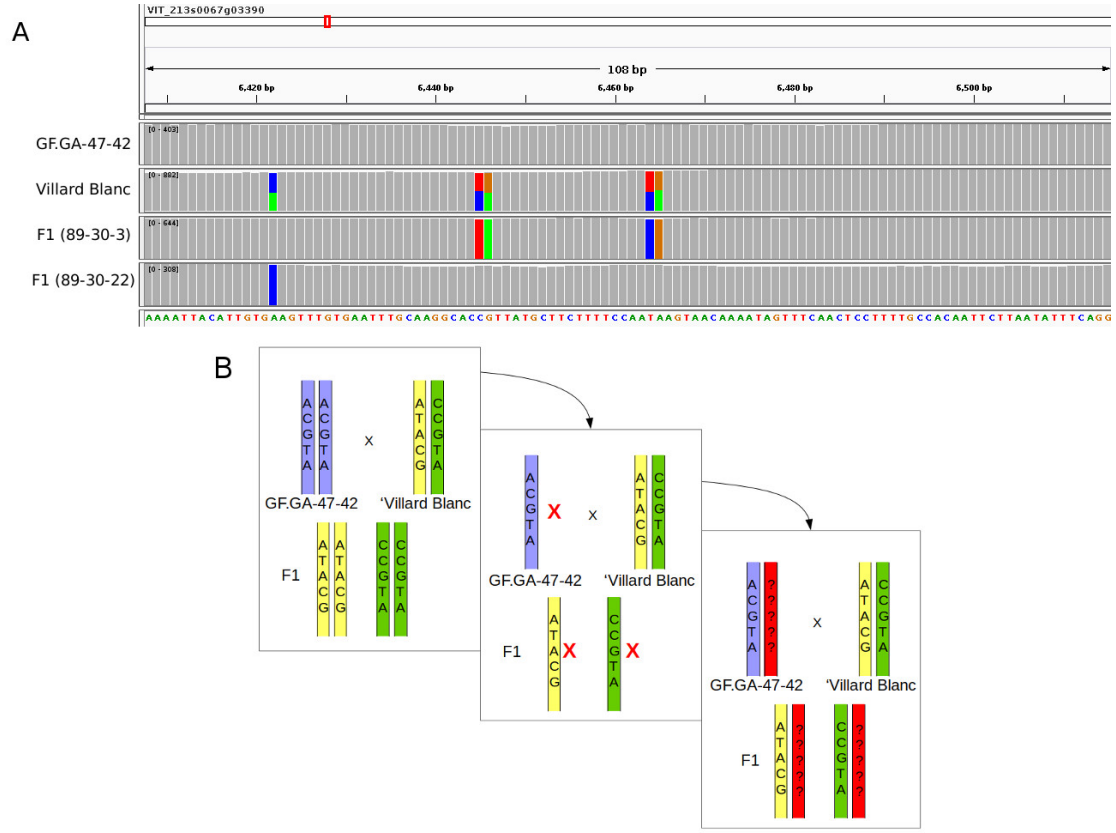

Figure 17: **A**: Example of an allele drop out. Upper two panels: parents, lower two panels: offsprings. GF.GA-47-42 is homozygous in the shown region for the sequence A-C-G-T-A, 'Villard Blanc' is heteroyzgous: 48-49% A-T-A-C-G: 51-52% C-C-G-T-A, the F1-individuals are homozygous for the sequences A-T-A-C-G and C-C-G-T-A. Non-variant sites over all shown samples are ignored. The reference sequence is shown below. Image from IGV (Robinson *et al.* 2011; Thorvaldsdóttir *et al.* 2013). **B**: Illustration of handling cases of allele drop out. GF.GA-47-42 and the two offsprings appear to be homozygous. After investigating inheritance patterns within the family, it is obvious, that an allele is missing. The three individuals are determined to be heterozygous but the sequence of the missing allele remains unknown.

Table 10: Genes that were affected from allele drop out and the name of the missing allele.

Gene-ID	Allele
VIT_210s0116g00750	L2
VIT_214s0083g01030	E2
VIT_214s0068g01800	L2
VIT_217s0000g10300	E2
VIT_200s1675g00010	L2

5.2.5 Correlation analysis

After identifying allele combinations of FTC candidate genes within the mapping population GF.GA-47-42 x 'Villard Blanc', the numbers of individuals harboring each of the alleles was determined. A correlation analysis between alleles of FTC candidate genes and the flowering time phenotype was performed for 43 genes. *VvAP1* (VIT_201s0011g00100) was not tested as sequence information was only available for seven F1-individuals. For *VvCUL1c* (VIT_207s0104g01590) both GF.GA-47-42 and 'Villard Blanc' were heterozygous but shared both alleles. Alleles common between GF.GA-47-42 and 'Villard Blanc' were not tested as it is not possible that a common allele would influence flowering time dominantly in view of the fact that GF.GA-47-42 flowers early while 'Villard Blanc' flowers late. Furthermore, only one allele was found for VIT_217s0000g00150 (*VvFL*) which was thus excluded from the correlation analysis.

A correlation between alleles and the timing of flowering was observed for genes in QTL-regions on chromosome 1, 4, 14, 17, and 18 (Fechter *et al.* 2014 and unpublished results), as well as outside of QTL-regions on chromosome 1, 14, and chrUn (Fig. 19, appendix table A10). The L2-alleles, inherited from the paternal line 'Villard Blanc', of *VvSEP4, VvBS2, VvHUA2a, Vitis vinifera Related to ABI3/VP1 1b (VvRAV1b)*, and *VvGAI1* (chr 1) correlate with late flowering. In addition to the already described floral homeotic genes *VvSEP4*, *VvHUA2a* and *VvBS2* are also a floral homeotic genes. In *A. thaliana, HUA2* regulates the expression of the floral homeotic class-C gene AGAMOUS (AG) and *FLC* (Chen *et al.* 1999; Jali *et al.* 2014). *BS2* belongs to the sister clade of class-B floral homeotic genes that are mainly expressed in female reproductive organs of flowers (Becker *et al.* 2002). The *A. thaliana* homologue of *VvRAV1b* - *TEM1* is a transcription factor that represses flowering promoter *FT* expression (Castillejo *et al.* 2008).

The E1-alleles of *VvCOL10, VvBRIa* and *VvFLC* correlate with early flowering. The

A. thaliana homologue of *VvBRIa - BRASSINOSTEROID INSENSITIVE 1 (BRI1)* - encodes a membrane-localized leucine-rich repeat receptor-like kinase and is an essential component of a receptor complex that binds Brassinolide (BL) (Caño-Delgado *et al.* 2004). *CONSTANS LIKE (COL)* genes are known to participate in the control of flowering via the photoperiodic pathway in *A. thaliana* (Putterill *et al.* 1995). *VvBS2, VvHUA2a, VvRAV1b, VvCOL10* and *VvGAI1* are located in an FTC related QTL-region, but only the E1-allele of *VvCOL10* correlates with early flowering. Otherwise, correlation with early flowering was not found in this QTL-region, but with late flowering for the L2-alleles of *VvBS2, VvHUA2a, VvRAV1b*, and *VvGAI1.*

An allele inherited from the early flowering maternal line (E1), of both *VvWNK6 (WITH NO LYSIN KINASE 6)* and *VvTM6 (TOMATO MADS 6)* on chromosome 4 showed an association to early flowering. *VvTM6* (or *VvAP3.2*) is a MADS-box B-class floral identity gene influencing the development of petals and stamen (Poupin *et al.* 2007; Theissen *et al.* 2000). *WNK6* has been shown to be involved in circadian rhythm in *A. thaliana* (Nakamichi *et al.* 2002). The E1-allele of *VvWNK6* (chromosome 4), associated with early flowering, was found to be harboring a variation leading to a stop codon in the terminal exon (SNP at chr4:21,997,435/ C $\rightarrow$ T, Arg $\rightarrow$ Stop). Figure 18 shows the distribution of allele combinations for *VvWNK6* among individuals of the mapping population GF.GA-47-42 x 'Villard Blanc'. *VvWNK6* and *VvTM6* are located 2,596,277 bp apart within a QTL-region for FTC on chromosome 4. F1-individuals carrying the E1-allele of *VvWNK6* (25) also carry the E1-allele of *VvTM6* in 88% of the cases. While all individuals carrying the E1-allele of *VvTM6* (22) also carry the E1-allele of *VvWNK6* (Table 11).

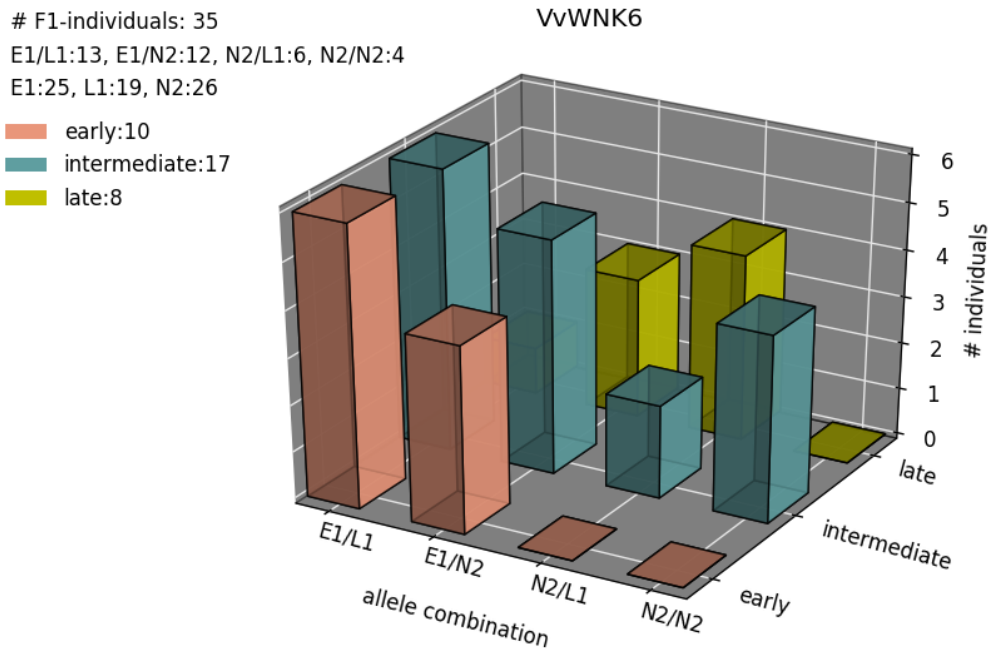

Figure 18: Distribution of allele combinations for *VvWNK6* (Chr4) among individuals of the mapping population GF.GA-47-42 x 'Villard Blanc'. For visualization flowering classes (section 4.4.1.1) 1 and 2, 3 and 4, and 5 and 6 were merged.

Table 11: Allele combinations of *VvTM6* and *VvWNK6* in F1-individuals of the mapping population GF.GA-47-42 x 'Villard Blanc'. 71% of F1-individuals carry the *VvWNK6*-E1-allele, 88% of which also carry the *VvTM6*-E1-allele.

Genes	**Allele combinations in F1-individuals**											
VvTM6	**E1/N2**	N2/N2	**E1/L1**	**E1/L1**	**E1/N2**	**E1/N2**	**E1/L1**	N2/N2	**E1/L1**	N2/N2	**E1/N2**	**E1/N2**
VvWNK6	**E1/N2**	E1/N2	**E1/L1**	**E1/L1**	**E1/N2**	**E1/N2**	**E1/L1**	N2/L1	**E1/L1**	N2/N2	**E1/N2**	**E1/N2**
VvTM6	**E1/L1**	**E1/L1**	N2/L1	**E1/L1**	N2/L1	**E1/N2**	**E1/L1**	N2/N2	**E1/L1**	**E1/L1**	**E1/L1**	N2/L1
VvWNK6	**E1/L1**	**E1/L1**	E1/L1	**E1/L1**	N2/L1	**E1/N2**	**E1/N2**	N2/N2	**E1/L1**	**E1/N2**	**E1/L1**	N2/L1
VvTM6	**E1/N2**	N2/L1	N2/N2	N2/N2	**E1/N2**	N2/N2	N2/L1	**E1/N2**	**E1/N2**	N2/N2	**E1/N2**	
VvWNK6	**E1/L1**	N2/L1	N2/N2	N2/N2	**E1/N2**	E1/L1	N2/L1	**E1/N2**	**E1/L1**	N2/L1	**E1/N2**	

On chromosome 14, E2-alleles of *VvGAIb, VvFUL2, VvSEP1*, and *VvFLC2* inside a QTL-region for flowering time were found to correlate with early flowering. *FUL* is also a floral homeotic gene that acts at later stages of flower development. Its *V. vinifera* homologue *VvFUL-L* is involved in the specification of inflorescence

and flower meristem identity and highly expressed in tendrils (Calonje *et al.* 2004; Diaz-Riquelme *et al.* 2009). The E1- and L2-alleles of *VvSVP2* in an FTC related QTL-region on chromosome 18 correlate with flowering time, as well as the L2-allele of *Vitis vinifera Cleavage Stimulating Factor 64 (VvCSTF64)* on the same chromosome but outside of the QTL-region. *SVP-like* genes control flowering time by repressing the expression of *FT* in *A. thaliana* (Lee *et al.* 2007) and are proposed to act similarly in *Vitis* (Díaz-Riquelme *et al.* 2012). The *VvCSTF64* *A. thaliana* homologue - *CSTF64* - belongs to the polyadenylation machinery of nuclear mRNAs (Yao *et al.* 2002).

For *VvGAI1* on chromosome 1 and *VvGAIb* on chromosome 14 significant correlation with late flowering was observed for alleles inherited from the paternal line 'Villard Blanc'.

Besides this, correlation was observed for genes outside of QTL-regions such as *VvSEP4, VvBRIa, VvMBD9a VvSCL21* and *VvFLC* (chr 1), *VvGID1B* and *VvFLKa* (chr 14) and *VvMFTb* (chr Un) (appendix table A10).

Chr	Gene	GF.GA-47-42	Villard Blanc	E1	E2	L1	L2	P-values	
1	VvSEP4	E1/N2	L1/N2	12 (0.435)	13 (0.6)	12 (0.48)	13 (0.69)	0.57	***0.044***
1	VvBS2	E1/E2	L1/L2	17 (0.46)	18 (0.61)	14 (0.41)	21 (0.61)	0.23	***0.019***
1	VvHUA2a	E1/E2	L1/L2	14 (0.41)	12 (0.665)	11 (0.41)	15 (0.66)	0.076	***0.04***
1	VvRAV1b	E0/E0	L1/L2	/	/	5 (0.41)	12 (0.665)	n.d.	***0.035***
1	VvPFT1a	E1/E2	L2/L1	13 (0.41)	14 (0.665)	12 (0.48)	15 (0.66)	0.077	0.12
1	VvCOL10	E1/E2	L1/L2	12 (0.41)	15 (0.67)	12 (0.48)	15 (0.66)	***0.018***	0.097
1	VvGAI1	E1/N	N/L2	15 (0.41)	10 (0.72)	14 (0.41)	11 (0.8)	0.071	***0.003***
1	VvBRIa	E1/E2	L0/L0	15 (0.41)	17 (0.61)	/	/	***0.05***	n.d.
1	VvMBD9a	E1/E2	L1/L2	18 (0.525)	17 (0.6)	19 (0.6)	16 (0.475)	0.45	0.2
1	VvTCP15b	N1/E2	N1/L2	17 (0.6)	16 (0.595)	18 (0.605)	15 (0.49)	0.7	0.22
1	VvCDF2a	E1/E2	L1/L2	17 (0.6)	12 (0.575)	18 (0.605)	14 (0.52)	0.66	0.66
1	VvbHLH49	E1/E2	L1/L2	13 (0.49)	14 (0.63)	12 (0.595)	15 (0.49)	0.44	0.77
1	VvSCL21a	E1/E2	L1/L2	13 (0.49)	13 (0.66)	12 (0.595)	8 (0.45)	0.053	0.49
1	VvFLC	N1/E2	N1/L2	14 (0.41)	19 (0.61)	19 (0.61)	14 (0.475)	***0.049***	0.14
4	VvTM6	E1/N2	L1/N2	22 (0.52)	13 (0.7)	16 (0.605)	19 (0.59)	***0.007***	0.6
4	VvWNK6	E1/N2	L1/N2	25 (0.55)	10 (0.725)	19 (0.61)	16 (0.575)	***0.007***	0.7
5	VvFPA	E1/E2	L1/L2	7 (0.55)	28 (0.6)	20 (0.6)	15 (0.55)	0.54	0.26
6	VvRAP2a	Na/Na	Na/L2	/	/	15 (0.59)	20 (0.6)	n.d.	0.51
6	VvGAMYBc	E1/E2	L1/L2	16 (0.57)	19 (0.61)	27 (0.6)	8 (0.625)	0.69	0.68
7	VvCUL1c	N1/N2	N1/N2	/	/	/	/	n.d.	n.d.
7	VvSPL1b	E1/E2	L1/L2	21 (0.6)	14 (0.55)	35 (0.6)	0	0.46	n.d.
8	VvCDF3b	Nb/Nb	L1/Nb	/	/	17 (0.61)	18 (0.505)	n.d.	0.22
8	ATHB51	Na/Na	Na/L2	/	/	21 (0.56)	14 (0.6)	n.d.	0.74
10	VvGESa	E0/E0	L1/L2	/	/	19 (0.61)	16 (0.555)	n.d.	0.31
11	VvCOL5	E1/E2	L1/L2	15 (0.66)	19 (0.6)	19 (0.59)	15 (0.6)	0.45	0.57
12	VvSPB1	E1/E2	L1/L2	12 (0.58)	22 (0.605)	14 (0.635)	20 (0.575)	0.46	0.44
13	VvPPR37b	E1/E2	L1/L2	14 (0.595)	14 (0.61)	9 (0.6)	19 (0.59)	0.73	0.43
13	VvTOE3	E1/E2	L1/L2	19 (0.56)	15 (0.61)	12 (0.605)	22 (0.57)	0.23	0.94
14	VvGID1Ba	E1/E2	L1/L2	6 (0.735)	17 (0.56)	11 (0.69)	12 (0.575)	0.13	0.33
14	VvFLKa	E1/E2	L1/L2	7 (0.67)	19 (0.49)	15 (0.6)	11 (0.55)	0.069	0.39
14	VvGAIb	E1/E2	L1/L2	12 (0.665)	16 (0.41)	13 (0.69)	15 (0.46)	***0.027***	0.059
14	VvFUL2	E1/E2	L0/L0	12 (0.725)	13 (0.41)	/	/	***0.003***	n.d.
14	VvSEP1	N1/E2	N1/L2	12 (0.725)	11 (0.41)	13 (0.69)	10 (0.435)	***0.034***	0.11
14	VvFLC2	E1/E2	L1/L2	14 (0.635)	14 (0.4)	15 (0.6)	13 (0.46)	***0.024***	0.68
15	VvAML3	E0/E0	L1/L2	/	/	16 (0.6)	18 (0.55)	n.d.	0.31
16	VvEMF2a	N1/E2	N1/L2	14 (0.635)	22 (0.61)	16 (0.6)	12 (0.61)	0.6	0.69
17	VvFL	N0/N0	N0/N0	/	/	/	/	n.d.	n.d.
17	VvMFTa	E1/E2	L1/L2	13 (0.61)	21 (0.55)	20 (0.595)	14 (0.58)	0.17	0.78
17	VvTOC1	N1/E2	N1/L2	13 (0.61)	19 (0.55)	20 (0.595)	22 (0.6)	0.15	0.52
17	VvWERb	N1/E2	N1/L2	13 (0.61)	17 (0.6)	20 (0.595)	14 (0.52)	0.28	0.27
17	VvGAIa	E1/E2	L1/L2	16 (0.6)	18 (0.555)	14 (0.575)	19 (0.6)	0.68	0.82
18	VvSVP2	E1/E2	L1/L2	17 (0.49)	17 (0.6)	15 (0.41)	19 (0.61)	***0.05***	0.064
18	VvCSTF64	E1/N2	L1/N2	16 (0.6)	9 (0.56)	10 (0.435)	15 (0.66)	0.32	***0.023***
Un	VvMFTb	E1/N2	L1/N2	1 (0.84)	23 (0.59)	15 (0.46)	9 (0.69)	0.19	0.074
Un	VvEMF2e	N1/E2	N1/L2	1 (0.84)	25 (0.6)	15 (0.46)	13 (0.6)	0.44	0.47
								GF.GA-47-42	Villard Blanc

Figure 19: Allele combinations of the parental lines of the mapping population GF.GA-47-42 x 'Villard Blanc' and the p-values of the correlation between alleles of FTC genes and the phenotype of flowering time in days after January 1st for the median of days after January 1st in the years 1999 and 2009 - 2016. Allele counts within the population and the median of flowering time in days after January 1st for each allele are shown in columns 4 - 8. Significant correlation values (0.05 and below) are in bold and italic. QTL-regions are marked in gray. "E"-alleles are inherited from GF.GA-47-42, while "L"-alleles originate from 'Villard Blanc'. "E0": E1=E2, "L0": L1=L2, "N1": E1=L1, "N2": E2=L2. "N": L2=E1 or E2=L1, "Na": E1=E2=L1, "Nb": E1=E2=L2. n.d.: not determined.

5.2.5.1 Variation of correlation values in dependence on the phenotypic data

Correlation values differed depending on whether days, or one of accumulated temperature or global radiation was used as phenotypic data as well as between the years (correlation heatmaps for all data sets can be viewed under the link: https://docs.cebitec.uni-bielefeld.de/index.php/s/kQLUhp5HAt8j4Qa). No differences in correlation were observed between accumulated temperature and global radiation.

The highest number of correlating alleles was found in 2014 using days after January 1st (16) and using accumulated temperature or global radiation in 2012 (16). The smallest number of correlating alleles was found in 2010 using days after January 1st (7). The genes of which alleles showed the most stable correlation are *VvGAI1* (Chr1) and *VvCSTF64* (Chr18) both of which the 'Villard Blanc'-L2 allele was found to significantly correlate with late flowering in all years except 2013, and *VvFUL-L* of which the E2-allele inherited from the maternal line GF.GA-47-42 correlates with early flowering in all year except 2010.

A significant correlation for the E1-alleles of genes in the QTL-region on chromosome 4 (*VvTM6* and *VvWNK6*) was observed in 1999, 2009, 2011 (only *VvWNK6*), and 2012 - 2014 (Table 12). In 2016 neither for days after January 1st, nor for both of accumulated temperature and global radiation was a correlation detectable. In 2015, the correlation was not significant for days after January 1st but, albeit only slightly, for the other two sets of phenotypic data. For *VvTM6*, however, correlation was significant in the years 1999, 2009, and 2012 - 2014 when considering days after January 1st, but only in 2014 when considering accumulated temperature or global radiation. None of the genes shows a correlation with the phenotype in 2010 using days after January 1st. However, using both the median and the average of flowering time in the years 1999 and 2009 - 2016 a significant correlation is observable for both genes. Table 12 shows p-values of the correlation analysis for different sets of phenotypic data.

Table 12: P-values of the correlation between the E1 allele of both *VvWNK6* and *VvTM6* in different sets of phenotypic data. Significant values are in bold and red.

Data set	Year	***VvWNK6* (E1)**	***VvTM6* (E1)**
Days after January 1st	1999	**0.032**	**0.023**
	2009	**0.012**	**0.009**
	2010	0.440	0.370
	2011	**0.033**	0.063
	2012	**0.047**	**0.041**
	2013	**0.008**	**0.012**
	2014	**0.015**	**0.029**
	2015	0.067	0.063
	2016	0.177	0.098
	Average 1999-2016	**0.012**	**0.009**
	Median 1999-2016	**0.007**	**0.006**
Accumulated Temp. above 3 °C	2011	**0.027**	0.109
	2012	**0.030**	0.091
	2013	**0.004**	0.058
	2014	**0.003**	**0.016**
	2015	**0.046**	0.186
	2016	0.177	0.098
Global radiation (KWh/m^2)	2011	**0.027**	0.109
	2012	**0.030**	0.091
	2013	**0.004**	0.058
	2014	**0.003**	**0.016**
	2015	**0.046**	0.186
	2016	0.177	0.098

Correlation for genes in the QTL-region on chromosome 1 was observed in all years but with differences in terms of the alleles showing significant correlation with the timing of flowering. The E1-allele of *VvCOL10* correlates with early flowering in 1999, 2009, and 2013 - 2016. In 2015, the E1-alleles of *VvBS2, VvHUA2a, VvRAV1b, VvCOL10*, and *VvGAI1* were also found to correlate with early flowering. The latter also showed a significant correlation in 2016. Apart from this, correlation for an allele inherited from the maternal line GF.GA-47-42 was not observed for the QTL-region on chromosome 1. However, L2-alleles inherited from 'Villard Blanc' of *VvBS2, VvHUA2, VvGAI1*, and *VvRAV1b* were found to correlate with late flowering in different years (*VvBS2*: 2011 and 2012, *VvHUA2*: 2012, *VvRAV1b*: 2012 and 2014, *VvGAI1*: all years except 2013).

For the QTL-region on chromosome 14 significant correlation for *VvFUL-L* was observed in all analyzed years except 2010 and for *VvFLC2* in 1999,

2009, 2012, 2014, and 2016. In 2012, 2014, and 2016 significant correlation was found for the E2-alleles of all analyzed genes within the QTL-region (*VvGAIb, VvFUL2, VvSEP1*, and *VvFLC2*). Moreover, for *VvGAIb*s L1-allele significant correlation with late flowering was detected in 2010, 2011, 2013 and 2015.

In the QTL-region on chromosome 17 significant correlation was found for *VvTOC1*, whose *A. thaliana* homologue *TIMING OF CAB EXPRESSION 1 (TOC1)* plays an important role in circadian clock regulation (Alabadi *et al.* 2001), in 1999, 2010, and 2012.

5.3 Marker development and analysis

From fifteen markers, twelve showed a segregation pattern matching the segregation pattern that was obtained through allele phasing (appendix table A11). Using the results of marker segregation across the above mentioned 151 F1-individuals, a correlation analysis between alleles and flowering time was performed. Table 13 shows the results of correlation analysis based on the median of flowering time in days after January 1st in the years 1999 and 2009 - 2016. Correlation results of marker analysis support those of allele phasing. Figure 20 shows the distribution of allele combinations of *VvWNK6* across all 151 F1-individuals of the mapping population GF.GA-47-42. The a-allele (Chr4:21,994,934..21,995,149) is derived from the same haplotype as the E1-allele (Chr4:21,997,042..21,997,532). This haplotype originates from the GF.GA-47-42-parent 'Bacchus'. Correlation results of both marker analysis and allele phasing show that the Bacchus-allele of *VvWNK6* highly correlates with early flowering.

Table 13: P-values of the correlation between alleles and the phenotype of flowering time from both the allele phasing workflow (first row) and marker analysis (second row) based on days after January 1st from the years 1999 and 2009 - 2016. Marker analysis was performed in all 151 F1-individuals of the mapping population GF.GA-47-42 x 'Villard Blanc', while allele phasing was performed in 35 F1-individuals. Number of alleles over the analyzed F1-individuals and the median of each are given in the same order as in column 3. "E0": E1=E2, "L0": L1=L2, "N1": E1=L1, "N2": E2=L2. "N": L2=E1 or E2=L1, "Na": E1=E2=L1, "Nb": E1=E2=L2. ab x cd: four alleles/ both parents heterozygous, ef x eg: three alleles/ parents heterozygous, lm x ll: two alleles/ mother heterozygous, nn x np: two alleles, father heterozygous.

Gene	**Marker name**	**Segregation GF x VB**	**Position**	**p-values**		**Allele numbers**				**Median**			
				GF	**VB**	**GF**		**VB**		**GF**		**VB**	
VvbHLH49		E1E2 x L1L2	Chr1:15904052..15907131	0.438	0.77	13	14	12	15	0.48	0.62	0.56	0.48
	GAVBInd_009	lm x ll (m = E1)	Chr1:15904095..15904241	0.32	/	64	67	/	/	0.48	0.5	/	/
VvHUA2		E1E2 x L1L2	Chr1:1797008..1802965	0.076	0.04	14	12	11	15	0.41	0.665	0.41	0.66
	GAVBInd_001	nn x np (p = L1)	Chr1:1797458..1797913	/	0.02	/	/	67	65	/	/	0.47	0.51
VvCOL10		E1E2 x L1L2	Chr1:3850488..3854068	0.018	0.097	12	15	12	15	0.41	0.67	0.48	0.66
	GAVBInd_005	lm x ll (m = E2)	Chr1:3849710..3849912	0.295		66	68			0.46	0.505	/	/
	GAVBInd_004	ef x eg (f=E2, g=L2)	Chr1:3853724..3853919	0.4	0.39	54	75	68	61	0.48	0.5	0.5	0.5
VvWNK6		E1N2 x L1N2	Chr4:21997042..21997532	0.007	0.703	25	10	19	16	0.55	0.725	0.61	0.575
	GAVBInd_019	abxcd	Chr4:21994934..21995149	0	0.573	81	67	66	82	0.45	0.54	0.475	0.5
	GAVBInd_019	lm x ll	Chr4:21994934..21995149	0	/	81	67	/	/	0.45	0.54	/	/
VvFPA		E1E2 x L1L2	Chr5:23301727..23304317	0.536	0.257	7	28	20	15	0.55	0.6	0.6	0.55
	GAVBInd_007	lm x ll (m = E2)	Chr5:23302979..23303340	0.7	/	53	79	/	/	0.48	0.5	/	/
VvGAMYBc		E1E2 x L1L2 ab x cd	Chr6:15021045..15016088	0.691	0.68	16	19	27	8	0.57	0.61	0.6	0.625
	GAVBInd_014	(a = E1, b=E2, c= L1, d L2)	Chr6:15019109..15018684	0.45	0.79	70	59	96	33	0.5	0.5	0.5	0.46
VvCOL5		E1E2 x L1L2	Chr11:19620851..19622370	0.445	0.567	15	19	19	15	0.66	0.6	0.59	0.6
	GAVBInd_015	lm x ll (m = E2)	Chr11:19621808..19622007	0.81	/	67	64	/	/	0.48	0.5	/	/
VvTOE3		E1E2 x L1L2	Chr13:4720125..4724087	0.231	0.943	19	15	12	22	0.56	0.61	0.605	0.57
	GAVBInd_016	ef x eg (f=E2, g=L2)	Chr13:4723743..4724017	0.23	0.85	69	66	68	67	0.5	0.49	0.5	0.5
	GAVBInd_017	lm x ll (m = E2)	Chr13:4724087..4721775	0.13	/	71	65	/	/	0.5	0.48	/	/
VvPRR37		E1E2 x L1L2	Chr13:1862179..1864516	0.73	0.431	14	14	9	19	0.595	0.61	0.6	0.59
	GAVBInd_018	lm x ll (m = E1)	Chr13:1863290..1863575	0.88	/	67	66	/	/	0.48	0.5	/	/
VvGAIb		E1E2 x L1L2	Chr14:14805995..14809259	0.027	0.059	12	16	13	15	0.665	0.41	0.69	0.46
	GAVBInd_006	ef x eg (f=E1, g=L2)	Chr14:14806045..14806275	0	0.25	80	55	70	65	0.45	0.58	0.48	0.51
VvFLKa		E1E2 x L1L2	Chr14:8304600..8312884	0.069	0.392	7	19	15	11	0.67	0.49	0.6	0.55
	GAVBInd_012	lm x ll (m = E2)	Chr14:8304930..8305109	0.01	/	79	59	/	/	0.46	0.56	/	/
VvFUL2		E1E2 x L0L0	Chr14:23324870..23319446	0.003	/	12	13	/	/	0.725	0.41	/	/
	GAVBInd_020	abxcd	Chr14:23323205..23322789	0	0.93	58	89	67	80	0.575	0.44	0.5	0.49
VvSVP2		E1E2 x L1L2	Chr18:5703037..5705883	0.05	0.064	17	17	15	19	0.49	0.6	0.41	0.61
	GAVBInd_008	lm x ll (m = E2)		0.86	/	62	66	/	/	0.48	0.5	/	/

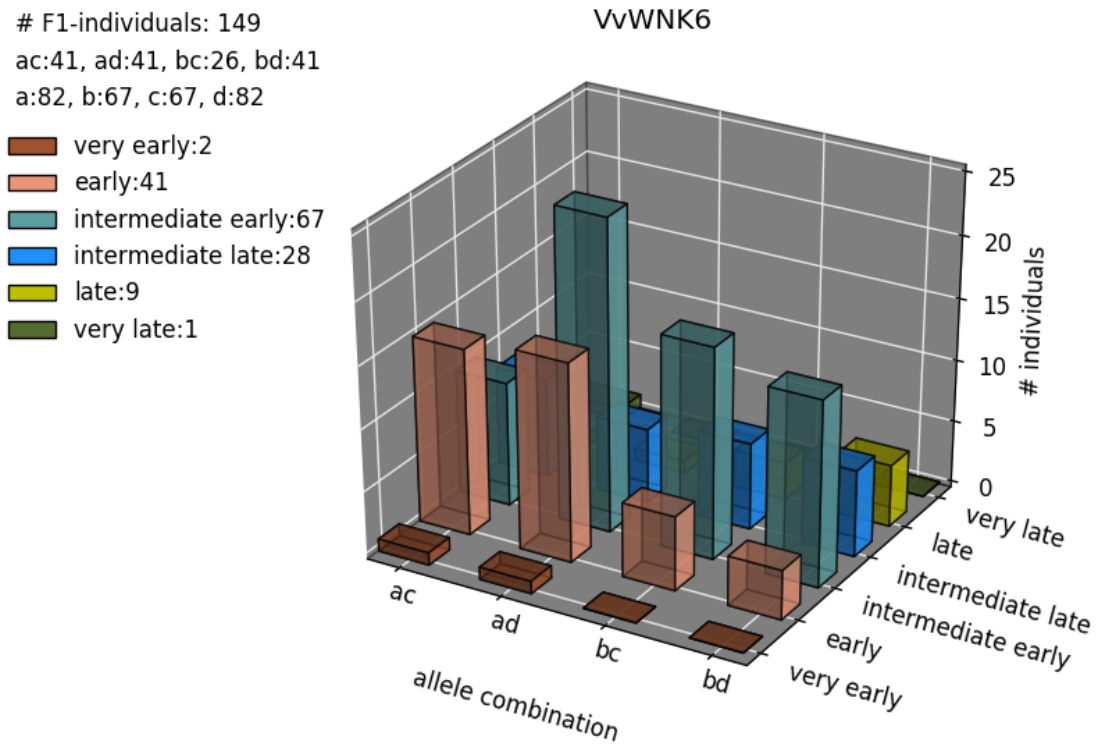

Figure 20: Distribution of allele combinations for *VvWNK6* among individuals of the mapping population GF.GA-47-42 x 'Villard Blanc' according to the median of flowering time in days after January 1st in the years 1999 and 2009 - 2016. The 'Bacchus' derived "a"-allele of GF.GA-47-42 correlates with early flowering. ab x cd: four alleles/ both parents heterozygous.

5.4 Investigating alleles of FTC candidate genes in grapevine cultivars

Different *Vitis* cultivars, such as 'Chardonnay', 'Riesling', and 'Pinot Noir' as well as some breeding lines (appendix table A3), were investigated for common alleles of FTC candidate genes. This aimed to ascertain whether alleles of FTC candidate genes, correlating with either early or late flowering time in the mapping population GF.GA-47-42 x 'Villard Blanc', lead to the same phenology in the timing of flowering in other cultivars. This would give further indications towards the role of these alleles in the control of flowering time.

For the detection of common alleles between the cultivars, allele sequences that were previously identified in the mapping population GF.GA-47-42 x 'Villard Blanc' were compared between different cultivars using blast (Camacho *et al.* 2009). For example, for *VvTM6*, three allelic sequences, that were present in several cultivars were found. For *VvWNK6*, GF.GA-47-42 and 'Villard Blanc' are identical for one allele (N2). This sequence was also found in 'Grenache Noir', for which no pedigree relationship is known. GF.GA-52-42 was found to share one allele of *VvWNK6* with 'Villard Banc' and the other with GF.GA-52-42. GF.GA-52-42 is an offspring of 'Villard Blanc' and a half-sibling of GF.GA-47-42 (Fig. 22). The L1-allele of *VvTM6* from 'Villard Blanc' was also found in GF.GA-52-42, 'Riesling', and 'Syrah' all of which are late flowering. GF.GA-47-42 and 'Villard Blanc' are identical for an allelic sequence of *VvTM6* (N2). The N2-sequence was also found in both of 'Syrah' and 'Riesling'. Riesling thus shared two alleles of *VvTM6* with individuals of the population Gf.GA-47-42 x 'Villard Blanc'. Moreover, 'Chardonnay', 'Pinot Noir précoce', GF.GA-52-42, 'Riesling', 'Silvaner', and 'Pinot Noir' were also found to share an allele of *VvTM6*, whereat 'Silvaner' is homozygous (Table 14). Figures 21 and 22 show pedigree trees of the analyzed cultivars and the inheritance pattern of alleles of *VvTM6* and *VvWNK6* (Chr4). All analyzed late flowering lines carry the allele *VvTM6*-L1.

Table 14: *V. vinifera* lines with common sequences of alleles of *VvTM6* and *VvWNK6* within intervals in which phasing was performed in the mapping population GF.GA-47-42 x 'Villard Blanc' (*VvTM6*: Chr4:19,395,984..19,397,583, *VvWNK6*: Chr4:21,997,042..21,997,532). Alleles present in only one cultivar are not shown as well as alleles with no overlap to the phased region of a gene with the mapping population GF.GA-47-42 x 'Villard Blanc'. hom: homozygous.

Gene	**Lines with common alleles within a defined interval**		
VvTM6	GF.GA-52-42 'Riesling' 'Syrah' 'Villard Blanc' (L1)	GF.GA-47-42 (N2) 'Grenache Noir' 'Villard Blanc' (N2)	'Chardonnay' 'Pinot Noir précoce' GF.GA-52-42 'Riesling' 'Silvaner' (hom) 'Pinot Noir'
VvWNK6	GF.GA-52-42 'Villard Blanc' (L1)	GF.GA-47-42 (N2) 'Grenache Noir' 'Villard Blanc' (N2)	GF.GA-47-42 (E1) GF.GA-52-42

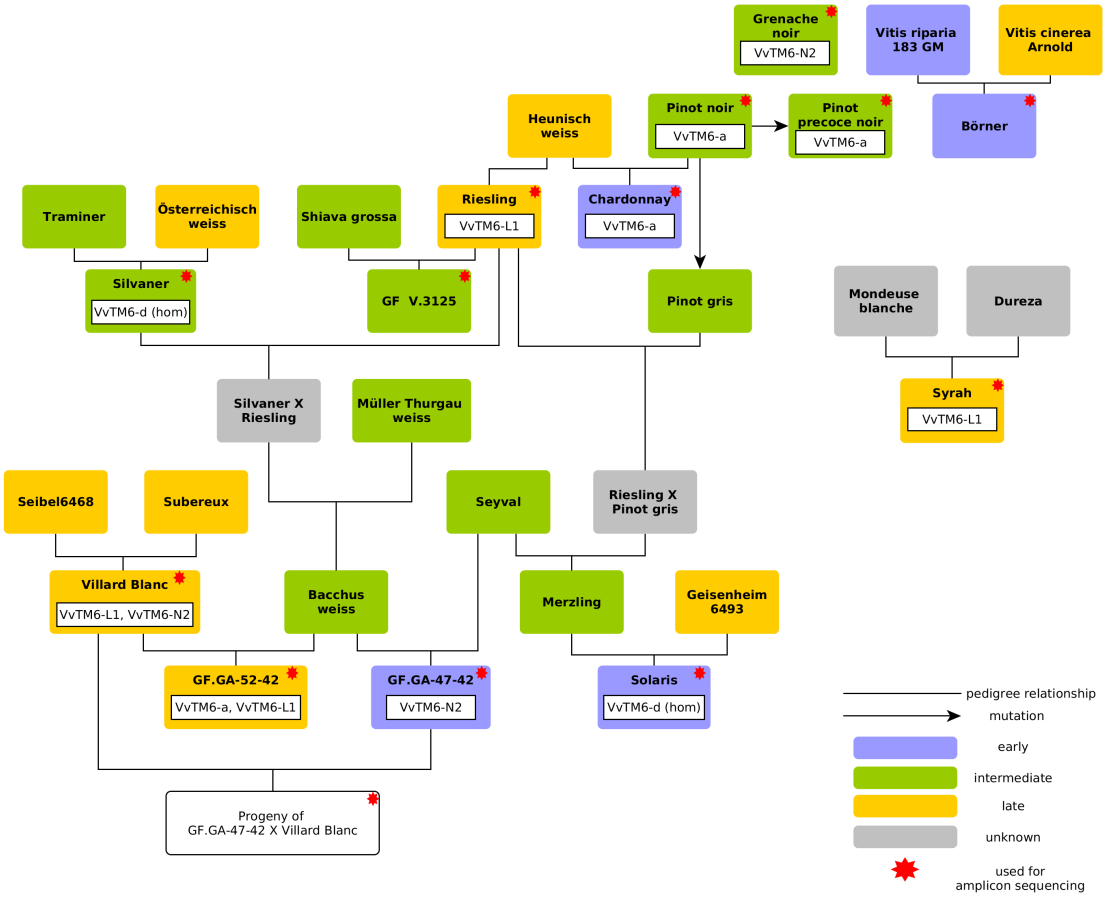

Figure 21: Pedigree tree of the different *Vitis* cultivars with alleles of *VvTM6* on chromosome 4. All analyzed late flowering lines carry the allele *VvTM6*-L1. Analyzed cultivars within this work, are marked with a red *. The flowering time classes intermediate-early, intermediate and intermediate-late were merged.

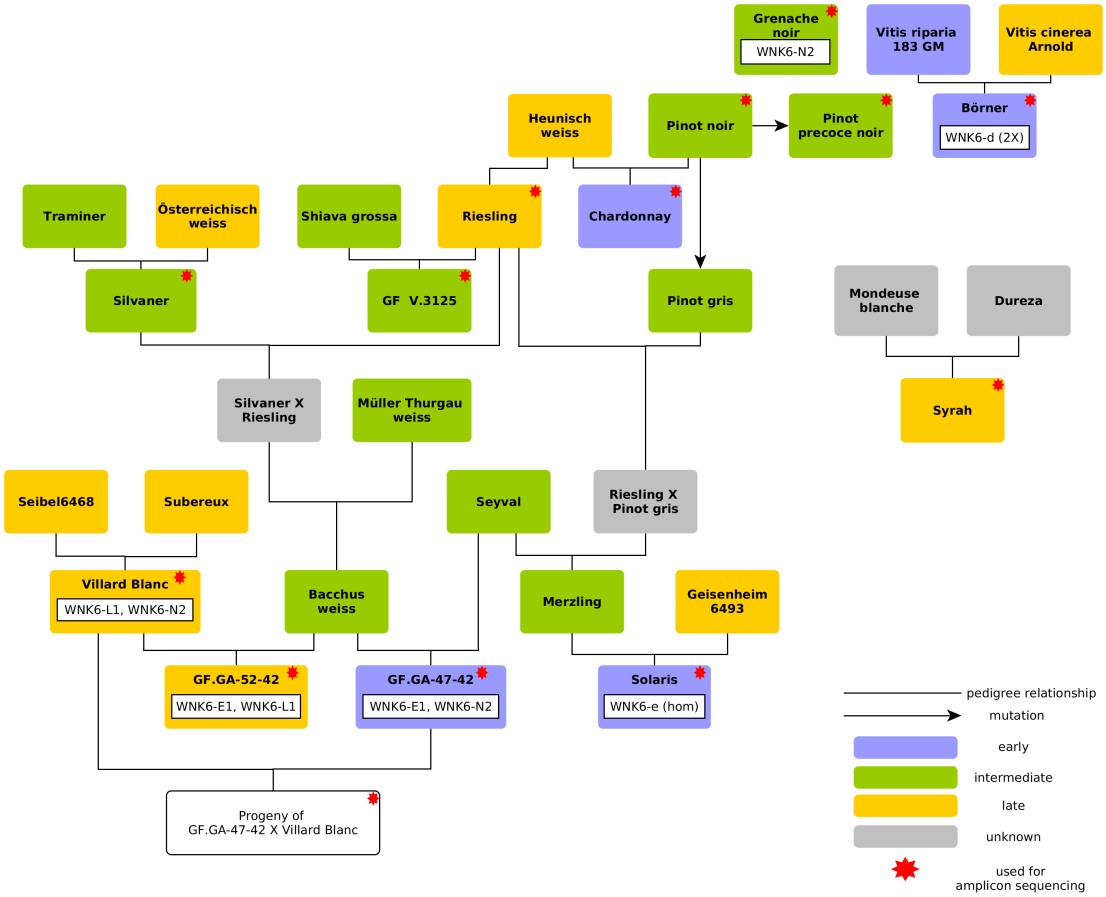

Figure 22: Pedigree tree of the different *Vitis* cultivars with alleles of *WNK6* on chromosome 4. Both early flowering GF.GA-47-42 and late flowering GF.GA-52-42 carry the *VvWNK6*-E1-allele that correlates with early flowering. Analyzed cultivars within this work, are marked with a red *. The flowering time classes intermediate-early, intermediate and intermediate-late were merged.

For 36 genes up to four sequences that were common between alleles of several cultivars were found (appendix table A12). This includes sequences shared between cultivars as well as between alleles of one cultivar. For *VvMFT* a common sequence within a defined interval of 588 bp was found in eight different cultivars, which were also homozygous for this allele. Other alleles, however, were common in as less as two cultivars.

Following alleles correlating with flowering time in the mapping population GF.GA-47-42 x 'Villard Blanc' were found in other cultivars:

- the E1-allele of GF.GA-47-42 for *VvBRIa* that correlates with early flowering was also found in 'Börner', GF.GA-52-42, 'Grenache Noir', and 'Solaris'. 'Börner' and 'Solaris' are early flowering, while 'Grenache Noir' is intermediate and GF-GA-52-42 late flowering,
- the L2-allele of *VvBS2* present in 'Villard Blanc' and correlating with late flowering was also found in early flowering 'Chardonnay', intermediate flowering 'Grenache Noir' and 'Silvaner', and late flowering 'Riesling' and 'Syrah',
- and eventually the strongly with early flowering correlating allele of *VvCOL10* was also found in GF V.3125 (intermediate flowering), 'GF.GA-52-42', 'Riesling', and 'Silvaner'.

5.5 Investigating expression of FTC candidate genes

Different grapevine varieties initiate flowering and hence reach full bloom at different dates under the same environmental conditions in dependence on the genetic background of the variety. In order to investigate the expression patterns of FTC candidate genes and further analyze their role, RNA-Seq experiments in resting and developing buds, inflorescences, and leaves of early flowering GF.GA-47-42 and late flowering 'Villard Blanc' were performed.

5.5.1 Processing of RNA-Seq data

5.5.1.1 Mapping

After trimming and quality control, reads were mapped to the *Vitis* reference sequence PN40024-12xv2. Between 46.2 and 80.8% of the reads were mapped to the reference (average: 70.72%, SD: 6.03%) (appendix table A13).

5.5.1.2 Normalization

Different libraries were sequenced at different depths but samples need to be comparable to each other for the calculation of differential gene expression between samples. Therefore, sizefactors were calculated. A sizefactor is the median ratio of each sample to a reference sample formed by the geometric mean of each single gene across samples (Love *et al.* 2014). Hence, it gives a scaling factor which makes gene counts comparable across samples. Normalization is reached by dividing the counts of each gene by the corresponding sizefactor of the library. If the sizefactors of all libaries are roughly equal to one, the libraries have been sequenced at equal depths. Table 15 shows the sizefactors of the analyzed samples. Sizefactors are

distributed between 0.285 and 1.703. Hence, libraries were not sequenced at equal depths. However, sizefactors vary in dependence of the samples that were compared with each other and are relative to each other. A successful normalization produces overlapping densities and brings the samples to a common scale (Fig. 23).

