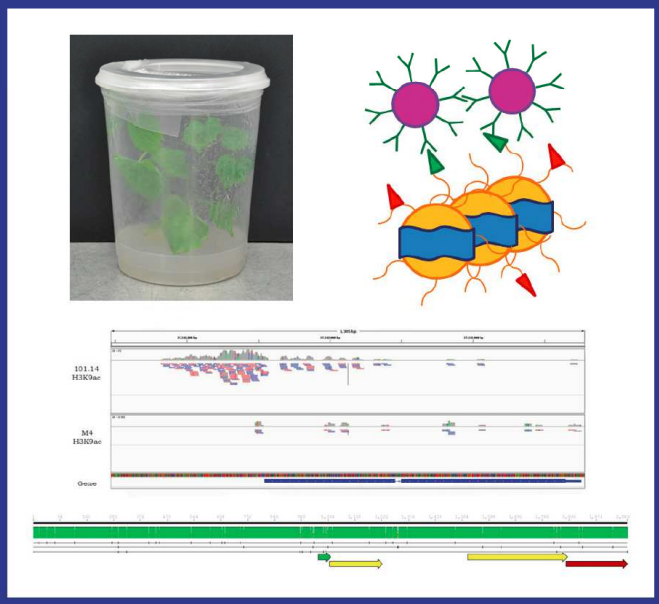
Università degli Studi di Padova

Michela Verna

An epigenetic approach to study gene regulation in two different grapevine rootstocks: from histone modifications to gene expression



Corso di Dottorato in Scienze delle Produzioni Vegetali

11/2014

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Agronomia Animali Alimenti Risorse Naturali e Ambiente
- DAFNAE -

CORSO DI DOTTORATO IN SCIENZE DELLE PRODUZIONI VEGETALI
CICLO XXVII

An epigenetic approach to study gene regulation in two different grapevine rootstocks: from histone modifications to gene expression

Coordinatore: Ch.mo Prof. Antonio Berti

Supervisore: Ch.mo Prof. Margherita Lucchin

Dottorando: Michela Verna

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Michela Verna February,2nd 2015

A copy of the thesis will be available at *http://paduaresearch.cab.unipd.it/*

Dichiarazione

Con la presente affermo che questa tesi è frutto del mio lavoro e che, per quanto io ne sia a conoscenza, non contiene materiale precedentemente pubblicato o scritto da un'altra persona né materiale che è stato utilizzato per l'ottenimento di qualunque altro titolo o diploma dell'università o altro istituto di apprendimento, a eccezione del caso in cui ciò venga riconosciuto nel testo.

Michela Verna 2 Febbraio 2015

Una copia della tesi sarà disponibile presso *http://paduaresearch.cab.unipd.it/*

Index

Chapter III: Chromatin Immunoprecipitation (ChIP) protocol for grapevine leaf material

Chapter IV: Correlation between histone modifications and gene expression in two grapevine rootstocks through NGS technologies: integration of ChIP Sequencing and mRNA Sequencing data

Summary

In the nineteenth century European viticulture was devastated by the introduction of phylloxera (*Daktuloshaira vitifoliae*) from North America. Since this moment a new era for viticulture started out based on grafting of a scion of *V. vinifera* varieties onto rootstocks from the pest's origin. The introduction of root system from American or non *vinifera* species was initially focused on the protection of viticulture from the pest but very soon it has been realized that grapevine rootstocks were not only capable to confer disease resistance, but they could imply a large range of advantages by influencing numerous physiological process at the scion level such as response to abiotic stresses. At the light of the climate changes that are affecting the earth, the viticulture, as well as the agriculture, need to develop plants able to cope with situations of long drought period. DiSAA research group of University of Milan establish till 1985 novel candidate grapevine genotypes that could be used as rootstocks with the objective to obtain new hybrids more performant in response to abiotic stress. In a preliminary screening, one of these, called M4 [(*V. vinifera* x *V. berlandieri*) x *V. berlandieri* x cv Resseguier n.1], was selected for its relatively high tolerance to water deficit and salt exposure. Biochemical and physiological studies were performed within Ager-Serres project 2010-2105 for the comparison between rootstock M4 and a commercial one susceptible to drought stress, 101.14 (*V. riparia x V. rupestris*). Furthermore the project used Next Generation Sequencing (NGS) techniques to investigate the genomes and transcriptome profiles of the two rootstocks grown in drought and in controlled conditions. The aim of this multidisciplinary project was the identification of marker genes for the selection of rootstocks with high performances under water stress conditions. It is within the Ager-Serres project that this PhD research project takes place with the objective to study the relation between transcriptional profile and histone modifications in the rootstocks 101.14 and M4. Our research represent the first study aimed to clarify the relation between histone modifications and gene expression in the genus *Vitis*. Literature data from other species showed, indeed, that chromatin remodeling, histone modifications and chromatin related

processes act on chromatin structure in a dynamic manner for the regulation of gene expression. Nucleosomes are the building block of chromatin that can be defined like a highly condensed structured that form the scaffold of fundamental nuclear processes such as transcription, replication and DNA repair. Nowadays it is clear in fact that the function of chromatin doesn't end with the packaging and the protection of DNA but it also act controlling gene expression. Cells have evolved mechanisms that alter the structure of chromatin allowing the access of transcriptional complexes.

To reach our objective a pool of plants of each genotype were grown *in vitro*, leaf materials were collected and gene expression was evaluated with mRNA-Sequencing (mRNA-seq) while distribution of acetylation of lysine 9 of histone three (H3K9ac) and trimethylation of lysine 4 of histone three (H3K4me3) were evaluated through a ChIP-Sequencing (ChIP-seq) approach. Nowadays mRNA-seq has become a routine analysis for the evaluation of transcriptome profile of a large number of plant species while the use of ChIP-seq is still limited to plant model species. During this PhD a lot of work was done for the development of a chromatin immunoprecipitation protocol suitable for the sequencing of chromatin extracted and immunoprecipitated from grapevine.

Literature data for other plant species established that H3K9ac and H3K4me3 histone modifications were positive related to gene expression and that these marks preferentially lied in the transcription starting site (TSS) of the genes but the identification of this feature is still missing in grapevine genome annotation. For this reason we decided to perform the enrichment analyses on immunoprecipitated chromatin keeping, in an arbitrary manner, a window of 1000 bp above the start codon (ATG) of the genes and considering for each gene the number of reads that aligned in this region. The reads upstream of each gene were normalized and the genes were studied in terms of enrichment or depletion respect their sequencing profile without immunoprecipitation. The groups of genes enriched in one genotype and depleted in the other were related to transcriptional profile discovering a positive correlation between histone mark H3K9ac and gene expression. This trend was not confirmed analyzing the mark of H3K4me3.

Considering, however, the genes with enrichment in one genotype and depletion in the other for both the histone marks at the same time we discovered a positive correlation with expression profile.

The results of this study are the first evidences of a positive correlation between histone modifications H3K9ac and H3K4me3, when present simultaneously with H3K9ac, and gene expression in grapevine.

Riassunto

Durante il diciannovesimo secolo la viticoltura europea fu devastata dall'introduzione di un insetto fitofago di origine nord americana la *Daktuloshaira vitifoliae* nota comunemente con il nome di "Phylloxera della vite". Da quel momento iniziò una nuova era per la viticoltura con l'utilizzo dell'innesto di varietà di *V. vinifera* in portinnesti provenienti dalle zone di origine del parassita. L'introduzione di sistemi radicali di specie americane e non *vinifera* si focalizzò inizialmente sugli aspetti della protezione della viticoltura dall'attacco del parassita. Molto presto, tuttavia, si capì che l'utilizzo del portinnesto non solo conferiva resistenza alla malattia ma implicava anche una serie di vantaggi influenzando numerosi processi fisiologici a livello del nesto, come ad esempio la resistenza agli stress abiotici. Alla luce dei cambiamenti climatici che stanno coinvolgendo il nostro pianeta, la viticoltura, così come l'intera agricoltura, necessita di sviluppare piante in grado di gestire situazioni di siccità prolungata. Il gruppo di ricerca del DiSAA dell'Università di Milano dal 1985 è coinvolto nello studio nuovi genotipi di vite con l'obbiettivo di ottenere nuovi ibridi utilizzabili come portinnesti e maggiormente performanti in situazioni di stress abiotico. In uno screening preliminare, uno di questi, denominato M4 [(*V. vinifera* x *V. berlandieri*) x *V. berlandieri* x cv Resseguier n.1], fu selezionato per la resistenza elevata allo stress idrico e salino. All'interno del progetto Ager-Serres 2010-2105 sono stati eseguiti degli studi biochimici e fisiologici comparando il genotipo sperimentale M4 con un portinnesto commerciale suscettibile allo stress idrico, 101.14 (*V. riparia x V. rupestris*). Questi dati sono stati integrati ad approcci di Next Generation Sequencing (NGS) volti a studiare i genomi ed i profili trascrizionali dei due portinnesti cresciuti in condizioni di stress idrico e di controllo. L'obiettivo principale del progetto Ager-Serres, risiede nell'identificazione di geni marcatori che possano essere utilizzati per la selezione di nuovi portinnesti con maggiori performance in situazioni di stress idrico.

Questo progetto di dottorato si inserisce all'interno del progetto Ager-Serres ed è volto a studiare la relazione tra profili trascrizionali e modifiche istoniche nei portinnesti 101.14 e M4. Dati di letteratura relativi ad altre specie hanno dimostrato infatti che il rimodellamento della

cromatina, le modifiche istoniche ed i processi correlati alla cromatina, agiscono in maniera dinamica sulla struttura della cromatina stessa per la regolazione dell'espressione genica. I nucleosomi sono le unità fondamentali della cromatina che può essere definita come una struttura altamente condensata che costituisce la base per processi nucleari fondamentali come trascrizione, replicazione e riparazione del DNA. Oggigiorno è chiaro appunto che la funzione della cromatina non si esaurisce nell'impacchettamento del DNA e nella sua protezione ma risiede anche nel controllo dell'espressione genica. Gli organismi hanno infatti evoluto dei meccanismi atti a alterare la struttura della cromatina consentendo l'accesso o l'esclusione di complessi trascrizionali.

Per condurre questa ricerca, un gruppo di piante di ciascun genotipo (101.14 e M4) è stato fatto crescere *in vitro*. Il materiale fogliare è stato campionato e l'espressione genica studiata attraverso un mRNA-Sequencing (mRNA-seq). La distribuzione di modifiche istoniche, come l'acetilazione della lisina 9 dell'istone H3 (H3K9ac) e la trimetilazione della lisina 4 dello stesso istone (H3K4me3), sono state valutate attraverso un approccio di ChIP-Sequencing (ChIP-seq). Questa ricerca rappresenta il primo studio di questo tipo condotto nel genere *Vitis*.

Oggigiorno i sequenziamenti di mRNA sono divenuti analisi di routine per la valutazione dei profili trascrizionali in un gran numero di specie vegetali mentre l'utilizzo di approcci di ChIP-seq rimane limitato alle piante modello. Durante questo progetto di dottorato molto lavoro è stato investito nello sviluppo di un protocollo di immunoprecipitazione che fosse performante per l'estrazione ed il sequenziamento della cromatina di materiale fogliare di piante di vite.

Dati di letteratura riguardanti altre specie vegetali hanno individuato una correlazione positiva tra le modifiche istoniche H3K9ac ed H3K4me3 e l'espressione genica. Si è inoltre stabilito che questi marchi giacciono preferenzialmente nella regione del sito di inizio della trascrizione (TSS) dei geni ma l'identificazione di questi siti è ancora mancante nell'annotazione del genoma di vite. Per questo motivo si è deciso di effettuare analisi di arricchimento sulla cromatina immunoprecipitata scegliendo in maniera arbitraria un intervallo di 1000 paia di basi a monte del codone d'inizio (ATG) dei geni e considerando, per ogni gene, il numero di reads che allineano in

questo intervallo. Le reads mappanti nella regione considerata sono state normalizzate ed i geni studiati in termini di arricchimento o deplezione della modifica rispetto alla cromatina non immunoprecipitata. I gruppi di geni arricchiti in un genotipo e depleti nell'altro, sono stati posti in relazione con i profili trascrizionali individuando così una correlazione positiva tra la modifica istonica H3K9ac e l'espressione genica. Questo andamento non è stato riscontrato analizzando la trimetilazione della lisina 4 dell'istone H3. Considerando, tuttavia, i geni che presentano un arricchimento in un genotipo e una deplezione nell'altro, contemporaneamente per entrambi i marchi istonici, è stata identificata una correlazione positiva con i profili trascrizionali.

I risultati di questo studio sono le prime evidenze di una correlazione positiva tra modifiche istoniche H3K9ac e H3K4me3, quando presente contemporaneamente ad H3K9ac ed espressione genica in vite.

Chapter I

General Introduction

1 Grapevine Rootstocks

1.1 Grapevine and viticulture

Grapes and their derivate have a large expanding worldwide market. In 2012 FAO estimates that grapes production was the first product for income in Italy and that our country was the third in the world agriculture incoming (http://faostat.fao.org/site/339/default.aspx).

Grapes can be grown at a very large range of latitudes, from 50°N to 40°S, and up to 3000 meters above the sea level. Furthermore the wine production could boast ancient origins, since 5000 years BC wine become an important component of many cultures (McGovern, 2003).

The genus *Vitis* is represented by several coexisting species in Europe. *Vitis vinifera* L. ssp. silvestris (Gmelin) Hegi is the only extant wild European taxon but many other grapevine genotype were also naturalized in Europe. They belong to *V. vinifera* L. subsp. *vinifera*, introduced for at least a thousand years when domesticated forms of grapevine were spread throughout Europe (Olmo, 1995). Furthermore during last century several American and Asian *Vitis* species have been introduced in Europe as rootstock taking more and more importance because of the use of them, together with other their hybrids, like bases for grafting. This practice took hold in Europe by the mid-nineteenth century because of the devastating effect of phylloxera (*Daktulosphaira vitifoliae*), a soilborne aphid pest, introduced into Europe from America. For this reason cultivated grapevines are the merge of two genotypes, one of the scion and one of the rootstock, which interact together to create the final phenotype (Marguerit *et al*., 2012). Improvement of genetic material of both grapevines and rootstocks is essential in the context of climate change events occurring in recent years and could be a crucial factor for the development of sustainable agricultural models (moderate irrigation and fertilization). Since 1982, in fact, Boyer theorized that water availability could be one of the major limiting factors for plant productivity (Boyer, 1982).

1.2 Rootstocks selection

Grapevine has thousands of varieties but only few cultivar are employed for wine production. Rootstocks used in Europe exhibit a relatively narrow genetic background due to the fact that their selection was essentially based on only few phenotypic traits like rooting ability, phylloxera resistance and scion-induced vigor.

In the second half of the IX century the Europe viticulture began to be affected by phylloxera attacks. In 1869 it was reported for the first time, by Léo Laliman, that some American grapevines were not susceptible to the attacks of phylloxera. Since that moment started the challenge for the selection of resistant genotypes that could be used like base for grafting of local varieties for the reconstructions of vineyards destroyed by the aphid. Although the simplicity of the idea of grafting the actuation of this practice wasn't rapid and straightforward because of different difficulties; first of all the different degrees of resistance of the American species to the phylloxera attacks. The first rootstocks imported in Europe showed scarce resistance to that pest and a good degree of resistance has been found in *Vitis riparia* and *Vitis rupestris* but these species were not able to cope with soils riches of limestone developing chlorosis pathologies more or less severe. To cope with this problem were developed breeding programs between American species and between American spp. and *V. vinifera*. This late option was chosen with the idea to develop a hybrid for the direct production keeping the positive characteristics of each species: resistance to phylloxera (American spp.) and good level of product quality (*V. vinifera*). This idea has remained a mirage and nowadays the viticulture is based on grafting of a *V. vinifera* sp. into a hybrid rootstock.

The choice of the best performance rootstock is very complex and it depends of different aspects especially linked to the composition of the soil. The rootstock has to be suitable to the soil chosen for the establishment of the vineyard and sometimes American species are more demanding respect of the indigenous *V. vinifera*. For instance the most part of American grapevine rootstocks, are not able to cope with alkaline and reach in limestone soils. In these conditions the plants develop more or less severe forms of iron chlorosis. American *Vitis* have

also other problems that regard the soil morphology, in fact they are less tolerable to soils excessively compacted respect of the European species. Moreover the American grapevine are, on one hand, sensible to soils permanently humid and, on the other hand, sensible to the drought of the substrate. Taking a look at the nutrition level it has been observed that some rootstocks show low absorption of some ions like K, Ca and Mg. Another aspect that have to be evaluated for the choice of the rootstocks is the destination of the production that is represented by the fact that the rootstocks can enhance or contain the characteristics of the *vinifera* graft production.

Certainty the main important American rootstocks are *V. riparia, V. rupestris* and *V. berlandieri* and their hybrid are the most widespread in viticulture. A brief description of each genotype characteristics are reported below.

- *V. riparia* was one of the first species used for the grafting; it shows high resistance to philloxera attacks, good affinity to the graft and it can be easily propagated by scion. There are several genotypes of this species and the most used for the vineyard reconstruction after phylloxera attack were *riparia* "Glorie de Montpellier" and *riparia* "Grand glabre". In conditions of fertile soils *V. riparia* can be considered a good rootstock but its performances in drought conditions are very limited;
- *V. rupestris* like *V. riparia* could be easily propagated by scion and it has good resistance to phylloxera and to graft. The most diffused *V. rupestris* is the selection *riparia* "du Lot".
- *V. berlandieri* shows high resistance to phylloxera, great resistance in soils alkaline and rich of limestone but *berlandieri* is not considered a good rootstock as pure species because of its great difficulty to be propagate by scion.

All these species show pros and cons, for this reason the breeders started to hybridize together plants of the three different species to obtain hybrids that are today the most used rootstocks in viticulture.

Breeders developed the idea of the creation of new plants direct producers by hybridizing American species with European *vinifera* but this remained a missed goal. The quality of the

euro-american hybrids fruits were very low and, sometimes, the rootstocks showed susceptibility to phylloxera attacks. Anyway in this group there are few members that worth mentioning for their employment like rootstocks: 196.17, Gravesac and 41B. The 196.17 has high vigor and low resistance to limestone; it is well suited to acid, dry and sandy soils. The Gravesac shows good vigor, low resistance to limestone and good adaptability to acid soils. 41B has been considered for a long time like the rootstock most resistant to chlorosis. Its resistance to phylloxera attacks is good although it shows limited capacity to scion propagation. (Dalmasso and Eynard, 1990).

The main commercialized hybrids rootstocks are reported in Table 1.1.

	Rootstocks	Vigor induced	Affinity with *V. vinifera* grafting
V. berlandieri x V. riparia	Kober 5 BB	++/+++	++++
	SO 4	++/+++	++
	420A	-/+	+/++
	161.49	-/+	-
	5C	+++	++
V. berlandieri x V. riparia	1103 Paulsen	+++	++++
	110 Richter	+++	+
	140 Ruggeri	++++	-/+
	775 Paulsen	+++	+
	779 Paulsen	++++	-/+
	Rupestris Du Lot	++/+++	++
V. berlandieri x V. riparia	101.14	-	++
	3309C	-	++
	Schwarzmann	-/+	+++
Euro-American hybrids	196.17	+++	++
	Gravesac	++/+++	++
	41 B	++/+++	--/-
Legend: very low --; low -; sufficient +; good ++; high +++; very high ++++			

Table 1.1: Synoptic table of characteristics of the main rootstocks used in viticulture (Data from Vivai Coperativi Rauscedo, VCR, http://www.vivairauscedo.com/portinnesti)

1.3 Interaction between rootstocks and scion

In addition to their ability to help the grafts to cope with biotic stresses, rootstocks can also confer tolerance to a large number of abiotic stress like drought and salinity that could have a large impact on the crop production. (Cramer *et al.,* 2007). For this reason the breeding of rootstocks that could use water more efficiently has become a key strategy for the improvement of agro systems (Marguerit *et al*., 2012). In 2013 Dai and coworker developed a global climate

model that predicts an increase of the aridity in the next future (Dai, 2013). At the light of climate changes the selection of new rootstocks that could use water in a more efficiently manner providing a better growth capacity and scion adaptation to stress may play a crucial role in the viticulture.

Each commercial rootstock shows different performances of tolerance to water stress, for instance 101.14 and Schwarzmann are considered low tolerant while Lider 116-60, 1103 Paulsen , Ramsey, 140 Ruggeri, Richter 110 and Kober 5BB are known to confer to scion higher drought tolerance (Flexas *et al*., 2009).

The ability of these rootstocks to confer tolerance to drought depends of different factors and the alteration of scion vigor was linked to differences in hydraulic parameters of the root system. In 2012 Gambetta and colleagues (Gambetta *et al*., 2012) hypnotized a key role of aquaporin proteins in relation to grapevine rootstocks vigor and in the management of the water used in drought conditions. Root system hydraulic capacity on delivering water to scion is related to the increase in root hydraulic conductivity (Lpr) and on whole-root-system surface area. In 2011 Alsina and coworkers (Alsina *et al*., 2011) discovered that grapevine grafted on a high vigor rootstock (1103P) had greater whole-root-system hydraulic conductance compared to the ones grafted on low vigor rootstock (101.14).

Stomata have a key role in the regulation of water loss during drought (Marguerit *et al*., 2012). and the stoma closure is, in fact, one of earliness responses to water deficit (Damour *et al*., 2010). At the light of these studies it became evident that a grapevine rootstock that increased the efficiency of stoma closure control and of water use efficiency had major tolerance to water stress.

1.4 Ager-Serres 2010-2105 project

The introduction of root system of American or non *vinifera* species was, as said before, initially focused on the protection of viticulture from the pest but very soon it has been realized that grapevine rootstocks were not only capable to confer disease resistance, but imply a large range of advantages by altering numerous physiological process at the scion level, such as biomass accumulation (Gregory *et al*., 2013), fruit quality (Walker *et al*., 2002 and 2004) and nevertheless ability to respond to many abiotic stresses (Marguerit *et al*., 2012). In this last field rootstocks were identified like carriers of many features, such as tolerance to high salinity (Fisarakis *et al.*, 2001), ion deficit (Covarrubias and Rombolà, 2013) and drought (Gambetta *et al.*, 2012). At the light of the continuous climate changes, the viticulture, as well as the agriculture, needs to develop plants able to cope with situations of long drought period.

DiSAA research group of University of Milan establish till 1985 novel candidate grapevine genotypes that could be used as rootstocks with the objective to obtain new hybrids more performant in response to abiotic stresses. In a preliminary screening, one of these, called M4 [(*V. vinifera* x *V. berlandieri*) x *V. berlandieri* x cv Resseguier n.1] was selected for its relatively high tolerance to water deficit and salt exposure. Rootstock M4 was the main characters of the project Ager-Serres 2010-2105 (Scienza *et al.*, 2013; http://users.unimi.it/serres/index.html). Ager-Serres was a multidisciplinary project that on one hand focused the attention on biochemical and physiological studies of a commercial rootstock susceptible to drought stress, 101.14 (*V. riparia x V. rupestris*), compared to the experimental one, M4 grown under controlled and drought conditions (Meggio *et al*., 2014). While on the other hand the project used Next Generation Sequencing (NGS) techniques to go in deep with genome and transcriptome analyses (Corso *et al*., submitted). The main goal of this project was to clarify the molecular and physiological bases of rootstock response to abiotic stresses and the identification of tolerance genes that could be used like marker for the screening and the selection of rootstocks with high performances in stress conditions. It is within the Ager-Serres project that this research takes place with the objective to point out the contribution of histone modifications to gene

expression and to establish the relation between transcriptional profile and histone modifications in the two rootstocks, M4 and 101.14, grown in controlled condition.

1.5 Grape genome

Arabidopsis thaliana was the first plant species whose genome was completely sequenced (The Arabidopsis Genome Initiative, 2000). Since 2000 the development of new sequencing technology allowed the publication of about 50 genome of plant species and about 135 are the genomes overall under whole genome investigation (http://www.ncbi.nlm.nih.gov/genome/browse/). The publication of these genomes marked a great improvement in the knowledge of plant biology helping to defining the function of a large number of genes. Grapevine was the fourth genome sequenced among of the flowering plants and the second, after *Polulus*, among wood plants (Michael and Jackson, 2013).

Grape genome was published in 2007 by two independent groups: the French-Italian Public Consortium For Grapevine Genome Characterization (Jaillon *et al*., 2007) and the group of the IASMA Research Center headed by Velasco (Velasco *et al*., 2007). The French - Italian consortium sequenced the PN40024 grape genome that originally derived from Pinot Noir that has been bred close to fully homozygosity (estimated about 93%) by successive selfings. This choice was made in order to by-pass the high heterozygosity that characterize grapevine, permitting a high quality whole genome shotgun assembly. The sequencing of PN40024 was performed with Sanger methodologies. On the other hand the group of IASMA Research Center utilized Sanger sequencing together with sequencing by synthesis (SBS) methods to perform the sequencing of heterozygous clone ENTAV115 on Pinot Noir cultivar. French-Italian consortium identified 30 434 genes for PN40024 and a genome size of 487 Mb. On the other hand genome size of ENTAV115 clone was slightly larger, 505 MB, with a prediction of 29 585 genes. More detailed information of the two grapevine genomes are reported in Table 1.2.

	Size [Mb]	Assembled [Mb]	Assembled [%]	Gene #	Repeat[%]	Sequencer Types
Jaillon *et al* .,2007	475	487	103	30434	41	Sanger
Velasco *et al* .,2007	505	477	95	29585	27	Sanger, Roche 454

Table 1.2: Grape published genomes.

Nowadays the genome sequence available online is the 12X coverage assembly of PN40024 where the genes annotations (V1 and V2) were performed by CRIBI, University of Padova (http://genomes.cribi.unipd.it/grape/).

The advent of high throughput sequencing technologies allowed to the comparison of genomic sequences related to different cultivar of *Vitis vinifera* species. These studies revealed the presence of a core genome containing genes that are present in all strains and a dispensable genome composed of partially shared and strain-specific DNA elements. This new concept was defined in 2007 by Morgante and coworkers like “pangenome” (Morgante *et al.*, 2007).

2 DNA and Chromatin

2.1 DNA packaging

Life depends on the ability of the cells in conservation, extraction and transduction of the genetic instructions needed for create and maintain a living organism. These instructions are contained in each leaving cell as genes that are the elements that determine the characters of a species and of the individuals who compose it. Starting from the 40s the researches understand that genetic information consists primarily of the instruction for the production of protein and that genetic information resides in deoxyribonucleic acid (DNA). The mechanism by which the information was hereditable remained unknown till 1953 when the structure of DNA was determined by James Watson and Francis Crick.

The structure of deoxyribonucleic acid has two helical chains each coiled around the same axis (Watson and Crick, 1953). Each chain consist of nucleotides, that are the key element of nucleic acid. A nucleotide is composed of a phosphate group bound to a five-carbon sugar, 2'deoxiribose, which is linked to a nitrogenous base through a glycoside bond between that and the hydroxyl group in position 1' of the sugar. The structure sugar-base is called nucleoside while the bond of a phosphate group at position 5' of the 2'deoxyribose through a phosphodiester bond create an organization called nucleotide. Nucleotides are linked one to other by the hydroxyl group present at 3' of 2'deoxyribose of one nucleotide and the phosphate linked to carbon 5' of the subsequent nucleotide. The two chain, but not their bases, are related by a dyad perpendicular to the fiber axis. Both chains follow the same direction, usually the right handed-helices, but owing to the dyad the sequences of the atoms in the two chain run in opposite direction. The bases are on the inside of the helix and the phosphate of the outside with the sugar that is roughly perpendicular to the attached base. Watson and Crick discovered that there was a residue on each chain every 3.4 Å in the z-direction supposing an angle of 36° between adjacent residues in the same chain. In this way the structure repeats every 10 residues or, in other terms, every 10 base pair. The two chains are held together by the pairing of the

nitrogenous bases that are divided in two groups: purine and pyrimidines; adenine and guanine constitute the first class and cytosine a thymine the second one. The planes of the bases are perpendicular to the fiber axis and are joined together in pairs, a single base from one chain is hydrogen-bounded to a single base from the other chain, so that they lie side by side with the same z-co-ordinates. The fundamental principle for the base pairing is that a purine must be coupled with a pyrimidines and vice versa, especially only specific pair of bases can bond together: adenine (A) with thymine (T) and guanine (G) with cytosine (C). The pairs A-T are keep together with two hydrogen bound while C-G can setting up three of this bounds. Through these discoveries Watson and Crick can assumed that the specific pairing is the fundamental events for the copying mechanism of the genetic material.

The eukaryotic DNA lies in the nucleus where it is distributed in different chromosomes. Each chromosome consist of a single long molecule of DNA associated with proteins those fold the DNA in a more compacted structure. The structure constituted of DNA and protein is called chromatin. The chromosomes are not only associated with protein involved in packaging of DNA but also with different kind of proteins needed for the management of the processes for gene expression, replication and DNA repair.

Eukaryotic DNA is tightly packaged around a core of structural proteins, the histones, to generate the chromatin nucleosome array (Kornberg, 1974). Two copies of each histone protein H2A, H2B, H3 and H4, are assembled into an octamer that has 145-147 bp of DNA wrapped around it to form a nucleosome core. In 1997 Luger and coworker (Luger *et al*., 1997) determined for the first time a high-resolution X-Ray structure of the nucleosome core particle from crystals made in bacteria and assembled after purification. They discovered that the 147 bp of DNA were wrapped around the histone octamer in 1.65 turn of a flat, left-handed super helix, Figure 1.1 (Luger *et al*., 2003). Subsequent works have solved the structure of the nucleosome core particle at higher resolution and in 2002 by Davey *et al.* studied the structure of the nucleosome core particle with the highest resolution for the structure of any protein-DNA complex of this size: 1.9Å (Davey *et al*., 2002).

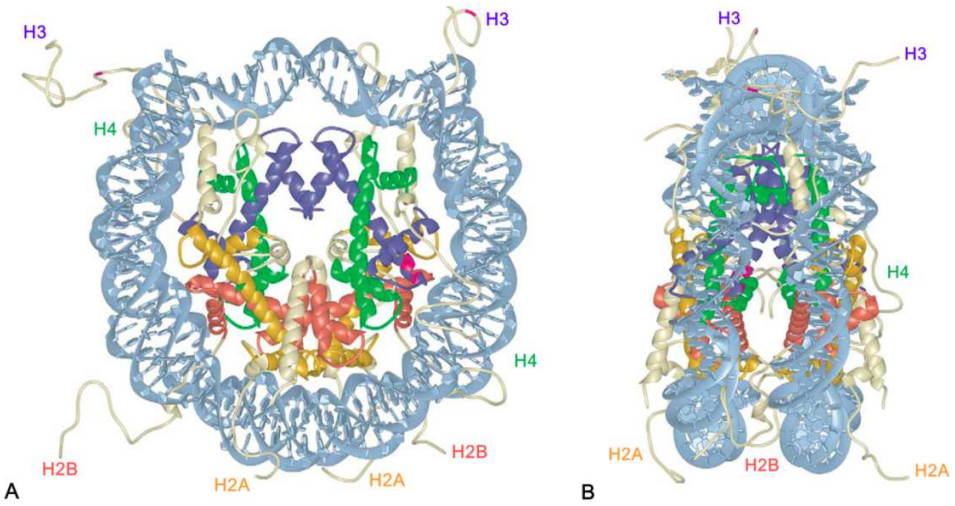


Figure 1.1.: Overall structure of the nucleosome core particle. (A) Front view of the nucleosome, viewed down the super helical axis. The histone fold domains of H2A, H2B, H3, and H4 are colored in yellow, red, blue, and green respectively, histone tails and extensions are shown in white and DNA is shown in light blue. (B) Side view of the same structure, obtained by a rotation of 90° around the vertical axis (adaptation from Luger *et al*., 2003)

The nucleosome core particle is a protein octamer divided into four histone-fold dimers created by H3-H4 and H2A-H2B pairs. The two dimers H3-H4 interact through a four-helix bundle formed only from H3 and H3' histone fold to define the H3-H4 tetramer. Each dimer H2A-H2B interacts with the tetramer through a second homologous four-helix bundle between H2B and H4 histone folds. The central histone core of the four protein shows a highly similar structure with three motif of α-helices connected by two loops, this structure is denoted as α1-L1- α2-L2- α3.

The way that histone proteins interact each other is quite complex and not very interesting for the purpose of our discussion while it is more interesting talk about of how occur the interactions between histone-fold and DNA. Luger and colleagues observed five predominant features for the interaction histone-fold/DNA each time the phosphodiester chains faces the histone core (Luger *et al*., 1997).

- the N-terminal of the α1-helices of H3, H4 and H2B and the four α2-helices of all the kinds of histone proteins, are used to fix the position of an individual phosphate group through the positive charge generate by the helix dipole;

- Hydrogen bonds to phosphates are made from main-chain amide nitrogen atoms of amino acid near or in the last turn of the α1 and α2 helices;
- An arginine side chain from an histone folds enters the minor groove 10 of the 14 times it faces of histone octamer. The other four occurrences have arginine side chains from tail regions penetrating in the minor groove;
- Extensive nonpolar contacts are made with the deoxyribose groups;
- Hydrogen bonds and salt links occur frequently between phosphate oxygen atoms of DNA and protein basic and hydroxyl side chain groups.

It has been estimated that about 28% of the mass of the core histone proteins is made up of sequences of histone N- and C- terminal tails those can interact with the DNA structure with the turn of the DNA super helix or creating a link with the minor groove of DNA. Biochemical studies of histone proteins have shown that they can be extensively modified post-translationally at their N-terminal tails through the addiction of different chemicals groups.

The DNA structure described above that consist of the wrapping of the DNA on the nucleosomes is called “beads on a string”. Each nucleosome is linked to the adjacent by a portion of DNA called linker that could be long 20 to 80 bp. This is only the first level of the packaging of DNA and could reduce the space occupied by DNA of 6-fold. The second level of packaging involves the histone H1 that brings together the nucleosomes stacking them in couples or grouping them in groups of eight nucleosome (solenoid) that can writhe creating loops that reduce the space occupied by DNA of other 10-folds. This is the level of DNA packaging maintained during active transcription or at the level of those genes that are usually expressed during the different phases of the development. During the differentiation a part of the chromatin could be further packaged to prevent its transcription. In this way occurs the forth levels of condensation that has been discovered in telomeric and centrometic regions where there are few genes and the chromatin is always condensate independently of the cellular type. This level of chromatin condensation is called heterochromatin.

The chromatin structure needs to be highly dynamic because it is involved in the regulation of all the cellular functions. The dynamism of the chromatin could be associated to the chromatin-based gene regulation that depends of interplay between sequence-specific DNA binding proteins, histone variants, histone modifying enzymes, chromatin-associated proteins an ATP-dependent nucleosome remodelers, but how all these components work together in still unclear (Li *et al*.,2007). Nowadays it seems clear that the composition of a nucleosome in terms of histone variants and histone post-translational modifications (PTMs) dictated its physical stability and propensity to be slid along the DNA or removed completely (Henikoff, 2008). In this way, pathway directed at altering these nucleosome characteristics can be used to regulate exposure and occlusion of the DNA.

2.2 Epigenetic

The term epigenetic was first used in 1942 by Waddington to describe "the interaction of genes with their environment that bring the phenotype into being" (Waddington, 1942 (2012)). Nowadays the definition includes all features such as DNA and chromatin modifications that are hereditable and stables over rounds of cell divisions without altering the nucleotide sequences of the underlying DNA. Since 1942 a wide variety of products and events have been defined as epigenetics and it has been proposed that the genome plasticity was determined by the epigenome. DNA methylation and the chromatin modifications profiles define the epigenomes in plants and animals in a way that define **DNA methylation, histone modifications** and **RNA-based mechanisms** as the "three pillars of epigenetics" (Grant-Downton *et al*., 2005 and 2006).

2.2.1 DNA methylation

In most eukaryotes DNA methylation consist of the transfer of a portion of A-adesylmethionine (SAM) to the 5-position of cytosine (Figure 1.2).

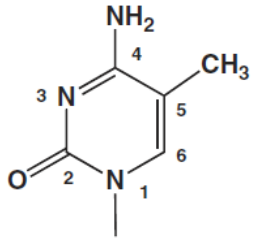


Figure 1.2: Structure of 5-methyl cytosine

In plants DNA methylation occurs at cytosine bases in all sequences contexts: symmetric where the cytosine lies in CG or CHG contexts (where H represent A,T or C) or asymmetric where the base methylated belongs to a CHH context. (Henderson and Jacobsen, 2007). Genome-wide DNA methylation levels analysis in *Arabidopsis thaliana* revealed CG, CHG and CHH contexts levels of 24%, 6.7% and 1.7%, respectively (Cokus *et al*., 2008) and the predominantly of DNA methylation on transposons and repetitive DNA elements (Zhang *et al*., 2006). In plants *de novo* methylation is catalyzed by DOMAINS REARRANGED METHYTRANSFERASE 2 (DRM2) and is maintained by three different pathway: **CG** methylation is maintained by DNA METHYLATRANSFERASE 1 (MET1, also called DMT1); **CHG** methylation is maintained by CHROMOMETHYLASE 3 (CMT3), specific of plants species, and asymmetric **CHH** methylation is maintained through persistent *de novo* methylation by DRM2. (Chan *et al.*, 2005). Although the characters of cytosine methylation are well defined, the pathways that controls the establishment and the maintenance of DNA methylation, as well as the ones involved in the removal of this modification, are less characterized. It became to be clear that methylation of cytosine is driven by RNA interference with RNA direct DNA methylation pathway, a complex mechanism purpose in 2010 by Law and Jacobsen (Law and Jacobsen, 2010) and, more recently revised, by Matzke and Mosher (Matzke and Mosher, 2014).

2.2.2 Histone modifications

Nucleosomes are the building block of chromatin that can be defined like a highly condensed structured that forms the scaffold of fundamental nuclear processes such as transcription, replication and DNA repair. Nowadays it is clear in fact that the function of chromatin doesn't end with the packaging and the protection of DNA but it also acts controlling gene expression (Fyodorov and Kadonaga, 2001). Cells have evolved mechanisms that alter the structure of chromatin allowing the access of cellular machinery for gene expression to chromatin DNA. Chromatin modifications can occur through histone posttranslational modifications (PTMs) and by the substitution with histone variants.

There are at least seven different kind of histone PTMs: proline isomerization, sumoylation, ubiquitination, ADP-rybosylation, phosphorylation, methylation and acetylation. These modifications, together with DNA methylation, control the folding of the nucleosome arrays into higher order structures and mediate signaling for cellular processes.

Proline Isomerization

The transformation of a molecule into a different isomere is defined isomerization. In the protein field a change in the conformation (*cis-trans*) of the protein can disrupt the secondary structure of the peptide. The isomerization can occur spontaneously but proline isomerases enzymes have evolved in order to accelerate the change between the different isomeres. In 2006 Nelson *et al*. (Nelson *et al*., 2006) described for the first time that histones can be isomerized in *Saccharomyces cerevisiae*; they identified Frp4 like histone isomerase of proline 30 (P30) and P38 on of Histone 3 tail. It is interesting to know that the conformational status of P38 is necessary for the induction of methylation of lysine 36 of histone H3 (H3K36me) where a change in isomerization of P38 seems to inhibit the possibility to methylate H3K36.

Sumoylation

Sumoylation consists in the addition of a "Small Ubiquitin-related Modifier protein" (SUMO) of about 100 amino acids. In 2003 it was reported that this kind of modification can affect the

histone H4 of HeLa cells and leads to transcriptional repression through the recruitment of histone deacetylases and HP1 protein (Shiio and Eisenman, 2003).

Ubiquitination

The modification of the ε-amino group of lysine residues by the covalent attachment of one or more ubiquitin monomers (76 amino acid protein highly conserved for all eukaryotes) is called ubiquitination. Typically monoubiquitination mark modifies the protein function whereas the polyubiquitination mark a protein to be degraded however the effects of histone ubiquitination on the transcription is still unclear (Dawson *et al*., 1991).

ADP-ribosylation

ADP-ribosylation is a post-transcriptional modification that consists of the addiction of an ADP-ribose onto a protein, mono-(ADP-ribosylation), or more of that, poly-(ADP-ribosylation) (PARation). Mono-(ADP-ribosylation) occurs in core histones and in the linker histone H1 either in response of stress condition or in physiological conditions. (Kreimeyer *et al*., 1984). Recent studies clarified the meaning of the PARation in response to DNA damage hypothesizing that after a damage the PARation leads to a quick and transient compaction of the chromatin that would protect the genetic material from additional damages and, after that, the same portion of chromatin is loosed allowing DNA repair to take place (Timinszky *et al*., 2009).

Phosphorylation

Protein phosphorylation is represented by the addiction of a phosphate group (PO_4) to a protein molecule, this action is catalyzed by specific protein kinases. In many eukaryotic organisms phosphorylation of histone H3 seems to be crucial for activating transcription, apoptosis, DNA repair and, although, cell-cycle with chromosome condensation and segregation. It has been propose that, also in plants, H3 phosphorylation has a role in the transcriptional activation of genes (Cheung *et al*., 2000). The primary function of histone H3 phosphorylation is still controversial, it is known that histone modifications control the binding of non-histone proteins to the chromatin fiber and that some chromodomains interact with methylated lysine while

bromodomains bind to acetylated lysine but until now no proteins that interact with phosphorylated histones have been described (Fuchs *et al.*, 2006).

Methylation

Protein methylation is a covalent modification constituted by the addition of a methyl group from donor S-adenosylmethionine (SAM) on carboxyl group of glutamate, leucine and cysteine, or on side-chain nitrogen atoms of lysine, arginine and histidine (Clarke, 1993). Histone methylation involves only arginine and lysine residues, the first can be mono- or di-methylated while the second, that is the most important, can be mono-, di- or tri-methylated (Kouzarides, 2007). Most of the histone modifications are conserved through eukaryotes but the establishment, the maintenance and, sometimes, the effects on gene regulation of these modifications may be different in the several kingdoms.

Histone lysine methylation is the most important histone modification and in *Arabidopsis* it involves mainly lysine 4 (K4), K9, K27 and K36 of histone 3 (H3). Generally, histone H3K9 and H3K27 methylation is associated with silenced regions, while H3K4 and H3K36 methylation is associated with active genes (Berger, 2007). The methylation of lysine residues does not change the net charge of the modified residue but rather it increases the hydrophobicity and may alter inter- and intra-molecular interactions that could create new binding surfaces for reader proteins that bind preferentially to the methylated domain.

The enzymes responsible for these modifications are called histone lysine methyltransferase (HKMTs) and SET domains proteins are putative candidates to be the writers of lysine methylation. *Arabidopsis* encodes for 41 SET domain proteins whereas rice encodes for 37 putative ones. SET domain proteins in plants are classified into four categories:

- SU(VAR3)-9 groups including SU(VAR3)-9 homologs (SUVH) and SU(VAR3)-9 related protein (SUVR);
- E(Z)b (enhancer of zeste) homologs;
- TRX (thithorax) groups (TRX);

- ASH1 (absent, small, or homeotic disc1) groups ASH1 homologs (ASHH) and ASH1-related protein (ASHR) (Baumbusch *et al*., 2001; Berger, 2007).

The enzymatic activity or specificity of these SET proteins in plant are still not well elucidated but genetic data suggest that they may act on the same lysine residue and pathway of the homolog in animals and yeast.

Whereas the most covalent histone modifications are reversible, histone methylation was considered as irreversible until 2004-2005 when two research groups discovered Lysine Specific Demethylase 1 (LSD1) that removes mono- and di-methyl groups from H3K4 (Shi *et al*., 2004) and from H3K9me in mammalian (Metzger *et al*., 2005). In *Arabidopsis* and in rice there are 4 LSD1-like genes; in *Arabidopsis* one of them, FLOWERING LOCUS D (FLD) has been shown to promote transition from vegetative to reproductive phase by repressing the FLOWERING LOCUS C (FLC). FLD acts partially redundantly with other two homologues: LSD1-LIKE1 (LDL1) and LSD1-LIKE2 (LDL2). Furthermore till 1995 Takeuchi *et al*. proposed Jumonji protein as potential histone demethylases (Takeuchi *et al.*, 1995). Structural analysis identified a conserved domain called JmjC-domain like the holder of demethylase activity (Tsukada *et al*., 2006). In animal cells many JmjC domain-containing histone demethylases have been identified; those were divided into distinct groups depending on sequences similarities (JARID/KDM5, JMJD1/JHDM2/KDM3, JMJD2/KDM4, JMJD3/KDM6, JHDM1/FBX/KDM2 and JmjC domain-only) and about 20 JmjC domain-containing proteins are identified in *Arabidopsis* (Chen *et al*., 2011).