Table 15: Sizefactors of the sequenced samples. Sizefactors were calculated for all samples together and are in relation to each other.

Tissue	**Date**	**Sizefactors GF.GA-47-42**	**Sizefactor 'Villard Blanc'**
bud	02.08.2013	1.193	1.193
	08.08.2013	1.278	1.073
	16.08.2013	1.680	1.320
	22.08.2013	1.459	1.422
	05.09.2013	1.310	1.168
	20.12.2012	0.985	1.095
	08.03.2013	1.454	1.378
	22.03.2013	1.744	1.493
	12.04.2013	1.550	1.703
	26.04.2013	1.338	1.376
	03.05.2013	1.163	1.337
inflorescence	07.06.2013	1.100	1.094
	14.06.2013	0.886	0.992
	17.06.2013	0.498	0.768
leaf	07.06.2013	0.285	0.601
	14.06.2013	0.808	0.686
	22.07.2013	0.710	0.655
	08.08.2013	0.773	0.785
	22.08.2013	0.427	0.701
	19.09.2013	0.792	0.815

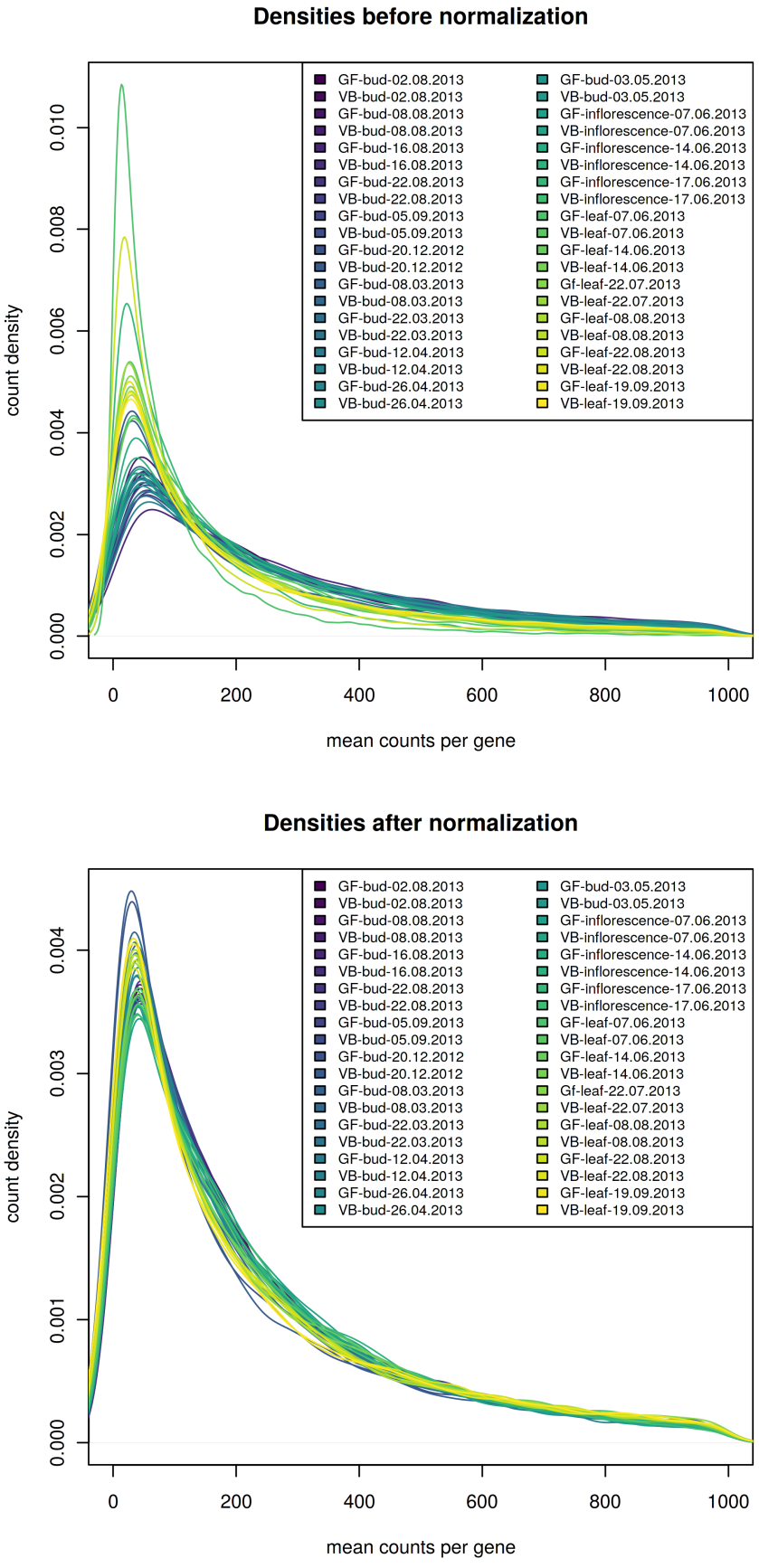

Figure 23: Density estimates of the sample counts before and after normalization. After normalization densities were overlapping and samples are on a common scale indicating the successful normalization.

5.5.1.3 Investigation of sample distances

In order to get an insight into the similarity of samples, the Euclidean distances between samples are calculated. In RNA–Seq data, variance grows with the mean. If samples are compared (by performing a PCA for instance) and matrices of normalized read counts are used, the result typically depends only on the few most strongly expressed genes because they show the greatest differences between samples. To avoid this, the logarithm of the normalized count values plus a small pseudocount is often used. However, this leads to genes with low counts dominating the result because they show the strongest relative differences between samples due to the strong Poisson noise of small count values. To overcome this issue, a regularized–logarithm (r-log) transformation of the count data was performed. For genes with high counts, the r-log-transformation is similar to a log2-transformation. For genes with lower counts, the count values are shrunken towards the average counts of genes across all samples. The r-log-transformed data is homoskedastic, which means that the variance of gene expression is independent from the mean (Love *et al.* 2014). Figure 24 shows a clustering map of sample distances. For most of the samples the next closest sample (the sample with the smallest distance to it) is either the sample from the next closest time point of the cultivar (either GF.GA-47-42 or 'Villard Blanc') or the sample from the same date of the other cultivars. This matches the expectations. However, the inflorescence and leaf samples of GF.GA-47-42 from June 14th 2013 have the smallest distance to each other. This is not expected as these samples are from different tissues. Perhaps, tissue material was mixed during sample collection. Also, this was not observed for 'Villard Blanc'. The inflorescence sample of GF.GA-47-42 from June 17th 2013 has a high distance to all other samples.

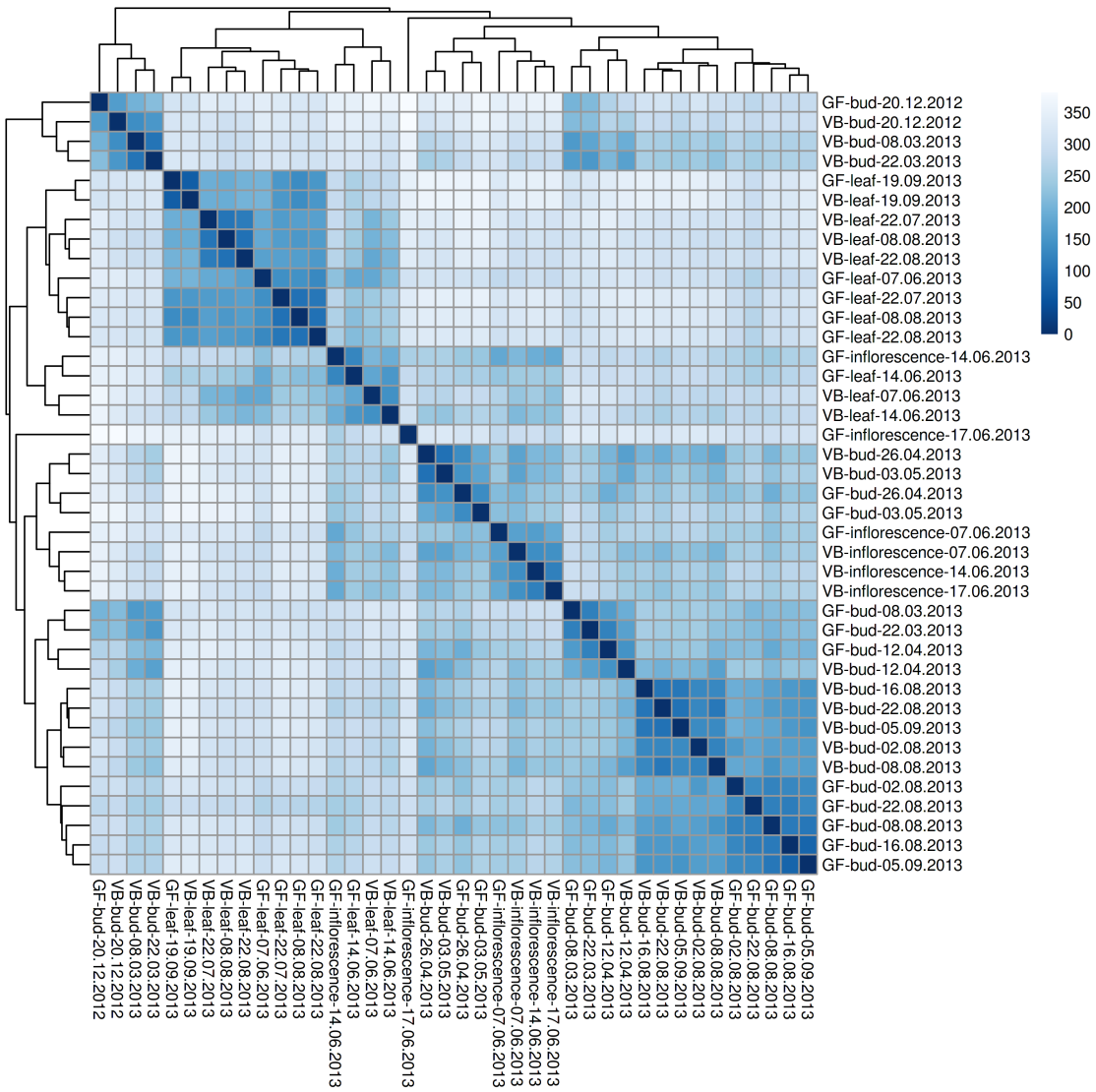

Figure 24: Clustering map of Euclidean distances between samples. For each sample the distance to all other samples is shown. The scale on the right side: Euclidean distances. GF: GF.GA-47-42. VB: 'Villard Blanc'.

Another way to visualize distances between samples is by performing a principal component analysis (PCA). An orthogonal transformation is performed to convert r-log-transformed count data to so called principal components, which are uncorrelated. The first principal component has the largest possible variance and every succeeding component has the highest variance possible while being orthogonal to the previous components (Jolliffe 2002). Samples are projected onto a two-dimensional plane so that they spread out in the two directions that explain

most of the differences. On the x-axis are the values of the first component which separates the data the most. The y-axis is the second component, which separates the data the second most. The percentages of the first and second principal component do not add to 100%, as there are more dimensions that contain the remaining variance (Love *et al.* 2014). The 500 most highly variable genes were selected for performing the PCA. The PCA plot (Fig. 25) shows that the differences between tissues is greater than the differences between time points. The sample of GF.GA-47-42 from 17th June 2013 was excluded from the plot since it shows a very high distance to all other samples and is hence the main contributor to the formation of the first component.

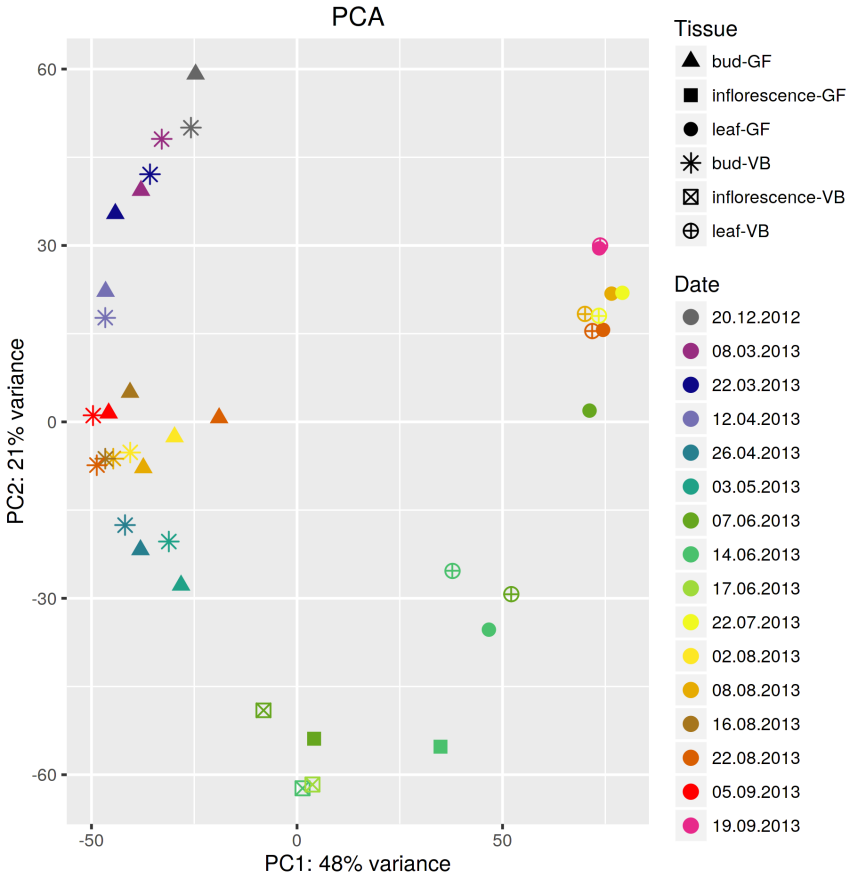

Figure 25: Visualization of a principal component analysis between RNA-Seq samples. The values of the first component are plotted on the x-axis while the values of the second component are plotted on the y-axis. The sample "GF.GA-47-42 - 17th June 2013" is not shown. GF: GF.GA-47-42, VB: 'Villard Blanc'.

5.5.2 Differential gene expression analysis

Differential gene expression was analyzed between early flowering GF.GA-47-42 and late flowering 'Villard Blanc' in several time courses of developmental stages of buds and inflorescences (Table 4). Samples were collected from both the developmental cycle 2012/2013 and 2013/2014. Time courses of the samples were created and differential gene expression was analyzed between the time points of the time courses as well as between GF.GA-47-42 and 'Villard Blanc' (Table 16). See appendix table A14 for the overall number of expressed genes for each of the samples.

Differential gene expression was aimed to be analyzed over one entire developmental cycle separated by a dormancy period to test for difference in expression between consecutive developmental stages of buds and inflorescences. However, samples were available from the dormancy period until shortly before full bloom for the developmental cycle 2012/2013 on one hand and prior to dormancy for the developmental cycle 2013/2014 on the other hand. Therefore, differential expression could not be tested over an entire developmental cycle but was instead tested over two separate time courses (Table 16).

Table 16: Time courses for the analysis of differential gene expression in buds, inflorescences and leaves.

Time course	Time span	Tissue
1	02.08.2013 – 05.09.2013	buds
2	20.12.2012 – 17.06.2013	buds & inflorescences

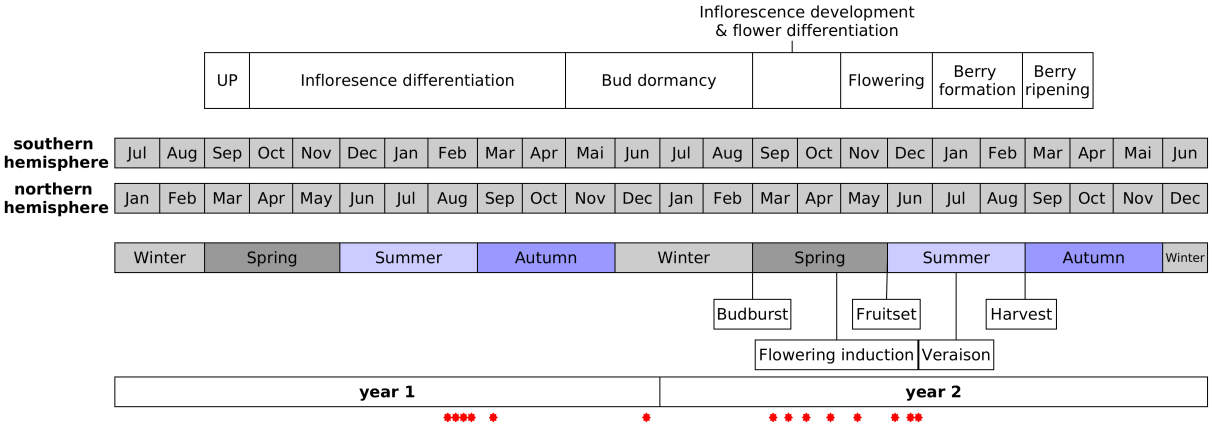

Figure 26: The developmental cycle of grapevine with the time points of which samples from buds and inflorescences were collected marked with a red *.

5.5.2.1 Gene expression prior to dormancy

Over consecutive stages of bud development prior to dormancy (Table 16 - time course 1), differential expression in GF.GA-47-42 was detected for the MADS transcription factor *Vitis vinifera TOMATO MADS 8 (VvTM8)* as well as for the protein kinase encoding gene *Vitis vinifera WITH NO LYSIN KINASE 5 (VvWNK5)* (Figure 27). *VvTM8* is a MIKC transcription factor, whose *A. thaliana* homologue, *TM8*, has been shown to be involved in the specification of flower organ identity (Diaz-Riquelme *et al.* 2009). Differential expression between GF.GA-47-42 and 'Villard Blanc' was found in consecutive developmental stages of bud differentiation before dormancy for a leucine rich repeat receptor like serine/threonine protein kinase - *VvLRR-RLK* (log2 fold change: 427.57, padj: 3.6E-8). This gene was also differentially expressed in 'Villard Blanc' over the same time course (Figure 28). LRR-RLKs are involved in development and stress responses in plants (Liu *et al.* 2017). Even in the genes with the highest variance across samples for GF.GA-47-42 and 'Villard Blanc' gene expression remains relatively stable over consecutive developmental stages of time course 1 (data not shown).

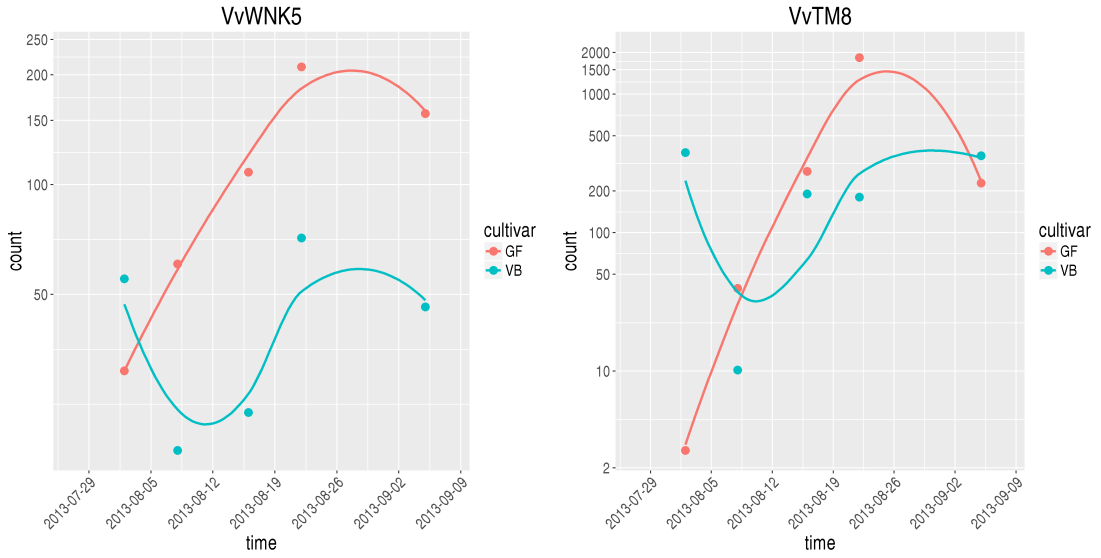

Figure 27: Gene expression profiles of *VvWNK5* and *VvTM8* in GF.GA-47-42 and 'Villard Blanc'. Both genes show significant differential expression over time in a time course of consecutive stages of bud development prior to dormancy in GF.GA-47-42.

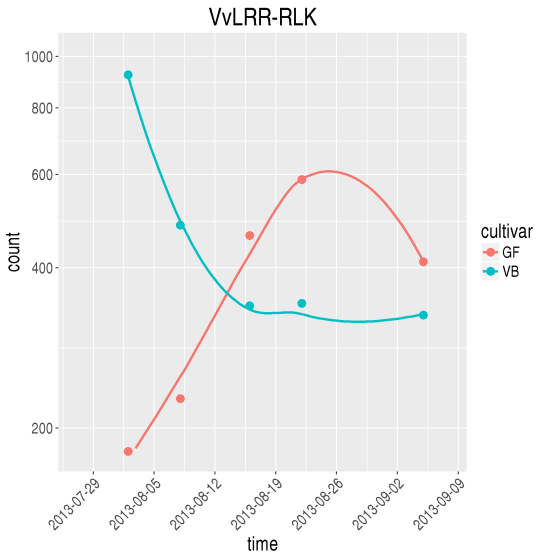

Figure 28: Gene expression profile of the leucine rich repeat receptor like serine/threonine protein kinase *VvLRR-RLK* in samples of consecutive developmental stages of buds before dormancy. *VvLRR-RLK* is differentially expressed between GF.GA-47-42 and 'Villard Blanc'.

5.5.2.2 Gene expression from dormancy to bloom

In consecutive developmental stages of buds/infloresences from bud dormancy until inflorescence stages shortly before full bloom (Table 16 - time course 2), *VvGAMYBa* (VIT_212s0059g00700) was found to be differentially expressed between the samples of 'Villard Blanc' and GF.GA-47-42 from 17th June 2013 (data not shown). However, the sample of GF.GA-47-42 from this time point exhibits a high distance to all other samples (section 5.5.1.3) and the differential expression might thus be an artifact.

When not testing for cultivar specific, but for time specific effects over consecutive developmental stages of buds and inflorescences from bud dormancy until shortly before full bloom (Table 16 - time course 2) for each of the cultivars, differential expression was detected in various FTC genes. In a time course of buds from dormancy (BBCH 0) until after bud burst when leaf formation had already begun (BBCH 11-13), 58 of the FTC candidate genes showed a time-dependent expression in GF.GA-47-42, 16 were found in 'Villard Blanc', and ten in both. Several of these genes are squamosa binding proteins, MADS – and MYC transcription factors that are known to influence floral development. Most of these genes are differentially expressed due to an up or down regulation towards inflorescence development. In order to test for differential expression between consecutive stages

of bud development before inflorescence maturation, inflorescences were excluded from the analysis. Genes that were differentially expressed when the time course was extended to include inflorescences, are those showing differential expression between buds and young to mature inflorescences. For GF.GA-47-42, 67 of such "inflorescence-specific genes" were found, 92 for 'Villard Blanc', and 35 genes common between both cultivars (appendix table A15).

After excluding infloresences, several genes showing a significant time dependent expression over time were found. They cluster into two groups: genes upregulated in winter during bud dormancy (Fig. 29, upper part) and genes upregulated towards inflorescence development (Fig. 29, lower part). Most of these genes are transcription factors, such as bzip-, MADS- or MYC-transcription factors, which regulate other flowering related proteins. Among the genes showing downregulation towards bud burst and inflorescence maturation are transcription factors involved in circadian rhythm such as *VvGRP2A (GLYCIN RICH PROTEIN 2A), VvRVE (REVEILLE), VvTICb (TIME FOR COFFEE)* and *VvELF3 (EARLY FLOWERING 3)*. Moreover, genes coding for transcription factors involved in GA biosynthesis were found to be upregulated during bud dormancy. *VvGAMYB* and *VvFL* show a higher expression in late flowering 'Villard Blanc' than in early flowering GF.GA-47-42 towards inflorescence development. This might indicate that floral development is delayed in 'Villard Blanc' in comparison to GF.GA-47-42. Numerous other genes, such as *VvHUA2*, involved in the repression of floral transition and flower development were found to be upregulated during bud dormancy. For most of the differentially expressed genes an up- or downregulation in expression is observed between the first and second time point during bud dormancy (Fig. 29). Many genes also show an up- or downregulation in expression between the third and the fourth time point when swelling buds are developing.

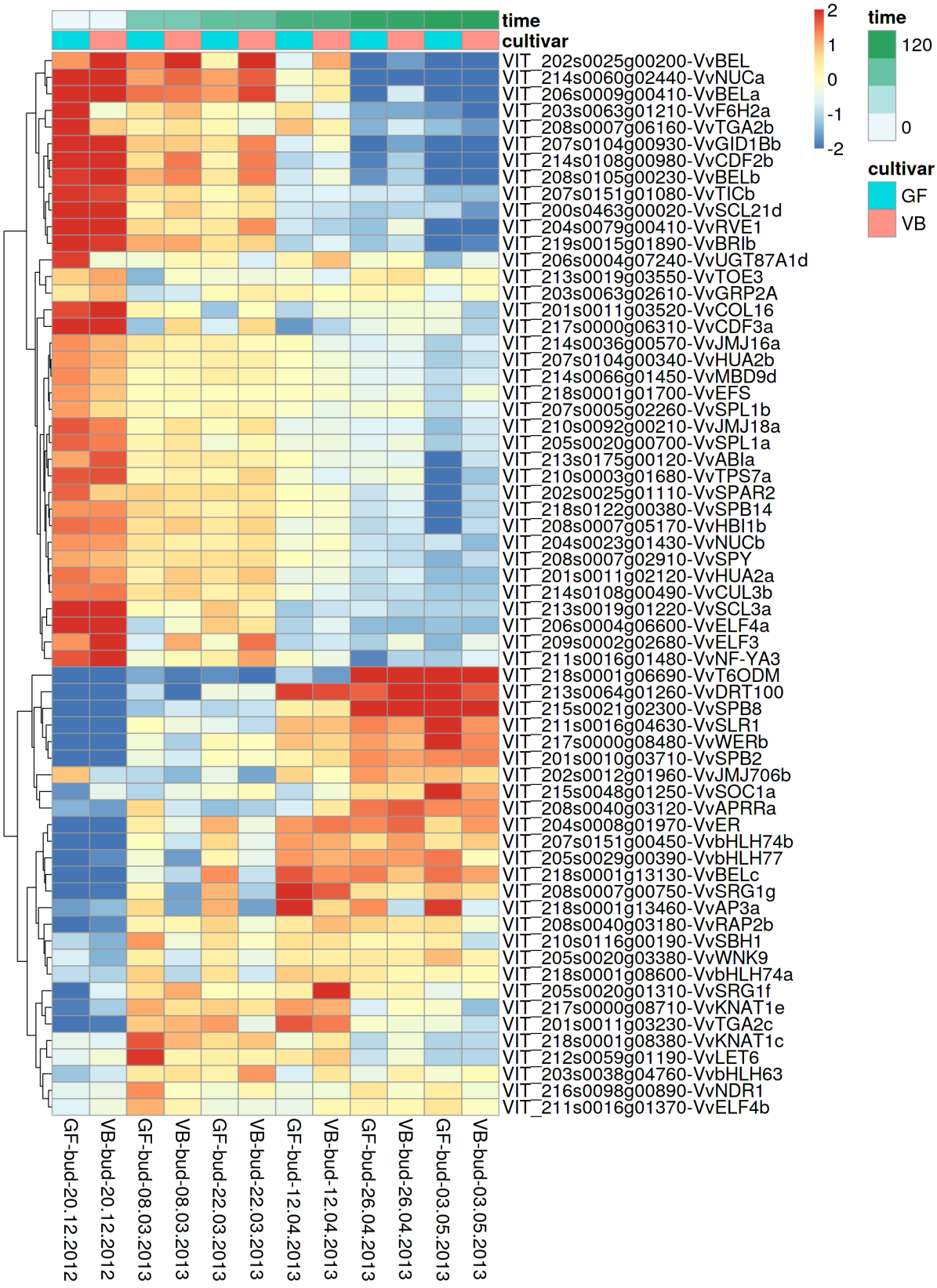

Figure 29: Heatmap of FTC candidate genes showing differential expression over consecutive timepoints of bud development from dormancy until before inflorescence development in both GF.GA-47-42 and 'Villard Blanc'. GF: GF.GA-47-42, VB: 'Villard Blanc'. Time: 20.12.2012 – 3.5.2013. LFC-threshold: 2 = expression fourfolded , -2 = expression quartered. r-log transformed counts are shown.

5.5.2.3 Gene expression in leaves corresponding to buds/inflorescences

For some bud and inflorescence samples, corresponding leaves in the axil of which the buds were formed, were collected and differential gene expression was analyzed between leaves and the associated bud/ inflorescence. Figure 30 shows a heatmap of the FTC candidate genes differentially expressed between leaves and buds/infloresences. The CONSTANS-LIKE genes *VvCOL16, VvCOL5*, and *VvCOL4* were upregulated in leaves. Other than this, most genes differentially expressed between leaves, and buds or inflorescences are downregulated or not expressed in leaves. All three B-class floral homeotic genes - *VvAP3a, VvTM6 (VvAP3b)*, and *VvPI* - are differentially expressed between leaves and inflorescences (Fig. 30). They are highly expressed in inflorescences while also absent in buds (Fig. 31).

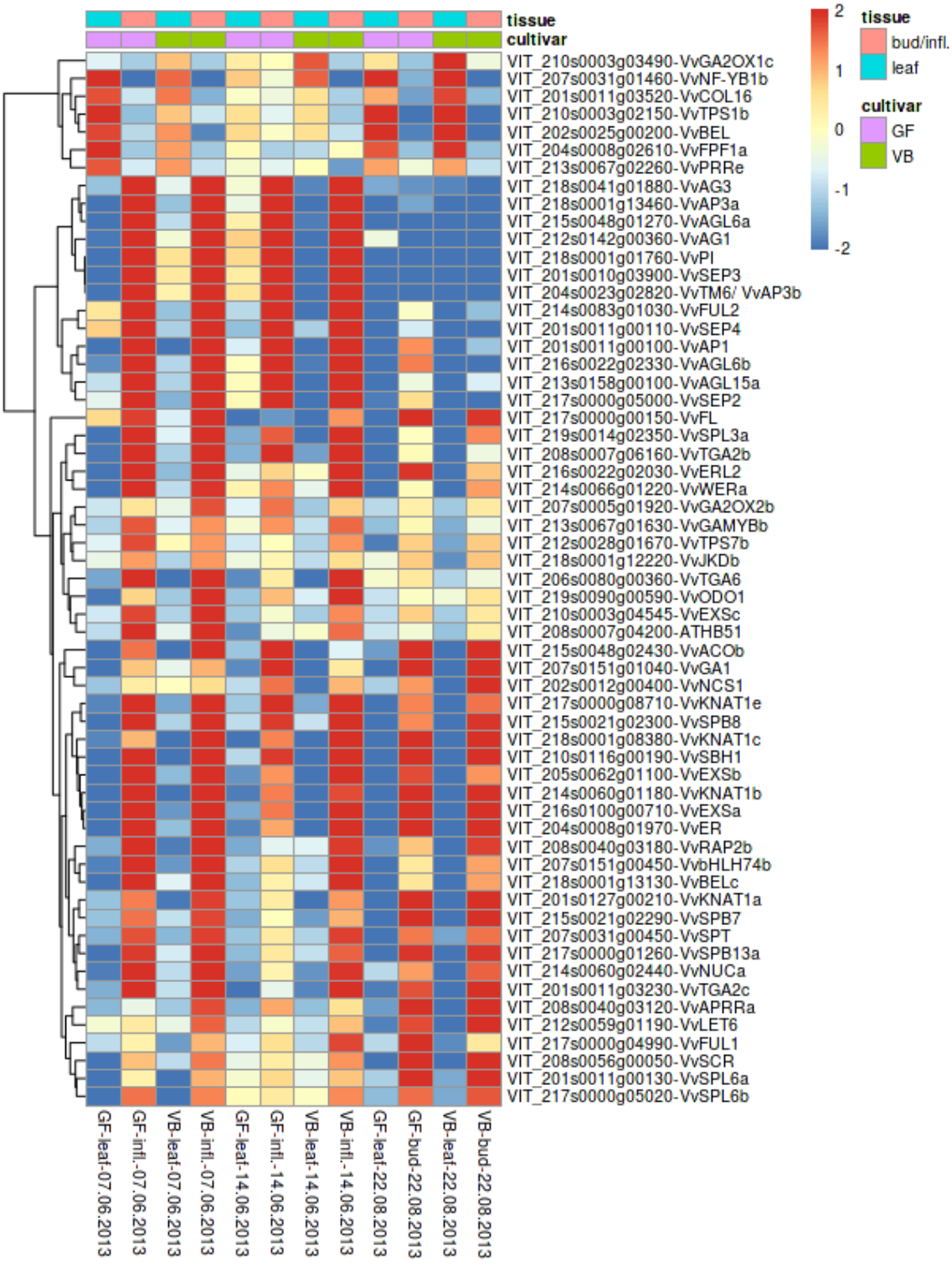

Figure 30: Heatmap of FTC candidate genes showing differential expression between leaves and associated buds/inflorescences in both GF.GA-47-42 and 'Villard Blanc'. GF: GF.GA-47-42, VB: 'Villard Blanc'. Time: 20.12.2012 – 3.5.2013. LFC-threshold: 2 = expression fourfolded , -2 = expression quartered. r-log transformed counts are shown.

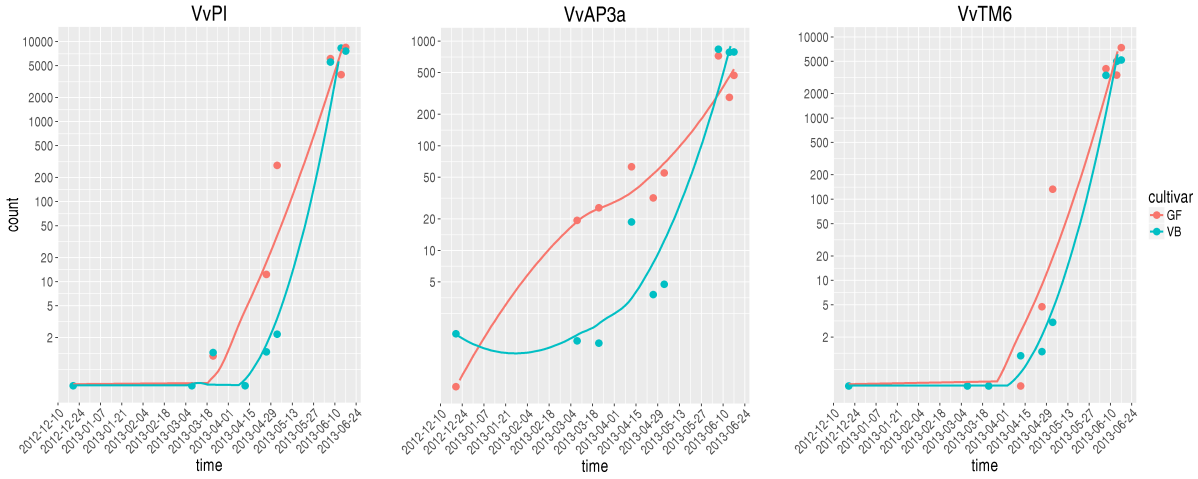

Figure 31: Expression profile of the three B-class floral homeotic genes *VvAP3*, *VvTM6* (or *VvAP3b*) and *VvPI* over consecutive developmental stages of bud- and inflorescence development in both GF.GA-47-42 and 'Villard Blanc'. GF: GF.GA-47-42, VB: 'Villard Blanc'.

5.5.2.4 Gene expression in amplicon sequenced FTC candidate genes

A closer look at the amplicon sequenced genes reveals that up- or downregulation in gene expression mainly occurs when swelling buds develop towards inflorescence maturation (between 22.03 and 26.04.2013) (Fig. 32). In developing buds before dormancy (Fig. 32, left part) obvious changes in gene expression are not observed. Genes involved in floral development, such as *VvSEP3 (SEPALLATA 3)* and *VvSEP4*, *VvAP1*, and *VvTM6* show an increased expression in developing inflorescences. Some genes were found not to be expressed in buds and inflorescences at all, such as *VvCOL2* and *VvGAIa* while some are only expressed before dormancy or in inflorescence tissue. *VvMFT*, of which an allele from 'Villard Blanc' correlates with late flowering, is not expressed in buds or inflorescences in neither of the cultivars.

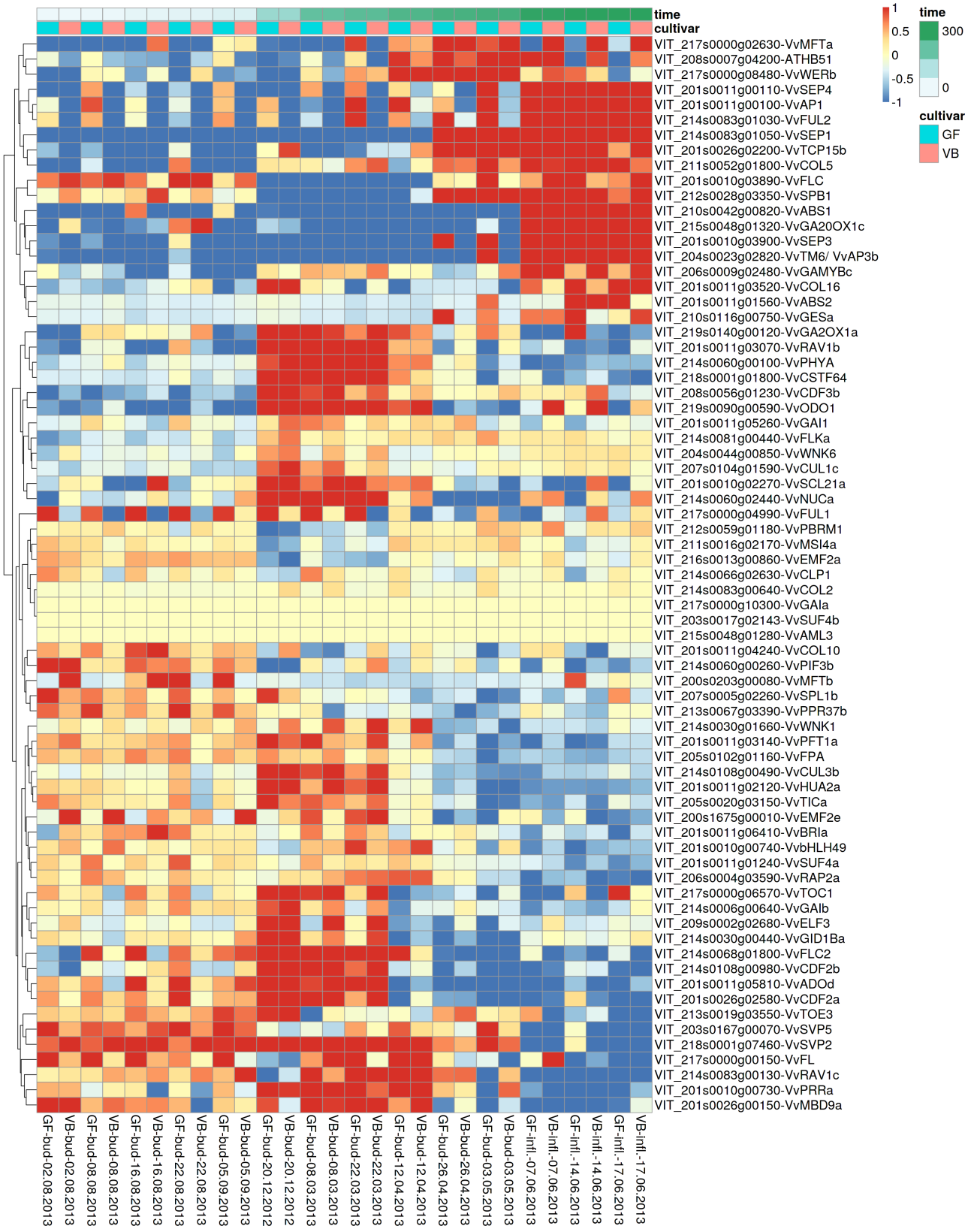

Figure 32: Heatmap of gene expression of amplicon sequenced FTC candidate genes in both GF.GA-47-42 and 'Villard Blanc' in different developmental stages of buds and inflorescences. LFC-threshold: 1 = expression doubled, -1 = expression halved. r-log transformed counts are shown. GF: GF.GA-47-42, VB: 'Villard Blanc'.

6 Discussion

6.1 Identification of more than four hundred flowering time control candidate genes in grapevine

More than four hundred FTC candidate genes were identified in the grapevine genome by using functional data from *A. thaliana* and defining homologues in the grapevine genome and by identifying candidate genes from previously defined QTL-regions for FTC. Chromosomes with previously identified QTL-regions involved in the control of flowering time were found to contain more candidate genes than other chromosomes. Several *Vitis* genes revealed more than one (up to eight) orthologues in *A. thaliana*. Remarkably, many SRF-proteins were found in grapevine. SRF proteins participate in cell cycle regulation, apoptosis, cell growth, and cell differentiation (Norman *et al.* 1988).

6.2 Investigating allele combinations of flowering time control candidate genes

In order to identify alleles of FTC candidate genes that are involved in the control of flowering time, a genetic association study between alleles of several of these genes and flowering time phenotypes was performed. Since haplotype based approaches are more informative and powerful in clarifying genotype-phenotype association, candidate genes were amplicon sequenced and a bioinformatic workflow for the phasing of alleles in amplicon sequenced genes was established. A genetic association analysis was subsequently performed in individuals of the pseudo F1-mapping population GF.GA-47-42 x 'Villard Blanc'.

6.2.1 Amplicon sequencing achieves a high level of targeting but suffers from technical difficulties

Targeted sequencing, such as amplicon sequencing, is a cost-effective method that enables the focus on investigating genomic regions that are likely to be involved in a particular phenotype (Lee *et al.* 2017). Nevertheless, the amplicon sequencing approach also suffered several technical difficulties, so that genotyping errors occurred due to the non-amplification of alleles or amplicons or strongly biased allele frequencies. Amplicon sequencing is prone to the typical problems of PCR-based approaches, like the incorrect determination of homopolymer lengths, nucleotide misincorporation, chimera formation, and allele drop out (Schirmer *et al.* 2015; Zucca *et al.* 2016).

Several problems can occur during sequencing of low diversity libraries, such as amplicon sequencing libraries. NGS-technologies are based on array formation. Clusters are generated during sequencing when the sequencing templates are amplified by emulsion PCR or solid-phase amplification (Lee *et al.* 2017; Schirmer *et al.* 2015). However, the amplification and imaging processes during sequencing can generate erroneous reads. During sequencing, reads are extended by one base per cycle. At each cycle, single molecules within a cluster may loose synchronity with each other. On the one hand, reads may fall a base behind (phasing) when the 3'-terminators are not completely removed. On the other hand, when the synthesis is too fast reads may run a base ahead (pre-phasing). The number of phased or pre-phased reads increases with sequencing, which generates noise and can result in higher error rates, especially towards the end of the reads (Cruaud *et al.* 2017).

To increase speed and reduce computational cost, Illumina's real time analysis software (RTA) uses images from the first four sequencing cycles to generate templates which are later used for the determination of cluster positions. In low-diversity samples, cluster recognition is unfortunately most difficult in the initial cycles. Thus, problems with overall data quality in Illumina sequencing of low sequence diversity samples may arise (Mitra *et al.* 2015). The uneven distribution of color intensities in low diversity samples is likely to cause a phasing problem. In order to overcome this issue, libraries are spiked in with a high amount of a higher diversity sample such as PhiX. This increases the diversity at the start of a read and leads to a more even distribution of color intensities.

Another factor influencing amplicon sequencing approaches, is the choice of primers. For haplotype phasing through an amplicon sequencing approach, not only is knowledge of the genomic region required, but primers for the amplification of a certain region also need to be unique. Non-unique primers lead to the

amplification of additional fragments; the sequence information is then falsified and genotyping errors arise. Not all primers used within this work were unique which might have been an error source. In some cases more than four alleles were found within the population GF.GA-47-42 x 'Villard Blanc' and their offspring or alleles were found, that were not detected in the parental lines. This happens when a variant is detected in an individual while not being detected in another due to a low allele frequency. Another cause is when primers are non-unique and more than two different sequences are found at a locus in a sample of one individual due to the amplification of additional fragments. DNA contamination can also lead to this phenomenon. However, it is not assumed that DNA contamination has occurred during sample preparation. Moreover, some amplimers could not be amplified in all cultivars (appendix table A5 and A6). This is likely due to a high diversity at the primer binding sites between the reference sequence and the affected cultivar. Primers then can not bind to the template and fragments are not amplified or allele drop out occurs. This can even happen for 'Pinot Noir' and 'Pinot Noir précoce'. Even though the grapevine reference sequence originates from a 'Pinot Noir' vine, the herewith analyzed 'Pinot Noir' vine harbors dissimilarities to the reference sequence, as the plant used for the generation of the reference sequence has been self-crossed various times. Since grapevine is very heterozygous and selfing leads to a loss of a great amount of heterozygosity, sequences are changed roughly through selfing.

Allele drop out also leads to genotyping errors due to the missing allele. However, since working with an F1-population derived from heterozygous parental lines and by applying Mendelian constraints, several cases in which an individual appeared artificially homozygous for a locus were identified. It was possible to rescue genes of which one allele was not amplified in five cases. Correlation was then calculated despite the sequence of the missing allele being unknown. In contrast, when analyzing unrelated individuals, it is not possible by the applied method to identify cases in which an individual is heteroyzgous for a locus but one allele of the locus is not amplified.

6.2.2 Allele phasing was successful in 46 genes of which many exhibit up to four alleles

Alleles in genomic regions of 46 genes on 16 different chromosomes with a length of up to 8.3 kb were phased by the means of the established workflow. By analyzing the inheritage pattern within a family of parents and F1-individuals, it was shown that the inheritage of alleles of closely neighbored genes remains largely constant

(e.g. table 11). This is consistent with the expectation of detecting only very few recombination points since working with an F1-population. This constancy indicates the functionality of the established workflow.

The use of an F1-population was beneficial for the validation of the established method since this enabled the validation of inheritage patterns of alleles by applying Mendelian constraints. Many of the analyzed and phased genes (89%) had four or three different alleles within the population GF.GA-47-42 x 'Villard Blanc'. This is consistent with the highly heterozygous nature of grapevine and its high genetic diversity. For the gene *VvFL* - the *A. thaliana* orthologue of *LFY* - only one allele was found within the population. *LFY* was shown to be highly conserved among angiosperms (Frohlich *et al.* 2000; Moyroud *et al.* 2009). In grapevine *VvFL* has introns at positions conserved with those described in others species (Carmona *et al.* 2002; Frohlich *et al.* 1997). This might explain the low genetic diversity detected for *VvFL*.

During variant calling, a strong allele bias was detected in several regions. A strand bias in Illumina sequenced data is not unusual (Schirmer *et al.* 2016) but DNA shearing prior to sequencing can also lead to fluctuations between variant frequencies. This can be due to a bias in the pattern of DNA shearing that depend on certain DNA sequence context, the shearing method used or combinations of both (Byeon *et al.* 2016). An amplicon can only be fragmented into a limited amount of fragments which can result in a bias in the quantity of sequenced fragments. A bias in allele amplification leads to a bias in allele frequency. When the bias is so strong that variant frequency drops below the threshold set for variant detection, variants fail to be detected. To overcome this issue, the ploidy was increased to 12 during variant calling. This made allele phasing in the genes, *VvSCL21a, VvSPB1, VvSEP1*, and *VvMFT* possible. For these genes allele phasing had failed under the default ploidy of two for diploid organisms. Neither the choice of primers, the length of amplimer, or the density of variations in these genes are significantly different to other genes. The amount of reads covering a variant can vary from one variant to the next, the frequency of these variants can thus differ despite being amplified from the same allele. When dealing with extremely biased allele frequencies, this can lead to some variants being detected while others from the same allele remain undetected. When the frequency of a variant is extremely low, another variant, originating from the same allele, can have an even lower allele frequency and hence remain undetected. In such cases, allele phasing was unsuccessful.

The limits of allele phasing can be attributed to the presence and density of

heterozygous variations, of which a mapped read pair must encompass at least two for physical phasing. These limits define the intervals in which allele phasing is possible. Heterozygous genotypes can be linked together since each read or pair of reads is obtained from one haplotype. Heterozygous blocks up- and downstream of homozygous regions that are longer than the maximum length spanned by a read pair can not be connected and phasing fails. Amplicon sequencing was performed in paired-end mode. The use of paired-end sequencing is highly advantageous in haplotype or allele phasing as it covers alleles that are spaced at distances longer than the sequencing technology's read length. Currently, read length of high-throughput sequencing technologies is constantly increasing and technologies are evolving rapidly. Per-base sequencing costs have dropped significantly during the last decade. Nevertheless, second-generation sequencing remains limited by short read length (∼300 bases) (Scheben *et al.* 2017). With the rise of third generation technologies such as the Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) many of the difficulties associated with haplotype phasing might soon be alleviated. These technologies are capable of producing even longer reads with long insert sizes. Long read lengths may permit direct and accurate haplotype phasing over multiple kilobase distances from sequence reads (Aguiar *et al.* 2012; Laver *et al.* 2016).

6.2.3 Correlation between alleles of candidate genes and flowering time phenotypes in- and outside of QTL-regions

A correlation between alleles of FTC candidate genes and flowering time was found in QTL-regions on chromosomes 1, 4, 14, 17, and 18. This supports the role of these regions in flowering time control. In some cases the p-value of correlation is significant although the medians are equal or nearly equal. This is because the Wilcoxon Rank-Sum test is a rank sum test and not a median test. It ranks all of the observations from both groups and then sums the ranks from one of the groups and compares it with the expected rank sum. It is possible, in rare cases, for groups to have different rank sums and yet have equal or nearly equal medians.

Within the FTC related QTL-region on chromosome 1 the L1-allele inherited from 'Villard Blanc' was found to correlate with late flowering and the E1-allele from GF.GA-47-42 showed correlation with early flowering. Only in 2015, the E1-alleles of all genes in the QTL-region on chromosome 1 correlated with flowering time. For QTL-regions on chromosomes 4 and 14 alleles inherited from GF.GA-47-42 (E1 and E2 respectively) were found to correlate with early flowering. All analyzed F1-individuals carrying alleles correlating with flowering time from two of the QTL-regions on chromosomes 1, 4, and 14 or all three of

them were either intermediate-early, early, or very early flowering. One exception was a late flowering cultivar that carried all alleles correlating with early flowering from chromosomes 1 and 14 (though not from chromosome 4). This indicates possible epistatic effects of these QTL-regions. Complex quantitative traits, such as flowering time, have been shown to be influenced by epistatic effects (Mackay 2014). Epistatsis, or locus interaction, is a crucial component of the genetic architecture of biologically relevant traits (Verhoeven *et al.* 2010). The effect of a QTL can be masked by the allelic effect of another. The phenotype of one locus can be enhanced or suppressed by genotypes at another locus. Hence, the effect of one locus might depend on the genotype at the interacting locus. The correlation for the QTL-regions on chromosomes 4 and 14 was more stable across the analyzed years than for chromosome 1 indicating a stronger affect of these QTLs in the timing of flowering.

6.2.4 Correlation values vary in dependence on phenotypic data

In grapevine, the timing of flowering and the length of the flowering period highly depend on the environmental conditions of the respective and the previous year. The timing of flowering in the analyzed population varied for each year and occurred between May and June depending on the temperature reached (Table 1, Fig. 33). Deviations in correlation values can be observed across the years and data sets used in dependence on the length and timing of the flowering period. A striking example of how the environmental conditions influence the timing of flowering is the year 2010, where first members of the mapping population GF.GA-47-42 x 'Villard Blanc' reached full bloom 151 days after January 1st while the last vines started flowering 180 days after January 1st. The second day of flowering was 170 days after January 1st leading to an unusual long flowering period. The 29th of May 2010 was the warmest day in that week, the following day was a rainy day followed by a drop in the average temperature 20 cm above ground by about 5°C[1]. Data from the year 2010 is thus likely not to be representative and would explain the high deviations for the correlation values of 2010 in comparison to other years.

[1]http://www.wetter.rlp.de

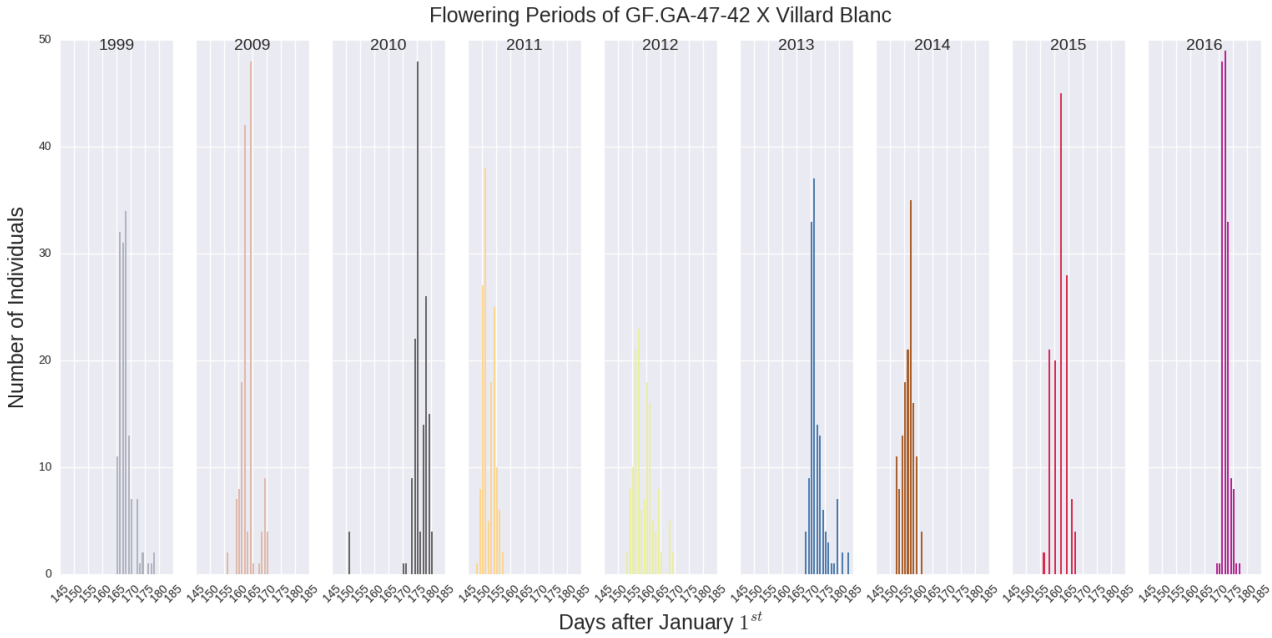

Figure 33: Flowering periods and the timing of flowering for individuals of the population GF.GA-47-42 x 'Villard Blanc' in the years 1999 and 2009 - 2016 in days after January 1st.

In the year 2016, flowering in the population GF.GA-47-42 x 'Villard Blanc' started on June 19th being very late compared to other years (Table 1). Correlation values for this year are also remarkably different from other years. Since full bloom occured very late in the season, the period between flowering initiation of the earliest and latest cultivar in the population was shorter than usual (10 days). Hence, cultivars that usually initiate flowering later, appeared to initiate flowering early. In warmer climates a higher variability in the duration of bloom between closely related cultivars within the same vineyard is observed (Gadoury *et al.* 2012). Global radiation in 2016 was distributed between approximately 502 and 536 KWh/m^2 at the beginning of the flowering period and between approximately 548 and 597 KWh/m^2 at the end of it. Figure 34 shows the global radiation and days until flowering from January 1st in the population GF.GA-47-42 x 'Villard Blanc' in 2016. While flowering occurred very late compared to other years, the amount of global radiation and solar radiation until the first day of the flowering period were smaller compared to other years. Flowering in 2016 was hence likely to have occurred very late since the amount of global and solar radiation until the beginning of the flowering period was comparably small.

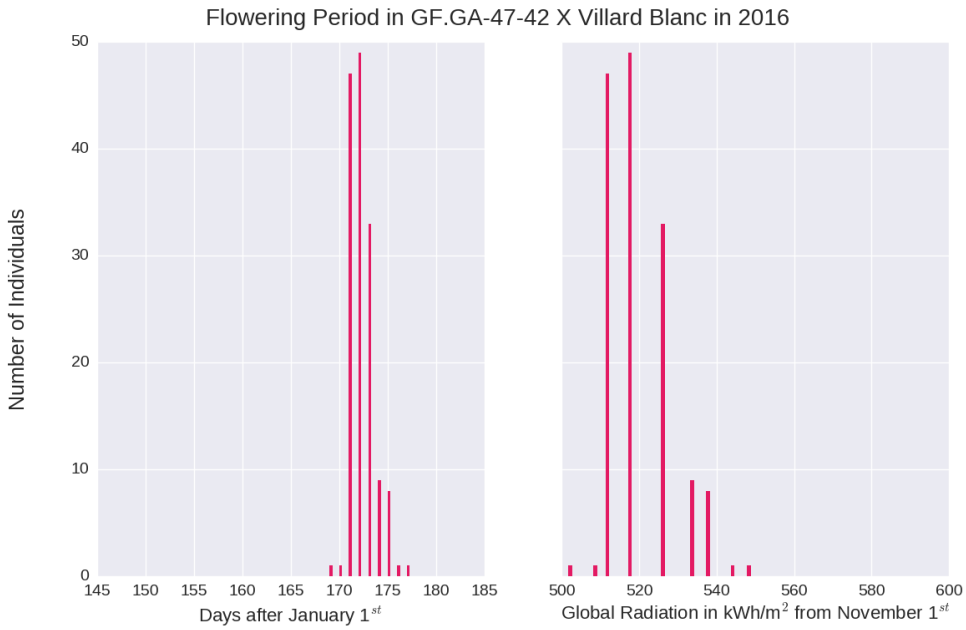

Figure 34: Days until flowering after January 1st, accumulated temperature above 3°C and global radiation in KWh/m^2 in 2016 in the population GF.GA-47-42 x 'Villard Blanc'.

The greatest portion of individuals within the population GF.GA-47-42 x 'Villard Blanc' reach full bloom in approximately the first third of the flowering period. Flowering time phenotypes within the population are thus not normally distributed throughout the flowering period. For correlation analysis between alleles of FTC candidate genes and flowering time phenotypes, a continuous variable for the timing of flowering was used. A Wilcoxon Rank-Sum test was applied because it is robust to non-normally distributed data. Nevertheless, the bias towards early flowering might still have an influence on the resulting correlation values. The number of individuals in flowering classes 5 and 6 was smaller than in other classes. Overall, a bias towards early flowering was observed (Fig. 35). A selection towards early flowering individuals in plant species may become generally stronger with the tendency to higher temperatures at earlier dates in the reproductive season in parts of the northern hemisphere (Munguía-Rosas *et al.* 2011).

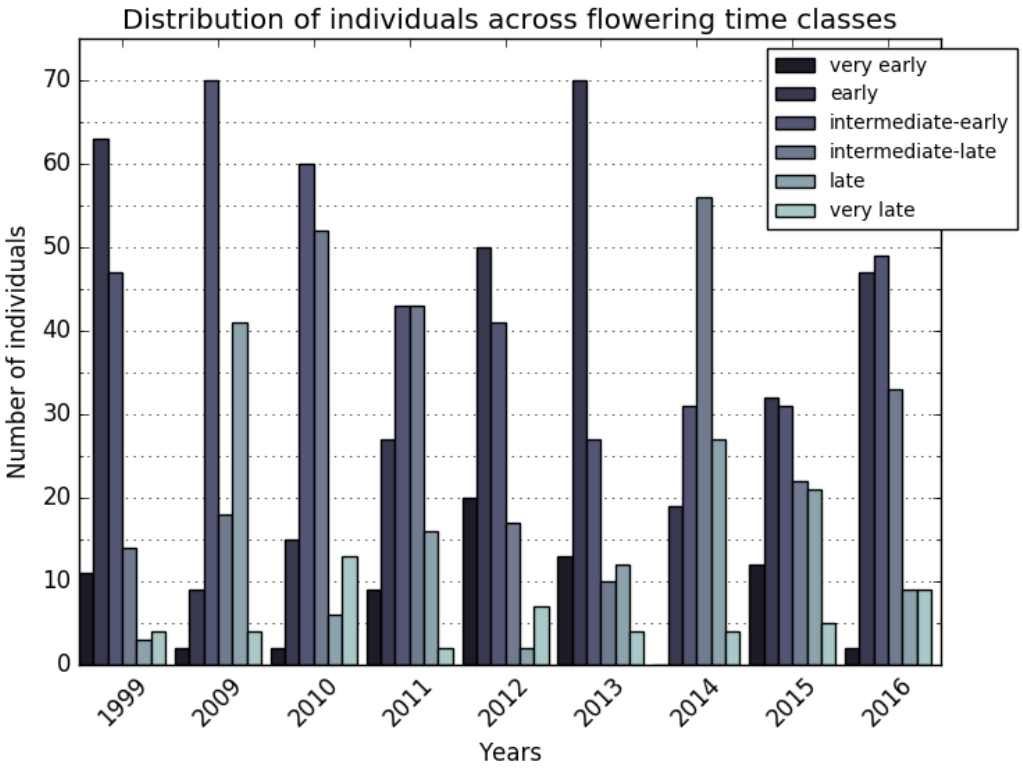

Figure 35: Distribution of individuals of the mapping population GF.GA-47-42 x 'Villard Blanc' across flowering classes in the years 1999 and 2009 - 2016 showing a bias towards early flowering.

6.2.5 Phenotyping influences genetic association analysis

Correlation analyses were performed using days after January 1st as well as temperature sums and global radiation from November 1st of the previous year. Phenotyping was performed in days after January 1st and values were then translated into temperature sums and global radiation. Consequently, the scale for the latter two was not finer than for days after January 1st. Remarkably, both the temperature sum and global radiation revealed the same correlation values in all tested cases (appendix table A10).

The outcome of the conducted genetic association study between alleles of FTC genes and flowering time phentoypes is on one hand based on the genotypic data, which is obtained through the allele phasing workflow, from amplicon sequencing, mapping, and variant calling to the creation of allele specific sequences. On the

other hand, it is highly dependent on the phenotypic data, which is prone to a number of errors. Firstly, it is generally prone to human errors, like mixing up samples or skipping a data point. Also, phenotyping of quantitative traits such as full bloom is to a certain degree dependent on the opinion of the person carrying it out. It is therefore recommended not to share this task among different people. However, digital technologies for the phenotyping of crop plants including grapevine have been arising recently (Kicherer *et al.* 2017; Klodt *et al.* 2015). This might meet the need for high-throughput and objective phenotyping in future grapevine research. Moreover, four vines are available at the vineyard of Geilweilerhof for every analyzed F1-individual. The possibility that one or the other vine has been mistakenly declared as a certain breeding line can not be excluded. Finally, since phenotyping of flowering time was performed on a daily basis, differences in the timing of flowering shorter than one day were not recorded.

6.2.6 E1-alleles of *VvTM6* and *VvWNK6* correlate with an early flowering phenotype

From the amplicon sequenced and early flowering individuals (according to the median) of the population GF.GA-47-42 x 'Villard Blanc', 90% were found to carry the *VvTM6*-E1-allele inherited from GF.GA-47-42. Only 10% of plants that carry the other maternal allele are early flowering. *VvTM6* is a MADS-box B-class floral identity gene and influences the development of petals and stamen. Mutants exhibit a transformation of petals to sepals and stamen to carpels. B-class floral homeotic genes either belong to the paleoAPETALA3 or to the PISTILLATA gene lineage which are paralogous and resulted from a duplication event before the emergence of angiosperms (Causier *et al.* 2010). The paleoAP3 lineage underwent a further duplication event at the base of the core eudicts resulting in the two sublineages euAP3 and TM6 (after the Tomato MADS-box gene 6) (Kramer *et al.* 1998). *TM6* is absent in *A. thaliana* (Causier *et al.* 2010; Kramer *et al.* 1998; Poupin *et al.* 2007). In Solanaceae, euAP3 might play a more direct role in petal development while TM6 might be more involved in stamen differentiation (Diaz-Riquelme *et al.* 2009). In *Vitis* all three B-class floral homeotic genes are highly expressed in inflorescences (Fig. 31) but not in leaves (Fig. 30). Diaz-Riquelme *et al.* (2009) showed that *VvTM6 (VvAP3.2)* is expressed in fruits, while the expression of *VvAP3 (VvAP3.1)* and *VvPI* is more restricted to flowers. Also, Poupin *et al.* (2007) showed that the expression of *VvTM6* is higher in carpels, fruits, and seeds than in petals. Due to the expression of *VvTM6* in carpels and during berry development and ripening, it is suggested that it plays an important role in grapevine fruit development (Diaz-Riquelme *et al.* 2009). The

expression of *VvTM6* increases towards inflorescence maturation which is followed by berry formation and ripening. This is consistent with its role in during berry development and ripening.

According to the median, all early flowering amplicon sequenced individuals of the population GF.GA-47-42 x 'Villard Blanc' were observed to carry the E1-allele of *VvWNK6* (Fig. 18). In *A. thaliana WNK6* has been shown to be involved in circadian rhythm (Nakamichi *et al.* 2002). WNKs are a subfamily of serine/threonine protein kinases, with a lysine residue, which is essential for ATP-binding, in kinase subdomain I instead of subdomain II as common among all other kinases (Hong-Hermesdorf *et al.* 2006). It has been suggested that the *WNK* gene family regulates flowering time in *A. thaliana* by modulating the photoperiod pathway. For instance, APRR3, a component of the clock-associated APRR1/TOC1 quintet is a substrate of WNK1 in *A. thaliana*. T-DNA knockout mutants of *AtWNK1* are delayed in flowering time while T-DNA knockout mutants of *AtWNK2, 5*, and *8* flower early (Wang *et al.* 2008). Moreover, it has been shown that *WNK6* transcription is downregulated in *AtABI4* mutants, that show an early flowering phenotype (Foyer *et al.* 2012). In *A. thaliana*, ABI4 negatively regulates flowering through directly promoting *FLC* transcription, a negative regulator of flowering (Shu *et al.* 2016). This might indicate that *VvWNK6* is involved in the delay of flowering. *VvWNK6* expression was detected in leaves, buds, and inflorescences of both the early and late flowering cultivar and no significant differences in the expression of *VvWNK6* were detected between late flowering 'Villard Blanc' and early GF.GA-47-42 that was found to be harboring a heterozygous variation leading to a premature stop codon. However, all individuals of the mapping population harboring a premature stop codon in one allele of *VvWNK6* (the E1-allele) flower early. This might be further evidence of the role of *VvWNK6* in the delay of flowering.

6.3 Variations in flowering time control candidate genes across cultivars

Variants in different grapevine cultivars were detected with a density between 0.79 and 2 variants per 100 bp, most of which were SNPs (Fig. 11). The highest variant density as well as the highest density of homozygous variations was identified in the cultivar 'Börner'. 'Börner' is derived from a crossing of cultivars belonging to the species *Vitis riparia* and *Vitis cinerea* and hence does not belong to the species *Vitis vinifera*. This explains the great amount of variability it displays to the *Vitis vinifera* derived reference sequence. 'Chardonnay' had the lowest

amount of variants and also the smallest density of homozygous and the second smallest density of heterozygous variants. This is a result of 'Chardonnay' being an offspring of 'Pinot Noir', a cultivated clone of which the grapevine reference genome sequence is derived from. 'Pinot Noir' also had a low genetic diversity in the analyzed amplicon data. Moreover, 'Pinot Noir précoce' is a result of a 'Pinot Noir' mutation, which explains its low variant density. However, the overall density of heterozygous variants in the analyzed panel of cultivars was higher than the density of homozygous variants, which refers to the heterozygous nature of grapevine.

GF.GA-47-42 was found to have a higher variant density than 'Villard Blanc'. From their progeny some have a higher and others a lower variant density than both of the parental line. Some have a variant density higher than 'Villard Blanc' but lower than GF.GA-47-42 (Fig. 11). The variant density in this F1-individuals can be be higher than in both parental lines. This can happen when the parental lines are heterozygous for different loci and the F1-individual inherits the loci alternate to the reference to an amount higher than both parental lines. Conversely, when an F1-individual inherits more loci identical to the reference, the result will be a lower density than in both parental lines.

Amplicon sequences were not equally available for all cultivars (Fig. 9). Although the calculated variant density refers to sequenced regions, the differences in the amount of sequenced amplicons between cultivars can make the variant densities less comparable. Especially in cases where amplification failed because primers were unable to bind to the DNA template due to a high variability in primer binding regions, variant density is likely to be higher than stated. However, the cultivar 'Börner' exhibits the highest variant density and 'Chardonnay' the lowest, which meets the expectations.

6.3.1 The L1-allele of *VvTM6* putatively correlates with late flowering

For 36 genes up to four allelic sequences that were common between several cultivars were found (appendix table A12). Some alleles were common between many cultivars, such as a detected sequence for *VvMFT*, while others were only common between a few. Since most of the analyzed cultivars are related, it is not expected to detect a larger number of different alleles in the analyzed panel. The defined intervals in which alleles were phased vary in length and are only part of a genes's sequence. Therefore, differences might still be present in unphased regions.

An allele of *VvTM6* (L1) was found in all four late flowering individuals and might hence correlate with late flowering. Nevertheless, the number of analyzed cultivars might be too small to detect a significant effect. In order to further analyze the role of this allele in late flowering and to a perform a correlation analysis, a larger panel of cultivars should be analyzed. Also, flowering time phenotypes in a cultivar of the analyzed panel is relative to the other cultivars in the panel. In the mapping population GF.GA-47-42 x 'Villard Blanc' a correlation for *VvTM6*-L1 with the timing of flowering was not detected. However, the number of late flowering individuals was limited to 7% on average.

Crossing cultivars in order to breed new varieties with new or enhanced qualities is very common in grapevine breeding. Since grapevine is highly heterozygous, crossing cultivars contributes to a high genetic diversity across varieties of the species. It is hence not expected to detect high levels of homozygozity in the analyzed grapevine cultivars. Nevertheless, for several FTC candidate genes, homozygous regions were found in several of the cultivars.

Due to epistatic effects between interacting loci, it is not unlikely to detect an allele that is associated with early flowering in a late flowering cultivar. For example, is the E1-allele of *VvWNK6*, that is correlated with early flowering, also found in the grapevine cultivar GF.GA-52-42. GF.GA-52-42 shares a parental line with early flowering GF.GA-47-42 and flowers late (Fig. 22). Alleles correlating with early flowering of *VvBRIa* and *VvCOL10* and with late flowering of *VvBS2* were detected in other cultivar (section 5.4). Since flowering time phenotypes of these cultivars were diverse, assumptions towards an association with the timing of flowering can not be made. When performing QTL-analyses, more precise testing for interactions should be performed since genotype-specific QTL-effects will not be observable in a different genetic background. To test the effects of an allele, it can be introgressed into several different wild-type genetic backgrounds. Typically, the expression of the mutant will be enhanced or suppressed, and the degree of dominance and pleiotropic effects on other traits can be observed. Moreover, the regulatory role of micro RNAs might have an influence of the timing of flowering, as previously shown in *A. thaliana* (Spanudakis *et al.* 2014).

6.4 Functional allele specific molecular markers were developed

The most prevalently used markers for genotyping in plant population genetic studies are SSR markers (Guichoux *et al.* 2011). SSRs, which are short repeated DNA motifs of 1-10 bp, are multi-allelic, highly informative, and codominant genetic markers(Vieira *et al.* 2016). Addition or deletion caused by polymerase strand-slippage in DNA replication or by recombination errors leads to repeat polymorphisms and hence different individuals exhibit variations in the number of repeats of the SSR motif (Vieira *et al.* 2016). The variation of the SSR repeat unit can be assessed through the PCR amplification with primers complementary to the region flanking the SSR motif (Huang *et al.* 2011).

The availability of the complete grape genome in 2007 (Jaillon *et al.* 2007) allowed access to thousands of SSRs (Cipriani *et al.* 2010). Nevertheless, although SSRs are the most prevalently applied markers for parentage analyses, the number of independently segregating markers usable for genetic analyses is limited (Glaubitz *et al.* 2003). Another disadvantage of SSRs is that their occurrence is generally low in gene regions, since they have a high mutation rate that could affect gene expression (Vieira *et al.* 2016). If candidate genes are available, SNP markers are informative for both parentage analysis and the investigation of the function of these genes. They bring along a wide range of advantages since they are plentiful in most genomes, easily accessible to high-throughput genotyping and their individual mutation rates are lower than those of SSR markers. Moreover, they can be developed with less effort and costs compared to SSR markers. Unfortunately, each SNP marker is typically bi-allelic and allele frequencies can be drastically unequal which results in a very low resolving power for a marker (Jones *et al.* 2009). In order to distinguish the maximum putative number of alleles at a single locus within a bi-parental F1-population of a diploid organism, the marker needs to be capable of distinguishing between four alleles. The usage of blocks of tightly linked polymorphisms and treating each haplotype of these blocks as a separate allele can produce highly polymorphic markers with properties similar to SSR markers (Guichoux *et al.* 2011). This has, for example, been reported previously by Jones *et al.* (2009) and Glaubitz *et al.* (2003).

Amplicon sequencing was performed in 35 F1-individuals and the parents of the mapping population GF.GA-47-42 x 'Villard Blanc'. In order to investigate the resulting allele distributions over all 151 heterozygous F1-individuals of the mapping population, markers that distinguish between the maximum number of different alleles, were designed. The established method has proven to deliver

functional molecular markers for parentage analysis suitable for analyzing polymorphisms located in genomic regions and exhibiting a functional significance. The information obtained from amplicon sequencing of the candidate genes proved usable for both deduction of segregation patterns and marker design. Correlation values obtained by performing a correlation analysis between alleles of FTC candidate genes and flowering time phenotypes were largely supported by marker analysis. However, the method used for the measurement of product sizes can not distinguish between differences in product sizes of up to two bp. Hence, the expected product size on the basis of the phased allelic sequences and the observed product size from marker analysis can differ to up to two bp. In addition, during variant filtration, variants with low coverage or quality were filtered out. In the case, that InDels were filtered out, the actual allele sequence length can be greater than the calculated one. Compared to marker analysis, the established allele phasing workflow has the advantage that it distinguishes between alleles of different sequences and equal length.

6.5 Differential gene expression analysis of FTC candidate genes indicates their role in the timing of flowering in grapevine

Many of the analyzed FTC candidate genes show differential expression over time. This strengthens their role in flowering time control. Transcripts coding for transcription factors and other proteins involved in inflorescence architecture, floral transition and flower development are usually upregulated after bud burst, while transcripts coding for proteins that repress flowering in diverse manners are usually upregulated during bud dormancy (Fig. 29).

Genes involved in circadian rhythm such as *VvGRP2A, VvRVE, VvTICb* and *VvELF3* were found to be downregulated towards bud burst and inflorescence maturation. *VvGRP2A* is highly expressed in developing buds with a peak during bud dormancy and falling expression towards infloresence maturation. The homologue of *GRP2A* in *A. thaliana* - *AtGRP7* - undergoes circadian oscillations with peak levels in the evening (Heintzen *et al.* 1997). *RVE* is a MYB-like transcription factor that controls auxin levels, promotes free auxin and hence plant growth during the day (Rawat *et al.* 2009). *TIC* and *ELF3* are components of the circadian clock in *A. thaliana*. *ELF3* is a circadian clock gene that contributes to photoperiod dependent flowering in plants (Boden *et al.* 2014; Ding *et al.* 2007; Hall 2003).

Genes coding for transcription factors involved in GA biosynthesis were found to be upregulated during bud dormancy. GAs are inhibitors of flowering in many fruit species but their role in grapevine varies with the stage of bud development. The initiation and development of lateral meristems is promoted by GAs as well as their development into tendrils, while inflorescence development is suppressed by GAs. Thus, GA is a promoter of flowering at an early stage but acts as an inhibitor of flowering later on and promotes vegetative growth (Carmona *et al.* 2007). SPINDLY (SPY) - homologue to *VvSPY*, which is upregulated during bud dormancy - is a negative regulator of GA response in *A. thaliana* and functions with GIGANTEA (GI) in pathways controlling flowering (Tseng 2004). In *Vitis* the role of SPY in GA-signaling is still unclear. It has been shown that GA-treatment of grapevine plants leads to rachis elongation and a downregulation of SPY in the rachis (Tseng 2004). In *A. thaliana* GA-signaling is initiated through its binding to GA INSENSITIVE DWARF1 (GID1) receptors. This allows subsequent interaction between GID1 and DELLA proteins (GAI, REPRESSOR OF GA (RGA), RGA-LIKE 1 (RGL1), RGL2, and RGL3). In the presence of GA, the stable GID1-GA-DELLA complexes are recognized by the SCF^{SLY1} complex, which ubiquitylates the DELLA proteins and causes their degradation by the 26S-proteasome (Cheng *et al.* 2015; Fleet *et al.* 2005). It has been reported previously (Khalil-Ur-Rehman *et al.* 2017) that GID1-transcripts are upregulated during bud dormancy in grapevine while DELLA protein transcripts are downregulated. Here, it was found that the GID1B-receptor transcript is being upregulated during bud dormancy while the DELLA protein SLENDER RICE 1 LIKE (SLR1-like) is being downregulated. This confirms the promoting role of GID1B in plant growth, and the initiation and development of lateral meristems in dormant buds and indicates that SLR1-like is responsive for the mediation of the suppression of inflorescence development through GA.

VvGAMYBa was found to be expressed in inflorescences while absent in buds. *GAMYB* regulates the expression of GA-regulated genes in cereal aleurone cells (Gubler 2002; Thomas 2004) and plays a role in anther development by regulating genes involved in the synthesis of sporopollenin components, and consequently contributes to Ubisch body formation in tapetal cells and exine formation in pollen (Aya *et al.* 2009; Murray *et al.* 2003). *GAMYB* may mediate a GA signaling role in flowering by its ability to bind to a specific 8-bp sequence in the promoter of *LFY*. This sequence is an important in the GA response of the *LFY* promoter (Gocal *et al.* 2001). The expression of the *Vitis LFY* orthologue *VvFL* has been reported to be correlated with inflorescence and flower development (Carmona *et al.* 2002). It is expressed in uncommitted primordia and later

in inflorescence- and tendril-meristems. The expression of *VvFL* reaches its highest level in the floral meristems after bud burst. This correlation might suggest that *VvFL* is required to reach a certain threshold in order to initiate reproductive development. In *A. thaliana* a threshold level of *LFY* expression has also been shown to promote flower initiation. It integrates flowering signals from different pathways and controls the expression of flower organ identity genes (Carmona *et al.* 2002). DELLA proteins have been shown to delay flowering in short-day photoperiods by repressing the upregulation of *LFY* and *SOC1* transcripts (Achard *et al.* 2007). Both *GA myeloblastosis (MYB) (GAMYB)* and *VvFL* show a higher expression in late flowering 'Villard Blanc' than in early flowering GF.GA-47-42 towards inflorescence maturation indicating that floral development is delayed in 'Villard Blanc' in comparison to GF.GA-47-42.

Numerous other genes involved in the repression of floral transition and flower development were found to be upregulated during bud dormancy. *ENHANCER OF AGAMOUS 2 (HUA2)-like* genes, that play a role in the repression of floral transition (Doyle *et al.* 2004), are upregulated during bud dormancy in *Vitis*. *KNOTTED1-like homeobox gene BREVIPEDICELLUS (BP)* was found to be upregulated towards bud burst and inflorescence maturation in *Vitis*. In *A. thaliana BP* controls distal pedicel growth and thus inflorescence architecture (Douglas *et al.* 2005; Venglat *et al.* 2002). *ERECTA (ER)* and other *KNOTTED-LIKE (KNAT)* genes, that in *A. thaliana* also play a role in inflorescence architecture (Douglas 2002; Shpak 2004), were also found to be upregulated towards bud burst. This indicates their function in inflorescence development. SQUAMOSA promoter-binding proteins, that are known to be involved in flower development (Yamasaki *et al.* 2006), were found to be downregulated during bud dormancy while upregulated during flower formation in grapevine. A *BEL*-like gene (*BLH1*) and the *Vitis SHOOT MERISTEMLESS (STM)* orthologue *VvSBH1* were also found to be upregulated during bud dormancy. *STM* and the *A. thaliana* homeobox-gene *BEL1* build a complex which maintains the indeterminacy of the inflorescence meristem (Bellaoui 2001).

The MADS-box B-class floral identity gene *AP3 (APETALA3)* influences the development of petals and stamen (Theissen *et al.* 2000). It was found to be upregulated after bud burst in grapevine. *RELATED TO APETALA 2 (RAP2)* shows a similar expression pattern. *AP2* and *RAP2* are transcription factors involved on flower development in *A. thaliana* (Okamuro *et al.* 1997) and their expression pattern in *Vitis* indicates a similar role.

WNK9 shows an overall higher expression in early flowering GF.GA-47-42

compared to late flowering 'Villard Blanc'. *WNK9* is the homologue of *A. thaliana WNK1.* In *A. thaliana WNK1* was shown to interact with ARABIDOPSIS PSEUDO RESPONSE REGULATOR 3 (APRR3) and phosphorylate the APRR3 component of the clock-associated APRR1/TOC1 quintet underlying clock-controlled circadian rhythms in plants (Nakamichi *et al.* 2002). A T-DNA knockout of *WNK1* had a much delayed flowering time (Wang *et al.* 2008) suggesting that *WNK1* promotes early flowering in *A. thaliana*. This indicates that *WNK9* might be involved in the promotion of flowering in grapevine.

MYC-transcription factors *VvbHLH74* and *VvbHLH63* are differentially expressed over time in GF.GA-47-42 with a peak in expression around March when buds are swelling. The *A. thaliana* homologue of *VvbHLH63* - *CRYPTOCHROME-INTERACTING basic-helix-loop-helix (CIB1)* - plays a role in CRYPTOCHROME 2 (CRY2)-dependent regulation of flowering time. CRYs are blue-light receptors that mediate light response. In yeast and *A. thaliana*, CIB1 interacts with CRY2 when blue-light is available. It promotes CRY2-dependent floral initiation together with additional CIB1-related proteins and stimulates *FT* transcripts (Liu *et al.* 2008). Hence, *VvbHLH74* and *VvbHLH63* might be involved in light dependent floral initiation in grapevine.

EARLY FLOWERING (ELF)-like genes as well as a constans-like gene (*COL16*) and *CYCLING DOF FACTORs (CDFs)* were upregulated during bud dormancy. DOF proteins delay flowering by repressing *CO* transcription (Fornara *et al.* 2009). *ELF3* and *ELF4*, as well as *TOC1* function in the primary, phytochrome-mediated light-input pathway to the circadian oscillator in *A. thaliana.* *TOC1* is necessary for light-induced *CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)/ LATE ELONGATED HYPOCOTYL (LHY)* expression (Doyle *et al.* 2002). *elf4* mutants show attenuated expression of *CCA1* and early flowering in non-inductive photoperiods, which is probably caused by elevated amounts of *CO*, a gene that promotes floral induction (Kikis *et al.* 2005). *ELF4* is a flowering pathway gene that may play a key role in signaling processes regulating dormancy induction in grapevine (Sreekantan *et al.* 2010).

METHYL-CPG-BINDING DOMAIN 9 (MBD9) is upregulated during bud dormancy in grapevine. Its *A. thaliana* homologue *AtMBD9* is related to the inhibition of flowering (Peng *et al.* 2006) and suggested to have a role in bud development through an interaction with *FLC* (Sreekantan *et al.* 2010). *SUPPRESSOR OF PHYA (SPA)* was found to be upregulated during bud dormancy and downregulated towards inflorescence development. It represses photomorphogenesis by targeting the photomorphogenesis promoting transcription

factor ELONGATED HYPOCOTYL 5 (HY5) for degradation (Huang *et al.* 2013; Srivastava *et al.* 2015). This indicates a similar flowering inhibiting role of *MBD9* and *SPA* in grapevine to their role in *A. thaliana*.

Many FTC candidate gene were found to be upregulated during dormancy or when swelling buds develop to mature inflorescences. These genes are mostly involved in floral development as either floral repressors or floral developmental genes. Floral repressors are upregulated during bud dormancy while floral developmental genes such as meristem formation and patterning genes are upregulated when flower formation takes place. Both analyzed cultivars - GF.GA-47-42 and 'Villard Blanc' - show similar expression patterns for most genes and are hence mutually supportive. This was beneficial considering that biological replicates were not available. Over the analyzed time course of consecutive developmental stages of bud development prior to dormancy in the summer of the first year differential expression was detected for only two genes in GF.GA-47-42 (*VvWNK5* and *VvTM8)* and for only one gene in 'Villard Blanc' (*VvLRR-RLK*). This shows that the expression of the analyzed FTC candidate genes remains largely constant over this time course.

Analysis of differential gene expression between the early and the late flowering cultivar was performed in order to detect genes with a time shifted expression between the cultivars. Differential expression between early flowering GF.GA-47.42 and late flowering 'Villard Blanc' was detected for only one gene (*VvLRR-RLK*) in the time course before dormancy. A time shift in the expression of FTC genes between GF.GA-47-42 and 'Villard Blanc' is likely to occur in a closer time window than the time period between samples of the analyzed time courses. Moreover, in order to investigate genes differentially expressed at certain time points between the early and the late flowering cultivar, the collection of biological replicates is an important aspect. For example, was *VvGAMYBa* found to be differentially expressed between the samples of GF.GA-47-42 and 'Villard Blanc' from June 17$^{\text{th}}$ 2013. The sample of GF.GA-47-42 from this time point was found to be an outlier with a great sample distance to all other samples. Thus, this results needs to be verified by the replication of the experiment or by conducting quantitative real time PCR experiments. The availability of biological replicates would compensate the presence of outliers.

7 Conclusion and Outlook

Identification of FTC candidate genes

Using a comparative genomics approach and by identifying candidate genes from previously defined FTC-QTL regions, more than four hundred FTC candidate genes were identified in the grapevine genome. This is a significantly higher amount of putative FTC genes compared to what is known from *A. thaliana* where about 180 FTC genes are identified so far. Several *Vitis* genes revealed more than one (up to eight) orthologues in *A. thaliana*. Further investigations should thus be conducted to determine the relevance of the identified FTC candidate genes in grapevine.

Association analysis of FTC candidate gene alleles with flowering time phenotypes

The established method for the analysis of the genetic association between alleles of FTC candidate genes and flowering time phenotypes in individuals of the F1-mapping population GF.GA-47-42 x 'Villard Blanc' has proven to be functional. Nevertheless, it suffered from several technical difficulties related to amplicon sequencing, allele phasing, and phenotyping. Working with an F1 population allowed the identification of technical issues like allele drop out and strongly biased allele frequencies. Moreover, the availability of the allelic sequences of F1-individuals was highly supportive for the confirmation of allele phasing results. By applying Mendelian rules of inheritance, it was possible to validate the established method for allele phasing. The number of alleles present at a single locus in an F1-population is limited to four and a parent-offspring pair must share one haplotype for each chromosome and thus one identical by descent allele for every gene. When sequencing individual cultivars instead of an F1-population there is no possibility to identify cases of allele drop out. Consequently, it is not possible to be sure if a homozygous locus is truly homozygous or affected from allele drop out. It was possible to overcome incorrect allele phasing due to biased allele frequencies by increasing the ploidy from the default value of two for diploid organisms to 12. Thus, for unrelated or single cultivars, it is recommendable to examine the data for biased allele frequencies and then use an increased ploidy for variant calling in amplicon sequenced data.

The genetic diversity of a population at a single locus is traditionally being assessed through SSR-markers, which are powerful due to high allelic diversity at a given locus. A disadvantage of SSRs is their low occurrence in gene regions. SNP-markers do not share this disadvantage. However, SNP-markers are bi-allelic and thus individually deliver little information about parentage. But using blocks of tightly linked SNPs and InDels and treating each haplotype as a separate allele can produce genotyping data with properties similar to those obtained from SSR marker, combining the advantages of both technologies. The herewith developed markers have proven to be functional molecular markers for parentage analysis. They can distinguish between the maximum putative number of alleles at a single locus within a bi-parental F1-population of a diploid organism and are suitable for analyzing polymorphisms located in genic regions and exhibiting a functional significance. For future research, the panel of cultivars that were analyzed in addition to the mapping population can be expanded and tested for correlating alleles using the developed markers.

In order to improve the efficiency of the established method, amplicon sequencing and subsequent allele phasing can be performed in only the parental lines of the population and, in addition, a few of F1-individuals. Amplicon sequencing of F1-individuals serves as control and confirmation of allele phasing results. The amplification and sequencing of amplicons in a smaller panel of individuals would be saving cost and effort. Afterwards marker analysis can be performed on the basis of the combinations of closely neighbored variations in a larger number of individuals of the population. The results of marker analysis can subsequently be used for correlation analysis.

Correlation between specified alleles of FTC candidate genes and the timing of flowering was found for several genes with some variation across the analyzed years. Genes within previously defined QTL-regions were found to correlate with flowering time phenotypes, which supports the role of these QTLs in the timing of flowering. Among others, alleles of *VvWNK6* and *VvTM6* inherited from early flowering GF.GA-47-42 were found to correlate with early flowering in individuals of the mapping population GF.GA-47-42 x 'Villard Blanc'. Marker analysis in the extended population confirmed and hence strengthen this result. Both genes are located in a previously defined flowering time QTL-region on chromosome 4. From the analyzed F1-individuals 86% (19/22) carry the E1-alleles of both genes. The investigation of allele combinations of further genes within and outside the QTL-region on chromosome 4 could further enlighten the role of this QTL in the timing of flowering.

Alleles of FTC candidate genes within a QTL region on chromosome 1 were found to rather correlate with late, while QTL-regions on chromosomes 4 and 14 were found to correlate with early flowering. With one exception, all analyzed F1-individuals carrying alleles correlating with flowering time from two of the QTL-regions on chromosomes 1, 4, and 14 or all three of them were either intermediate-early, early, or very early flowering. The correlation for the QTL-regions on chromosomes 4 and 14 was more stable than for chromosome 1 indicating a stronger affect of these QTLs in the timing of flowering. The investigation of epistatic effects between these QTL-regions could contribute to the clarification of the genetic factors that influence and control flowering time in grapevine. Subsequently, molecular markers for the selection of cultivars with a particular flowering time phenotype can be developed. These marker can be used to predict flowering time phenotypes of grapevine seedlings and would thus eliminate the need to wait for several years until first reproductive cycles are completed.

Phenotyping and environmental conditions play a tremendous role in genetic association studies of complex traits in crop plants grown in the field. Over the analyzed time period (1999 - 2016), a tendency to shorter flowering periods was observed. A certain amount of variation in the correlation between alleles of FTC candidate genes and the timing of flowering across the analyzed years was found. To overcome such fluctuations in the genotype/phentoype correlation between years in dependence on the environmental conditions, plants can be grown in greenhouses. Since the population GF.GA-47-42 x 'Villard Blanc' was established long before this study was conducted, growing in the greenhouse to exclude environmental conditions was not an option. Another possibility to overcome the difficulties resulting from the influence of environmental conditions, was to perform phenotyping over as many years as possible and perform a correlation analysis using the median.

FTC candidate genes in important grapevine cultivars

The frequency of variations in amplicon sequenced FTC genes varies across different grapevine cultivars. Cultivars with a genetic background closer to the reference cultivar, such as 'Pinot Noir', 'Pinot Noir précoce', and 'Chardonnay' showed the lowest variant frequency. 'Börner' showed the highest frequency of variations since it is not derived from *Vitis vinifera*. Common alleles of FTC candidate genes between different cultivars were found, which allows the tracking of the inheritage of alleles between related cultivars. Most alleles were not common between more than a few individuals since grapevine exhibits a high genetic diversity. Among others, an allele of *VvTM6* (L1) was found in all four late flowering individuals indicating a putative association with late flowering. However, the number of

analyzed cultivars might be too small to detect a significant effect. In order to further analyze the role of this allele in late flowering, a larger panel of cultivars should be analyzed.

Differential gene expression analysis of FTC candidate genes

Differential gene expression analysis gave indications for the confirmation of the role of many of the FTC candidate genes in the control of flowering time. The analysis of differential gene expression over consecutive stages of bud and inflorescence development, revealed significant differential expression for many FTC candidate genes in both early flowering GF.GA-47-42 and late flowering 'Villard Blanc'. A shift between an up- or downregulation in expression mostly occurred between dormant and swelling buds, or toward inflorescence maturation when inflorescence structures become visible. Moreover, many FTC candidate genes were found to be expressed in buds and inflorescences but not in leaves which further confirms their role in flowering time and floral development. Both analyzed cultivars - GF.GA-47-42 and 'Villard Blanc' - show similar expression patterns for most genes and are hence mutually supportive. This was beneficial considering that biological replicates were not available. Analysis of differential gene expression between the early and the late flowering cultivars was performed with the goal to find genes with a time shifted expression between the cultivars but could hardly be detected. A time shift in the expression of FTC genes between GF.GA-47-42 and 'Villard Blanc' is likely to occur in a closer time window than the time period between samples of the analyzed timeseries'. Genes upregulated during dormancy or when swelling buds develop to mature inflorescences are mostly involved in floral development as either floral repressors that are upregulated during bud dormancy or floral developmental genes such as meristem formation and patterning genes that are upregulated when flower formation takes place. Over the analyzed time course in the summer of the first year differential expression was detected for only a very few genes. It is likely that the expression of genes involved in the timing of flowering occurs even earlier in the developmental cycle. Differential gene expression should therefore be analyzed in a closer time window in the spring of the first year when inflorescence differentiation begins. It would be ideal to analyze gene expression between the early and the late flowering cultivar on a daily or nearly daily basis since a time shift in expression is expected to occur in a very close time window. Another important aspect is the collection of at least three biological replicates which would include either samples from the same shoot of one vine or from different vines.