The most studied histone methylation marks are: H3K9 and H3K27 with repression activity, H3K36 with activation function and H3K4 with divalent feature.

Methylation of H3K9

In *Arabidopsis* H3K9 histone methylation exist predominantly like mono- and di-methylation, while the amount of H3K9me3 is low. (Johnson *et al*., 2004). H3K9me2 is enriched in repeated sequences and in transposons and, together with H3K9me2, take part in silencing of the chromocenters. The tri-methylation on Lys9 of H3 shows a complete different outcome respect

of mono- and di-methylation, in fact this mark seems to be enriched in euchromatin where most active genes are found (Mathieu *et al*., 2005).

Drosophila SU(VAR)3-9 was the first identified histone methyltransferase specific for H3K9 (Rea *et al.*, 2000). Plant genome encoded 10-12 of Su(VAR)3-9 homologues and KRYPTONITE (KYP), also known as SU(VAR)3-9 homolog 4 (SUVH4), was the first plant histone H3K9 methyltransferase identified and it works together with SUVH5 and SUVH6 (Jackson *et al*., 2002 and 2004) . Two genes belonging to KDM3/JMJD1 and JMJD2/KDM4 groups were identified in plants like involved in H3K9 demethylation, they were INCREASE IN BONSAI METHYLATION 1 (IBM1/JMJ25) *Arabidopsis* and JMJ706in rice. (Chen *et al*., 2011; Sun *et al.*, 2008).

In 2007 Vaillant and coworkers characterized the interplay between H3K9 methylation and DNA methylation showing that this histone mark is fundamental for the maintenance of genome wide transcriptional gene silencing ad genome stability (Vaillant *et al*., 2007). Especially the maintenance of CG and non-CG DNA methylation requires H3K9 methylation for the creation of a performing instrument of positive feedback loop that allow the repression of sensible region of genome like transposable elements in accord with the RDdM that will be described subsequently.

Methylation of H3K27

H3K27 methylation is a repressive mark that can act in different way respect the previously described H3K9 methylation. In *Arabidopsis* H3K9me1-2 and H3K27me1 are enriched at constitutive silenced heterochromatin but the way that they interact with the DNA environment is completely different. While H3K9me2 is strongly associated with DNA methylation, H3K27me1 is independent of DNA state, this evidence suggest that their disposition, maintenance and, maybe, function can be mediated by different pathways (Mathieu *et al.*, 2005). The current opinion is that DNA-methylation dependent H3K9me2 pathway and DNA-methylation independent H3K27me1 pathway control constitutive heterochromatin formation in parallel (Liu *et al*., 2010). Whereas the molecular mechanisms that trigger and maintain H3K27me1 at the

constitutive heterochromatin is not completely clear instead more is known about tri-methylation of this residue.

H3K27me3 has been implicated in developmental regulation since it provides a cellular memory to maintain the repressed transcriptional states of target genes during cell division. Because of its fundamental function this mark has been studied in detail identifying in animals E(Z) SET domain histone methyltransferase within polycomb repressive complex 2 (PRC2) that catalyzes tri-methylation of H3K27 which in turn is recognized by the chromodomains of POLYCOMB (Pc), a core component of the PRC1 complex. The *Arabidopsis* genomes encodes several homologues of *Drosophila* E(Z) like CURLY LEAF (CLF), SWINGER (SWN), MEDEA (MEA) and other components of PRC2 in *Arabidopsis* have been shown to behave as regulator of plant developmental transitions (Pien and Grossniklaus, 2007).

In *Arabidopsis* the mark H3K27me3 is preferentially localized to the transcribed region of genes, with a bias toward those immediately upstream the promoters and at 5' end of transcribed region of genes. This modification has a consistent role in transcriptional repression and a large number of coding genes (about 17%) was found to be marked with tri-methylation on Lys27, indicating that H3K27me3 is a major gene silencing mechanisms in *Arabidopsis* (Zhang *et al*., 2007).

Methylation of H3K36

Lysine 36 of histone 3 can be di- or tri-methylated (H3K36me2, H3K36me3), these marks are both related to an increase in transcription of genes marked on gene bodies. (Lauria and Rossi, 2011). Little is known about the mechanisms of methylation and demethylation for this mark, recent data suggest that the activity of a histone methyltransferase (SDG8/EARLY FLOWERING IN SHORT DAYS, EFS) maybe specific to H3K36 together with SDG4. It has been shown that methylation on FLC could be balanced by REF6 (JMJ12), that suggest an involvement of REF6 in H3K36 demethylation (Ko *et al*., 2010).

Methylation of H3K4

Lysine 4 of histone 3 could be mono- di- or tri-methylated, these modifications occurs in at least two-thirds of genes of *Arabidopsis*. Histone H3K4 methylation is mediated by thrithorax group proteins (TRX) that are represented in *Arabidopsis* by ARABIDOPSIS TRITHORAX 1,2 (ATX1,2) (Saleh *et al*., 2008). ATX1 and ATX2 are not the only methyltransferases of H3K4; also SDG4 and SDG2, members of SET-domain genes, are responsible of this mark (Berr *et al*., 2010). For what concern demethylation of this mark it has been discovered a couple of *Arabidopsis* JmjC genes encoding demethylases that are members of the group JARID1/KDM5 and called JMJ14 and JMJ15.

Zhang and colleagues applied genome-wide analysis like chromatin immunoprecipitation (ChIP) and whole-genome tilling microarrays (ChIP-on-chip) for establish that all three types of H3K4me are highly enriched in the gene-rich euchromatin and absent in pericentromeric heterochromatin regions where transposons and other repetitive sequences cluster. It is estimated that 96.7%, 93.3% and 95.7% of all H3K4me1, H3K4me2 and H3K4me3, respectively, lied on transcribed regions of genes or in their promoters. In particular H3K4me3, H3K4me2 and H3K4me1 are distributed with a 5'-to-3' gradient along genes; H3K4me3 are enriched in promoters and in 5' region of transcribed genes with an occurrence of tri-methylation lightly above of the region interested by the marks H3K4me2. H3K4me1 is instead poorly represented in promoter regions but enriched in the transcribed regions with a bias with 3' end. (Zhang *et al*., 2009; van Dijk *et al*., 2010). To investigate in detail the distribution of the H3K4me in *Arabidopsis* genes Zhang and coworkers selected genes longer than 1 kb and located at distance of 1 kb or more from other genes. Using these criteria they obtained about 6000 genes that were divided into four classes according to their length: long genes (>4 kb), intermediate genes (3-4kb and 2-3kb) and short genes (1-2kb). For long and intermediate genes the distribution is quite similar with a distribution of K3K4me1 along all gene length and with the peak of H3K4me2 slightly downstream of H3K3me3 (600-800bp and 300-600bp respectively from transcription starting

site). These results suggest that the length of gene may affect the association with histone modification especially for H3K4me1 (Figure 1.3).

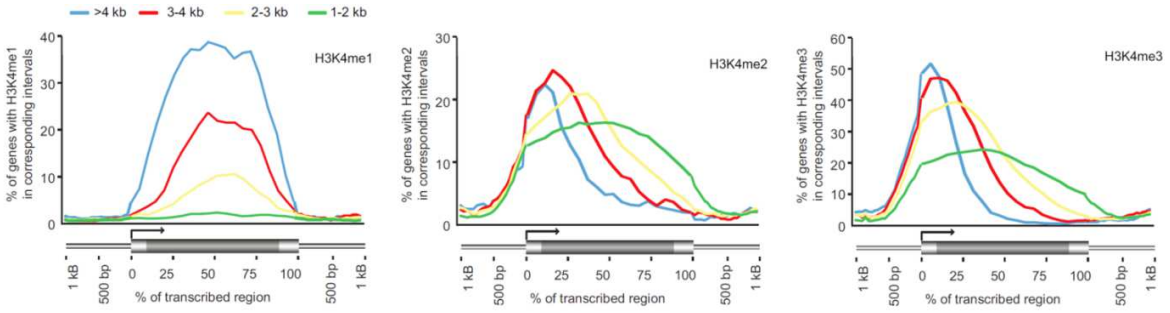


Figure 1.3: Distribution of H3K4me across Arabidopsis genes. Genes are divided into four groups according to their length. On the x-axis is represented each gene (thick horizontal bar), divided in perceptual of the transcribed region, and the 1kb region upstream and downstream. On the y-axis is represented the perceptual of genes with histone mark considered (adaptation from Zhang *et al*., 2009)

Once the researchers considered not only the distribution of the marks but also their relation with the expression profile it appeared that H3K4me3 was generally associated with actively transcribed genes while the correlation with transcription for mono- and di-methylation of H3K4 was not well defined. Furthermore this study shown that H3K4me2 often overlaps with H3K27me3. Interesting H3K4me2 and H3K4me3 appear to be mutually exclusive with methylated Cytosine, while H3K4me1 co-localizes with CG methylation within the gene body suggesting a functional interplay between these two marks.

Acetylation

Acetylation describes a reaction that introduce a acetyl group (CH_3COO^-) into an organic compound. In histones amino-terminal tails have positively charged lysine and arginine and the high binding affinity between DNA and histones is based on their opposite charge. In 2001 Lusser and colleagues proposed three alternative explanations for the effects of lysine acetylation on chromatin structure. First, acetylation might neutralize a positive charge and thus weaken the interaction of the histone octamer with the negatively charged DNA. This would destabilize nucleosomes, allowing transcriptional regulators to access to the DNA. Second, acetylation might interfere with higher-order packing of chromatin and thus alter the accessibility of large

chromatin areas for regulatory proteins (Lusser *et al*.,2001). Third, acetylation could act as a specific signal that alters histone-protein interactions (Loidl, 1994)

Histone deacetylases (HATs) could be divided into two classes, HAT-A and HAT-B (Roth *et al*., 2001). Type A HATs are located in cell nucleus and acetylate nucleosome core histones; they act as transcriptional co-activators playing an important role in the regulation of gene expression. Type B HATs catalyze acetylation events linked to the transport of newly synthesized histones from the cytoplasm to the nucleus for deposition onto newly replicated DNA.

Sequence characterization reveals several distinct families of HATs that were described below.

- The GNAT (GCN5-related N-terminal acetyltransferase)-MYST family whose members have sequence motifs shared with enzymes that acetylate small molecules and non-histone proteins. The GNAT family is generally considered composed by four subfamilies designated GCN5, ELP3, HAT1 and HPA2. Pandey and coworkers identified, in the *Arabidopsis* genome, a single homolog of each of the GCN5, ELP3 and HAT1 subfamily (respectively HAG1, HAG3 and HAG2) and no homolog for HPA2. Furthermore *Arabidopsis* genome encode two MYST family proteins: HAG4 and HAG5 (Pandey *et al*., 2002);
- The p300/CREB-binding protein (CBP) co-activator family that, in animals, is required for cell cycle control, differentiation and apoptosis. The CBP family of HATs (CBP-type HAT) is comprised of large, multidomains proteins and there is no relation between the histone acetylation domain of GNAT-MYST and CBP-type HAT. *Arabidopsis* encodes for five CBP-type HAT domain proteins named HAC1, HAC2, HAC4, HAC5 and HAC12.
- The family related to mammalian $TAF_{II}250$ that is the largest of the TATA binding protein-associated factors (TAFs) within the transcription factor complex TFIID (Mizzen *et al*., 1996). Pandey and colleagues studied the domain architecture of $TAF_{II}250$-type proteins in plants (*A. thaliana*), animals (*D. melanogaster*, *H. sapiens*) and fungi (*S. cerevisiae*, *S. pombe*) discovering three interesting features:

- $TAF_{II}250$-type HAT domain in animals shows two bromodomains whereas *Arabidopsis* proteins possess only one of these in the same region, C-terminal side of HAT domain.
- A zinc-finger-type domain in present downstream of HAT domain in all the proteins suggesting a role in DNA binding or protein-protein interactions.
- A conserved ubiquitin signature was found in N-terminal side of HAT domain in *Arabidopsis* proteins but not in the animal or fungi proteins(Pandey *et al*., 2002).

The main acetylation sites in most species involve lysine 9, 14, 18 and 23 of histone 3 (Strahl and Allis, 2000). In particular acetylation of H3K9 is the most characterized epigenetic marks associated with active transcription and has been shown to influence numerous developmental and biological processes in higher plants (Benhamed *et al*., 2006; Ausìn *et al*., 2004; Zhou *et al*., 2010). These mark is invariably correlated with transcriptional activation in all the species analyzed so far (Kurdistani *et al*., 2004; Pokholok *et al*., 2005;Sinha *et al*., 2006; Roh *et al*., 2005).

Histone variants of histones H2A and H3

In addition to the major histones all eukaryotes have variant types of H2A and H3 that are incorporated into chromatin during interphase and that can impart particular properties to the nucleosome they occupy (Talbert and Henikoff, 2010). H2A shows two histone variants H2AX and H2A.Z while H3 has two variants CenH3 and H3.3.

Histone H2.AX is phosphorylated on C-terminal serine when a DNA damage happened and it is involved in the orchestration of the DNA repair pathway (van Attikum *et al*., 2009). The role of this variant in repairing pathway has been well elucidated in animals and it is presumed to play a related role also in plants (Rybaczek *et al*., 2007).

The variant H2A.Z differs to the canonical H2A by many amino acid substitutions widespread along the protein and in particular in the C-terminal α-helical region (Suto *et al*., 2000). This variants have been implicated in many different cell processes including maintenance of genome integrity, transcriptional regulation and formation of heterochromatin boundaries (Raisner and Madhani, 2006). Several studies have localized this variant widespread into the genomes in

nucleosome flanking the transcriptional start site (TSS) where it appears to play a role in transcription regulation by helping to prevent methylation (Zilberman *et al*., 2008). In 2003 Mizuguchi and coworkers discovered that H2A.Z was inserted into nucleosome by the yeast Swr1 ATP-dependent nucleosome remodeling complex or by related complexes in plant and animals, these complexes act unwrapping in part the nucleosome and replacing and H2A/H2B dimer with and H2A.Z/H2B dimer (Mizuguchi *et al*., 2004).

The two variants of histone 3 that were found in all eukaryotes are: CenH3 and H3.3

CenH3 is incorporated at centromeres and it is essential for chromosome segregation. It is still unclear what determines the CenH3 deposition in centromere regions but the available evidences indicate that a specific sequence, or arrangement of DNA, is neither necessary nor sufficient to organize the deposition of CenH3. The most consistent finding is that the presence of CenH3 fosters the deposition of new CenH3 molecules at the same position at each cell division. If this cycle is broken, then inactivation of the centromere occurs (Birchler *et al*., 2011).

The variant H3.3 differs from canonical H3 only for three to four amino acids in plants and animals and it is deposed in the nucleosome outside of DNA replication by the action of different histone chaperones, like HirA and Daxx, depending on the genomic location. H3.3 is located predominantly within promoters, transcribed regions of expressed genes and at gene regulatory elements where nucleosomes are being rapidly disrupted and replaced. (Mito *et al*., 2005). Multiple isoform of H3.3 exits in plants but none of these have been yet mapped genome wide or studied in detail (Deal and Henikoff, 2011).

H2A.Z and H3.3 in animals show partially overlapping in their distribution among the genome, both variants are enriched, in fact, near the TSS and in gene body nucleosomes at the 5’ end of expressed genes. Nucleosomes containing H2A.Z but not H3.3 are relatively stable, but *in vivo* studies reveal that those that contain both variants are prone to disassembly (Henikoff, 2008). These unstable double-variant on non-canonical nucleosomes are found near TSSs and so may modulate exposure of promoter DNA by promoting nucleosome turnover (Jin *et al*., 2009).

Correlation between histone modifications and chromatin states

The histone code hypothesis was defined by Strahl and Allis like the hypothesis that multiple histone modifications could act in a combinatorial or sequential fashion on one or multiple histone tails (Strahl and Allis, 2000; Henikoff, 2011). Genetic and biochemical studies suggest that histone methylation controls DNA methylation (Tamaru and Selker, 2001) and histone acetylation affects histone methylation (Lawrence *et al.*, 2004). To study the relationships among various histone modifications Ha and colleagues analyzed H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27me3,and DNA methylation patterns within 2-kb upstream and downstream of the transcription start site (TSS) using 100-bp sliding windows identifying that the correlation between H3K4me3 and H3K9ac at the same loci is highly significant (Ha *et al*., 2011). Chromatin regions with H3K4me3 are targeted for histone acetylation by histone acetyltransferases (Wang *et al*., 2009), and these modifications may reinforce the activity of constitutively expressed genes. Furthermore the location of histone marks like H3K9ac and H3K4me3 near the TSS destabilize the interaction between histones and DNA, leading to nucleosome loss (Boeger *et al*. 2003; Reinke and Horz, 2003).

At the light of the recently published works we describe briefly the chromatin state of whole *Arabidopsis thaliana* genome. At first sight chromatin could be labeled as a static and repetitive structure but it is far from reality because there are at least three major sources of variation. One is DNA modifications, represented primary by cytosine methylation; the second is the posttranslational histone modifications and, finally, the individual histone molecules that could be replaced within nucleosome with other histone variants such as H2A.Z and H3.3. Altogether, as said before, these variations provide a very high combinatorial diversity at individual genomic loci and this large diversity of chromatin composition has significant consequences in transcription and genes replication (Berger, 2007). The first effort to identify chromatin states was performed in *Drosophila melanogaster* using genomic information of 53 chromatin proteins (Filion *et al*., 2010). This study allowed to the identification of five major chromatin states, namely heterochromatin, Polycomb, repressed and two types of active chromatin regions and a

more recent study based on 18 histone modifications in *Drosophila* cultured cells identified nine chromatin states (Kharchenko *et al*., 2011), whose functional significance was investigated by integrating chromosome organization with data of DNaseI hypersensitivity, RNA transcripts and non-histone protein binding. A similar approach was performed in *A. thaliana* by Routier and coworker that, using information from histone marks across tilling arrays of chromosome 4, identified four major chromatin states: heterochromatin, Polycomb, active genes and intergenic regions (Roudier *et al*., 2011). In 2014 Sequeira-Mendes and colleagues reported nine chromatin states for *Arabidopsis* providing a ground for a better understanding of the linear organization of the genome and the relevance and/or the preference that certain signatures of genomic elements may have for either establishing gene expression patterns or specifying DNA replication origins (Sequeira-Mendes *et al*., 2014). Their work regarded data obtained from the published profiles of nine histone modification marks (H3K9me2, H3K27me1, H4K5ac, H3K4me1, H2Bub, H3K36me3,H3K4me2, H3K4me3 and H3K27me3) and three histone variants (H2A.Z, H3.1 and H3.3). They also evaluated the nucleosome density obtained from total H3 histone content, the genomic CG content and the CG methylated residues (Bernatavichute *et al*., 2008; Zilberman *et al*., 2008; Stroud *et al*., 2012). Moreover the Spanish researchers generated genome-wide chromatin immunoprecipitation ChIP-on-chip data for H3K9ac and H3K14ac. Analyzing this huge amount of data they have concluded that nine chromatin states render a solid and coherent biological interpretation of the genome of *Arabidopsis* (see chapter IV for details about the nine states identified).

The relationships between different histone marks have without doubts a functional relevance as previously observed by Henikoff; these relations and the investigation of eventual similar modifications profile in different species may became the field of study of the scientific community (Deal and Henikoff, 2011; Henikoff, 2011).

Complex works like the ones described above remain confined to *Arabidopsis*. ChIP approach were, in fact, performed only to a handful of plant species like maize (He *et al*., 2014, Wang *et al*., 2009), rice (Du *et al*., 2013; He *et al*., 2010), tomato (Ricardi *et al*., 2010 and 2014) and poplar

(Li *et al*., 2014). Although the available data are sometimes very preliminary it become clear that an epigenetic regulation of gene expression performed by histone modification has a key role in plant development and stress response.

Many open questions remain regarding the role of chromatin modifications and epigenetics in plant developmental processes. Development in multicellular organisms often involves differentiation of cells into specialized cell types that express different sets of genes. The silencing of certain genes in specific cell types could be considered to occur through facultative heterochromatin, with repressive chromatin indifferent locations based on cell type (Eichten *et al*., 2014). On the other hand epigenetic may contribute to plant development by stably maintaining gene expression states that are initially directed by sequence-specific factors, such as transcription factors or small RNAs. Direct analyses of chromatin modifications in different cell types have provided evidence that the profiles of some histone modifications change substantially during development (Roudier *et al*., 2009). For example, H3K27me3 contributes to regulation of important transcription factors in some cell types and shows clear tissue-specific patterns that are often associated with tissue-specific expression of the target genes in *Arabidopsis* (Lafos *et al*., 2011; Zheng and Chen, 2011) and maize (Makarevitch *et al*., 2013). Other marks, such as histone acetylation and H3K4me3, show tissues specific patterns (Berr *et al*., 2010; He *et al*., 2014) but may be an effect, rather than a cause, of tissue-specific gene expression.

All the strong evidences present in literature show that chromatin varies among different cell types and that some of these modifications play important roles in plant development, but whether specific examples are epigenetic (meant as transmissible) during plant development remains less clear.

A large number of studies have shown that epigenetic modifications of DNA and histone play a key role in gene expression and plant development under stress (Chinnusamy and Zhu, 2009; Luo *et al*., 2012). Tobacco and *Arabidopsis* cells show dynamic changes in histone modifications in response to high salinity and cold stress, manifested by transient up-regulation of H3

phosphoacetylation and histone H4 acetylation (Tsuji *et al*., 2006; Sokol *et al*., 2007). Dynamic changes in genome-wide histone H3K4 methylation patterns in response to dehydration stress in *Arabidopsis* were also observed by van Dijk and colleagues (van Dijk *et al*., 2010). In the same year Chen *et al.* investigated the effects of abscisic acid (ABA) and salt stress on the histone acetylation and methylation of abiotic stress response genes (Chen *et al*., 2010). Both ABA and salt stress can induce histone H3K9K14 acetylation and H3K4 trimethylation but decrease H3K9 dimethylation of some ABA and abiotic stress responsive genes, suggesting that functionally related gene groups are regulated coordinately through histone modifications in response to abiotic stress in plant cells.

Kim and coworker evaluated the enrichments of trimethylation of histone H3 Lys4 (H3K4me3) and acetylation of histone H3 Lys9 (H3K9ac), often used as a positive marker of histone modifications associated with gene activity, and identified a correlation with gene activation in response to drought stress in all drought-inducible genes considered ((RD)29A, RD29B and RD20, and At2g20880) (Kim *et al*., 2008).These are some examples of undeniable relation between histone modifications and gene expression in plant stress responses.

2.2.3 RNA interference

RNA-direct DNA methylation (RdDM) is the major small RNA-mediated epigenetic pathway (Matzke, and Mosher, 2014).

Small RNAs can act in different ways depending on the cellular compart in which they lie; in the cytoplasm small RNAs induce post-transcriptional gene silencing (PTGS) through the targeting of complementary mRNAs for its degradation or translational repression. In the nucleus small RNAs induce transcriptional gene silencing (TGS) establishing repressive epigenetic modification represented by DNA cytosine methylation and histone modification to homologous region of the genome. Small RNAs are 20-30 nucleotides in length and they are classified into two major group: the small interfering RNAs (siRNAs) and the PIWI-interacting RNAs (piRNAs) that are not present in plant and fungi (Castel and Martienssen, 2013).

The core of RNAi machinery consist of two proteins: Dicer (DCR) and Argonaute (AGO). Dicer is a ribonuclease III enzyme that cleaves double-stranded RNAs (dsRNAs) precursors onto siRNAs. In plants Dicers homologous are named DICER-LIKE (DCL). In *Arabidopsis* the DICER-LIKE consists of four genes:

- DCL1 cleaves endogenous dsRNAs to produce siRNA and miRNA;
- DCL2 and DCL4 process dsRNA precursors into 21nt and 22nt siRNAs which combined with AGO proteins act in the guide of the degradation of homologous RNA in PTGS;
- DCL3 produces 24-nucleotide siRNAs involved in the RNA-direct DNA methylation pathway that leads to TGS (Bernstein *et al*., 2001; Bouche *et al*., 2006; Deleris *et al*., 2006).

Argonaute (AGO) proteins are involved in all known small RNA-direct regulatory pathway. Of the three group of the AGO family Plants encode only for group 1 that is the one that bind both miRNAs and siRNAs. All AGO proteins show a variable N-terminal domain and three conserved domains at the C-terminal: PAZ, MID (middle) and PIWI. The PAZ domain recognized the 3’ end of small RNAs, the MID domain binds to the 5’ phosphate of small RNAs and the PIWI domain adopt a folder structure similar to that of RNAseH with further endonuclease activity (slicer)

which is carried out by an active site usually carrying an Asp-Asp-Lys (DDH) motif (Vaucheret, 2008). Moreover the PIWI domains specifically interact with Gly-Trp (GW) repeat proteins (i.e. the *At*AGO4 interacts with *At*NRPD1b, the largest subunit of nuclear RNA polymerase IV) (El-Shami *et al*., 2007). *Arabidopsis* encode ten AGO proteins but only AGO4 and AGO6 operate in the DCL3-siRNAs TGS pathway while the role of the other AGO proteins are less clear.

RdDM in plants depends on a specialized transcriptional machinery that is centered on two plant-specific RNA polymerase II (Pol II) related enzymes called Pol IV and Pol V (Rivas *et al*., 2005). The current opinion identify the components of the RdDM in Pol IV-dependent siRNA biogenesis and Pol V mediated *de novo* DNA methylation or other chromatin alterations like histone modifications, nucleosome positioning and higher-order chromatin conformation.

Pol IV-dependent siRNA biogenesis

Pol IV is responsible for the production of the precursors of more than 90% of 24-nucleotide siRNAs, which drive methylation in the canonical RdDM pathway (Zhang *et al*., 2007). This polymerase is recruited to a subset of its genomics target by Pol IV-interacting protein SAWADEE HOMEODOMAIN HOMOLOGUE1 (SHH1), which binds to H3K9me and unmethylated H3K4 through its unique tandem Tudor-like (Law *et al*., 2013). It is in the current opinion that Pol IV transcribes single-stranded RNAs (ssRNAs) at its target loci. The ssRNAs are copied by the RNA-dependent RNA polymerase 2 (RDR2), that is linked to Pol IV, to produce dsRNAs. RDR2i is in fact an enzyme able to copy ssRNAs to produce dsRNA precursors that are processed by Dicer like protein to produce siRNAs.

Pol V-mediated *de novo* methylation

Pol V transcripts are thought to constitute the scaffold RNAs that interact with siRNAs and that recruit other factors for silencing machinery (Wierzbicki *et al*., 2008). Late studies identified, through ChIP-seq analysis, that most of Pol V are located at transposons and at repeats associated with 24-nt siRNAs and, further, with cytosine methylation indicating that Pol V acts in the RdDM of these sites (Wierzbicki *et al*., 2012).

In Figure 1.4 is briefly described the canonical RdDM pathway using an explanatory model propose by Pikaard and colleagues in 2012 and quoted also by Matzke in 2014 (Pikaard *et al*., 2012; Matzke and Mosher, 2014).

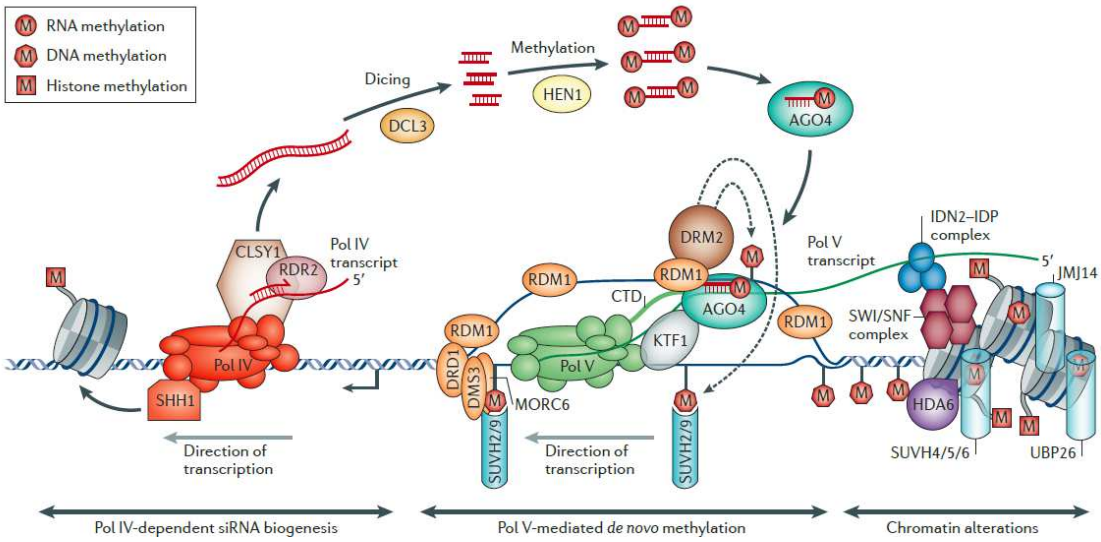


Figure 1.4: Canonical RdDM pathway. In RNA Pol IV-dependent siRNA biogenesis Pol IV transcribed ssRNA which is copied in a dsRNA through the action of RNA-dependent RNA polymerase (RDR2) with the mediation of CLASSY (CLSY1) chromatin remodeler. The nascent dsRNA is processed by DICER-LIKE 3 (DCL3) into 24-nt siRNA that shows a 3' ends methylation marked by HUA ENHANCER 1 (HEN1) and which is incorporated into ARGONAUTE 4 (AGO4). SAWADEE HOMEODOMAIN HOMOLOGUE1 (SHH1) binds to H3K9me and interact whit Pol IV recruits it to target loci. In Pol V-mediated *de novo* methylation the polymerase transcribes a scaffold RNA complementary with AGO4-bound siRNAs. AGO4 is recruited in the proximity of Pol V by the interaction between AGO hook region, in the C-terminal domain of the largest subunit of Pol V, and KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1). RNA-DIRECT DNA METHYLATION 1 (RDM1) links AGO4 and DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) which act catalyzing *de novo* methylation of DNA. The transcription of Pol V may be enabled by the duplex unwinding activity of DEFECTIVE IN RNA-DIRECTED DNA METHYLATION (DRD1) chromatin remodelers whereas the single-stranded DNA-binding activity of RDM1 and the putative cohesin-like roles of DEFECTING IN MERISTEM SILENCING3 (DRS3) and MICRORCHIDIA 6 (MORC6) may help to create and stabilize a unwound state. The recruitment of Pol V may be helped by SUV2 or SUV9 which bind to methylate DNA.
Another interesting pathway involved nucleosome positioning and heterochromatin formation that is adjusted by SWI/SNF chromatin remodeling complex which interact with IDN2-IDP (INVOLVED IN DE NOVO-IDN2 PARALOGUE) complex that bind to scaffold of RNA made by Pol V. The imposition of repressive mark, such as H3K9me made by SUVH4, 5 and 6; is facilitated by the action of the HISTONE DEACETYLASE 6 (HDA6), the histone demethylase JUMUNJI 14 (JMJ14) and the UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26) which remove the active histone marks (adaptation from Matzke and Mosher, 2014).

In plants RNA direct DNA methylation controls the establishment of DNA methylation in three different sequence contexts: CG, CHG and CHH (where H represent A, T or C). Three DNA methyltransferase cooperate to establish and maintain the genome methylation profile: CHROMOMETHYLASE3 (CMT3) and DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) produce *de novo* cytosine methylation while MET1 is considered the maintenance methylase that controls the symmetrical CG methylation on both DNA strands (Chan *et al*., 2005).

The CMT3-like genes encodes methyltransferase proteins containing chromodomain and are specific for plants (Henikoff and Comai, 1998). Through the chromodomain CMT3 binds demethylate lysine 9 on histone 3 (H3K9me2) and with SUVH4 (the activity that establishes the mark H3K9me2, also known as KRYPTONITE, KYP) it generates a feedforward loop for the maintaining the methylation on CHG (Henderson and Jacobsen, 2007). It is a siRNAs based mechanism the one that guide CMT3 to the sequences for non-CG methylation.

The DRM genes are required for the establishment of preexisting CG methylation. DRM, like CMT3, targeted sequences by siRNA and may act redundantly with CMT3 to establish and maintain CHH and CHG methylations (Henderson and Jacobsen, 2007).

MET1 together with chromatin remodeling factor DECREASED DNA METHYLATION (DDM1) and methylcitosine binding-protein VARIANT IN METHYLATION 1 (VIM1), is the main character for symmetric (CG) methylation.

2.3 Techniques for the evaluation of epigenetic marks

2.3.1 Bisulfite-mediated cytosine conversion to uracil

The principal method for determining the 5-methylcitosine sites in genes at the sequence level is the bisulfite genomic sequencing. The principle of this method resides in the selectivity of sodium bisulfite in the conversion of cytosines in uracils, while 5-methylcytosines remain unconverted (Figure 1.5).

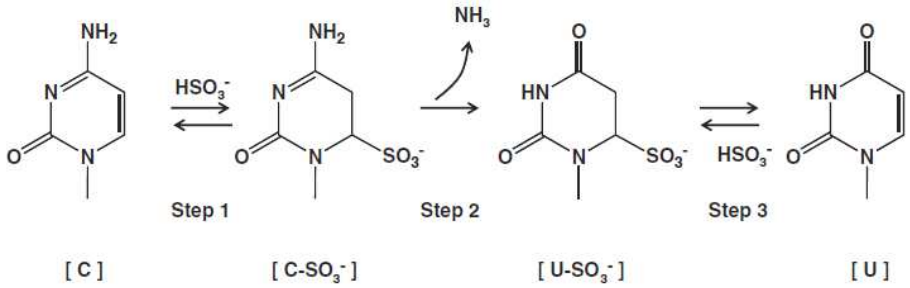


Figure 1.5: Bisulfite-mediated cytosine conversion to uracil

The bisulfite-mediated specific deamination of cytosine was discovered in 1970s by Shapiro and coworker in New York (Shapiro *et al.*, 1974) and, at the same time, in Tokyo by Hayatsu's group (Hayatsu *et al*., 1970) but only in 1992 Frommer *et al*. used this method for the determination of the sites of 5-methylcytosine in DNA (Frommer *et al*., 1992). They treated denatured DNA with bisulfite to convert cytosine to uracil. Subsequently an amplification by PCR was performed so that modified DNA produced a group of DNA fragments in which original cytosines were replaced by thymines. The cloning into a vector of this DNA followed by sequencing gave DNA sequences in which all the original cytosines have been changed into thymines. In these processes, 5-methylcytosine of the original DNA molecules stayed unchanged and the descendant DNA and, once sequenced, presents still a cytosine in that position, this allows to determine the sites of the 5-methylcytosines. Nowadays, in the NGS era, the traditional bisulfite-sequencing performed on a single locus has been exceeded and whole genome approach is preferred: the Bisulfite sequencing (BS-seq). In this technology after bisulfite conversion genomic

DNA libraries were created, sequenced and mapped on the genome of the species of interest (Cokus *et al*., 2008). The BS-seq is more sensitive and powerful than the previously used method but the genome of the species of interest had to be sequenced.

2.3.2 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) is a technique whereby a protein of interest is immunoprecipitated in a selectively manner from a chromatin preparation to determine the DNA sequences associated with it. ChIP has become the technique of choice to evaluate the interactions protein-DNA that occur inside the cells (O'Neill *et al*., 1996). ChIP technique is used for many applications in instance for mapping the localization of post-translational modified histones and histone variants in the genome, for mapping DNA target sites of transcription factors or other DNA associated proteins.

Chromatin immuonoprecipitation assay, as we know it since mid-1990s, has occurred over many years. In fact the use of formaldehyde to cross-link proteins with other proteins or with DNA was first reported in the 1960s and its application on study histone-DNA interactions goes back till the mid-late 1970s. In the 1990s the development of anti-histone antibodies led the path to the ChIP assay allowing the investigation of the association of histone with DNA in relation to transcription (Solomon *et al*., 1988). For over a decade, ChIP has remained a cumber-some protocol requiring a great deal of work and material for immunoprecipitation. These characteristics have limited the application of ChIP also considering that ChIP assay involve extensive sample handling that is inevitably a source of loss of material and an opportunity for technical errors. Recently some improvement have been introduced into the ChIP protocols to made them simpler, shorter and more efficient allowing the use of limited amount of starting material (Collas, 2010). However the development of these smarter protocols were limited to animal samples and plant model species while for other species it is still difficult obtain an efficient protocol for ChIP assay.

Chromatin immunoprecipitation is capable of providing high-resolution spatial and temporal information about the interaction between proteins and DNA in living cells. In traditional ChIP

assays the immunoprecipitated obtained during the experiments is evaluated by quantitative PCR (qPCR) using primers designed to amplify specific regions of interest. In the last decade the power of ChIP has been tremendously increased by its coupling with DNA microarray technology (ChIP-on-chip) or with Next Generation Sequencing (NGS) techniques (ChIP-seq).

ChIP-on-chip is an assay where immunoprecipitated material and its counterpart that weren't subjected to the immunoprecipitation (input), are labeled with two different fluorescent dyes and hybridized to DNA microarrays containing several hundred thousand, to several million probes. Binding of the precipitated protein to a target site is inferred when intensity of the ChIP DNA significantly exceeds that of the input DNA on the array.

Performing ChIP coupled with DNA microarrays has several significant advantages over traditional ChIP (evaluation of enriched DNA with PCR amplifications):

- ChIP-on-chip allow the investigation of a large number of genomic regions that could be probed in a single experiment, eliminating the bias of researches and permitting discovery of site of protein binding in region unexpected;
- Localization of protein binding can be accomplished with optimized available platform, eliminating time spent designing and testing couples of primers and running expensive large-scale quantitative PCR assays;
- The use of the same platform by different research groups can facilitate to directly compare the data obtained in different labs;
- The parallel analysis of thousands of genes allows the labelling of genes into different class based on different binding distribution, or behavior, and permits statistical comparison between classes (Gilchrist *et al.*, 2009).

ChIP-sequencing (**ChIP-seq**) offers an appealing complementary or alternative method of ChIP-on-chip. The strategy is similar to ChIP-on-chip but instead of labelling immunoprecipitated material and hybridize it on a microarray, the enriched DNA is used to construct libraries of millions of DNA fragments which are amplified and sequenced. Unlike microarray method, the vast majority of single-copies sites in the genome is accessible for ChIP-seq assay rather than a

subset selected to be tested. In addition, ChIP-seq technology avoids the constraints imposed by the chemistry intrinsically connected to array technology, such as base composition constraints related to T_m (the temperature at which 50% of double-stranded DNA, or DNA-RNA, is denatured); cross-hybridization and secondary structure interference. Moreover ChIP-seq technology could be applied to any sequenced genome, rather to be restricted to species for which genome tilling arrays have been produced. ChIP-seq fit well with the new sequencing platforms, like Illumina and SOLiD, in fact the large numbers of short individual sequence reads produced by these instruments are well suited to making direct digital measurements of the enriched DNA. The reads are mapped on the reference genome to identify the position and the frequency of the DNA enriched in the immunoprecipitated material. The desired level of sensitively and statistical needed to detect rare molecular species can be achieved by adjusting the total number of reads or the parameter used for their tilling. Another positive aspect of ChIP-seq is that it doesn't require to know in advantage if a sequence of interest is a promoter, an enhancer or a RNA-coding domain, as most current microarray design do (Johnson *et al*., 2007).

3 Next Generation Sequencing technologies

Ten years ago next-generation sequencing (NGS) technologies appeared in the scientific world. During this decade great progresses have been made in term of speed, read length and throughputs and NGS has been developed with a large number of novel applications in difference science fields (van Dijk *et al*., 2014).

In the 1970s two methods for DNA sequencing were developed: one based on side chain termination, developed by Sanger and coworkers (Sanger *et al*., 1977), and one based on fragmentation techniques created by Maxam and Gilbert (Maxam and Gilbert, 1977). The techniques developed by Sanger and colleagues, commonly referred like Sanger sequencing, became the prevailing DNA sequencing method for the next 30 years. Advantages like laboratory automation and process parallelization with hundreds of sequencing instruments allowed the implementation of the Sanger sequencing and, in 2004, it led to the completion of the first human genome sequence (International Human Genome Sequencing Consortium, 2004). The Human Genome Project, however, required vast amounts of time and resources and it was immediately clear that faster, higher throughput, and cheaper technologies were required. For this reason, in the same year the National Human Genome Research Institute (NHGRI) initiated a funding program with the objective of reducing the cost of human genome sequencing to US $ 1000 in ten years (Schloss, 2008). This target stimulated the development and the commercialization of NGS technologies.

All Next Generation Sequencing methods developed share three major characteristics that make the power of these technologies:

- The preparation of NGS libraries is performed in a cell free system;
- NGS performs thousands-to-many-millions of sequencing reactions in parallel;
- The sequencing output is directly detected and the base interrogation is performed cyclically and in parallel.

The enormous number of reads generated by these technologies enables the sequencing of entire genomes at an unprecedented speed. However a drawback of the first NGS technologies

was their relatively short reads that made genome assembly difficult and required the development of novel alignment algorithms (van Dijk *et al*., 2014; Buermans *et al*., 2014).

The first NGS technology was the pyrosequencing methods developed by 454 Life Science/Roche in 2005. The 454 sequencer generated about 200 000 reads of 110 bp of lenght (Margulies *et al*., 2005). One year later, the Solexa/Illumina sequencing platform was commercialized and, in 2007, the third NGS technology was released: the Sequencing by Oligo Ligation Detection (SOLiD) by Applied Biosystems/Life Technologies (Valouev *et al*., 2008).

The Illumina and SOLiD sequencers generated much larger numbers of reads than 454 (30 and 100 million reads, respectively) but the reads produced were only 35 bp of lenght. The fourth NGS technology was proposed in 2010 and it was the Personal Genome Machine (PGM) released by Ion Torrent/Life Technologies. This technology is based on semiconductor and does not rely on the optical detection of incorporated nucleotide using fluorescence and camera scanning.

When NGS technologies made their appearance in the scientific world Illumina and SOLiD sequencing were more suitable than 454 approach for quantitative application like transcriptome profiles and ChIP-seq studies because of their higher throughput. By contrast, the reads generated by these technologies were initially too short for *de novo* genome assembly thus for this application were preferred 454 approach. During the years Illumina upgrades its sequencing machines, base-calling software and sequencing chemistries and now it can generate reads of several hundred of base pair of length allowing its employment also for *de novo* genomes assembly and nowadays Illumina offers the highest throughput per-run and the lowest cost per-base. Illumina recently released new sequencing machinery called HiSeq X, and coupled together ten of these sequencers (HiSeq X Ten) generating up to 1.8 Tb of sequences for run. With this approach Illumina claims to have broken the barrier of the US $ 1000 for genome sequencing that corresponding to the original goal of the NHGRI funding program (National Human Genome Research Institute) (Figure 1.6).

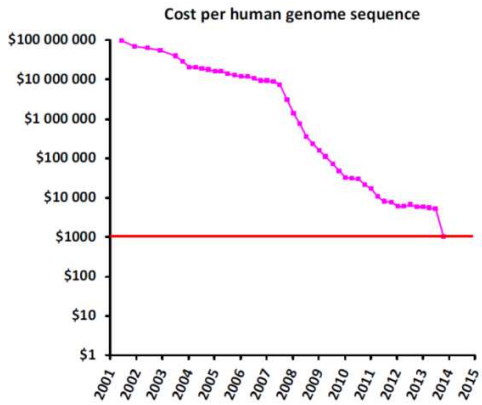


Figure 1.6:Evolution of the cost of sequencing a human genome from 2001 until today.

The NGS technology has a wide diversity of novel applications. Advantages in higher throughput and cost reduction have made Whole Genome Sequencing (WGS) a great instrument for the researches. Since the first large-scale human genetic variation study, the 1000 genomes project (1000 Genomes Project Consortium, 2010), other projects have been promoted involving the sequencing of thousand or even millions of genomes. The number of registered sequencing projects raises continuously, nowadays the number of complete project available at the Genomes Online Database (GOLD) is 6649. A great improvement led by NGS reside in the genetic disease diagnosis, WGS, in fact can accelerate molecular diagnosis and minimize the duration of empirical study (Kingsmore *et al.*, 2011). Another exciting new field is single cells genomics whose major objective is to reconstruct cell lineage trees using somatic mutations that arise to DNA replication errors. These cell lineages trees can have applications in developmental and tumor biology (Frumkin *et al.*, 2005). The NGS technologies have also been successful applied to gene expression profiling sequencing mRNA. This application plays an important role in molecular biology because the activity of thousands of genes could be measured on parallel. Moreover the NGS enables the absolute quantification of the transcripts, this allows to detect changes of gene expression levels under different biological conditions or between different cell types or tissues. NGS has also been applied to small non-coding RNA (ncRNAs) discovery and profiling and also on sequencing of immunoprecipitated chromatin (ChIP-seq) (Morozova *et al.*, 2008).