Acronyms

2OGD 2-oxoglutarate-dependent dioxygenase

A Arabidopsis

ABI4 ABSCISIC ACID-INSENSITIVE 4

AG AGAMOUS

AP APETALA

AP1 V. vinifera APETALA1

APRR3 ARABIDOPSIS PSEUDO RESPONSE REGULATOR 3

BAM compressed sam format

BC Before Christ

bHLH basic-helix-loop-helix

BL Brassinolide

bp base pairs

BP BREVIPEDICELLUS

BRI1 BRASSINOSTEROID INSENSITIVE 1

BS B(sister)

bzip basic leucine zipper domain

CCA1 CIRCADIAN CLOCK ASSOCIATED 1

CDF CYCLING DOF FACTOR

Chr Chromosome

CIB1 CRYPTOCHROME-INTERACTING basic-helix-loop-helix

CO CONSTANS

COL CONSTANS LIKE

CRIBI *Centro di Ricerca Interdipartimentale per le Biotecnologie Innovative*

CRY2 CRYPTOCHROME 2

CSTF64 Cleavage Stimulating Factor 64

CUL1 CULLIN 1C

DNA deoxyribonucleic Acid

dNTP Desoxyribonucleotide

DOF DNA-binding-with-One-Finger

EHD1 EARLY HEADING DATE

EDI early day-length insensitive

EDTA Ethylenediaminetetraacetic acid

ELF EARLY FLOWERING

ER ERECTA

FAO Food and Agriculture Organization of the United Nations

FL FLORICAULA

FLC FLOWERING LOCUS C

FLM FLOWERING LOCUS M

FRI FRIGIGDA

FT FLOWERING LOCUS T

FTC Flowering Time Control

FUL FRUITFULL

GA Giberellic acid

GAI GIBBERELLIC ACID INSENSITIV

GAMYB GA MYB

GATK Genome Analysis Toolkit

GF GF.GA-47-42

GI GIGANTEA

GID1 GA INSENSITIVE DWARF1

GQ genotype quality

GRP2A GLYCIN RICH PROTEIN 2A

GVCF genomic variant call format

ha hectare

HD3a HEADING DATE 3a

HUA2 ENHANCER OF AGAMOUS 2

HY5 ELONGATED HYPOCOTYL 5

IGV Integrative Genomics Viewer

InDel Insertion/ Deletion

JKI Julius Kühn Institute

JmjC Jumonji C

KH K-Homology

KNAT KNOTTED-LIKE

KNOX Knotted1-like homeobox

LD Long day

LFC log fold change

LFY LEAFY

LHY LATE ELONGATED HYPOCOTYL

LOD logarithmic odds ratio

log logarithm

LRR-RLK Leucine-rich repeat receptor-like protein kinase

m meter

mM milli molar

M Million

MADS MCM1 AGAMOUS DEFICIENS SRF

MAGIC multi-parent advanced generation inter-cross

Mb Mega bases

MBD9 METHYL-CPG-BINDING DOMAIN 9

MFT MOTHER OF FT

mg milligram

MPRIL multi parent RIL

MYB myeloblastosis

MYC myelocytomatosis viral oncogene homolog

MIKC MADS-I K and C

n number of chromosome in the haploid set of chromosomes

NGS Next Generation Sequencing

NILs near isogenic lines

OsMADS51 Oryza sativa MADS51

PAS Per-Arnt-Sim

PCA principal component analysis

PCR polymerase chain reaction

PHYD Phytochrome D

PIF4 PHYTOCHROME INTERACTING FACTOR 4

QTL quantitative trait locus

RAP2 RELATED TO APETALA 2

RAV Related to ABI3/VP1

RBH reciprocal best hit

RGA REPRESSOR OF GA

RGL1 RGA-LIKE 1

RILs recombinant inbred lines

RNA ribonucleic acid

RVE REVEILLE

SAM Sequence Alignment/Map format

SBP SQUAMOSA PROMOTER BINDING PROTEIN

SDS Sodium dodecyl sulfate

SD standard deviation

SEP SEPALLATA

SGA shotgun genome assembly

SLR1-like SLENDER RICE 1 LIKE

SNP single nucleotide polymorphisms

SOC1 SUPRESSOR OF OVEREXPRESSION OF CONSTANS 1

SPA SUPPRESSOR OF PHYA

SPY SPINDLY

SRE serum response element

SRF serum response factor

SSR Single Sequence Repeat

STM SHOOT MERISTEMLESS

SVP Short Vegetative Phase

TCP TB1, CYC, and PCF

TEM1 TEMPRANILLO 1

TIC TIME FOR COFFEE

TM TOMATO MADS BOX GENE

TOC1 TIMING OF CAB EXPRESSION 1

UTR untranslated region

V *Vitis*

v version

VB *'Villard Blanc'*

VCF Variant Call Format

VRN2 VERNALIZATION2

VIN3 VERNALIZATION INSENSITIVE 3

VIVC Vitis International Variety Catalogue

Vv *Vitis vinifera*

WNK WITH NO LYSIN KINASE

List of Figures

List of Tables

Bibliography

Achard, P., M. Baghour, A. Chapple, P. Hedden, D. Van Der Straeten, *et al.* (2007). "The plant stress hormone ethylene controls floral transition via DELLA-dependent regulation of floral meristem-identity genes". In: *Proceedings of the National Academy of Sciences* 104.15, pp. 6484–6489.

Adam-Blondon, A.-F., O. Jaillon, S. Vezzulli, A. Zharkikh, M. Troggio, *et al.* (2011). "Genome Sequence Initiatives". In: *Genetics, Genomics, and Breeding of Grapes*. Science Publishers, pp. 211–234.

Aguiar, D. and S. Istrail (2012). "HapCompass: A Fast Cycle Basis Algorithm for Accurate Haplotype Assembly of Sequence Data". In: *Journal of Computational Biology* 19.6, pp. 577–590.

Alabadi, D., T. Oyama, M. J. Yanovsky, F. G. Harmon, P. Más, *et al.* (2001). "Reciprocal Regulation Between TOC1 and LHY/CCA1 Within the Arabidopsis Circadian Clock". In: *Science* 293.5531, pp. 880–883.

Anders, S., P. T. Pyl, and W. Huber (2015). "HTSeq-A Python framework to work with high-throughput sequencing data". In: *Bioinformatics* 31.2, pp. 166–169.

Andrés, F. and G. Coupland (2012). "The genetic basis of flowering responses to seasonal cues." In: *Nature reviews. Genetics* 13.9, pp. 627–39.

El-Assal, E., C. Alonso-Blanco, A. J. Peeters, V. Raz, and M. Koornneef (2001). "A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2." In: *Nature genetics* 29.4, pp. 435–440.

Atwell, S., Y. S. Huang, B. J. Vilhjálmsson, G. Willems, M. Horton, *et al.* (2010). "Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines". In: *Nature* 465.7298, pp. 627–631.

Aya, K., M. Ueguchi-Tanaka, M. Kondo, K. Hamada, K. Yano, *et al.* (2009). "Gibberellin Modulates Anther Development in Rice via the Transcriptional Regulation of GAMYB". In: *the Plant Cell Online* 21.5, pp. 1453–1472.

Bansal, V. and V. Bafna (2008). "HapCUT: An efficient and accurate algorithm for the haplotype assembly problem". In: *Bioinformatics* 24.16, pp. 153–159.

Becker, A., K. Kaufmann, A. Freialdenhoven, C. Vincent, M. Li, *et al.* (2002). "A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes". In: *Molecular Genetics and Genomics* 266.6, pp. 942–950.

Bellaoui, M. (2001). "The Arabidopsis BELL1 and KNOX TALE Homeodomain Proteins Interact through a Domain Conserved between Plants and Animals". In: *the Plant Cell Online* 13.11, pp. 2455–2470.

Bhat, J. A., S. Ali, R. K. Salgotra, Z. A. Mir, S. Dutta, *et al.* (2016). "Genomic selection in the era of next generation sequencing for complex traits in plant breeding". In: *Frontiers in Genetics* 7.DEC, pp. 1–11.

Blackman, B. K. (2017). "Changing Responses to Changing Seasons: Natural Variation in the Plasticity of Flowering Time." In: *Plant physiology* 173.1, pp. 16–26.

Boden, S. A., D. Weiss, J. J. Ross, N. W. Davies, B. Trevaskis, *et al.* (2014). "EARLY FLOWERING3 Regulates Flowering in Spring Barley by Mediating Gibberellin Production and FLOWERING LOCUS T Expression". In: *The Plant Cell* 26.4, pp. 1557–1569.

Bolger, A. M., M. Lohse, and B. Usadel (2014). "Trimmomatic: A flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15, pp. 2114–2120.

Boss, P. K., E. J. Buckeridge, A. Poole, M. R. Thomas, P. K. Boss, *et al.* (2003). "New insights into grapevine flowering". In: *Functional Plant Biology* 30.6, p. 593.

Boss, P. K., L. Sreekantan, and M. R. Thomas (2006). "A grapevine TFL1 homologue can delay flowering and alter floral development when overexpressed in heterologous species". In: *Functional Plant Biology* 33.1, pp. 31–41.

Boss, P. K. and M. R. Thomas (2002). "Association of dwarfism and floral induction with a grape green revolution' mutation". In: *Nature* 416.6883, pp. 847–850.

Bosu, D. R. and E. T. Kipreos (2008). "Cullin-RING ubiquitin ligases: global regulation and activation cycles". In: *Cell Division* 3.1, p. 7.

Browning, S. and B. Browning (2011). "Haplotype phasing: existing methods and new developments". In: *Nature Reviews Genetics* 12.10, pp. 703–714.

Byeon, B. and I. Kovalchuk (2016). "Pattern Recognition on Read Positioning in Next Generation Sequencing." In: *PloS one* 11.6, e0157033.

Calonje, M., P. Cubas, J. M. Martínez-Zapater, and M. J. Carmona (2004). "Floral Meristem Identity Genes Are Expressed during Tendril Development in Grapevine". In: *PLANT PHYSIOLOGY* 135.3, pp. 1491–1501.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, *et al.* (2009). "BLAST+: architecture and applications". In: *BMC Bioinformatics* 10.1, p. 421.

Caño-Delgado, A., Y. Yin, C. Yu, D. Vafeados, S. Mora-García, *et al.* (2004). "BRL1 and BRL3 are novel brassinosteroid receptors that function in vascular differentiation in Arabidopsis." In: *Development (Cambridge, England)* 131.21, pp. 5341–51.

Carmo Vasconcelos, M., M. Greven, C. S. Winefield, M. C. T. Trought, and V. Raw (2009). "The flowering process of vitis vinifera: A review". In: *American Journal of Enology and Viticulture* 60.4, pp. 411–434.

Carmona, M. J., J. Chaïb, J. M. Martínez-Zapater, and M. R. Thomas (2008). "A molecular genetic perspective of reproductive development in grapevine". In: *Journal of Experimental Botany* 59.10, pp. 2579–2596.

Carmona, M. J., P. Cubas, M. Calonje, and J. M. Martínez-Zapater (2007). "Flowering transition in grapevine (Vitis vinifera L.)" In: *Canadian Journal of Botany* 85.8, pp. 701–711.

Carmona, M. J., P. Cubas, and J. M. Martínez-Zapater (2002). "VFL, the Grapevine FLORICAULA/LEAFY Ortholog, Is Expressed in Meristematic Regions Independently of Their Fate 1". In: 130.1.

Castillejo, C. and S. Pelaz (2008). "The Balance between CONSTANS and TEMPRANILLO Activities Determines FT Expression to Trigger Flowering". In: *Current Biology* 18.17, pp. 1338–1343.

Causier, B., R. Castillo, Y. Xue, Z. Schwarz-Sommer, and B. Davies (2010). "Tracing the evolution of the floral homeotic B-and C-function genes through genome synteny". In: *Molecular Biology and Evolution* 27.11, pp. 2651–2664.

Chen, X. and E. M. Meyerowitz (1999). "HUA1 and HUA2 are two members of the floral homeotic AGAMOUS pathway." In: *Molecular cell* 3.3, pp. 349–60.

Chen, X., Z. Zhang, D. Liu, K. Zhang, A. Li, *et al.* (2010). "SQUAMOSA Promoter-Binding Protein-Like Transcription Factors: Star Players for Plant Growth and Development". In: *Journal of Integrative Plant Biology* 52.11, pp. 946–951.

Cheng, C., C. Jiao, S. D. Singer, M. Gao, X. Xu, *et al.* (2015). "Gibberellin-induced changes in the transcriptome of grapevine (Vitis labrusca V. vinifera) cv. Kyoho flowers". In: *BMC Genomics* 16.1, p. 128.

Christenhusz, M. J. M. and J. W. Byng (2016). "The number of known plants species in the world and its annual increase". In: *Phytotaxa* 261.3, pp. 201–217.

Cipriani, G., A. Spadotto, I. Jurman, G. D. Gaspero, M. Crespan, *et al.* (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". In: *Theoretical and Applied Genetics* 121.8, pp. 1569–1585.

Clissold, P. M. and C. P. Ponting (2001). "JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2beta." In: *Trends in biochemical sciences* 26.1, pp. 7–9.

Craufurd, P. Q. and T. R. Wheeler (2009). "Climate change and the flowering time of annual crops". In: *Journal of Experimental Botany* 60.9, pp. 2529–2539.

Cruaud, P., J.-Y. Rasplus, L. J. Rodriguez, and A. Cruaud (2017). "High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy." In: *Scientific reports* 7, p. 41948.

Cubas, P., N. Lauter, J. Doebley, and E. Coen (1999). "The TCP domain: a motif found in proteins regulating plant growth and development". In: *The Plant Journal* 18.2, pp. 215–222.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, *et al.* (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." In: *Nature genetics* 43.5, pp. 491–8.

Diaz-Riquelme, J., D. Lijavetzky, J. M. Martinez-Zapater, and M. J. Carmona (2009). "Genome-Wide Analysis of MIKCC-Type MADS Box Genes in Grapevine". In: *Plant Physiology* 149.1, pp. 354–369.

Díaz-Riquelme, J., J. Grimplet, J. M. Martínez-Zapater, and M. J. Carmona (2012). "Transcriptome variation along bud development in grapevine (Vitis vinifera L.)." In: *BMC plant biology* 12, p. 181.

Ding, Z., A. J. Millar, A. M. Davis, and S. J. Davis (2007). "TIME FOR COFFEE Encodes a Nuclear Regulator in the Arabidopsis thaliana Circadian Clock". In: *the Plant Cell Online* 19.5, pp. 1522–1536.

Douglas, S. J. (2002). "KNAT1 and ERECTA Regulate Inflorescence Architecture in Arabidopsis". In: *the Plant Cell Online* 14.3, pp. 547–558.

Douglas, S. J. and C. D. Riggs (2005). "Pedicel development in Arabidopsis thaliana: Contribution of vascular positioning and the role of the BREVIPEDICELLUS and ERECTA genes". In: *Developmental Biology* 284.2, pp. 451–463.

Doyle, M. R., C. M. Bizzell, M. R. Keller, S. D. Michaels, J. Song, *et al.* (2004). "HUA2 is required for the expression of floral repressors in Arabidopsis thaliana". In: *The Plant Journal* 41.3, pp. 376–385.

Doyle, M. R., S. J. Davis, R. M. Bastow, H. G. McWatters, L. Kozma-Bognár, *et al.* (2002). "The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana". In: *Nature* 419.6902, pp. 74–77.

Duchêne, E. (2016). "How can grapevine genetics contribute to the adaptation to climate change?" In: *Journal International des Sciences de la Vigne et du Vin* 50.3, pp. 113–124.

Duchêne, E., G. Butterlin, V. Dumas, and D. Merdinoglu (2012). "Towards the adaptation of grapevine varieties to climate change: QTLs and candidate genes for developmental stages". In: *Theoretical and Applied Genetics* 124.4, pp. 623–635.

Ehrenreich, I. M., Y. Hanzawa, L. Chou, J. L. Roe, P. X. Kover, *et al.* (2009). "Candidate gene association mapping of Arabidopsis flowering time". In: *Genetics* 183.1, pp. 325–335.

Eichorn, K. W. and D. H. Lorenz (1977). "Phänologische Entwicklungsstadien der Rebe." In: *Nachrichtenbl. Deut. Pflanzenschutz* 29, pp. 119–120.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller (2016). "MultiQC: summarize analysis results for multiple tools and samples in a single report". In: *Bioinformatics* 32.19, pp. 3047–3048.

Fechter, I., L. Hausmann, M. Daum, P. Vieho, T. R. So, *et al.* (2012). "Candidate genes within a 143 kb region of the flower sex locus in Vitis". In:

Fechter, I., L. Hausmann, E. Zyprian, M. Daum, D. Holtgräwe, *et al.* (2014). "QTL analysis of flowering time and ripening traits suggests an impact of a genomic region on linkage group 1 in vitis". In: *Theoretical and Applied Genetics* 127.9, pp. 1857–1872.

Fennell, A. Y., K. A. Schlauch, S. Gouthu, L. G. Deluc, V. Khadka, *et al.* (2015). "Short day transcriptomic programming during induction of dormancy in grapevine." In: *Frontiers in plant science* 6.November, p. 834.

Fernandez, L., A. Ageorges, and L. Torregrosa (2006). "A putative NAP homolog specifically expressed during grapevine flower and berry development". In: *Vitis - Journal of Grapevine Research* 45.1, pp. 51–52.

Fleet, C. M. and T. P. Sun (2005). "A DELLAcate balance: The role of gibberellin in plant morphogenesis". In: *Current Opinion in Plant Biology* 8.1, pp. 77–85.

Forcato, C. (2010). "Gene prediction and functional annotation in the Vitis vinifera genome". In:

Fornara, F., K. C. S. Panigrahi, L. Gissot, N. Sauerbrunn, M. Rühl, *et al.* (2009). "Arabidopsis DOF Transcription Factors Act Redundantly to Reduce CONSTANS Expression and Are Essential for a Photoperiodic Flowering Response". In: *Developmental Cell* 17.1, pp. 75–86.

Foyer, C. H., P. I. Kerchev, and R. D. Hancock (2012). "The ABA-INSENSITIVE-4 (ABI4) transcription factor links redox, hormone and sugar signaling pathways". In: *Plant Signaling & Behavior* 7.2, pp. 276–281.

Fraga, H., A. C. Malheiro, J. Moutinho-Pereira, and J. A. Santos (2013). "Future scenarios for viticultural zoning in Europe: Ensemble projections and uncertainties". In: *International Journal of Biometeorology* 57.6, pp. 909–925.

Frohlich, M. W. and E. M. Meyerowitz (1997). *The Search for Flower Homeotic Gene Homologs in Basal Angiosperms and Gnetales: A Potential New Source of Data on the Evolutionary Origin of Flowers*.

Frohlich, M. W. and D. S. Parker (2000). "The Mostly Male Theory of Flower Evolutionary Origins: From Genes to Fossils". In: *Systematic Botany* 25.2, p. 155.

Gadoury, D. M., R. C. Seem, and M. Moyer (2012). "Climate, Duration of Bloom, and the Window of Risk for Grapevine Diseases". In: *Research News from Cornell's Viticulture and Enology Program* 22.7, pp. 1–6.

García-Mayoral, M. F., D. Hollingworth, L. Masino, I. Díaz-Moreno, G. Kelly, *et al.* (2007). "The structure of the C-terminal KH domains of KSRP reveals a noncanonical motif important for mRNA degradation." In: *Structure (London, England : 1993)* 15.4, pp. 485–98.

Gascuel, Q., G. Diretto, A. J. Monforte, A. M. Fortes, and A. Granell (2017). "Use of Natural Diversity and Biotechnology to Increase the Quality and Nutritional Content of Tomato and Grape." In: *Frontiers in plant science* 8, p. 652.

Gehring, W. J. (1992). "The homeobox in perspective." In: *Trends in biochemical sciences* 17.8, pp. 277–80.

Glaubitz, J. C., O. E. Rhodes, and J. A. Dewoody (2003). "Prospects for inferring pairwise relationships with single nucleotide polymorphisms". In: *Molecular Ecology* 12.4, pp. 1039–1047.

Gocal, G. F., C. C. Sheldon, F. Gubler, T. Moritz, D. J. Bagnall, *et al.* (2001). "GAMYB-like Genes, Flowering, and Gibberellin Signaling in Arabidopsis". In: *Plant Physiology* 127.4, pp. 1682–1693.

Gökbayrak, Z., C. Özer, and G. Söylemezo lu (2006). "Preliminary results on genome mapping of an Italia x Mercan grapevine population". In: *Turkish Journal of Agriculture and Forestry* 30.4, pp. 273–280.

Grattapaglia, D. and R. Sederoff (1994). "Genetic linkage maps of Eucalyptus grandis and Eucalyptus urophylla using a pseudo-testcross: mapping strategy and RAPD markers." In: *Genetics* 137.4, pp. 1121–37.

Greenup, A., W. J. Peacock, E. S. Dennis, and B. Trevaskis (2009). "The molecular biology of seasonal flowering-responses in Arabidopsis and the cereals." In: *Annals of botany* 103.8, pp. 1165–72.

Gubler, F. (2002). "Gibberellin Signaling in Barley Aleurone Cells. Control of SLN1 and GAMYB Expression". In: *Plant Physiology* 129.1, pp. 191–200.

Guichoux, E., L. Lagache, S. Wagner, P. Chaumeil, P. Léger, *et al.* (2011). "Current trends in microsatellite genotyping". In: *Molecular Ecology Resources* 11.4, pp. 591–611. arXiv: S0169-5347(01)02432-6.

Hack, H., H. Bleichholder, L. Buhr, U. Meier, U. Schnock-Fricke, *et al.* (1992). "Einheitliche Codierung der phänologischen Entwicklungsstadien mono- und dikotyler Pflanzen Erweiterte BBCH-Skala." In: *Nachrichtenbl. Deut. Pflanzenschutzd.* 44, pp. 265–270.

Hake, S., H. M. Smith, H. Holtan, E. Magnani, G. Mele, *et al.* (2004). "THE ROLE OF KNOX GENES IN PLANT DEVELOPMENT". In: *Annual Review of Cell and Developmental Biology* 20.1, pp. 125–151.

Hall, A. (2003). "The TIME FOR COFFEE Gene Maintains the Amplitude and Timing of Arabidopsis Circadian Clocks". In: *the Plant Cell Online* 15.11, pp. 2719–2729.

Heintzen, C., M. Nater, K. Apel, and D. Staiger (1997). "AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in Arabidopsis thaliana". In: *Proceedings of the National Academy of Sciences* 94.16, pp. 8515–8520.

Hong-Hermesdorf, A., A. Brüx, A. Grüber, G. Grüber, and K. Schumacher (2006). "A WNK kinase binds and phosphorylates V-ATPase subunit C". In: *FEBS Letters* 580.3, pp. 932–939.

Huang, H., J. Lu, Z. Ren, W. Hunter, S. E. Dowd, *et al.* (2011). "Mining and validating grape (Vitis L.) ESTs to develop EST-SSR markers for genotyping and mapping." In: *Molecular breeding : new strategies in plant improvement* 28.2, pp. 241–254.

Huang, X., X. Ouyang, P. Yang, O. S. Lau, L. Chen, *et al.* (2013). "Conversion from CUL4-based COP1-SPA E3 apparatus to UVR8-COP1-SPA complexes underlies a distinct biochemical function of COP1 under UV-B". In: *Proceedings of the National Academy of Sciences* 110.41, pp. 16669–16674.

Jagadish, S. V. K., R. N. Bahuguna, M. Djanaguiraman, R. Gamuyao, P. V. V. Prasad, *et al.* (2016). "Implications of High Temperature and Elevated CO2 on Flowering Time in Plants." In: *Frontiers in plant science* 7, p. 913.

Jaillon, O., J.-M. Aury, B. Noel, A. Policriti, C. Clepet, *et al.* (2007). "The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla". In: *Nature* 449.7161, pp. 463–467.

Jakoby, M., B. Weisshaar, W. Dröge-Laser, J. Vicente-Carbajosa, J. Tiedemann, *et al.* (2002). "bZIP transcription factors in Arabidopsis." In: *Trends in plant science* 7.3, pp. 106–11.

Jali, S. S., S. M. Rosloski, P. Janakirama, J. G. Steffen, V. Zhurov, *et al.* (2014). "A plant-specific HUA2-LIKE (HULK) gene family in Arabidopsis thaliana is essential for development." In: *The Plant journal : for cell and molecular biology* 80.2, pp. 242–54.

Johanson, U., J. West, C. Lister, S. Michaels, R. Amasino, *et al.* (2000). "Molecular analysis of FRIGIDA, a major determinant of natural variation in Arabidopsis flowering time." In: *Science (New York, N.Y.)* 290.5490, pp. 344–7.

Jolliffe, I. T. (2002). *Principal component analysis*. Springer, p. 487, ISBN: 0387954422.

Joly, D., M. Perrin, C. Gertz, J. Kronenberger, G. Demangeat, *et al.* (2004). "Expression analysis of flowering genes from seedling-stage to vineyard life of grapevine cv. Riesling". In: *Plant Science* 166.6, pp. 1427–1436.

Jones, B., D. Walsh, L. Werner, and A. Fiumera (2009). "Using blocks of linked single nucleotide polymorphisms as highly polymorphic genetic markers for parentage analysis". In: *Molecular Ecology Resources* 9.2, pp. 487–497.

Khalil-Ur-Rehman, M., L. Sun, C.-X. Li, M. Faheem, W. Wang, *et al.* (2017). "Comparative RNA-seq based transcriptomic analysis of bud dormancy in grape". In: *BMC Plant Biology* 17.1, p. 18.

Kicherer, A., K. Herzog, N. Bendel, H.-C. Klück, A. Backhaus, *et al.* (2017). "Phenoliner: A New Field Phenotyping Platform for Grapevine Research". In: *Sensors* 17.7, p. 1625.

Kikis, E. A., R. Khanna, and P. H. Quail (2005). "ELF4 is a phytochrome-regulated component of a negative-feedback loop involving the central oscillator components CCA1 and LHY". In: *Plant Journal* 44.2, pp. 300–313.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, *et al.* (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions". In: *Genome Biology* 14.4, R36.

Kim, S. A., S. Y. Ahn, and H. K. Yun (2016). "Analysis of Structure and Expression of Grapevine 2-oxoglutarate Oxygenase Genes in Response to Low Temperature". In: *Korean Journal of Horticultural Science & Technology* 34.1.

Klodt, M., K. Herzog, R. Töpfer, and D. Cremers (2015). "Field phenotyping of grapevine growth using dense stereo reconstruction." In: *BMC bioinformatics* 16, p. 143.

Kramer, E. M., R. L. Dorit, and V. F. Irish (1998). "Molecular evolution of genes controlling petal and stamen development: Duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages". In: *Genetics* 149.2, pp. 765–783.

Kumar, S. V., D. Lucyshyn, K. E. Jaeger, E. Alós, E. Alvey, *et al.* (2012). "Transcription factor PIF4 controls the thermosensory activation of flowering." In: *Nature* 484.7393, pp. 242–5.

Kwon, J. and A. Goate (2000). "The candidate gene approach." In: *Alcohol Res Health* 24.3, pp. 164–168.

Lancia, G., V. Bafna, S. Istrail, R. Lippert, and R. Schwartz (2001). "SNPs Problems, Complexity, and Algorithms". In: *Algorithms ESA 2001* 2161, pp. 182–193.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2". In: *Nature Methods* 9.4, pp. 357–359.

Laver, T. W., R. C. Caswell, K. A. Moore, J. Poschmann, M. B. Johnson, *et al.* (2016). "Pitfalls of haplotype phasing from amplicon-based long-read sequencing." In: *Scientific reports* 6, p. 21746.

Lebon, G., G. Wojnarowiez, B. Holzapfel, F. Fontaine, N. Vaillant-Gaveau, *et al.* (2008). "Sugars and flowering in the grapevine (Vitis vinifera L.)." In: *Journal of experimental botany* 59.10, pp. 2565–78.

Lee, B., T. Moon, S. Yoon, T. Weissman, T. Weissman, *et al.* (2017). "DUDE-Seq: Fast, flexible, and robust denoising for targeted amplicon sequencing". In: *PLOS ONE* 12.7. Ed. by J. Wang, e0181463.

Lee, J. H., S. J. Yoo, S. H. Park, I. Hwang, J. S. Lee, *et al.* (2007). "Role of SVP in the control of flowering time by ambient temperature in Arabidopsis." In: *Genes & development* 21.4, pp. 397–402.

Leeuwen, C. van, H. R. Schultz, I. Garcia de Cortazar-Atauri, E. Duchene, N. Ollat, *et al.* (2013). "Why climate change will not dramatically decrease viticultural suitability in main wine-producing areas by 2050". In: *Proceedings of the National Academy of Sciences* 110.33, E3051–E3052.

LePaslier, M.-C., N. Choisne, S. Scalabrin, R. Bacilieri, A. Berard, *et al.* (2013). "The GRAPERESEQ 18K Vitis genotyping chip". In: *IX International Symposium on Grapevine Physiology & Biotechnology*, p. 18.

Li, H. (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: arXiv: 1303.3997.

Li, H. and R. Durbin (2009a). "Fast and accurate short read alignment with Burrows-Wheeler transform". In: *Bioinformatics* 25.14, pp. 1754–1760. arXiv: 1303.3997.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.* (2009b). "The Sequence Alignment/Map format and SAMtools". In: *Bioinformatics* 25.16, pp. 2078–2079. arXiv: 1006.1266v2.

Liu, H., X. Yu, K. Li, J. Klejnot, H. Yang, *et al.* (2008). "Photoexcited CRY2 Interacts with CIB1 to Regulate Transcription and Floral Initiation in Arabidopsis". In: *Science* 322.5907, pp. 1535–1539.

Liu, P.-L., L. Du, Y. Huang, S.-M. Gao, and M. Yu (2017). "Origin and diversification of leucine-rich repeat receptor-like protein kinase (LRR-RLK) genes in plants." In: *BMC evolutionary biology* 17.1, p. 47.

Lorenz, D. H., K. W. Eichhorn, H. Bleiholder, R. Klose, U. Meier, *et al.* (1995). "Growth Stages of the Grapevine: Phenological growth stages of the grapevine (Vitis vinifera L. ssp. vinifera) Codes and descriptions according to the extended BBCH scale". In: *Australian Journal of Grape and Wine Research* 1.2, pp. 100–103.

Love, M. I., W. Huber, and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12, p. 550.

Mackay, T. F. C. (2011). "The Genetic Architecture of Quantitative Traits". In: *Annu. Rev. Genet* 35, pp. 303–39.

Mackay, T. F. C. (2014). "Epistasis and quantitative traits: using model organisms to study gene-gene interactions." In: *Nature reviews. Genetics* 15.1, pp. 22–33.

Martin, M., M. Patterson, S. Garg, S. O. Fischer, N. Pisanti, *et al.* (2016). "WhatsHap : fast and accurate read-based phasing". In: pp. 1–18.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, *et al.* (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." In: *Genome research* 20.9, pp. 1297–303.

Mitra, A., M. Skrzypczak, K. Ginalski, and M. Rowicka (2015). "Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using illumina platform." In: *PloS one* 10.4, e0120520.

Mosedale, J. R., R. J. Wilson, and I. M. D. Maclean (2015). "Climate change and crop exposure to adverse weather: Changes to frost risk and grapevine flowering conditions". In: *PLoS ONE* 10.10, pp. 1–16.

Moyroud, E., G. Tichtinsky, and F. Parcy (2009). "The LEAFY Floral Regulators in Angiosperms: Conserved Proteins with Diverse Roles". In: *Journal of Plant Biology* 52.3, pp. 177–185.

Mullins, M. G., A. Bouquet, and L. Williams (1992). *Biology of the Grapevine*.

Munguía-Rosas, M. A., J. Ollerton, V. Parra-Tabla, and J. A. De-Nova (2011). "Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured". In: *Ecology Letters* 14.5, pp. 511–521.

Murray, F., R. Kalla, J. Jacobsen, and F. Gubler (2003). "A role for HvGAMYB in anther development". In: *Plant Journal* 33.3, pp. 481–491.

Murre, C., G. Bain, M. A. van Dijk, I. Engel, B. A. Furnari, *et al.* (1994). "Structure and function of helix-loop-helix proteins." In: *Biochimica et biophysica acta* 1218.2, pp. 129–35.

Nakamichi, N., M. Murakami-Kojima, E. Sato, Y. Kishi, T. Yamashino, *et al.* (2002). "Compilation and characterization of a novel WNK family of protein kinases in Arabiodpsis thaliana with reference to circadian rhythms." In: *Bioscience, biotechnology, and biochemistry* 66.11, pp. 2429–2436.

Norman, C., M. Runswick, R. Pollock, and R. Treisman (1988). "Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element." In: *Cell* 55.6, pp. 989–1003.

Okamuro, J. K., B. Caster, R. Villarroel, M. Van Montagu, and K. D. Jofuku (1997). "The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis". In: *Proceedings of the National Academy of Sciences* 94.13, pp. 7076–7081.

Park, S. T. and J. Kim (2016). "Trends in next-generation sequencing and a new era for whole genome sequencing". In: *International Neurourology Journal* 20, pp. 76–83.

Parker, A. K. (2012). "Modelling phenology and maturation of the grapevine Vitis vinifera L .: varietal differences and the role of leaf area to fruit weight ratio manipulations". In: p. 237.

Patnala, R., J. Clements, J. Batra, J. Kwon, A. Goate, *et al.* (2013). "Candidate gene association studies: a comprehensive guide to useful in silico tools". In: *BMC Genetics* 14.1, p. 39.

Pavlousek, P. (2011). "Evaluation of drought tolerance of new grapevine rootstock hybrids". In: *Journal of Environmental Biology* 32.5, pp. 543–549.

Peng, F. Y., Z. Hu, and R.-C. Yang (2015). "Genome-Wide Comparative Analysis of Flowering-Related Genes in Arabidopsis, Wheat, and Barley." In: *International journal of plant genomics* 2015, p. 874361.

Peng, M., Y. Cui, Y. M. Bi, and S. J. Rothstein (2006). "AtMBD9: A protein with a methyl-CpG-binding domain regulates flowering time and shoot branching in Arabidopsis". In: *Plant Journal* 46.2, pp. 282–296.

Ponting, C. P. and L. Aravind (1997). "PAS: a multifunctional domain family comes to light." In: *Current biology : CB* 7.11, R674–7.

Poupin, M. J., F. Federici, C. Medina, J. T. Matus, T. Timmermann, *et al.* (2007). "Isolation of the three grape sub-lineages of B-class MADS-box TM6,

PISTILLATA and APETALA3 genes which are differentially expressed during flower and fruit development". In: *Gene* 404.1-2, pp. 10–24.

Putterill, J., F. Robson, K. Lee, R. Simon, and G. Coupland (1995). "The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors." In: *Cell* 80.6, pp. 847–57.

Pysh, L. D., J. W. Wysocka-Diller, C. Camilleri, D. Bouchez, and P. N. Benfey (1999). "The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes". In: *The Plant Journal* 18.1, pp. 111–119.

Rawat, R., J. Schwartz, M. A. Jones, I. Sairanen, Y. Cheng, *et al.* (2009). "REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways". In: *Proceedings of the National Academy of Sciences* 106.39, pp. 16883–16888.

Riechmann, J. L. and E. M. Meyerowitz (1998). "The AP2/EREBP Family of Plant Transcription Factors". In: *Biol. Chem* 379, pp. 633–646.

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, *et al.* (2011). "Integrative genomics viewer." In: *Nature biotechnology* 29.1, pp. 24–6.

Samad, S., T. Kurokura, E. Koskela, T. Toivainen, V. Patel, *et al.* (2017). "Additive QTLs on three chromosomes control flowering time in woodland strawberry (Fragaria vesca L.)" In: *Horticulture Research* 4.September 2016, p. 17020.

Scheben, A., J. Batley, and D. Edwards (2017). "Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application". In: *Plant Biotechnology Journal* 15.2, pp. 149–161.

Schirmer, M., R. D 'amore, U. Z. Ijaz, N. Hall, and C. Quince (2016). "Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data". In: *BMC Bioinformatics* 17.

Schirmer, M., U. Z. Ijaz, R. D'Amore, N. Hall, W. T. Sloan, *et al.* (2015). "Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform". In: *Nucleic Acids Research* 43.6. arXiv: arXiv:1312.0570v2.

Shpak, E. D. (2004). "Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation". In: *Development* 131.7, pp. 1491–1501.

Shu, K., Q. Chen, Y. Wu, R. Liu, H. Zhang, *et al.* (2016). "ABSCISIC ACID-INSENSITIVE 4 negatively regulates flowering through directly promoting Arabidopsis FLOWERING LOCUS C transcription". In: *Journal of Experimental Botany* 67.1, pp. 195–205.

Spanudakis, E. and S. Jackson (2014). "The role of microRNAs in the control of flowering time". In: *Journal of Experimental Botany* 65.2, pp. 365–380.

Sreekantan, L., K. Mathiason, J. Grimplet, K. Schlauch, J. A. Dickerson, *et al.* (2010). "Differential floral development and gene expression in grapevines during long and short photoperiods suggests a role for floral genes in dormancy transitioning". In: *Plant Molecular Biology* 73.1-2, pp. 191–205.

Sreekantan, L. and M. R. Thomas (2006). "VvFT and VvMADS8, the grapevine homologues of the floral integrators FT and SOC1, have unique expression patterns in grapevine and hasten flowering in Arabidopsis". In: *Functional Plant Biology* 33.12, pp. 1129–1139.

Srivastava, A. K., D. Senapati, A. K. Srivastava, M. Chakraborty, S. N. Gangappa, *et al.* (2015). "SHW1 Interacts with HY5 and COP1, and Promotes COP1-mediated Degradation of HY5 During Arabidopsis Seedling Development". In: *Plant Physiology* 169.4, pp.01184.2015.

Stracke, R., M. Werber, and B. Weisshaar (2001). "The R2R3-MYB gene family in Arabidopsis thaliana." In: *Current opinion in plant biology* 4.5, pp. 447–56.

Sweetman, C., V. O. Sadras, R. D. Hancock, K. L. Soole, and C. M. Ford (2014). "Metabolic effects of elevated temperature on organic acid degradation in ripening Vitis vinifera fruit". In: *Journal of Experimental Botany* 65.20, pp. 5975–5988.

Tabor, H. K., N. J. Risch, and R. M. Myers (2002). "Candidate-gene approaches for studying complex genetic traits: Practical considerations." In: *Nature Reviews. Genetics* 3.5, pp. 391–397.

Terral, J.-F., E. Tabard, L. Bouby, S. Ivorra, T. Pastor, *et al.* (2010). "Evolution and history of grapevine (Vitis vinifera) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars." In: *Annals of botany* 105.3, pp. 443–55.

Theissen, G., A. Becker, A. Di Rosa, A. Kanno, J. T. Kim, *et al.* (2000). "A short history of MADS-box genes in plants". In: *Plant Molecular Biology* 42.1, pp. 115–149.

Thomas, S. G. (2004). "Update on Gibberellin Signaling. A Tale of the Tall and the Short". In: *Plant Physiology* 135.2, pp. 668–676.

Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov (2013). "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." In: *Briefings in bioinformatics* 14.2, pp. 178–92.

Tseng, T.-S. (2004). "SPINDLY and GIGANTEA Interact and Act in Arabidopsis thaliana Pathways Involved in Light Responses, Flowering, and Rhythms in Cotyledon Movements". In: *the Plant Cell Online* 16.6, pp. 1550–1563.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, *et al.* (2013). "From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline." In: *Current protocols in bioinformatics* 43.1110, pp. 11.10.1–33.

Van Dijk, E. L., H. Lè Ne Auger, Y. Jaszczyszyn, and C. Thermes (2014). "Ten years of next-generation sequencing technology". In: *Trends in Genetics* 30.9, pp. 418–426. arXiv: arXiv:1312.0570v2.

Velasco, R., A. Zharkikh, M. Troggio, D. A. Cartwright, A. Cestaro, *et al.* (2007). "A high quality draft consensus sequence of the genome of a heterozygous grapevine variety". In: *PLoS ONE* 2.12.

Venglat, S. P., T. Dumonceaux, K. Rozwadowski, L. Parnell, V. Babic, *et al.* (2002). "The homeobox gene BREVIPEDICELLUS is a key regulator of inflorescence architecture in Arabidopsis". In: *Proceedings of the National Academy of Sciences* 99.7, pp. 4730–4735.

Verhoeven, K. J. F., G. Casella, and L. M. Mcintyre (2010). "Epistasis: Obstacle or Advantage for Mapping Complex Traits?" In: *PLoS ONE* 5.8.

Vieira, M. L. C., L. Santini, A. L. Diniz, and C. d. F. Munhoz (2016). "Microsatellite markers: what they mean and why they are so useful." In: *Genetics and molecular biology* 39.3, pp. 312–28.

Vitulo, N., C. Forcato, E. Carpinelli, A. Telatin, D. Campagna, *et al.* (2014). "A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype". In: *BMC Plant Biology* 14.1, p. 99.

Wan, Y., H. R. Schwaninger, A. M. Baldo, J. A. Labate, G.-Y. Zhong, *et al.* (2013). "A phylogenetic analysis of the grape genus (Vitis L.) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change". In: *BMC Evolutionary Biology* 13.

Wang, Y., K. Liu, H. Liao, C. Zhuang, H. Ma, *et al.* (2008). "The plant WNK gene family and regulation of flowering time in Arabidopsis". In: *Plant Biology* 10.5, pp. 548–562.

Ward, N. and G. Moreno-Hagelsieb (2014). "Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: How much do we miss?" In: *PLoS ONE* 9.7, pp. 1–6.

Wilcoxon, F. (1945). "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6, p. 80.

Yamasaki, K., T. Kigawa, M. Inoue, T. Yamasaki, T. Yabuki, *et al.* (2006). "An Arabidopsis SBP-domain fragment with a disrupted C-terminal zinc-binding site retains its tertiary structure". In: *FEBS Letters* 580.8, pp. 2109–2116.

Yao, Y., L. Song, Y. Katz, and G. Galili (2002). "Cloning and characterization of Arabidopsis homologues of the animal CstF complex that regulates 3' mRNA cleavage and polyadenylation." In: *Journal of experimental botany* 53.378, pp. 2277–8.

Zhu, M. and S. Zhao (2007). "Candidate gene identification approach: progress and challenges." In: *Int J Biol Sci* 3.7, pp. 420–427.

Zucca, S., M. Villaraggia, S. Gagliardi, G. S. Grieco, M. Valente, *et al.* (2016). "Analysis of amplicon-based NGS data from neurological disease gene panels: a new method for allele drop-out management." In: *BMC bioinformatics* 17.Suppl 12, p. 339.

Zyprian, E., R. Eibach, and R. Töpfer (2006). "Eine neue genetische Karte der Weinrebe aus der Kreuzung GF.GA-47-42 x Villard Blanc". In: *Deutsches Weinbau-Jahrbuch* 57, pp. 151–158.

Zyprian, E., I. Ochßner, F. Schwander, S. Simon, L. Hausmann, *et al.* (2016). "Quantitative trait loci affecting pathogen resistance and ripening of grapevines". In: *Molecular Genetics and Genomics* 291.4, pp. 1573–1594.

Appendix

Table A1: FTC candidate genes chosen for amplicon sequencing in a panel of 48 diverse individuals. QTL regions are highlighted in gray.

Gene ID	gene name	Chr	start	end	length	strand
VIT_201s0011g00100	*VvAP1*	1	150,391	175,330	24,939	-
VIT_201s0011g00110	*VvSEP4*	1	193,865	206,588	12,723	-
VIT_201s0011g01240	*VvSUF4a*	1	1,071,613	1,078,482	6,869	-
VIT_201s0011g01560	*VvBS2*	1	1,353,271	1,355,959	2,688	+
VIT_201s0011g02120	*VvHUA2a*	1	1,797,826	1,815,325	17,499	+
VIT_201s0011g03070	*VvRAV1b*	1	2,751,566	2,753,036	1,470	+
VIT_201s0011g03140	*VvPFT1a*	1	2,806,738	2,878,676	71,938	+
VIT_201s0011g03520	*VvCOL16*	1	3,190,592	3,193,373	2,781	+
VIT_201s0011g04240	*VvCOL10*	1	3,849,170	3,854,088	4,918	+
VIT_201s0011g05260	*VvGAI1*	1	4,895,037	4,897,415	2,378	-
VIT_201s0011g05810	*VvADOd*	1	5,580,698	5,585,530	4,832	-
VIT_201s0011g06410	*VvBRIa*	1	6,227,185	6,230,660	3,475	-
VIT_201s0026g00150	*VvMBD9a*	1	8,787,896	8,805,571	17,675	-
VIT_201s0026g02200	*VvTCP15b*	1	11,609,059	11,611,589	2,530	-
VIT_201s0026g02580	*VvCDF2a*	1	12,211,948	12,215,668	3,720	+
VIT_201s0010g00730	*VvPRRa*	1	15,898,381	15,900,732	2,351	-
VIT_201s0010g00740	*VvbHLH49*	1	15,902,319	15,907,218	4,899	-
VIT_201s0010g02270	*VvSCL21a*	1	19,230,746	19,232,935	2,189	+
VIT_201s0010g03890	*VvFLC*	1	21,240,373	21,261,785	21,412	-
VIT_201s0010g03900	*VvSEP3*	1	21,368,759	21,386,789	18,030	-
VIT_203s0017g02143	*VvSVP5*	3	12,256,603	12,270,440	13,837	-
VIT_203s0167g00070	*VvSUF4b*	3	18,864,770	18,865,390	620	+
VIT_204s0023g02820	*VvTM6/ VvAP3b*	4	19,395,299	19,397,804	2,505	+
VIT_204s0044g00850	*VvWNK6*	4	21,993,660	21,997,968	4,308	+
VIT_205s0020g03150	*VvTICa*	5	4,894,145	4,901,644	7,499	+
VIT_205s0102g01160	*VvFPA*	5	23,292,031	23,303,676	11,645	-
VIT_206s0004g03590	*VvRAP2a*	6	4,494,029	4,498,663	4,634	-
VIT_206s0009g02480	*VvGAMYBc*	6	15,015,885	15,020,621	4,736	-
VIT_207s0104g01590	*VvCUL1c*	7	2,559,573	2,569,054	9,481	+
VIT_207s0005g02260	*VvSPL1b*	7	4,628,005	4,638,050	10,045	-
VIT_208s0056g01230	*VvCDF3b*	8	1,937,022	1,940,172	3,150	-
VIT_208s0007g04200	*ATHB51*	8	18,188,022	18,188,890	868	+
VIT_209s0002g02680	*VvELF3*	9	2,435,821	2,441,150	5,329	+
VIT_210s0042g00820	*VvABS1*	10	14,127,878	14,133,902	6,024	-
VIT_211s0016g02170	*VvMSI4a*	11	1,774,420	1,784,665	10,245	+
VIT_211s0052g01800	*VvCOL5*	11	19,620,744	19,622,772	2,028	+
VIT_212s0028g03350	*VvSPB1*	12	4,108,169	4,110,257	2,088	-
VIT_212s0059g01180	*VvPBRM1*	12	6,103,036	6,111,456	8,420	-
VIT_213s0067g03390	*VvPPR37b*	13	1,857,954	1,891,641	33,687	+
VIT_213s0019g03550	*VvTOE3*	13	4,719,781	4,724,350	4,569	+
VIT_214s0060g00100	*VvPHYA*	14	114,200	121,866	7,666	+
VIT_214s0060g00260	*VvPIF3b*	14	262,210	267,727	5,517	-

Gene ID	gene name	Chr	start	end	length	strand
VIT_214s0060g02440	*VvNUCa*	14	2,059,291	2,063,317	4,026	-
VIT_214s0030g00440	*VvGID1Ba*	14	4,293,572	4,298,380	4,808	-
VIT_214s0030g01660	*VvWNK1*	14	6,461,467	6,466,112	4,645	-
VIT_214s0081g00440	*VvFLKa*	14	8,304,722	8,321,747	17,025	-
VIT_214s0006g00640	*VvGAIb*	14	14,806,872	14,809,122	2,250	+
VIT_214s0083g00130	*VvRAV1c*	14	22,050,931	22,051,785	854	-
VIT_214s0083g00640	*VvCOL2*	14	22,695,967	22,698,956	2,989	+
VIT_214s0083g01030	*VvFUL-L*	14	23,320,025	23,341,255	21,230	-
VIT_214s0083g01050	*VvSEP1*	14	23,363,172	23,379,670	16,498	-
VIT_214s0068g01800	*VvFLC2*	14	25,509,913	25,511,279	1,366	+
VIT_214s0066g02630	*VvCLP1*	14	28,790,578	28,797,347	6,769	+
VIT_214s0108g00490	*VvCUL3b*	14	29,254,976	29,265,032	10,056	+
VIT_214s0108g00980	*VvCDF2b*	14	29,617,258	29,620,832	3,574	+
VIT_215s0048g01280	*VvAML3*	15	15,432,479	15,432,618	139	+
VIT_215s0048g01320	*VvGA20OX1c*	15	15,453,586	15,455,519	1,933	+
VIT_216s0013g00860	*VvEMF2a*	16	6,368,434	6,436,206	67,772	-
VIT_217s0000g00150	*VvFL*	17	71,718	73,739	2,021	+
VIT_217s0000g02630	*VvMFT*	17	2,414,076	2,416,006	1,930	-
VIT_217s0000g04990	*VvFUL1*	17	5,434,018	5,452,725	18,707	-
VIT_217s0000g06570	*VvTOC1*	17	7,100,783	7,110,058	9,275	+
VIT_217s0000g08480	*VvWERb*	17	9,703,225	9,704,469	1,244	-
VIT_217s0000g10300	*VvGAIa*	17	12,556,730	12,558,304	1,574	-
VIT_218s0001g07460	*VvSVP2*	18	5,694,296	5,711,407	17,111	+
GSVIVT01001405001	*VvGI*	18	18675720	18680201	4482	-
VIT_219s0140g00120	*VvODO1*	19	6,691,483	6,696,029	4,546	-
VIT_219s0090g00590	*VvGA2OX1a*	19	15,497,652	15,500,059	2,407	+
VIT_210s0116g00750	*VvGESa*	10_random	340,057	344,422	4,365	-
VIT_218s0001g01800	*VvCSTF64*	18_random	2,318,221	2,332,606	14,385	-
VIT_200s0203g00080	*VvMFT*	Un	11,888,686	11,889,802	1,116	+
VIT_200s1675g00010	*VvEMF2e*	Un	40,052,725	40,055,878	3,153	-

Table A2: Amplimers and primer used for amplicon sequencing and their genomic position. fw: forward. rv: reverse. P: primer

Amplimer	fw-P	Sequence	Position fw-Primer	rv-P	Sequence	Position rv-Primer
Vv00064a	v19p	TTGCTGCTGTTTGTGCTCTT	chr18:18,680,319..18,680,300	v19q	CCTTGGCTCTCAGTCTGCTA	chr18:18,675,535..18,675,554
Vv00060a	v19h	CATTTTGAAAATCGGGCCGC	chrUn:11,888,203..11,888,222	v19i	TCTAGATTGAGCTGGGCCTG	chrUn:11,888,203..11,888,222
Vv00100a	v23a	ACCAATTCCCTCCTTCCCTG	chr2:35,343..35,362	v23b	CTAAAGCCTACAACGCCGTC	chr2:39,337..39,318
Vv00046a	BA44	AGCACTTTCCAGAAATACCCTG	chr1:15,900,798..15,900,777	BA45	GTTTTCTGATGCTCTGCCCTG	chr1:15,898,413..15,898,433
Vv00044a	BA36	TGGGTATAAATGCAGGGCCTT	chr1:15,907,193..15,907,173	BA73	GGGTGCTGACTCCTTGTACC	chr1:15,903,920..15,903,939
Vv00040a	BA28	TCAGAGGTACCTTGCAGTTGAG	chr1:19,231,131..19,231,152	BA29	TCTCCATGGATCACACTAGTTTT	chr1:19,232,440..19,232,418
vv00033d	v07O	AAAAGCTAGGTGTTATGGGCAA	chr1:21,257,213..21,257,234	v07P	GGGAATTGAGAGATGGCTAACA	chr1:21,263,113..21,263,092
Vv00034a	v07R	TTGAATTTGTTACCGGTTGATG	chr1:21,367,943..21,367,964	v07S	GATTTAAATGTCCCAAGCCAAA	chr1:21,374,427..21,374,406
Vv00034b	v07T	TCAAAGTTCAAAGGAGTTGTGC	chr1:21,374,333..21,374,354	v07U	TTTAGGAAAAGGACAAGGACCA	chr1:21,380,953..21,380,932
Vv00012a	v08X	GTATGGCCAGGTTTTACTCAGC	chr1:149,474..149,495	v08Y	TTAGAGTCCCTGCAACTAACCC	chr1:156,231..156,210
Vv00012b	v08Z	GGGTTAGTTGCAGGGACTCTAA	chr1:156,210..156,231	v09A	CTATGTAGATCCCTCGAAGCCT	chr1:163,047..163,026
Vv00009a	v08R	GGTGGCTGCTCCAGTTAAAGTA	chr1:193,108..193,129	v08S	CTGTCTAAGCCAATGTTGGATATT	chr1:198,021..197,998
Vv00007a	v08N	CCACATCGAGTGATTATTTAGATTT	chr1:1,070,799..1,070,823	v08O	CACCCAGCAATCTCTATTCCTC	chr1:1,075,058..1,075,037
Vv00007b	v08P	GAGGAATAGAGATTGCTGGGTG	chr1:1,075,037..1,075,058	v08Q	GACCAATTCAAGACGGAGAAAG	chr1:1,079,319..1,079,298
Vv00039a	BA26	CCAAGCAAGGAAAAGGGCATAA	chr1:1,353,185..1,353,206	BA27	GGAACAAAAGTGACATAGGGAGA	chr1:1,356,069..1,356,047
Vv00008a	v08J	ACAGCAAACCGTAGAGAAGAGG	chr1:1,803,021..1,803,042	v08K	CCTGGTTAGAATAAAGCAGGAGG	chr1:1,809,378..1,809,356
Vv00008b	v12T	AGACAATCGTTCAATTTCGGAT	chr1:1,796,969..1,796,990	v12U	ATTTCCGAGGACTATGGGAGAT	chr1:1,803,088..1,803,067
Vv00013a	v12P	AGGAGGTTTGTGAACGTTGG	chr1:2,750,739..2,750,758	v12Q	CCCATGACACCTCTGGACTT	chr1:2,753,836..2,753,817
Vv00016b	v13L	TTGTGTTTATTCCTGAAAGGGG	chr1:2,832,803..2,832,824	v13M	AATCGAAAGTAGTGTTTGGGGA	chr1:2,839,326..2,839,305
Vv00015a	v12J	TAGCCAGGAAAGGATGGATG	chr1:3,190,434..3,190,453	v12K	ACGATTCCCCACTGAGTTTG	chr1:3,192,067..3,192,048
Vv00015b	v12L	GGCCTGACTGCTTGGTATGT	chr1:3,191,876..3,191,895	v12M	GGCCTGGCCATAAATCATTA	chr1:3,193,245..3,193,226
Vv00014a	v030	ACGAAACTCGGACTTGCACT	chr1:3,848,721..3,848,740	v031	GAAGCACTCCCATGAAAAGC	chr1:3,855,311..3,855,292
Vv00011a	v08F	ACATAGTCCAATCCATAGGGGA	chr1:4,894,044..4,894,065	v08G	GAGAGTGTCAGTCAGTTGGGG	chr1:4,898,616..4,898,596
Vv00010a	v08D	GAGGATTTTGGGTGAGATCTGT	chr1:5,580,291..5,580,312	v08E	TTTTCTTTCTCTTATTGTCTCCCAA	chr1:5,586,262..5,586,238
Vv00041a	BA30	ATTCTCATCATTCTCGCCACTC	chr1:6,230,731..6,230,710	BA31	AATTCGCATGTGGATCTGTGA	chr1:6,226,919..6,226,939
Vv00038a	v16K	CCCTTGGGTTGCTGTTTCAA	chr1:8,786,973..8,786,992	v16L	TTAGTTCACCCATGCTTGCC	chr1:8,792,139..8,792,120
Vv00043a	BA34	TTGAGTCGAGGGTCATCTACAG	chr1:11,611,459..11,611,438	BA72	CCCACTTGACATGGTCCCTC	chr1:11,610,533..11,610,552
Vv00042a	BA32	CTGACCTGTCTTCCCTCTTTCT	chr1:12,212,051..12,212,072	BA33	ATAGTGCTAGGAAGGGGAGTTG	chr1:12,215,566..12,215,545
Vv00083a	v205	TATTGTTCCTTCCCCGTGCT	chr3:18,907,194..18,907,213	v206	GCACAAGTACAACCACCCTG	chr3:18,917,559..18,917,540
Vv00083b	v207	CGGGAACTGCATTTTAGTGTCA	chr3:18,906,449..18,906,470	v208	CCAACACCCATGCAAACTCA	chr3:18,912,671..18,912,652
Vv00081c	v20s	AGGTCGGTTAGGCGTGTAAA	chr3:12,288,583..12,288,564	v20t	CGGAAGGTTTTGAAGGCGAA	chr3:19,562,834..19,562,815
Vv00081e	v20w	TGCATCATTTTCCATTCCCTCT	chr3:12,275,801..12,275,780	v20x	GCGAAATCTCTCTTGCAGCA	chr3:12,270,157..12,270,176
Vv00086a	v21x	ACTGACCAGGGACGTGAATT	chr4:19,394,945..19,394,964	v21y	ACATCCCAACATCCAAGCATG	chr4:19,398,045..19,398,025
Vv00096a	v22m	TGGGAGTGACTCGAATGCAT	chr4:21,994,322..21,994,341	v22n	CTTCAATGGTGTCTGGGTGC	chr4:21,998,205..21,998,186
Vv00101a	v23e	AGCGGAGATGGACAGAAACA	chr5:4,894,743..4,894,762	v23f	CTGAGGTTGAGGGTTCGAGT	chr5:4,899,433..4,899,414
Vv00082a	v203	TCGGTGCGTTGGATTGTTTT	chr5:23,304,850..23,304,831	v204	GTGAAGGGACAGGCAAGTTG	chr5:23,296,885..23,296,904
Vv00102a	v23g	ATTTGGTGGTCTTGGCAACC	chr6:4,498,966..4,498,947	v23h	TGCAAAGAAACGTGTCAGGG	chr6:4,494,149..4,494,168
Vv00091a	v22a	GGGATTGTGAGAAACGGTGG	chr6:15,021,068..15,021,049	v22b	AAGTACCCTTGCTCTGCTGT	chr6:15,015,975..15,015,994

Amplimer	fw-P	Sequence	Position fw-Primer	rv-P	Sequence	Position rv-Primer
Vv00104a	v23m	GCTCTTCACTGCCAAACACA	chr7:4,635,154..4,635,135	v23n	CGAGCACTCTTCCATGCTTC	chr7:4,629,108..4,629,127
Vv00103a	v23k	ATTCCGTTGCTTGACCGATG	chr7:2,560,922..2,560,941	v23l	TTCCTGCCAAAAGTGACAGC	chr7:2,567,071..2,567,052
Vv00108a	v232	GGACACACACACAAGCACAT	chr8:18,187,348..18,187,367	v233	GCCAGCCAAATTAAACCCCA	chr8:18,188,931..18,188,912
Vv00092a	v22c	ATTTGCTCCAACTCCCCTCA	chr8:1,939,911..1,939,892	v22d	GTGCACCAACCACGAATTCT	chr8:1,937,060..1,937,079
Vv00105a	v23t	TTTAGATACAGGCCGCCACA	chr9:2,435,494..2,435,513	v23u	CAACGTCTTGTGGTCCAGTG	chr9:2,440,791..2,440,772
Vv00069a	v19z	GAAGAGCTCATTGGACAGCA	chr10:14,134,419..14,134,400	v191	CTCTTGTGCAACTGTGGAACA	chr10:14,127,773..14,127,793
Vv00070a	v192	GATGGGTTCTGGGGAGAGTC	chr10_random:344,218..344,199	v193	GAAGAACATGGCATCAGGGG	chr10_random:339,717..339,736
Vv00071a	v194	ATGAGGGAGAGGGCAAAGTC	chr11:1,775,089..1,775,108	v195	GGGGTTGGTGTGGTCAAATC	chr11:1,783,580..1,783,561
Vv00072a	v196	CCACACCTCAGTCGTCAGTA	chr11:19,620,833..19,620,852	v197	AAGGTGGGTGAGAGCCAATT	chr11:19,622,464..19,622,445
Vv00074a	v20a	GTGGGGCTCTTGGGAATTTG	chr12:4,110,705..4,110,686	v20b	ACCTGCCCACAAACGATTTC	chr12:4,108,700..4,108,719
Vv00106a	v23v	GTGAGAACGGAGCCAAACAG	chr12:6,110,647..6,110,628	v23w	CACACCCTGAAGCTAGACCA	chr12:6,104,022..6,104,041
Vv00075a	v20c	CGGACCCCTTCAGATTGTCT	chr13:4,719,797..4,719,816	v20d	ATTGTGAATGGGGCTGTTCC	chr13:4,724,124..4,724,105
Vv00099a	v22w	CTCAGCTCGCATGTTGTCTC	chr13:1,862,156..1,862,175	v22x	TAGATCAGCCTGCCAACCAA	chr13:1,866,105..1,866,086
Vv00026a	v09Z	TTTGGTGCGTACCACTTACATC	chr14:14,805,916..14,805,937	v10A	ATCTATGCCAAGTCACGCTTCT	chr14:14,809,295..14,809,274
Vv00035a	v16O	CCAGAGCAGCACATTCGTTT	chr14:4,293,121..4,293,140	v16P	CCGTTCACCCTAGTCGTTCT	chr14:4,298,339..4,298,320
Vv00021a	v09V	TGGGCATGACTAAATACATTCG	chr14:6,461,502..6,461,523	v09W	CCTCAAAGGACCCTTCTCATAA	chr14:6,466,967..6,466,946
Vv00023b	v10H	CATGCATGTGTTTGGTAACTGA	chr14:118,333..118,354	v10I	ATTGCGTATTCTCAATTCTGGC	chr14:122,599..122,578
Vv00022a	v10D	GGTAGGGAGGATTAGGACGTTT	chr14:261,879..261,900	v10E	CCCACCTCTCTCTCAATCAAAA	chr14:268,677..268,656
Vv00036a	v16Q	AGAGAAGCTGACCCAAGACC	chr14:2,058,357..2,058,376	v16R	ATGGGCACAAAGGTCCTACA	chr14:2,063,488..2,063,469
Vv00109a	v234	AGCTTCTGATCACCATGCCT	chr14:/../	v235	AAAATGCGGGGTCCACAATC	chr14:/../
Vv00024d	v10T	TGATCTAACTCTTGCTAGAGACCCT	chr14:25,525,842..25,525,866	v10U	AGTCTGGTTGCTTGAGAATGGT	chr14:25,531,610..25,531,589
Vv00025a	v17A	GAGGAGTCAGAGGCGTAGAC	chr14:8,304,570..8,304,589	v17B	TGCCCTCTTTGCTGTTCATG	chr14:8,310,133..8,310,114
Vv00025b	v10Z	AAGTCATCCAAAAGGACCTTGA	chr14:8,310,401..8,310,422	V11A	AAACCATTACCTTGAGCCCTCT	chr14:8,316,668..8,316,647
Vv00037a	v17E	TCGACGGAGAATCAGGACTG	chr14:22,050,144..22,050,163	v17F	ACCGTGACTTGACAAGACCA	chr14:22,056,188..22,056,169
Vv00017a	V11F	GGTTTTGTTGTATGGGTAGGGA	chr14:22,694,839..22,694,860	v06Z	GAGTCCGCTCTGCAGTAAATAGT	chr14:22,696,324..22,696,302
Vv00018a	V11H	TTACAACAACAATCTGTGTGGCT	chr14:23,319,408..23,319,430	V11I	AATTTTAACGCCACCCCAAG	chr14:23,325,134..23,325,115
Vv00019a	V11P	ATGAATGAAAGCTTCGGGTCT	chr14:23,362,282..23,362,302	V11Q	AAGCTTTAAGATCACGCTTCAAA	chr14:23,368,382..23,368,360
Vv00019b	V11R	GAAAGCTTTGAAGCGTGATCTT	chr14:23,368,354..23,368,375	V11S	TCATATGAGGCCACACATATCAG	chr14:23,374,518..23,374,496
Vv00029a	v07X	TTACAAATGGGACATGGCATAA	chr14:29,254,281..29,254,302	v07Y	GCAATAGGTCTGATTTGGTTCAG	chr14:29,260,159..29,260,137
Vv00030a	v08B	TGAAAATTAGTATTATGCAAAAGCAA	chr14:29,616,477..29,616,502	v08C	GTATCAAAAGCCACCTGCCTAC	chr14:29,621,629..29,621,608
Vv00110a	v238	TTGGTCTTCAGGCCATGGAT	chr15:15,432,380..15,432,399	v239	CCATTCCCTCACCACCTCTT	chr15:15,438,814..15,438,795
Vv00107a	v23z	AATGAACGGCCTTGGCTTTT	chr16:14,863,151..14,863,132	v231	AGCTGACCTCATTTGCCTCT	chr16:15,455,615..15,455,596
Vv00085c	v21p	GGTTTGAAGGGCTGTGTGTT	chr16:6,429,831..6,429,812	v21q	GCTCATGCCAATTCTAGGCC	chr16:6,423,330..6,423,349
Vv00085d	v21r	ATGCATGGCCTAGAATTGGC	chr16:6,423,355..6,423,336	v21s	CATGGATGCAAGTCAACAGTT	chr16:6,416,753..6,416,773
vv00032a	v14J	GGGGAACAATAGCAATAGCAAA	chr17:70,722..70,743	v14K	TACCTGTTCAGGCTGTGATCTG	chr17:74,726..74,705
Vv00061a	v19j	CGCAGAGGGTGTTGATCATG	chr17:2,416,337..2,416,318	v19k	AGGATGGAGCAGTTATGGGG	chr17:2,414,026..2,414,045
Vv00048a	v18s	CGCTCGGACGATTGAATCAG	chr17:5,453,263..5,453,244	v18t	ACTCAACTCAGTCCAACCTCA	chr17:5,449,238..5,449,258
Vv00049a	v18u	TGAGGTTGGACTGAGTTGAGT	chr17:5,449,258..5,449,238	v18v	GGAAGCATGTGTCACGGAAG	chr17:5,442,334..5,442,353
Vv00050a	v18w	CTTCCGTGACACATGCTTCC	chr17:5,442,353..5,442,334	v18x	CCATCACAACCAAAGGCACA	chr17:5,437,689..5,437,708
Vv00051a	v18y	ACCCCTCAGCTAGTTGTTGT	chr17:5,438,111..5,438,092	v18z	CAATATAGCGGTGTGTGGCC	chr17:5,433,520..5,433,539

Amplimer	fw-P	Sequence	Position fw-Primer	rv-P	Sequence	Position rv-Primer
Vv00052a	v181	TAACGCCTCCACAGTTCCTT	chr17:7,101,034..7,101,053	v182	AAGTGGAACTGGAGCCTACC	chr17:7,108,989..7,108,970
Vv00065a	v19r	GGGAGAATGGGTGATTTGAGAC	chr17:9,704,872..9,704,851	v19s	CCCAATTTACGGCCCCAATT	chr17:9,703,139..9,703,158
Vv00066	v19t	AAAGTTCGTGGGGAGGGAAA	chr17:12,558,550..12,558,531	v19u	TGAGTTACAGTTGCCAGGCT	chr17:12,556,385..12,556,404
VV00084b	v21d	CATATCCCCACCCCAATCCT	chr18_random:2,328,257..2,328,276	v21b	CGCCTCCAATAAGACCCAGA	chr18_random:2,332,642..2,332,623
Vv00084d	v21g	CTGTGGTGTCCTGGCCTT	chr18_random:2,322,655..2,322,638	v21h	CTGGGAAGATTTGTGGTGGG	chr18_random:2,317,315..2,317,334
Vv00084f	v21k	ACGTGGAGAGACATAGGCAC	chr18_random:2,312,936..2,312,917	v21c	TGTGCGGGAATAGGAAGGTT	chr18_random:2,307,895..2,307,914
Vv00053a	v183	GGCTAACTGTACGGAAAGCG	chr18:5,694,328..5,694,347	v184	TGGTAGTTTTGGCTCCCGAT	chr18:5,702,000..5,701,981
Vv00054a	v185	CTAGCATCGGGAGCCAAAAC	chr18:5,701,976..5,701,995	v186	GATGGTGAGGGGTGACAAGA	chr18:5,706,377..5,706,358
Vv00055a	v187	TCTTGTCACCCCTCACCATC	chr18:5,706,358..5,706,377	v188	CCACCCCAAGAGCAGAGTAA	chr18:5,711,670..5,711,651
Vv00079a	v20k	CCATCTGAGGCCCAATCTGA	chr19:6,694,918..6,694,899	v20l	AAAGAGCGCGACCCAAATTT	chr19:6,691,502..6,691,521
Vv00077a	v20g	CGTGTTCTTCCAAAGCCCTC	chr19:15,497,813..15,497,832	v20h	CACAGGTCGCTGAAATGGTT	chr19:15,499,851..15,499,832

Table A3: Analyzed cultivars and the lines they originate from, usage of the cultivar, the flowering time phenotype and the genetic background.

Line	**Phenotype of flowering time**	**Usage**	**genetic background**
'Bacchus'	intermediate early	wine grape	*Vitis vinifera*
'Chardonnay'	early	wine grape	*Vitis vinifera*
'Cinerea Arnold'	late	wine grape	*Vitis cinerea*
'Dureza'	unknown	wine grape	*Vitis vinifera*
'Grenache Noir'	late	wine grape	*Vitis vinifera*
'Heunisch'	late	wine grape	*Vitis vinifera*
'Merzling'	intermediate early	wine grape	new variety
'Mondeuse Blanche'	unknown	wine grape	*Vitis vinifera*
'Müller-Thurgau'	intermediate early	wine grape	*Vitis vinifera*
'Östereichisch weiss'	unknown	wine grape	*Vitis vinifera*
'Pinot Blanc'	intermediate	wine grape	*Vitis vinifera*
'Pinot Noir'	intermediate	wine grape	*Vitis vinifera*
'Pinot Noir précoce'	intermediate early	wine grape	*Vitis vinifera*
'Riesling'	late	wine grape	*Vitis vinifera*
'Riparia Gm 183'	early	wine grape	*Vitis riparia*
'Seyval'	intermediate early	wine grape	new variety
'Shiava Grossa'	intermediate late	wine grape	*Vitis vinifera*
'Silvaner'	intermediate early	wine grape	*Vitis vinifera*
'Solaris'	early	wine grape	new variety
'Subereux'	intermediate late	wine grape	complex breeding line
'Syrah'	late	wine grape	*Vitis vinifera*
'Traminer'	intermediate late	wine grape	*Vitis vinifera*
'Villard Blanc'	late	wine grape	new variety
'Börner'	intermediate late	rootstock	interspecies crossing
Geisenheim 6493	unknown	breeding line	complex breeding line
GF V.3125	intermediate	wine grape & breeding line	*Vitis vinifera*
GF.GA-47-42	early	wine grape & breeding line	*Vitis vinifera*
GF.GA-52-42	late	breeding line	complex breeding line

Table A4: F1-individuals of the mapping population GF.GA-47-42 x 'Villard Blanc' chosen for amplicon sequencing of FTC candidate genes and the mean of their flowering time classes according to data from 1999 and 2009 - 2016.

Line	**flowering time class**
89-030-022	4
89-030-029	2
89-030-052	2
89-030-054	2
89-030-057	3
89-030-095	2
89-030-098	3
89-030-152	5
89-030-185	3
89-030-198	3
89-030-201	5
89-030-247	3
89-030-257	1
89-030-273	2
89-030-309	4
89-030-328	4
89-030-329	5
89-030-349	2
89-030-371	6
89-030-378	4
89-030-384	5
89-030-434	2
89-030-461	4
89-030-503	5
89-030-508	2
89-030-544	5
89-030-560	4
89-030-566	4
89-030-575	3
89-030-578	4
89-030-584	4
89-030-594	5
89-030-596	2
89-030-632	4
89-030-640	4

Table A5: Sequenced amplicons in 35 F1-individuals and the parental lines of the mapping population GF.GA-47-42 x 'Villard Blanc'

Gene-ID/Sample-ID	89-030-22	89-030-29	89-030-52	89-030-54	89-030-57	89-030-95	89-030-98	89-030-152	89-030-185	89-030-198	89-030-201	89-030-247	89-030-257	89-030-273	89-030-309	89-030-328	89-030-329	89-030-349	89-030-371	89-030-378	89-030-384	89-030-434	89-030-461	89-030-503	89-030-508	89-030-544	89-030-560	89-030-566	89-030-575	89-030-578	89-030-584	89-030-594	89-030-596	89-030-632	89-030-640	GF.GA-47-42	'Villard Blanc'
GSVIVT01001405001	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_200s0203g00080	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_200s1675g00010	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0010g00730	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0010g00740	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0010g02270	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0010g03890	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X	X	X	✓	✓
VIT_201s0010g03900	✓	X	X	X	X	✓	✓	X	✓	✓	✓	X	X	✓	X	X	X	X	✓	✓	X	X	X	✓	X	X	X	X	X	X	X	X	X	X	✓	✓
VIT_201s0011g00100	✓	X	✓	X	✓	✓	X	X	X	X	X	X	X	✓	X	✓	X	X	✓	✓	X	X	X	✓	X	✓	✓	X	X	X	X	X	X	X	✓	✓	✓
VIT_201s0011g00110	✓	✓	✓	✓	X	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X	X	X	✓	✓
VIT_201s0011g01240	X	X	X	X	X	✓	X	X	X	X	X	X	X	✓	X	✓	X	X	✓	✓	X	X	X	✓	X	✓	X	X	X	X	X	X	X	X	✓	✓
VIT_201s0011g01560	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0011g02120	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X	X	X	✓	✓
VIT_201s0011g03070	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	X	X	X	✓	✓
VIT_201s0011g03140	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0011g03520	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X	X	X	✓	✓
VIT_201s0011g04240	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X	X	X	✓	✓
VIT_201s0011g05260	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X	X	X	✓	✓
VIT_201s0011g05810	X	X	X	X	X	✓	✓	X	X	X	✓	X	X	✓	X	X	X	X	✓	✓	X	X	X	X	X	✓	X	X	X	X	X	X	X	X	X	✓	✓
VIT_201s0011g06410	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0026g00150	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	X	✓	X	X	X	✓	✓
VIT_201s0026g02200	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0026g02580	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_203s0017g02143	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_203s0167g00070	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_204s0023g02820	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_204s0044g00850	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_205s0020g03150	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_205s0102g01160	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_206s0004g03590	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_206s0009g02480	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_207s0005g02260	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_207s0104g01590	✓	✓	✓	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	X	✓	✓	✓	✓	✓	✓

Gene-ID/Sample-ID	89-030-22	89-030-29	89-030-52	89-030-54	89-030-57	89-030-95	89-030-98	89-030-152	89-030-185	89-030-198	89-030-201	89-030-247	89-030-257	89-030-273	89-030-309	89-030-328	89-030-329	89-030-349	89-030-371	89-030-378	89-030-384	89-030-434	89-030-461	89-030-503	89-030-508	89-030-544	89-030-560	89-030-566	89-030-575	89-030-578	89-030-584	89-030-594	89-030-596	89-030-632	89-030-640	GF.GA-47-42	'Villard Blanc'
VIT_208s0007g04200	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_208s0056g01230	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_209s0002g02680	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_210s0042g00820	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_210s0116g00750	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_211s0016g02170	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓
VIT_211s0052g01800	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_212s0028g03350	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_212s0059g01180	✓	✓	✓	✓	✓	✓	X	✓	X	✓	X	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓
VIT_213s0019g03550	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_213s0067g03390	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_214s0006g00640	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_214s0030g00440	X	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_214s0030g01660	✓	X	X	✓	X	✓	✓	✓	X	✓	✓	✓	X	X	✓	X	✓	X	✓	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_214s0060g00100	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_214s0060g00260	✓	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	X	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
VIT_214s0060g02440	X	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	X	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
VIT_214s0066g02630	✓	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	X	✓	✓	✓	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
VIT_214s0068g01800	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓
VIT_214s0081g00440	X	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_214s0083g00130	X	X	X	X	X	X	X	X	✓	X	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_214s0083g00640	✓	✓	✓	X	X	✓	X	X	X	X	X	✓	✓	✓	X	✓	✓	X	✓	✓	X	X	X	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓
VIT_214s0083g01030	✓	✓	✓	✓	✓	X	X	✓	X	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	X	X	X	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_214s0083g01050	✓	✓	✓	✓	✓	X	✓	✓	X	✓	✓	✓	X	✓	✓	✓	X	✓	✓	✓	X	X	X	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_214s0108g00490	✓	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	X	✓	✓	✓	X	X	X	X	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓
VIT_214s0108g00980	✓	X	X	X	X	X	X	X	X	X	X	X	X	✓	X	✓	X	✓	✓	✓	X	X	X	X	✓	X	✓	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓	✓
VIT_215s0048g01280	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_215s0048g01320	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_216s0013g00860	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g00150	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g02630	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g04990	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g06570	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g08480	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g10300	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Gene-ID/Sample-ID	89-030-22	89-030-29	89-030-52	89-030-54	89-030-57	89-030-95	89-030-98	89-030-152	89-030-185	89-030-198	89-030-201	89-030-247	89-030-257	89-030-273	89-030-309	89-030-328	89-030-329	89-030-349	89-030-371	89-030-378	89-030-384	89-030-434	89-030-461	89-030-503	89-030-508	89-030-544	89-030-560	89-030-566	89-030-575	89-030-578	89-030-584	89-030-594	89-030-596	89-030-632	89-030-640	GF.GA-47-42	'Villard Blanc'
VIT_218s0001g01800	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_218s0001g07460	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_219s0090g00590	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_219s0140g00120	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table A6: Sequenced amplicons in other analyzed cultivars.

Gene-ID/Sample-ID	'Börner'	'Chardonnay'	GF V.3125	GF.GA-52-42	'Grenache Noir'	'Pinot Noir'	'Pinot Noir précoce'	'Riesling'	'Silvaner'	'Solaris'	'Syrah'
GSVIVT01001405001	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_200s0203g00080	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_200s1675g00010	X	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0010g00730	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0010g00740	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0010g02270	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0010g03890	X	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0010g03900	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0011g00100	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0011g00110	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0011g01240	X	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0011g01560	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0011g02120	✓	✓	✓	✓	✓	X	X	✓	✓	✓	X
VIT_201s0011g03070	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_201s0011g03140	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_201s0011g03520	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0011g04240	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0011g05260	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0011g05810	X	X	X	X	X	X	X	X	X	X	✓
VIT_201s0011g06410	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0026g00150	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_201s0026g02200	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_201s0026g02580	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_203s0017g02143	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_203s0167g00070	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_204s0023g02820	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_204s0044g00850	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_205s0020g03150	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_205s0102g01160	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_206s0004g03590	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_206s0009g02480	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_207s0005g02260	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_207s0104g01590	✓	✓	✓	✓	✓	X	X	✓	✓	X	✓
VIT_208s0007g04200	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_208s0056g01230	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_209s0002g02680	✓	X	✓	✓	✓	X	✓	✓	✓	✓	✓
VIT_210s0042g00820	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_210s0116g00750	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Gene-ID/Sample-ID	'Börner'	'Chardonnay'	GF V.3125	GF.GA-52-42	'Grenache Noir'	'Pinot Noir'	'Pinot Noir précoce'	'Riesling'	'Silvaner'	'Solaris'	'Syrah'
VIT_211s0016g02170	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_211s0052g01800	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_212s0028g03350	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_212s0059g01180	✓	X	✓	X	X	X	X	X	X	✓	✓
VIT_213s0019g03550	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_213s0067g03390	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_214s0006g00640	X	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0030g00440	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_214s0030g01660	X	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0060g00100	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0060g00260	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0060g02440	✓	✓	X	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0066g02630	X	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0068g01800	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0081g00440	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_214s0083g00130	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_214s0083g00640	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0083g01030	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_214s0083g01050	✓	✓	✓	✓	✓	✓	X	✓	✓	✓	✓
VIT_214s0108g00490	✓	✓	✓	X	✓	X	X	✓	✓	✓	✓
VIT_214s0108g00980	✓	✓	✓	X	✓	X	X	✓	✓	✓	✓
VIT_215s0048g01280	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_215s0048g01320	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_216s0013g00860	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g00150	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓
VIT_217s0000g02630	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g04990	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g06570	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g08480	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_217s0000g10300	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_218s0001g01800	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_218s0001g07460	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_219s0090g00590	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VIT_219s0140g00120	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table A7: Percentage of samples that passed different attributes according to FastQC's quality control in seven the seven runs performed for amplicon sequencing.

Attribute		**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**	**Run 6**	**Run 7**
Kmer Content	raw	25%	13%	0%	29%	20%	16%	39%
	trimmed	11%	6%	0%	29%	20%	16%	3%
Overrepresented sequences	raw	0%	6%	7%	32%	88%	71%	72%
	trimmed	0%	6%	4%	21%	88%	71%	84%
Per base GC content	raw	36%	18%	0%	46%	5%	16%	31%
	trimmed	39%	26%	18%	43%	5%	16%	78%
Per base N content	raw	94%	28%	0%	50%	50%	53%	41%
	trimmed	100%	100%	100%	100%	50%	53%	100%
Per base sequence content	raw	22%	0%	0%	0%	0%	0%	0%
	trimmed	11%	0%	0%	0%	0%	0%	0%
Per sequence GC content	raw	19%	35%	86%	43%	100%	100%	61%
	trimmed	17%	36%	50%	29%	100%	100%	91%
Per sequence quality scores	raw	100%	100%	100%	100%	100%	100%	100%
	trimmed	100%	100%	100%	100%	100%	100%	100%
Sequence Duplication Levels	raw	0%	0%	4%	0%	20%	31%	2%
	trimmed	0%	0%	29%	0%	20%	31%	6%
Sequence Length Distribution	raw	100%	100%	100%	100%	100%	100%	100%
	trimmed	100%	100%	100%	100%	100%	100%	100%

Table A8: Statistics of read mappings for amplicon sequenced cutivars.

Cultivar	Read pair unique	Unpaired read unique	Read pair not optical duplicate	Read pair optical duplicate	Unpaired read duplicate	Unmapped reads	Mapped reads	Total reads	Percent mapped	Percent duplicates	Percent unmapped
89-30-544	422,894	1,677	429,161	525	8,117	1,008,218	847,465	1,855,683	46%	24%	54%
'Silvaner'	330,309	1,191	332,154	131	5,353	707,266	661,809	1,369,075	48%	25%	52%
89-30-328	398,995	960	423,865	120	6,272	805,200	798,950	1,604,150	50%	27%	50%
'Börner'	369,760	1,682	324,603	52	8,166	707,550	741,202	1,448,752	51%	23%	49%
89-30-273	365,503	882	460,413	2,573	5,892	653,742	731,888	1,385,630	53%	34%	47%
'Chardonnay'	302,421	925	420,113	1,912	6,338	506,309	605,767	1,112,076	54%	39%	46%
GF V.3125	473,700	895	432,646	474	3,968	692,267	948,295	1,640,562	58%	27%	42%
'Riesling'	323,084	746	351,105	569	3,770	447,390	646,914	1,094,304	59%	32%	41%
'Solaris'	401,705	938	356,395	57	6,867	550,785	804,348	1,355,133	59%	27%	41%
GF.GA-52-52	417,798	1,851	400,945	1,075	7,783	546,274	837,447	1,383,721	61%	30%	39%
'Pinot Noir précoce'	122,250	408	41,035	8	1,592	159,452	244,908	404,360	61%	11%	39%
89-30-371	455,277	764	440,110	228	2,849	542,061	911,318	1,453,379	63%	30%	37%
'Pinot Noir'	209,040	521	90,050	27	3,528	228,469	418,601	647,070	65%	14%	35%
89-30-384	53,707	220	9,651	1	382	54,350	107,634	161,984	66%	6%	34%
89-30-503	556,826	911	506,695	238	6,383	500,142	1,114,563	1,614,705	69%	32%	31%
89-30-632	142,804	525	35,583	13	1,317	119,612	286,133	405,745	71%	9%	29%
GF.GA-47-42	550,082	869	517,172	90	7,283	454,936	1,101,033	1,555,969	71%	34%	29%
'Grenache Noir'	701,153	1,218	561,290	1,465	11,470	570,738	1,403,524	1,974,262	71%	29%	29%
89-30-596	113,099	438	26,195	9	816	88,962	226,636	315,598	72%	9%	28%
89-30-640	233,445	802	60,091	18	1,914	179,176	467,692	646,868	72%	10%	28%
89-30-098	244,165	784	119,562	42	2,349	184,123	489,114	673,237	73%	18%	27%
89-30-095	471,670	688	499,820	426	7,029	348,933	944,028	1,292,961	73%	39%	27%
'Syrah'	504,538	731	577,864	159	3,680	363,077	1,009,807	1,372,884	74%	42%	26%
89-30-201	580,633	1,471	233,800	68	5,224	402,911	1,162,737	1,565,648	74%	15%	26%
89-30-461	406,670	656	134,784	33	5,527	270,221	813,996	1,084,217	75%	13%	25%
89-30-578	107,473	441	21,960	7	954	70,805	215,387	286,192	75%	8%	25%
89-30-584	269,284	601	70,862	15	1,920	173,705	539,169	712,874	76%	10%	24%
89-30-185	353,525	700	119,429	35	4,015	224,965	707,750	932,715	76%	13%	24%
89-30-054	131,226	636	70,448	21	935	82,181	263,088	345,269	76%	21%	24%
89-30-247	154,517	612	65,273	25	1,190	93,058	309,646	402,704	77%	17%	23%
89-30-198	159,534	748	56,591	19	982	93,436	319,816	413,252	77%	14%	23%
89-30-508	147,595	839	36,699	13	1,370	86,141	296,029	382,170	77%	10%	23%
89-30-057	176,529	1,011	58,201	17	2,283	102,590	354,069	456,659	78%	13%	22%
89-30-378	239,110	386	53,624	19	1,380	136,476	478,606	615,082	78%	9%	22%
89-30-022	675,037	779	749,623	426	7,704	376,733	1,350,853	1,727,586	78%	44%	22%

Cultivar	Read pair unique	Unpaired read unique	Read pair not optical duplicate	Read pair optical duplicate	Unpaired read duplicates	Unmapped reads	Mapped reads	Total reads	Percent mapped	Percent duplicates	Percent unmapped
89-30-257	150,586	461	51,431	20	1,639	83,838	301,633	385,471	78%	14%	22%
89-30-560	116,058	240	21,202	10	554	63,790	232,356	296,146	78%	7%	22%
89-30-029	152,149	628	68,274	26	1,170	82,350	304,926	387,276	79%	18%	21%
89-30-152	132,146	421	113,259	88	998	70,703	264,713	335,416	79%	34%	21%
89-30-575	347,300	862	107,853	48	1,651	181,819	695,462	877,281	79%	12%	21%
89-30-434	149,906	422	34,927	14	755	72,847	300,234	373,081	80%	10%	20%
89-30-052	153,121	398	79,019	33	515	68,923	306,640	375,563	82%	21%	18%
89-30-566	152,830	369	37,801	19	805	68,622	306,029	374,651	82%	10%	18%
89-30-309	115,816	287	73,065	14	464	51,723	231,919	283,642	82%	26%	18%
89-30-329	169,913	645	36,255	11	710	71,313	340,471	411,784	83%	9%	17%
'Villard Blanc'	532,961	787	545,748	128	3,165	208,838	1,066,709	1,275,547	84%	43%	16%
89-30-349	124,815	312	32,482	14	367	45,403	249,942	295,345	85%	11%	15%
89-30-594	147,202	154	32,398	16	358	46,014	294,558	340,572	86%	10%	14%

Table A9: Distribution of alleles of phased FTC genes within the 35 amplicon sequenced F1-individuals of the mapping population GF.GA-47-42 x 'Villard Blanc'.

Gene-ID/Sample-ID	89-030-22	89-030-29	89-030-52	89-030-54	89-030-57	89-030-95	89-030-98	89-030-152	89-030-185
VIT_201s0011g00110	N2/L1	E1/L1	E1/N2	E1/L1	/	N2/L1	E1/N2	N2/N2	/
VIT_201s0011g01560	E2/L1	E1/L1	E1/L2	E1/L1	E1/L2	E2/L1	E1/L2	E1/L2	E1/L2
VIT_201s0011g02120	/	E1/L1	E1/L2	E1/L1	E1/L2	E2/L1	E1/L2	E1/L2	/
VIT_201s0011g03070	E0/L1	/	E0/L2	/	E0/L2	E0/L1	/	/	/
VIT_201s0011g03140	E2/L1	E1/L1	E1/L2	E1/L1	E1/L2	E2/L1	E1/L2	E1/L2	/
VIT_201s0011g04240	E2/L1	E1/L1	E1/L1	/	E1/L2	E2/L1	E1/L2	E1/L2	/
VIT_201s0011g05260	N/L2	E1/N	E1/N	E1/N	E1/N	N/N	E1/N	E1/L2	/
VIT_201s0011g06410	E2/L0	E1/L0	E2/L0	E1/L0	E1/L0	/	/	E1/L0	E2/L0
VIT_201s0026g00150	E2/L1	E1/L1	E2/L2	E1/L2	E1/L2	E2/L1	E1/L2	E1/L2	E2/L1
VIT_201s0026g02200	E2/L1	E1/L1	E2/L2	E2/L1	E1/L2	E2/L1	E1/L2	E1/L2	E2/L1
VIT_201s0026g02580	E2/N1	N1/N1	N1/L2	E2/N1	N1/L2	E2/N1	N1/L2	N1/L2	/
VIT_201s0010g00740	E2/L1	E1/L1	E2/L2	/	E1/L2	E2/L1	E1/L2	E1/L2	/
VIT_201s0010g02270	/	E1/L1	E2/L2	E1/L2	E2/L2	E2/L1	E1/L2	E2/L1	E2/L1
VIT_201s0010g03890	E2/N1	N1/N1	/	N1/L2	N1/L2	E2/N1	E2/N1	N1/L2	E2/N1
VIT_204s0023g02820	E1/N2	N2/N2	E1/L1	E1/L1	E1/N2	E1/N2	E1/L1	N2/N2	E1/L1
VIT_204s0044g00850	E1/N2	E1/N2	E1/L1	E1/L1	E1/N2	E1/N2	E1/L1	N2/L1	E1/L1
VIT_205s0102g01160	E1/L1	E2/L2	E2/L1	E1/L2	E2/L1	E2/L2	E2/L2	E2/L2	E1/L1
VIT_206s0004g03590	Na/Na	Na/L2	Na/L2	Na/L2	Na/Na	Na/Na	Na/Na	Na/Na	Na/L2
VIT_206s0009g02480	E1/L1	E1/L1	E2/L1	E2/L1	E1/L1	E2/L1	E2/L2	E2/L2	E2/L1
VIT_207s0104g01590	N1/N2	/	N1/N2	N2/N2	/	N1/N2	/	/	N1/N2
VIT_207s0005g02260	E1/L1	E1/L1	E1/L1	E1/L1	E1/L1	E1/L1	E2/L1	E1/L1	E2/L1
VIT_208s0056g01230	Nb/L1	Nb/Nb	Nb/Nb	Nb/L1	L1/L1	Nb/Nb	Nb/Nb	Nb/Nb	Nb/Nb
VIT_208s0007g04200	Na/L2	Na/Na	Na/Na	Na/Na	Na/L2	Na/L2	Na/Na	Na/Na	Na/Na
VIT_210s0116g00750	E0/L1	E0/L2	E0/L2	E0/L1	E0/L2	E0/L1	E0/L2	E0/L2	E0/L2
VIT_211s0052g01800	E1/L1	E2/L2	E1/L1	E1/L1	E2/L1	E2/L1	E1/L2	E1/L1	E2/L1
VIT_212s0028g03350	E2/L2	E1/L2	E1/L2	E2/L2	E2/L1	E2/L2	E1/L2	E1/L2	E2/L1
VIT_213s0067g03390	E1/L2	E1/L1	/	E1/L2	E2/L1	E1/L1	E1/L2	E1/L2	E2/L2
VIT_213s0019g03550	E1/L2	E1/L1	E1/L2	E1/L2	E2/L1	E1/L1	E1/L2	E1/L1	E2/L2
VIT_214s0030g00440	/	E2/L2	E2/L1	/	E2/L1	/	E1/L2	E1/L1	/
VIT_214s0081g00440	/	E2/L2	E2/L1	E2/L1	E2/L1	/	E1/L2	E1/L1	/
VIT_214s0006g00640	E1/L1	E2/L2	E2/L2	E2/L2	E2/L1	E1/L2	E1/L2	E1/L1	/
VIT_214s0083g01030	E1/L0	E2/L0	E2/L0	E2/L0	E2/L0	/	E1/L0	E1/L0	/
VIT_214s0083g01050	N1/N1	E2/L2	N1/L2	E2/N1	/	/	N1/L2	N1/N1	/
VIT_214s0068g01800	E1/L1	E2/L2	E2/L1	E2/L1	E2/L1	E1/L2	E1/L2	E1/L1	/
VIT_215s0048g01280	E0/L1	E0/L1	E0/L1	E0/L1	E0/L1	E0/L2	E0/L2	E0/L1	E0/L2
VIT_216s0013g00860	N1/N1	E2/N1	N1/L2	E2/L2	N1/N1	N1/L2	E2/N1	E2/N1	N1/N1
VIT_217s0000g00150	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	/
VIT_217s0000g02630	E2/L1	E2/L1	E2/L1	E2/L1	E1/L2	E2/L1	E1/L1	E1/L1	E2/L1
VIT_217s0000g06570	E2/L2	E2/L2	E2/N1	E2/L2	N1/L2	E2/L2	N1/L2	N1/L2	E2/L2
VIT_217s0000g08480	E2/N1	E2/N1	N1/L2	/	E2/L2	/	N1/L2	E2/N1	/
VIT_217s0000g10300	E1/L2	E1/L2	E2/L1	E2/L2	E1/L1	E2/L2	E1/E1	E2/L2	E2/L2
VIT_218s0001g07460	E2/L2	E1/L1	E1/L1	E2/L2	E1/L2	E1/L1	E2/L1	E2/L2	E2/L2
VIT_218s0001g01800	E1/N2	N2/L1	N2/L1	E1/N2	E1/N2	N2/L1	E1/L1	/	E1/N2
VIT_200s0203g00080	N2/N2	N2/L1	/	N2/L1	/	/	N2/L1	N2/L1	N2/L1
VIT_200s1675g00010	/	E2/L2	E2/N1	N1/N1	E2/L2	E2/N1	E2/L2	E2/N1	E2/L2