4 Research objective

This PhD research project aims to clarify the contribution of histone modifications to gene expression in two grapevine genotypes (M4 and 101.14) grown in controlled conditions. To reach this objective we used the most recently and high performance NGS techniques that were applied for sequencing of mRNA (mRNA-seq) and DNA immunoprecipitated with antibodies against acetylation of lysine 9 of histone 3 and trimethylation of lysine 4 of the same histone (ChIP-seq).

Nowadays mRNA-seq has become a routine analysis for the evaluation of transcriptome profile for a large number of plant species, while the use of ChIP-seq is still limited to plant model species.

During this PhD a lot of work was done for the development of a chromatin immunoprecipitation protocol suitable for the sequencing of chromatin extracted and immunoprecipitated from grapevine.

In this study we clarified and demonstrated the correlation between histone modifications and gene expression for the first time in grapevine.

5 References

1000 Genomes Project Consortium, Abecasis GR, Altshuler D et al (2010) A map of human genome variation from population-scale sequencing. Nature 467(7319):1061-1073

Alsina MM, Smart DR, Bauerle T et al (2011) Seasonal changes of whole root system conductance by a drought-tolerant grape root system. J Exp Bot 62(1):99-109

Ausin I, Alonso-Blanco C, Jarillo JA, Ruiz-Garcia L, Martinez-Zapater JM(2004) Regulation of flowering time by FVE, a retinoblastoma-associated protein. Nat Genet 36(2):162-166

Baumbusch LO, Thorstensen T, Krauss V et al (2001) The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res 29(21):4319-4333

Benhamed M, Bertrand C, Servet C, Zhou DX(2006) Arabidopsis GCN5, HD1, and TAF1/HAF2 interact to regulate histone acetylation required for light-responsive gene expression. Plant Cell 18(11):2893-2903

Berger SL(2007) The complex language of chromatin regulation during transcription. Nature 447(7143):407-412

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE(2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One 3(9):e3156

Bernstein E, Caudy AA, Hammond SM, Hannon GJ(2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409(6818):363-366

Berr A, McCallum EJ, Menard R et al (2010) Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development. Plant Cell 22(10):3232-3248

Birchler JA, Gao Z, Sharma A, Presting GG, Han F(2011) Epigenetic aspects of centromere function in plants. Curr Opin Plant Biol 14(2):217-222

Boeger H, Griesenbeck J, Strattan JS, Kornberg RD(2003) Nucleosomes unfold completely at a transcriptionally active promoter. Mol Cell 11(6):1587-1598

Bouche N, Lauressergues D, Gasciolli V, Vaucheret H(2006) An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. EMBO J 25(14):3347-3356

Boyer JS(1982) Plant productivity and environment. Science 218(4571):443-448

Buermans HP, den Dunnen JT(2014) Next generation sequencing technology: Advances and applications. Biochim Biophys Acta 1842(10):1932-1941

Castel SE, Martienssen RA(2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet 14(2):100-112

Chan SW, Henderson IR, Jacobsen SE(2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet 6(5):351-360

Chen X, Hu Y, Zhou DX(2011) Epigenetic gene regulation by plant Jumonji group of histone demethylase. Biochim Biophys Acta 1809(8):421-426

Chen LT, Luo M, Wang YY, Wu K(2010) Involvement of Arabidopsis histone deacetylase HDA6 in ABA and salt stress response. J Exp Bot 61(12):3345-3353

Cheung P, Tanner KG, Cheung WL, Sassone-Corsi P, Denu JM, Allis CD(2000) Synergistic coupling of histone H3 phosphorylation and acetylation in response to epidermal growth factor stimulation. Mol Cell 5(6):905-915

Clarke S(1993) Protein methylation. Curr Opin Cell Biol 5(6):977-983

Cokus SJ, Feng S, Zhang X et al (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452(7184):215-219

Collas P(2010) The current state of chromatin immunoprecipitation. Mol Biotechnol 45(1):87-100

Corso M, Vannozzi A, Maza E, Vitulo N, Meggio F, Pitacco A, Telatin A, D'Angelo M, Schiavon R, Negri AS, Prinsi B, Valle G, Ramina A, Bouzayen M, Bonghi C and Lucchin M (2015) Comprehensive transcripts profiling of two contrasting grapevine rootstock genotypes links phenylpropanoid pathways to enhanced drought tolerance. Submitted.

Covarrubias J, Rombolà AD (2013) Physiological and biochemical responses of the iron chlorosis tolerant grapevine rootstock 140 Ruggeri to iron deficiency and bicarbonate. Plant Soil 370:305–315

Cramer GR, Ergul A, Grimplet J et al (2007) Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. Funct Integr Genomics 7(2):111-134

Dai A(2013), Increasing drought under global warming in observations and models Nature Climate Change 3, 52–58

Dalmaso G, Eynard I (1990) Viticoltura moderna manuale pratico, IX Edizione

Damour G, Simonneau T, Cochard H, Urban L(2010) An overview of models of stomatal conductance at the leaf level. Plant Cell Environ 33(9):1419-1438

Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ(2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. J Mol Biol 319(5):1097-1113

Dawson BA, Herman T, Haas AL, Lough J(1991) Affinity isolation of active murine erythroleukemia cell chromatin: uniform distribution of ubiquitinated histone H2A between active and inactive fractions. J Cell Biochem 46(2):166-173

Deal RB, Henikoff S(2011) Histone variants and modifications in plant gene regulation. Curr Opin Plant Biol 14(2):116-122

Deleris A, Gallego-Bartolome J, Bao J, Kasschau KD, Carrington JC, Voinnet O(2006) Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. Science 313(5783):68-71

Du Z, Li H, Wei Q et al (2013) Genome-wide analysis of histone modifications: H3K4me2, H3K4me3, H3K9ac, and H3K27ac in Oryza sativa L. Japonica. Mol Plant 6(5):1463-1472

Eichten SR, Schmitz RJ, Springer NM(2014) Epigenetics: Beyond Chromatin Modifications and Complex Genetic Regulation. Plant Physiol 165(3):933-947

El-Shami M, Pontier D, Lahmy S et al (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. Genes Dev 21(20):2539-2544

Filion GJ, van Bemmel JG, Braunschweig U et al (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143(2):212-224

Fisarakis I, Chartzoulakis K, Stavrakas D (2001) Response of Sultana vines (V. vinifera L.) on six rootstocks to NaCl salinity exposure and recovery. Agric Water Manage 51: 13–27

Flexas J, Baron M, Bota J et al (2009) Photosynthesis limitations during water stress acclimation and recovery in the drought-adapted Vitis hybrid Richter-110 (V. berlandierixV. rupestris). J Exp Bot 60(8):2361-2377

Frommer M, McDonald LE, Millar DS et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 89(5):1827-1831

Frumkin D, Wasserstrom A, Kaplan S, Feige U, Shapiro E(2005) Genomic variability within an organism exposes its cell lineage tree. PLoS Comput Biol 1(5):e50

Fuchs J, Demidov D, Houben A, Schubert I(2006) Chromosomal histone modification patterns--from conservation to diversity. Trends Plant Sci 11(4):199-208

Fyodorov DV, Kadonaga JT(2001) The many faces of chromatin remodeling: SWItching beyond transcription. Cell 106(5):523-525

Gambetta GA, Manuck CM, Drucker ST et al (2012) The relationship between root hydraulics and scion vigour across Vitis rootstocks: what role do root aquaporins play? J Exp Bot 63(18):6445-6455

Gilchrist DA, Fargo DC, Adelman K(2009) Using ChIP-chip and ChIP-seq to study the regulation of gene expression: genome-wide localization studies reveal widespread regulation of transcription elongation. Methods 48(4):398-408

Grant-Downton RT, Dickinson HG(2005) Epigenetics and its implications for plant biology. 1. The epigenetic network in plants. Ann Bot 96(7):1143-1164

Grant-Downton RT, Dickinson HG(2006) Epigenetics and its implications for plant biology 2. The 'epigenetic epiphany': epigenetics, evolution and beyond. Ann Bot 97(1):11-27

Gregory PJ, Atkinson CJ, Bengough AG et al (2013) Contributions of roots and rootstocks to sustainable, intensified crop production. J Exp Bot 64(5):1209-1222

Ha M, Ng DW, Li WH, Chen ZJ(2011) Coordinated histone modifications are associated with gene expression variation within and between species. Genome Res 21(4):590-598

Hayatsu H, Wataya Y, Kazushige K(1970) The addition of sodium bisulfite to uracil and to cytosine. J Am Chem Soc 92(3):724-726

He G, Zhu X, Elling AA et al (2010) Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell 22(1):17-33

He S, Yan S, Wang P et al (2014) Comparative analysis of genome-wide chromosomal histone modification patterns in maize cultivars and their wild relatives. PLoS One 9(5):e97364

Henderson IR, Jacobsen SE(2007) Epigenetic inheritance in plants. Nature 447(7143):418-424

Henikoff S(2008) Nucleosome destabilization in the epigenetic regulation of gene expression. Nat Rev Genet 9(1):15-26

Henikoff S, Comai L(1998) A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in Arabidopsis. Genetics 149(1):307-318

Henikoff S, Shilatifard A(2011) Histone modification: cause or cog? Trends Genet 27(10):389-396

International Human Genome Sequencing Consortium(2004) Finishing the euchromatic sequence of the human genome. Nature 431(7011):931-945

Jackson JP, Johnson L, Jasencakova Z et al (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma 112(6):308-315

Jackson JP, Lindroth AM, Cao X, Jacobsen SE(2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature 416(6880):556-560

Jaillon O, Aury JM, Noel B et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463-467

Jin C, Zang C, Wei G et al (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet 41(8):941-945

Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316(5830):1497-1502

Johnson L, Mollah S, Garcia BA et al (2004) Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res 32(22):6511-6518

Kharchenko PV, Alekseyenko AA, Schwartz YB et al (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471(7339):480-485

Kim JM, To TK, Ishida J et al (2008) Alterations of lysine modifications on the histone H3 N-tail under drought stress conditions in Arabidopsis thaliana. Plant Cell Physiol 49(10):1580-1588

Kingsmore SF, Saunders CJ(2011) Deep sequencing of patient genomes for disease diagnosis: when will it become routine? Sci Transl Med 3(87):87ps23

Ko JH, Mitina I, Tamada Y et al (2010) Growth habit determination by the balance of histone methylation activities in Arabidopsis. EMBO J 29(18):3208-3215

Kornberg RD(1974) Chromatin structure: a repeating unit of histones and DNA. Science 184(4139):868-871

Kouzarides T(2007) Chromatin modifications and their function. Cell 128(4):693-705

Kreimeyer A, Wielckens K, Adamietz P, Hilz H(1984) DNA repair-associated ADP-ribosylation in vivo. Modification of histone H1 differs from that of the principal acceptor proteins. J Biol Chem 259(2):890-896

Kurdistani SK, Tavazoie S, Grunstein M(2004) Mapping global histone acetylation patterns to gene expression. Cell 117(6):721-733

Lafos M, Kroll P, Hohenstatt ML, Thorpe FL, Clarenz O, Schubert D(2011) Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. PLoS Genet 7(4):e1002040

Lauria M, Rossi V(2011) Epigenetic control of gene regulation in plants. Biochim Biophys Acta 1809(8):369-378

Law JA, Du J, Hale CJ et al (2013) Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature 498(7454):385-389

Law JA, Jacobsen SE(2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11(3):204-220

Li B, Carey M, Workman JL(2007) The role of chromatin during transcription. Cell 128(4):707-719

Li W, Lin YC, Li Q et al (2014) A robust chromatin immunoprecipitation protocol for studying transcription factor-DNA interactions and histone modifications in wood-forming tissue. Nat Protoc 9(9):2180-2193

Liu C, Lu F, Cui X, Cao X(2010) Histone methylation in higher plants. Annu Rev Plant Biol 61:395-420

Loidl P(1994) Histone acetylation: facts and questions. Chromosoma 103(7):441-449

Luger K(2003) Structure and dynamic behavior of nucleosomes. Curr Opin Genet Dev 13(2):127-135

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ(1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389(6648):251-260

Lusser A, Kolle D, Loidl P(2001) Histone acetylation: lessons from the plant kingdom. Trends Plant Sci 6(2):59-65

Makarevitch I, Eichten SR, Briskine R et al (2013) Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. Plant Cell 25(3):780-793

Marguerit E, Brendel O, Lebon E, Van Leeuwen C, Ollat N(2012) Rootstock control of scion transpiration and its acclimation to water deficit are controlled by different genes. New Phytol 194(2):416-429

Margulies M, Egholm M, Altman WE et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057):376-380

Mathieu O, Probst AV, Paszkowski J(2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J 24(15):2783-2791

Matzke MA, Mosher RA(2014) RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet 15(6):394-408

Maxam AM, Gilbert W(1992) A new method for sequencing DNA. 1977. Biotechnology 24:99-103

McGovern PE (2003) Ancient Wine: The Search for the Origins of Viniculture

Meggio F, Prinsi B, Negri AS, Simone Di Lorenzo G, Lucchini G, Pitacco A, Failla O, Scienza A, Cocucci M and Espen L Biochemical and physiological responses of two grapevine rootstock genotypes to drought and salt treatments. Australian Journal of Grape and Wine Research 20:310-323

Metzger E, Wissmann M, Yin N et al (2005) LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature 437(7057):436-439

Michael TP, Jackson S (2013) The first 50 plant genomes. Plant gen vol 6, n° 2

Mito Y, Henikoff JG, Henikoff S(2005) Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet 37(10):1090-1097

Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C(2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science 303(5656):343-348

Mizzen CA, Yang XJ, Kokubo T et al (1996) The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. Cell 87(7):1261-1270

Morgante M, De Paoli E, Radovic S(2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10(2):149-155

Morozova O, Marra MA(2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92(5):255-264

Nelson CJ, Santos-Rosa H, Kouzarides T(2006) Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell 126(5):905-916

Olmo HP (1995) Grapes. In Evolution of crop plants Grapes. Ed. Smartt J, Simmonds NW, 485-490. Harlow: Longman 2nd editon

O'Neill LP, Turner BM(1996) Immunoprecipitation of chromatin. Methods Enzymol 274:189-197

Pandey R, Muller A, Napoli CA et al (2002) Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. Nucleic Acids Res 30(23):5036-5055

Pien S, Grossniklaus U(2007) Polycomb group and trithorax group proteins in Arabidopsis. Biochim Biophys Acta 1769(5-6):375-382

Pikaard CS, Haag JR, Pontes OM, Blevins T, Cocklin R(2012) A transcription fork model for Pol IV and Pol V-dependent RNA-directed DNA methylation. Cold Spring Harb Symp Quant Biol 77:205-212

Pokholok DK, Harbison CT, Levine S et al (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122(4):517-527

Raisner RM, Madhani HD(2006) Patterning chromatin: form and function for H2A.Z variant nucleosomes. Curr Opin Genet Dev 16(2):119-124

Rea S, Eisenhaber F, O'Carroll D et al (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406(6796):593-599

Reinke H, Horz W(2003) Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. Mol Cell 11(6):1599-1607

Ricardi MM, Gonzalez RM, Iusem ND(2010) Protocol: fine-tuning of a Chromatin Immunoprecipitation (ChIP) protocol in tomato. Plant Methods 6:11-4811-6-11

Ricardi MM, Gonzalez RM, Zhong S et al (2014) Genome-wide data (ChIP-seq) enabled identification of cell wall-related and aquaporin genes as targets of tomato ASR1, a drought stress-responsive transcription factor. BMC Plant Biol 14:29-2229-14-29

Rivas FV, Tolia NH, Song JJ et al (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol 12(4):340-349

Roh TY, Cuddapah S, Zhao K(2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev 19(5):542-552

Roth SY, Denu JM, Allis CD(2001) Histone acetyltransferases. Annu Rev Biochem 70:81-120

Roudier F, Teixeira FK, Colot V(2009) Chromatin indexing in Arabidopsis: an epigenomic tale of tails and more. Trends Genet 25(11):511-517

Roudier F, Ahmed I, Berard C et al (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J 30(10):1928-1938

Rybaczek D, Bodys A, Maszewski J(2007) H2AX foci in late S/G2- and M-phase cells after hydroxyurea- and aphidicolin-induced DNA replication stress in Vicia. Histochem Cell Biol 128(3):227-241

Saleh A, Alvarez-Venegas R, Yilmaz M et al (2008) The highly similar Arabidopsis homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions. Plant Cell 20(3):568-579

Sanger F, Nicklen S, Coulson AR(1992) DNA sequencing with chain-terminating inhibitors. 1977. Biotechnology 24:104-108

Schloss JA(2008) How to get genomes at one ten-thousandth the cost. Nat Biotechnol 26(10):1113-1115

Scienza A et al, (2013), Nuovi portinnesti della vite. L'informatore agrario 14/2013: 36-44

Sequeira-Mendes J, Araguez I, Peiro R et al (2014) The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell 26(6):2351-2366

Shapiro R, DiFate V, Welcher M(1974) Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction. J Am Chem Soc 96(3):906-912

Shi Y, Lan F, Matson C et al (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119(7):941-953

Shiio Y, Eisenman RN(2003) Histone sumoylation is associated with transcriptional repression. Proc Natl Acad Sci U S A 100(23):13225-13230

Sinha I, Wiren M, Ekwall K(2006) Genome-wide patterns of histone modifications in fission yeast. Chromosome Res 14(1):95-105

Sokol A, Kwiatkowska A, Jerzmanowski A, Prymakowska-Bosak M(2007) Up-regulation of stress-inducible genes in tobacco and Arabidopsis cells in response to abiotic stresses and ABA treatment correlates with dynamic changes in histone H3 and H4 modifications. Planta 227(1):245-254

Solomon MJ, Larsen PL, Varshavsky A(1988) Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. Cell 53(6):937-947

Strahl BD, Allis CD(2000) The language of covalent histone modifications. Nature 403(6765):41-45

Stroud H, Otero S, Desvoyes B, Ramirez-Parra E, Jacobsen SE, Gutierrez C(2012) Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc Natl Acad Sci U S A 109(14):5370-5375

Sun Q, Zhou DX(2008) Rice jmjC domain-containing gene JMJ706 encodes H3K9 demethylase required for floral organ development. Proc Natl Acad Sci U S A 105(36):13679-13684

Suto RK, Clarkson MJ, Tremethick DJ, Luger K(2000) Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. Nat Struct Biol 7(12):1121-1124

Takeuchi T, Yamazaki Y, Katoh-Fukui Y et al (1995) Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. Genes Dev 9(10):1211-1222

Talbert PB, Henikoff S(2010) Histone variants--ancient wrap artists of the epigenome. Nat Rev Mol Cell Biol 11(4):264-275

Tamaru H, Selker EU(2001) A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature 414(6861):277-283

The Arabidopsis Genome Initiative (2000), Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796-815

Timinszky G, Till S, Hassa PO et al (2009) A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. Nat Struct Mol Biol 16(9):923-929

Tsukada Y, Fang J, Erdjument-Bromage H et al (2006) Histone demethylation by a family of JmjC domain-containing proteins. Nature 439(7078):811-816

Tsuji H, Saika H, Tsutsumi N, Hirai A, Nakazono M(2006) Dynamic and reversible changes in histone H3-Lys4 methylation and H3 acetylation occurring at submergence-inducible genes in rice. Plant Cell Physiol 47(7):995-1003

Vaillant I, Paszkowski J(2007) Role of histone and DNA methylation in gene regulation. Curr Opin Plant Biol 10(5):528-533

Valouev A, Ichikawa J, Tonthat T et al (2008) A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res 18(7):1051-1063

van Attikum H, Gasser SM(2009) Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol 19(5):207-217

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C(2014) Ten years of next-generation sequencing technology. Trends Genet 30(9):418-426

van Dijk K, Ding Y, Malkaram S et al (2010) Dynamic changes in genome-wide histone H3 lysine 4 methylation patterns in response to dehydration stress in Arabidopsis thaliana. BMC Plant Biol 10:238-2229-10-238

Vaucheret H(2008) Plant ARGONAUTES. Trends Plant Sci 13(7):350-358

Velasco R, Zharkikh A, Troggio M et al (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS One 2(12):e1326

Waddington CH(2012) The epigenotype. 1942. Int J Epidemiol 41(1):10-13

Walker RR, Blackmore DH, Clingeleffer PR, Correll RL (2002) Rootstock effects on salt tolerance of irrigated field-grown grapevines (Vitis vinifera L. cv. Sultana). 1. Yield and vigour inter-relationships. Australian Journal of Grape and Wine Research 8:3-14

Walker RR, Blackmore DH, Clingeleffer PR, Correll RL (2004) Rootstock effects on salt tolerance of irrigated field-grown grapevines (Vitis vinifera L. cv. Sultana). 2. Ion concentration in leaves and juice. Australian Journal of Grape and Wine Research 10:90-99.

Wang X, Elling AA, Li X et al (2009) Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. Plant Cell 21(4):1053-1069

Wang Z, Zang C, Cui K et al (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell 138(5):1019-1031

Watson JD, Crick FH(2003) A structure for deoxyribose nucleic acid. 1953. Nature 421(6921):397-8; discussion 396

Wierzbicki AT, Cocklin R, Mayampurath A et al (2012) Spatial and functional relationships among Pol V-associated loci, Pol IV-dependent siRNAs, and cytosine methylation in the Arabidopsis epigenome. Genes Dev 26(16):1825-1836

Wierzbicki AT, Haag JR, Pikaard CS(2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135(4):635-648

Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE(2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol 10(6):R62-2009-10-6-r62. Epub 2009 Jun 9

Zhang X, Clarenz O, Cokus S et al (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol 5(5):e129

Zhang X, Henderson IR, Lu C, Green PJ, Jacobsen SE(2007) Role of RNA polymerase IV in plant small RNA metabolism. Proc Natl Acad Sci U S A 104(11):4536-4541

Zhang X, Yazaki J, Sundaresan A et al (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126(6):1189-1201

Zheng B, Chen X(2011) Dynamics of histone H3 lysine 27 trimethylation in plant development. Curr Opin Plant Biol 14(2):123-129

Zhou J, Wang X, He K, Charron JB, Elling AA, Deng XW(2010) Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in Arabidopsis reveals correlation between multiple histone marks and gene expression. Plant Mol Biol 72(6):585-595

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S(2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456(7218):125-129

Chapter II

Transcriptome analyses of two grapevine rootstocks grown in vitro

1 Introduction

In the nineteenth century European viticulture was devastated by the introduction of phylloxera (*Daktuloshaira vitifoliae*) from North America. Ever since a new era for viticulture started out based on grafting of a scion of *V. vinifera* varieties, used commercially for wine production in Europe, onto rootstocks from the pest's origin. The introduction of root system from American or non *vinifera* species was initially focused on the protection of viticulture from the pest, but very soon it has been realized that grapevine rootstocks were not only capable to confer disease resistance, but they could imply a large range of advantages by influencing numerous physiological process at the scion level, such as biomass accumulation (Gregory *et al*., 2013), fruit quality (Walker *et al*., 2002, 2004) and nevertheless the response to many abiotic stresses (Marguerit *et al*., 2012; Meggio *et al.*, 2014). In this last field rootstocks were identified like carriers of many features, such as tolerance to salinity (Fisarakis *et al.*, 2001), ion deficit (Covarrubias and Rombolà, 2013) and drought (Gambetta *et al.*, 2012, Meggio *et al.*, 2014).

Nowadays water availability is one of the major limiting factor for viticulture (Cramer *et al.,* 2007; Flexas *et al.,* 2009; Chaves *et al.,* 2010) also considering that the most important wine-producing regions in the world are subjected to seasonal drought. The Intergovernmental Panel on Climate Change (IPCC, 2007 http://www.ipcc.ch/) climate models predict an increase of the aridity and water deficit in the future, then these changes could became the major limiting factors for grapevine production and wine quality. Generally, drought stress in plants is associated with many physiological and morphological changes over a spatial and temporal range (Chaves *et al.,* 2002). These effects could be described like lack of root growth (Sharp and Davies, 1979), reduced expansion of aerial organs and accumulation of osmotic compounds and ions (Cramer *et al.*, 2007), decrease in transpiration and photosynthesis (Chaves *et al.*, 2009; Chaves *et al.*, 2010) and activation of detoxifying processes. All these features are tightly related to the transcriptional regulation of a wide number of genes like reported in studies performed by Cramer and Tillett (Cramer *et al.*, 2007; Tillett *et al.*, 2011).

The enhanced pressure on water resources had to be translate into a reduction of the amount of water used for crops irrigation (Cominelli *et al.*, 2013). Grapevine is already well adapted to semi-arid climate regions, like Mediterranean area, and it is generally considered able to cope with relatively water deficit. The large and deep root system, together with physiological drought avoidance mechanisms, such as stomatal control of transpiration, xylem embolism (Lovisolo *et al.,* 2002) and the ability to adjust osmotic pressure, mean that these plants are also able to grow in sub-optimal water conditions. Nevertheless a large amount of vineyard are located in areas where seasonal drought coincides with grapevine grooving season and the effect of water deficit added to high air temperature and evaporative demand could limits the yield and produce negative effects on berries and, consequently, on wine quality. (Chaves *et al.,* 2009; Flexas *et al.,* 2009; Chaves *et al.,* 2010)

Because the cultivation of grapevine is very tightly related to areas that could be subjected to drought stress, its photosynthetic process is quite resistant to mild water stress (Souza *et al.,* 2003; Chaves *et al.,* 2009). When the drought became severe, net CO_2 assimilation (A_n) and other metabolic processes operating in the mesophyll are inhibited and water use efficiency declines. Changes in the photochemistry of chloroplasts are due to the imbalance between light capture and its utilization and may involve the generation of reactive oxygen species (ROS) such as H_2O_2, O_2^-, -OH, RO_2, and NO. These ROS are responsible for most of the oxidative damage in biological systems and cellular components (Apel and Hirt, 2004; Kar, 2011).

A wide range of response to abiotic stresses are mediated by phytormones (Santner and Estelle, 2009; Peleg and Blumwald, 2011; Kelley and Estelle, 2012). Abscissic acid (ABA) is probably the most studied stress-responsive phytohormone, especially for what concerns the plant response to water deficit (Novikova *et al.,* 2009; Fujita *et al.*, 2011; Qin *et al.*, 2011). Its biosynthesis and accumulation represent one of the fastest plant response to abiotic stress, ABA acts triggering ABA-responsive genes and inducing stomatal closure to reduce water loss and limiting cellular growth (Lata *et al*., 2011; Peleg and Blumwald, 2011). Recently many different studies on drought stress have revealed some others hormones related to drought response such as

ethylene, jasmonates (JAs), auxins, gibberellins (GAs), salicylic acid (SA) and brassinosteroids (BRs) (Peleg and Blumwald, 2011). The adaptation of plant to water stress could be considered like a concerted action of these phythormones (Kohli *et al.*, 2013) that not only interact each other but also with ROS that could act like secondary messangers in the adaptation mechanism (Kar 2011). It is well know that ABA-mediated stomal closure is regulated by H_2O_2 that acts on Ca^{2+} levels and inactivating protein phosphatase 2C (Meinhard *et al.*, 2002). Stomatal closure is also mediated by ethylene via ETR1, one of its receptors, which is involved in H_2O_2-sensing (Desikan *et al.*, 2005).

Meggio and colleagues recently published a biochemical and physiological study of a commercial rootstock 101.14 (*V. riparia x V. rupestris*), compared to an experimental one, M4 [(*V. vinifera* x *V. berlandieri*) x *V. berlandieri* x cv Resseguier n.1] grown under controlled and stressed conditions. The rootstock M4 was established in 1985 by the DiSAA research group of University of Milan and was selected for its high tolerance to water deficit and salt exposure. Rootstock M4 was the main characters of the project Ager-Serres 2010-2105 (http://users.unimi.it/serres/index.html). This project focused the attention not only on biochemical and physiological aspects of rootstock drought response, but go in deep with a genome and transcriptome analysis performed using techniques of Next Generation Sequencing (NGS) (Corso *et al*., submitted). The main goal of the project was the identification of molecular markers suitable for the selection of rootstocks showing high performances in stress conditions. It is within Ager-Serres project that our research takes place with the objective to establish the relation between transcriptional profile and histone modifications in the two rootstocks under consideration.

101.14 and M4 are both interspecific hybrids (*V. riparia* x *V. rupestris* and [(*V. vinifera* x *V. berlandieri*) x *V. berlandieri* cv. Resseguier n. 1] and this raised the question about the genetic differences between PN40024 (Jailon *et al*., 2007) and the hybrids. Could genetic differences compromise the robustness of mRNA-seq analyses at the level of reads-mapping? In order to

refute if PN40024 could be a good reference for the two rootstocks their genomes were sequenced and analyzed for SNPs distribution, In/Del and genes prediction.

In this chapter we will report the data emerged from genomic analyses of the two rootstocks (Corso *et al*., submitted) and we will study the transcriptome profiles of two pools of plants, one for each genotype, that have been grown *in vitro* under controlled and stable conditions into a growth camber. The object of this analysis is the comparison of the transcriptome profiles of the two rootstocks grown under controlled conditions for the individuation of genes that were differentially expressed (DEGs) between 101.14 and M4. Furthermore we want to focus on the comparison between differentially expressed genes present in plants grown in *in vitro* and in greenhouse. However we have to remember that in our work the most important consideration that could be done on transcriptome profiles is the subsequent correlation with ChIP-seq analysis (see Chapter IV). In this study indeed transcriptome analysis that is functional to the identification of the correlation between gene expression and histone modifications.

2 Materials and methods

2.1 *In vitro* plant material and experimental design

In the month of June 2013 explants with two or three internodes of rootstock 101.14 (*V. riparia* x *V. rupestris*) and M4 [(*V. vinifera* x *V. berlandieri*) x *V. berlandieri* cv. Resseguier n. 1] were collected from two-years old plants grown into the greenhouse of the Azienda agraria sperimentale L. Toniolo in Legnaro, PD. The explants were treated with a sterilization procedure in order to make them suitable for *in vitro* grown. The handling and the sterilization of the explants were performed in horizontal fume hood (Faster Bio48). The explants were cut to produce fragment with one internode each that were treated for ten minutes with a solution of commercial sodium hypoclorite *in ratio* 1:2 with sterilized milliQ water (Milli-Q Academic, Millipore). After this step the materials were washed with sterile milliQ water in three subsequent steps of five minutes each. After these washes the explants were planted in cylindrical hermetic vessel of crystal-clear polypropylene with breathing strip (vessel base: 90mm of diameter, vessel top and cover: 140mm, vessel height: 115 mm, Duchefa) filled with 150 ml of "*in vitro* medium" (Figure 2.1) .The composition of the medium and of its components were reported in Table 2.1-5. Plants were placed into a growth chamber in controlled condition of temperature (25°C±1°C) and photoperiod (light/dark 16h/8h), after few weeks the plants started to produce new buds and roots. When new leaves reached the diameter of about three centimeter, about 6 week after the start of *in vitro* growth, new explants were used for subsequent micropropagation .

In vitro Medium	
Stock solution	**Final Concentration**
10x SM macrosalt	1x
1000x MS microsalt	1x
1000x B5 vitamins	1x
200x FeEDTA	1x
Sucrose	1.5%
Agar	0.8%
2000x PPM	1x
1000x Cefotaxime	1x
pH	5.7

Table 2.1: Composition of *in vitro* medium

10x SM Macrosalt		
Substance	**Chemical Formula**	**gr/l**
Ammonium nitrate	NH_4NO_3	1.60
Calcium chloride	$CaCl_2 \cdot 2H_2O$	4.10
Magnesium sulfate	$MgSO_4 \cdot 7H_2O$	3.70
Potassium nitrate	KNO_3	23.25
Potassium monophosphate	KH_2PO_4	3.40

Table 2.2: Composition of SM Macrosalt stock solution

1000x SM Microsalt		
Substance	**Chemical Formula**	**gr/l**
Boric acid	H_3BO_3	6.20
Manganese sulfate	$MnSO_4 \cdot 4H_2O$	22.30
Zinc sulfate	$ZnSO_4 \cdot 7H_2O$	4.30
Potassium iodide	KI	0.42
Sodium molybdate	$Na_2MoO_4 \cdot 2H_2O$	0.13
Cuper sulphate	$CuSO_4 \cdot 5H_2O$	$12.5 \cdot 10^{-3}$
Cobalt chloride	$CoCl_2 \cdot 6H_2O$	$12.5 \cdot 10^{-3}$

Table 2.3: Composition of SM Microsalt stock solution

1000x B5 Vitamins	
Substance	**gr/l**
Myo-Inositol	100.00
Thiamine-Hcl	10.00
Nicotinic Acid	1.00
Pyridoxine HCl	1.00

Table 2.4: Composition of B5 Vitamin stock solution

200x FeEDTA	
Substance	**gr/l**
$Na_2EDTA{\cdot}2H_2O$	7.44
$FeSO_4{\cdot}7H_2O$	0.46

Table 2.5: Composition of FeEDTA stock solution

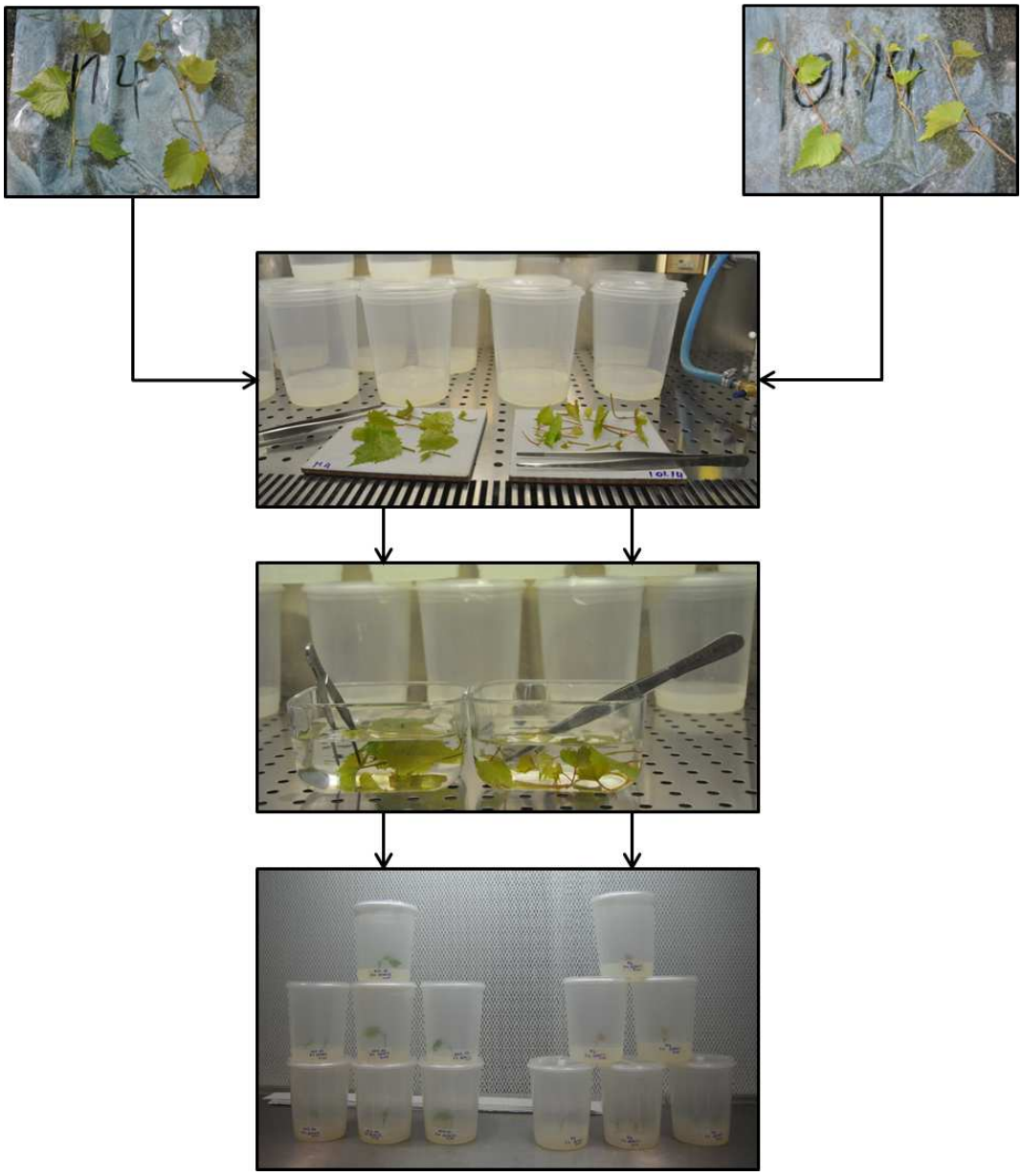


Figure 2.1: Schematic representation of plants explant and *in vitro* propagation

In vitro plants that were subsequently propagated for eight months were used to produce explants with two internodes used in this experiment. Twenty five plants for each rootstock where grown *in vitro* for three months in condition of controlled temperature (25°C±1°C) and photoperiod (light/dark 16h/8h). After this time the plants appeared with a good leaf area and were ready to be sampled. In April 2014 leaf material of twenty three plants of 101.14 and twenty five plants of M4 were collected in two biological replicate pools and frozen in liquid nitrogen. The tissues were maintained at -80°C until the moment of the utilize were mRNA was extracted and sequenced.

2.2 Transcriptome analysis of the two rootstocks

Total RNA was extracted from frozen grapevine leaves using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) according to the manufacturer's user guide. Total RNA was analyzed with NanoDrop instrument (Thermo Fisher) and, once established the goodness of the RNA extracted, it has been frozen and send to the sequencing center, IGA Technology Services, where the RNA samples were treated for library preparation using the Illumina mRNA Sample Prep kit v2.0. The polyA-mRNA were fragmented 3 minutes at 94°C and purification steps were performed using 1X Agencourt AMPure XP beads (Beckman Coulter). The expression profiles of the two genotypes were analyzed by Next Generation Sequencing (NGS) technologies at IGA technology services, Udine. Sequencing was made with Illumina technology HiSeq2000, single-ends reads of 50bp of length. The sequencing was performed on two biological replicas for genotype, called A and B. Reads from single-read runs were put through the sequencing pipeline of IGA technology Services . The base calling was performed using the Illumina Pipeline with default parameters. The trimming of the sequences were performed with ERNE v 1.4 with minimum value used by Mott-like trimming 20, minimum mean value to accept a trimmed sequence 20 and minimum sequence length after trimming 40 (Vezzi *et al.,* 2012). The removal of the adapter sequences were performed with standard settings of Cutadapt software (https://pypi.python.org/pypi/cutadapt/). The mapping and the annotation on the reference genome PN40024 were performed with Tophat2 v2.0.6 using default parameters. Differential

expression analysis was performed by Cufflinks software v2.0.2 using all default parameters (Trapnell *et al.,* 2012). The transcriptome analyses were performed on two biological replicas for each genotype and the value of the two replicas were mediated for the subsequent analyses.

Transcriptome analyses on plants grown in greenhouse were performed within Ager-Serres project and were described in details in Corso (2014). In this work we used their data, already analyzed, to relate DEGs observed under *in vitro* growth with DEGs emerged from greenhouse plants identifying the genes that could be related to genotype rather than grown condition.

2.3 Differentially Expressed Genes and Gene Ontology

In order to make our analysis more robust we selected genes with value of FPKM, mediated for the two biological replicas, superior of 1 for at least one of the genotype in exam. Genes in which the logarithm to the base 2 of the ratio between FPKM of the genes in the two genotypes were major or equal to 1 or minor or equal to -1($\log_2 \frac{101.14}{M4} \geq 1 \; or \leq -1$) were classified like differentially expressed genes (DEGs).

In order to classify the genes differentially expressed between the two genotypes grown in *in vitro* we performed a Gene Ontology (GO) terms analysis. GO terms studied were the ones for molecular function and biological processes that were considered as defined by Gene Ontology Consortium. Molecular function GO terms describes activities that occur at the molecular level and represented these activities rather than the entities that perform the actions while biological process GO terms describes a series of events accomplished by one or more organized assemblies of molecular functions (http://geneontology.org/page/ontology-documentation).

Grapevine GO terms for molecular function and process were retrieved from website of grape Genome browser developed by CRIBI Biotechnology Services, University of Padova (http://genomes.cribi.unipd.it/DATA/V1/ANNOTATION/GO.tab) and imported in Blast2GO v2.8.0 and each gene were associate to its Gene Ontology terms (Gotz *et al*., 2008).

3 Results

3.1 mRNA sequencing and reads mapping on reference genome

mRNA sequencing produced about 89 million of raw reads for M4 (39.5 Million (M) of raw reads for M4_A and 49.8 M for M4_B) and 76 million of raw reads for 101.14 (35.3 M of raw reads for 101.14_A and 40.4 M for 101.14_B). Once trimmed about 83% of the single reads aligned on reference genome PN40024 12x V1 (Jaillon *et al.,* 2007) producing a number of unique mapping reads between 29 and 41 million depending on the sample (data are reported in Table 2.6). 101.14 and M4 reads were mapped onto *Vitis vinifera* reference genome PN40024 (Jailon *et al.*, 2007) taking into account to results obtained by Corso (2014) by comparing genome PN40024, 101.14 and M4 genomes. Corso (2014) founds an average of about one variant every 200 bases both in M4 and in 101.14, suggesting that the PN40024 genome should be a suitable reference for mapping 101.14 and M4 reads without compromising the robustness of the analysis.

Sample	Raw reads [M]	Trimmed reads [M]	Aligned reads [%]
M4_A	39.5	39.2	84.0
M4_B	49.8	49.4	82.8
101.14_A	35.3	35.0	83.2
101.14_B	40.4	40.1	82.8

Table 2.6: mRNA-Seq data obtained from each replica of the two rootstock genotypes. The number of reads are reported in million

The values of FPKM of the two replicas for each genotype were mediate and allowed to the identification of 29 693 genes. In order to make our analysis more robust, we selected genes with value of mean FPKM superior to one for at least one of the rootstocks, the number of elements considered decreased in this way to 18 976 genes.

3.2 Differentially Expressed Genes and Ontology analyses

3.2.1 Differentially Expressed Genes (DEGs)

The identification of Differentially Expressed Genes (DEGs) in the two rootstocks grown *in vitro*, were performed by calculating the logarithm to the base 2 of the ratio between FPKM of the genes in the two genotypes, called from hereinafter Fold Change (FC, $FC = log_2 \frac{101.14}{M4}$). Genes that had values of FC major or equal to 1 or minor or equal to -1 were considered differentially expressed genes. Genes with FC values major to 0 were overexpressed in the commercial susceptible rootstock 101.14 while FC values negatives identified genes that were overexpressed in the genotype M4. We have chosen the approach based on the FC rather than more stringent approaches that considered also p-value and FDR to obtain a list of genes wider that the one obtained with the other restrictions. In this way we didn't risk to loose many interesting and relevant genes.

This analysis allowed the identification of 3205 DEGs of which 1603 overexpressed in 101.14 and 1602 overexpressed in M4.

3.2.2 Gene Ontology terms analyses

In order to classify the differentially expressed genes between the two genotypes grown in the same controlled conditions we performed a Gene Ontology (GO) terms analysis as described in materials and methods. The analyses for molecular function GO terms of DEGs are reported like supplementary materials (Section 6.1 Analyses of molecular function GO terms) while analyses for biological process are described below.

Analysis of GO terms relative to biological process identified in 101.14 up regulated genes five most specific GO terms over represented and nine GO terms underrepresented respect of the reference dataset (Blast2GO) The most specific GO terms are represented in Figure 2.2 and they are listed in Table 2.7. GO terms over represented and underrepresented in 101.14 up regulated DEGs were distributed on 222 genes and 15 genes respectively.

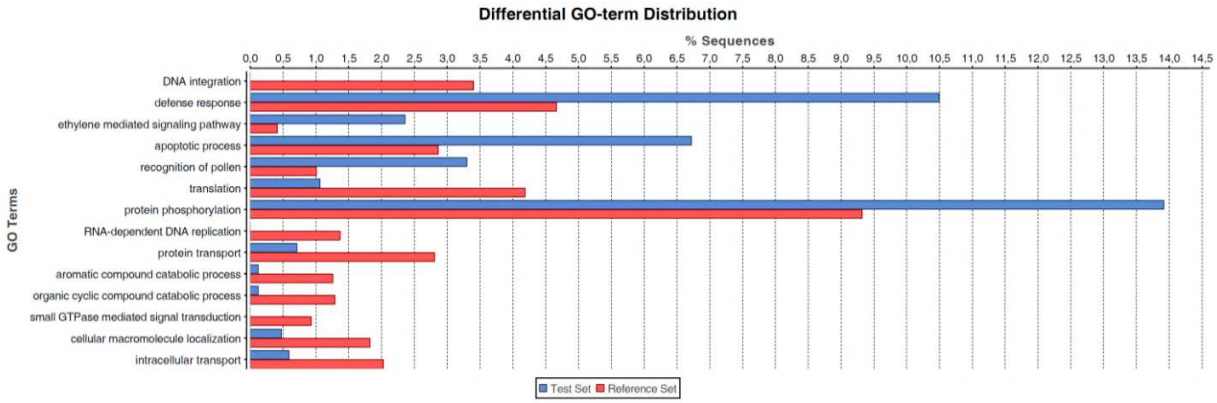


Figure 2.2: Bar chart of 101.14 overexpressed genes most specific biological process GO terms

GO terms analysis for biological processes of DEGs overexpressed in M4 identified ten most specific GO terms over represented in these elements in respect to the reference dataset and eight underrepresented (Figure 2.3). The most specific GO terms are listed in Table 2.7. GO terms over represented and underrepresented in M4 upregulated DEGs were distributed on 327 genes and 97 genes respectively.

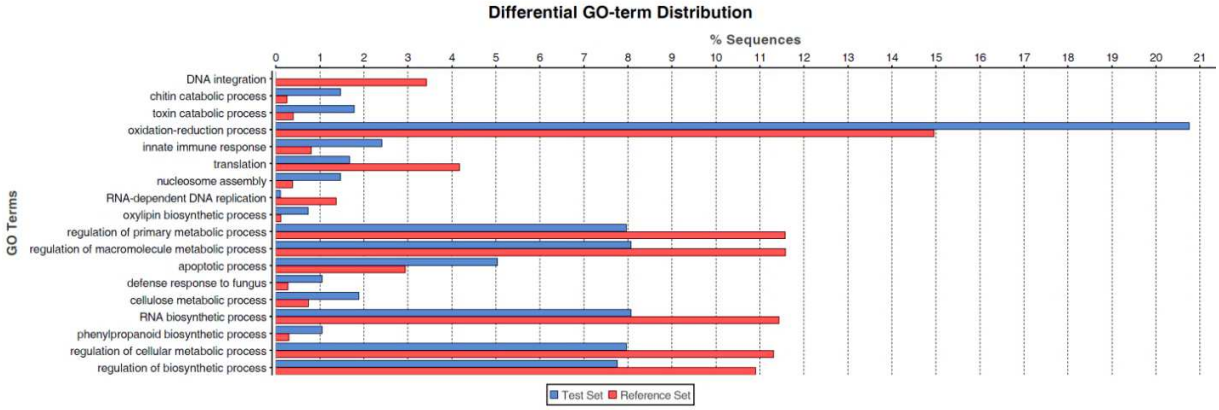


Figure 2.3: Bar chart of M4 overexpressed genes most specific biological process GO terms

101.14_DEGs_most specific biological process GO terms over/under represented	
GO:0006468	protein phosphorylation
GO:0006915	apoptotic process
GO:0006952	defense response
GO:0009873	ethylene mediated signaling pathway
GO:0048544	recognition of pollen
GO:0006278	*RNA-dependent DNA replication*
GO:0006412	*translation*
GO:0007264	*small GTPase mediated signal transduction*
GO:0015031	*protein transport*
GO:0015074	*DNA integration*
GO:0019439	*aromatic compound catabolic process*
GO:0046907	*intracellular transport*
GO:0070727	*cellular macromolecule localization*
GO:1901361	*organic cyclic compound catabolic process*

M4_DEGs_most specific biological process GO terms over/under represented	
GO:0006032	chitin catabolic process
GO:0006334	nucleosome assembly
GO:0006915	apoptotic process
GO:0009407	toxin catabolic process
GO:0009699	phenylpropanoid biosynthetic process
GO:0030243	cellulose metabolic process
GO:0031408	oxylipin biosynthetic process
GO:0045087	innate immune response
GO:0050832	defense response to fungus
GO:0055114	oxidation-reduction process
GO:0006278	*RNA-dependent DNA replication*
GO:0006412	*translation*
GO:0009889	*regulation of biosynthetic process*
GO:0015074	*DNA integration*
GO:0031323	*regulation of cellular metabolic process*
GO:0032774	*RNA biosynthetic process*
GO:0060255	*regulation of macromolecule metabolic process*
GO:0080090	*regulation of primary metabolic process*

Table 2.7: The most represented biological process GO terms of *in vitro* 101.14 and M4 overexpressed DEGs over and underrepresented. Terms underrepresented are in italics.

Four biological process GO terms were present like most specific terms in both the genotypes and could be related to *in vitro* growth rather than difference between genotypes. These terms were DNA integration, translation and RNA-dependent DNA replication that were underrepresented respect to the reference set and the term relative to apoptotic process that was over represented.

3.3.3 Differentially Expressed Gene: correlation between *in vitro* and in greenhouse behavior of rootstocks

In order to evaluate if *in vitro* growth could be associated, in terms of gene expression, to the behavior of plants grown in greenhouse under controlled conditions, we compared the transcriptome profiles of *in vitro* plants, with the expression profiles obtained from a mRNA-seq approach of plants grown in greenhouse that were performed by Corso and colleagues within Ager-Serres project (Corso, 2014). For these analyses we considered genes identified like differentially expressed, with the same trend, in both the growth conditions. DEGs detected in greenhouse plants (called from hereinafter *in vivo*) were filtered for the identification of only the genes already considered in the mRNA-seq analyses of *in vitro* plants (18 976 genes). *In vivo* detected DEGs were 879: 399 overexpressed in 101.14 and 480 genes up regulated in M4. We

related 101.14 *in vitro* with *in vivo* DEGs and we identified 230 genes in 101.14 and 279 in M4 that were overexpressed in both the growth conditions. We saw that about 58% of *in vivo* DEGs (230/399 for 101.14 and 279/480 for M4) were differentially expressed also in *in vitro* condition (Figure 2.4). This data is quite interesting if we consider that when we analyzed different transcriptome profiles of plants grown in greenhouse at the same controlled conditions but in different time points, the percentage of overlapping DEGs was about 61% (239/399 for 101.14 and 302/480 for M4).

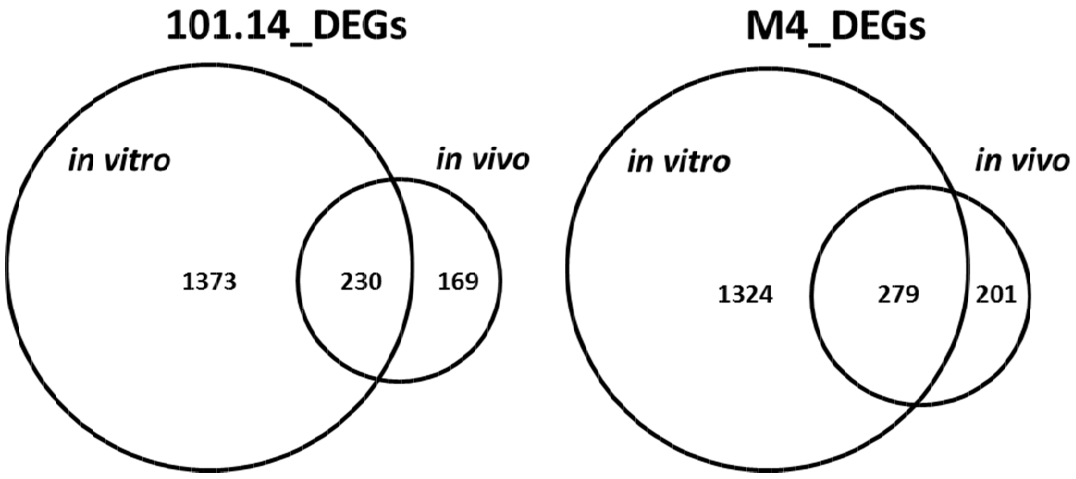


Figure 2.4: Venn's diagram of DEGs shared between *in vitro* and *in vivo* growth conditions.

At the light of these data we decided to evaluate if there was a relationship between the most specific GO terms identified in DEGs of *in vitro* plants and the data obtained from the plants grown in greenhouse.

GO terms analysis for biological process of DEGs overexpressed in 101.14 *in vivo* plants identified six most represented GO terms overrepresented and five underrepresented respect to the reference set. GO terms analysis for biological process of DEGs overexpressed in M4 rootstocks grown in greenhouse identified indeed nine most specific GO terms over represented and seven underrepresented respect to the reference dataset. The most specific biological process GO terms over and underrepresented for each genotype were listed in Table 2.8.

in vivo 101.14_overexpressed_DEGs_most specific biological process GO terms	
GO:0006915	apoptotic process
GO:0009607	response to biotic stimulus
GO:0009873	ethylene mediated signaling pathway
GO:0045087	innate immune response
GO:0048544	recognition of pollen
GO:0080167	response to karrikin
GO:0006996	*organelle organization*
GO:0008104	*protein localization*
GO:0015074	*DNA integration*
GO:0044249	*cellular biosynthetic process*
GO:1901576	*organic substance biosynthetic process*

in vivo M4_overexpressed_DEGs_most specific biological process GO terms	
GO:0006032	chitin catabolic process
GO:0006694	steroid biosynthetic process
GO:0006855	drug transmembrane transport
GO:0006857	oligopeptide transport
GO:0009407	toxin catabolic process
GO:0009813	flavonoid biosynthetic process
GO:0019419	sulfate reduction
GO:0042128	nitrate assimilation
GO:0045087	innate immune response
GO:0006351	*transcription, DNA-dependent*
GO:0006412	*translation*
GO:0006508	*proteolysis*
GO:0010468	*regulation of gene expression*
GO:0015074	*DNA integration*
GO:2000112	*regulation of cellular macromolecule biosynthetic process*
GO:2001141	*regulation of RNA biosynthetic process*

Table 2.8: The most represented biological process GO terms of *in vivo* 101.14 and M4 overexpressed DEGs over and underrepresented. Terms underrepresented are in italics

The comparison between most specific biological process GO terms of *in vitro* and *in vivo* plants revealed four common terms in 101.14 DEGs and in M4 that are reported in Table 2.9

101.14_DEGs_most specific biological process GO terms over/under represented common *in vitro* and *in vivo*	
GO:0006915	apoptotic process
GO:0009873	ethylene mediated signaling pathway
GO:0048544	recognition of pollen
GO:0015074	*DNA integration*

M4_DEGs_most specific biological process GO terms over/under represented common *in vitro* and *in vivo*	
GO:0006032	chitin catabolic process
GO:0009407	toxin catabolic process
GO:0006412	*translation*
GO:0015074	*DNA integration*

Table 2.9: The most represented biological process GO terms of 101.14 and M4 overexpressed DEGs over and underrepresented in both growth conditions. Terms underrepresented are in italics.

4 Discussion

In this study we evaluated leaf transcriptional profiles of two different grapevine rootstocks: 101.14 (*V. riparia* x *V. rupestris*) and M4 ((*V. vinifera* x *V. berlandieri*) x *V. berlandieri* cv. Resseguier n. 1) grown in *in vitro* condition.

DEGs detected on *in vitro* plants transcriptome was related with data of *in vivo* plants and we could establish that 58% of DEGs detected in plants grown in greenhouse were shared with DEGs of *in vitro* plants. This data, compared to the percentage of overlapping DEGs of greenhouse plants maintained at controlled conditions but sampled in different time point (61%), led us to assume that the experimental conditions of *in vitro* growth could be approximated, at transcriptome level, to plants grown in greenhouse under controlled conditions. In other words we established that using *in vitro* growth we could mimic the behaviors, in terms of gene expressions, of plants grown in controlled condition in greenhouse. Our hypothesis could be confirmed by the study of Chupeau and coworkers (Chupeau *et al*., 2013) that identified only few differences in the transcriptome profiles of *Arabidopsis*' plants grown in soil and *in vitro*. The researcher used the same growth conditions for plants in soil and *in vitro* and the comparison between transcriptome profiles identified only 355 DEGs (30 genes overexpressed and 325 down regulated *in vitro*). Considering that these data came from *Arabidopsis* and that in our experiment we compared *in vitro* plants (growth chamber) with two years old plants (greenhouse) the identification of a high percentage of overlapping DEGs between *in vivo* and *in vitro* DEGs for both the genotypes is quite interesting.

Biological process GO term analysis were performed on *in vitro* and *in vivo* DEGs for the identification of term and genes expressed that could be genotype specific rather that related to grown condition. We studied in details GO terms for apoptotic process (GO:0006915) and chitin catabolic process (GO:0006032) that were over represented in DEGs of 101.14 and M4, respectively. 101.14 showed a high number of genes for resistance protein in particular genes (5) that encoded nucleotide-binding site leucine-rich repeat (NBS-LRR) proteins. These proteins are characterized by nucleotide binding site (NBS) and leucine-rich repeat (LRR) domains as well as

variable amino- and carboxy-terminal domains and they are involved in the detection of diverse pathogens, including bacteria, viruses, fungi, nematodes, insects and oomycetes (McHale *et al*., 2006). When we considered the M4 GO category chitin catabolic process we identified six genes related to chitinase rather than NBS-LRR. We hypothesized that the mechanisms that the two genotypes implement in response of pathogens are different in both growth conditions encoding constitutively for two different defense mechanisms. 101.14 entrusts its defense to NBS-LRR while M4 used chitinase expression. Van Sluyter and colleagues compared grape chitinase activities in Chardonnay and Cabernet Sauvignon, both *V. vinifera*, with *V. rotundifolia* cv. Fry (Van Sluyterand *et al*., 2005). They identified that in *vinifera* varieties chitinase activities in berries were approximately 130-fold (Cabernet Sauvignon) and 80-fold (Chardonnay) higher than in *rotundifolia*. The researcher suggests that the disparity in constitutive chitinase activity between the two species could be related to differed strategies of defense based on other pathogenesis-related proteins in non-*vinifera* species. Although this study analyzed only the chitinase activity and not the gene expression of chitinases we can speculate that the constitutive overexpression of chitinase genes in M4 in both grown conditions could be related to the characteristics of the hybrid (*V. vinifera* x *V. berlandieri*) x *V. berlandieri* respect to 101.14 that is an hybrid *V. riparia* x *V. rupestris*.

The results analyzed in this discussion are only some of the aspects that could be studied more in details to characterize transcriptome profiles of the two grapevine genotypes. However we have to remember that the study of transcriptional profiles of *in vitro* plants takes relevance in our research once related to ChIP-seq data analyses (Chapter IV). The main target of this PhD project is indeed the study of the relations between histone modifications and genes expression in grapevine and data like these allowed us to conclude that *in vitro* plants could be good representatives of *in vivo* gene expression.

5 References

Apel K, Hirt H(2004) Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol 55:373-399

Campagna D, Albiero A, Bilardi A et al (2009) PASS: a program to align short sequences. Bioinformatics 25(7):967-968

Chaves MM, Flexas J, Pinheiro C(2009) Photosynthesis under drought and salt stress: regulation mechanisms from whole plant to cell. Ann Bot 103(4):551-560

Chaves MM, Pereira JS, Maroco J et al (2002) How plants cope with water stress in the field. Photosynthesis and growth. Ann Bot 89 Spec No:907-916

Chaves MM, Zarrouk O, Francisco R et al (2010) Grapevine under deficit irrigation: hints from physiological and molecular data. Ann Bot 105(5):661-676

Chupeau MC, Granier F, Pichon O, Renou JP, Gaudin V, Chupeau Y(2013) Characterization of the early events leading to totipotency in an Arabidopsis protoplast liquid culture by temporal transcript profiling. Plant Cell 25(7):2444-2463

Cominelli E, Conti L, Tonelli C, Galbiati M(2013) Challenges and perspectives to improve crop drought and salinity tolerance. N Biotechnol 30(4):355-361

Corso M (2014) A transcriptomic approach to dissect the effect of grapevine rootstocks on plant tolerance to abiotic stresses and berry ripening (PhD thesis, http://paduaresearch.cab.unipd.it/6393/)

Corso M, Vannozzi A, Maza E, Vitulo N, Meggio F, Pitacco A, Telatin A, D'Angelo M, Schiavon R, Negri AS, Prinsi B, Valle G, Ramina A, Bouzayen M, Bonghi C and Lucchin M (2015) Comprehensive transcripts profiling of two contrasting grapevine rootstock genotypes links phenylpropanoid pathways to enhanced drought tolerance. Submitted.

Covarrubias J, Rombolà AD (2013) Physiological and biochemical responses of the iron chlorosis tolerant grapevine rootstock 140 Ruggeri to iron deficiency and bicarbonate. Plant Soil 370:305–315

Cramer GR, Ergul A, Grimplet J et al (2007) Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. Funct Integr Genomics 7(2):111-134

Desikan R, Hancock JT, Bright J et al (2005) A role for ETR1 in hydrogen peroxide signaling in stomatal guard cells. Plant Physiol 137(3):831-834

Fisarakis I, Chartzoulakis K, Stavrakas D (2001) Response of Sultana vines (V. vinifera L.) on six rootstocks to NaCl salinity exposure and recovery. Agric Water Manage 51: 13–27

Flexas J, Baron M, Bota J et al (2009) Photosynthesis limitations during water stress acclimation and recovery in the drought-adapted Vitis hybrid Richter-110 (V. berlandierixV. rupestris). J Exp Bot 60(8):2361-2377

Fujita Y, Fujita M, Shinozaki K, Yamaguchi-Shinozaki K(2011) ABA-mediated transcriptional regulation in response to osmotic stress in plants. J Plant Res 124(4):509-525

Gambetta GA, Manuck CM, Drucker ST et al (2012) The relationship between root hydraulics and scion vigour across Vitis rootstocks: what role do root aquaporins play? J Exp Bot 63(18):6445-6455

Gotz S, Garcia-Gomez JM, Terol J et al (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36(10):3420-3435

Gregory PJ, Atkinson CJ, Bengough AG et al (2013) Contributions of roots and rootstocks to sustainable, intensified crop production. J Exp Bot 64(5):1209-1222

Jaillon O, Aury JM, Noel B et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463-467

Kar RK(2011) Plant responses to water stress: role of reactive oxygen species. Plant Signal Behav 6(11):1741-1745

Kohli A, Sreenivasulu N, Lakshmanan P, Kumar PP(2013) The phytohormone crosstalk paradigm takes center stage in understanding how plants respond to abiotic stresses. Plant Cell Rep 32(7):945-957

Lata C, Yadav A, Prasad M (2011) Role of the plant transcription factors in abiotic stress tolerance. Abiotic Stress Response in Plants - Physiological, Biochemical and Genetic Perspectives

Lovisolo C, Schubert A, Sorce C(2002) Are xylem radial development and hydraulic conductivity in downwardly-growing grapevine shoots influenced by perturbed auxin metabolism? New Phytol 156(1):65-74

Marguerit E, Brendel O, Lebon E, Van Leeuwen C, Ollat N(2012) Rootstock control of scion transpiration and its acclimation to water deficit are controlled by different genes. New Phytol 194(2):416-429

McHale L, Tan X, Koehl P, Michelmore RW(2006) Plant NBS-LRR proteins: adaptable guards. Genome Biol 7(4):212

Meggio F, Prinsi B, Negri AS, Simone Di Lorenzo G, Lucchini G, Pitacco A, Failla O, Scienza A, Cocucci M and Espen L Biochemical and physiological responses of two grapevine rootstock genotypes to drought and salt treatments. Australian Journal of Grape and Wine Research 20:310-323

Meinhard M, Rodriguez PL, Grill E(2002) The sensitivity of ABI2 to hydrogen peroxide links the abscisic acid-response regulator to redox signalling. Planta 214(5):775-782

Novikova GV, Stepanchenko NS, Nosov AV, Moshkov IE (2009) At the beginning of route: ABA perception and signal transduction in plants. Russina Journal of Plant Physiology 56:727-741

Peleg Z, Blumwald E(2011) Hormone balance and abiotic stress tolerance in crop plants. Curr Opin Plant Biol 14(3):290-295

Peleg Z, Blumwald E(2011) Hormone balance and abiotic stress tolerance in crop plants. Curr Opin Plant Biol 14(3):290-295

Qin F, Shinozaki K, Yamaguchi-Shinozaki K(2011) Achievements and challenges in understanding plant abiotic stress responses and tolerance. Plant Cell Physiol 52(9):1569-1582

Santner A, Estelle M(2009) Recent advances and emerging trends in plant hormone signalling. Nature 459(7250):1071-1078

Sharp RE, Davies WJ(1979) Solute regulation and growth by roots and shoots of water-stressed maize plants. Planta 147(1):43-49

Souza CRd, Maroco JP, Santos TPd, Rodrigues ML, Lopes CM, Pereira JS, Chaves MM (2003) Partial rootzonedrying: regulation of stomatal aperture and carbon assimilation in filed-grown grapevines (Vitis vinifera cv Moscatel). Functional Plant Biology4:447-456

Tillett RL, Ergul A, Albion RL, Schlauch KA, Cramer GR, Cushman JC(2011) Identification of tissue-specific, abiotic stress-responsive gene expression patterns in wine grape (Vitis vinifera L.) based on curation and mining of large-scale EST data sets. BMC Plant Biol 11:86-2229-11-86

Trapnell C, Roberts A, Goff L et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7(3):562-578

Van Sluyter S, Durako MJ, Halkides CJ(2005) Comparison of Grape Chitinase Activities in Chardonnay and Cabernet Sauvignon with *Vitis rotundifolia* cv. Fry. Am. J. Enol. Vitic. March 2005 56:81-85

Vezzi F, Del Fabbro C, Tomescu AI, Policriti A(2012) rNA: a fast and accurate short reads numerical aligner. Bioinformatics 28(1):123-124

Walker RR, Blackmore DH, Clingeleffer PR, Correll RL (2002) Rootstock effects on salt tolerance of irrigated field-grown grapevines (*Vitis vinifera* L. cv. Sultana).: 1. Yield and vigour inter-relationships. Australian Journal of Grape and Wine Research 8:3-14

Walker RR, Blackmore DH, Clingeleffer PR, Correll RL (2004) Rootstock effects on salt tolerance of irrigated field-grown grapevines (*Vitis vinifera* L. cv. Sultana).:2. Ion concentration in leaves and juice. Australian Journal of Grape and Wine Research 10:90-99.

Zerbino DR, Birney E(2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18(5):821-829

6 Supplementary Materials

6.1 Analyses of molecular function GO terms

GO terms analysis for molecular function of DEGs overexpressed in *in vitro* allowed the identification of ten most specific GO terms in 101.14 and twenty five in M4.

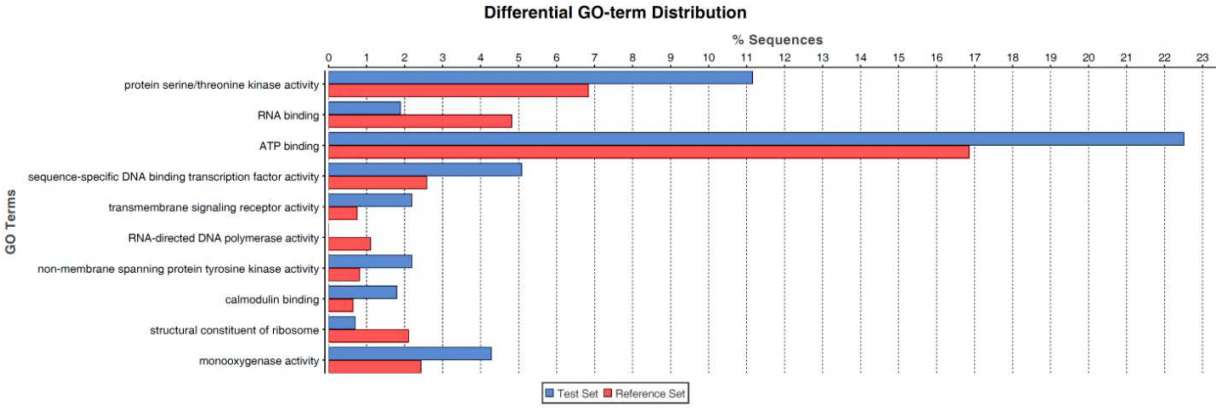


Figure 2.5: Bar chart of 101.14 overexpressed genes most specific molecular function GO terms

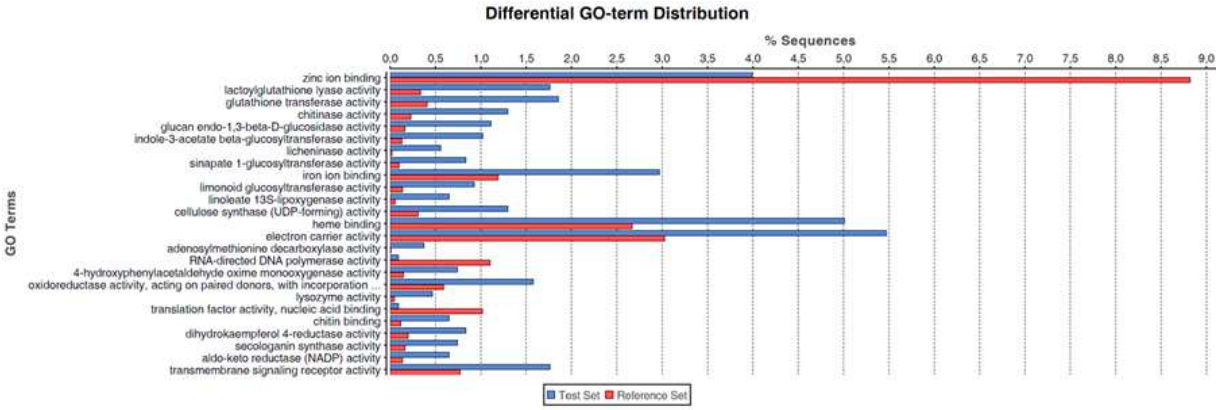


Figure 2.6: Bar chart of M4 overexpressed genes most specific molecular function GO terms

101.14_DEGs_most specific molecular function GO terms over/under represented	
GO:0003700	sequence-specific DNA binding transcription factor activity
GO:0004497	monooxygenase activity
GO:0004674	protein serine/threonine kinase activity
GO:0004715	non-membrane spanning protein tyrosine kinase activity
GO:0004888	transmembrane signaling receptor activity
GO:0005516	calmodulin binding
GO:0005524	ATP binding
GO:0003723	*RNA binding*
GO:0003735	*structural constituent of ribosome*
GO:0003964	*RNA-directed DNA polymerase activity*

M4_DEGs_most specific molecular function GO terms over/under represented	
GO:0003796	lysozyme activity
GO:0004014	adenosylmethionine decarboxylase activity
GO:0004033	aldo-keto reductase (NADP) activity
GO:0004364	glutathione transferase activity
GO:0004462	lactoylglutathione lyase activity
GO:0004568	chitinase activity
GO:0004888	transmembrane signaling receptor activity
GO:0005506	iron ion binding
GO:0008061	chitin binding
GO:0009055	electron carrier activity
GO:0016165	linoleate 13S-lipoxygenase activity
GO:0016706	oxidoreductase activity
GO:0016760	cellulose synthase (UDP-forming) activity
GO:0020037	heme binding
GO:0042972	licheninase activity
GO:0042973	glucan endo-1,3-beta-D-glucosidase activity
GO:0045552	dihydrokaempferol 4-reductase activity
GO:0047215	indole-3-acetate beta-glucosyltransferase activity
GO:0050284	sinapate 1-glucosyltransferase activity
GO:0050592	4-hydroxyphenylacetaldehyde oxime monooxygenase activity
GO:0050616	secologanin synthase activity
GO:0050645	limonoid glucosyltransferase activity
GO:0003964	*RNA-directed DNA polymerase activity*
GO:0008135	*translation factor activity, nucleic acid binding*
GO:0008270	*zinc ion binding*

Table 2.10: The most represented molecular function GO terms of *in vitro* 101.14 and M4 overexpressed DEGs over and underrepresented. Terms underrepresented are in italics.

in vivo 101.14_overexpressed_DEGs_most specific molecular function GO terms	
GO:0004721	phosphoprotein phosphatase activity
GO:0004888	transmembrane signaling receptor activity
GO:0009055	electron carrier activity
GO:0018685	alkane 1-monooxygenase activity
GO:0020037	heme binding
GO:0033788	leucoanthocyanidin reductase activity
GO:0070330	aromatase activity
GO:0003723	*RNA binding*

in vivo M4_overexpressed_DEGs_most specific molecular function GO terms	
GO:0003854	3-beta-hydroxy-delta5-steroid dehydrogenase activity
GO:0004014	adenosylmethionine decarboxylase activity
GO:0004032	alditol:NADP+ 1-oxidoreductase activity
GO:0004364	glutathione transferase activity
GO:0004462	lactoylglutathione lyase activity
GO:0004568	chitinase activity
GO:0004604	phosphoadenylyl-sulfate reductase (thioredoxin) activity
GO:0008061	chitin binding
GO:0009703	nitrate reductase (NADH) activity
GO:0015238	drug transmembrane transporter activity
GO:0015297	antiporter activity
GO:0016621	cinnamoyl-CoA reductase activity
GO:0033741	adenylyl-sulfate reductase (glutathione) activity
GO:0035251	UDP-glucosyltransferase activity
GO:0042972	licheninase activity
GO:0042973	glucan endo-1,3-beta-D-glucosidase activity
GO:0045552	dihydrokaempferol 4-reductase activity
GO:0003677	*DNA binding*
GO:0003723	*RNA binding*
GO:0008270	*zinc ion binding*

Table 2.11:The most represented molecular function GO terms of *in vivo* 101.14 and M4 overexpressed DEGs over and underrepresented. Terms underrepresented are in italics.

The comparison between most specific molecular function GO terms of *in vitro* and *in vivo* plants revealed two common terms in 101.14 DEGs and nine common terms in M4 (Table 2.12).

101.14_DEGs_most specific molecular function GO terms over/under represented common *in vitro* and *in vivo*	
GO:0004888	transmembrane signaling receptor activity
GO:0003723	*RNA binding*

M4_DEGs_most specific molecular function GO terms over/under represented common *in vitro* and *in vivo*	
GO:0004014	adenosylmethionine decarboxylase activity
GO:0004364	glutathione transferase activity
GO:0004462	lactoylglutathione lyase activity
GO:0004568	chitinase activity
GO:0008061	chitin binding
GO:0042972	licheninase activity
GO:0042973	glucan endo-1,3-beta-D-glucosidase activity
GO:0045552	dihydrokaempferol 4-reductase activity
GO:0008270	*zinc ion binding*

Table 2.12: The most represented molecular function GO terms of 101.14 and M4 overexpressed DEGs over and underrepresented in *in vitro* and *in vivo* growth conditions. Terms underrepresented are in italics

6.2 Description of most specific molecular function and biological process GO terms of Differentially Expressed Genes

6.2.1 Biological Process most specific GO terms of DEGs overexpressed in 101.14 and M4

GO:0015074 Term: DNA integration Definition: The process in which a segment of DNA is incorporated into another, usually larger, DNA molecule such as a chromosome. NO genes

GO:0006952 Term: defense response Definition: Reactions, triggered in response to the presence of a foreign body or the occurrence of an injury, which result in restriction of damage to the organism attacked or prevention/recovery from the infection caused by the attack.

GO:0009873 Term: ethylene mediated signaling pathway Definition: A series of molecular signals mediated by ethylene (ethene).

GO:0006915 Term: apoptotic process Definition: A programmed cell death process which begins when a cell receives an internal (e.g. DNA damage) or external signal (e.g. an extracellular death ligand), and proceeds through a series of biochemical events (signaling pathways) which typically lead to rounding-up of the cell, retraction of pseudopodes, reduction of cellular volume (pyknosis), chromatin condensation, nuclear fragmentation (karyorrhexis), plasma membrane blebbing and fragmentation of the cell into apoptotic bodies. The process ends when the cell has died. The process is divided into a signaling pathway phase, and an execution phase, which is triggered by the former.

GO:0048544 Term: recognition of pollen Definition: The process, involving the sharing and interaction of the single locus incompatibility haplotypes, involved in the recognition or rejection of the self-pollen by cells in the stigma. This process ensures out-breeding in certain plant species.

GO:0006412 Term: translation Definition: The cellular metabolic process in which a protein is formed, using the sequence of a mature mRNA molecule to specify the sequence of amino acids in a polypeptide chain. Translation is mediated by the ribosome, and begins with the formation of a ternary complex between aminoacylated initiator methionine tRNA, GTP, and initiation factor 2, which subsequently associates with the small subunit of the ribosome and an mRNA. Translation ends with the release of a polypeptide chain from the ribosome.

GO:0006468 Term: protein phosphorylation Definition: The process of introducing a phosphate group on to a protein.

GO:0006278 Term: RNA-dependent DNA replication Definition: A DNA replication process that uses RNA as a template for RNA-dependent DNA polymerases (e.g. reverse transcriptase) that synthesize the new strands. NO genes

GO:0015031 Term: protein transport Definition: The directed movement of proteins into, out of or within a cell, or between cells, by means of some agent such as a transporter or pore.

GO:0019439 Term: aromatic compound catabolic process Definition: The chemical reactions and pathways resulting in the breakdown of aromatic compounds, any substance containing an aromatic carbon ring.

GO:1901361 Term: organic cyclic compound catabolic process Definition: The chemical reactions and pathways resulting in the breakdown of organic cyclic compound.

GO:0007264 Term: small GTPase mediated signal transduction Definition: Any series of molecular signals in which a small monomeric GTPase relays one or more of the signals. NO genes

GO:0070727 Term: cellular macromolecule localization Definition: Any process in which a macromolecule is transported to, and/or maintained in, a specific location at the level of a cell. Localization at the cellular level encompasses movement within the cell, from within the cell to the cell surface, or from one location to another at the surface of a cell.

GO:0046907 Term: intracellular transport Definition: The directed movement of substances within a cell.

GO:0015074 Term: DNA integration Definition: The process in which a segment of DNA is incorporated into another, usually larger, DNA molecule such as a chromosome. NO genes

GO:0006032 Term: chitin catabolic process Definition: The chemical reactions and pathways resulting in the breakdown of chitin, a linear polysaccharide consisting of beta-(1->4)-linked N-acetyl-D-glucosamine residues.

GO:0009407 Term: toxin catabolic process Definition: The chemical reactions and pathways resulting in the breakdown of toxin, a poisonous compound (typically a protein) that is produced by cells or organisms and that can cause disease when introduced into the body or tissues of an organism.

GO:0055114 Term: oxidation-reduction process Definition: A metabolic process that results in the removal or addition of one or more electrons to or from a substance, with or without the concomitant removal or addition of a proton or protons.

GO:0045087 Term: innate immune response Definition: Innate immune responses are defense responses mediated by germline encoded components that directly recognize components of potential pathogens.

GO:0006412 Term: translation Definition: The cellular metabolic process in which a protein is formed, using the sequence of a mature mRNA molecule to specify the sequence of amino acids in a polypeptide chain. Translation is mediated by the ribosome, and begins with the formation of a ternary complex between aminoacylated initiator methionine tRNA, GTP, and initiation factor 2, which subsequently associates with the small subunit of the ribosome and an mRNA. Translation ends with the release of a polypeptide chain from the ribosome.

GO:0006334 Term: nucleosome assembly Definition: The aggregation, arrangement and bonding together of a nucleosome, the beadlike structural units of eukaryotic chromatin composed of histones and DNA.

GO:0006278 Term: RNA-dependent DNA replication Definition: A DNA replication process that uses RNA as a template for RNA-dependent DNA polymerases (e.g. reverse transcriptase) that synthesize the new strands.

GO:0031408 Term: oxylipin biosynthetic process Definition: The chemical reactions and pathways resulting in the formation of any oxylipin, any of a group of biologically active compounds formed by oxidative metabolism of polyunsaturated fatty acids.

GO:0080090 Term: regulation of primary metabolic process Definition: Any process that modulates the frequency, rate or extent of the chemical reactions and pathways within a cell or an organism involving those compounds formed as a part of the normal anabolic and catabolic processes. These processes take place in most, if not all, cells of the organism.

GO:0060255 Term: regulation of macromolecule metabolic process Definition: Any process that modulates the frequency, rate or extent of the chemical reactions and pathways involving macromolecules, any molecule of high relative molecular mass, the structure of which essentially comprises the multiple repetition of units derived, actually or conceptually, from molecules of low relative molecular mass.

GO:0006915 Term: apoptotic process Definition: A programmed cell death process which begins when a cell receives an internal (e.g. DNA damage) or external signal (e.g. an extracellular death ligand), and proceeds through a series of biochemical events (signaling pathways) which typically lead to rounding-up of the cell, retraction of pseudopodes, reduction of cellular volume (pyknosis), chromatin condensation, nuclear fragmentation (karyorrhexis), plasma membrane blebbing and fragmentation of the cell into apoptotic bodies. The process ends when the cell has

died. The process is divided into a signaling pathway phase, and an execution phase, which is triggered by the former.

GO:0050832 Term: defense response to fungus Definition: Reactions triggered in response to the presence of a fungus that act to protect the cell or organism.

GO:0030243 Term: cellulose metabolic process Definition: The chemical reactions and pathways involving cellulose, a linear beta1-4 glucan of molecular mass 50-400 kDa with the pyranose units in the -4C1 conformation.

GO:0032774 Term: RNA biosynthetic process Definition: The chemical reactions and pathways resulting in the formation of RNA, ribonucleic acid, one of the two main type of nucleic acid, consisting of a long, unbranched macromolecule formed from ribonucleotides joined in 3',5'-phosphodiester linkage. Includes polymerization of ribonucleotide monomers.

GO:0009699 Term: phenylpropanoid biosynthetic process Definition: The chemical reactions and pathways resulting in the formation of aromatic derivatives of trans-cinnamic acid.

GO:0031323 Term: regulation of cellular metabolic process Definition: Any process that modulates the frequency, rate or extent of the chemical reactions and pathways by which individual cells transform chemical substances.

GO:0009889 Term: regulation of biosynthetic process Definition: Any process that modulates the frequency, rate or extent of the chemical reactions and pathways resulting in the formation of substances.

6.2.2 Molecular Function most specific GO terms of DEGs overexpressed in 101.14 and M4

GO:0004674 Term: protein serine/threonine kinase activity Definition: Catalysis of the reactions: ATP + protein serine = ADP + protein serine phosphate, and ATP + protein threonine = ADP + protein threonine phosphate.

GO:0003723 Term: RNA binding Definition: Interacting selectively and non-covalently with an RNA molecule or a portion thereof.

GO:0005524 Term: ATP binding Definition: Interacting selectively and non-covalently with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator.

GO:0003700 Term: sequence-specific DNA binding transcription factor activity Definition: Interacting selectively and non-covalently with a specific DNA sequence in order to modulate transcription. The transcription factor may or may not also interact selectively with a protein or macromolecular complex.

GO:0004888 Term: transmembrane signaling receptor activity Definition: Combining with an extracellular or intracellular signal and transmitting the signal from one side of the membrane to the other to initiate a change in cell activity.

GO:0003964 Term: RNA-directed DNA polymerase activity Definition: Catalysis of the reaction: deoxynucleoside triphosphate + DNA(n) = diphosphate + DNA(n+1). Catalyzes RNA-template-directed extension of the 3'- end of a DNA strand by one deoxynucleotide at a time. NO genes

GO:0004715 Term: non-membrane spanning protein tyrosine kinase activity Definition: Catalysis of the reaction: ATP + a non-membrane spanning protein L-tyrosine = ADP + a non-membrane spanning protein L-tyrosine phosphate.

GO:0005516 Term: calmodulin binding Definition: Interacting selectively and non-covalently with calmodulin, a calcium-binding protein with many roles, both in the calcium-bound and calcium-free states.

GO:0003735 Term: structural constituent of ribosome Definition: The action of a molecule that contributes to the structural integrity of the ribosome.

GO:0004497 Term: monooxygenase activity Definition: Catalysis of the incorporation of one atom from molecular oxygen into a compound and the reduction of the other atom of oxygen to water.

GO:0008270 Term: zinc ion binding Definition: Interacting selectively and non-covalently with zinc (Zn) ions.

GO:0004462 Term: lactoylglutathione lyase activity Definition: Catalysis of the reaction: (R)-S-lactoylglutathione = glutathione + methylglyoxal.

GO:0004364 Term: glutathione transferase activity Definition: Catalysis of the reaction: R-X + glutathione = H-X + R-S-glutathione. R may be an aliphatic, aromatic or heterocyclic group; X may be a sulfate, nitrile or halide group.

GO:0004568 Term: chitinase activity Definition: Catalysis of the hydrolysis of (1->4)-beta linkages of N-acetyl-D-glucosamine (GlcNAc) polymers of chitin and chitodextrins.

GO:0042973 Term: glucan endo-1,3-beta-D-glucosidase activity Definition: Catalysis of the hydrolysis of (1->3)-beta-D-glucosidic linkages in (1->3)-beta-D-glucans.

GO:0047215 Term: indole-3-acetate beta-glucosyltransferase activity Definition: Catalysis of the reaction: (indol-3-yl)acetate + UDP-D-glucose = 1-O-(indol-3-ylacetyl)-beta-D-glucose + UDP.

GO:0042972 Term: licheninase activity Definition: Catalysis of the hydrolysis of (1->4)-beta-D-glucosidic linkages in beta-D-glucans containing (1->3) and (1->4) bonds.

GO:0050284 Term: sinapate 1-glucosyltransferase activity Definition: Catalysis of the reaction: UDP-glucose + sinapate = UDP + 1-sinapoyl-D-glucose.

GO:0005506 Term: iron ion binding Definition: Interacting selectively and non-covalently with iron (Fe) ions.

GO:0050645 Term: limonoid glucosyltransferase activity Definition: Catalysis of the reaction: UDP-glucose + limonin = glucosyl-limonin + UDP.

GO:0016165 Term: linoleate 13S-lipoxygenase activity Definition: Catalysis of the reaction: linoleate + O2 = (9Z,11E)-(13S)-13-hydroperoxyoctadeca-9,11-dienoate.

GO:0016760 Term: cellulose synthase (UDP-forming) activity Definition: Catalysis of the reaction: UDP-glucose + ((1,4)-beta-D-glucosyl)(n) = UDP + ((1,4)-beta-D-glucosyl)(n+1).

GO:0020037 Term: heme binding Definition: Interacting selectively and non-covalently with heme, any compound of iron complexed in a porphyrin (tetrapyrrole) ring.

GO:0009055 Term: electron carrier activity Definition: Any molecular entity that serves as an electron acceptor and electron donor in an electron transport chain. An electron transport chain

is a process in which a series of electron carriers operate together to transfer electrons from donors to any of several different terminal electron acceptors to generate a transmembrane electrochemical gradient.

GO:0004014 Term: adenosylmethionine decarboxylase activity Definition: Catalysis of the reaction: S-adenosyl-L-methionine + H(+) = S-adenosylmethioninamine + CO(2).

GO:0003964 Term: RNA-directed DNA polymerase activity Definition: Catalysis of the reaction: deoxynucleoside triphosphate + DNA(n) = diphosphate + DNA(n+1). Catalyzes RNA-template-directed extension of the 3'- end of a DNA strand by one deoxynucleotide at a time.

GO:0050592 Term: 4-hydroxyphenylacetaldehyde oxime monooxygenase activity Definition: Catalysis of the reaction: (Z)-(4-hydroxyphenyl)acetaldehyde oxime + H(+) + NADPH + O(2) = (S)-4-hydroxymandelonitrile + 2 H(2)O + NADP(+).

GO:0016706 Term: oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors Definition: Catalysis of the reaction: A + 2-oxoglutarate + O2 = B + succinate + CO2. This is an oxidation-reduction (redox) reaction in which hydrogen or electrons are transferred from 2-oxoglutarate and one other donor, and one atom of oxygen is incorporated into each donor.

GO:0003796 Term: lysozyme activity Definition: Catalysis of the hydrolysis of the beta-(1->4) linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in a peptidoglycan and between N-acetyl-D-glucosamine residues in chitodextrins.