Gene-ID/Sample-ID	89-030-198	89-030-201	89-030-247	89-030-257	89-030-273	89-030-309	89-030-328	89-030-329	89-030-349
VIT_201s0011g00110	E1/L1	E1/N2	E1/L1	N2/L1	N2/L1	N2/L1	N2/N2	E1/L1	/
VIT_201s0011g01560	E1/L1	E2/L1	E1/L2	E1/L1	E2/L1	E2/L2	E2/L2	E1/L1	E1/L1
VIT_201s0011g02120	E1/L1	E2/L1	E1/L2	E1/L1	E2/L1	/	E2/L2	E1/L1	E1/L1
VIT_201s0011g03070	/	/	E0/L2	/	E0/L1	E0/L2	E0/L2	/	E0/L1
VIT_201s0011g03140	E1/L1	E2/L1	E1/L2	/	E2/L1	E2/L1	E2/L2	E1/L1	E1/L1
VIT_201s0011g04240	E1/L1	E2/L1	E1/L2	E1/L1	E2/L2	E2/L1	E2/L2	E1/L1	E1/L1
VIT_201s0011g05260	E1/N	/	E1/N	E1/N	N/L2	E1/L2	E1/N	E1/L2	E1/N
VIT_201s0011g06410	E1/L0	E2/L0	E1/L0	E1/L0	E2/L0	E1/L0	E1/L0	E1/L0	E1/L0
VIT_201s0026g00150	E1/L1	E2/L1	E1/L1	E1/L2	E2/L2	E1/L2	E1/L1	E1/L2	E1/L2
VIT_201s0026g02200	E2/N1	E2/N1	E2/N1	N1/L2	L2/E2	N1/L2	N1/N1	N1/L2	N1/L2
VIT_201s0026g02580	/	E2/L1	E2/L1	E1/L2	E2/L2	/	E1/L1	E1/L2	E1/L2
VIT_201s0010g00740	E2/L2	E2/L1	E1/L1	E1/L2	E2/L2	E1/L1	E1/L1	E1/L2	E1/L2
VIT_201s0010g02270	E1/L1	E2/L1	E1/L1	E1/L2	E2/L2	E2/L1	E1/L1	E1/L2	E1/L2
VIT_201s0010g03890	N1/N1	E2/N1	N1/N1	N1/L2	/	E2/N1	N1/N1	N1/L2	N1/L2
VIT_204s0023g02820	N2/N2	E1/N2	E1/N2	E1/L1	E1/L1	N2/L1	E1/L1	N2/L1	E1/N2
VIT_204s0044g00850	N2/N2	E1/N2	E1/N2	E1/L1	E1/L1	E1/L1	E1/L1	N2/L1	E1/N2
VIT_205s0102g01160	E2/L1	E2/L1	E2/L1	E2/L1	E1/L2	E2/L1	E2/L2	E2/L2	E2/L2
VIT_206s0004g03590	Na/L2	Na/Na	Na/L2	Na/L2	Na/Na	Na/L2	Na/L2	Na/L2	Na/Na
VIT_206s0009g02480	E1/L1	E1/L2	E2/L2	E1/L1	E2/L2	E2/L1	E1/L1	E1/L1	E1/L1
VIT_207s0104g01590	N2/N2	N2/N2	N2/N2	N1/N2	/	/	N1/N2	N2/N2	N1/N2
VIT_207s0005g02260	E1/L1	E1/L1	E1/L1	E2/L1	E2/L1	E1/L1	E2/L1	E1/L1	E2/L1
VIT_208s0056g01230	Nb/Nb	Nb/L1	Nb/L1	Nb/Nb	Nb/Nb	Nb/L1	Nb/Nb	Nb/Nb	Nb/L1
VIT_208s0007g04200	Na/Na	Na/Na	Na/Na	Na/L2	Na/L2	Na/L2	Na/L2	Na/L2	Na/L2
VIT_210s0116g00750	E0/L1	E0/L1	E0/L2	E0/L2	E0/L2	E0/L1	E0/L2	E0/L1	E0/L1
VIT_211s0052g01800	E2/L1	E2/L1	E2/L2	E2/L1	E2/L2	E2/L2	E2/L2	E1/L2	/
VIT_212s0028g03350	E2/L1	E2/L1	E1/L2	E2/L2	/	E2/L1	E1/L2	E2/L1	E2/L2
VIT_213s0067g03390	/	E1/L2	E1/L2	E2/L2	E2/L2	E1/L2	E1/L1	E2/L2	E1/L2
VIT_213s0019g03550	E1/L2	E1/L2	E1/L1	E1/L2	E2/L2	/	E1/L1	E2/L2	E1/L2
VIT_214s0030g00440	E2/L1	E2/L1	E2/L1	/	E2/L2	E2/L1	E2/L2	E2/L1	E2/L2
VIT_214s0081g00440	E2/L1	E2/L1	E2/L1	E2/L2	E2/L2	E2/L1	E1/L2	E2/L1	E2/L2
VIT_214s0006g00640	E2/L1	E2/L1	E2/L1	E2/L2	E2/L2	E2/L1	E1/L2	E2/L1	E2/L2
VIT_214s0083g01030	E2/L0	E2/L0	E2/L0	E2/L0	/	E2/L0	E1/L0	E2/L0	E2/L0
VIT_214s0083g01050	E2/N1	E2/N1	E2/N1	E2/L2	/	E2/N1	N1/N1	E2/N1	E2/L2
VIT_214s0068g01800	E1/L2	E2/L1	E2/L1	E2/L2	E2/L1	E2/L1	E1/L1	E2/L1	E2/L2
VIT_215s0048g01280	E0/L2	E0/L2	E0/L1	E0/L2	E0/L2	E0/L2	E0/L1	E0/L1	E0/L2
VIT_216s0013g00860	E2/L2	E2/N1	N1/N1	E2/N1	N1/N1	E2/N1	N1/N1	N1/L2	N1/N1
VIT_217s0000g00150	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0
VIT_217s0000g02630	E2/L2	E2/L2	E2/L2	E1/L1	E2/L1	/	E1/L1	E2/L1	E2/L2
VIT_217s0000g06570	E2/N1	N1/N1	E2/N1	N1/L2	E2/L2	N1/L2	N1/L2	N1/L2	E2/N1
VIT_217s0000g08480	N1/L2	E2/L2	/	E2/N1	/	E2/N1	E2/N1	E2/N1	N1/L2
VIT_217s0000g10300	E2/L1	E1/L1	E2/L2	E1/L2	E1/L2	/	E1/L2	E1/L2	E2/L1
VIT_218s0001g07460	E2/L1	E2/L2	E2/L1	E1/L2	E1/L1	/	E2/L2	E2/L2	E2/L1
VIT_218s0001g01800	/	E1/N2	N2/L1	/	/	E1/N2	E1/N2	E1/N2	N2/L1
VIT_200s0203g00080	N2/L1	N2/N2	N2/L1	N2/L1	N2/L1	N2/L1	/	E1/N2	/
VIT_200s1675g00010	E2/N1	/	N1/N1	/	N1/N1	E2/N1	E2/L2	E2/L2	E2/L2

Gene-ID/Sample-ID	89-030-371	89-030-378	89-030-384	89-030-434	89-030-461	89-030-503	89-030-508	89-030-544	89-030-560
VIT_201s0011g00110	N2/N2	N2/N2	/	E1/N2	N2/N2	E1/N2	E1/N2	E1/N2	/
VIT_201s0011g01560	E2/L2	E2/L2	E1/L1	E1/L2	E2/L2	E1/L2	E1/L2	E1/L2	E2/L2
VIT_201s0011g02120	E2/L2	E2/L2	/	E1/L2	E2/L2	E2/L2	E1/L2	E1/L2	E2/L2
VIT_201s0011g03070	E0/L2	E0/L2	/	E0/L2	E0/L2	E0/L2	/	/	E0/L2
VIT_201s0011g03140	E2/L2	E2/L2	/	E1/L2	E2/L2	E2/L2	E1/L2	E1/L2	E2/L2
VIT_201s0011g04240	E2/L2	E2/L2	/	E1/L2	E2/L2	E2/L2	E1/L2	E2/L2	E2/L2
VIT_201s0011g05260	N/L2	N/N	/	E1/N	N/L2	N/L2	E1/L2	N/L2	/
VIT_201s0011g06410	/	E2/L0	E2/L0	E1/L0	E2/L0	E2/L0	E1/L0	E2/L0	E2/L0
VIT_201s0026g00150	E2/L1	E2/L2	E2/L1	E1/L1	E2/L1	E2/L1	E1/L2	E2/L1	E2/L2
VIT_201s0026g02200	E2/N1	N1/L2	/	N1/L2	N1/N1	E2/N1	N1/N1	E2/N1	L2/E2
VIT_201s0026g02580	E2/L1	E1/L2	E2/L1	E1/L2	E1/L1	E2/L1	E1/L1	E2/L1	E2/L2
VIT_201s0010g00740	E2/L2	E2/L2	/	E1/L1	E2/L1	E2/L2	E1/L2	E2/L1	E2/L2
VIT_201s0010g02270	E2/L1	E2/L2	E2/L1	E1/L1	E2/L1	E2/L1	E1/L1	E2/L1	/
VIT_201s0010g03890	E2/N1	/	E2/N1	N1/N1	E2/N1	E2/N1	N1/N1	E2/N1	/
VIT_204s0023g02820	E1/L1	N2/N2	E1/L1	E1/L1	E1/L1	N2/L1	E1/N2	N2/L1	N2/N2
VIT_204s0044g00850	E1/N2	N2/N2	E1/L1	E1/N2	E1/L1	N2/L1	E1/L1	N2/L1	N2/N2
VIT_205s0102g01160	E2/L1	E1/L2	E2/L1	E1/L1	E2/L2	E2/L1	E2/L1	E1/L1	E2/L1
VIT_206s0004g03590	Na/L2	Na/Na	Na/Na	Na/Na	Na/L2	Na/L2	Na/L2	Na/L2	Na/Na
VIT_206s0009g02480	E2/L1	E2/L2	E2/L1	E1/L2	E2/L1	E1/L1	E1/L1	E1/L2	E2/L1
VIT_207s0104g01590	N2/N2	/	/	N1/N1	N2/N2	N1/N1	N1/N2	N1/N2	/
VIT_207s0005g02260	E1/L1	E2/L1	E1/L1	E2/L1	E1/L1	E2/L1	E2/L1	E2/L1	E1/L1
VIT_208s0056g01230	Nb/Nb	Nb/L1	Nb/L1	Nb/Nb	Nb/L1	Nb/L1	Nb/Nb	Nb/Nb	L1/L1
VIT_208s0007g04200	Na/Na	Na/L2	Na/Na	Na/Na	Na/L2	Na/L2	Na/Na	Na/Na	Na/Na
VIT_210s0116g00750	E0/L1	E0/L2	E0/L1	E0/L1	E0/L2	E0/L1	E0/L2	E0/L2	E0/L1
VIT_211s0052g01800	E2/L2	E1/L2	E2/L1	E1/L1	E1/L1	E1/L1	E1/L1	E1/L2	E1/L1
VIT_212s0028g03350	E2/L2	E2/L2	E2/L2	E1/L1	E1/L1	E1/L2	E1/L1	E2/L2	E2/L1
VIT_213s0067g03390	E2/L2	/	E2/L1	E2/L1	E1/L1	E2/L2	E2/L2	E2/L2	E2/L1
VIT_213s0019g03550	E2/L2	E1/L2	E2/L1	E2/L1	E1/L1	E2/L2	E2/L2	E2/L2	E2/L1
VIT_214s0030g00440	E2/L1	E2/L2	/	E2/L2	E1/L2	E1/L2	/	E2/L2	E1/L2
VIT_214s0081g00440	E2/L1	E2/L2	/	E2/L1	E1/L2	E1/L2	E2/L1	E2/L1	E1/L2
VIT_214s0006g00640	E1/L1	E1/L2	/	E2/L1	E1/L2	E1/L2	E2/L1	E2/L1	E1/L2
VIT_214s0083g01030	E1/L0	E1/L0	/	E2/L0	E1/L0	E1/L0	E2/L0	E1/L0	E1/L0
VIT_214s0083g01050	/	N1/L2	/	E2/N1	N1/L2	N1/L2	E2/L2	N1/N1	N1/N1
VIT_214s0068g01800	E1/L2	E1/L2	/	E2/L1	E2/L2	E1/L2	E2/L2	E1/L1	E1/L1
VIT_215s0048g01280	E0/L1	E0/L1	E0/L2	E0/L1	/	E0/L1	E0/L2	E0/L2	E0/L2
VIT_216s0013g00860	E2/L2	E2/L2	E2/L2	E2/L2	E2/N1	N1/N1	E2/N1	N1/N1	E2/N1
VIT_217s0000g00150	N0/N0	N0/N0	/	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0	N0/N0
VIT_217s0000g02630	E1/L2	E1/L2	E2/L1	E2/L2	E1/L1	E1/L1	E2/L2	E2/L1	E2/L2
VIT_217s0000g06570	N1/N1	E2/L2	N1/L2	N1/N1	E2/L2	N1/L2	E2/L2	E2/L2	E2/N1
VIT_217s0000g08480	E2/L2	/	E2/N1	E2/L2	/	/	E2/L2	E2/N1	N1/L2
VIT_217s0000g10300	E2/L1	E2/L2	E1/L2	E1/L2	E2/L2	E2/L2	E2/L1	E1/L2	E2/L1
VIT_218s0001g07460	E2/L1	E1/L2	E1/L1	E1/L1	E2/L2	E2/L2	E1/L1	E2/L2	E1/L2
VIT_218s0001g01800	N2/L1	/	N2/N2	E1/L1	E1/N2	E1/N2	/	E1/N2	/
VIT_200s0203g00080	N2/N2	N2/L1	N2/N2	N2/L1	/	N2/L1	N2/L1	/	N2/L1
VIT_200s1675g00010	E2/L2	N1/N1	E2/L2	N1/N1	N1/N1	E2/L2	E2/N1	E2/N1	N1/N1

Gene-ID/Sample-ID	89-030-566	89-030-575	89-030-578	89-030-584	89-030-594	89-030-596	89-030-632	89-030-640
VIT_201s0011g00110	N2/L1	N2/L1	/	/	N2/N2	/	/	/
VIT_201s0011g01560	E2/L1	E2/L1	E2/L2	E2/L2	E2/L2	E2/L1	E2/L2	E2/L2
VIT_201s0011g02120	E2/L1	E2/L1	/	/	E2/L2	/	/	/
VIT_201s0011g03070	/	E0/L1	/	/	E0/L2	/	/	/
VIT_201s0011g03140	E2/L1	E2/L1	/	/	E2/L2	/	/	/
VIT_201s0011g04240	E2/L1	E2/L1	/	/	E2/L2	/	/	/
VIT_201s0011g05260	/	N/N	/	/	N/L2	/	/	/
VIT_201s0011g06410	E2/L0	E2/L0	E1/L0	E2/L0	E2/L0	E1/L0	E2/L0	E2/L0
VIT_201s0026g00150	E2/L1	E2/L2	E1/L2	E1/L1	E1/L1	E2/L2	E1/L1	E2/L1
VIT_201s0026g02200	N1/N1	L2/E2	/	/	N1/L2	/	/	/
VIT_201s0026g02580	E2/L1	E2/L2	E1/L1	E1/L1	E1/L2	E1/L2	E1/L1	E2/L2
VIT_201s0010g00740	E2/L1	E2/L1	/	/	E1/L2	/	/	/
VIT_201s0010g02270	E2/L1	E2/L2	E1/L1	E2/L1	E1/L2	E1/L2	E2/L2	E2/L2
VIT_201s0010g03890	E2/N1	/	N1/N1	E2/N1	N1/L2	N1/L2	/	/
VIT_204s0023g02820	N2/N2	E1/N2	N2/N2	N2/L1	E1/N2	E1/N2	N2/N2	E1/N2
VIT_204s0044g00850	N2/N2	E1/N2	E1/L1	N2/L1	E1/N2	E1/L1	N2/L1	E1/N2
VIT_205s0102g01160	E2/L1	E2/L2	E2/L1	E2/L2	E2/L1	E2/L2	E2/L2	E2/L1
VIT_206s0004g03590	Na/L2	Na/L2	Na/Na	Na/Na	Na/L2	Na/Na	Na/L2	Na/L2
VIT_206s0009g02480	E1/L1	E1/L1	E1/L1	E2/L1	E2/L1	E2/L1	E2/L1	E2/L1
VIT_207s0104g01590	N1/N1	/	/	/	N2/N2	N2/N2	/	/
VIT_207s0005g02260	E1/L1	E2/L1	E1/L1	E2/L1	E1/L1	E1/L1	E2/L1	E1/L1
VIT_208s0056g01230	L1/L1	Nb/Nb	Nb/L1	Nb/L1	Nb/Nb	L1/L1	Nb/Nb	Nb/L1
VIT_208s0007g04200	Na/L2	Na/Na	Na/Na	Na/Na	Na/L2	Na/Na	Na/Na	Na/Na
VIT_210s0116g00750	E0/L2	E0/L1	E0/L1	E0/L1	E0/L1	E0/L1	E0/L2	E0/L1
VIT_211s0052g01800	E2/L2	E2/L2	E2/L1	E2/L2	E1/L1	E1/L2	E2/L1	E2/L2
VIT_212s0028g03350	E1/L2	E2/L2	E2/L1	E2/L2	E2/L1	E2/L2	E1/L1	E2/L1
VIT_213s0067g03390	/	/	E2/L1	E1/L2	E1/L2	E2/L2	/	/
VIT_213s0019g03550	E1/L2	E1/L2	E2/L1	E1/L2	E1/L2	E2/L2	E2/L1	E2/L2
VIT_214s0030g00440	E2/L1	E2/L2	/	/	E1/L1	/	/	/
VIT_214s0081g00440	E2/L1	E2/L2	/	/	E1/L1	/	/	/
VIT_214s0006g00640	E1/L2	E2/L2	/	/	E1/L1	/	/	/
VIT_214s0083g01030	E1/L0	/	/	/	E1/L0	/	/	/
VIT_214s0083g01050	N1/N1	N1/L2	/	/	N1/N1	/	/	/
VIT_214s0068g01800	E1/L1	E1/L2	/	/	E1/L2	/	/	/
VIT_215s0048g01280	E0/L1	E0/L2	E0/L1	E0/L2	E0/L2	E0/L2	E0/L1	E0/L2
VIT_216s0013g00860	N1/N1	E2/N1	E2/N1	E2/L2	E2/L2	E2/N1	E2/L2	E2/N1
VIT_217s0000g00150	N0/N0	N0/N0	/	/	N0/N0	/	/	/
VIT_217s0000g02630	E1/L2	E2/L2	E2/L2	E1/L1	E1/L1	E1/L1	E2/L1	E2/L2
VIT_217s0000g06570	E2/N1	E2/N1	N1/N1	E2/N1	E2/L2	N1/N1	N1/L2	N1/L2
VIT_217s0000g08480	E2/L2	N1/L2	/	N1/L2	E2/N1	E2/N1	/	N1/L2
VIT_217s0000g10300	E2/L1	E2/L1	E1/L2	E1/L1	E1/L1	E2/L1	E2/L2	E1/L1
VIT_218s0001g07460	E2/L1	E1/L2	E1/L1	E2/L1	E1/L2	E1/L2	E1/L2	E1/L2
VIT_218s0001g01800	N2/L1	E1/N2	/	N2/L1	E1/N2	/	/	E1/N2
VIT_200s0203g00080	N2/N2	/	/	/	N2/N2	N2/N2	/	N2/N2
VIT_200s1675g00010	E2/L2	E2/N1	E2/N1	E2/N1	E2/L2	E2/L2	E2/N1	E2/N1

Table A10: Correlation results of allele phased FTC candidate genes using different sets of phenotypic data including days after January 1st in the year 1999 and 2009 - 2016, as well as temperature sums above 3°C and global radiation from November 1st of the previous year. DotY: days of the year, TS: temperature sum, GR: global radiation. #: Number alleles. Md: Median

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
VIT_201s0011g00110	DotY 1999	0.84	-	0.09	-	12	12	12	12	167	168.5	167	170.5
	DotY 2009	0.3	-	0.17	-	12	13	12	13	162	164	162.5	164
	DotY 2010	0.91	-	0.03	-	12	13	12	13	177	178	175.5	178
	DotY 2011	0.91	-	0.04	-	12	13	12	13	151	153	150.5	155
	DotY 2012	0.85	-	0.05	-	12	13	12	13	159	161	157	161
	DotY 2013	0.06	-	0.42	-	11	13	12	12	172	178	172	174.5
	DotY 2014	0.5	-	0.17	-	12	13	12	13	156.5	157	156	157
	DotY 2015	0.07	-	0.18	-	12	12	11	13	161	164	162	164
	DotY 2016	0.29	-	0.12	-	12	13	12	13	171.5	173	172	173
	Average 1999-2016	0.66	-	0.05	-	12	13	12	13	0.47	0.6	0.49	0.7
	Median 1999-2016	0.57	-	0.04	-	12	13	12	13	0.44	0.6	0.48	0.69
	TS 2011	0.98	-	0.03	-	12	13	12	13	1170.61	1209.22	1165.79	1251.21
	TS 2012	0.98	-	0.03	-	12	13	12	13	1192.64	1207.58	1167.18	1218.09
	TS 2013	0.17	-	0.21	-	11	13	12	12	1180.44	1202.28	1180.44	1209.35
	TS 2014	0.87	-	0.09	-	12	13	12	13	1257.8	1270.24	1201.47	1270.24
	TS 2015	0.21	-	0.06	-	12	11	10	13	1233.01	1291.07	1213.66	1291.07
	TS 2016	0.29	-	0.12	-	12	13	12	13	1358.04	1392.58	1369.31	1392.58
	GR 2011	0.98	-	0.03	-	12	13	12	13	543.1	556.81	542.34	569.43
	GR 2012	0.98	-	0.03	-	12	13	12	13	542.42	550.2	530.74	554.09
	GR 2013	0.17	-	0.21	-	11	13	12	12	541.09	549.33	541.09	551.78
	GR 2014	0.87	-	0.09	-	12	13	12	13	557.05	561.04	537.65	561.04
	GR 2015	0.21	-	0.06	-	12	11	10	13	559.38	580.01	553.37	580.01
	GR 2016	0.29	-	0.12	-	12	13	12	13	514.22	525.6	517.14	525.6
VIT_201s0011g01560	DotY 1999	0.14	0.14	0.06	0.06	16	18	14	20	167	168.5	166.5	168
	DotY 2009	0.16	0.16	0.17	0.17	17	18	14	21	162	164	163.5	164
	DotY 2010	0.51	0.51	0.08	0.08	17	18	14	21	177	178	175	178
	DotY 2011	0.39	0.39	0.04	0.04	17	17	14	20	151	154	150.5	154
	DotY 2012	0.41	0.41	0.01	0.01	17	18	14	21	160	161	156.5	161
	DotY 2013	0.09	0.09	0.2	0.2	16	18	14	20	172	174	172	173.5
	DotY 2014	0.36	0.36	0.06	0.06	17	18	14	21	157	157	155	157
	DotY 2015	0.02	0.02	0.1	0.1	16	18	13	21	162	164	160	164
	DotY 2016	0.09	0.09	0.3	0.3	17	18	14	21	172	173	172	173
	Average 1999-2016	0.23	0.23	0.04	0.04	17	18	14	21	0.47	0.62	0.42	0.64
	Median 1999 - 2016	0.23	0.23	0.02	0.02	17	18	14	21	0.46	0.61	0.41	0.61
	TS 2011	0.47	0.47	0.05	0.05	17	17	14	20	1170.61	1209.22	1165.79	1220.32
	TS 2012	0.53	0.53	0.02	0.02	17	18	14	21	1207.58	1218.09	1159.02	1218.09
	TS 2013	0.21	0.21	0.36	0.36	16	18	14	20	1180.44	1202.28	1180.44	1191.16
	TS 2014	0.33	0.33	0.03	0.03	16	18	13	21	1245.36	1270.24	1201.47	1270.24
	TS 2015	0.05	0.05	0.14	0.14	15	17	12	20	1252.35	1291.07	1213.66	1271.71
	TS 2016	0.09	0.09	0.3	0.3	17	18	14	21	1369.31	1392.58	1369.31	1392.58
	GR 2011	0.47	0.47	0.05	0.05	17	17	14	20	543.1	556.81	542.34	560.23
	GR 2012	0.53	0.53	0.02	0.02	17	18	14	21	550.2	554.09	528.3	554.09
	GR 2013	0.21	0.21	0.36	0.36	16	18	14	20	541.09	549.33	541.09	543.96
	GR 2014	0.33	0.33	0.03	0.03	16	18	13	21	553.06	561.04	537.65	561.04
	GR 2015	0.05	0.05	0.14	0.14	15	17	12	20	565.4	580.01	553.37	572.7
	GR 2016	0.09	0.09	0.3	0.3	17	18	14	21	517.14	525.6	517.14	525.6
	DotY 1999	0.09	0.09	0.07	0.07	13	12	11	14	166	168.5	166	168
	DotY 2009	0.06	0.06	0.38	0.38	14	12	11	15	162	164	163	164

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
VIT_201s0011g02120	DotY 2010	0.24	0.24	0.17	0.17	14	12	11	15	175	178	175	178
	DotY 2011	0.12	0.12	0.12	0.12	14	11	11	14	151	154	150	153.5
	DotY 2012	0.13	0.13	0.03	0.03	14	12	11	15	156.5	161	156	161
	DotY 2013	0.08	0.08	0.2	0.2	14	11	11	14	171.5	177	172	172.5
	DotY 2014	0.08	0.08	0.08	0.08	14	12	11	15	155.5	158	155	157
	DotY 2015	0.02	0.02	0.35	0.35	13	12	10	15	160	164	160	162
	DotY 2016	0.05	0.05	0.13	0.13	14	12	11	15	171.5	173	172	173
	Average 1999-2016	0.08	0.08	0.05	0.05	14	12	11	15	0.42	0.65	0.37	0.64
	Median 1999 - 2016	0.08	0.08	0.04	0.04	14	12	11	15	0.41	0.67	0.41	0.66
	TS 2011	0.12	0.12	0.12	0.12	14	11	11	14	1170.61	1231.41	1160.96	1220.32
	TS 2012	0.13	0.13	0.03	0.03	14	12	11	15	1159.02	1218.09	1150.85	1218.09
	TS 2013	0.08	0.08	0.2	0.2	14	11	11	14	1174.05	1238.26	1180.44	1185.8
	TS 2014	0.09	0.09	0.08	0.08	14	12	11	15	1223.42	1282.26	1201.47	1270.24
	TS 2015	0.02	0.02	0.23	0.23	13	11	9	15	1213.66	1291.07	1213.66	1252.35
	TS 2016	0.05	0.05	0.13	0.13	14	12	11	15	1358.04	1392.58	1369.31	1392.58
	GR 2011	0.12	0.12	0.12	0.12	14	11	11	14	543.1	563.64	541.57	560.23
	GR 2012	0.13	0.13	0.03	0.03	14	12	11	15	528.3	554.09	525.86	554.09
	GR 2013	0.08	0.08	0.2	0.2	14	11	11	14	538.84	562.47	541.09	542.53
	GR 2014	0.09	0.09	0.08	0.08	14	12	11	15	545.36	564.02	537.65	561.04
	GR 2015	0.02	0.02	0.23	0.23	13	11	9	15	553.37	580.01	553.37	565.4
	GR 2016	0.05	0.05	0.13	0.13	14	12	11	15	514.22	525.6	517.14	525.6
VIT_201s0011g03070	DotY 1999	-	-	0.07	0.07			5	12	-	-	166	168.5
	DotY 2009	-	-	0.29	0.29			5	12	-	-	164	164
	DotY 2010	-	-	0.07	0.07			5	12	-	-	175	178
	DotY 2011	-	-	0.09	0.09			5	11	-	-	151	154
	DotY 2012	-	-	0.02	0.02			5	12	-	-	154	161
	DotY 2013	-	-	0.28	0.28			5	11	-	-	172	173
	DotY 2014	-	-	0.03	0.03			5	12	-	-	153	157
	DotY 2015	-	-	0.17	0.17			5	12	-	-	160	163
	DotY 2016	-	-	0.46	0.46			5	12	-	-	172	172.5
	Average 1999-2016	-	-	0.05	0.05			5	12	-	-	0.37	0.65
	Median 1999-2016	-	-	0.04	0.04			5	12	-	-	0.41	0.67
	TS 2011	-	-	0.14	0.14			5	11	-	-	1170.61	1209.22
	TS 2012	-	-	0.03	0.03			5	12	-	-	1122.89	1218.09
	TS 2013	-	-	0.53	0.53			5	11	-	-	1180.44	1180.44
	TS 2014	-	-	0.07	0.07			5	12	-	-	1172.17	1270.24
	TS 2015	-	-	0.34	0.34			5	12	-	-	1213.66	1252.35
	TS 2016	-	-	0.46	0.46			5	12	-	-	1369.31	1380.95
	GR 2011	-	-	0.14	0.14			5	11	-	-	543.1	556.81
	GR 2012	-	-	0.03	0.03			5	12	-	-	515.68	554.09
	GR 2013	-	-	0.53	0.53			5	11	-	-	541.09	541.09
	GR 2014	-	-	0.07	0.07			5	12	-	-	522.2	561.04
	GR 2015	-	-	0.34	0.34			5	12	-	-	553.37	565.4
	GR 2016	-	-	0.46	0.46			5	12	-	-	517.14	521.37
VIT_201s0011g03140	DotY 1999	0.08	0.08	0.28	0.28	12	14	12	14	166.5	169	167	168
	DotY 2009	0.06	0.06	0.56	0.56	13	14	12	15	162	164	164	164
	DotY 2010	0.27	0.27	0.27	0.27	13	14	12	15	175	178	175.5	178
	DotY 2011	0.16	0.16	0.21	0.21	13	13	12	14	151	154	151	153.5
	DotY 2012	0.15	0.15	0.07	0.07	13	14	12	15	157	161	157	161
	DotY 2013	0.07	0.07	0.46	0.46	13	13	12	14	172	178	173	172.5
	DotY 2014	0.1	0.1	0.18	0.18	13	14	12	15	156	157	156	157
	DotY 2015	0.01	0.01	0.43	0.43	13	14	12	15	160	164	161	162
	DotY 2016	0.06	0.06	0.35	0.35	13	14	12	15	172	173	172	173
	Average 1999-2016	0.08	0.08	0.18	0.18	13	14	12	15	0.46	0.65	0.5	0.64
	Median 1999 - 2016	0.08	0.08	0.12	0.12	13	14	12	15	0.41	0.67	0.48	0.66
	TS 2011	0.21	0.21	0.15	0.15	13	13	12	14	1170.61	1209.22	1170.61	1220.32
	TS 2012	0.21	0.21	0.03	0.03	13	14	12	15	1167.18	1218.09	1167.18	1218.09
	TS 2013	0.19	0.19	0.23	0.23	13	13	12	14	1180.44	1238.26	1180.44	1185.8

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	TS 2014	0.23	0.23	0.1	0.1	13	14	12	15	1245.36	1270.24	1201.47	1270.24
	TS 2015	0.06	0.06	0.15	0.15	13	13	11	15	1213.66	1291.07	1213.66	1252.35
	TS 2016	0.06	0.06	0.35	0.35	13	14	12	15	1369.31	1392.58	1369.31	1392.58
	GR 2011	0.21	0.21	0.15	0.15	13	13	12	14	543.1	556.81	543.1	560.23
	GR 2012	0.21	0.21	0.03	0.03	13	14	12	15	530.74	554.09	530.74	554.09
	GR 2013	0.19	0.19	0.23	0.23	13	13	12	14	541.09	562.47	541.09	542.53
	GR 2014	0.23	0.23	0.1	0.1	13	14	12	15	553.06	561.04	537.65	561.04
	GR 2015	0.06	0.06	0.15	0.15	13	13	11	15	553.37	580.01	553.37	565.4
	GR 2016	0.06	0.06	0.35	0.35	13	14	12	15	517.14	525.6	517.14	525.6
VIT_201s0011g04240	DotY 1999	0.01	0.01	0.34	0.34	11	15	12	14	166	169	167	168
	DotY 2009	0.01	0.01	0.46	0.46	12	15	12	15	162	164	164	164
	DotY 2010	0.08	0.08	0.19	0.19	12	15	12	15	175	178	175	178
	DotY 2011	0.06	0.06	0.12	0.12	12	14	12	14	151	154	151	153.5
	DotY 2012	0.05	0.05	0.12	0.12	12	15	12	15	156.5	161	157	161
	DotY 2013	0.03	0.03	0.64	0.64	12	14	12	14	171.5	178.5	173	172.5
	DotY 2014	0.04	0.04	0.2	0.2	12	15	12	15	155.5	157	156	157
	DotY 2015	0.01	0.01	0.6	0.6	11	15	11	15	160	164	162	162
	DotY 2016	0.02	0.02	0.31	0.31	12	15	12	15	171.5	173	172	173
	Average 1999-2016	0.02	0.02	0.17	0.17	12	15	12	15	0.42	0.65	0.5	0.64
	Median 1999 - 2016	0.02	0.02	0.1	0.1	12	15	12	15	0.41	0.67	0.48	0.66
	TS 2011	0.09	0.09	0.08	0.08	12	14	12	14	1170.61	1220.32	1170.61	1220.32
	TS 2012	0.08	0.08	0.06	0.06	12	15	12	15	1159.02	1218.09	1167.18	1218.09
	TS 2013	0.11	0.11	0.37	0.37	12	14	12	14	1174.05	1244.64	1180.44	1185.8
	TS 2014	0.07	0.07	0.12	0.12	12	15	12	15	1223.42	1270.24	1201.47	1270.24
	TS 2015	0.04	0.04	0.22	0.22	11	14	10	15	1213.66	1291.07	1213.66	1252.35
	TS 2016	0.02	0.02	0.31	0.31	12	15	12	15	1358.04	1392.58	1369.31	1392.58
	GR 2011	0.09	0.09	0.08	0.08	12	14	12	14	543.1	560.23	543.1	560.23
	GR 2012	0.08	0.08	0.06	0.06	12	15	12	15	528.3	554.09	530.74	554.09
	GR 2013	0.11	0.11	0.37	0.37	12	14	12	14	538.84	565.18	541.09	542.53
	GR 2014	0.07	0.07	0.12	0.12	12	15	12	15	545.36	561.04	537.65	561.04
	GR 2015	0.04	0.04	0.22	0.22	11	14	10	15	553.37	580.01	553.37	565.4
	GR 2016	0.02	0.02	0.31	0.31	12	15	12	15	514.22	525.6	517.14	525.6
VIT_201s0011g05260	DotY 1999	0.08	-	-	0.02	14	10	14	10	166.5	170.5	166.5	172.5
	DotY 2009	0.06	-	-	0	15	10	14	11	162	164	162	168
	DotY 2010	0.13	-	-	0.01	15	10	14	11	175	178	175	178
	DotY 2011	0.09	-	-	0	15	10	14	11	151	154.5	150.5	155
	DotY 2012	0.16	-	-	0.01	15	10	14	11	157	161	156.5	164
	DotY 2013	0.08	-	-	0.05	15	9	14	10	172	179	172	179
	DotY 2014	0.07	-	-	0.01	15	10	14	11	156	158.5	155	158
	DotY 2015	0.03	-	-	0.02	14	10	13	11	161	164	160	164
	DotY 2016	0.04	-	-	0.01	15	10	14	11	172	173.5	171.5	174
	Average 1999-2016	0.07	-	-	0	15	10	14	11	0.46	0.67	0.42	0.81
	Median 1999-2016	0.07	-	-	0	15	10	14	11	0.41	0.72	0.41	0.8
	TS 2011	0.07	-	-	0.01	15	10	14	11	1170.61	1241.31	1165.79	1251.21
	TS 2012	0.1	-	-	0.03	15	10	14	11	1167.18	1218.09	1159.02	1258.49
	TS 2013	0.04	-	-	0.23	15	9	14	10	1167.66	1267.05	1180.44	1267.05
	TS 2014	0.04	-	-	0.05	15	10	14	11	1201.47	1294.27	1201.47	1294.27
	TS 2015	0.01	-	-	0.09	14	10	13	11	1213.66	1291.07	1213.66	1291.07
	TS 2016	0.04	-	-	0.01	15	10	14	11	1369.31	1400.95	1358.04	1409.32
	GR 2011	0.07	-	-	0.01	15	10	14	11	543.1	566.53	542.34	569.43
	GR 2012	0.1	-	-	0.03	15	10	14	11	530.74	554.09	528.3	568.36
	GR 2013	0.04	-	-	0.23	15	9	14	10	536.59	576.24	541.09	576.24
	GR 2014	0.04	-	-	0.05	15	10	14	11	537.65	566.99	537.65	566.99
	GR 2015	0.01	-	-	0.09	14	10	13	11	553.37	580.01	553.37	580.01
	GR 2016	0.04	-	-	0.01	15	10	14	11	517.14	529.38	514.22	533.17
	DotY 1999	0.04	0.04	-	-	14	17			166.5	168	-	-
	DotY 2009	0.04	0.04	-	-	15	17			162	164	-	-
	DotY 2010	0.25	0.25	-	-	15	17			177	178	-	-

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
VIT_201s0011g06410	DotY 2011	0.07	0.07	-	-	15	16			151	154	-	-
	DotY 2012	0.09	0.09	-	-	15	17			157	161	-	-
	DotY 2013	0.06	0.06	-	-	15	16			172	175.5	-	-
	DotY 2014	0.01	0.01	-	-	15	17			157	158	-	-
	DotY 2015	0.02	0.02	-	-	14	17			161	164	-	-
	DotY 2016	0.05	0.05	-	-	15	17			172	173	-	-
	Average 1999-2016	0.04	0.04	-	-	15	17			0.46	0.64	-	-
	Median 1999-2016	0.05	0.05	-	-	15	17			0.41	0.61	-	-
	TS 2011	0.04	0.04	-	-	15	16			1170.61	1231.41	-	-
	TS 2012	0.04	0.04	-	-	15	17			1167.18	1218.09	-	-
	TS 2013	0.02	0.02	-	-	15	16			1167.66	1220.27	-	-
	TS 2014	0.02	0.02	-	-	15	16			1201.47	1294.27	-	-
	TS 2015	0.02	0.02	-	-	14	15			1213.66	1291.07	-	-
	TS 2016	0.05	0.05	-	-	15	17			1369.31	1392.58	-	-
	GR 2011	0.04	0.04	-	-	15	16			543.1	563.64	-	-
	GR 2012	0.04	0.04	-	-	15	17			530.74	554.09	-	-
	GR 2013	0.02	0.02	-	-	15	16			536.59	555.9	-	-
	GR 2014	0.02	0.02	-	-	15	16			537.65	566.99	-	-
	GR 2015	0.02	0.02	-	-	14	15			553.37	580.01	-	-
	GR 2016	0.05	0.05	-	-	15	17			517.14	525.6	-	-
VIT_201s0026g00150	DotY 1999	0.37	0.37	0.13	0.13	17	17	19	15	167	168	168	167
	DotY 2009	0.32	0.32	0.37	0.37	18	17	19	16	163.5	164	164	163
	DotY 2010	0.53	0.53	0.04	0.04	18	17	19	16	177	178	178	175
	DotY 2011	0.51	0.51	0.19	0.19	18	16	19	15	152	153.5	153	151
	DotY 2012	0.46	0.46	0.23	0.23	18	17	19	16	160.5	161	161	160
	DotY 2013	0.21	0.21	0.24	0.24	18	16	18	16	172	175.5	174	172
	DotY 2014	0.36	0.36	0.01	0.01	18	17	19	16	157	157	158	155.5
	DotY 2015	0.38	0.38	0.12	0.12	17	17	19	15	162	164	164	162
	DotY 2016	0.25	0.25	0.02	0.02	18	17	19	16	172	173	173	172
	Average 1999-2016	0.35	0.35	0.13	0.13	18	17	19	16	0.5	0.6	0.6	0.48
	Median 1999 - 2016	0.45	0.45	0.2	0.2	18	17	19	16	0.53	0.6	0.6	0.48
	TS 2011	0.44	0.44	0.13	0.13	18	16	19	15	1170.61	1220.32	1209.22	1170.61
	TS 2012	0.31	0.31	0.12	0.12	18	17	19	16	1201.17	1218.09	1218.09	1187.38
	TS 2013	0.1	0.1	0.09	0.09	18	16	18	16	1180.44	1220.27	1202.28	1174.05
	TS 2014	0.49	0.49	0.01	0.01	18	16	18	16	1257.8	1270.24	1282.26	1201.47
	TS 2015	0.31	0.31	0.09	0.09	17	15	17	15	1252.35	1291.07	1291.07	1252.35
	TS 2016	0.25	0.25	0.02	0.02	18	17	19	16	1369.31	1392.58	1392.58	1369.31
	GR 2011	0.44	0.44	0.13	0.13	18	16	19	15	543.1	560.23	556.81	543.5
	GR 2012	0.31	0.31	0.12	0.12	18	17	19	16	547.18	554.09	554.09	540.47
	GR 2013	0.1	0.1	0.09	0.09	18	16	18	16	541.09	555.9	549.33	538.84
	GR 2014	0.49	0.49	0.01	0.01	18	16	18	16	557.05	561.04	564.02	537.65
	GR 2015	0.31	0.31	0.09	0.09	17	15	17	15	565.4	580.01	580.01	565.4
	GR 2016	0.25	0.25	0.02	0.02	18	17	19	16	517.14	525.6	525.6	517.14
VIT_201s0026g02200	DotY 1999	-	0.48	-	0.7	15	12	14	13	167	168	167.5	168
	DotY 2009	-	0.71	-	0.93	16	12	14	14	164	164	164	164
	DotY 2010	-	0.59	-	0.18	16	12	14	14	175	177	177	175
	DotY 2012	-	0.52	-	0.75	16	12	14	14	160.5	160.5	160.5	160.5
	DotY 2013	-	0.73	-	0.73	16	11	13	14	172	174	172	172.5
	DotY 2014	-	0.96	-	0.15	16	12	14	14	157	157	157	156
	DotY 2015	-	0.85	-	0.77	15	12	14	13	162	162	162	162
	DotY 2016	-	0.49	-	0.16	16	12	14	14	172	172.5	173	172
	Average 1999-2016	-	0.55	-	0.58	16	12	14	14	0.51	0.57	0.56	0.51
	Median 1999-2016	-	0.66	-	0.66	16	12	14	14	0.55	0.58	0.6	0.52
	TS 2011	-	0.37	-	0.63	16	11	14	13	1170.61	1209.22	1189.92	1170.61
	TS 2012	-	0.39	-	0.57	16	12	14	14	1187.38	1212.84	1212.84	1207.58
	TS 2013	-	0.47	-	0.88	16	11	13	14	1180.44	1202.28	1180.44	1180.44
	TS 2014	-	0.63	-	0.06	16	12	14	14	1245.36	1270.24	1270.24	1223.42
	TS 2015	-	0.47	-	0.61	14	12	13	13	1252.35	1252.35	1252.35	1252.35

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	TS 2016	-	0.49	-	0.16	16	12	14	14	1369.31	1380.95	1392.58	1369.31
	GR 2011	-	0.37	-	0.63	16	11	14	13	543.1	556.81	549.96	543.5
	GR 2012	-	0.39	-	0.57	16	12	14	14	540.47	552.14	552.14	550.5
	GR 2013	-	0.47	-	0.88	16	11	13	14	541.09	549.33	541.09	541.09
	GR 2014	-	0.63	-	0.06	16	12	14	14	553.06	561.04	561.04	545.36
	GR 2015	-	0.47	-	0.61	14	12	13	13	565.4	565.4	565.4	565.4
	GR 2016	-	0.49	-	0.16	16	12	14	14	517.14	521.37	525.6	517.14
VIT_201s0026g02580	DotY 1999	0.62	0.62	0.25	0.25	16	16	18	14	168	168	168.5	167.5
	DotY 2009	0.63	0.63	0.56	0.56	17	16	18	15	164	164	164	164
	DotY 2010	0.56	0.56	0.08	0.08	17	16	18	15	175	177	178	175
	DotY 2011	0.69	0.69	0.33	0.33	17	15	18	14	153	153	153	151
	DotY 2012	0.48	0.48	0.44	0.44	17	16	18	15	161	160.5	161	160
	DotY 2013	0.37	0.37	0.65	0.65	17	15	17	15	172	174	174	172
	DotY 2014	0.71	0.71	0.02	0.02	17	16	18	15	157	157	158	156
	DotY 2015	0.84	0.84	0.09	0.09	16	16	18	14	163	163	164	162
	DotY 2016	0.6	0.6	0.07	0.07	17	16	18	15	173	173	173	172
	Average 1999-2016	0.61	0.61	0.2	0.2	17	16	18	15	0.58	0.57	0.6	0.48
	Median 1999 - 2016	0.71	0.71	0.22	0.22	17	16	18	15	0.6	0.6	0.61	0.49
	TS 2011	0.69	0.69	0.33	0.33	17	15	18	14	1209.22	1209.22	1209.22	1170.61
	TS 2012	0.48	0.48	0.44	0.44	17	16	18	15	1218.09	1212.84	1218.09	1207.58
	TS 2013	0.37	0.37	0.65	0.65	17	15	17	15	1180.44	1202.28	1202.28	1180.44
	TS 2014	0.97	0.97	0.01	0.01	17	15	17	15	1270.24	1270.24	1294.27	1245.36
	TS 2015	0.69	0.69	0.13	0.13	16	14	16	14	1271.71	1252.35	1291.07	1252.35
	TS 2016	0.6	0.6	0.07	0.07	17	16	18	15	1392.58	1392.58	1392.58	1369.31
	GR 2011	0.69	0.69	0.33	0.33	17	15	18	14	556.81	556.81	556.81	543.5
	GR 2012	0.48	0.48	0.44	0.44	17	16	18	15	554.09	552.14	554.09	550.5
	GR 2013	0.37	0.37	0.65	0.65	17	15	17	15	541.09	549.33	549.33	541.09
	GR 2014	0.97	0.97	0.01	0.01	17	15	17	15	561.04	561.04	566.99	553.06
	GR 2015	0.69	0.69	0.13	0.13	16	14	16	14	572.7	565.4	580.01	565.4
	GR 2016	0.6	0.6	0.07	0.07	17	16	18	15	525.6	525.6	525.6	517.14
VIT_201s0010g00740	DotY 1999	0.34	0.34	0.74	0.74	12	14	12	14	167	168	168	167.5
	DotY 2009	0.45	0.45	0.7	0.7	13	14	12	15	163	164	164	164
	DotY 2010	0.72	0.72	0.75	0.75	13	14	12	15	175	176.5	176.5	175
	DotY 2011	0.61	0.61	0.66	0.66	13	13	12	14	151	153	153	151
	DotY 2012	0.61	0.61	0.7	0.7	13	14	12	15	161	160.5	160.5	161
	DotY 2013	0.33	0.33	0.88	0.88	13	13	12	14	172	177	173	172.5
	DotY 2014	0.27	0.27	0.96	0.96	13	14	12	15	157	157	157	157
	DotY 2015	0.22	0.22	0.57	0.57	12	14	12	14	162	164	163	162
	DotY 2016	0.36	0.36	0.51	0.51	13	14	12	15	172	173	173	172
	Average 1999-2016	0.37	0.37	0.92	0.92	13	14	12	15	0.48	0.62	0.56	0.48
	Median 1999 - 2016	0.44	0.44	0.77	0.77	13	14	12	15	0.49	0.63	0.6	0.49
	TS 2011	0.52	0.52	0.59	0.59	13	13	12	14	1170.61	1209.22	1189.92	1170.61
	TS 2012	0.42	0.42	0.54	0.54	13	14	12	15	1207.58	1212.84	1207.58	1218.09
	TS 2013	0.15	0.15	0.57	0.57	13	13	12	14	1180.44	1238.26	1180.44	1185.8
	TS 2014	0.13	0.13	0.94	0.94	13	14	12	15	1245.36	1270.24	1270.24	1245.36
	TS 2015	0.14	0.14	0.93	0.93	12	13	11	14	1252.35	1291.07	1252.35	1252.35
	TS 2016	0.36	0.36	0.51	0.51	13	14	12	15	1369.31	1392.58	1392.58	1369.31
	GR 2011	0.52	0.52	0.59	0.59	13	13	12	14	543.1	556.81	549.96	543.5
	GR 2012	0.42	0.42	0.54	0.54	13	14	12	15	550.2	552.14	550.2	554.09
	GR 2013	0.15	0.15	0.57	0.57	13	13	12	14	541.09	562.47	541.09	542.53
	GR 2014	0.13	0.13	0.94	0.94	13	14	12	15	553.06	561.04	561.04	553.06
	GR 2015	0.14	0.14	0.93	0.93	12	13	11	14	565.4	580.01	565.4	565.4
	GR 2016	0.36	0.36	0.51	0.51	13	14	12	15	517.14	525.6	525.6	517.14
	DotY 1999	0.08	0.08	0.34	0.34	14	18	18	14	166.5	168	168	167.5
	DotY 2009	0.07	0.07	0.29	0.29	14	19	19	14	162	164	164	163
	DotY 2010	0.17	0.17	0.14	0.14	14	19	19	14	176	178	178	175.5
	DotY 2011	0.08	0.08	0.17	0.17	14	19	19	14	151	154	154	151
	DotY 2012	0.09	0.09	0.23	0.23	14	19	19	14	156.5	161	161	160