GO:0008135 Term: translation factor activity, nucleic acid binding Definition: Functions during translation by interacting selectively and non-covalently with nucleic acids during polypeptide synthesis at the ribosome.

GO:0008061 Term: chitin binding Definition: Interacting selectively and non-covalently with chitin, a linear polysaccharide consisting of beta-(1->4)-linked N-acetyl-D-glucosamine residues.

GO:0045552 Term: dihydrokaempferol 4-reductase activity Definition: Catalysis of the reaction: cis-3,4-leucopelargonidin + NADP+ = (+)-dihydrokaempferol + NADPH + H+.

GO:0050616 Term: secologanin synthase activity Definition: Catalysis of the reaction: loganin + NADPH + H+ + O2 = secologanin + NADP+ + 2 H2O.

GO:0004033 Term: aldo-keto reductase (NADP) activity Definition: Catalysis of the reaction: an alcohol + NADP+ = an aldehyde or a ketone + NADPH + H+.

GO:0004888 Term: transmembrane signaling receptor activity Definition: Combining with an extracellular or intracellular signal and transmitting the signal from one side of the membrane to the other to initiate a change in cell activity.

Chapter III

Chromatin Immunoprecipitation (ChIP) protocol for grapevine leaf material

1 Introduction

At first sight chromatin could be labeled as a static and repetitive structure but it is far from reality. Chromatin remodeling, histone modifications and chromatin related processes act on chromatin structure in a dynamic manner for the regulation of gene expression. The epigenetic control of gene expression is based on an intricate interplay between three main molecular mechanisms: DNA methylation, histone modifications and RNA-based mechanisms (RdDM).

Whereas DNA methylation has been studied in great detail for several decades, the role of histone modifications has only been fully appreciated from about 15 years (Jenuweinet *et al*., 2001). Since then a large number of studies on new histone modifications and on their possible function has been performed especially in model organisms. The most important procedure used for the study of histone modifications is the Chromatin Immunoprecipitation (ChIP) assay, a techniques first established for *Drosophila* cultured cell in 1988 by Solomon (Solomon *et al*., 1988). Very briefly Chromatin Immunoprecipitation (ChIP) is a technique whereby an antibody against a specific feature of interest is used to immunoprecipitated in a selectively manner that element from a chromatin preparation. The ChIP protocol consist of a series of subsequent steps that start with the crosslinking between proteins and DNA sequences through the employment of formaldehyde (Figure 3.1). Formaldehyde crosslinks proteins and DNA molecules within approximately 2 Å of each other, and thus it is suitable for looking at proteins that directly bind DNA but it is not suitable for examining proteins that are indirectly associated with DNA like the ones of big complexes. After crosslinking the chromatin is sheared in fragments of 250-750 bp of length either by enzymatic digestion with micrococcal nuclease or, more frequently, by mechanical sonication. The lysate is cleared by sedimentation and the complexes protein-DNA are immunoprecipitated using antibodies against the protein of interest. After the appropriate treatments, the crosslink is reversed (we referred to this step as reverse crosslinking or decrossing), the proteins are removed and the precipitated ChIP-enriched DNA is purified and analyzed by PCR and quantitative PCR or, more recently, with technology based on microarray or Next Generation Sequencing (NGS) (Collas, 2010). NGS approach could be very useful in this

kind of studies because gene expressions may depend on the distribution of histone modifications through some regions more than on an exact position on gene. Using approach of NGS, like ChIP sequencing, scientist don't have to choose in advance specific genes and regions to test in PCR on immunoprecipitated material but rather the distribution of the modifications in whole genome can be studied.

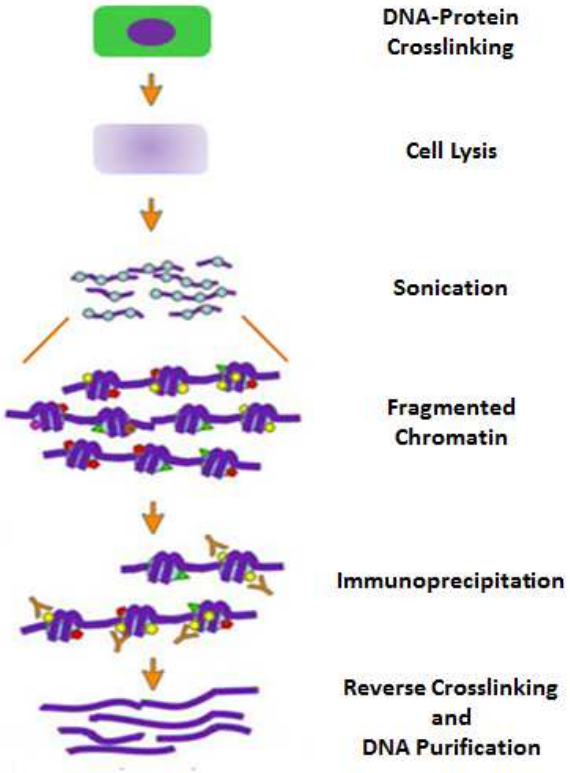


Figure 3.1: Schematic representation of Chromatin Immunoprecipitation assay (adaptation from Collas, 2010)

Nowadays we are witnessing a paradoxical situation where a large amount of data on histone modifications derived from genome wide analyses are available for animal and plant model species while for non-model species no ChIP protocols are available.

The greater difficulty in the development of the first ChIP experiments in plants was related to the large amount of material needed for chromatin extraction, early protocols (Ascenzi and Gantt, 1999; Chua *et al*., 2001) required indeed 100 g of *Arabidopsis* tissues for a single ChIP experiment. This problem was circumvented by recent studies that successfully combined

various protocols and optimizations (Gendrel *et al*., 2002; Johnson *et al*., 2002; Tariq *et al*., 2003; Bowler *et al*., 2004) but sometimes they are still not applicable in non-model species like for example in grapevine.

In order to study the relation between histone modifications and gene expression in two grapevine rootstocks we needed to develop a protocol for Chromatin Immunoprecipitation specific for *Vitis*. Once immunoprecipitated the enriched DNA was managed for ChIP sequencing as describe in Chapter IV.

2 Material and methods

Nowadays there isn't any chromatin immunoprecipitation assay available for grapevine. For the development of a ChIP protocol specific for *Vitis* we tested different protocols available for other species (Rossi *et al*., 2007, Haring *et al.*, 2007, Locatelli *et al*., 2009 and Ricardi *et al*., 2010) and different kind of grapevine tissues. Here we reported the definitive protocol optimized during this PhD research project.

2.1 Chose of starting material

We started our protocol development from fresh young leaf and root materials that were sampled from two-years old plants grown into the greenhouse of the Azienda agraria sperimentale L. Toniolo in Legnaro, PD. Once collected the tissues were washed with water, dyed with paper towel and placed in a 4°C TBS solution (Tris pH7.6 20mM, NaCl 200mM, stock reagents by Sigma-Aldrich). After some trials we have shelved the idea of using fresh materials for three main reasons: the first was the large amount of material needed for the protocol optimization, the second was the seasonality of the plants and the third was the necessity to keep the plants in controlled and stable conditions. We decided to carry out our experiments on leaf materials of plants grown *in vitro* that were immediately frozen in liquid nitrogen after the sampling. The *in vitro* plants (Figure 3.2) were micropropagated and grown as described in Chapter II

Figure 3.2: Examples of the plants used in our studies

2.2 DNA-Protein crosslinking

Crosslinking of the starting material by formaldehyde was performed to ensure that the chromatin structure was preserved during its isolation and ChIP assay (Solomon *et al.*, 1985 and 1988). The crosslink step should be optimized, since too little crosslinking is not be sufficient to preserve the chromatin structure during ChIP while too much crosslinking would hamper the procedure (Das *et al.*, 2004).

As said before we started our setting on fresh roots and leaves of plants grown in greenhouse. Samples were cut in pieces of 0.4 x 0.4 mm and 1.5 grams of materials were placed in a 50 ml centrifuge tube (Sarstedt) and treated with a solution of cold TBS (Tris pH7.6 20mM, NaCl 200mM) added of three different amount of formaldehyde to achieve a final concentration of 0%, 1% and 3% of formaldehyde to test which of these conditions were the most effective in our experiment. The samples were maintained at 4°C and treated with different cycles of vacuum infiltration. Initially we have tried to use two cycles each consisting of 2 minutes of vacuum infiltration plus 8 minutes of rest in ice, for a total of 20 minutes after of which the reaction was blocked using glycine (0.1M) (Locatelli *et al.*, 2009) but no good results were obtained. We tried to modify the number of vacuum cycles (three cycles with the same method), the length of the infiltration (10 and 15 minutes of vacuum in continuous) and the temperature of the incubation (different trials at 4°C and at room temperature), but no good results were again obtained. The

large amount of materials required for all these proofs, the plant seasonality and the necessity to maintain the plants in controlled conditions made us divert to use tissues derived from *in vitro* plants immediately frozen and grinded in liquid nitrogen after the sampling, that were stored at -80°C till the moment of the use.

More than one year of work was spent to set a protocol suitable for the fixation of chromatin in *in vitro* leaf material, but no good results were however obtained for roots.

2.2.1 Fixation protocol

In this section is reported the fixation protocol optimized for grapevine *in vitro* leaf material.

- 1.5 grams of frozen and grinded tissues were placed in 50 ml centrifuge tube (Sarstedt) and were treated with 25 ml of Nuclear Isolation Buffer at room temperature (NIB^+: 10mM Hepes pH7.6, 1M Sucrose, 5mM KCl, 5mM $MgCl_2$, 5mM EDTA, 1% Formadehyde, 0.6% Triton X-100 and 0.4mM of Phenylmethanesulfonyl fluoride -PMSF-, all stock reagents by Sigma-Aldrich);
- fixation was performed at room temperature for 15 minutes, after this incubation the crosslinking reaction was blocked using 3.4 ml of Glycine 1M (Sigma-Aldrich) and the sample was left 5 minutes at room temperature;
- the solution was filtered on single layer of Miracloth (Merck-Millipore) for the elimination of the residues of the grinded tissues. The permeated was centrifuged at 4000Xg for 10 minutes at 10°C (Universal 32R, Hettich Zentrifugen). After centrifugation the supernatant were discarded by focusing on the pellet where lay the nuclei;
- the pellet was resuspended with 2 ml of Nuclear Isolation Buffer without Triton X-100 and PMSF (NIB^-) with gentle shaking;
- the solution was centrifuged at 3000Xg for 10 minutes at 10°C. After the centrifugation the supernatant was discarded and the pellet was resuspended with gentle shaking in 10 ml of NIB^-. Once the solution was homogeneous it was filtered on single layer of Miracloth;

- the permeated was centrifuged for 10 minutes at 3000Xg at 10°C and the supernatant was removed;
- the pellet was finally resuspended in 3 ml of Nuclear Lysis Buffer (NLB: Hepes 50mM, NaCl 150mM, EDTA 1mM, Triton X-100 1%, Sodium Deoxycholate 0.1% and SDS 0.1%, all stock reagents by Sigma-Aldrich).

Once the chromatin was fixed a control of its quality was needed, for this reason a step of reverse crosslinking and a subsequent DNA extraction was necessary.

In the reverse crosslinking phase an aliquot of fixed chromatin was treated to disconnect DNA from the histone core using an incubation of 15 hours at 65°C after the addiction of NaCl with final concentration of 0.2M. After incubation an aliquot of decrossed DNA (+DC) was used in parallel with one not decrossed (-DC) and the DNA extraction was performed using Phenol : Chloroform : Isoamyl Alcohol (FCIA 25:24:1, all stock reagents by Sigma-Aldrich).

An equal volume of FCIA was added to the samples and, after 5 minutes of agitation, a centrifugation at 12000Xg (EBA12R, Hettich Zentrifugen) for 20 minutes was performed. The aqueous phase was then collected and treated with equal volume of a solution of Chloroform : Isoamyl Alcohol (CIA 24:1); after 5 minutes of agitation and 20 minutes of centrifugation at 12000Xg the aqueous phase was again collected and the DNA was precipitated using 2.5 volumes of cold absolute ethanol and 1/150 of volume of glycogen (Glycogen from *Mytilus edulis* (Blue mussel), Sigma-Aldrich). After an incubation of 3 hours in ice, a centrifugation at 12000Xg for 20 minutes at 10°C was performed. The supernatant was removed and the pellet was treated with 500 µl of cold ethanol 70% and left overnight at 4°C. After the incubation a further centrifugation at 12000Xg for 20 minutes at 10°C was performed, the supernatant was removed and the pellet was dried by applying vacuum for 5 minutes. The pellet was resuspended in 30 µl of sterile milliQ wate, treated with 1/10 of volume of RNAseA 1 µg/µl (Roche Diagnostics) and incubated for 30 minutes at 37°C (Thermobloc TD150P3, International PBI). The performances of crosslinking and reverse crosslinking were evaluated by gel electrophoresis.

2.3 Shearing of chromatin

The resolution of the ChIP procedure is determined by the size of the chromatin fragments used as input material. Two methods are commonly used to fragment the chromatin: sonication (hydrodynamic shearing) and micrococcal nuclease (MNase) digestion (Haring *et al.*, 2007). When using formaldehyde crosslinking, sonication is the preferred method, because crosslinking restricts the access of MNase to chromatin (Bellard *et al.*, 1989). The ideal fragmentation length of chromatin is between 250 and 750 bp and this fragmentation should be obtained with sonication at low power in combination with several pulses rather than high power and few pulses (Haring *et al.*, 2007).

Optimal fragmentation can be achieved by testing various sonication conditions on chromatin so, once established that the method for the crosslinking was fine, the shearing of chromatin is been set testing different combination of power, number and length of pulses.

The shearing was performed using a sonicator Sonoplus GM200 (Bandelin) with a microtip MS73 (Bandelin). The optimal sonication conditions on chromatin was achieved by trying various volume of sample, different tubes, different sonicator power setting, pulses length and pulses number. Each attempt was followed by reverse crosslinking, DNA isolation and the estimation of the shearing efficiency was evaluated by gel electrophoresis. The best setting that we have achieved to obtain chromatin fragments between 750 and 250bp required the use of aliquots of 500 µl of chromatin placed in 1.5 ml microcentrifuge tubes (Sarstedt), a sonicator power set at 30% and four pulses of 15 seconds of length alternated with 15 minutes of incubation on ice.

2.4 Quality and quantification of chromatin

The chromatin, once fixed and sheared, must be subjected to two quantification that had to be most accurate possible because a high yield of immunoprecipitation depends of the use of the right amount of chromatin. The first quantification was performed by a spectrophotometric method and was called Input RC PK. The second method involved the use of electrophoresis gel visualizations and was called Input RC.

The quantification were performed on 500 µl of fixed and sheared chromatin treated for reverse crosslinking. After decrossing the chromatin was split in two aliquots, one of 300 µl for Input RC PK analysis and one of 200 µl for Input RC.

2.4.1 Input RC PK

300 µl of sheared and reverse crosslinked chromatin were precipitated for 3 hour at -20°C using 750 µl of cold absolute ethanol (Sigma-Aldrich) and 2 µl of glycogen (Sigma-Aldrich). After incubation the sample was centrifuged for 15 minutes at 4°C with a speed of 17000Xg (EBA12R, Hettich Zentrifugen), the supernatant was removed and the pellet was washed with 1000 µl of cold ethanol 70%. The sample was again centrifuged at 17000Xg for 15 min at 4°C, the supernatant was removed and the pellet was dried by applying 5 minutes of vacuum. The pellet was resuspended in 300 µl of TE buffer for proteinase K (40mM Tris pH8, 10mM EDTA, all stock reagents by Sigma-Aldrich) and incubated at 37°C for 30 minutes after the addiction of 3 µl of RNAseA 1 µg/µl (Roche Diagnostics). After the incubation 2 µl of proteinase K 10 µg/µl (Roche Diagnostics) was added to the sample that was subjected to an incubation of one hour at 42°C. After these steps 332 µl of FCIA was added to the sample and the aqueous phase was recovered after centrifugation at 12000Xg for 20 minutes at 18°C. The aqueous phase was then purified using MinElute PCR Purification Kit (Qiagen) following the manufacturer's instructions and performing two elution of 13.5 µl each. The concentration of the DNA of the sample was estimated using a NanoDrop spectrophotometer (Thermo Scientific).

2.4.2 Input RC

In Input RC analysis the estimation of the concentration of DNA was performed through gel electrophoresis.

200 µl of the decrossed DNA was treated with equal volume of FCIA, the sample was centrifuged for 7 minutes at 12000Xg at 18°C and the aqueous phase was collected. 100 µl of this solution was treated with 10 µl of RNAseA 1 µg/µl (Roche Diagnostics) and incubated 1 hour at 37°C. Known volumes of the sample was then charged on a 0.8% agarose (Agarose Ultrapure, Invitrogen) gel (1x TAE, all stock reagents by Sigma-Aldrich) and compared with the signal

obtained from a sample of chromatin of known concentration called hereinafter calibrator. At the light of the data obtained from the first gel, different hypothesis of the concentration of the sample was tested and validated in a second gel with the same characteristics of the previous. This second gel allowed the identification of the more robust hypothesis that was validated in a third and fourth gel.

2.5 Chromatin immunoprecipitation procedure

The immunoprecipitation protocol that we used provides for the processing of aliquots of 10 µg of chromatin and was partitioned in three days of work.

2.5.1 Bound between chromatin and Dynabeads

The first day of chromatin immunoprecipitation was focused on the creation of the bound between antibodies and Dynabeads Protein G and subsequently between these complexes and the chromatin.

The treatment of the Dynabeads Protein G (Life technologies) was the first fundamental step of the immunoprecipitation procedure. 50 µl of Dynabeads protein G were necessary for each sample and for each sample two aliquots of Dynabeads protein G were used: one for the pre-clearing of the sample and one for the bound of antibodies.

2.5.1.1 Beads preparation

Each aliquot of Dynabeads was centrifuged at 700Xg for 2 minutes at room temperature (EBA12R, Hettich Zentrifugen) and accommodated in the DynaMag-2 Magnet (Life Technologies), the supernatant was removed and replaced by 1000 µl of Nuclear Lysis Buffer (NLB: Hepes 50mM, NaCl 150mM, EDTA 1mM, Triton X-100 1%, Sodium Deoxycholate 0.1% and SDS 0.1%, all stock reagents by Sigma-Aldrich) added of Inhibitors: Protease Inhibitor Cocktail 200x (Sigma-Aldrich), 0.01M Phenylmethanesulfonyl fluoride (PMSF, Sigma-Aldrich) and 0.01M Sodium butyrate (Sigma-Aldrich). Once added of the NLB the sample was placed 5 minutes on rotation (Mini-rotator Bio RS-24, BioSan). After incubation the sample was centrifuged at 700Xg for 2 minutes at room temperature and accommodated in the DynaMag-2 Magnet, the

supernatant was removed and the washing procedure was repeated for three times. Once added the NLB after the third centrifugation and elimination of the supernatant the beads were resuspended in NLB without inhibitors to obtain a final volume of 100 µl. Once get an homogeneous solution, the beads were split in two aliquots of 50 µl, the first aliquot was used for the bound of the antibodies and second for the pre-clearing of the sample.

2.5.1.2 Bound Antibodies-Dynabeads Protein G

In this step of the protocol we promoted the bound between antibodies and Dynabeads.

The *ratio* between chromatin and antibody is fundamental for the success of the experiment, for this reason the right amount of antibodies was necessary. We used 10 µl of antibodies against acetylation of lysine 9 of histone 3 (αH3K9Ac, Millipore Cat. 07-352), 5 µl of antibodies against C-terminal tail of histone 3 (αH3 C-term, Abcam Cat. ab1791) and 4 µl of antibodies against trimethylation of lysine 4 of histone 3 (αH3K4me3, Active Motif Cat. 39159). The right amount of antibodies were added to NLB with inhibitors to reach a final volume of 250 µl. The Antibodies solution were added to the Dynabeads to promote the ligation between them. Hereinafter, in parallel with the treatments of the beads with antibodies, a control was inserted into the procedure, it was represented by an aliquot of beads incubated without antibodies but only with NLB; this control was called NoAb.

For each sample an aliquot of beads (prepared like described in 2.5.1.1) was centrifuged at 700Xg for 2 minutes at 4°C and accommodated in the DynaMag-2 Magnet, the supernatant was removed and replaced by the solution Ab/NLB prepared as describe above. After that, the sample was left in rotation for two hours at 4°C. Once finished incubation the sample was centrifuged at 700Xg for 2 minutes at 4°C and accommodated in the DynaMag-2 Magnet, the supernatant was removed and 500 µl of NLB were added to the sample that was left in rotation for five minutes at room temperature. After that the sample was centrifuged at 700Xg for 2 minutes at 4°C and accommodated in the DynaMag-2 Magnet, the supernatant was removed and two steps of washes were performed with 500 µl of NLB added of BSA 5µg/µl (Sigma-Aldrich). The sample was placed in rotation for five minutes at room temperature and

centrifuged at 700Xg for 2 minutes at 4°C. Then it was accommodated in the DynaMag-2 Magnet and the supernatant was removed. After the second wash the sample was resuspended in 500 µl of NLB added of BSA and it was left in rotation for 2 hours at 4°C.

2.5.1.3 Pre-clearing of the sample

In parallel with the ligation Antibodies/Dynabeads the pre-clearing of the chromatin sample was performed. This step consisted on the elimination of the chromatin that were bounded to the Dynabeads with nonspecific interactions.

The sheared chromatin, that was stored at -20°C, was defrosted in ice and centrifuged at 9000Xg for 2 minutes, the supernatant was recovered and 10 µg of chromatin were placed in 15 ml centrifuge tubes (Sarstedt) and the volume of 8 ml was reached adding NLB with inhibitors. The aliquot of beads previously prepared (2.5.1.1) was centrifuged, 700Xg for 2 minutes at 4°C, accommodated in the DynaMag-2 Magnet, cleared from supernatant and then the beads were added to the chromatin and incubated in rotation for 4 hours at 4°C.

2.5.1.4 Immunoprecipitation

The chromatin pre-cleared was splitted in five 2 ml microcentrifuge tubes (Sarstedt), that were centrifuged at 700 Xg for 2 minutes at 4°C and accommodated in the DynaMag-2 Magnet. The supernatants were recovered in a new 15 ml centrifuge tubes and left in ice till the Dynabeads bounded to the antibodies finish their incubation and were ready to be centrifuged for the removal of the supernatant. The Dynabeas/Antibodies were added to the chromatin pre-clearified and the samples were immunoprecipitated through an overnight incubation in rotation at 4°C.

2.5.2 Isolation of enriched DNA

After the incubation between Dynabeas/Antibodies and chromatin we obtained a sample in which the part of chromatin characterized by the histone modification of interest resulted linked to the beads while the remaining part of chromatin was free in solution. During this phase of the

immunoprecipitation procedure, the chromatin enriched of the modification of interest linked to the beads was separated from the one free in solution.

2.5.2.1 Washes

After the overnight incubation the sample was treated with subsequent cycles of centrifugation (n-times 700Xg for 2 minutes at 4°C) designed to bring the beads of the sample from a 15 ml centrifuge tubes to a 1.5ml microcentrifuge tube. Once that all the Dynabeads of a sample were collected in the same 1.5ml microcentrifuge tube the beads were processed with subsequent washes.

The washes were performed with 500 µl of each of the buffers described below and the incubation was carried out in rotation for 5 minutes at room temperature. After each wash the sample was centrifuged at 700Xg for 2 minutes at room temperature and accommodated in the DynaMag-2 Magnet where the supernatant was removed. The sequence of washing buffers was reported below:

- One wash with Low Salt Buffer (Tris 20mM pH8, NaCl 50mM, EDTA 2mM, Triton X-100 1%, SDS 0.1%, all stock reagents by Sigma-Aldrich);
- One wash with High Salt Buffer (Tris 20mM pH8, NaCl 500 mM, EDTA 2mM, Triton X-100 1%, SDS 0.1%);
- One wash with LNDET Buffer (Tris 10mM, LiCl 250mM, EDTA 1mM, Nonidet P-40 1%, Sodium Deoxycholate 1%, all stock reagents by Sigma-Aldrich);
- Two washes with TE buffer (Tris 10mM pH8, EDTA 1mM)

2.5.2.2 Elution

After the last wash all the supernatant was removed from the sample taking care of not remove the Dynabeads. 260 µl of Elution Buffer ($NaHCO_3$ 0.1M, SDS 1%, all stock reagents by Sigma-Aldrich) were added to the sample that was vortexed and left in rotation at 65°C for 30 minutes. After the incubation the sample was centrifuged at 700Xg for 2 minutes at room temperature and accommodated in the DynaMag-2 Magnet, the supernatant was recovered in a new

microcentrifuge tube and placed immediately on ice. A second elution with the same procedure was performed and the supernatant obtained was added to the same microcentrifuge tube.

2.5.2.3 Reverse Crosslinking

The sample was treated for reverse crosslinking adding NaCl with final concentration of 0.2M and incubated at 65°C overnight for 15 hours.

2.5.3 Purification of enriched DNA

After the reverse crosslinking phase the sample was purified to obtain a DNA that can be used for all the subsequent analyses.

The decrossed sample was precipitate 2 hour at -20°C using 1000 µl of cold absolute ethanol (Sigma-Aldrich) and 2 µl of glycogen (Sigma-Aldrich). After the precipitation the sample was centrifuged for 15 minutes at 4°C at 17000Xg. After this step the supernatant was removed, the pellet was washed with 1000 µl of cold ethanol 70% and another centrifugation was performed. The supernatant was removed and the pellet was dried by applying vacuum for 5 minutes. The pellet was resuspended in 300 µl TE buffer for proteinase K (Tris 40mM pH8, EDTA 10mM) and incubated at 37°C for 30 minutes after the addiction of 3 µl of RNAseA 1 µg/µl (Roche Diagnostics) . After this phase 2 µl of proteinase K 10 µg/µl (Roche Diagnostics) was added to the sample that was subjected to 1 hour of incubation at 42°C. Subsequently equal volume of FCIA was added to the sample and the aqueous phase was recovered after centrifugation (12000Xg for 20 minutes at 18°C). The DNA was then purified using MinElute PCR Purification Kit (Qiagen), following manufacturer's instructions and performing two elution of 13.5 µl each. The material collected was then available for all the subsequent analyses.

2.5.4 Evaluation of immunoprecipation performances

As for the precedent steps of the ChIP protocol development (2.2 DNA-Protein crosslinking, 2.3 Shearing of chromatin, 2.4 Quality and quantification of chromatin), also the immunoprecipitation procedure (2.5) had to be validated tested. The evaluation was performed

studying the real-time PCR amplification plots of the sample immunoprecipitated with different antibodies in comparison with the internal control NoAb.

The set of primers used for the evaluation of amplification capacity of immunoprecipitated DNA were primers designed on 5'UTR of three genes involved in the jasmonate pathway (VIT_09s0002g00890, VIT_11s0016g00710 and VIT_18s0041g02020). The sequence of the primers are reported in Table 3.1

Gene ID		Primer Name	Sequence	Lenght of Amplification Fragment
VIT_09s0002g00890	Jasmonate ZIM-domain protein 1	JZIM09_5'UTR_F JZIM09_5'UTR_R	TCACATTCAAAGAACCCAAGC GTTGAACGACGTCCCAATC	127
VIT_11s0016g00710	Jasmonate ZIM-domain protein 1	JZIM11_5'UTR_Ex1_F JZIM11_5'UTR_Ex1_R	CGAGAACTTCGATTCTGTGGA ACCTTCCGGAGTCTGCA	143
VIT_18s0041g02020	12-oxophytodienoate reductase 2	Oxo18_20_5'UTR_Ex1_F Oxo18_20_5'UTR_Ex1_R	TCTAAATGGGCTTCAACCTGT CAGAGACGCCATTCATTCT	139

Table 3.1: Primers designed on genes of jasmonate pathway. Each couple of primer was previously tested in PCR and the PCR products were sequenced to verify the specificity of the amplification.

3 Results

The development of a Chromatin Immunoprecipitation assay suitable for grapevine leaf tissues was be a long and laborious process. For these reasons different checkpoints were inserted into the protocol to test the performance of the assay that we were developing. The checkpoints were set for: DNA-Protein crosslinking, sonication, quantification of chromatin and, finally, for the immunoprecipitation procedure (Figure 3.3).

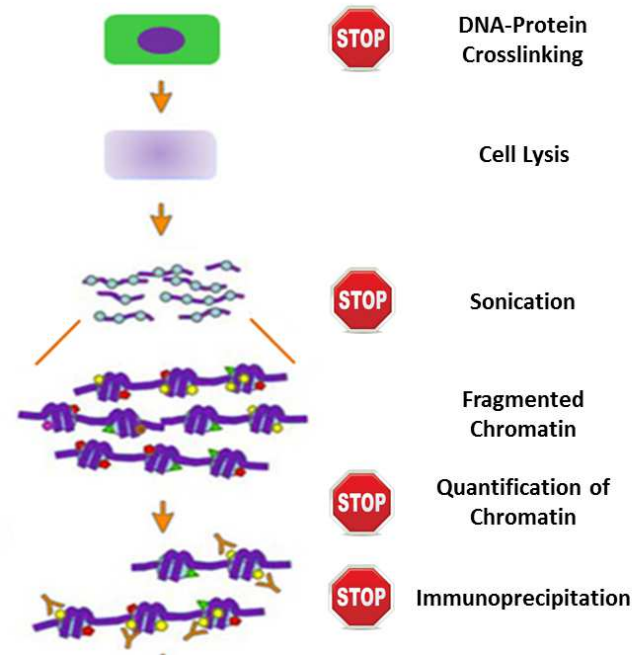


Figure 3.3: Schematic representation of ChIP protocol where the checkpoints inserted during development of the assay for grapevine leaf material were highlighted.

3.1 DNA-Protein crosslinking

After numerous test of different buffer, formaldehyde concentration, vacuum infiltration cycles, temperature of reaction, we were able to set the optimal conditions for crosslinking between DNA and protein as described in details in the section 2.2.1 of this chapter. In Figure 3.4 we reported the visualization of fixed material not treated for reverse crosslinking (-DC) and treated

for it (+DC). This gel electrophoresis image allowed us to establish that the fixation protocol developed and the subsequent stages of extraction were suitable for grapevine leaf material. In fact no signal was present in the -DC sample indicating that we were not able to extract the DNA from the sample because it was bound tightly to the histone core. On the other hand a good signal was obtained from the sample processed for the reverse crosslinking, suggesting that the fixation and the reverse crosslinking phase were fine for the material processed.

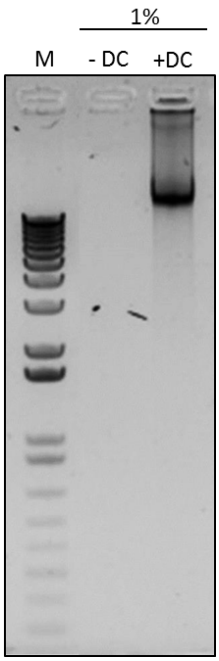


Figure 3.4: Gel electrophoresis image of 5µl of fixed material non treated (-DC) and treated (+DC) for reverse crosslinking. [Gel 1xTAE, 0.8% agarose (Agarose Ultrapure, Life Technologies), molecular weight marker (M): 1Kb Plus Ladder (Life Technologies)]

From hereinafter all the results reported derived from fixed chromatin DNA treated for reverse crosslinking and purified before gel visualization or others analyses.

3.2 Chromatin fragmentation

Since for the subsequent phase of immunoprecipitation a fragmentation between 250 and 750bp was recommended, we established that the best way to obtain this result was the use of aliquots of 500µl of chromatin in placed in 1.5 ml microcentrifuge tubes, sonicator power set at 30% and four pulses of 15 seconds of length, alternated with 15 minutes of incubation on ice. We reported in Figure 3.5 the gel electrophoresis image that allowed us to assert that the optimal setting condition was the one described above.

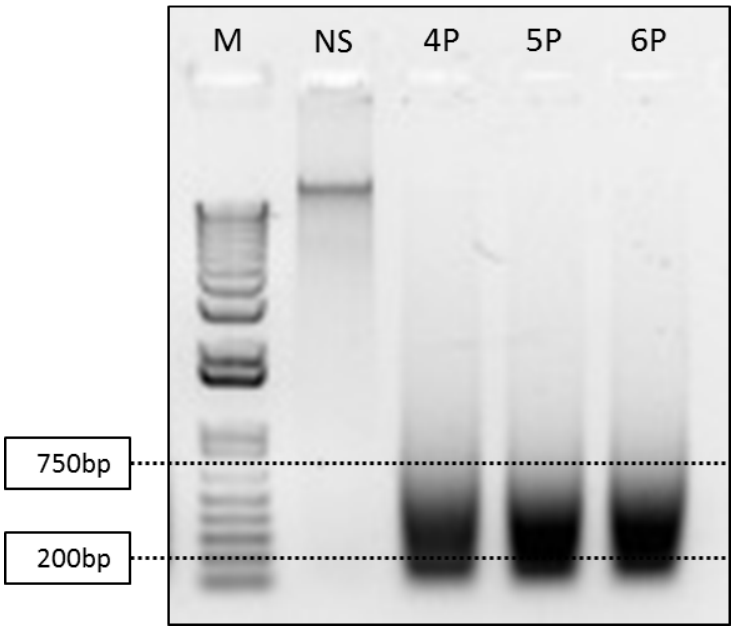


Figure 3.5: Visualization on gel electrophoresis of fixed chromatin not sonicated (NS, 5 µl) and subjected to 4, 5 and 6 pulses (respectively: 4P, 5P and 6P, 15 µl). [Gel 1xTAE, 0.8% agarose (Agarose Ultrapure, Life Technologies), molecular weight marker (M): 1Kb Plus Ladder(Life Technologies)]

When we compared the smear obtained from four cycles of sonication with the ones from five and six cycles, we saw that the four cycles smear appeared more homogeneous than the others. Furthermore we decided to use four cycles, rather than a higher number, because chromatin was in solution with SDS that could be a great matter in the sonication procedure because of its predisposition to the formation of bubbles that can affect the efficiency of sonication.

3.3 Input RC PK and Input RC

Although Input RC PK may seem a robust methodology because based on an instrumental quantification (section 2.4.1 of this chapter), it is necessary to remember that this quantification method is very sensible to the impurities present in the sample that may cause aberration in the measurement. Furthermore the high number of steps of the input RC PK protocol could make difficult to estimate the real DNA concentration of the chromatin sample. Nevertheless this analysis was very useful for determining the quality of the material obtained in terms of contaminants evaluated like *ratio* 260/280nm and 260/230nm. In Figure 3.6 and Table 3.2 we reported the absorbance profile of two chromatin samples analyzed through Input RC PK method.

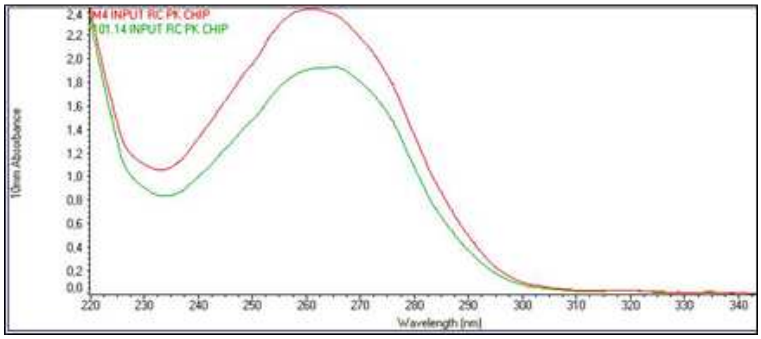


Figure 3.6 : NanoDrop (Thermo Fisher) absorbance plots of Input RC PK samples

Sample ID	Nucleic Acid Concentration [ng/µl]	A 260	A 280	A 260/280	A 260/230
M4 Input RC PK	120.8	2.416	1.355	1.8	2.2
101.14 Input RC PK	94.9	1.898	1.081	1.8	2.1

Table 3.2: NanoDrop (Thermo Fisher) values of absorbance and *ratio* of two Input RC PK samples

These results indicated that the DNA extracted from fixed chromatin of the two samples were of high quality, in fact the *ratio* 260/280nm was about 1.8 and the *ratio* 260/230 was above 2.1 for both the samples.

A more robust method, although more laborious, for the quantification of fixed chromatin was the analysis of Input RC. Input RC samples were prepared like described in the section 2.4.2 of this chapter. The quantification of fixed chromatin through the methods of Input RC involved gel electrophoresis analyses and the use of a sample of chromatin of known concentration called calibrator. Input RC quantification consist of four subsequent gel electrophoresis assays.
In the **first gel** were charged 30 ng of a chromatin sample with known concentration (calibrator) and known volumes, 10µl and 20µl, of sample with unknown concentration (Figure 3.7)

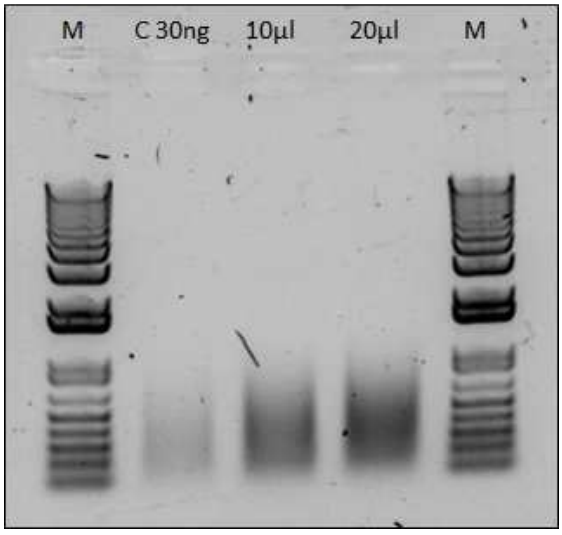


Figure 3.7: Agarose gel of 30 ng of calibrator's shared chromatin (C 30ng) and volume variables (10 µl and 20 µl) of chromatin with unknown concentration. [Gel 1xTAE, 0.8% agarose (Agarose Ultrapure, Life Technologies), molecular weight marker (M): 1Kb Plus Ladder(Life Technologies)].

These data allowed us to estimate that the signal obtained with 10 µl of unknown concentration chromatin could be 2.5 times more intense that the one of the calibrator. In this way we estimated that the concentration of the sample in exams could be 7.5 ng/µl.
After the data obtained from the first visualization on gel, we needed to evaluate if the estimation done could be correct so we produced a **second gel** charging 30 ng of the calibrator and theoretical 30 ng of the sample in exams assuming that its signal could be two (Hypothesis 1: H1), three (H2) or four (H3) times highest than the calibrator (Figure 3.8).

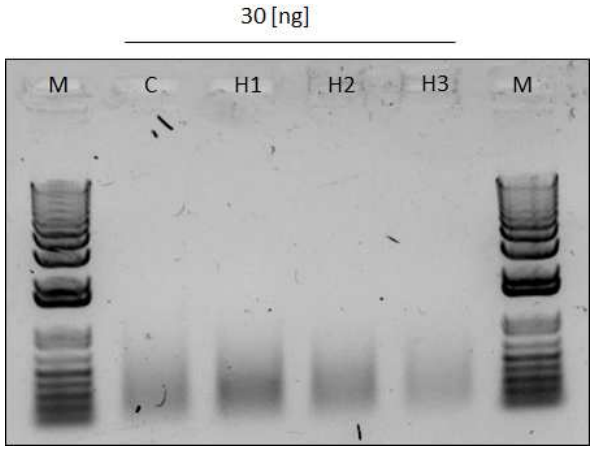


Figure 3.8: Visualization on agarose gel of shared chromatin of 30 ng of a calibrator (C) and 30 ng theoretically charged according to the three hypothesis that considered the unknown sample concentrated 6 ng/µl (H1), 9 ng/µl (H2) and 12 ng/µl (H3). [Gel 1xTAE, 0.8% agarose (Agarose Ultrapure, Life Technologies), molecular weight marker (M): 1Kb Plus Ladder (Life Technologies)] .

This second gel allowed us to estimate that the concentration of the sample could be 10 ng/µl.

A **third gel** electrophoresis analysis were performed. In this case we charged the calibrator with increasing concentrations of chromatin (30, 60 and 90 ng) and increasing concentrations of the sample in exams whose concentration was estimated like 10 ng/µl (Figure 3.9)

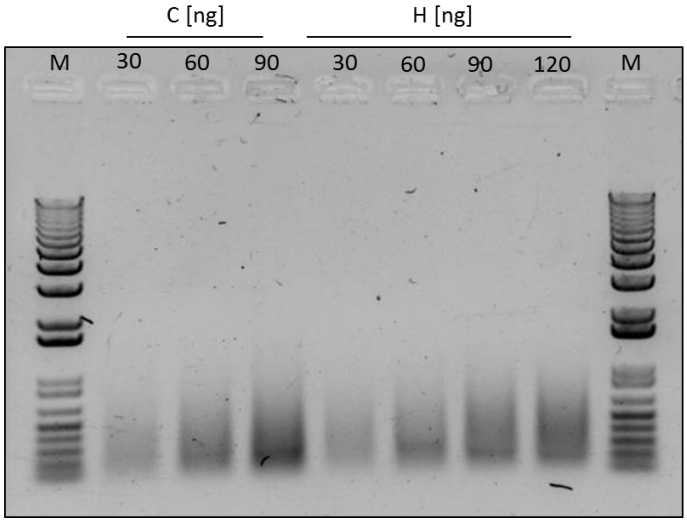


Figure 3.9: Visualization on agarose gel of increasing concentrations of calibrator chromatin (C: 30, 60, 90 ng) and of the sample with hypothetical concentration of 10 ng/µl (H: 30, 60, 90 and 120ng). [Gel 1xTAE, 0.8% agarose (Agarose Ultrapure, Life Technologies), molecular weight marker (M): 1Kb Plus Ladder (Life Technologies)]

The third gel indicated that the hypothesis of a concentration of 10 ng/µl was overestimated so we established that the concentration of the sample could be 8 ng/µl. To verifying this final hypothesis we charged on a **forth electrophoresis gel** 30 ng of the calibrator and of the sample (Figure 3.10).

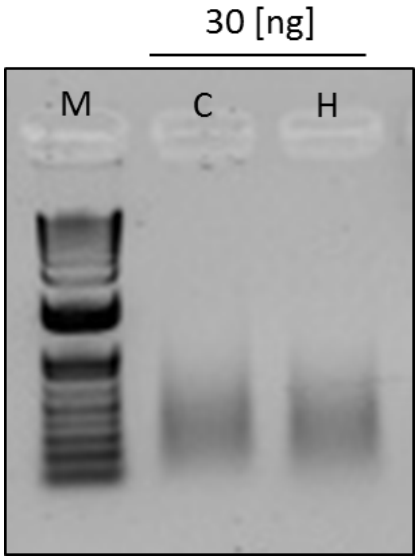


Figure 3.10: Visualization on agarose gel of 30 ng of chromatin from calibrator (C) and from the sample in exams with estimated hypothetically concentration of 8 ng/µl (H). [Gel 1xTAE, 0.8% agarose (Agarose Ultrapure, Life Technologies), molecular weight marker (M): 1Kb Plus Ladder (Life Technologies)] .

This last analysis allowed us to conclude that, for this particular sample fixed and sheared, the estimation of 8 ng/µl was correct. We calculated the yield of chromatin of this sample using the following formula

$$3000\ \mu l \cdot 8\frac{ng}{\mu l} = 24000\ ng \rightarrow \frac{24\ \mu g}{1.5\ g} = 16\ \frac{\mu g}{g}$$

The analysis of Input RC was performed every time that we fixed and sheared chromatin aliquots. We have to remember, in fact, that the success of the immunoprecipitation procedure is based upon the corrected *ratio* between chromatin and antibodies.

3.4 Evaluation of immunoprecipitated DNA and antibodies performances

In this section we reported the results obtained when three different antibodies were used for the immunoprecipitation of fixed and sheared chromatin. Immunoprecipitation with antibodies against C-terminal tail of histone 3 was used like positive control of the immunoprecipitation procedure while immunoprecipitation performances of antibodies against trimethylation of lysine 4 of histone 3 and against acetylation of lysine 9 of histone 3 were tested in order to use them in our later analysis of ChIP sequencing (see Chapter IV). The functionality of the immunoprecipitation procedure was evaluated tested by rea-time PCR (StepOnePlus real-time PCR system, Sybr Green chemistry, Applied Biosystem). Real-time PCR analyses were aimed to the evaluation of the enrichment of the amplification in the immunoprecipitated sample in respect to the internal control NoAb.

Immunoprecipitated enriched DNA was amplified using three couple of primers of genes involved in the jasmonate pathways and designed on the 5'UTR of the genes: Jzim09_5'UTR, Jzim11_5'UTR_Ex1 and Oxo18_20_5'UTR_Ex1.

The amplification plots of DNA enriched by immunoprecipitation and the relative control NoAb are reported in Figure 3.11 (H3 C-term immunoprecipitation), Figure 3.12 (H3K9ac immunoprecipitation) and Figure 3.13 (H3K4me3 immunoprecipitation).

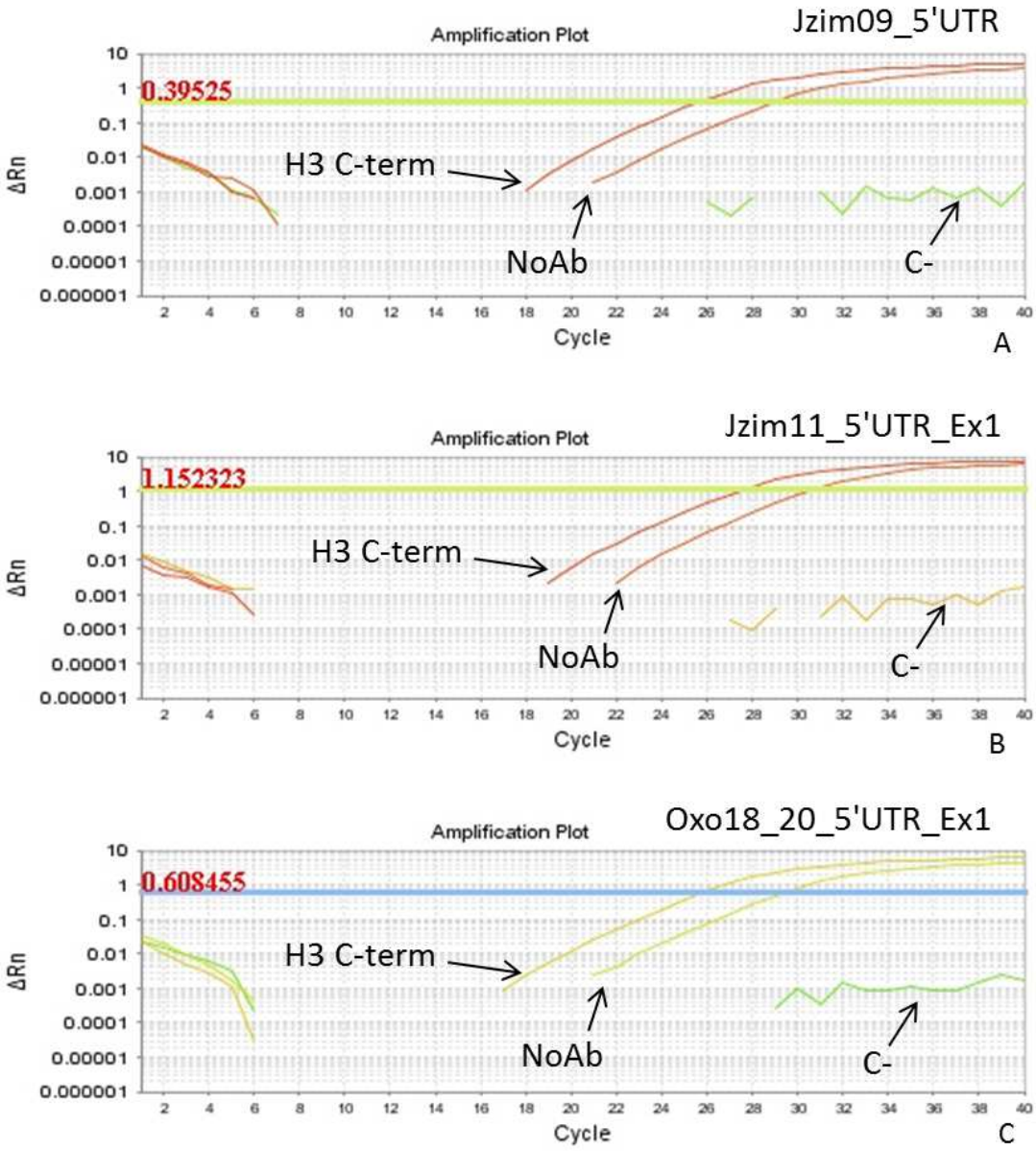


Figure 3.11: Real-time PCR amplification plots of the DNA enriched for H3 C-term and NoAb sample obtained with primers Jzim09_5'UTR (A), Jzim11_5'UTR_Ex1 (B) and Oxo18_20_5'UTR_Ex1 (C).

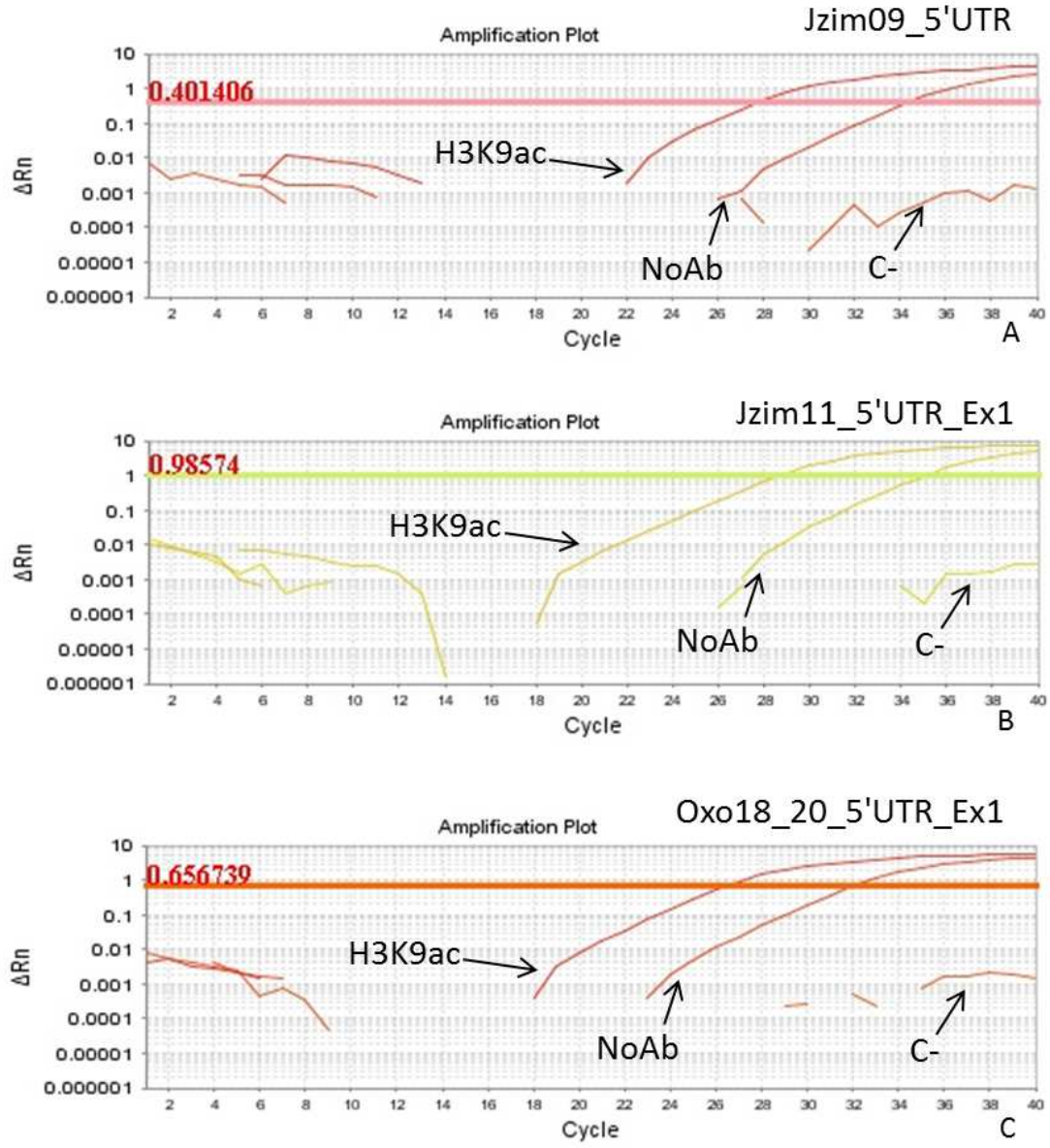


Figure 3.12: Real-time PCR amplification plots of the DNA enriched for histone modification H3K9ac and NoAb sample obtained with primers Jzim09_5'UTR (A), Jzim11_5'UTR_Ex1 (B) and Oxo18_20_5'UTR_Ex1 (C).

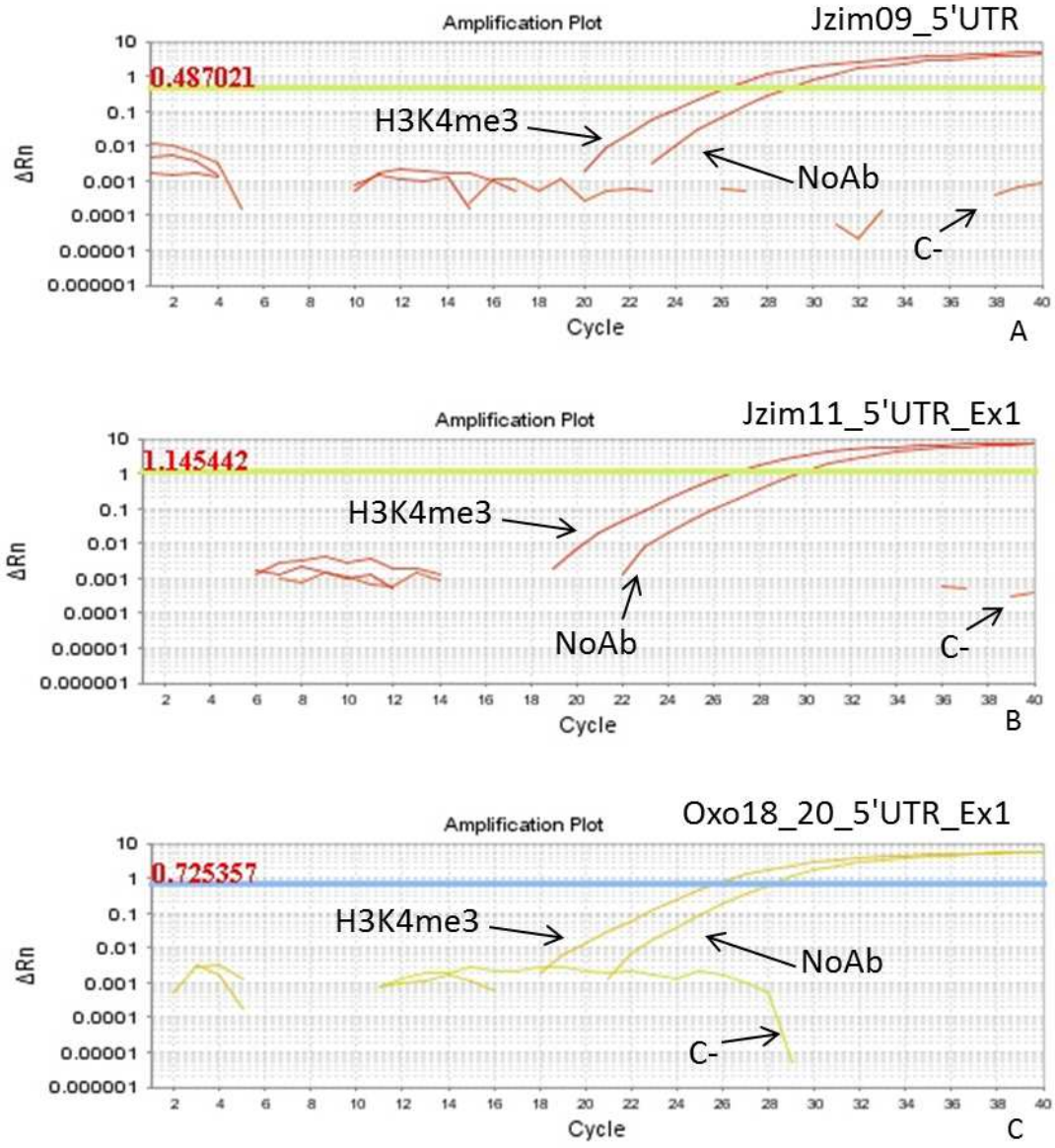


Figure 3.13: Real-time PCR amplification plots of the DNA enriched for tri methylation of lysine4 of histone 3 and NoAb sample obtained with primers Jzim09_5'UTR (A), Jzim11_5'UTR_Ex1 (B) and Oxo18_20_5'UTR_Ex1 (C).

When we compared C_T values obtained in the enriched DNA with the data from NoAb samples and the values of ΔC_T we identified a minimum enrichment of 2.3 ΔC_T and a maximum enrichment of 6.8 ΔC_T (Table 3.3).

	ΔC_T Ab vs NoAb		
	H3 C-term	H3K9ac	H3K4me3
Jzim09_5'UTR	3.3	6.1	2.7
Jzim11_5'UTR_Ex1	3.1	6.8	2.9
Oxo18_20_5'UTR_Ex1	3.3	5.8	2.3
Mean ΔC_T	3.2	6.2	2.6

Table 3.3: Real-time PCR ΔC_T values of DNA enriched with three different antibodies vs internal control (NoAb) of each immunoprecipitation reaction (NoAb)

The data obtained allowed us to assert that the immunoprecipitation of chromatin and the yield of enriched DNA were fine with both the three types of antibodies against the histone features considered and especially for H3K9ac immunoprecipitated DNA.

4 Discussion

In this chapter we described the development of a Chromatin Immunoprecipitation protocol specific for grapevine leaf material. To our knowledge, this protocol represents the first ChIP assay developed for analyzing the distribution of histone modifications in grapevine genome. For its novelty and the long time required for its development, this ChIP protocol represents an important goal in this PhD research project. Going back in our schedules, we can see that the first steps in ChIP field was performed in February 2012 and the first fine real-time PCR on enriched DNA was achieved in January 2014. This gives an idea of the difficulties that we have encountered during the development of this assay.

Comparing our protocol to the more recent and robust protocols developed for plant material, like for example the one recently released by Villar and Köhler (Villar and Köhler, 2010), we modified completely the fixation and extraction phases for the adaptation to the characteristics of recalcitrant tissues like those of grapevine. Despite the large use of vacuum infiltration of formaldehyde on fresh material in published protocols, not only in *Arabidopsis* (Johnson *et al*., 2002; Bowler *et al*., 2004; Makerevich *et al*., 2006; Sequeira-Mendes *et al*., 2014) but also in maize (Locatelli *et al*., 2009), in tomato (Ricardi *et al*., 2010) and *Populus* (Li *et al*., 2014), this method is not suitable for grapevine. The best performance of cross-linking in grapevine was achieved on frozen grinded leaf materials. Moreover the most used protocol for chromatin extraction required the use of sucrose gradient for the purification of nuclei while we used constant concentration of sucrose. This was done with the purpose of limiting the variants on protocol development. The subsequent phases of immunoprecipitation were very similar in each protocol and also in the one developed for grapevine. Nevertheless we had the foresight to use Dynabeads linked to Protein G rather than protein A Agarose beads that could give some contamination problems in the sequencing phase, because of their pre-blocking with salmon sperm DNA.

After all the adjustments of the protocol we obtained good results in each check point that we inserted at the fundamental steps of the procedure (Figure 3.3). We established that the DNA-

protein crosslink and the subsequent reverse crosslinking were optimized for grapevine leaf material (Section 3.1, Figure 3.4) as well as the shearing of the chromatin in fragment between 250 bp and 750 bp (Section 3.2, Figure 3.5). The quality and the yield of fixed DNA was tested through the analyses called Input RC PK and Input RC (Section 3.3) that showed how the DNA obtained after crosslinking, shearing and decrossing was of high quality (Figure 3.6) and the concentration was comparable with another chromatin calibrator (Figure 3.7-10)

The data obtained from the final step of ChIP, the real-time PCRs analyses on immunoprecipitated material, allowed us to assert that the immunoprecipitation of chromatin and the yield of enriched DNA in real-time amplification were fine for both the three types of antibodies against the histone features considered (Figure 3.11-13). In particular antibody against H3K9ac seems to be the one with the best performance in the tested situation. In fact it was the one with the highest level of enrichment, a mean ΔC_T of 6.2 (Table 3.3) corresponds in fact to a fold difference of 74 times. The fold difference could be assimilated to the enrichment level of sample treated with the antibody respect the sample that was linked to the beads in a nonspecific manner.

5 Acknowledgement

I want to thanks Dr. Vincenzo Rossi (CRA, Consiglio per la ricerca e la sperimentazione in agricoltura, Bergamo) for the chromatin calibrator and the precious information about antibodies and methodologies of immunoprecipitation procedure.

6 References

Ascenzi R, Gantt JS(1999) Subnuclear distribution of the entire complement of linker histone variants in Arabidopsis thaliana. Chromosoma 108(6):345-355

Bellard M, Dretzen G, Giangrande A, Ramain P(1989) Nuclease digestion of transcriptionally active chromatin. Methods Enzymol 170:317-346

Bowler C, Benvenuto G, Laflamme P et al (2004) Chromatin techniques for plant cells. Plant J 39(5):776-789

Chua YL, Brown AP, Gray JC(2001) Targeted histone acetylation and altered nuclease accessibility over short regions of the pea plastocyanin gene. Plant Cell 13(3):599-612

Collas P(2010) The current state of chromatin immunoprecipitation. Mol Biotechnol 45(1):87-100

Das PM, Ramachandran K, vanWert J, Singal R(2004) Chromatin immunoprecipitation assay. BioTechniques 37(6):961-969

Gendrel AV, Lippman Z, Yordan C, Colot V, Martienssen RA(2002) Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science 297(5588):1871-1873

Haring M, Offermann S, Danker T, Horst I, Peterhansel C, Stam M(2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. Plant Methods 3:11

Jenuwein T, Allis CD(2001) Translating the histone code. Science 293(5532):1074-1080

Johnson L, Cao X, Jacobsen S(2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr Biol 12(16):1360-1367

Li W, Lin YC, Li Q et al (2014) A robust chromatin immunoprecipitation protocol for studying transcription factor-DNA interactions and histone modifications in wood-forming tissue. Nat Protoc 9(9):2180-2193

Locatelli S, Piatti P, Motto M, Rossi V(2009) Chromatin and DNA modifications in the Opaque2-mediated regulation of gene transcription during maize endosperm development. Plant Cell 21(5):1410-1427

Makarevich G, Leroy O, Akinci U et al (2006) Different Polycomb group complexes regulate common target genes in Arabidopsis. EMBO Rep 7(9):947-952

Ricardi MM, Gonzalez RM, Iusem ND(2010) Protocol: fine-tuning of a Chromatin Immunoprecipitation (ChIP) protocol in tomato. Plant Methods 6:11-4811-6-11

Sequeira-Mendes J, Araguez I, Peiro R et al (2014) The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell 26(6):2351-2366

Rossi V, Locatelli S, Varotto S et al (2007) Maize histone deacetylase hda101 is involved in plant development, gene transcription, and sequence-specific modulation of histone modification of genes and repeats. Plant Cell 19(4):1145-1162

Solomon MJ, Larsen PL, Varshavsky A(1988) Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. Cell 53(6):937-947

Solomon MJ, Varshavsky A(1985) Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. Proc Natl Acad Sci U S A 82(19):6470-6474

Tariq M, Saze H, Probst AV, Lichota J, Habu Y, Paszkowski J(2003) Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc Natl Acad Sci U S A 100(15):8823-8827

Villar CB, Köhler C(2010) Plant chromatin immunoprecipitation. Methods Mol Biol 655:401-411

7 Supplementary material

Target Name	Sample Name	CT Mean	ΔCT	Sample Name	CT Mean	ΔCT	Sample Name	CT Mean	ΔCT
Jzim_09_5'utr	H3 C-term	25.7	3.3	H3K9ac	27.8	6.1	H3K4me3	26.3	2.7
	NoAb	29.0		NoAb	33.9		NoAb	29.0	
Jzim_11_5'utr-ex1	H3 C-term	27.6	3.1	H3K9ac	28.7	6.8	H3K4me3	27.0	2.9
	NoAb	30.7		NoAb	35.5		NoAb	29.9	
Oxo_12_20_5'utr_ex1	H3 C-term	25.9	3.3	H3K9ac	26.3	5.8	H3K4me3	25.7	2.3
	NoAb	29.2		NoAb	32.1		NoAb	28.0	

Table 3.4: Real-time PCR C_T mean values of immunoprecipitate material and control NoAb amplified with three couple of primers Jzim09_5'UTR (A), Jzim11_5'UTR_Ex1 (B) and Oxo18_20_5'UTR_Ex1

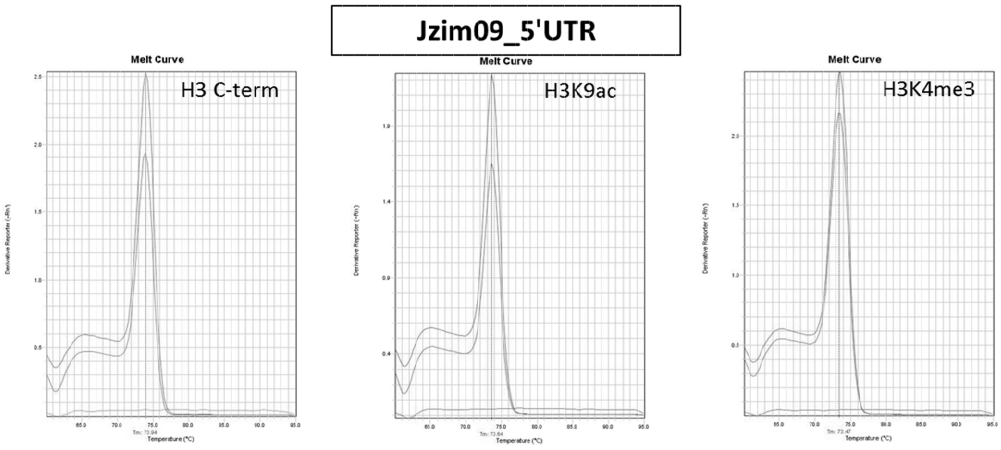


Figure 3.14: Melt curve of Jzim09_5'UTR in the three real-time pCR performed

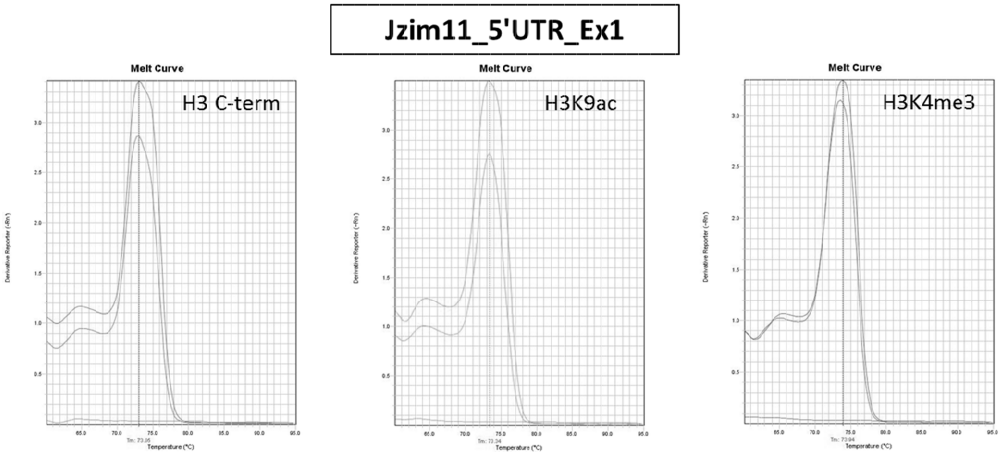


Figure 3.15: Melt curve of Jzim11_5'UTR_Ex1 in the three real-time pCR performed

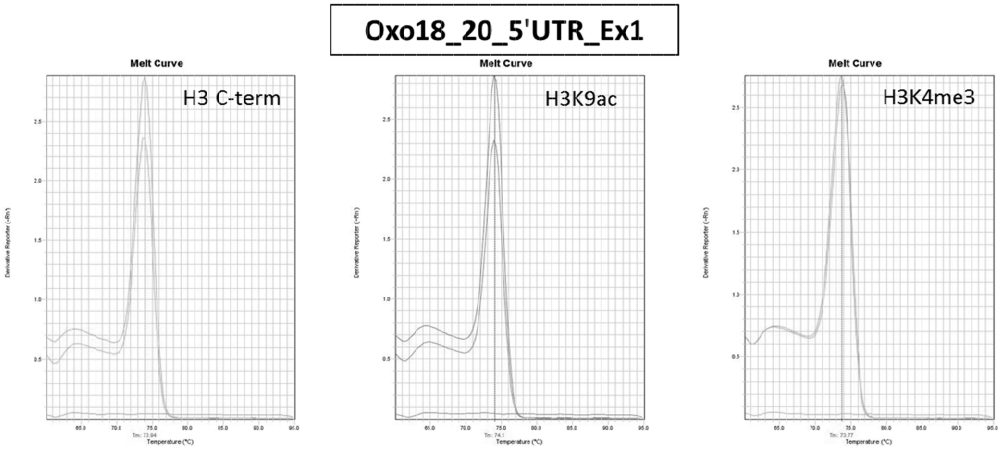


Figure 3.16: Melt curve of Oxo18_20_5'UTR_Ex1 in the three real-time pCR performed

Chapter IV

Correlation between histone modifications and gene expression in two grapevine rootstocks through NGS technologies: integration of ChIP Sequencing and mRNA Sequencing data

1 Introduction

Nowadays the viticulture is based on grafting, a practice in which a section of stem of *V. vinifera* with leaf buds is inserted into a resistant rootstock derived from either North American *Vitis* species or interspecific *Vitis* hybrids; this process is necessary to protect *V. vinifera* varieties from two main root pests, phylloxera and parasitic nematodes. On the other hand water availability is one of the major limiting factor for viticulture (Cramer *et al.,* 2007; Flexas *et al.,* 2009; Chaves *et al.,* 2010) also considering that the most important wine-producting regions in the world are subjected to seasonal drought and that many commercial rootstocks show scarce resistance to environmental stresses. DiSAA research group of University of Milan established till 1985 novel candidate rootstock, called M4, more performant in situation of drought stress. Rootstock M4 was the main characters of the project Ager-Serres 2010-2105 (http://users.unimi.it/serres/index.html). Ager-Serres was a multidisciplinary project; on one hand it focused the attention on biochemical and physiological behavior of the commercial rootstock 101.14 (*V. riparia x V. rupestris*) compared to the experimental one, M4 [(*V. vinifera* x *V. berlandieri*) x *V. berlandieri* x cv Resseguier n.1] grown under controlled and stressed conditions (Meggio *et al*., 2014). On the other hand the project used Next Generation Sequencing (NGS) techniques to go in deep with genome and transcriptome analyses (Corso *et al*., submitted). The main aim of the project was the identification of marker genes for the selection of rootstocks with high performances under stress conditions. It is within the Ager-Serres project that this PhD research study takes place with the objective to study the relation between transcriptional profile and histone modifications in the two rootstocks (Figure 4.1).

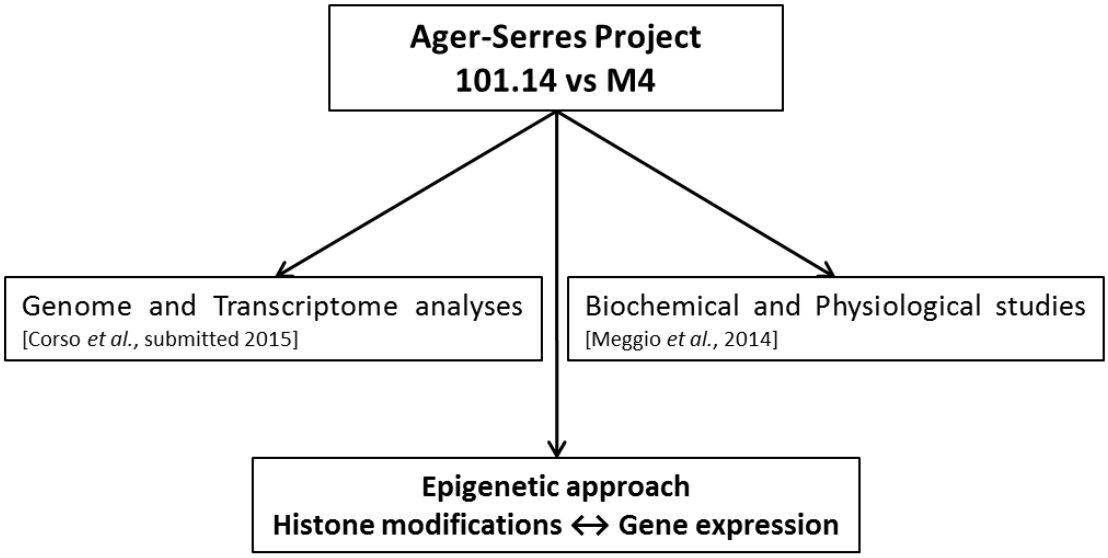


Figure 4.1: Schematic representation of Ager-Serres multidisciplinary project

Our research represents the first study aimed to clarify the relation between histone modifications and gene expression in the genus *Vitis*. To reach this objective we used Next Generation Sequencing (NGS) technologies for transcriptome analysis (mRNA-seq, discussed in details in Chapter II) and for sequencing of chromatin (ChIP-seq) immunoprecipitated with antibodies against histone modifications, known in literature like transcription activating: acetylation of lysine 9 of histone 3 (H3K9ac) and trimethylation of lysine 4 of histone 3 (H3K4me3) (Lauria and Rossi, 2011; Sequeira-Mendes *et al*., 2014).

In the study of the relation between histone modifications and chromatin transcriptional state a big step forward was recently done by Sequeira-Mendes and colleagues (Sequeira-Mendes *et al*., 2014). The research group identified nine chromatin states for *Arabidopsis thaliana* providing a ground for better understanding the linear organization of the genome, the relevance and/or the preference that certain signatures of genomic elements had and the effects of these features in gene expression.

This study started from the published profiles of nine histone modification marks (H3K9me2, H3K27me1, H4K5ac, H3K4me1, H2Bub, H3K36me3, H3K4me2, H3K4me3 and H3K27me3), three

histone variants (H2A.Z, H3.1 and H3.3), nucleosome density obtained from total H3 histone content, the genomic CG content and the CG methylated residues (Bernatavichute *et al*., 2008; Zilberman *et al*., 2008; Costas *et al*.; 2011; Stroud *et al*., 2012). Moreover the Spanish researchers generated genome-wide chromatin immunoprecipitation ChIP-on-chip data for H3K9ac and H3K14ac. Analyzing this huge amount of data they concluded that nine chromatin states could give a solid and coherent biological interpretation of the organization of *Arabidopsis* genome.

The features of each of the nine chromatin states like described in the Sequeria-Mendes's work are reported below:

- Chromatin state 1 is typically associated with transcribed regions and transcription start sites (TSSs) and it is characterized by a relatively low nucleosome density. In this state there are a high amount of H3K4me2/3, H3 acetylation, H3K36 and H2Bub; furthermore it is enriched in H3.3 and H2A.Z.;
- Chromatin state 2 shows a similar set of active marks in respect to state 1, but it also has high levels of the repressive modification H3K27me3. This state presents low levels of H3K36me3, H2Bub, H3ac and low nucleosome density;
- Chromatin state 3 is defined by high levels of histone marks H3K4me1, H2Bub, H3K36me3, H3K4me2/3 and it is highly depleted in Polycomb marks;
- Chromatin state 4 is similar to state 2, maintaining H3.3, H2A.Z and high levels of H3K27me3 but with reduced levels of marks typical of active transcription;
- Chromatin state 5 corresponds to the typical Polycomb regulated chromatin and is defined by a lower than average amount of all marks analyzed except for a high level of H3K27me3 and a moderate enrichment of H2A.Z within a nucleosome context enriched in H3.1. Whereas states 1 and 3 trend to be present at the 5' half of genic regions, states 2 and 4 seem to be more linked to intergenic regions containing proximal (state 2) and distal (state 4) promoter elements and perhaps regulatory regions. Chromatin state 5 primarily colocalizes in intergenic regions and only in part with genetic regions;

- Chromatin state 6 is characterized by a slight enrichment in H2A.Z, a higher than average nucleosome density and H3K4me1 marks within gene bodies. This state is present usually in intragenic regions;
- Chromatin state 7 has marks like H3K4me1, H2Bub and H3K36me3. H3K36me3 is the most diffuse mark and it is almost exclusively related to intragenic regions;
- Like chromatin state 7, also state 8 preferentially colocalizes with intergenic regions but also with transposon elements;
- Chromatin state 9 is the most GC-rich chromatin state, it is mostly located at intergenic region and on transposable elements.

The analyses of gene expression levels performed by Kurihara and collaborators (Kurihara *et al*., 2012) identified that the four states tagged like "open chromatin" by Sequeria-Mendes (state 1, 3, 6 and 7) are the ones with the highest levels of transcription. The scientists attributed the low amount of RNA transcripts of states 2, 4 and 5 to their preferential location in intergenic regions and on genes enriched for the mark H3K27me3, while for states 8 and 9 the low transcriptional level is due to their own nature of heterochromatic signatures (Figure 4.2).

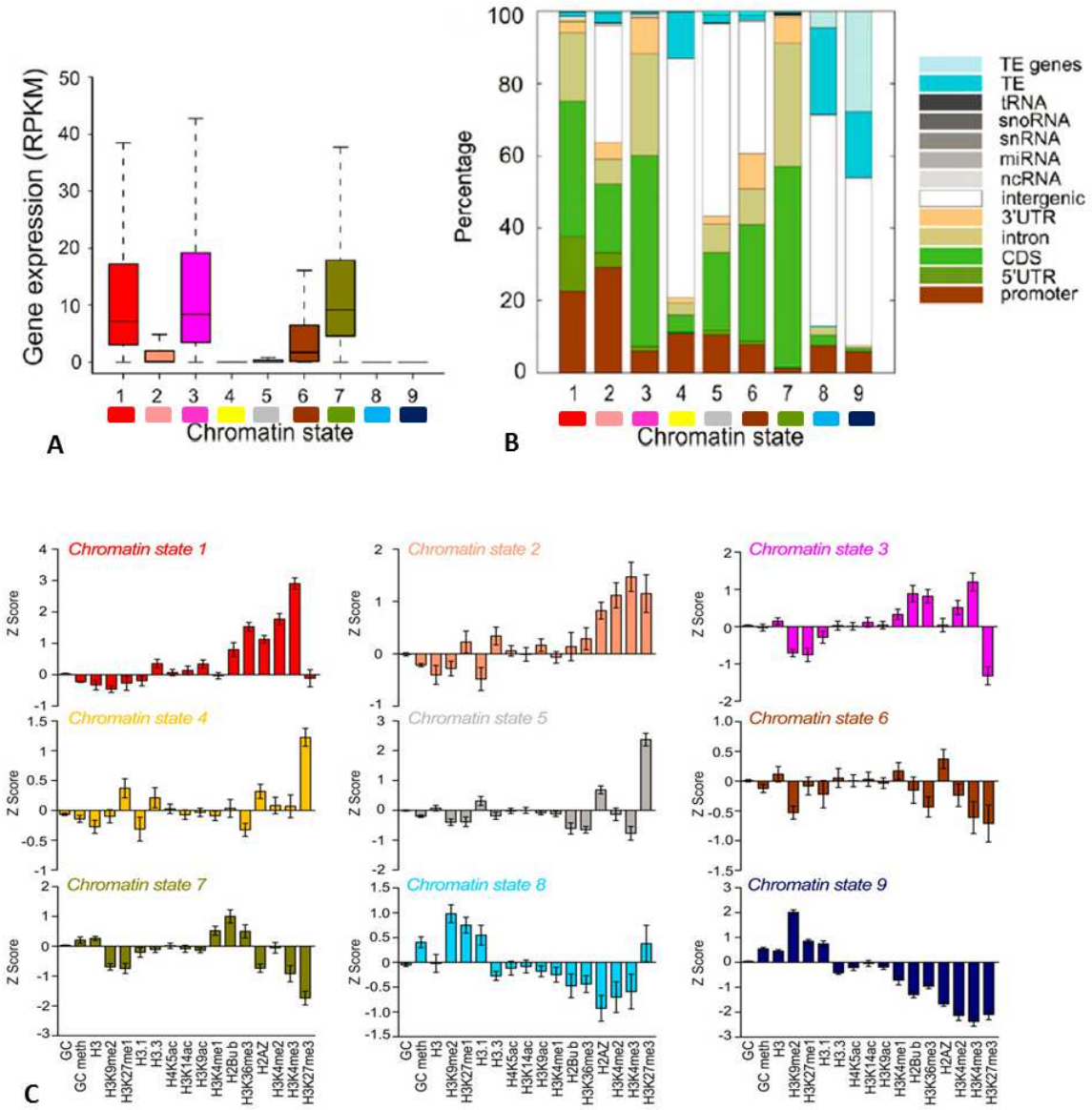


Figure 4.2: A) Relationship between gene expression level and chromatin state; B) Relationship between genomic elements and chromatin states; C) Genome-wide annotation of the *Arabidopsis* chromatin defined by specific signatures. The chromatin features considered were characterized by an unique combination of values (positive and negative z-score indicate values above or below the average in the genome, respectively). Error bars represent the SE of the mean. (Adaptation from Sequeria-Mendes *et al*., 2014)

Moreover, Sequeria-Mendes *et al*. went in deep and considered the distribution of the histone modifications along specific regions like transcriptional starting site (TSS), gene bodies and

transcriptional terminator site (TTS). Since our research objective is to determine the relation between histone modifications, in particular the activating modifications H3K9ac and H3K4me3, and gene expression, we studied the distribution of these marks in published data of *Arabidopsis*. The data published in 2014 by Gutierrez group confirmed previous literature data (Lauria and Rossi, 2011) and identified in TSS of genes the region of interest (Figure 4.3). In this region not only the marks H3K9ac and H3K4me3, but also chromatin state 1 and state 3 were identified like open chromatin states. In fact, the researchers found that each of the nine chromatinic states had very strong propensities to associate with only a subset of other states with similar expression features.

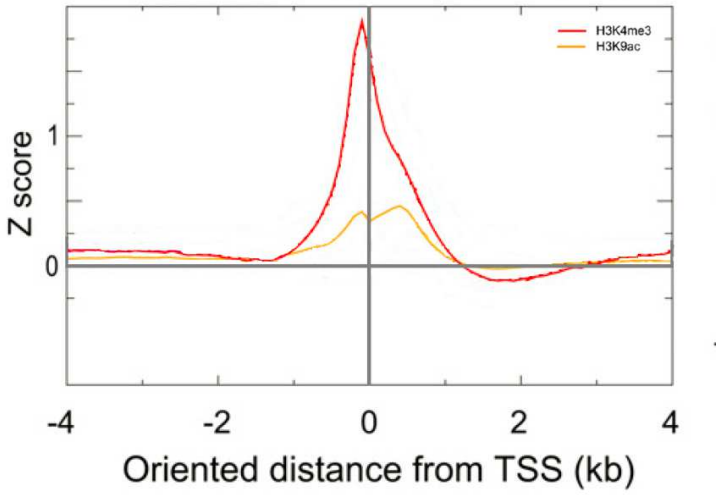


Figure 4.3: Estimation of the relative enrichment of the histone marks H3K4me3 and H3K9ac around the TSS (Adaptation from Sequeria-Mendes *et al.*, 2014)

Starting from available literature data for model plant species described above, we used a genome wide ChIP sequencing approach to study the relations between histone modifications and gene expression in two different grapevine rootstocks one susceptible to drought stress (101.14) and one resistant (M4) grown *in vitro* controlled conditions.

So far, our study is the first ChIP-seq ever produced for grapevine tissues.

2 Materials and methods

2.1 Plant material and experimental design

In order to maintain grapevine rootstocks 101.14 and M4 in controlled and stable growth conditions, a pool of plants for each genotype was grown *in vitro* like described in details in Chapter II. In April 2014, leaf material from twenty three plants of 101.14 and twenty five plants of M4, respectively, was collected in two biological replicates and frozen in liquid nitrogen. The tissues were maintained at -80°C until RNA extraction (Chapter II) and Chromatin Immunoprecipitation (ChIP) assay. ChIP protocol was described in details in Chapter III, in this section we focalize the attention on the results obtained from chromatin immunoprecipitation with antibodies against H3K9ac and H3K4me3 modifications and the subsequent sequencing. In particular, we used different approaches to clarify the relation between histone modification profiles (ChIP-seq) and transcriptome analysis (mRNA-seq) in the two grapevine rootstocks.

2.2 ChIP-Seq of the two grapevine rootstocks

Chromatin extracted from leaf materials of the two grapevine rootstocks was immunoprecipitated with antibodies against histone modifications H3K9ac (Millipore Cat. 07-352) and H3K4me3 (Active Motif Cat. 39159). Immunoprecipitated enriched DNA and two Input RC PK samples, one for each genotype, were frozen and sent to the sequencing center, IGA Techology Services. For each ChIP-Seq library construction, three ChIPs were performed in parallel and merged together to obtain a sufficient amount of enriched DNA (between 1 to 10 ng). The adapter-ligated libraries were prepared using the Paired-End Sample Preparation kit (NuGEN) and single-read sequencing was performed on the HiSeq2000 (Illumina) generating 50-baselenght reads. Immunoprecipitated samples were sequenced with a 6-plex sequencing (20M of reads) while Input RC PK were sequenced with a 2-plex sequencing (60M of reads). Reads from single-read runs were processed with a sequencing pipeline consisting of base calling using Illumina Pipeline. The trimming of the sequences was performed using ERNE v1.4 with minimum

value used by Mott-like trimming 20, the minimum mean value to accept a trimmed sequence was 20 and minimum sequence length after trimming 40 (Vezzi *et al*., 2012). The mapping of the reads on the reference genome PN40024 was performed with Bowtie2 v2.0.2 (Langmead *et al*., 2009) with all default parameters. Unique mapping read selection was performed on alignments and only reads with map quality (mapq) major or equal to 10 were considered for the subsequent analyses.

The use of PN40024 12x V1 like reference genome for reads alignment rather than the genomes of the two rootstocks has long been discussed in transcriptome section (Chapter II). A control of the alignments of the reads of the different genotypes was anyway performed through the sequencing and alignment of the two samples Input RC PK (one for each genotype) versus PN40024, in fact Input RC PK is none other than the genome sequencing of each rootstock without any enrichment.

2.3 Selection of region 1000bp above start codon (ATG)

As previously discussed in the introduction of this chapter, literature data of histone modifications studied in our research project established that these marks lied preferentially in the transcriptional starting site (TSS) of the genes. However, in the grapevine reference genome PN40024 12x V1 the annotations both of TSSs and often the untrantranslated regions (UTR) are still missing. We arbitrarly decided to keep a window of 1000 base pair above the translation start codon (ATG) and the number of reads that aligned in this region was considered for each gene in each of the sample sequenced. Hereinafter, we referred to the region 1000 bp above the ATG calling it “upstream region” or more simply “upstream”.

2.3.1 Normalization of reads upstream

The normalization of reads upstream for each gene was performed by calculating the *ratio* between the reads of one gene and the number of total reads present in sequencing considered. An example of the formula used is reported in Equation 4.1.

$$Reads\ upstream\ normalized\ for\ Gene\ A\_H3K9ac = \frac{Reads\ upstream\ Gene\ A\_H3K9ac}{Total\ reads\ upstream\ \_H3K9ac}$$

Equation 4.1: Example of normalization performed for an hypothetically gene called "gene A" in the H3K9ac sequencing

2.3.2 Evaluation of enrichment and depletion

Once normalized the reads value of each gene respect its sequencing experiment, we performed an evaluation of enrichment, or depletion, of the immunoprecipitated in respect to the Input RC PK data (Equation 4.2).

$$Evaluation\ of\ enrichment\ or\ depletion = \frac{Reads\ upstream\ normalized\ for\ Gene\ A\_H3K9ac}{Reads\ upstream\ normalized\ for\ Gene\ A\ Input\ RC\ PK}$$

Equation 4.2: Example of evaluation of enrichment or depletion for an hypothetic gene called "gene A" immunoprecipitated with αH3K9ac in respect to Input RC PK sequencing

This approach was very useful for two main reasons:

- the comparison of the immunoprecipitated gene value in respect to Input RC PK situation could confirm that the gene considered was covered by the alignment using PN40024 as reference;
- the evaluation of the enrichment, or the depletion, allowed us to perform a comparison between the same gene in samples immunoprecipitated with different antibodies and in different genotypes.