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
VIT_201s0010g02270	DotY 2013	0.01	0.01	0.17	0.17	14	18	18	14	171.5	174	174	172
	DotY 2014	0.03	0.03	0.12	0.12	14	19	19	14	155.5	158	157	155.5
	DotY 2015	0.01	0.01	0.21	0.21	13	19	19	13	160	164	164	162
	DotY 2016	0.14	0.14	0.09	0.09	14	19	19	14	172	173	173	172
	Average 1999-2016	0.04	0.04	0.15	0.15	14	19	19	14	0.42	0.64	0.59	0.48
	Median 1999 - 2016	0.05	0.05	0.14	0.14	14	19	19	14	0.41	0.61	0.61	0.48
	TS 2011	0.11	0.11	0.21	0.21	14	19	19	14	1170.61	1231.41	1209.22	1170.61
	TS 2012	0.13	0.13	0.33	0.33	14	19	19	14	1159.02	1218.09	1218.09	1207.58
	TS 2013	0.03	0.03	0.31	0.31	14	18	18	14	1174.05	1202.28	1196.72	1180.44
	TS 2014	0.15	0.15	0.13	0.13	14	18	18	14	1223.42	1282.26	1270.24	1223.42
	TS 2015	0.04	0.04	0.46	0.46	13	17	17	13	1213.66	1291.07	1291.07	1252.35
	TS 2016	0.14	0.14	0.09	0.09	14	19	19	14	1369.31	1392.58	1392.58	1369.31
	GR 2011	0.11	0.11	0.21	0.21	14	19	19	14	543.1	563.64	556.81	543.5
	GR 2012	0.13	0.13	0.33	0.33	14	19	19	14	528.3	554.09	554.09	550.5
	GR 2013	0.03	0.03	0.31	0.31	14	18	18	14	538.84	549.33	546.64	541.09
	GR 2014	0.15	0.15	0.13	0.13	14	18	18	14	545.36	564.02	561.04	545.36
	GR 2015	0.04	0.04	0.46	0.46	13	17	17	13	553.37	580.01	580.01	565.4
	GR 2016	0.14	0.14	0.09	0.09	14	19	19	14	517.14	525.6	525.6	517.14
VIT_201s0010g03890	DotY 1999	-	0.02	-	0.25	14	13	20	7	166.5	169	168	166
	DotY 2009	-	0.07	-	0.9	15	13	20	8	162	164	164	163
	DotY 2010	-	0.28	-	0.52	15	13	20	8	177	178	177.5	176
	DotY 2012	-	0.06	-	0.51	15	13	20	8	157	161	161	157
	DotY 2013	-	0.02	-	0.63	15	12	19	8	172	177.5	174	172
	DotY 2014	-	0.02	-	0.18	15	13	20	8	157	158	157	156
	DotY 2015	-	0.01	-	0.33	14	13	20	7	161	164	164	162
	DotY 2016	-	0.13	-	0.51	15	13	20	8	172	173	173	172
	Average 1999-2016	-	0.04	-	0.46	15	13	20	8	0.46	0.64	0.59	0.43
	Median 1999-2016	-	0.05	-	0.49	15	13	20	8	0.41	0.66	0.6	0.45
	TS 2011	-	0.13	-	0.46	15	13	20	8	1170.61	1231.41	1209.22	1165.79
	TS 2012	-	0.1	-	0.56	15	13	20	8	1167.18	1218.09	1212.84	1165.24
	TS 2013	-	0.06	-	0.81	15	12	19	8	1180.44	1220.27	1191.16	1179.41
	TS 2014	-	0.06	-	0.18	15	12	19	8	1245.36	1294.27	1270.24	1223.42
	TS 2015	-	0.04	-	0.61	14	11	18	7	1233.01	1291.07	1252.35	1252.35
	TS 2016	-	0.13	-	0.51	15	13	20	8	1369.31	1392.58	1392.58	1369.31
	GR 2011	-	0.13	-	0.46	15	13	20	8	543.1	563.64	556.81	542.34
	GR 2012	-	0.1	-	0.56	15	13	20	8	530.74	554.09	552.14	532.94
	GR 2013	-	0.06	-	0.81	15	12	19	8	541.09	555.9	543.96	540.27
	GR 2014	-	0.06	-	0.18	15	12	19	8	553.06	566.99	561.04	545.36
	GR 2015	-	0.04	-	0.61	14	11	18	7	559.38	580.01	565.4	565.4
	GR 2016	-	0.13	-	0.51	15	13	20	8	517.14	525.6	525.6	517.14
VIT_204s0023g02820	DotY 1999	0.02	-	0.36	-	22	12	16	18	167	169.5	168	167.5
	DotY 2009	0.01	-	0.74	-	22	13	16	19	163	164	164	164
	DotY 2010	0.37	-	0.5	-	22	13	16	19	175.5	178	178	176
	DotY 2011	0.06	-	0.43	-	22	12	16	18	151	154	153.5	152
	DotY 2012	0.04	-	0.6	-	22	13	16	19	160	161	161	160
	DotY 2013	0.01	-	0.65	-	22	12	15	19	172	178.5	172	174
	DotY 2014	0.03	-	0.53	-	22	13	16	19	156.5	158	157.5	157
	DotY 2015	0.06	-	0.64	-	21	13	15	19	162	164	164	162
	DotY 2016	0.1	-	0.4	-	22	13	16	19	172	173	172	173
	Average 1999-2016	0.01	-	0.7	-	22	13	16	19	0.5	0.66	0.59	0.54
	Median 1999-2016	0.01	-	0.6	-	22	13	16	19	0.52	0.7	0.61	0.59
	TS 2011	0.11	-	0.56	-	22	12	16	18	1170.61	1231.41	1209.22	1189.92
	TS 2012	0.09	-	0.86	-	22	13	16	19	1207.58	1218.09	1212.84	1207.58
	TS 2013	0.06	-	0.34	-	22	12	15	19	1180.44	1226.65	1180.44	1202.28
	TS 2014	0.02	-	0.9	-	21	13	15	19	1245.36	1294.27	1270.24	1270.24
	TS 2015	0.19	-	0.97	-	20	12	14	18	1252.35	1291.07	1271.71	1252.35
	TS 2016	0.1	-	0.4	-	22	13	16	19	1369.31	1392.58	1369.31	1392.58
	GR 2011	0.11	-	0.56	-	22	12	16	18	543.1	563.64	556.81	549.96

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	GR 2012	0.09	-	0.86	-	22	13	16	19	550.2	554.09	552.14	550.5
	GR 2013	0.06	-	0.34	-	22	12	15	19	541.09	558.61	541.09	549.33
	GR 2014	0.02	-	0.9	-	21	13	15	19	553.06	566.99	561.04	561.04
	GR 2015	0.19	-	0.97	-	20	12	14	18	565.4	580.01	572.7	565.4
	GR 2016	0.1	-	0.4	-	22	13	16	19	517.14	525.6	517.14	525.6
VIT_204s0044g00850	DotY 1999	0.03	-	0.76	-	25	9	18	16	167	170	168	167.5
	DotY 2009	0.01	-	0.88	-	25	10	19	16	164	166	164	164
	DotY 2010	0.44	-	0.49	-	25	10	19	16	176	178	178	175.5
	DotY 2011	0.03	-	0.64	-	25	9	19	15	151	155	153	151
	DotY 2012	0.05	-	0.78	-	25	10	19	16	160	162.5	161	160
	DotY 2013	0.01	-	0.31	-	25	9	18	16	172	179	172.5	174
	DotY 2014	0.02	-	0.58	-	25	10	19	16	157	158.5	157	157
	DotY 2015	0.07	-	0.44	-	24	10	18	16	162	164	164	162
	DotY 2016	0.18	-	0.53	-	25	10	19	16	172	173	172	173
	Average 1999-2016	0.01	-	0.74	-	25	10	19	16	0.51	0.68	0.59	0.54
	Median 1999-2016	0.01	-	0.7	-	25	10	19	16	0.55	0.73	0.61	0.58
	TS 2011	0.03	-	0.74	-	25	9	19	15	1170.61	1251.21	1209.22	1170.61
	TS 2012	0.03	-	0.99	-	25	10	19	16	1207.58	1238.29	1218.09	1207.58
	TS 2013	0	-	0.15	-	25	9	18	16	1180.44	1267.05	1180.44	1202.28
	TS 2014	0	-	0.74	-	24	10	18	16	1223.42	1294.27	1270.24	1257.8
	TS 2015	0.05	-	0.72	-	23	9	17	15	1252.35	1291.07	1291.07	1252.35
	TS 2016	0.18	-	0.53	-	25	10	19	16	1369.31	1392.58	1369.31	1392.58
	GR 2011	0.03	-	0.74	-	25	9	19	15	543.1	569.43	556.81	543.5
	GR 2012	0.03	-	0.99	-	25	10	19	16	550.2	561.23	554.09	550.5
	GR 2013	0	-	0.15	-	25	9	18	16	541.09	576.24	541.09	549.33
	GR 2014	0	-	0.74	-	24	10	18	16	545.36	566.99	561.04	557.05
	GR 2015	0.05	-	0.72	-	23	9	17	15	565.4	580.01	580.01	565.4
	GR 2016	0.18	-	0.53	-	25	10	19	16	517.14	525.6	517.14	525.6
VIT_205s0102g01160	DotY 1999	0.75	0.75	0.32	0.32	7	27	20	14	167	168	168	168
	DotY 2009	0.56	0.56	0.38	0.38	7	28	20	15	164	164	164	164
	DotY 2010	0.92	0.92	0.15	0.15	7	28	20	15	177	177	177.5	176
	DotY 2011	0.62	0.62	0.44	0.44	7	27	19	15	151	153	153	153
	DotY 2012	0.17	0.17	0.18	0.18	7	28	20	15	160	161	161	160
	DotY 2013	0.75	0.75	0.19	0.19	7	27	19	15	173	173	174	172
	DotY 2014	0.73	0.73	0.26	0.26	7	28	20	15	157	157	157	156
	DotY 2015	0.22	0.22	0.72	0.72	7	27	19	15	162	164	164	162
	DotY 2016	0.27	0.27	0.39	0.39	7	28	20	15	171	173	173	172
	Average 1999-2016	0.52	0.52	0.24	0.24	7	28	20	15	0.54	0.59	0.6	0.53
	Median 1999 - 2016	0.54	0.54	0.26	0.26	7	28	20	15	0.55	0.6	0.6	0.55
	TS 2011	0.67	0.67	0.54	0.54	7	27	19	15	1170.61	1209.22	1209.22	1209.22
	TS 2012	0.22	0.22	0.27	0.27	7	28	20	15	1207.58	1218.09	1212.84	1207.58
	TS 2013	0.88	0.88	0.35	0.35	7	27	19	15	1191.16	1191.16	1202.28	1180.44
	TS 2014	1	1	0.54	0.54	7	27	19	15	1270.24	1270.24	1270.24	1245.36
	TS 2015	0.19	0.19	0.91	0.91	6	26	17	15	1216.64	1252.35	1252.35	1252.35
	TS 2016	0.27	0.27	0.39	0.39	7	28	20	15	1346.77	1392.58	1392.58	1369.31
	GR 2011	0.67	0.67	0.54	0.54	7	27	19	15	543.1	556.81	556.81	556.81
	GR 2012	0.22	0.22	0.27	0.27	7	28	20	15	550.2	554.09	552.14	550.5
	GR 2013	0.88	0.88	0.35	0.35	7	27	19	15	543.96	543.96	549.33	541.09
	GR 2014	1	1	0.54	0.54	7	27	19	15	561.04	561.04	561.04	553.06
	GR 2015	0.19	0.19	0.91	0.91	6	26	17	15	554.43	565.4	565.4	565.4
	GR 2016	0.27	0.27	0.39	0.39	7	28	20	15	511.31	525.6	525.6	517.14
	DotY 1999	-	-	-	0.47			14	20	-	-	168	168
	DotY 2009	-	-	-	0.4			15	20	-	-	164	164
	DotY 2010	-	-	-	0.1			15	20	-	-	175	178
	DotY 2011	-	-	-	0.59			14	20	-	-	151	153
	DotY 2012	-	-	-	0.31			15	20	-	-	160	161
	DotY 2013	-	-	-	0.56			15	19	-	-	174	172
	DotY 2014	-	-	-	0.16			15	20	-	-	157	157

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
VIT_206s0004g03590	DotY 2015	-	-	-	0.6			15	19	-	-	162	164
	DotY 2016	-	-	-	0.79			15	20	-	-	173	172.5
	Average 1999-2016	-	-	-	0.46			15	20	-	-	0.59	0.56
	Median 1999-2016	-	-	-	0.5			15	20	-	-	0.59	0.6
	TS 2011	-	-	-	0.65			14	20	-	-	1170.61	1209.22
	TS 2012	-	-	-	0.46			15	20	-	-	1207.58	1212.84
	TS 2013	-	-	-	0.34			15	19	-	-	1202.28	1180.44
	TS 2014	-	-	-	0.2			14	20	-	-	1245.36	1270.24
	TS 2015	-	-	-	0.98			15	17	-	-	1252.35	1252.35
	TS 2016	-	-	-	0.79			15	20	-	-	1392.58	1380.95
	GR 2011	-	-	-	0.65			14	20	-	-	543.1	556.81
	GR 2012	-	-	-	0.46			15	20	-	-	550.2	552.14
	GR 2013	-	-	-	0.34			15	19	-	-	549.33	541.09
	GR 2014	-	-	-	0.2			14	20	-	-	553.06	561.04
	GR 2015	-	-	-	0.98			15	17	-	-	565.4	565.4
	GR 2016	-	-	-	0.79			15	20	-	-	525.6	521.37
VIT_206s0009g02480	DotY 1999	0.96	0.96	0.73	0.73	16	18	27	7	167.5	168	168	168
	DotY 2009	0.7	0.7	0.78	0.78	16	19	27	8	164	164	164	163
	DotY 2010	0.95	0.95	0.89	0.89	16	19	27	8	175.5	178	177	177.5
	DotY 2011	0.68	0.68	0.35	0.35	16	18	26	8	151	153.5	153	154
	DotY 2012	0.49	0.49	0.62	0.62	16	19	27	8	160	161	160	161
	DotY 2013	0.43	0.43	0.7	0.7	15	19	26	8	172	174	173	174.5
	DotY 2014	0.46	0.46	0.49	0.49	16	19	27	8	157	157	157	157
	DotY 2015	0.53	0.53	0.81	0.81	15	19	26	8	162	164	163	163
	DotY 2016	0.73	0.73	0.49	0.49	16	19	27	8	173	172	173	172
	Average 1999-2016	0.73	0.73	0.81	0.81	16	19	27	8	0.54	0.6	0.58	0.61
	Median 1999 - 2016	0.69	0.69	0.68	0.68	16	19	27	8	0.57	0.61	0.6	0.63
	TS 2011	0.8	0.8	0.32	0.32	16	18	26	8	1170.61	1209.22	1189.92	1230.22
	TS 2012	0.63	0.63	0.52	0.52	16	19	27	8	1207.58	1218.09	1207.58	1218.09
	TS 2013	0.72	0.72	0.56	0.56	15	19	26	8	1180.44	1191.16	1191.16	1209.35
	TS 2014	0.92	0.92	0.9	0.9	16	18	26	8	1270.24	1257.8	1257.8	1270.24
	TS 2015	0.73	0.73	0.95	0.95	14	18	24	8	1252.35	1252.35	1252.35	1271.71
	TS 2016	0.73	0.73	0.49	0.49	16	19	27	8	1392.58	1369.31	1392.58	1369.31
	GR 2011	0.8	0.8	0.32	0.32	16	18	26	8	543.1	556.81	549.96	563.12
	GR 2012	0.63	0.63	0.52	0.52	16	19	27	8	550.2	554.09	550.2	554.09
	GR 2013	0.72	0.72	0.56	0.56	15	19	26	8	541.09	543.96	543.96	551.78
	GR 2014	0.92	0.92	0.9	0.9	16	18	26	8	561.04	557.05	557.05	561.04
	GR 2015	0.73	0.73	0.95	0.95	14	18	24	8	565.4	565.4	565.4	572.7
	GR 2016	0.73	0.73	0.49	0.49	16	19	27	8	525.6	517.14	525.6	517.14
VIT_207s0005g02260	DotY 1999	0.64	0.64	nan	nan	20	14	34	0	168	168	168	nan
	DotY 2009	0.88	0.88	nan	nan	21	14	35	0	164	164	164	nan
	DotY 2010	0.59	0.59	nan	nan	21	14	35	0	177	175.5	177	nan
	DotY 2011	0.9	0.9	nan	nan	20	14	34	0	153.5	153	153	nan
	DotY 2012	0.23	0.23	nan	nan	21	14	35	0	161	160	161	nan
	DotY 2013	0.08	0.08	nan	nan	21	13	34	0	174	172	173	nan
	DotY 2014	0.45	0.45	nan	nan	21	14	35	0	157	156.5	157	nan
	DotY 2015	0.93	0.93	nan	nan	21	13	34	0	162	164	163	nan
	DotY 2016	0.19	0.19	nan	nan	21	14	35	0	173	172	173	nan
	Average 1999-2016	0.34	0.34	nan	nan	21	14	35	0	0.6	0.54	0.58	nan
	Median 1999 - 2016	0.46	0.46	nan	nan	21	14	35	0	0.6	0.55	0.6	nan
	TS 2011	0.93	0.93	nan	nan	20	14	34	0	1189.92	1209.22	1209.22	nan
	TS 2012	0.33	0.33	nan	nan	21	14	35	0	1218.09	1207.58	1207.58	nan
	TS 2013	0.17	0.17	nan	nan	21	13	34	0	1202.28	1180.44	1191.16	nan
	TS 2014	0.86	0.86	nan	nan	20	14	34	0	1270.24	1257.8	1270.24	nan
	TS 2015	0.92	0.92	nan	nan	20	12	32	0	1252.35	1271.71	1252.35	nan
	TS 2016	0.19	0.19	nan	nan	21	14	35	0	1392.58	1369.31	1392.58	nan
	GR 2011	0.93	0.93	nan	nan	20	14	34	0	549.96	556.81	556.81	nan
	GR 2012	0.33	0.33	nan	nan	21	14	35	0	554.09	550.2	550.2	nan

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	GR 2013	0.17	0.17	nan	nan	21	13	34	0	549.33	541.09	543.96	nan
	GR 2014	0.86	0.86	nan	nan	20	14	34	0	561.04	557.05	561.04	nan
	GR 2015	0.92	0.92	nan	nan	20	12	32	0	565.4	572.7	565.4	nan
	GR 2016	0.19	0.19	nan	nan	21	14	35	0	525.6	517.14	525.6	nan
VIT_208s0056g01230	DotY 1999	-	-	0.34	-			17	17	-	-	168	167
	DotY 2009	-	-	0.55	-			17	18	-	-	164	163.5
	DotY 2010	-	-	0.37	-			17	18	-	-	177	176.5
	DotY 2011	-	-	0.41	-			16	18	-	-	154	152
	DotY 2012	-	-	0.52	-			17	18	-	-	161	160
	DotY 2013	-	-	0.42	-			16	18	-	-	174	172
	DotY 2014	-	-	0.28	-			17	18	-	-	157	156.5
	DotY 2015	-	-	0.29	-			17	17	-	-	164	162
	Average 1999-2016	-	-	0.28	-			17	18	-	-	0.6	0.5
	Median 1999-2016	-	-	0.22	-			17	18	-	-	0.61	0.51
	TS 2011	-	-	0.52	-			16	18	-	-	1220.32	1189.92
	TS 2012	-	-	0.66	-			17	18	-	-	1218.09	1207.58
	TS 2013	-	-	0.72	-			16	18	-	-	1202.28	1180.44
	TS 2014	-	-	0.63	-			16	18	-	-	1270.24	1257.8
	TS 2015	-	-	0.44	-			16	16	-	-	1252.35	1252.35
	TS 2016	-	-	0.07	-			17	18	-	-	1392.58	1358.04
	GR 2011	-	-	0.52	-			16	18	-	-	560.23	549.96
	GR 2012	-	-	0.66	-			17	18	-	-	554.09	550.5
	GR 2013	-	-	0.72	-			16	18	-	-	549.33	541.09
	GR 2014	-	-	0.63	-			16	18	-	-	561.04	557.05
	GR 2015	-	-	0.44	-			16	16	-	-	565.4	565.4
	GR 2016	-	-	0.07	-			17	18	-	-	525.6	514.22
VIT_208s0007g04200	DotY 1999	-	-	-	0.77			20	14	-	-	168	168.5
	DotY 2009	-	-	-	0.46			21	14	-	-	164	164
	DotY 2010	-	-	-	0.7			21	14	-	-	177	176.5
	DotY 2011	-	-	-	0.94			20	14	-	-	153	152
	DotY 2012	-	-	-	1			21	14	-	-	161	160.5
	DotY 2013	-	-	-	0.43			21	13	-	-	173	174
	DotY 2014	-	-	-	0.61			21	14	-	-	157	157
	DotY 2015	-	-	-	0.83			21	13	-	-	162	164
	DotY 2016	-	-	-	0.84			21	14	-	-	173	172.5
	Average 1999-2016	-	-	-	0.91			21	14	-	-	0.54	0.59
	Median 1999-2016	-	-	-	0.74			21	14	-	-	0.56	0.6
	TS 2011	-	-	-	0.75			20	14	-	-	1209.22	1170.61
	TS 2012	-	-	-	0.71			21	14	-	-	1218.09	1207.58
	TS 2013	-	-	-	0.82			21	13	-	-	1191.16	1191.16
	TS 2014	-	-	-	0.99			20	14	-	-	1257.8	1270.24
	TS 2015	-	-	-	0.85			20	12	-	-	1252.35	1252.35
	TS 2016	-	-	-	0.84			21	14	-	-	1392.58	1380.95
	GR 2011	-	-	-	0.75			20	14	-	-	556.81	543.5
	GR 2012	-	-	-	0.71			21	14	-	-	554.09	550.5
	GR 2013	-	-	-	0.82			21	13	-	-	543.96	543.96
	GR 2014	-	-	-	0.99			20	14	-	-	557.05	561.04
	GR 2015	-	-	-	0.85			20	12	-	-	565.4	565.4
	GR 2016	-	-	-	0.84			21	14	-	-	525.6	521.37
VIT_210s0116g00750	DotY 1999	-	-	0.15	0.15			19	15	-	-	168	167
	DotY 2009	-	-	0.23	0.23			19	16	-	-	164	163.5
	DotY 2010	-	-	0.2	0.2			19	16	-	-	177	176
	DotY 2011	-	-	0.16	0.16			18	16	-	-	154	152
	DotY 2012	-	-	0.42	0.42			19	16	-	-	161	160.5
	DotY 2013	-	-	0.22	0.22			18	16	-	-	174	172
	DotY 2014	-	-	0.93	0.93			19	16	-	-	157	157
	DotY 2015	-	-	0.69	0.69			19	15	-	-	164	162
	DotY 2016	-	-	0.14	0.14			19	16	-	-	173	172

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	Average 1999-2016	-	-	0.25	0.25			19	16	-	-	0.6	0.53
	Median 1999-2016	-	-	0.31	0.31			19	16	-	-	0.61	0.56
	TS 2011	-	-	0.23	0.23			18	16	-	-	1220.32	1189.92
	TS 2012	-	-	0.61	0.61			19	16	-	-	1207.58	1212.84
	TS 2013	-	-	0.43	0.43			18	16	-	-	1202.28	1180.44
	TS 2014	-	-	0.62	0.62			18	16	-	-	1257.8	1270.24
	TS 2015	-	-	0.73	0.73			19	13	-	-	1252.35	1252.35
	TS 2016	-	-	0.14	0.14			19	16	-	-	1392.58	1369.31
	GR 2011	-	-	0.23	0.23			18	16	-	-	560.23	549.96
	GR 2012	-	-	0.61	0.61			19	16	-	-	550.2	552.14
	GR 2013	-	-	0.43	0.43			18	16	-	-	549.33	541.09
	GR 2014	-	-	0.62	0.62			18	16	-	-	557.05	561.04
	GR 2015	-	-	0.73	0.73			19	13	-	-	565.4	565.4
	GR 2016	-	-	0.14	0.14			19	16	-	-	525.6	517.14
VIT_211s0052g01800	DotY 1999	0.61	0.61	0.29	0.29	14	19	18	15	168	168	167.5	168
	DotY 2009	0.59	0.59	0.49	0.49	15	19	19	15	164	164	164	164
	DotY 2010	0.78	0.78	1	1	15	19	19	15	177	177	177	177
	DotY 2011	0.61	0.61	0.64	0.64	14	19	18	15	152.5	153	152	153
	DotY 2012	0.49	0.49	0.6	0.6	15	19	19	15	161	160	160	161
	DotY 2013	0.88	0.88	0.7	0.7	14	19	18	15	175.5	173	173	174
	DotY 2014	0.74	0.74	0.97	0.97	15	19	19	15	157	157	157	157
	DotY 2015	0.28	0.28	0.86	0.86	15	18	18	15	162	164	163	164
	DotY 2016	0.85	0.85	0.88	0.88	15	19	19	15	172	173	173	172
	Average 1999-2016	0.47	0.47	0.72	0.72	15	19	19	15	0.64	0.54	0.6	0.58
	Median 1999 - 2016	0.45	0.45	0.57	0.57	15	19	19	15	0.66	0.6	0.59	0.6
	TS 2011	0.56	0.56	0.76	0.76	14	19	18	15	1201.01	1209.22	1189.92	1209.22
	TS 2012	0.39	0.39	0.84	0.84	15	19	19	15	1218.09	1207.58	1207.58	1218.09
	TS 2013	0.68	0.68	0.9	0.9	14	19	18	15	1223.75	1191.16	1191.16	1180.44
	TS 2014	0.83	0.83	0.94	0.94	15	18	18	15	1270.24	1270.24	1270.24	1270.24
	TS 2015	0.58	0.58	0.87	0.87	15	16	17	14	1252.35	1252.35	1252.35	1252.35
	TS 2016	0.85	0.85	0.88	0.88	15	19	19	15	1369.31	1392.58	1392.58	1369.31
	GR 2011	0.56	0.56	0.76	0.76	14	19	18	15	553.37	556.81	549.96	556.81
	GR 2012	0.39	0.39	0.84	0.84	15	19	19	15	554.09	550.2	550.2	554.09
	GR 2013	0.68	0.68	0.9	0.9	14	19	18	15	558.67	543.96	543.96	541.09
	GR 2014	0.83	0.83	0.94	0.94	15	18	18	15	561.04	561.04	561.04	561.04
	GR 2015	0.58	0.58	0.87	0.87	15	16	17	14	565.4	565.4	565.4	565.4
	GR 2016	0.85	0.85	0.88	0.88	15	19	19	15	517.14	525.6	525.6	517.14
VIT_212s0028g03350	DotY 1999	0.24	0.24	0.51	0.51	11	22	14	19	167	168	168	168
	DotY 2009	0.2	0.2	0.36	0.36	12	22	14	20	163.5	164	164	164
	DotY 2010	0.38	0.38	0.22	0.22	12	22	14	20	176	177.5	178	175.5
	DotY 2011	0.31	0.31	0.25	0.25	12	21	13	20	152	154	154	152
	DotY 2012	0.63	0.63	0.35	0.35	12	22	14	20	161	160.5	161	160
	DotY 2013	0.05	0.05	0.68	0.68	11	22	14	19	172	174	173.5	172
	DotY 2014	0.83	0.83	0.29	0.29	12	22	14	20	157	157	157	157
	DotY 2015	0.39	0.39	0.37	0.37	12	21	14	19	162	164	164	162
	DotY 2016	0.84	0.84	0.54	0.54	12	22	14	20	173	172.5	172.5	173
	Average 1999-2016	0.39	0.39	0.37	0.37	12	22	14	20	0.53	0.6	0.65	0.54
	Median 1999 - 2016	0.46	0.46	0.44	0.44	12	22	14	20	0.58	0.61	0.64	0.58
	TS 2011	0.4	0.4	0.34	0.34	12	21	13	20	1189.92	1209.22	1209.22	1189.92
	TS 2012	0.79	0.79	0.54	0.54	12	22	14	20	1218.09	1207.58	1218.09	1207.58
	TS 2013	0.11	0.11	0.94	0.94	11	22	14	19	1180.44	1202.28	1191.16	1180.44
	TS 2014	0.88	0.88	0.47	0.47	12	21	14	19	1270.24	1270.24	1270.24	1270.24
	TS 2015	0.48	0.48	0.69	0.69	11	20	13	18	1252.35	1271.71	1252.35	1252.35
	TS 2016	0.84	0.84	0.54	0.54	12	22	14	20	1392.58	1380.95	1380.95	1392.58
	GR 2011	0.4	0.4	0.34	0.34	12	21	13	20	549.96	556.81	556.81	549.96
	GR 2012	0.79	0.79	0.54	0.54	12	22	14	20	554.09	550.2	554.09	550.5
	GR 2013	0.11	0.11	0.94	0.94	11	22	14	19	541.09	549.33	543.96	541.09
	GR 2014	0.88	0.88	0.47	0.47	12	21	14	19	561.04	561.04	561.04	561.04

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	GR 2015	0.48	0.48	0.69	0.69	11	20	13	18	565.4	572.7	565.4	565.4
	GR 2016	0.84	0.84	0.54	0.54	12	22	14	20	525.6	521.37	521.37	525.6
VIT_213s0067g03390	DotY 1999	0.88	0.88	0.52	0.52	13	14	9	18	168	167.5	168	168
	DotY 2009	0.66	0.66	0.29	0.29	14	14	9	19	164	164	164	164
	DotY 2010	0.61	0.61	0.38	0.38	14	14	9	19	177	176.5	175	177
	DotY 2011	0.87	0.87	0.23	0.23	14	13	8	19	153	153	151.5	153
	DotY 2012	0.93	0.93	0.49	0.49	14	14	9	19	161	160.5	161	161
	DotY 2013	0.94	0.94	0.82	0.82	14	13	9	18	173	173	173	173.5
	DotY 2014	0.7	0.7	0.98	0.98	14	14	9	19	157	157.5	157	157
	DotY 2015	1	1	0.63	0.63	14	13	9	18	162	164	162	164
	DotY 2016	0.71	0.71	0.79	0.79	14	14	9	19	173	172	173	172
	Average 1999-2016	0.78	0.78	0.45	0.45	14	14	9	19	0.59	0.6	0.58	0.59
	Median 1999 - 2016	0.73	0.73	0.43	0.43	14	14	9	19	0.6	0.61	0.6	0.59
	TS 2011	0.79	0.79	0.28	0.28	14	13	8	19	1189.92	1209.22	1185.09	1209.22
	TS 2012	0.77	0.77	0.68	0.68	14	14	9	19	1212.84	1212.84	1218.09	1207.58
	TS 2013	0.75	0.75	0.86	0.86	14	13	9	18	1180.44	1191.16	1191.16	1185.8
	TS 2014	0.88	0.88	0.73	0.73	14	13	8	19	1270.24	1245.36	1245.36	1270.24
	TS 2015	0.84	0.84	0.91	0.91	14	12	9	17	1252.35	1271.71	1252.35	1252.35
	TS 2016	0.71	0.71	0.79	0.79	14	14	9	19	1392.58	1369.31	1392.58	1369.31
	GR 2011	0.79	0.79	0.28	0.28	14	13	8	19	549.96	556.81	549.19	556.81
	GR 2012	0.77	0.77	0.68	0.68	14	14	9	19	552.14	552.14	554.09	550.5
	GR 2013	0.75	0.75	0.86	0.86	14	13	9	18	541.09	543.96	543.96	542.53
	GR 2014	0.88	0.88	0.73	0.73	14	13	8	19	561.04	553.06	553.06	561.04
	GR 2015	0.84	0.84	0.91	0.91	14	12	9	17	565.4	572.7	565.4	565.4
	GR 2016	0.71	0.71	0.79	0.79	14	14	9	19	525.6	517.14	525.6	517.14
VIT_213s0019g03550	DotY 1999	0.29	0.29	0.76	0.76	18	15	11	22	167.5	168	168	168
	DotY 2009	0.27	0.27	0.47	0.47	19	15	12	22	164	164	164	164
	DotY 2010	0.13	0.13	0.72	0.72	19	15	12	22	176	178	177.5	176.5
	DotY 2011	0.27	0.27	0.67	0.67	19	14	11	22	151	153.5	153	152
	DotY 2012	0.3	0.3	0.68	0.68	19	15	12	22	160	161	161	160
	DotY 2013	0.66	0.66	0.76	0.76	19	14	12	21	172	173.5	173	173
	DotY 2014	0.32	0.32	0.83	0.83	19	15	12	22	157	158	157	157
	DotY 2015	0.68	0.68	0.84	0.84	18	15	12	21	162	164	162	164
	DotY 2016	0.48	0.48	0.4	0.4	19	15	12	22	173	173	173	172
	Average 1999-2016	0.2	0.2	0.97	0.97	19	15	12	22	0.53	0.65	0.61	0.54
	Median 1999 - 2016	0.23	0.23	0.94	0.94	19	15	12	22	0.56	0.61	0.61	0.57
	TS 2011	0.27	0.27	0.67	0.67	19	14	11	22	1170.61	1220.32	1209.22	1189.92
	TS 2012	0.3	0.3	0.68	0.68	19	15	12	22	1207.58	1218.09	1218.09	1207.58
	TS 2013	0.66	0.66	0.76	0.76	19	14	12	21	1180.44	1196.72	1191.16	1191.16
	TS 2014	0.62	0.62	0.89	0.89	19	14	11	22	1270.24	1269.82	1270.24	1270.24
	TS 2015	0.69	0.69	0.7	0.7	17	14	12	19	1252.35	1271.71	1252.35	1252.35
	TS 2016	0.48	0.48	0.4	0.4	19	15	12	22	1392.58	1392.58	1392.58	1369.31
	GR 2011	0.27	0.27	0.67	0.67	19	14	11	22	543.1	560.23	556.81	549.96
	GR 2012	0.3	0.3	0.68	0.68	19	15	12	22	550.2	554.09	554.09	550.5
	GR 2013	0.66	0.66	0.76	0.76	19	14	12	21	541.09	546.64	543.96	543.96
	GR 2014	0.62	0.62	0.89	0.89	19	14	11	22	561.04	560.03	561.04	561.04
	GR 2015	0.69	0.69	0.7	0.7	17	14	12	19	565.4	572.7	565.4	565.4
	GR 2016	0.48	0.48	0.4	0.4	19	15	12	22	525.6	525.6	525.6	517.14
VIT_214s0030g00440	DotY 1999	0.35	0.35	0.53	0.53	5	17	10	12	168	167	169.5	168
	DotY 2009	0.13	0.13	0.85	0.85	6	17	11	12	168.5	164	165	164
	DotY 2010	0.67	0.67	0.25	0.25	6	17	11	12	178	177	178	175.5
	DotY 2011	0.18	0.18	0.45	0.45	5	17	11	11	155	153	154	153
	DotY 2012	0.05	0.05	0.27	0.27	6	17	11	12	164	160	164	161
	DotY 2013	0.27	0.27	0.06	0.06	5	17	11	11	179	172	178	172
	DotY 2014	0.29	0.29	0.14	0.14	6	17	11	12	158	157	157	156
	DotY 2015	0.25	0.25	0.28	0.28	6	17	11	12	164	162	164	162
	DotY 2016	0.09	0.09	0.98	0.98	6	17	11	12	173.5	172	172	173
	Average 1999-2016	0.14	0.14	0.27	0.27	6	17	11	12	0.73	0.53	0.72	0.56

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	Median 1999 - 2016	0.13	0.13	0.32	0.32	6	17	11	12	0.74	0.56	0.69	0.58
	TS 2011	0.16	0.16	0.6	0.6	5	17	11	11	1251.21	1170.61	1209.22	1209.22
	TS 2012	0.03	0.03	0.44	0.44	6	17	11	12	1258.49	1207.58	1218.09	1218.09
	TS 2013	0.18	0.18	0.19	0.19	5	17	11	11	1267.05	1180.44	1238.26	1180.44
	TS 2014	0.22	0.22	0.37	0.37	6	17	11	12	1282.26	1270.24	1270.24	1245.36
	TS 2015	0.13	0.13	0.64	0.64	6	16	10	12	1291.07	1252.35	1271.71	1252.35
	TS 2016	0.09	0.09	0.98	0.98	6	17	11	12	1400.95	1369.31	1369.31	1392.58
	GR 2011	0.16	0.16	0.6	0.6	5	17	11	11	569.43	543.1	556.81	556.81
	GR 2012	0.03	0.03	0.44	0.44	6	17	11	12	568.36	550.2	554.09	554.09
	GR 2013	0.18	0.18	0.19	0.19	5	17	11	11	576.24	541.09	562.47	541.09
	GR 2014	0.22	0.22	0.37	0.37	6	17	11	12	564.02	561.04	561.04	553.06
	GR 2015	0.13	0.13	0.64	0.64	6	16	10	12	580.01	565.4	572.7	565.4
	GR 2016	0.09	0.09	0.98	0.98	6	17	11	12	529.38	517.14	517.14	525.6
VIT_214s0081g00440	DotY 1999	0.13	0.13	0.55	0.55	6	19	14	11	168.5	167	167	168
	DotY 2009	0.08	0.08	0.86	0.86	7	19	15	11	168	163	164	164
	DotY 2010	0.4	0.4	0.13	0.13	7	19	15	11	178	176	177	175
	DotY 2011	0.11	0.11	0.51	0.51	6	19	15	10	154.5	151	153	152
	DotY 2012	0.03	0.03	0.47	0.47	7	19	15	11	164	157	161	161
	DotY 2013	0.2	0.2	0.24	0.24	6	19	15	10	175.5	172	177	172
	DotY 2014	0.14	0.14	0.28	0.28	7	19	15	11	157	157	157	156
	DotY 2015	0.21	0.21	0.82	0.82	7	18	15	10	164	162	164	162
	DotY 2016	0.03	0.03	0.94	0.94	7	19	15	11	173	172	172	173
	Average 1999-2016	0.07	0.07	0.39	0.39	7	19	15	11	0.65	0.48	0.54	0.53
	Median 1999 - 2016	0.07	0.07	0.39	0.39	7	19	15	11	0.67	0.49	0.6	0.55
	TS 2011	0.09	0.09	0.64	0.64	6	19	15	10	1241.31	1170.61	1170.61	1189.92
	TS 2012	0.02	0.02	0.66	0.66	7	19	15	11	1258.49	1167.18	1207.58	1218.09
	TS 2013	0.13	0.13	0.47	0.47	6	19	15	10	1223.75	1180.44	1191.16	1180.44
	TS 2014	0.09	0.09	0.44	0.44	7	19	15	11	1270.24	1201.47	1270.24	1245.36
	TS 2015	0.1	0.1	0.77	0.77	7	17	14	10	1291.07	1213.66	1252.35	1252.35
	TS 2016	0.03	0.03	0.94	0.94	7	19	15	11	1392.58	1369.31	1369.31	1392.58
	GR 2011	0.09	0.09	0.64	0.64	6	19	15	10	566.53	543.1	543.1	549.96
	GR 2012	0.02	0.02	0.66	0.66	7	19	15	11	568.36	530.74	550.2	554.09
	GR 2013	0.13	0.13	0.47	0.47	6	19	15	10	558.67	541.09	543.96	541.09
	GR 2014	0.09	0.09	0.44	0.44	7	19	15	11	561.04	537.65	561.04	553.06
	GR 2015	0.1	0.1	0.77	0.77	7	17	14	10	580.01	553.37	565.4	565.4
	GR 2016	0.03	0.03	0.94	0.94	7	19	15	11	525.6	517.14	517.14	525.6
VIT_214s0006g00640	DotY 1999	0.08	0.08	0.07	0.07	11	16	12	15	169	166.5	170.5	167
	DotY 2009	0.09	0.09	0.23	0.23	12	16	13	15	164	162	165	164
	DotY 2010	0.43	0.43	0.02	0.02	12	16	13	15	178	175.5	178	175
	DotY 2011	0.16	0.16	0.03	0.03	11	16	13	14	154	151	154	150.5
	DotY 2012	0.03	0.03	0.08	0.08	12	16	13	15	161	156.5	164	157
	DotY 2013	0.01	0.01	0.04	0.04	11	16	13	14	179	171.5	179	172
	DotY 2014	0.01	0.01	0.12	0.12	12	16	13	15	158	155	157	155
	DotY 2015	0.04	0.04	0.29	0.29	12	15	13	14	164	160	164	162
	DotY 2016	0.01	0.01	0.31	0.31	12	16	13	15	173	172	172	172
	Average 1999-2016	0.03	0.03	0.05	0.05	12	16	13	15	0.65	0.42	0.72	0.47
	Median 1999 - 2016	0.03	0.03	0.06	0.06	12	16	13	15	0.67	0.41	0.69	0.46
	TS 2011	0.13	0.13	0.05	0.05	11	16	13	14	1231.41	1170.61	1209.22	1165.79
	TS 2012	0.02	0.02	0.13	0.13	12	16	13	15	1218.09	1159.02	1218.09	1167.18
	TS 2013	0	0	0.14	0.14	11	16	13	14	1267.05	1160.49	1238.26	1180.44
	TS 2014	0.01	0.01	0.24	0.24	12	16	13	15	1282.26	1201.47	1270.24	1201.47
	TS 2015	0.02	0.02	0.43	0.43	11	15	13	13	1291.07	1213.66	1252.35	1252.35
	TS 2016	0.01	0.01	0.31	0.31	12	16	13	15	1392.58	1369.31	1369.31	1369.31
	GR 2011	0.13	0.13	0.05	0.05	11	16	13	14	563.64	543.1	556.81	542.34
	GR 2012	0.02	0.02	0.13	0.13	12	16	13	15	554.09	528.3	554.09	530.74
	GR 2013	0	0	0.14	0.14	11	16	13	14	576.24	533.57	562.47	541.09
	GR 2014	0.01	0.01	0.24	0.24	12	16	13	15	564.02	537.65	561.04	537.65
	GR 2015	0.02	0.02	0.43	0.43	11	15	13	13	580.01	553.37	565.4	565.4

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	GR 2016	0.01	0.01	0.31	0.31	12	16	13	15	525.6	517.14	517.14	517.14
VIT_214s0083g01030	DotY 1999	0.01	0.01	-	-	11	13			169	166	-	-
	DotY 2009	0.01	0.01	-	-	12	13			166	162	-	-
	DotY 2010	0.12	0.12	-	-	12	13			178	175	-	-
	DotY 2011	0.04	0.04	-	-	11	13			155	151	-	-
	DotY 2012	0.01	0.01	-	-	12	13			162.5	156	-	-
	DotY 2013	0.01	0.01	-	-	11	13			179	171	-	-
	DotY 2014	0.01	0.01	-	-	12	13			158.5	155	-	-
	DotY 2015	0.01	0.01	-	-	12	12			164	160	-	-
	DotY 2016	0	0	-	-	12	13			173.5	172	-	-
	Average 1999-2016	0	0	-	-	12	13			0.68	0.37	-	-
	Median 1999-2016	0	0	-	-	12	13			0.73	0.41	-	-
	TS 2011	0.03	0.03	-	-	11	13			1251.21	1170.61	-	-
	TS 2012	0	0	-	-	12	13			1238.29	1150.85	-	-
	TS 2013	0	0	-	-	11	13			1267.05	1153.32	-	-
	TS 2014	0	0	-	-	12	13			1294.27	1201.47	-	-
	TS 2015	0	0	-	-	11	12			1291.07	1197.3	-	-
	TS 2016	0	0	-	-	12	13			1400.95	1369.31	-	-
	GR 2011	0.03	0.03	-	-	11	13			569.43	543.1	-	-
	GR 2012	0	0	-	-	12	13			561.23	525.86	-	-
	GR 2013	0	0	-	-	11	13			576.24	530.55	-	-
	GR 2014	0	0	-	-	12	13			566.99	537.65	-	-
	GR 2015	0	0	-	-	11	12			580.01	548.42	-	-
	GR 2016	0	0	-	-	12	13			529.38	517.14	-	-
VIT_214s0083g01050	DotY 1999	-	0.06	-	0.11	12	11	13	10	168.5	166	169	167
	DotY 2009	-	0.08	-	0.21	13	11	14	10	164	162	164.5	163.5
	DotY 2010	-	0.54	-	0.14	13	11	14	10	178	177	177.5	175
	DotY 2012	-	0.05	-	0.14	13	11	14	10	161	156	162.5	158
	DotY 2013	-	0.04	-	0.04	12	11	14	9	178.5	171	178.5	171
	DotY 2014	-	0.05	-	0.17	13	11	14	10	157	155	157	155
	DotY 2015	-	0.07	-	0.59	13	10	14	9	164	160	163	162
	DotY 2016	-	0.02	-	0.28	13	11	14	10	173	172	173	172
	Average 1999-2016	-	0.04	-	0.11	13	11	14	10	0.64	0.37	0.63	0.42
	Median 1999-2016	-	0.03	-	0.11	13	11	14	10	0.66	0.41	0.64	0.44
	TS 2011	-	0.19	-	0.34	12	11	13	10	1220.32	1170.61	1209.22	1170.61
	TS 2012	-	0.01	-	0.22	13	11	14	10	1218.09	1150.85	1218.09	1179.22
	TS 2013	-	0.01	-	0.11	12	11	14	9	1259.03	1153.32	1244.64	1167.66
	TS 2014	-	0.02	-	0.27	13	11	14	10	1270.24	1201.47	1270.24	1201.47
	TS 2015	-	0.03	-	0.92	12	10	13	9	1291.07	1197.3	1252.35	1252.35
	TS 2016	-	0.02	-	0.28	13	11	14	10	1392.58	1369.31	1392.58	1369.31
	GR 2011	-	0.19	-	0.34	12	11	13	10	560.23	543.1	556.81	543.5
	GR 2012	-	0.01	-	0.22	13	11	14	10	554.09	525.86	554.09	538.03
	GR 2013	-	0.01	-	0.11	12	11	14	9	572.06	530.55	565.18	536.59
	GR 2014	-	0.02	-	0.27	13	11	14	10	561.04	537.65	561.04	537.65
	GR 2015	-	0.03	-	0.92	12	10	13	9	580.01	548.42	565.4	565.4
	GR 2016	-	0.02	-	0.28	13	11	14	10	525.6	517.14	525.6	517.14
VIT_214s0068g01800	DotY 1999	0.03	0.03	0.61	0.61	13	14	14	13	169	166	167.5	168
	DotY 2009	0.04	0.04	0.91	0.91	14	14	15	13	164	162	164	164
	DotY 2010	0.3	0.3	0.98	0.98	14	14	15	13	177.5	175	177	176
	DotY 2011	0.13	0.13	0.96	0.96	13	14	14	13	153	150.5	152	151
	DotY 2012	0.03	0.03	0.8	0.8	14	14	15	13	161	156	161	160
	DotY 2013	0	0	0.53	0.53	13	14	15	12	179	170.5	177	172
	DotY 2014	0.04	0.04	1	1	14	14	15	13	157	155	157	156
	DotY 2015	0.08	0.08	0.68	0.68	14	13	15	12	164	160	162	163
	DotY 2016	0.03	0.03	0.89	0.89	14	14	15	13	173	172	172	172
	Average 1999-2016	0.03	0.03	0.65	0.65	14	14	15	13	0.63	0.37	0.58	0.47
	Median 1999 - 2016	0.02	0.02	0.68	0.68	14	14	15	13	0.64	0.4	0.6	0.46
	TS 2011	0.09	0.09	0.83	0.83	13	14	14	13	1209.22	1165.79	1170.61	1170.61