We decided to restrict our analyses only to the 18 796 genes selected during the transcriptome analysis that presented values of mean FPKM superior to one for at least one of the two genotypes. On this pool of genes we focalized our analyses on evaluation of enrichment and depletion for the identification of genes that could be associated with histone modification in different manner in the two genotypes. To achieve this objective we selected genes to set up three main gene groups, one for each histone modification and one for the merge of the two marks, where were clustered genes enriched in 101.14 and depleted in M4 and vice versa. Each of these groups was subdivided in 3 sub-groups: a gene was labeled like “enriched” when the value of immunoprecipitated reads (Ab) was higher or equal to three respect to the Input RC PK

(Input) value, while we considered three levels of depletions. The situation was schematically represented in Figure 4.4.

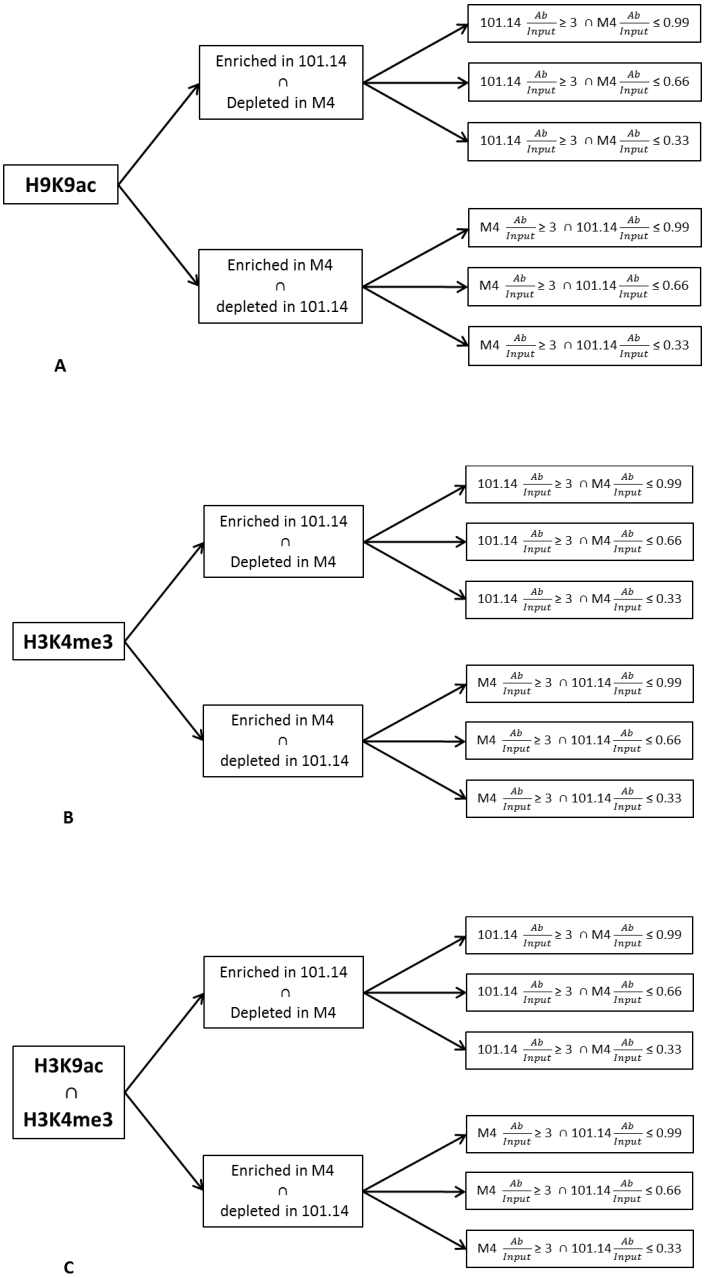


Figure 4.4: Representation of the groups set up for the enrichment/depletion analyses of histone modification H3K9ac (A), H3K4me3 (B) and for the merge of the two marks (C).

The depletion 0.99 was the one that made the comparison between genotypes robust. It was the evidence that in the genotype not enriched the value of the reads in immunoprecipitated and in Input RC PK samples was about the same indicating that the gene was coverage by the alignment in PN40024. The depletion 0.66 was considered for the individuation of genes enriched three times in one genotypes that presented a *ratio* $\frac{Ab}{Input} \leq \frac{2}{3}$ on the other rootstock. The depletion of 0.33 was the highest stringent range and identified genes enriched in one genotype and heavily depleted in the other $(\frac{Ab}{Input} \leq \frac{1}{3})$.

Once identified these twelve subgroups of genes we related the expression level of each gene in the two genotypes and we studied if the fold change ($\log_2 \frac{101.14}{M4}$) could be associated to the modifications in a statistically significant manner within each subgroup. For this purpose we used the Welch's t-test and we performed a study of cumulated distribution function and percentage frequency .

The Welch's t-test (Welch, 1947) tests if the difference in means between two groups is equal to a hypothesized value and it is preferred to Student's t-test when several different population variances are involved. Welch's t-test assumes that the populations are normally distributed. Due to the central limits theorem the test may still be useful when the assumption is violated if the sample sizes are equals and the distribution of values presents similar shape. As said before the test does not assume that the population variances were equals. Statistical analyses of t-Welch were performed with Analyse-it for Microsoft Excel 3.90.2.

2.3.3 Study of Differentially Expressed Genes (DEGs) in relation to histone modifications

Genes that were enriched or depleted for both the marks considered at the same time could represent the group more interesting for the identification of relationships between histone modifications and gene expression. At the light of the statistical analyses performed, we identified in H3K9ac ∩ H3K4me3 3x_enrichment/0.99_depletion the most interesting subgroup of genes and the Differentially Expressed Genes (DEGs) within this group.

These genes were studied in details through different approaches:

- alignments of gene sequences (inclusive of upstream regions) between the experimental genotypes and the reference genome to confirm good alignments among genotypes (Geneious Pro v3.6.2);
- visualization of the ChIP-Seq reads alignments by Integrative Genomes Viewer (IGV v 2.3.40) software;
- genes functional annotation and comparison with *Arabidopsis thaliana* orthologues genes and Gene Ontology (GO) term analyses;
- comparisons between enriched DEGs of *in vitro* grown plants and DEGs of plants grown in controlled conditions in greenhouse (as previously described in details in Chapter II).

3 Results

3.1 ChIP protocol

In this section are reported the results (Figure 4.5) for each checkpoint of chromatin immunoprecipitation protocol (described in details in Chapter III) of each chromatin sample that was subsequently used for sequencing.

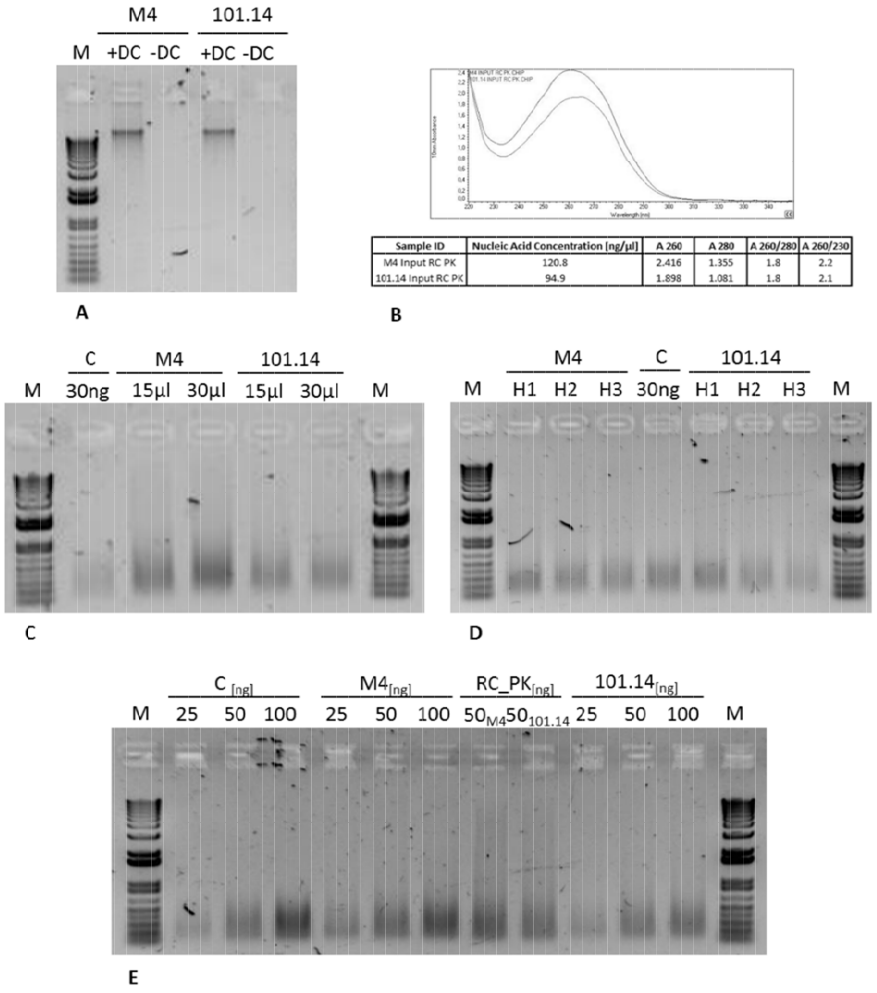


Figure 4.5: Quality control and quantification of chromatin used for ChIP-seq. A) Control of the fixation and reverse crosslinking procedures, 15 µl of sample were charged for each situation (decrossed, +DC, or non decrossed, -DC). B) NanoDrop plot and absorption data of Input RC PK of the samples. C, D, E) Input RC analyses: C) 30 ng of calibrator's shared chromatin (C 30 ng) and volume variables (15 µl and 30 µl) of chromatin with unknown concentrations for each sample were charged on gel. D) Visualization on agarose gel of shared chromatin of 30 ng of a calibrator (C) and 30 ng theoretically charged with the three hypothesis for each genotype that considered the unknown samples

concentrated 4 ng/µl (H1), 5.5 ng/µl (H2) and 7 ng/µl (H3). E) Visualization on agarose gel of different amount of calibrator (25,50 and 100 ng) and the same amount of the samples' chromatin charged considering an hypothetically concentration of 5 ng/µl for M4 and 4 ng/µl for 101.14 and 50 ng of Input RC PK of each sample whose concentrations were estimate from NanoDrop data (Fig 4.5B). [Gels 1xTAE, 0.8% agarose (Agarose Ultrapure, Life Technologies), molecular weight marker (M): 1Kb Plus Ladder (Life Technologies)]

Input RC PK and Input RC allowed us to assert that the concentration of the two sample was 4.5 ng/µl for M4 and 2.7 ng/µl for 101.14.

Immunoprecipitations were performed like described in Chapter III and real-time PCR amplifications were studied to establish the quality and quantity of the enriched DNA. The amplifications were performed with a couple of primer designed on 5'UTR of the gene 12-oxophytodienoate reductase 2 called Oxo18_20_5'UTR_Ex1 (Fw:TCTAAATGGGCTTCAACCTGT, Rev:CAGAGACGCCATTCATTCT). The immunoprecipitation and the real-time PCRs were done without the sample of control of immunoprecipitation procedure (NoAb). This choice was necessary for the large amount of material that was used for the immunoprecipitation process leading to sequencing.

Real-time PCRs were executed on 1 µl of immunoprecipitated enriched DNA, obtained by pooling three parallel immunoprecipitations, and on 0.9 ng of Input RC PK that was used like calibration for the quantification of enriched DNA. The amplification plots are reported in Figure 4.6, while the estimations of enriched DNA performed through the comparison with Input RC PK are reported in Table 4.1.

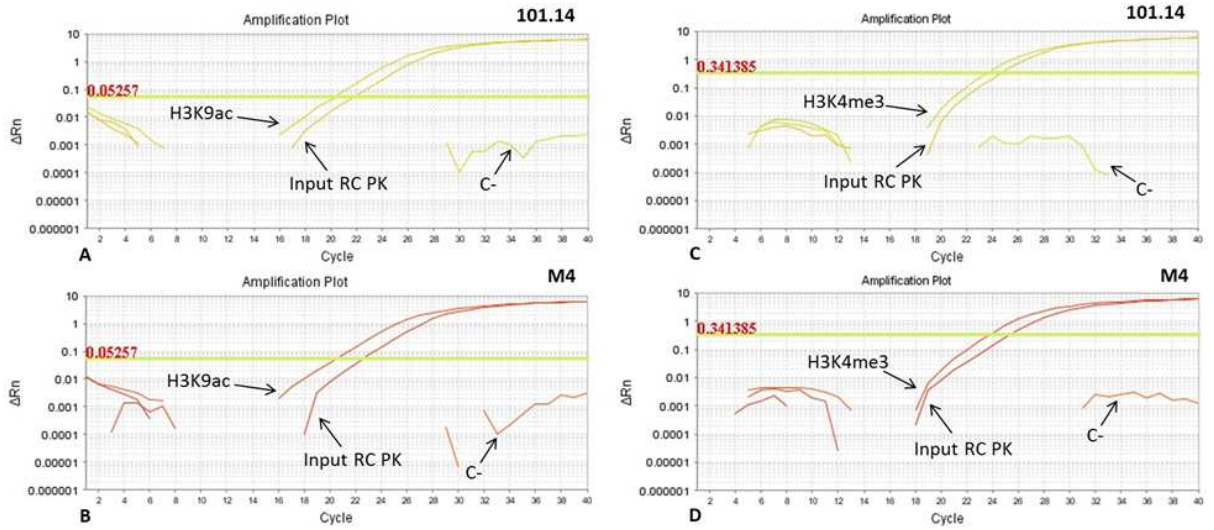


Figure 4.6: Real-time PCR amplification plots of 0.9 ng of Input RC PK and 1 µl of immunoprecipitated DNA αH3K9ac for 101.14 (A) and M4 (B) and αH3K4me3 for 101.14 (C) and M4 (D). The amplifications were performed with primer Oxo18_20_5'UTR_Ex1

Antibody	Sample Name	Cт Mean	ΔCт Input RC PK vs Ab	Fold Difference	Concentration of Enriched DNA [ng/µl]
H3K9ac	101.14_Input RC PK	22.006	1.457	2.746	2.472
	101.14_Ab	20.549			
	M4_Input RC PK	22.732	2.111	4.320	3.888
	M4_Ab	20.620			
H3K4me3	101.14_Input RC PK	24.791	1.068	2.097	1.887
	101.14_Ab	23.723			
	M4_Input RC PK	25.451	1.656	3.151	2.836
	M4_Ab	23.795			

Table 4.1: Real-time PCR data on Input RC PK and enriched DNA. C_T mean value, ΔC_T, Fold difference and estimation of enriched DNA. The amplifications were performed with primer Oxo18_20_5'UTR_Ex1.

3.2 ChIP sequencing and reads mapping on reference genome

Chromatin immunoprecipitations were executed using antibodies against acetylation of lysine 9 of histone 3 (H3K9ac) and against tri-methylation of lysine 4 of histone 3 (H3K4me3). Next Generation Sequencing technologies were used to perform a whole genome study of histone modifications distribution in the two rootstocks. ChIP-sequencing was performed at IGA technology services applying Illumina Technology HiSeq2000, single-end reads of 50bp of length. For each histone modification of each genotype only one sequencing was done on a sample that

derived from the merge of three different ChIP assays carried out in parallel. It is worth to remember that each immunoprecipitated sample derived from 12 parallel fixation and about 70 aliquots that were sonicated in parallel, to obtain three aliquots of 10 µg of chromatin. For each genotype was sequenced also an Input RC PK sample that was none other than the genome sequencing of each rootstock without any enrichment.

The sequencing produced for 101.14 H3K9ac 9.9 million of unique reads with mapq ≥10 (24.8 M of raw reads) for 101.14 H3K9ac, 23.8 M of unique reads with mapq ≥10 for 101.14 H3K4me3 (61.9 M of raw reads) and 23.2 M for 101.14 Input RC PK (65.7 M of raw reads). For M4 were obtained 15.9 M of unique reads with mapq ≥10 (40.8 M of raw reads) for H3K9ac, 10.4 M for H4K4me3 (26.7 M of raw reads) and 24.1 M for Input RC PK. All data are reported in Table 4.2.

Sample	Raw reads (M)	Unique reads (M)	Unique reads mapq≥10 (M)
101.14_H3K9ac	24.8	26.5	9.9
101.14_H3K4me3	61.9	40.5	23.8
101.14_InputRC_PK	65.7	24.6	23.2
M4_H3K9ac	40.8	66.5	15.9
M4_H3K4me3	26.7	61.5	10.4
M4_InputRC_PK	66.7	65.3	24.1

Table 4.2: ChIP-Sequencing reads data

The reads obtained from the sequencing were analyzed for the identification of a possible presence of contaminants. The data were of good quality, in fact a mean of about 98% of the reads were attributable to the genus *Vitis* (Table 4.3).

Sample	Genus	% of reads
101.14_H3K9ac	*Vitis*	99.1
101.14_H3K4me3	*Vitis*	98.8
101.14_InputRC_PK	*Vitis*	96.1
M4_H3K9ac	*Vitis*	97.9
M4_H3K4me3	*Vitis*	99.2
M4_InputRC_PK	*Vitis*	97.1

Table 4.3: Percentage of reads attributable to the genus *Vitis* in each of the samples sequenced

The number of genes coverage by Input RC PK was 24 708 for 101.14 and 24 910 for the experimental rootstock M4. In order to make our analyses more robust, we decided to limit the studies of histone modifications distribution and correlation to gene expression to the 18 796

genes selected during mRNA-seq analyses for value of mean FPKM superior to one in at least one of the rootstocks in exam (see Chapter II for details).

3.3 Evaluation of enrichment and depletion in relation to gene expression

In section 2.3.2 were defined the choices made for the identification of different subgroup of genes enriched ($\frac{Ab}{Input}$ ≥3) in the histone modification of interest in one genotype and depleted in the other ($\frac{Ab}{Input}$ ≤0.99, $\frac{Ab}{Input}$ ≤0.66 and $\frac{Ab}{Input}$ ≤0.33). In this section we describe in details the results obtained for each subgroup once that the histone modification was kept in relation to gene expression as fold change (logarithm to the base 2 of the *ratio* between FPKM of the genes in the two genotypes, $FC = \log_2 \frac{101.14}{M4}$).

The analyses were based on the comparisons between fold change of genes in the subgroup created with the same filter enriched in one genotype *versus* the other, for instance: FC of genes " $\frac{Ab}{Input}$ ≥3_101.14 ∩ $\frac{Ab}{Input}$ ≤0.99_M4" *versus* FC of genes " $\frac{Ab}{Input}$ ≥3_M4 ∩ $\frac{Ab}{Input}$ ≤0.99_101.14").

The analyses were performed with two complementary approaches: one involved the use of a Welch's t-test for the evaluation of the significance of the difference between the means of FC values of the two groups compared. The second approach involved the study of cumulated distribution function and frequency distribution for the identification of different cumulate frequency and percentage frequency distributions.

3.3.1 H3K9ac

Genes enriched of histone modification H3K9ac and depleted of 0.99 were 258 for 101.14 and 340 for M4. The depletion value of 0.66 identified 148 genes enriched in 101.14 and 280 in M4. The subgroup of genes with the highest level of depletion (0.33) identified 80 genes enriched in 101.14 and 221 in M4. The values of statistical analyses within each group are reported in Table 4.4. The p-value allowed us to reject the null hypothesis in favor of the alternative hypothesis at the 5% significance level for the depletion range 0.99-0.66; these results indicated a significative difference between the means of groups considered. For the most strict depletion level, we could not reject the null hypothesis, maybe for the different numerosity of the two groups with subsequent inefficiency of the test rather than for an effective lack of difference between FCs. The different distribution of fold change values within the three subgroups were also visualized with the cumulative distribution function and frequency distribution that are reported in Figure 4.7.

log_2FC (101.14/M4)	N	Mean	Mean SE	Variance	SD	Welch's t-test	DF	p-value
101.14_3x_M4≤0.99	258	0.11	0.07	1.19	1.09	-1.73	591.64	**0.04**
M4_3x_101.14≤0.99	340	-0.08	0.09	2.47	1.57			
101_3x_M4≤0.66	148	0.15	0.10	1.48	1.22	-1.80	388.15	**0.04**
M4_3x_101≤0.66	280	-0.11	0.10	2.87	1.69			
101.14_3x_M4≤0.33	80	0.15	0.12	1.23	1.11	-1.37	215.70	0.09
M4_3x_101.14≤0.33	221	-0.08	0.11	2.91	1.70			
μΔ = μM4_enriched_101.14_depleted- μ101.14_enriched_M4_depleted H0: δ ≥ 0The difference between the means of the populations is greater than or equal to 0 H1: δ < 0 The difference between the means of the populations is less than 0								

Table 4.4: Statistical analyses of t-Welch performed with Analyse-it for Microsoft Excel 3.90.2 on genes enriched 3x and depleted for histone modification H3K9ac.

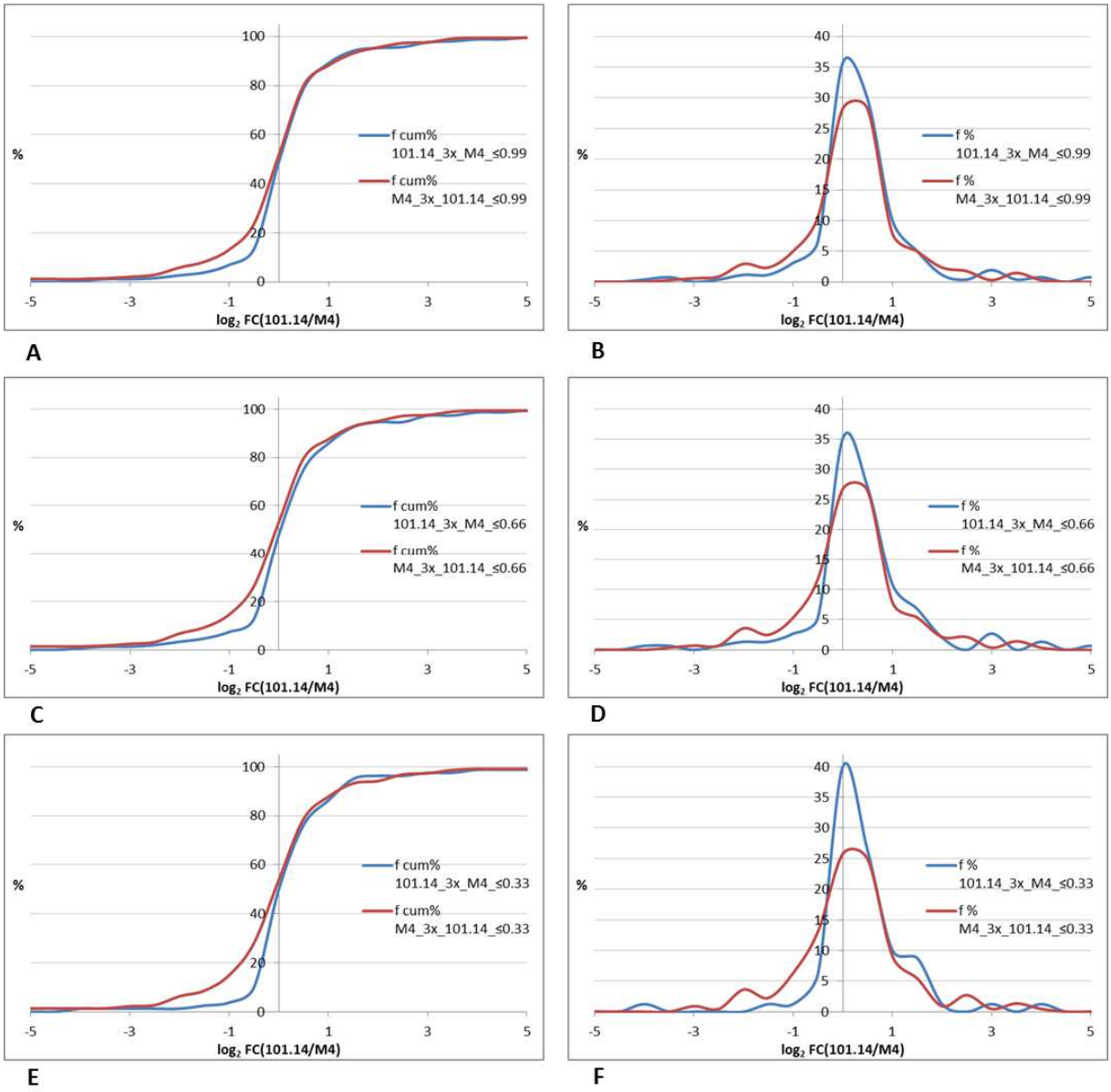


Figure 4.7: Cumulative distribution (A, C, E) and frequency function (B, D, F) of fold change plots of the three subgroups of genes enriched and depleted for H3K9ac. A, B) Distributions of genes enriched 3x and depleted 0.99; C, D) Distributions of genes enriched 3x and depleted 0.66; E, F) Distributions of genes enriched 3x and depleted 0.33.

The cumulative distribution function showed an effective deviation for each subgroup in genes enriched in M4 whose distribution resulted shifted to negative values of fold change. That could indicate that for these genes there was a propensity to negative values of fold change

(remembering that FC was calculated like $\log_2 \frac{101.14}{M4}$) and then to an overexpression in M4 genotype. This trend clearly emerged once observing the values of cumulative distribution and frequency function of fold change reported in Table 4.5. For instance, the values of cumulated percentage frequency for M4 at fold change value of -1 were 13.2% (3x _0.99), 14.6% (3x_0.66) and 14.9% (3x_0.33) which deviated from the corresponding values of 101.14: 7.0%, 7.4% and 3.8%. The relationship between histone modification H3K9ac and gene expression showed by M4 enriched genes didn't seem to be present with the same strength in 101.14 enriched genes.

Class	f cum% 101.14_3x_ M4_ ≤0.99	f cum% M4_3x_ 101.14_≤0.99	f % 101.14_3x_ M4_≤0.99	f % M4_3x_ 101.14_≤0.99	f cum% 101.14_3x_ M4_≤0.66	f cum% M4_3x_ 101.14_≤0.66	f % 101.14_3x_ M4_≤0.66	f % M4_3x_ 101.14_≤0.66	f cum% 101.14_3x_ M4_≤0.33	f cum% M4_3x_ 101.14_≤0.33	f % 101.14_3x_ M4_≤0.33	f % M4_3x_ 101.14_≤0.33
-5.0	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-4.5	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-4.0	0.4	1.2	0.4	0.0	0.7	1.4	0.7	0.0	1.3	1.4	1.3	0.0
-3.5	1.2	1.5	0.8	0.3	1.4	1.8	0.7	0.4	1.3	1.4	0.0	0.0
-3.0	1.2	2.1	0.0	0.6	1.4	2.5	0.0	0.7	1.3	2.3	0.0	0.9
-2.5	1.6	2.9	0.4	0.9	2.0	3.2	0.7	0.7	1.3	2.7	0.0	0.5
-2.0	2.7	5.9	1.2	2.9	3.4	6.8	1.4	3.6	1.3	6.3	0.0	3.6
-1.5	3.9	8.2	1.2	2.4	4.7	9.3	1.4	2.5	2.5	8.6	1.3	2.3
-1.0	7.0	13.2	3.1	5.0	7.4	14.6	2.7	5.4	3.8	14.9	1.3	6.3
-0.5	13.6	23.8	6.6	10.6	12.8	26.4	5.4	11.8	10.0	28.1	6.3	13.1
0.0	**49.2**	**52.1**	**35.7**	**28.2**	**48.0**	**53.2**	**35.1**	**26.8**	**50.0**	**53.8**	**40.0**	**25.8**
0.5	79.1	80.3	29.8	28.2	75.0	79.6	27.0	26.4	76.3	78.7	26.3	24.9
1.0	89.1	88.2	10.1	7.9	85.8	87.5	10.8	7.9	86.3	87.8	10.0	9.0
1.5	94.2	93.2	5.0	5.0	92.6	92.9	6.8	5.4	95.0	93.2	8.8	5.4
2.0	95.3	95.6	1.2	2.4	94.6	95.0	2.0	2.1	96.3	94.1	1.3	0.9
2.5	95.7	97.4	0.4	1.8	94.6	97.1	0.0	2.1	96.3	96.8	0.0	2.7
3.0	97.7	97.6	1.9	0.3	97.3	97.5	2.7	0.4	97.5	97.3	1.3	0.5
3.5	98.1	99.1	0.4	1.5	97.3	98.9	0.0	1.4	97.5	98.6	0.0	1.4
4.0	98.8	99.4	0.8	0.3	98.6	99.3	1.4	0.4	98.8	99.1	1.3	0.5
4.5	98.8	99.4	0.0	0.0	98.6	99.3	0.0	0.0	98.8	99.1	0.0	0.0
5.0	99.6	99.4	0.8	0.0	99.3	99.3	0.7	0.0	98.8	99.1	0.0	0.0

Table 4.5: Cumulative distribution and frequency function of fold change values in the three subgroups of genes enriched and depleted for H3K9ac

In Figure 4.7 and Table 4.5 are reported the values of cumulated frequency and frequency percentage between fold change -5 and +5; the representation of all the fold change classes (-12 +12) are reported in supplementary materials (Figure 4.15 and Table 4.15).

3.3.2 H3K4me

Genes enriched in histone modification H3K4me3 and depleted of 0.99 were 355 for 101.14 and 286 for M4. The value of depletion 0.66 identified 281 genes enriched in 101.14 and 180 in M4. The subgroup of genes with the highest level of depletion (0.33) identified 226 genes enriched in 101.14 and 98 in M4. The values of statistical analyses within each group are reported in Table 4.6. The p-value force to not reject the null hypothesis at the 5% significance level for both the three subgroups; these results indicate a no-significative difference between the means of groups considered.

The distribution of fold change values within the three subgroups, their cumulative distribution function and frequency distribution are reported in Figure 4.8.

log_2FC (101.14/M4)	N	Mean	Mean SE	Variance	SD	Welch's t-test	DF	p-value
101.14_3x_M4≤0.99	355	0.04	0.07	1.92	1.39	-0.48	597.02	0.32
M4_3x_101.14≤0.99	286	0.00	0.05	0.71	0.84			
101.14_3x_M4≤0.66	281	0.04	0.09	2.26	1.50	-0.70	443.51	0.24
M4_3x_101.14≤0.66	180	-0.03	0.06	0.62	0.79			
101.14_3x_M4≤0.33	226	0.08	0.11	2.61	1.61	-1.12	313.76	0.13
M4_3x_101.14≤0.33	98	-0.07	0.08	0.68	0.82			
μΔ = μM4_enriched_101.14_depleted- μ101.14_enriched_M4_depleted H0: δ ≥ 0The difference between the means of the populations is greater than or equal to 0 H1: δ < 0 The difference between the means of the populations is less than 0								

Table 4.6: Statistical analyses of t-Welch performed by Analyse-it for Microsoft Excel 3.90.2 on genes enriched 3x and depleted for histone modification H3K4me3

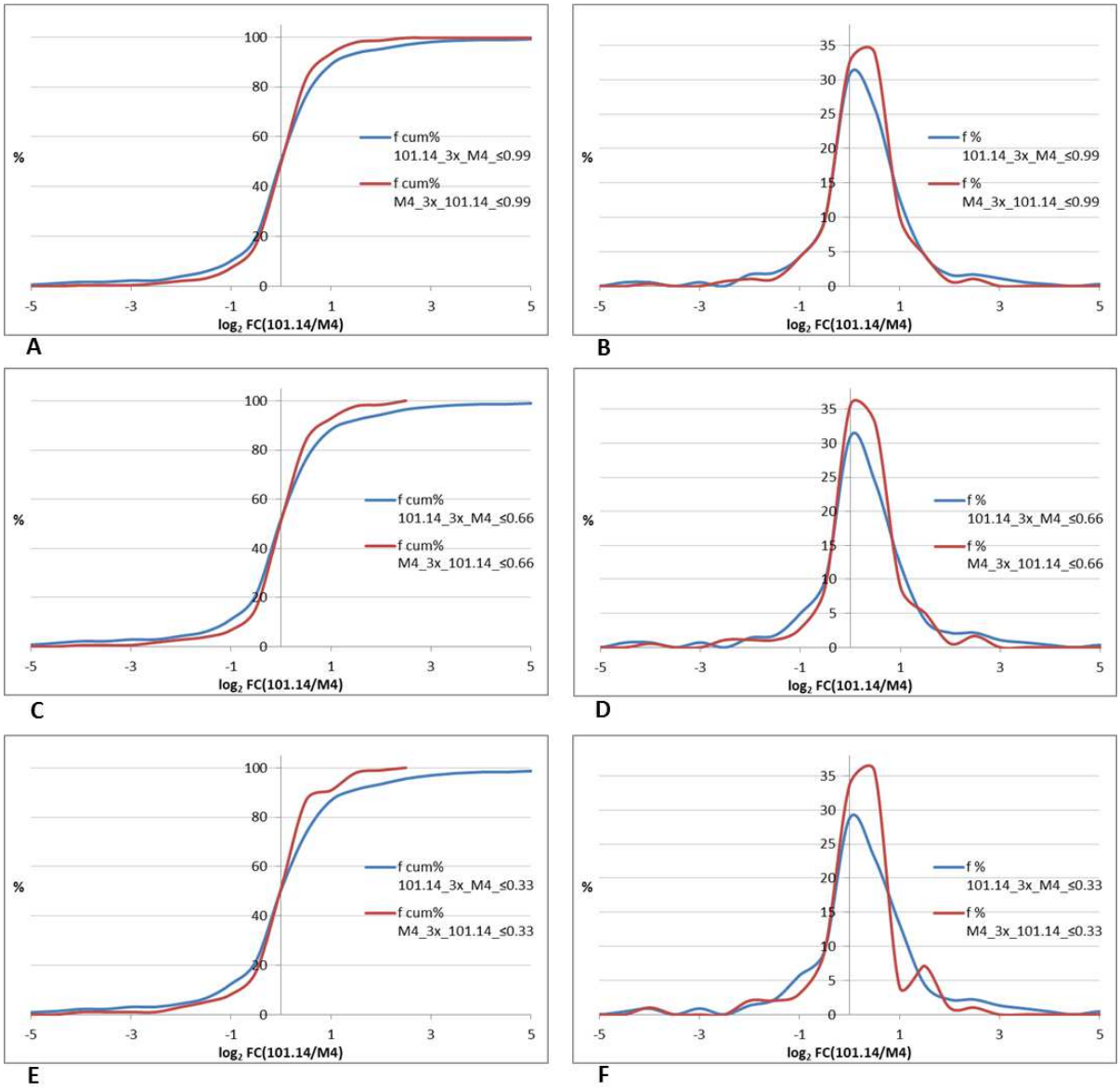


Figure 4.8: Cumulative distribution (A, C, E) and frequency function (B, D, F) of fold change plots of the three subgroups of genes enriched and depleted for H3K4me3. A, B) Distributions of genes enriched 3x and depleted 0.99; C, D) Distributions of genes enriched 3x and depleted 0.66; E, F) Distributions of genes enriched 3x and depleted 0.33.

Although statistical analyses shown a lack of significance between the difference of fold change values of genes enriched in one genotype respect the ones enriched in the other at all the ranges of depletion, the plot of cumulative distribution function had different trends. Genes enriched in

101.14 and depleted in M4 (with both the three values of depletion's cut-off) showed values of cumulated frequencies shifted to higher values of fold change, but this trend was not confirmed by statistical analysis. The values of cumulative distribution and frequency function of Fold Change are reported in Table 4.7.

Class	f cum% 101.14_3x_ M4_ ≤0.99	f cum% M4_3x_ 101.14_ ≤0.99	f % 101.14_3x_ M4_ ≤0.99	f % M4_3x_ 101.14_≤0.99	f cum% 101.14_3x_ M4_≤0.66	f cum% M4_3x_ 101.14_≤0.66	f % 101.14_3x_ M4_≤0.66	f % M4_3x_ 101.14_≤0.66	f cum% 101.14_3x_ M4_≤0.33	f cum% M4_3x_ 101.14_≤0.33	f % 101.14_3x_ M4_≤0.33	f % M4_3x_ 101.14_≤0.33
-5.0	0.6	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.9	0.0	0.0	0.0
-4.5	1.1	0.0	0.6	0.0	1.4	0.0	0.7	0.0	1.3	0.0	0.4	0.0
-4.0	1.7	0.3	0.6	0.3	2.1	0.6	0.7	0.6	2.2	1.0	0.9	1.0
-3.5	1.7	0.3	0.0	0.0	2.1	0.6	0.0	0.0	2.2	1.0	0.0	0.0
-3.0	2.3	0.3	0.6	0.0	2.8	0.6	0.7	0.0	3.1	1.0	0.9	0.0
-2.5	2.3	1.0	0.0	0.7	2.8	1.7	0.0	1.1	3.1	1.0	0.0	0.0
-2.0	3.9	2.1	1.7	1.0	4.3	2.8	1.4	1.1	4.4	3.1	1.3	2.0
-1.5	5.9	3.1	2.0	1.0	6.0	3.9	1.8	1.1	6.6	5.1	2.2	2.0
-1.0	10.1	7.3	4.2	4.2	11.0	6.7	5.0	2.8	12.4	8.2	5.8	3.1
-0.5	19.7	16.8	9.6	9.4	21.0	15.6	10.0	8.9	21.7	17.3	9.3	9.2
0.0	**50.4**	**49.3**	**30.7**	**32.5**	**52.0**	**51.1**	**31.0**	**35.6**	**50.4**	**51.0**	**28.8**	**33.7**
0.5	76.3	83.2	25.9	33.9	76.2	83.9	24.2	32.8	73.5	86.7	23.0	35.7
1.0	89.0	93.4	12.7	10.1	88.3	92.8	12.1	8.9	86.7	90.8	13.3	4.1
1.5	93.5	97.9	4.5	4.5	92.2	97.8	3.9	5.0	91.2	98.0	4.4	7.1
2.0	95.2	98.6	1.7	0.7	94.3	98.3	2.1	0.6	93.4	99.0	2.2	1.0
2.5	96.9	99.7	1.7	1.0	96.4	100.0	2.1	1.7	95.6	100.0	2.2	1.0
3.0	98.0	99.7	1.1	0.0	97.5		1.1	0.0	96.9		1.3	0.0
3.5	98.6	99.7	0.6	0.0	98.2		0.7	0.0	97.8		0.9	0.0
4.0	98.9	99.7	0.3	0.0	98.6		0.4	0.0	98.2		0.4	0.0
4.5	98.9	99.7	0.0	0.0	98.6		0.0	0.0	98.2		0.0	0.0
5.0	99.2	99.7	0.3	0.0	98.9		0.4	0.0	98.7		0.4	0.0

Table 4.7: Cumulative distribution and frequency function of Fold Change values in the three subgroups of genes enriched and depleted for H3K4me3.

In Figure 4.8 and Table 4.7 are reported the values of cumulated frequency and frequency percentage between fold change -5 and +5; the representation of all the fold change classes (-12 +12) are reported in supplementary materials (Figure 4.16 and Table 4.16).

3.3.3 H3K9ac ∩ H3K4me

The distributions of fold change values in genes enriched/depleted at the same time for both the histone modifications under study were certainly the best approach for the identification of the relationship between histone modifications and gene expression.

Genes enriched (3x) and depleted of 0.99 for both histone modifications H3K9ac and H3K4me3 were 67 for 101.14 and 52 for M4. The value of depletion 0.66 identified 41 genes enriched in 101.14 and 31 in M4. The subgroup of genes with the highest level of depletion (0.33) identified 23 genes enriched in 101.14 and 18 in M4. The values of statistical analyses within each group are reported in Table 4.8. The p-value allowed us to reject the null hypothesis in favor of the alternative hypothesis at the 5% significance level for the depletion range 0.99-0.66; these results indicated a significative difference between the means of groups considered. For the most strict depletion we could not reject the null hypothesis maybe for low number of genes that where identified and that could not be good statistically samples.

The distribution of fold change values within the three subgroups was also studied through the cumulative distribution function and frequency distribution. Results are plotted in Figure 4.9.

log_2FC (101.14/M4)	N	Mean	Mean SE	Variance	SD	Welch's t-test	DF	p-value
101.14_3x_M4≤0.99	67	0.22	0.16	1.62	1.27	-1.89	109.90	**0.03**
M4_3x_101.14≤0.99	52	-0.13	0.10	0.56	0.75			
101.14_3x_M4≤0.66	41	0.25	0.22	2.01	1.42	-1.93	61.51	**0.03**
M4_3x_101.14≤0.66	31	-0.24	0.13	0.49	0.70			
101.14_3x_M4≤0.33	23	0.19	0.29	1.98	1.41	-1.25	36.21	0.11
M4_3x_101.14≤0.33	18	-0.25	0.19	0.66	0.81			
μΔ = μM4_enriched_101.14_depleted- μ101.14_enriched_M4_depleted H0: δ ≥ 0The difference between the means of the populations is greater than or equal to 0 H1: δ < 0 The difference between the means of the populations is less than 0								

Table 4.8: Statistical analyses of t-Welch performed by Analyse-it for Microsoft Excel 3.90.2 on genes enriched 3x and depleted for histone modification H3K9ac ∩ H3K4me3

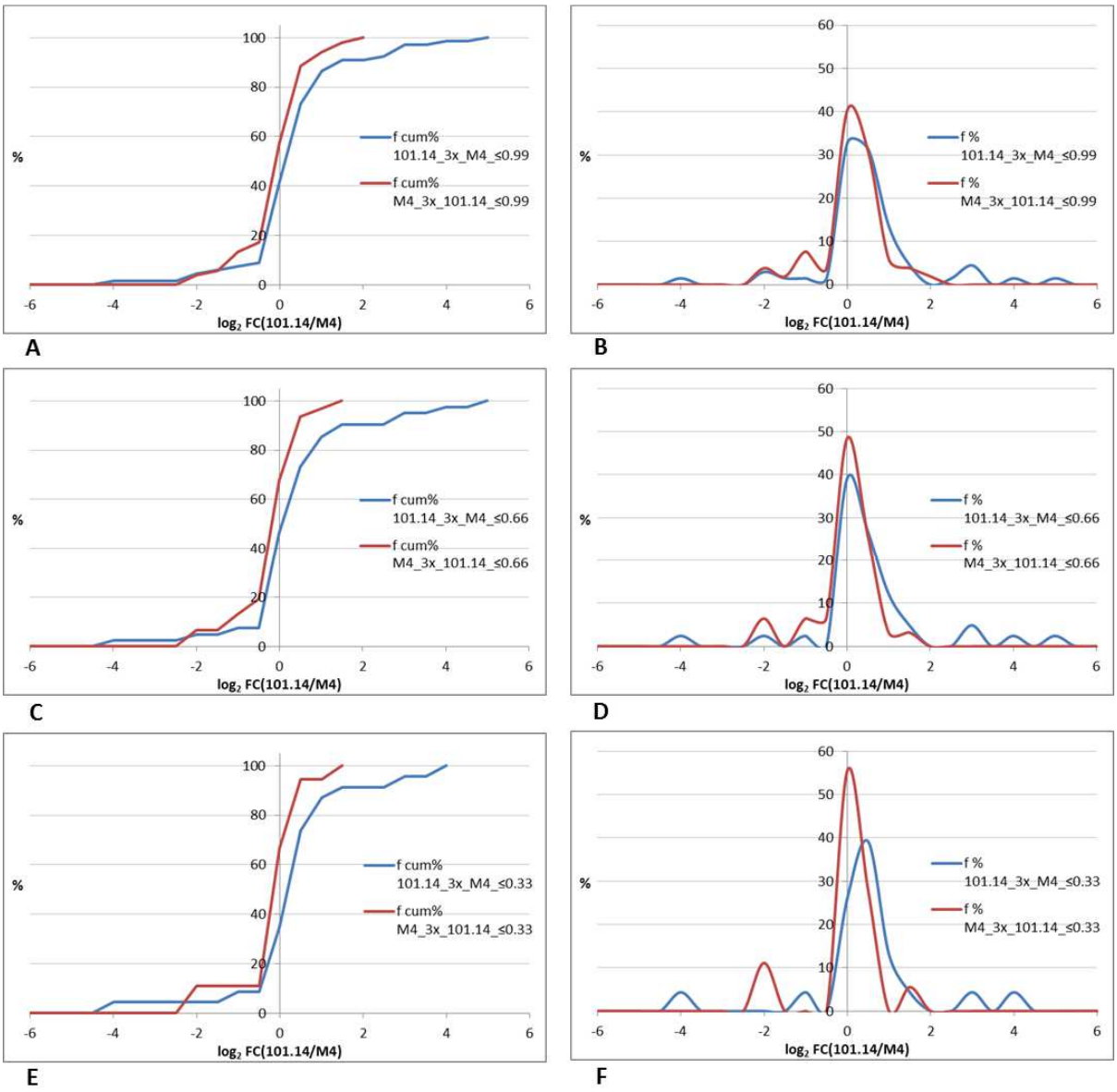


Figure 4.9: Cumulative distribution (A, C, E) and frequency function (B, D, F) of fold change plots of the three subgroups of genes enriched and depleted at the same time for both the marks in exams H3K9ac and H3K4me3. A, B) Distributions of genes enriched 3x and depleted 0.99; C, D) Distributions of genes enriched 3x and depleted 0.66; E, F) Distributions of genes enriched 3x and depleted 0.33.