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	TS 2012	0.01	0.01	0.96	0.96	14	14	15	13	1218.09	1150.85	1207.58	1207.58
	TS 2013	0	0	0.85	0.85	13	14	15	12	1267.05	1153.32	1191.16	1180.44
	TS 2014	0.01	0.01	0.82	0.82	14	14	15	13	1270.24	1201.47	1270.24	1245.36
	TS 2015	0.03	0.03	0.37	0.37	13	13	14	12	1291.07	1180.93	1252.35	1271.71
	TS 2016	0.03	0.03	0.89	0.89	14	14	15	13	1392.58	1369.31	1369.31	1369.31
	GR 2011	0.09	0.09	0.83	0.83	13	14	14	13	556.81	542.34	543.1	543.5
	GR 2012	0.01	0.01	0.96	0.96	14	14	15	13	554.09	525.86	550.2	550.5
	GR 2013	0	0	0.85	0.85	13	14	15	12	576.24	530.55	543.96	541.09
	GR 2014	0.01	0.01	0.82	0.82	14	14	15	13	561.04	537.65	561.04	553.06
	GR 2015	0.03	0.03	0.37	0.37	13	13	14	12	580.01	543.46	565.4	572.7
	GR 2016	0.03	0.03	0.89	0.89	14	14	15	13	525.6	517.14	517.14	517.14
VIT_215s0048g01280	DotY 1999	-	-	0.4	0.4			15	18	-	-	169	168
	DotY 2009	-	-	0.86	0.86			16	18	-	-	164	164
	DotY 2010	-	-	0.42	0.42			16	18	-	-	177.5	175.5
	DotY 2011	-	-	0.84	0.84			16	17	-	-	153	153
	DotY 2012	-	-	0.51	0.51			16	18	-	-	161	160
	DotY 2013	-	-	0.52	0.52			15	18	-	-	173	173.5
	DotY 2014	-	-	0.17	0.17			16	18	-	-	157	156.5
	DotY 2015	-	-	0.77	0.77			16	17	-	-	163	162
	DotY 2016	-	-	0.21	0.21			16	18	-	-	173	172
	Average 1999-2016	-	-	0.33	0.33			16	18	-	-	0.59	0.54
	Median 1999-2016	-	-	0.31	0.31			16	18	-	-	0.6	0.55
	TS 2011	-	-	0.73	0.73			16	17	-	-	1209.22	1170.61
	TS 2012	-	-	0.35	0.35			16	18	-	-	1218.09	1201.17
	TS 2013	-	-	0.31	0.31			15	18	-	-	1191.16	1185.8
	TS 2014	-	-	0.08	0.08			16	17	-	-	1270.24	1245.36
	TS 2015	-	-	0.55	0.55			15	16	-	-	1252.35	1252.35
	TS 2016	-	-	0.21	0.21			16	18	-	-	1392.58	1369.31
	GR 2011	-	-	0.73	0.73			16	17	-	-	556.81	543.5
	GR 2012	-	-	0.35	0.35			16	18	-	-	554.09	547.18
	GR 2013	-	-	0.31	0.31			15	18	-	-	543.96	542.53
	GR 2014	-	-	0.08	0.08			16	17	-	-	561.04	553.06
	GR 2015	-	-	0.55	0.55			15	16	-	-	565.4	565.4
	GR 2016	-	-	0.21	0.21			16	18	-	-	525.6	517.14
VIT_216s0013g00860	DotY 1999	-	0.66	-	0.49	13	21	22	12	167	168	168	169
	DotY 2009	-	0.8	-	0.5	13	22	23	12	164	164	164	164
	DotY 2010	-	0.44	-	0.58	13	22	23	12	175	177.5	176	177.5
	DotY 2012	-	0.67	-	0.94	13	22	23	12	160	161	161	160
	DotY 2013	-	0.87	-	0.32	12	22	22	12	173	173.5	172.5	174
	DotY 2014	-	0.97	-	0.2	13	22	23	12	157	157	157	158.5
	DotY 2015	-	0.25	-	0.84	13	21	22	12	162	164	162	164
	DotY 2016	-	0.88	-	0.6	13	22	23	12	172	173	173	172
	Average 1999-2016	-	0.57	-	0.72	13	22	23	12	0.54	0.62	0.54	0.63
	Median 1999-2016	-	0.6	-	0.69	13	22	23	12	0.56	0.61	0.59	0.61
	TS 2011	-	0.51	-	0.49	13	21	22	12	1170.61	1209.22	1189.92	1220.32
	TS 2012	-	0.81	-	0.93	13	22	23	12	1207.58	1218.09	1207.58	1206.42
	TS 2013	-	0.86	-	0.21	12	22	22	12	1191.16	1185.8	1180.44	1202.28
	TS 2014	-	0.72	-	0.32	13	21	23	11	1270.24	1245.36	1245.36	1294.27
	TS 2015	-	0.28	-	0.61	11	21	20	12	1252.35	1291.07	1252.35	1291.07
	TS 2016	-	0.88	-	0.6	13	22	23	12	1369.31	1392.58	1392.58	1369.31
	GR 2011	-	0.51	-	0.49	13	21	22	12	543.1	556.81	549.96	560.23
	GR 2012	-	0.81	-	0.93	13	22	23	12	550.2	554.09	550.2	549.13
	GR 2013	-	0.86	-	0.21	12	22	22	12	543.96	542.53	541.09	549.33
	GR 2014	-	0.72	-	0.32	13	21	23	11	561.04	553.06	553.06	566.99
	GR 2015	-	0.28	-	0.61	11	21	20	12	565.4	580.01	565.4	580.01
	GR 2016	-	0.88	-	0.6	13	22	23	12	517.14	525.6	525.6	517.14
	DotY 1999	0.3	0.3	0.91	0.91	12	21	19	14	168	167	168	167.5
	DotY 2009	0.34	0.34	1	1	13	21	20	14	164	164	164	164

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
VIT_217s0000g02630	DotY 2010	0.55	0.55	0.71	0.71	13	21	20	14	178	177	177.5	176.5
	DotY 2011	0.5	0.5	0.66	0.66	13	20	20	13	154	152	153	153
	DotY 2012	0.21	0.21	0.92	0.92	13	21	20	14	161	160	160.5	160.5
	DotY 2013	0.28	0.28	0.72	0.72	12	21	19	14	173.5	173	173	173.5
	DotY 2014	0.15	0.15	0.73	0.73	13	21	20	14	158	157	157	157
	DotY 2015	0.05	0.05	0.99	0.99	12	21	19	14	164	162	164	162
	DotY 2016	0.26	0.26	0.87	0.87	13	21	20	14	173	172	173	172.5
	Average 1999-2016	0.26	0.26	0.99	0.99	13	21	20	14	0.59	0.53	0.59	0.54
	Median 1999 - 2016	0.17	0.17	0.78	0.78	13	21	20	14	0.61	0.55	0.6	0.58
	TS 2011	0.5	0.5	0.66	0.66	13	20	20	13	1231.41	1189.92	1209.22	1209.22
	TS 2012	0.21	0.21	0.92	0.92	13	21	20	14	1218.09	1207.58	1212.84	1212.84
	TS 2013	0.28	0.28	0.72	0.72	12	21	19	14	1196.72	1191.16	1191.16	1196.72
	TS 2014	0.13	0.13	0.8	0.8	13	20	19	14	1294.27	1245.36	1270.24	1257.8
	TS 2015	0.05	0.05	0.97	0.97	11	20	18	13	1291.07	1252.35	1271.71	1252.35
	TS 2016	0.26	0.26	0.87	0.87	13	21	20	14	1392.58	1369.31	1392.58	1380.95
	GR 2011	0.5	0.5	0.66	0.66	13	20	20	13	563.64	549.96	556.81	556.81
	GR 2012	0.21	0.21	0.92	0.92	13	21	20	14	554.09	550.2	552.14	552.14
	GR 2013	0.28	0.28	0.72	0.72	12	21	19	14	546.64	543.96	543.96	546.64
	GR 2014	0.13	0.13	0.8	0.8	13	20	19	14	566.99	553.06	561.04	557.05
	GR 2015	0.05	0.05	0.97	0.97	11	20	18	13	580.01	565.4	572.7	565.4
	GR 2016	0.26	0.26	0.87	0.87	13	21	20	14	525.6	517.14	525.6	521.37
VIT_217s0000g06570	DotY 1999	-	0.05	-	0.49	15	19	13	21	169	167	167	168
	DotY 2009	-	0.35	-	0.53	16	19	13	22	164	164	164	164
	DotY 2010	-	0.03	-	0.43	16	19	13	22	178	175	175	178
	DotY 2012	-	0.05	-	0.37	16	19	13	22	161	159	159	161
	DotY 2013	-	0.27	-	0.5	15	19	13	21	174	172	172	173
	DotY 2014	-	0.5	-	0.25	16	19	13	22	157	157	157	157
	DotY 2015	-	0.11	-	0.94	15	19	13	21	164	162	162	164
	DotY 2016	-	0.32	-	0.96	16	19	13	22	173	172	173	172.5
	Average 1999-2016	-	0.1	-	0.42	16	19	13	22	0.66	0.53	0.53	0.6
	Median 1999-2016	-	0.15	-	0.52	16	19	13	22	0.65	0.55	0.56	0.6
	TS 2011	-	0.18	-	0.63	16	18	12	22	1220.32	1170.61	1189.92	1209.22
	TS 2012	-	0.09	-	0.49	16	19	13	22	1218.09	1194.75	1194.75	1218.09
	TS 2013	-	0.54	-	0.78	15	19	13	21	1191.16	1180.44	1180.44	1191.16
	TS 2014	-	0.89	-	0.68	15	19	13	21	1270.24	1270.24	1270.24	1270.24
	TS 2015	-	0.15	-	0.82	15	17	12	20	1291.07	1252.35	1252.35	1252.35
	TS 2016	-	0.32	-	0.96	16	19	13	22	1392.58	1369.31	1392.58	1380.95
	GR 2011	-	0.18	-	0.63	16	18	12	22	560.23	543.1	549.96	556.81
	GR 2012	-	0.09	-	0.49	16	19	13	22	554.09	544.16	544.16	554.09
	GR 2013	-	0.54	-	0.78	15	19	13	21	543.96	541.09	541.09	543.96
	GR 2014	-	0.89	-	0.68	15	19	13	21	561.04	561.04	561.04	561.04
	GR 2015	-	0.15	-	0.82	15	17	12	20	580.01	565.4	565.4	565.4
	GR 2016	-	0.32	-	0.96	16	19	13	22	525.6	517.14	525.6	521.37
VIT_217s0000g08480	DotY 1999	-	0.33	-	0.15	8	16	10	14	168	169	170.5	167.5
	DotY 2009	-	0.41	-	0.18	8	17	11	14	164	164	164	164
	DotY 2010	-	0.13	-	0.3	8	17	11	14	175	178	178	175
	DotY 2012	-	0.41	-	0.22	8	17	11	14	159.5	161	162	159.5
	DotY 2013	-	0.24	-	0.18	8	17	11	14	172	178	179	172.5
	DotY 2014	-	0.2	-	0.37	8	17	11	14	156	157	157	156.5
	DotY 2015	-	0.39	-	0.58	8	16	10	14	162	164	164	162
	DotY 2016	-	0.22	-	0.15	8	17	11	14	172.5	173	173	172.5
	Average 1999-2016	-	0.27	-	0.19	8	17	11	14	0.5	0.6	0.72	0.51
	Median 1999-2016	-	0.28	-	0.27	8	17	11	14	0.51	0.6	0.73	0.52
	TS 2011	-	0.73	-	0.69	7	17	11	13	1170.61	1170.61	1209.22	1170.61
	TS 2012	-	0.6	-	0.4	8	17	11	14	1201.17	1207.58	1218.09	1201.17
	TS 2013	-	0.47	-	0.49	8	17	11	14	1180.44	1238.26	1267.05	1185.8
	TS 2014	-	0.5	-	0.79	8	16	10	14	1245.36	1270.24	1270.24	1245.36
	TS 2015	-	0.72	-	0.83	8	15	10	13	1252.35	1252.35	1271.71	1252.35

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	TS 2016	-	0.22	-	0.15	8	17	11	14	1380.95	1392.58	1392.58	1380.95
	GR 2011	-	0.73	-	0.69	7	17	11	13	543.1	543.1	556.81	543.5
	GR 2012	-	0.6	-	0.4	8	17	11	14	547.18	550.2	554.09	547.18
	GR 2013	-	0.47	-	0.49	8	17	11	14	541.09	562.47	576.24	542.53
	GR 2014	-	0.5	-	0.79	8	16	10	14	553.06	561.04	561.04	553.06
	GR 2015	-	0.72	-	0.83	8	15	10	13	565.4	565.4	572.7	565.4
	GR 2016	-	0.22	-	0.15	8	17	11	14	521.37	525.6	525.6	521.37
VIT_217s0000g10300	DotY 1999	0.25	0.25	0.64	0.68	16	17	14	18	168.5	167	167.5	168.5
	DotY 2009	0.57	0.57	0.7	0.92	16	18	14	19	164	164	164	164
	DotY 2010	0.76	0.76	0.62	0.36	16	18	14	19	176.5	177	175	178
	DotY 2011	0.58	0.58	0.73	0.62	16	17	13	19	153.5	153	151	153
	DotY 2012	0.6	0.6	0.73	0.84	16	18	14	19	161	160	159.5	161
	DotY 2013	0.53	0.53	0.9	0.77	16	17	14	18	174	172	173.5	173
	DotY 2014	0.99	0.99	0.47	0.38	16	18	14	19	157	157	157	157
	DotY 2015	0.93	0.93	0.83	0.74	15	18	14	18	162	163	162	164
	DotY 2016	0.44	0.44	0.82	0.86	16	18	14	19	173	172	173	173
	Average 1999-2016	0.64	0.64	0.78	0.64	16	18	14	19	0.6	0.54	0.54	0.6
	Median 1999 - 2016	0.68	0.68	0.82	0.68	16	18	14	19	0.6	0.56	0.58	0.6
	TS 2011	0.58	0.58	0.73	0.62	16	17	13	19	1220.32	1209.22	1170.61	1209.22
	TS 2012	0.6	0.6	0.73	0.84	16	18	14	19	1218.09	1207.58	1201.17	1218.09
	TS 2013	0.53	0.53	0.9	0.77	16	17	14	18	1202.28	1180.44	1196.72	1191.16
	TS 2014	0.91	0.91	0.5	0.43	15	18	14	18	1270.24	1270.24	1245.36	1270.24
	TS 2015	0.77	0.77	0.79	0.74	15	16	13	17	1252.35	1252.35	1252.35	1291.07
	TS 2016	0.44	0.44	0.82	0.86	16	18	14	19	1392.58	1369.31	1392.58	1392.58
	GR 2011	0.58	0.58	0.73	0.62	16	17	13	19	560.23	556.81	543.1	556.81
	GR 2012	0.6	0.6	0.73	0.84	16	18	14	19	554.09	550.2	547.18	554.09
	GR 2013	0.53	0.53	0.9	0.77	16	17	14	18	549.33	541.09	546.64	543.96
	GR 2014	0.91	0.91	0.5	0.43	15	18	14	18	561.04	561.04	553.06	561.04
	GR 2015	0.77	0.77	0.79	0.74	15	16	13	17	565.4	565.4	565.4	580.01
	GR 2016	0.44	0.44	0.82	0.86	16	18	14	19	525.6	517.14	525.6	525.6
VIT_218s0001g07460	DotY 1999	0.07	0.07	0.03	0.03	17	16	15	18	167	168	167	168.5
	DotY 2009	0.07	0.07	0.02	0.02	17	17	15	19	163	164	162	164
	DotY 2010	0.09	0.09	0.06	0.06	17	17	15	19	175	178	175	178
	DotY 2011	0.04	0.04	0.09	0.09	16	17	15	18	150.5	153	151	153.5
	DotY 2012	0.08	0.08	0.03	0.03	17	17	15	19	160	161	157	161
	DotY 2013	0.23	0.23	0.13	0.13	17	16	15	18	173	173.5	172	173.5
	DotY 2014	0.07	0.07	0.21	0.21	17	17	15	19	155	157	156	157
	DotY 2015	0.11	0.11	0.08	0.08	16	17	15	18	162	164	160	164
	DotY 2016	0.21	0.21	0.31	0.31	17	17	15	19	172	173	172	173
	Average 1999-2016	0.06	0.06	0.03	0.03	17	17	15	19	0.48	0.59	0.46	0.64
	Median 1999 - 2016	0.05	0.05	0.06	0.06	17	17	15	19	0.49	0.6	0.41	0.61
	TS 2011	0.04	0.04	0.09	0.09	16	17	15	18	1165.79	1209.22	1170.61	1220.32
	TS 2012	0.08	0.08	0.03	0.03	17	17	15	19	1207.58	1218.09	1167.18	1218.09
	TS 2013	0.23	0.23	0.13	0.13	17	16	15	18	1191.16	1196.72	1180.44	1196.72
	TS 2014	0.01	0.01	0.18	0.18	16	17	14	19	1201.47	1270.24	1223.42	1270.24
	TS 2015	0.15	0.15	0.06	0.06	16	15	14	17	1252.35	1291.07	1213.66	1291.07
	TS 2016	0.21	0.21	0.31	0.31	17	17	15	19	1369.31	1392.58	1369.31	1392.58
	GR 2011	0.04	0.04	0.09	0.09	16	17	15	18	542.34	556.81	543.1	560.23
	GR 2012	0.08	0.08	0.03	0.03	17	17	15	19	550.2	554.09	530.74	554.09
	GR 2013	0.23	0.23	0.13	0.13	17	16	15	18	543.96	546.64	541.09	546.64
	GR 2014	0.01	0.01	0.18	0.18	16	17	14	19	537.65	561.04	545.36	561.04
	GR 2015	0.15	0.15	0.06	0.06	16	15	14	17	565.4	580.01	553.37	580.01
	GR 2016	0.21	0.21	0.31	0.31	17	17	15	19	517.14	525.6	517.14	525.6
	DotY 1999	0.08	-	0.01	-	16	9	10	15	168.5	167	166.5	169
	DotY 2009	0.17	-	0.02	-	16	9	10	15	164	164	162.5	164
	DotY 2010	0.09	-	0	-	16	9	10	15	178	175	175	178
	DotY 2011	0.17	-	0.02	-	16	9	10	15	153.5	151	150.5	154
	DotY 2012	0.12	-	0.02	-	16	9	10	15	161	157	156.5	161

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
VIT_200s0203g00080	DotY 2013	0.77	-	0.13	-	15	9	10	14	173	174	172	175.5
	DotY 2014	0.69	-	0.05	-	16	9	10	15	157	157	155.5	157
	DotY 2015	0.55	-	0.11	-	16	9	10	15	164	162	161	164
	DotY 2016	0.69	-	0.24	-	16	9	10	15	172.5	173	172.5	173
	Average 1999-2016	0.2	-	0.01	-	16	9	10	15	0.6	0.51	0.42	0.64
	Median 1999-2016	0.32	-	0.02	-	16	9	10	15	0.6	0.56	0.44	0.66
	TS 2011	0.2	-	0.02	-	16	9	10	15	1209.22	1170.61	1165.79	1231.41
	TS 2012	0.17	-	0.03	-	16	9	10	15	1218.09	1167.18	1159.02	1218.09
	TS 2013	0.88	-	0.32	-	15	9	10	14	1191.16	1202.28	1180.44	1196.72
	TS 2014	0.6	-	0.21	-	16	8	10	14	1270.24	1235.86	1223.42	1270.24
	TS 2015	0.65	-	0.17	-	15	8	9	14	1252.35	1233.01	1213.66	1291.07
	TS 2016	0.69	-	0.24	-	16	9	10	15	1380.95	1392.58	1380.95	1392.58
	GR 2011	0.2	-	0.02	-	16	9	10	15	556.81	543.1	542.34	563.64
	GR 2012	0.17	-	0.03	-	16	9	10	15	554.09	530.74	528.3	554.09
	GR 2013	0.88	-	0.32	-	15	9	10	14	543.96	549.33	541.09	546.64
	GR 2014	0.6	-	0.21	-	16	8	10	14	561.04	549.35	545.36	561.04
	GR 2015	0.65	-	0.17	-	15	8	9	14	565.4	559.38	553.37	580.01
	GR 2016	0.69	-	0.24	-	16	9	10	15	521.37	525.6	521.37	525.6
VIT_200s0203g00080	DotY 1999	0.13	-	0.04	-	1	22	14	9	176	167.5	167	172
	DotY 2009	0.17	-	0.1	-	1	23	15	9	169	164	162	164
	DotY 2010	0.61	-	0.05	-	1	23	15	9	178	177	177	179
	DotY 2011	0.1	-	0.16	-	1	22	14	9	157	152	151	155
	DotY 2012	0.19	-	0.13	-	1	23	15	9	168	161	160	162
	DotY 2013	0.2	-	0.08	-	1	22	14	9	181	173.5	172	179
	DotY 2014	0.28	-	0.05	-	1	23	15	9	159	157	156	159
	DotY 2015	0.45	-	0.05	-	1	22	14	9	164	162	162	164
	DotY 2016	0.83	-	0.01	-	1	23	15	9	172	172	171	173
	Average 1999-2016	0.17	-	0.04	-	1	23	15	9	0.83	0.54	0.47	0.73
	Median 1999-2016	0.19	-	0.07	-	1	23	15	9	0.84	0.59	0.46	0.69
	TS 2011	0.1	-	0.14	-	1	22	14	9	1289.44	1170.61	1170.61	1251.21
	TS 2012	0.19	-	0.08	-	1	23	15	9	1327.95	1207.58	1167.18	1229.49
	TS 2013	0.2	-	0.04	-	1	22	14	9	1300.1	1185.8	1174.05	1267.05
	TS 2014	0.23	-	0.11	-	1	22	15	8	1294.27	1257.8	1245.36	1282.26
	TS 2015	0.36	-	0.02	-	1	20	13	8	1291.07	1252.35	1213.66	1291.07
	TS 2016	0.83	-	0.01	-	1	23	15	9	1369.31	1369.31	1346.77	1392.58
	GR 2011	0.1	-	0.14	-	1	22	14	9	579.43	543.1	543.1	569.43
	GR 2012	0.19	-	0.08	-	1	23	15	9	591.57	550.2	530.74	558.13
	GR 2013	0.2	-	0.04	-	1	22	14	9	586.4	542.53	538.84	576.24
	GR 2014	0.23	-	0.11	-	1	22	15	8	566.99	557.05	553.06	564.02
	GR 2015	0.36	-	0.02	-	1	20	13	8	580.01	565.4	553.37	580.01
	GR 2016	0.83	-	0.01	-	1	23	15	9	517.14	517.14	511.31	525.6
VIT_200s1675g00010	DotY 1999	-	0.24	-	0.35	7	24	18	13	167	168	168	168
	DotY 2009	-	0.2	-	0.27	7	25	19	13	162	164	164	164
	DotY 2010	-	0.25	-	0.49	7	25	19	13	177	178	177	178
	DotY 2012	-	0.29	-	0.31	7	25	19	13	161	161	161	161
	DotY 2013	-	0.26	-	0.7	7	24	19	12	172	173.5	173	173
	DotY 2014	-	0.35	-	0.16	7	25	19	13	156	157	157	159
	DotY 2015	-	0.25	-	0.72	7	25	19	13	162	164	162	164
	DotY 2016	-	0.14	-	0.12	7	25	19	13	172	173	172	173
	Average 1999-2016	-	0.24	-	0.37	7	25	19	13	0.51	0.58	0.59	0.54
	Median 1999-2016	-	0.44	-	0.47	7	25	19	13	0.56	0.6	0.6	0.6
	TS 2011	-	0.6	-	0.69	6	25	18	13	1185.09	1209.22	1209.22	1209.22
	TS 2012	-	0.37	-	0.22	7	25	19	13	1218.09	1207.58	1207.58	1218.09
	TS 2013	-	0.36	-	0.47	7	24	19	12	1180.44	1191.16	1180.44	1191.16
	TS 2014	-	0.49	-	0.39	7	24	19	12	1245.36	1270.24	1245.36	1282.26
	TS 2015	-	0.4	-	0.64	7	23	19	11	1252.35	1252.35	1252.35	1252.35
	TS 2016	-	0.14	-	0.12	7	25	19	13	1369.31	1392.58	1369.31	1392.58
	GR 2011	-	0.6	-	0.69	6	25	18	13	549.19	556.81	556.81	556.81

Gene	Data set	E1	E2	L1	L2	# E1	# E2	# L1	# L2	Md E1	Md E2	Md L1	Md L2
	GR 2012	-	0.37	-	0.22	7	25	19	13	554.09	550.2	550.2	554.09
	GR 2013	-	0.36	-	0.47	7	24	19	12	541.09	543.96	541.09	543.96
	GR 2014	-	0.49	-	0.39	7	24	19	12	553.06	561.04	553.06	564.02
	GR 2015	-	0.4	-	0.64	7	23	19	11	565.4	565.4	565.4	565.4
	GR 2016	-	0.14	-	0.12	7	25	19	13	517.14	525.6	517.14	525.6

Table A11: Segregation of markers in several FTC-candidate genes over the mapping population GF.GA-47-42 x 'Villard Blanc', the expected segregation from the allele phasing workflow as well as expected and observed product sizes. The markers GAVBInd_019 and GAVBInd_020 were not designed using the obtained allele sequences of GF.GA-47-42 and 'Villard Blanc', since suitable InDels were not available but instead based on InDels upstream of the phased regions. An expected segregation and expected product sizes are therefore not available. Observed product sizes can deviate from the expected ones by 1-2 bp due to the used measuring method. The markers GAVBInd_004, GAVBInd_014, and GAVBInd_019 therefore have two different segregation patterns. ab x cd: four alleles/ both parents heterozygous, hk x hk: 2 alleles/ both parents heterozygous, ef x eg: 3 alleles/ parents heterozygous, lm x ll: 2 alleles/ mother heterozygous, nn x np: 2 alleles, father heterozygous. x: amplification failed. GF: GF.GA-47-42, VB: 'Villard Blanc'.

Gene	**Marker**	**Expected product sizes**		**Expected segregation**	**Observed product sizes**		**Observed segregation**
		GF	**VB**		**GF**	**VB**	
VvbHLH74	GAVBInd_009	147/155	155/155	lmxll	137/146	146/146	lmxll
	GAVBInd_010	230/233	234/230	hkxhk	231/237	231/237	hkxhk
VvHUA2	GAVBInd_001	452/455	437/456	abxcd	454/454	441/454	nnxnp
VvCOL10a	GAVBInd_004	197/195	197/196	hkxhk	195/197	195/197	hkxhk
	GAVBInd_004	197/195	197/196	efxeg	194/196	195/196	efxeg
	GAVBInd_005	146/115	155/155	lmxll	110/142	110/110	lmxll
VvWNK6	GAVBInd_019	/	/	/	210/x	217/218	abxcd
	GAVBInd_019	/	/	/	210/x	x	lmxll
VvFPA	GAVBInd_007	362/353	363/365	efxeg	363/x	x	lmxll
VvGAMYBc	GAVBInd_014	401/422	435/426	abxcd	407/428	432/437	abxcd
	GAVBInd_014	401/422	435/426	abxcd	457/x	x	lmxll
VvCOL5	GAVBInd_015	309/312	312/312	lmxll	195/198	198/198	lmxll
VvTOE3	GAVBInd_016	276/275	268/275	efxeg	279/282	272/279	efxeg
	GAVBInd_017	139/144	144/144	lmxll	136/140	140/1740	lmxll
VvPRR37b	GAVBInd_018	281/286	286/286	lmxll	281/286	286/286	lmxll
VvGAIb	GAVBInd_006	231/241	231/236	efxeg	231/245	231/237	efxeg
VvFLKa	GAVBInd_012	133/182	182/182	lmxll	128/178	178/178	lmxll
	GAVBInd_013	213/217	213/213	abxcd	211/215	211/215	hkxhk
VvFUL2	GAVBInd_020	/	/	/	443/433	414/444	abxcd
VvSVP2	GAVBInd_008	236/247	248/248	lmxll	238/244	x	lmxll

Table A12: *V. vinifera* lines with common sequences of alleles of FTC genes within the mapping population GF.GA-47-42 x 'Villard Blanc'. 'hom' means the region is homozygous. Alleles grouped together are identical. Allelic sequences common to only one varietal are not shown as well as allelic sequences with no overlap to the phased region of a gene with the mapping population GF.GA-47-42 x 'Villard Blanc'.

Gene-ID	**Lines with common alleles within a defined interval**			
VIT_201s0011g00110	'Solaris' 'Villard Blanc' (L1)	GF.GA-47-42 (E1) GF.GA-52-42		
VIT_201s0011g01560	'Chardonnay' 'Grenache Noir' 'Riesling' 'Silvaner' 'Syrah' 'Villard Blanc' (L2)			
VIT_201s0011g02120	'Grenache Noir' (hom)			
VIT_201s0011g03070	'Silvaner' 'Syrah' (hom) 'Villard Blanc' (L1)	GF.GA-47-42 (E1,E2)		
VIT_201s0011g04240	GF.GA-47-42 (E2) GF.GA-52-42 'Riesling' 'Silvaner' GF V.3125	'Chardonnay' GF.GA-52-42 'Grenache Noir' GF V.3125 'Villard Blanc' (L2)	'Solaris' 'Villard Blanc' (L1)	'Grenache Noir' 'Riesling' 'Silvaner' 'Syrah' (hom)
VIT_201s0011g05260	'Börner' (hom)	GF.GA-47-42 (N) 'Solaris' 'Villard Blanc' (N)		
VIT_201s0011g06410	'Börner' GF.GA-52-42 'Grenache Noir' 'Villard Blanc' (L1) 'Villard Blanc' (L2)	'Börner' GF.GA-47-42 (E1) GF.GA-52-42 'Grenache Noir' 'Solaris'		
VIT_201s0026g00150	GF.GA-52-42 'Riesling' 'Silvaner' 'Solaris' 'Villard Blanc' (L1)	GF.GA-47-42 (E1) GF.GA-52-42 'Riesling' 'Silvaner'		
VIT_201s0026g02200	'Börner' (hom) GF.GA-47-42 (E1,E2) 'Solaris' (hom) 'Villard Blanc' (L1,L2)			
VIT_201s0026g02580	'Chardonnay' 'Syrah' 'Villard Blanc' (L2)			
VIT_201s0010g00740	'Chardonnay' 'Riesling' 'Silvaner'	GF.GA-52-42 'Silvaner' 'Syrah'		
VIT_201s0010g02270	'Chardonnay' GF.GA-47-42 (N1) GF.GA-52-42 'Syrah' 'Villard Blanc' (N1)	GF.GA-52-42 'Riesling' (hom) GF V.3125 'Villard Blanc' (L2)		
VIT_201s0010g03890	'Chardonnay' (hom) 'Riesling'	GF.GA-47-42 (E2) 'Villard Blanc' (L2)		

Gene-ID	Lines with common alleles within a defined interval			
VIT_204s0023g02820	'Chardonnay' 'Pinot Noir précoce' GF.GA-52-42 'Riesling' 'Silvaner' (hom) 'Pinot Noir'	GF.GA-52-42 'Riesling' 'Syrah' 'Villard Blanc' (L1)	GF.GA-47-42 (N2) 'Grenache Noir' 'Villard Blanc' (N2)	
VIT_204s0044g00850	'Börner' (hom)	GF.GA-52-42 'Villard Blanc' (L1)	GF.GA-47-42 (N2) 'Grenache Noir' 'Villard Blanc' (N2)	GF.GA-47-42 (E1) GF.GA-52-42
VIT_205s0102g01160	'Silvaner' GF V.3125	'Chardonnay' 'Grenache Noir' 'Syrah'		
VIT_206s0004g03590	GF.GA-47-42 (Na,Na) 'Grenache Noir' 'Villard Blanc' (Na)	GF.GA-52-42 (hom) GF V.3125 'Villard Blanc' (L2)		
VIT_207s0005g02260	GF.GA-47-42 (E2) GF V.3125	GF.GA-47-42 (E1) 'Silvaner'		
VIT_207s0104g01590	GF.GA-47-42 (E1) 'Villard Blanc' (L1)	'Chardonnay' GF.GA-52-42	GF.GA-47-42 (E2) 'Villard Blanc' (L2)	'Chardonnay' 'Silvaner' GF V.3125
VIT_208s0056g01230	'Pinot Noir précoce' (hom) GF.GA-47-42 (Nb,Nb) 'Grenache Noir' 'Pinot Noir' 'Syrah' GF V.3125 'Villard Blanc' (Nb)			
VIT_210s0116g00750	'Chardonnay' 'Pinot Noir précoce' GF.GA-47-42 (E0,E0) 'Riesling' 'Silvaner' 'Pinot Noir' GF V.3125			
VIT_211s0052g01800	GF.GA-52-42 'Villard Blanc' (L2)	'Chardonnay' GF.GA-47-42 (E2) GF.GA-52-42		
VIT_213s0019g03550	'Pinot Noir précoce' 'Syrah'			
VIT_213s0067g03390	'Börner' (hom)	'Riesling' 'Silvaner' GF V.3125	'Chardonnay' 'Silvaner'	
VIT_214s0030g00440	'Solaris' 'Villard Blanc' (L2)			
VIT_214s0068g01800	'Villard Blanc' (L1,L2)	'Silvaner' 'Syrah'	'Börner' (hom)	'Chardonnay' 'Riesling' 'Silvaner' 'Syrah' GF V.3125
VIT_214s0083g01030	'Villard Blanc' (L1,L2)			
VIT_214s0083g01050	GF.GA-47-42 (N1) 'Villard Blanc' (N1)			
VIT_215s0048g01280	'Pinot Noir précoce' GF.GA-47-42 (E0,E0) GF.GA-52-42 'Solaris' 'Pinot Noir'			

Gene-ID	Lines with common alleles within a defined interval			
VIT_216s0013g00860	'Chardonnay' 'Pinot Noir précoce' 'Pinot Noir'	GF.GA-47-42 (N1) 'Villard Blanc' (N1)		
VIT_217s0000g00150	'Silvaner' (hom)	GF.GA-47-42 (N0,N0) 'Villard Blanc' (N0,N0)	GF.GA-52-42 (hom)	
VIT_217s0000g06570	GF.GA-47-42 (N1) GF.GA-52-42 'Grenache Noir' 'Villard Blanc' (N1)			
VIT_217s0000g08480	GF.GA-47-42 (N1) 'Villard Blanc' (N1)			
VIT_218s0001g01800	GF.GA-47-42 (E1) GF.GA-52-42	GF.GA-47-42 (E2) 'Villard Blanc' (L2)		
VIT_218s0001g07460	'Riesling' 'Syrah'			
VIT_200s0203g00080	GF.GA-47-42 (E2) 'Villard Blanc' (L2)	'Pinot Noir précoce' (hom) 'Grenache Noir' (hom) 'Riesling' (hom) 'Silvaner' (hom) 'Solaris' (hom) 'Pinot Noir' (hom) 'Syrah' (hom) GF V.3125 (hom)		
VIT_200s1675g00010	'Chardonnay' 'Riesling'	'Chardonnay' GF.GA-47-42 (E2) GF.GA-52-42 (hom) 'Riesling' 'Silvaner' (hom) 'Solaris' 'Syrah' GF V.3125 (hom)	GF.GA-47-42 (N1) 'Syrah' 'Villard Blanc' (N1)	

Table A13: Amount of total, mapped, and unmapped, and multimapped reads of RNA-Seq samples.

Sample	Total	Mapped	Multimapped	Unmapped
GF.GA-47-42-bud-02082013	9,817,926	7,663,916	269,427	1,884,583
GF.GA-47-42-bud-08082013	10,154,673	7,780,149	250,110	2,124,414
GF.GA-47-42-bud-16082013	12,903,264	10,097,057	325,530	2,480,677
GF.GA-47-42-bud-22082013	11,445,397	8,860,242	297,303	2,287,852
GF.GA-47-42-bud-05092013	10,455,865	8,156,042	258,961	2,040,862
GF.GA-47-42-bud-20122012	11,249,516	8,321,523	290,357	2,637,636
GF.GA-47-42-bud-08032013	13,118,051	9,021,214	275,422	3,821,415
GF.GA-47-42-bud-22032013	14,532,486	10,470,755	306,836	3,754,895
GF.GA-47-42-bud-12042013	12,643,312	9,508,844	284,421	2,850,047
GF.GA-47-42-bud-26042013	13,497,487	9,957,238	277,411	3,262,838
GF.GA-47-42-bud-03052013	11,422,437	8,781,858	240,942	2,399,637
GF.GA-47-42-inflorescence-07062013	10,293,064	7,428,752	211,778	2,652,534
GF.GA-47-42-inflorescence-14062013	9,514,173	6,884,231	199,387	2,430,555
GF.GA-47-42-inflorescence-17062013	5,107,723	3,572,500	110,235	1,424,988
GF.GA-47-42-leaf-07062013	6,046,705	3,252,743	88,205	2,705,757
GF.GA-47-42-leaf-14062013	10,367,630	7,442,666	218,553	2,706,411
GF.GA-47-42-leaf-22072013	9,805,519	6,589,470	195,467	3,020,582
GF.GA-47-42-leaf-08082013	10,000,060	6,850,638	208,159	2,941,263
GF.GA-47-42-leaf-22082013	9,167,887	4,586,507	117,489	4,463,891
GF.GA-47-42-leaf-19092013	9,765,702	6,823,185	211,268	2,731,249
Villard Blanc-bud-02082013	10,281,338	7,532,393	254,871	2,494,074
Villard Blanc-bud-08082013	8,948,376	6,564,385	201,525	2,182,466
Villard Blanc-bud-16082013	11,544,052	8,365,329	278,615	2,900,108
Villard Blanc-bud-22082013	12,203,608	9,238,683	290,164	2,674,761
Villard Blanc-bud-05092013	10,696,745	7,854,277	254,600	2,587,868
Villard Blanc-bud-20122012	13,429,955	9,674,812	285,554	3,469,589
Villard Blanc-bud-08032013	12,275,204	8,826,728	274,514	3,173,962
Villard Blanc-bud-22032013	13,542,256	9,669,376	278,116	3,594,764
Villard Blanc-bud-12042013	13,891,094	10,418,271	314,680	3,158,143
Villard Blanc-bud-26042013	12,597,785	9,538,519	272,672	2,786,594
Villard Blanc-bud-03052013	12,375,675	9,108,003	266,281	3,001,391
Villard Blanc-inflorescence-07062013	10,613,490	7,267,770	217,860	3,127,860
Villard Blanc-inflorescence-14062013	10,248,371	6,922,899	192,193	3,133,279
Villard Blanc-inflorescence-17062013	8,496,994	5,507,873	158,593	2,830,528
Villard Blanc-leaf-08082013	10,023,523	6,601,176	205,489	3,216,858
Villard Blanc-leaf-14062013	9,056,078	6,088,334	187,478	2,780,266
Villard Blanc-leaf-17062013	9,599,442	6,503,760	198,905	2,896,777
Villard Blanc-leaf-19092013	9,848,068	6,924,549	219,224	2,704,295
Villard Blanc-leaf-22072013	9,805,374	6,009,056	189,549	3,606,769
Villard Blanc-leaf-22082013	10,710,929	6,724,437	192,087	3,794,405

Table A14: Amount of expressed genes in each of the analyzed RNA-Seq samples. Total number of genes: 31,846.

Tissue	**Time point**	**GF.GA-47-42**	**'Villard Blanc'**
bud	02.08.2013	22,620	22,366
bud	08.08.2013	22,878	22,140
bud	16.08.2013	22,918	22,695
bud	22.08.2013	22,832	22,344
bud	05.09.2013	22,688	22,118
bud	20.12.2012	20,977	20,740
bud	08.03.2013	21,842	21,346
bud	22.03.2013	22,062	21,232
bud	12.04.2013	22,254	22,255
bud	26.04.2013	22,280	22,205
bud	03.05.2013	22,293	22,061
inflorescence	07.06.2013	22,640	22,340
inflorescence	14.06.2013	22,267	22,206
inflorescence	17.06.2013	21,189	21,851
leaf	22.07.2013	20,305	20,477
leaf	08.08.2013	20,611	20,788
leaf	22.08.2013	19,969	20,510
leaf	19.09.2013	20,637	20,650

Table A15: Genes differentially expressed at the shift to visible inflorescence structure in both GF.GA-47-42 and 'Villard Blanc'.

Gene-ID	Gene name	Chr	Start	End
VIT_201s0011g00130	*VvSPL6a*	chr1	230,152	236,567
VIT_201s0011g03070	*VvRAV1b*	chr1	2,751,566	2,753,036
VIT_201s0011g05810	*VvADOd*	chr1	5,580,698	5,585,530
VIT_201s0011g05960	*VvTPS6*	chr1	5,722,102	5,727,217
VIT_202s0025g03560	*VvJMJ5*	chr2	3,074,289	3,119,467
VIT_203s0038g00050	*VvBELe*	chr3	40,891	48,218
VIT_203s0038g00160	*VvEBS*	chr3	119,955	125,526
VIT_204s0008g00660	*VvELF3*	chr4	595,113	602,256
VIT_204s0008g07340	*VvCOL4*	chr4	7,669,233	7,671,703
VIT_204s0043g00760	*VvSECb*	chr4	14,943,616	14,962,525
VIT_206s0004g02580	*VvBELi*	chr6	3,097,994	3,107,810
VIT_206s0004g05120	*VvLUX*	chr6	6,067,481	6,071,559
VIT_206s0004g07230	*VvUGT87A1c*	chr6	7,983,729	7,986,100
VIT_207s0031g00320	*VvCLFa*	chr7	16,560,116	16,572,925
VIT_207s0031g01320	*VvTGA1*	chr7	17,404,970	17,410,595
VIT_208s0007g01290	*VvBELj*	chr8	15,381,600	15,387,977
VIT_208s0007g03880	*VvPCFS4e*	chr8	17,859,611	17,869,331
VIT_208s0007g06370	*VvPIE1a*	chr8	20,145,531	20,170,881
VIT_208s0007g06520	*VvACTe*	chr8	20,249,662	20,252,713
VIT_208s0007g08250	*VvNF-YA1c*	chr8	21,596,967	21,603,488
VIT_209s0002g04340	*VvGL2*	chr9	3,967,729	3,972,081
VIT_210s0003g00050	*VvSPB1*	chr10	1,346,191	1,348,771
VIT_210s0003g03710	*VvCUL4*	chr10	6,283,790	6,322,024
VIT_210s0116g00160	*VvBAHCC*	chr10_random	77,087	83,420
VIT_211s0052g00730	*VvADOb*	chr11	18,287,912	18,290,912
VIT_212s0028g03350	*VvSPB1*	chr12	4,108,169	4,110,257
VIT_212s0059g01420	*VvCOP1a*	chr12	6,341,424	6,352,095
VIT_213s0067g01020	*VvF3Ha*	chr13	584,205	586,569
VIT_214s0030g00440	*VvGID1Ba*	chr14	4,293,572	4,298,380
VIT_215s0021g02140	*VvBREb*	chr15	12,988,877	13,025,618
VIT_215s0021g02140	*VvBREc*	chr15	12,988,877	13,025,618
VIT_215s0048g02540	*VvPRR95*	chr15	16,686,112	16,692,048
VIT_216s0098g00900	*VvAPRP5*	chr16	21,231,035	21,238,345
VIT_217s0000g00430	*VvbHLH137*	chr17	299,560	302,242
VIT_217s0053g00780	*VvHDG11*	chr17	16,226,906	16,231,203
VIT_219s0014g02350	*VvSPL3a*	chr19	2,484,255	2,485,739

Additional data:

The following data is accessible under the link: https://docs.cebitec.uni-bielefeld.de/index.php/s/kQLUhp5HAt8j4Qa:

Files:

1. ftc_candidate_genes.xls: detailed list of FTC candidate genes
2. amplicon_sequenced_genes.xls: detailed list of amplicon sequenced genes
3. amplified_amplicons.xls: amplicon sequenced genes and cultivars for which amplification was performed
4. amplimers_and_primer_with_position.xls: positions of amplimers for amplicon sequencing and primer for amplification
5. analyzed_cultivars_and_phenotypes.xls:
   - sheet 1: F1-individuals of the mapping population GF.GA-47-42 x 'Villard Blanc' and their flowering time phenotypes in days after January 1st in 1999 and 2009 - 2016
   - sheet 2: sequenced grapevine cultivars, related cultivars and their phenotypes of flowering time in relation to each other
6. correlation_results.xls: results of correlation analysis of amplicon sequenced genes in amplicon sequenced individuals of the population GF.GA-47-42 x 'Villard Blanc'
7. correlation_extended_population.xls: results of correlation analysis of all 151 F1-individuals of the population GF.GA-47-42 x 'Villard Blanc' based on STR-marker analysis
8. differential_expression_ftc_genes.xls: significantly differentially expressed FTC-candidate genes in both of GF.GA-47-42 and 'Villard Blanc' in buds, inflorescences, and leaves over different time courses (see sheets)
9. ftc_gene_alleles_in_grapevine_cultivars.xls: alleles of FTC-candidate genes common between different grapevine cultivars
10. inflorescence_specific_genes.xls: genes differentially expressed between buds and inflorescences shortly before full bloom

Folder:

1. allele_distribution_plots: allele distribution plots of amplicon sequenced genes for which correlation analysis was performed
2. allele_sequences: allele sequences of genomic regions of amplicon sequenced genes in which phasing was performed
3. correlation_plots: correlation heatmaps of all analyzed datasets. gr: global radiation, ts: temperature sum, days: days after January 1st

Danksagung

Meine Promotionszeit erlebte ich sowohl wissenschaftlich als auch persönlich als Bereicherung und Herausforderung. Dafür möchte ich an erster Stelle meinem Doktorvater, Prof. Dr. Bernd Weisshaar, für die Möglichkeit diese Arbeit in seiner Gruppe anzufertigen, danken. Vor allem bin ich dankbar für die Möglichkeit mit meinem damals biologischen Hintergrund bioinformatisch arbeiten zu können. Auch wenn es manchmal schwierig war, habe ich mich selber, mein Wissen und meine Fähigkeiten in dieser Zeit sehr weiterentwickeln können.

Ein ganz besonders herzlicher Dank geht an Dr. Daniela Holtgräwe für die wissenschaftliche Betreuung dieser Arbeit, die konstruktiven Ratschläge und die großartige und immerwährende liebevolle Unterstützung.

Auch meinen weiteren Kollegen und ehemaligen Kollegen am Lehrstuhl für Bioinformatik und Genomforschung danke ich für die prägende Zeit. Johanna Müschner und Willy Keller danke ich herzlich für die vielen Amplikon-PCRs und Prisca Viehöver für die Sequenzierung meiner Daten. Besondere Erwähnung soll außerdem David finden, dem ich danke, für gute Zeiten neben und während des Promovierens, dafür dass er mir stets ein interessanter und angenehmer Diskussionspartner war und ganz besonders dafür, dass er mir als Freund vor allem in schwierigen Phasen hilfsbereit und motivierend zur Seite stand.

Der BRF danke ich für die Bereitstellung der Infrastruktur für die Bearbeitung meiner Daten und meinen Kollegen dort für jegliche Unterstützung.

Desweiteren danke ich Dr. Iris Ochßner, Anna Werner, Dr. Ludger Hausmann und Prof. Reinhard Töpfer vom Julius-Kühn-Institut für Rebenzüchtung in Siebeldingen für die erfolgreiche Zusammenarbeit im Rahmen unseres Projektes, ihrer Kooperation und Hilfsbereitschaft.

Natürlich danke ich auch all meinen lieben Freunden und Bekannten aus der nicht Genomforschungswelt. Erwähnung finden sollen hier Anne, die mich daran erinnert, dass man mit Wein auch anderes machen kann als dessen Gene zu erforschen, Daniel, der immer hilfsbereit ist und ganz besonders Lea, die mich mit viel Verständnis sehr unterstützt und motiviert hat, geduldig und einfühlsam ist und immer ein

offenes Ohr für mich hat.

Meiner Familie danke ich für ihre Unterstützung und ihr Verständnis. Mein größter Dank geht schließlich an meine Eltern, die stets viel Wert auf Bildung gelegt und mir die Anfänge meines Ausbildungsweges geebnet haben. Meiner Mutter bin ich für ihre maßlose Unterstützung zutiefst dankbar.

Selbstständigkeitserklärung

Ich, Nadia Kamal, erkläre hiermit folgendes:

1. Die Promotionsordnung der Fakultät für Biologie vom 1. Juni 2012 ist mir bekannt.

2. Die, hier vorligende, Dissertationsarbeit habe ich selbstständig angefertigt. Ich habe keine Textabschnitte von Dritten oder aus eigenen Prüfungsarbeiten ohne Kennzeichnung übernommen. Alle von mir genutzten Hilfsmittel und Quellen habe ich angegeben.

3. Ich habe weder unmittelbar noch mittelbar geldwerte Leistungen für Vermittlungstätigkeiten oder für Arbeiten, die im Zusammenhang mit dem Inhalt der vorliegenden Dissertation stehen, an Dritte gezahlt.

4. Ich habe die vorliegende Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht hat.

5. Ich habe weder eine der vorliegende Dissertation gleiche, oder eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht.

Bielefeld, October 19, 2017