The cumulative distribution function showed for each subgroup an effective deviation between genes enriched in M4 that resulted shifted to negative value of fold change and genes enriched

in 101.14 whose distribution was shifted to positive value of fold change. These trends clearly emerged once observed the values of cumulative distribution and frequency function of fold change that are reported in Table 4.9.

Class	f cum% 101.14_3x_ M4_ ≤0.99	f cum% M4_3x_ 101.14_≤0.99	f % 101.14_3x_ M4_≤0.99	f % M4_3x_ 101.14_≤0.99	f cum% 101.14_3x_ M4_≤0.66	f cum% M4_3x_ 101.14_≤0.66	f % 101.14_3x_ M4_≤0.66	f % M4_3x_ 101.14_≤0.66	f cum% 101.14_3x_ M4_≤0.33	f cum% M4_3x_ 101.14_≤0.33	f % 101.14_3x_ M4_≤0.33	f % M4_3x_ 101.14_≤0.33
-6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-5.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-4.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-4.0	1.5	0.0	1.5	0.0	2.4	0.0	2.4	0.0	4.3	0.0	4.3	0.0
-3.5	1.5	0.0	0.0	0.0	2.4	0.0	0.0	0.0	4.3	0.0	0.0	0.0
-3.0	1.5	0.0	0.0	0.0	2.4	0.0	0.0	0.0	4.3	0.0	0.0	0.0
-2.5	1.5	0.0	0.0	0.0	2.4	0.0	0.0	0.0	4.3	0.0	0.0	0.0
-2.0	4.5	3.8	3.0	3.8	4.9	6.5	2.4	6.5	4.3	11.1	0.0	11.1
-1.5	6.0	5.8	1.5	1.9	4.9	6.5	0.0	0.0	4.3	11.1	0.0	0.0
-1.0	7.5	13.5	1.5	7.7	7.3	12.9	2.4	6.5	8.7	11.1	4.3	0.0
-0.5	9.0	17.3	1.5	3.8	7.3	19.4	0.0	6.5	8.7	11.1	0.0	0.0
0.0	**41.8**	**57.7**	**32.8**	**40.4**	**46.3**	**67.7**	**39.0**	**48.4**	**34.8**	**66.7**	**26.1**	**55.6**
0.5	73.1	88.5	31.3	30.8	73.2	93.5	26.8	25.8	73.9	94.4	39.1	27.8
1.0	86.6	94.2	13.4	5.8	85.4	96.8	12.2	3.2	87.0	94.4	13.0	0.0
1.5	91.0	98.1	4.5	3.8	90.2	100.0	4.9	3.2	91.3	100.0	4.3	5.6
2.0	91.0	100.0	0.0	1.9	90.2		0.0	0.0	91.3		0.0	0.0
2.5	92.5		1.5	0.0	90.2		0.0	0.0	91.3		0.0	0.0
3.0	97.0		4.5	0.0	95.1		4.9	0.0	95.7		4.3	0.0
3.5	97.0		0.0	0.0	95.1		0.0	0.0	95.7		0.0	0.0
4.0	98.5		1.5	0.0	97.6		2.4	0.0	100.0		4.3	0.0
4.5	98.5		0.0	0.0	97.6		0.0	0.0			0.0	0.0
5.0	100.0		1.5	0.0	100.0		2.4	0.0			0.0	0.0
5.5			0.0	0.0			0.0	0.0			0.0	0.0
6.0			0.0	0.0			0.0	0.0			0.0	0.0

Table 4.9: Cumulative distribution and frequency function of fold change values in the three subgroups of genes enriched and depleted for H3K9ac and, at the same time, for H3K4me3.

The values of cumulated percentage frequency for M4 at fold change value of -1 were 13.5% (3x_0.99), 12.9% (3x_0.66) and 11.1% (3x_0.33) which deviated from the relative values of 101.14: 7.5%, 7.3% and 8.7%. The cumulated percentage frequency at fold change value 0 for M4 were 57.7% (3x _0.99), 67.7% (3x_0.66) and 66.7% (3x_0.33), while for 101.14 the values were lower: 41.8% (3x _0.99), 46.3% (3x_0.66) and 34.8% (3x_0.33). These data clearly indicated a shifting of cumulate frequency values to lower values of FC for M4 and to higher values of fold change for 101.14. Also at fold change 1 the cumulated frequency for M4 was higher than 101.14 (94.2% M4 vs 86.6% 101.14 for 3x _0.99, 96.8% M4 vs 85.4% 101.14 3x _0.66 and 94.4% M4 vs 87% 101.14 3x _0.33).

3.4 Differentially Expressed Genes (DEGs) in relation to histone modifications

We decided to focalize our attention on the groups of genes enriched 3x and depleted 0.99 at the same time for both the histone modifications under consideration. To clarify the relationships between histone modifications that act together to modulate the gene expression, we selected differentially expressed genes (genes that had values of FC ≥ 1 or ≤ -1 were considered like differentially expressed genes). This filtration identified 14 DEGs enriched of both the histone marks in 101.14 and depleted, at the same way, in M4 and 10 DEGs enriched in M4 and depleted in 101.14. In Table 4.10 the gene ID of the two DEG groups are reported.

DEGs 101.14_3x_M4≤0.99	DEGs M4_3x_101≤0.99
VIT_00s0207g00130	VIT_00s0163g00050
VIT_02s0012g00830	VIT_00s1764g00020
VIT_02s0025g02170	VIT_07s0031g00260
VIT_02s0154g00300	VIT_08s0058g00040
VIT_10s0003g04870	VIT_11s0037g00290
VIT_12s0057g00950	VIT_12s0035g00860
VIT_14s0030g00140	VIT_14s0083g00920
VIT_14s0036g01360	VIT_16s0013g00210
VIT_14s0068g00140	VIT_16s0050g01430
VIT_15s0021g00850	VIT_18s0001g11570
VIT_15s0046g01060	
VIT_17s0000g07030	
VIT_18s0001g10850	
VIT_18s0001g12190	

Table 4.10: Differentially Expressed Genes enriched/depleted for both the histone modifications, H3K9ac and H3K4me3

3.4.1 Alignment of genes and upstream regions of PN40024, 101.14 and M4

Differentially expressed gene sequences and upstream regions of 101.14, M4 and PN40024 were visualized through Geneious software for the analysis of SNPs distribution, In/Dels and other problems of alignment that could affect the quality of the analyses. This approach identified three genes, VIT_10s0003g04870, VIT_14s0030g00140 and VIT_17s0000g07030, that showed alignment problems because they mapped on more than one region of the genome. For this

reason we decided to exclude them from subsequent analyses. In Figure 4.10 are reported by way of example two alignments between gene sequences, inclusive of 1000 bp upstream region, of the two experimental genotypes and the reference genome. The alignments of all DEGs are reported in Figure 4.17-18 of supplementary materials.

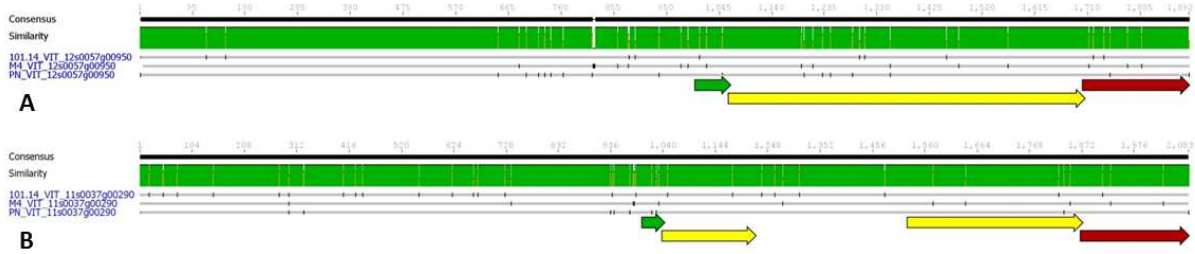


Figure 4.10: Alignment snapshots of two DEGs genes and their upstream regions in 101.14, M4 and reference genome PN40024. In figure A VIT_12s0057g00950 (DEG 101.14_3x M4≤0.99) is reported, in figure B VIT_11s0037g00290 (DEG M4_3x 101.14≤0.99) is showed

Fold change values and the logarithm to the base 2 of the *ratio* between normalized values of histone modifications ($\log_2 \frac{101.14\_H3K9ac}{M4\_H3K9ac}$; $\log_2 \frac{101.14\_H3K4me3}{M4\_H3K4me3}$) of the DEGs were kept in relation and results are reported in Table 4.11. 101.14 DEGs showed positive modulation of expression according to literature data (Sequeira-Mendes *et al*., 2014; Ha *et al*., 2013) in seven of eleven DEGs (64%), while M4 DEGs showed positive modulation of expression in seven genes of ten DEGs (70%).

	$\log_2$(101.14/M4)		
DEGs 101.14_3x_M4≤0.99	Expression	H3K9ac	H3K4me3
VIT_18s0001g10850	-4.088	20.284	17.687
VIT_15s0021g00850	-2.120	2.825	11.664
VIT_02s0025g02170	-2.100	5.429	14.839
VIT_02s0154g00300	-1.852	8.336	10.626
VIT_18s0001g12190	1.189	13.598	14.163
VIT_12s0057g00950	1.308	2.768	4.313
VIT_15s0046g01060	1.355	12.672	2.553
VIT_02s0012g00830	2.311	3.147	4.026
VIT_00s0207g00130	2.548	6.192	7.392
VIT_14s0036g01360	3.589	14.965	5.924
VIT_14s0068g00140	4.825	11.983	12.293

A

	$\log_2$(101.14/M4)		
DEGs M4_3x_101.14≤0.99	Expression	H3K9ac	H3K4me3
VIT_18s0001g11570	-2.313	-12.331	-18.551
VIT_16s0013g00210	-2.022	-4.587	-5.097
VIT_00s0163g00050	-1.729	-3.237	-13.745
VIT_11s0037g00290	-1.341	-11.796	-2.248
VIT_07s0031g00260	-1.076	-17.140	-9.911
VIT_00s1764g00020	-1.028	-17.935	-9.974
VIT_16s0050g01430	-1.019	-19.496	-10.466
VIT_14s0083g00920	1.283	-19.256	-18.126
VIT_08s0058g00040	1.366	-12.527	-2.408
VIT_12s0035g00860	1.755	-3.193	-13.032

B

Table 4.11: Fold change and value of histone modifications for DEGs enriched in 101.14 (A) and in M4 (B)

The expression levels and the histone modification values in the two groups of DEGS are plotted in Figure 4.11 (A,B). Furthermore the relationship between values of the two histone modifications were studied (C). These data allowed us the identification of a high level of correlation (0.8322, R^2=0.6926) between the values of the two histone modifications in differentially expressed genes.

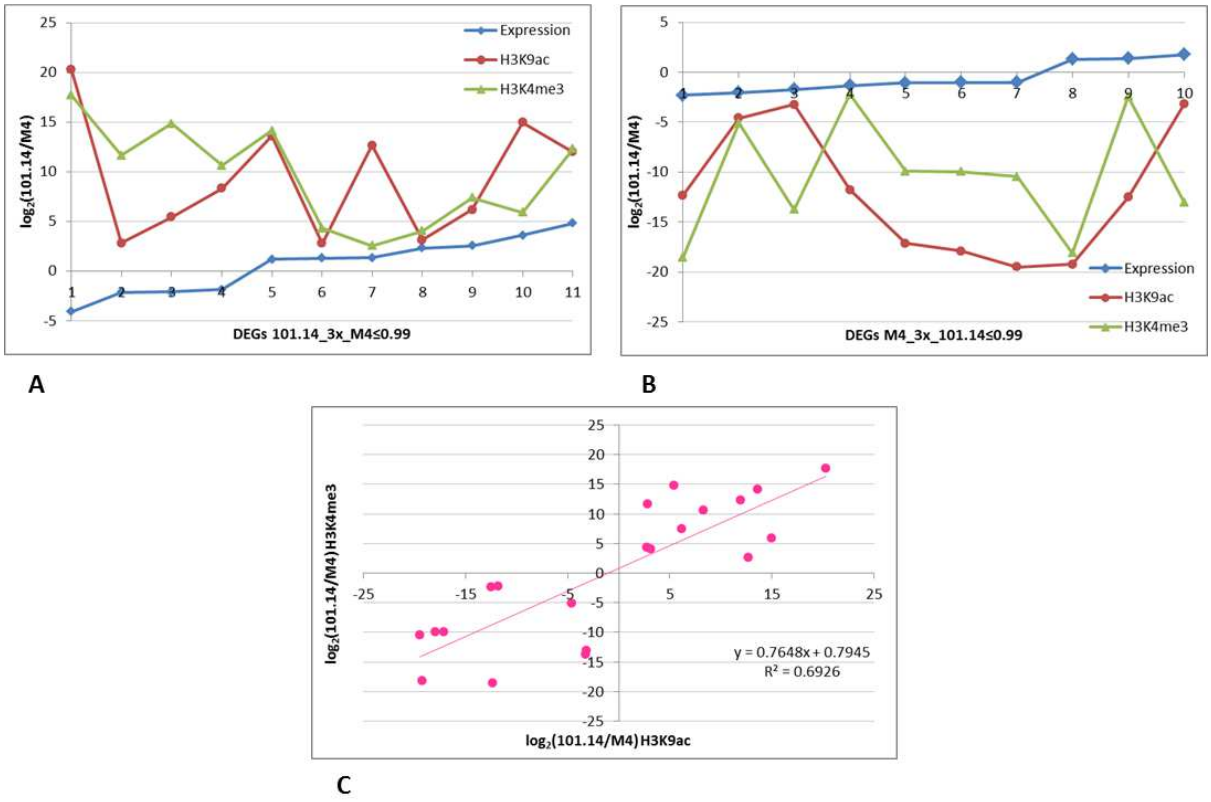


Figure 4.11: Plot of expression levels, H3K9ac and H3K4me3 values for DEGs 101.14 enriched_M4 depleted (A) and for DEGs M4 enriched_101.14 depleted (B). C) Plot of values $\log_2 \frac{101.14}{M4}$ of H3K9ac *versus* H3K4me3.

3.4.2 ChIP-seq reads alignments visualization (IGV v2.3)

Once identified the set of DEGs of interest, ChIP-seq reads alignments were analyzed in details in order to validate the data through the use of software IGV (Integrative Genomes Viewer, IGV v2.3.40). The visualizations of reads alignments confirmed the previously data of the enrichment/depletion. Two examples of reads visualization are reported in Figure 4.12.

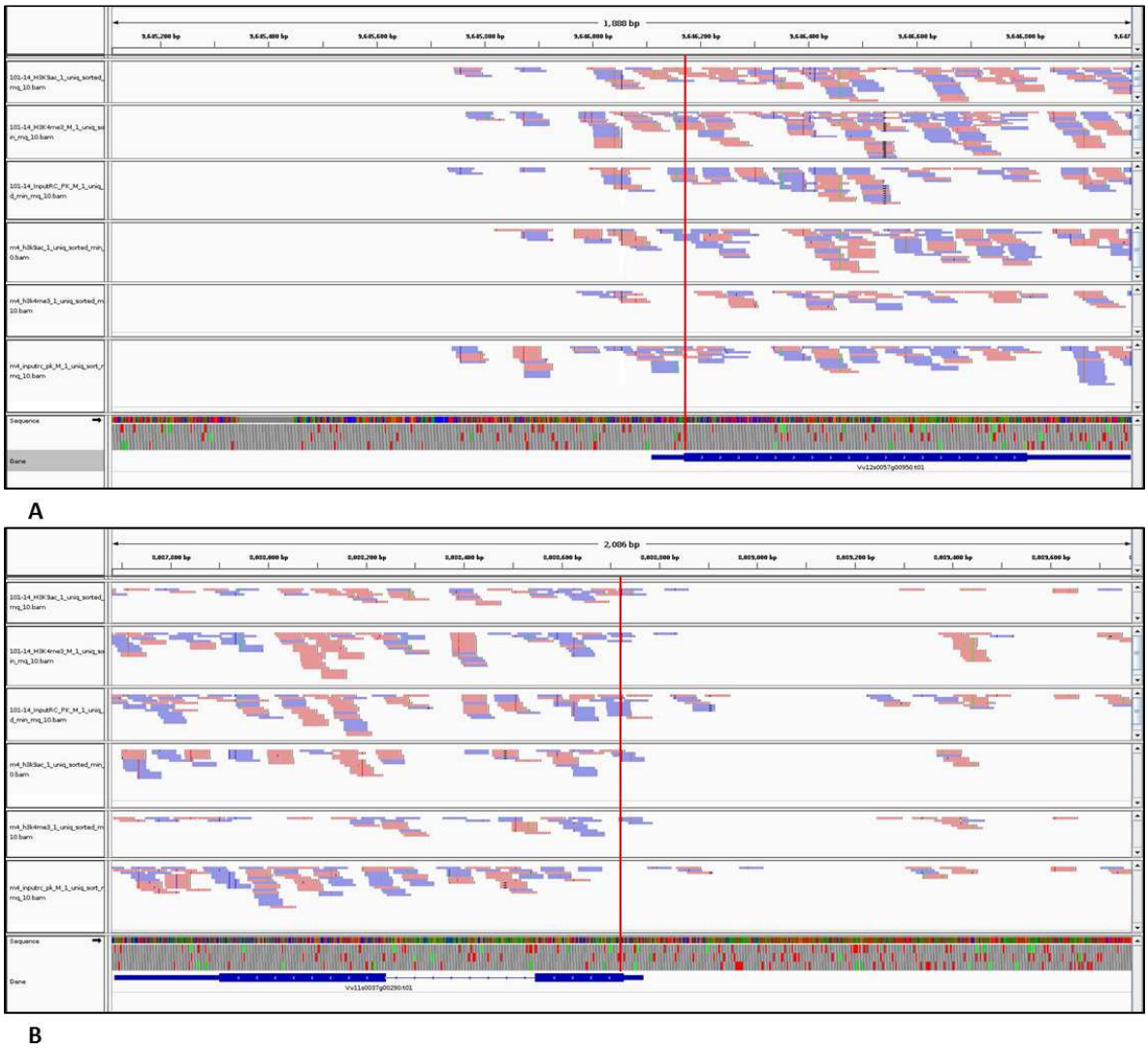


Figure 4.12: IGV snapshot for a DEG 101.14_3x M4<0.99 (A) (VIT_12s0057g00950) and for a DEG M4_3x 101.14<0.99 (B) (VIT_11s0037g00290). In the windows are reported the genes and the 1000 bp above start codons those are identified by the red line. In this windows we can see the reads alignments for 101.14 H3K9ac, H3K4me3, Input RC PK and for M4 H3K9ac, H3K4me3 and Input RC PK starting from the top downwards

3.4.3 Functional annotation of genes and Gene Ontology terms analyses

Functional annotations of genes were performed through PN40024 12x V2 annotation developed by CRIBI (University of Padova, http://genomes.cribi.unipd.it/gb2/gbrowse/public/vitis_vinifera_v2/) and by the functional annotation established by Grimplet (Grimplet *et al.*, 2012). Gene sequences were also aligned to *Arabidopsis thaliana* genome for the identification of orthologues genes using BLAST P, Basic Local Alignment Search Tools, NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=&LINK_LOC=blasttab&LAST_PAGE=blastn) (Table 4.12). Furthermore, Enriched DEGs were analyzed for Gene Ontology terms of Biological Process (BP) and Molecular Function (MF) in agreement with CRIBI GO annotation (http://genomes.cribi.unipd.it/DATA/V1/ANNOTATION/GO.tab) (Table 4.13)

DEGs 101.14_3x_M4≤0.99	Annotation V2	Functional annotation Grimplet	Blast P
VIT_18s0001g10850	Protein, PLA2A	Patatin	Phospholipase A 2A [At]
VIT_15s0021g00850	---	Unknown	Uncharacterized protein [At]
VIT_02s0025g02170	Protein, TRE1	Trehalase precursor, E2_4_1_15	Trehalase 1 [At]
VIT_02s0154g00300	14 kda proline-rich protein	Small nuclear ribonucleoprotein Sm D3, K04796	Protease inhibitor/seed storage/lipid transfer protein (LTP) family protein [At]
VIT_18s0001g12190	Cytochrome P450, CYP721A1	CYP721A1	Cytochrome P450, family 721, subfamily A, polypeptide 1 [At]
VIT_12s0057g00950	---	No hit	Hypothetical protein [At]
VIT_15s0046g01060	Pentatricopeptide repeat-containing protein mitochondrial-like	Pentatricopeptide (PPR) repeat	Pentatricopeptide repeat-containing protein [At]
VIT_02s0012g00830	Expansin-like b1,EXLB1	Expansin-like B1, EXPB1	No At match
VIT_00s0207g00130	Sts14 protein	Pathogenesis-related protein 1 precursor,PRP 1	CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein) superfamily protein [At] Sts14 [At]
VIT_14s0036g01360	Uncharacterized protein loc100240893	Disease resistance protein RPS2	No At match
VIT_14s0068g00140	---	No hit	No At match

A

DEGs M4_3x_101.14≤0.99	Annotation V2	Functional annotation Grimplet	Blast P
VIT_18s0001g11570	Cytochrome p450 82a3-like, CYP82C4	CYP82C1p	No At match
VIT_16s0013g00210	Metacaspase 1, AMC1	Metacaspase 1	Metacaspase 1 [At]
VIT_00s0163g00050	---	Argininosuccinate synthase	No At match
VIT_11s0037g00290	Protein, ENOD L15	Plastocyanin domain-containing protein	Early nodulin-like protein 15 [At]
VIT_07s0031g00260	cyclin delta-3	Cyclin delta-3	Uncharacterized protein [At]
VIT_00s1764g00020	Protein kinase like protein	Leucine-rich repeat transmembrane protein kinase	Protein kinase like protein [At]
VIT_16s0050g01430	Ornithine cyclodeaminase	Ornithine cyclodeaminase	NAD(P)-binding Rossmann-fold superfamily protein [At]
VIT_14s0083g00920	---	No hit	No At match
VIT_08s0058g00040	Leucine-rich repeat receptor protein kinase exs-like, GSO1	Verticillium wilt disease resistance protein Ve2	No At match
VIT_12s0035g00860	Ubiquitin carboxyl-terminal hydrolase-like protein	RPD1 (root primordium defective 1)	Ubiquitin carboxyl-terminal hydrolase family protein [At]

B

Table 4.12: Gene annotations and *Arabidopsis thaliana* orthologues for DEGs 101.14 3x_M4 ≤0.99 (A) and M4 3x_101.14 ≤0.99 (B).

DEGs 101.14_3x_M4≤0.99	GO terms		
VIT_18s0001g10850	GO:0006952	defense response	BP
	GO:0016042	lipid catabolic process	BP
	GO:0016787	hydrolase activity	MF
	GO:0045735	nutrient reservoir activity	MF
VIT_02s0025g02170	GO:0005991	trehalose metabolic process	BP
	GO:0004555	alpha,alpha-trehalase activity	MF
VIT_02s0154g00300	GO:0006869	lipid transport	BP
VIT_18s0001g12190	GO:0055114	oxidation-reduction process	BP
	GO:0004497	monooxygenase activity	MF
	GO:0009055	electron carrier activity	MF
	GO:0020037	heme binding	MF
VIT_15s0046g01060	GO:0005488	binding	MF
VIT_02s0012g00830	GO:0007047	cellular cell wall organization	BP
	GO:0019953	sexual reproduction	BP
VIT_00s0207g00130	GO:0006952	defense response	BP
	GO:0009607	response to biotic stimulus	BP
VIT_14s0036g01360	GO:0006915	apoptosis	BP
	GO:0006952	defense response	BP
	GO:0004721	phosphoprotein phosphatase activity	MF
	GO:0005515	protein binding	MF
	GO:0005524	ATP binding	MF
	GO:0017111	nucleoside-triphosphatase activity	MF

A

DEGs M4_3x_101.14≤0.99	GO terms		
VIT_18s0001g11570	GO:0055114	oxidation-reduction process	BP
	GO:0009055	electron carrier activity	MF
	GO:0020037	heme binding	MF
	GO:0033772	flavonoid 3',5'-hydroxylase activity	MF
VIT_16s0013g00210	GO:0006508	proteolysis	BP
	GO:0004197	cysteine-type endopeptidase activity	MF
VIT_11s0037g00290	GO:0005507	copper ion binding	MF
	GO:0009055	electron carrier activity	MF
VIT_16s0050g01430	GO:0055114	oxidation-reduction process	BP
	GO:0004764	shikimate 5-dehydrogenase activity	MF
	GO:0005488	binding	MF
VIT_08s0058g00040	GO:0005515	protein binding	MF
VIT_12s0035g00860	GO:0006511	ubiquitin-dependent protein catabolic process	BP
	GO:0006979	response to oxidative stress	BP
	GO:0008285	negative regulation of cell proliferation	BP
	GO:0010102	lateral root morphogenesis	BP
	GO:0004221	ubiquitin thiolesterase activity	MF

B

Table 4.13: Gene ontology biological process (BP) and molecular function (MF) terms of DEGs 101.14 3x_M4 ≤0.99 (A) and M4 3x_101.14 ≤0.99 (B).

Rootstock 101.14 showed a high number of enriched DEGs belonging to the BP Gene Ontology term of defense response (GO:0006952)˟. For genotype M4 the BP GO term most represented in enriched DEGs belonged to the class of oxidation-reduction process (GO: 0055114)˟.

˟ Blast2GO v2.8.0 definition of GO terms:

GO:0006952: reaction trigged in response to the presence of a foreign body or the occurrence of an injury which results in restriction of damage to the organism attacked or prevention/recovery from the infection caused by the attack.

GO0055114: metabolic process that results in removal or addiction of one or more electrons to or from a substance, with or without the concomitant removal or addiction of a proton or protons.

3.4.4 Comparisons between enriched *in vitro* DEGs and *in vivo* DEGs

The expression levels of enriched DEGs were also studied in plants grown in greenhouse (*in vivo*). The analysis showed that four of the DEGs enriched genes were also differentially expressed genes in *in vivo* plants. These data allowed us to hypothesize that the regulation of

these DEGs could be related to histone modifications also in *in vivo* condition. In Table 4.14 the expression level of this set of genes is reported.

		log_2(101.14/M4)			
Group of enrichment	**Gene ID**	**Expression *in vivo***	**Expression *in vitro***	**H3K9ac**	**H3K4me3**
101.14 3x_M4≤0.99	VIT_00s0207g00130	2.492	2.548	6.192	7.392
M4 3x_101.14≤0.99	VIT_18s0001g11570	-4.531	-2.313	-12.331	-18.551
	VIT_16s0013g00210	-2.396	-1.019	-19.496	-10.466
	VIT_08s0058g00040	1.563	1.366	-12.527	-2.408

Table 4.14: Expression levels comparison between *in vivo* DEGs and *in vitro* enriched DEGs.

The four genes differentially expressed in both grown conditions showed the same trend in transcriptional profiles, in particular VIT_00s0207g00130, VIT_18s0001g11570 and VIT_08s0058g00040 had expression levels in accord to histone modifications enrichment, while transcription profile of VIT_08s0058g00040 did not seem to have been influenced by histone modifications.

4 Discussion

In this study the histone modification profiles obtained from leaves of two genotypes of grapevine rootstocks grown in controlled conditions were compared and related to gene expression. This was the first work performed in grapevine aim at clarifying the relationship between histone modifications and gene expression.

In order to maintain grapevine rootstocks in controlled and stable conditions a pool of plants of each genotype were grown *in vitro* (25°C ±1°C, 16h/8h light/dark), leaf materials were collected and histone modifications distribution of H3K9ac and H3K4me3 were studied through a ChIP-Sequencing approach (HiSeq2000 Illumina, 1x50bp reads length). The reads were mapped on reference genome PN40024 that showed good reliability like reference for the genomes 101.14 and M4 (like discussed in details in Chapter II and by Corso (2014)). Reads were filtered and only the ones with map quality (mapq) ≥ 10 were considered reliable for our analyses. Each Input RC PK samples sequenced cover about 25 000 genes but our analyses were focused on the 18 796 genes selected during mRNA-Seq analyses for values of mean FPKM superior to one in at least one of the two rootstocks .

Literature data for other plant species demonstrate that the H3K9ac and H3K4me3 histone modifications preferentially lied in the transcription starting site (TSS) of the genes (Ha *et al*., 2013; Sequeira-Mendes *et al*., 2014). In grapevine reference genome PN40024, however, the annotations of TSSs and often also the ones of untranslated region (UTR) were not determined. For these reasons in an arbitrary manner, we decided, to keep a window of 1000bp above the start codon (ATG) and for each gene the number of reads that aligned in this region was considered in each of the sample sequenced.

The reads upstream of each gene were normalized and the genes were studied in terms of enrichment or depletion. To evaluate the enrichment in one genotype and the depletion, of the same gene in the other genotype, was used an approach with different cut off values of depletion: gene enriched 3x ($\frac{Ab}{Input}$ ≥3) of the histone modification of interest in one genotype

and depleted in the other of 0.99, 0.66 and 0.33 ($\frac{Ab}{Input} \leq \frac{2.9}{3}$ (0.99), $\frac{Ab}{Input} \leq \frac{2}{3}$ (0.66) and $\frac{Ab}{Input} \leq \frac{1}{3}$ (0.33)). The expression level of each gene enriched and depleted in one genotype and vice versa was evaluated using a statistical approach, namely the Welch's t-test, and a study of cumulative distribution function and frequency in relation to gene expression was performed. The most significative value of relation between histone modification and gene expression was obtained considering the genes enriched in one genotype and depleted in the other for both the histone modification marks at the same time. In particular, the groups of genes enriched 3x in one genotype and depleted with 0.99 in the other were studied in details. They consisted of 119 genes, 67 enriched in 101.14 and 52 enriched in M4. The fold change difference between these groups was statistically significative (p-value 0.03). Furthermore, the study of cumulated frequency in these set of genes showed cumulative distribution function that greatly differs between the two groups showing a propensity to negative values of fold changes ($\log_2 \frac{101.14}{M4}$) in genes enriched in M4 and a trend to positive values in genes enriched in 101.14. A clear example of this trend is provided by the analysis of the distribution of cumulated frequency at fold change 0: the 58% of the genes enriched in M4 showed a cumulated frequency value lower than this FC while only the 42% of the genes enriched in 101.14 shown the same distribution. These data represent evidences that the histone modifications here considered act in concert for the activation of gene expression.

Within the group 3x_≤0.99 H3K9ac ∩ H3K4me3 14 differentially expresses genes for 101.14 and 10 for M4 were respectively identified. The genotype specific gene and their upstream regions where aligned to the reference genome PN40024. Out of all the genes, three genes showed alignment problems and were excluded from our analyses. The study was restricted to 11 DEGs of 101.14 and 9 DEGs of M4, respectively. The reads of upstream region of the genes were visualized through a reads visualization software to confirm the coverage. Furthermore we analyzed the annotation of the genes, the ontologies with *Arabidopsis thaliana* and the Gene Ontology terms attribution. Finally were performed a comparison of expression levels between *in vivo* DEGs of plants grown in controlled conditions and *in vitro* enriched DEGs. Four genes

were identified like conserved DEGs with the same trends in both grown conditions and three of these shown expression level in agreement to histone modification enrichment evaluated in *in vitro* plants.

Till now ChIP-seq approaches, like those presented in this work, were performed in a few plant species, such as *Arabidopsis* (Sequeria-Mendes *et al*., 2014), maize (He *et al*., 2014, Wang *et al*., 2009), rice (Du *et al*., 2013; He *et al*., 2010), tomato (Ricardi *et al*., 2010 and 2014) and poplar (Li *et al*., 2014) but for these last two crops data about histone modifications are still missing because ChIP-seq was used for transcription factor characterization.

The distribution of *Arabidopsis* histone modifications at genome wide level and their correlation with gene expression (Sequeira-Mendes *et al*., 2014) were substantially confirmed in maize (Wang *et al*., 2009) and rice (Du *et al*., 2013). Du and coworkers performed a genome-wide analysis of four histone modifications (H3K4me2-3, H3K9ac and H3K27ac) in seedling of *Oryza sativa* and observed that active gene promoters contained the histone modifications analyzed in our study: H3K9ac and H3K4me3. Both these histone modifications showed a positive correlation with gene expression. Furthermore, they evaluated the concurrence of histone marks within transcribed regions discovering the concurrence frequencies of H3K4me3-H3K9ac and vice versa (Figure 4.13).

	H3K4me3	H3K9ac
H3K4me3		57.8%
H3K9ac	83.9%	

Figure 4.13: Concurrence frequencies for histone modifications H3K9ac and H3K4me3. The percentage number indicates the possibility that a histone modification peak on the *x*-axis exists in a histone peak on the *y*-axis (Adaptation from Du *et al*., 2013)

These analyses revealed high concurrence frequencies between H3K4me3 and H3K9ac in a high number of genes (about 45 000) expressed in the seedling. These high concurrence frequencies are likely to depend upon the high number of expressed genes that characterize the seedling and concurrence frequencies might not be as high in other tissues such as the mature leaf. Furthermore marks, such as histone acetylation and H3K4me3, showed tissues specific patterns (Berr *et al*., 2010; He *et al*., 2014) and this could be another reasons why the results obtained from seedlings are not comparable with those obtained in differentiated tissues.

Altogether the ChIP studies performed in plants showed similar conserved profile of histone modifications distribution and relation with transcriptome profile, particularly for H3K4me3 and H3K9ac. Considering the results of our work we can conclude that the positive correlation between these histone modification distribution and gene expression could be conserved in grapevine.

Data from literature of *Arabidosis*, maize and rice identify in the Transcription Starting Site (TSS) the genic region where H3K9ac and H3K4me3 histone marks are enriched (Sequeria-Mendes *et al*., 2014; Du *et al*., 2013) (Figure 4.14). In grapevine reference genome PN40024, however, the annotation of TSS is still missing and the selected window of 1000 bp above the ATG might not be the optimal range for the distribution analyses of these marks in relation to transcription.

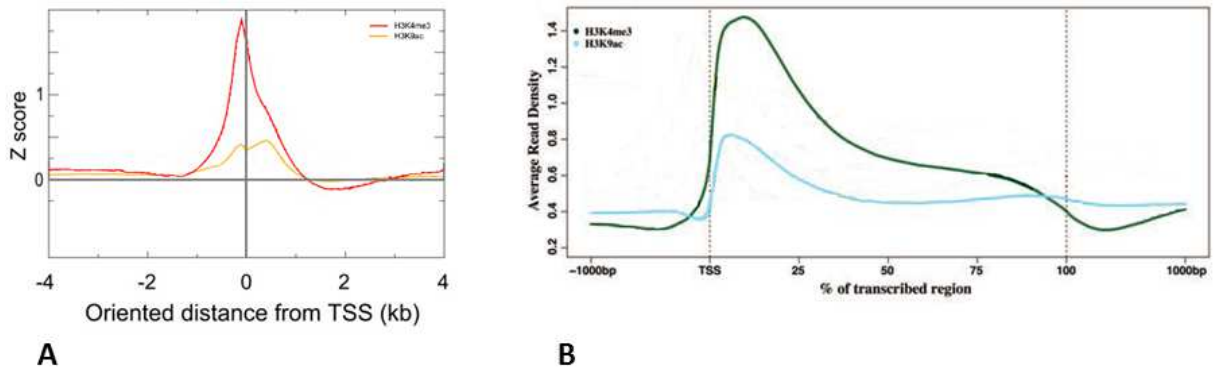


Figure 4.14: Estimation of the relative enrichment of the histone marks H3K4me3 and H3K9ac around the TSS in *Arabidopsis* (A) and in *Oryza sativa* (B) (A: Adaptation from Sequeria-Mendes *et al*., 2014; B: Adaptation from Du *et al*., 2013)

5 Conclusions and perspective

The results discussed in this chapter represent only a little part of the information that may arise from the merger of ChIP-seq and mRNA-seq data for the clarification of the correlation between histone modifications and gene expression.

Despite the earliness of the analyses performed until now the results of this study are the first evidences of a positive correlation between histone modifications H3K9ac and H3K4me3, when present simultaneously with H3K9ac, and gene expression in grapevine.

An improvement in this research project will be done with the annotation of the TSS in 101.14 and M4 grapevine genome. This work will be long and difficult but in this way the region of TSS could be studied in details for the identification of enrichment in histone modifications obtaining data comparable with other plant species.

6 References

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE(2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One 3(9):e3156

Chaves MM, Zarrouk O, Francisco R et al (2010) Grapevine under deficit irrigation: hints from physiological and molecular data. Ann Bot 105(5):661-676

Costas C, de la Paz Sanchez M, Stroud H et al (2011) Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. Nat Struct Mol Biol 18(3):395-400

Corso M (2014) A transcriptomic approach to dissect the effect of grapevine rootstocks on plant tolerance to abiotic stresses and berry ripening (PhD thesis, http://paduaresearch.cab.unipd.it/6393/)

Corso M, Vannozzi A, Maza E, Vitulo N, Meggio F, Pitacco A, Telatin A, D'Angelo M, Schiavon R, Negri AS, Prinsi B, Valle G, Ramina A, Bouzayen M, Bonghi C and Lucchin M (2015) Comprehensive transcripts profiling of two contrasting grapevine rootstock genotypes links phenylpropanoid pathways to enhanced drought tolerance. Submitted.

Cramer GR, Ergul A, Grimplet J et al (2007) Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. Funct Integr Genomics 7(2):111-134

Flexas J, Baron M, Bota J et al (2009) Photosynthesis limitations during water stress acclimation and recovery in the drought-adapted Vitis hybrid Richter-110 (V. berlandierixV. rupestris). J Exp Bot 60(8):2361-2377

Grimplet J, Van Hemert J, Carbonell-Bejerano P et al (2012) Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. BMC Res Notes 5:213-0500-5-213

Ha M, Ng DW, Li WH, Chen ZJ(2011) Coordinated histone modifications are associated with gene expression variation within and between species. Genome Res 21(4):590-598

Kurihara Y, Schmitz RJ, Nery JR et al (2012) Surveillance of 3' Noncoding Transcripts Requires FIERY1 and XRN3 in Arabidopsis. G3 (Bethesda) 2(4):487-498

Langmead B, Trapnell C, Pop M, Salzberg SL(2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25-2009-10-3-r25. Epub 2009 Mar 4

Lauria M, Rossi V(2011) Epigenetic control of gene regulation in plants. Biochim Biophys Acta 1809(8):369-378

Meggio F, Prinsi B, Negri AS, Simone Di Lorenzo G, Lucchini G, Pitacco A, Failla O, Scienza A, Cocucci M and Espen L Biochemical and physiological responses of two grapevine rootstock genotypes to drought and salt treatments. Australian Journal of Grape and Wine Research 20:310-323

Sequeira-Mendes J, Araguez I, Peiro R et al (2014) The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell 26(6):2351-2366

Stroud H, Otero S, Desvoyes B, Ramirez-Parra E, Jacobsen SE, Gutierrez C(2012) Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc Natl Acad Sci U S A 109(14):5370-5375

Vezzi F, Del Fabbro C, Tomescu AI, Policriti A(2012) rNA: a fast and accurate short reads numerical aligner. Bioinformatics 28(1):123-124

Welch BL(1947) The generalization of "Student's" problem when several different population variances are involved. Biometrika 34 (1–2):28–35

Zilberman D(2008) The evolving functions of DNA methylation. Curr Opin Plant Biol 11(5):554-559

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S(2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456(7218):125-129

7 Supplementary materials

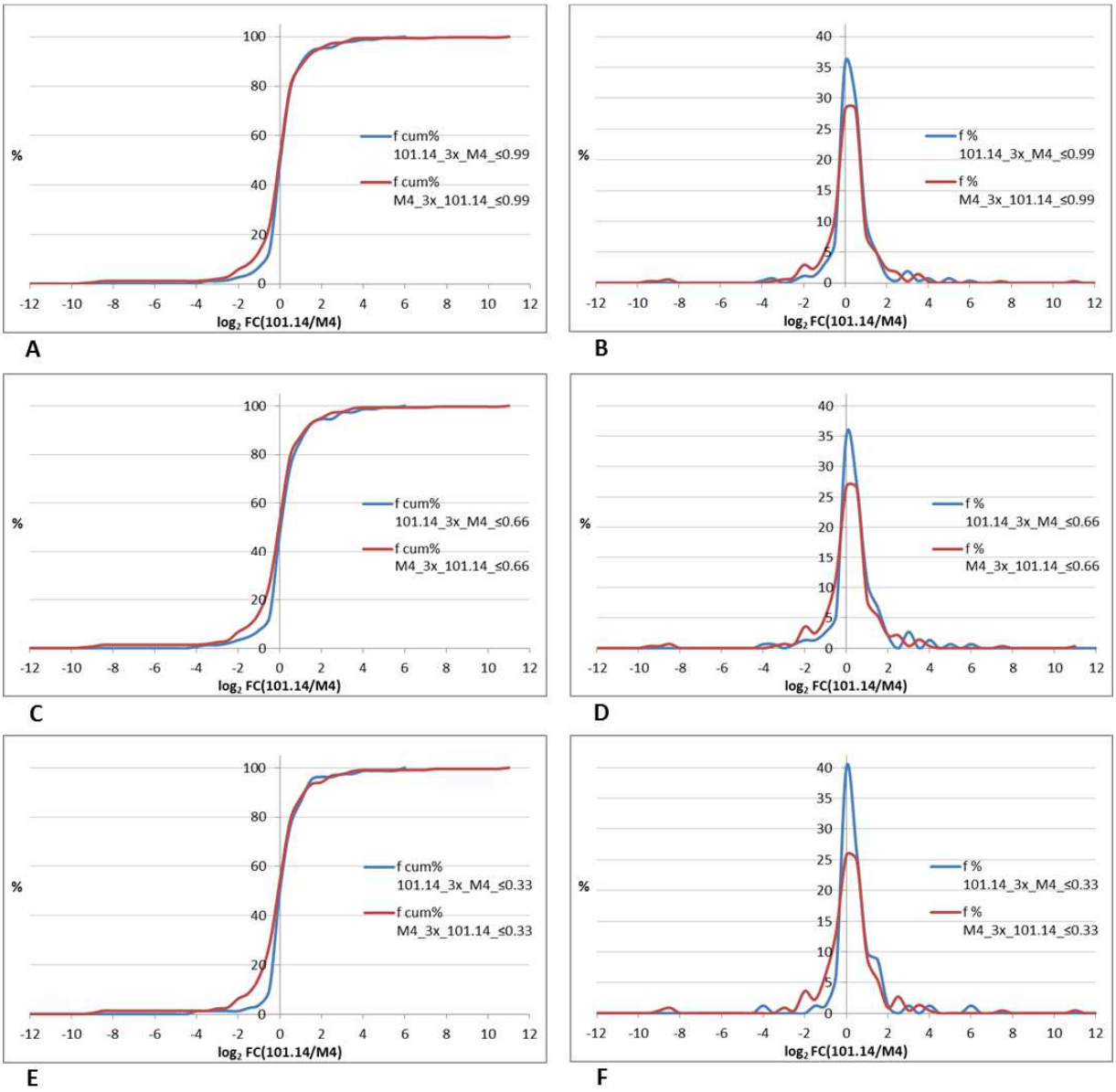


Figure 4.15: Cumulative distribution (A, C, E) and frequency function (B, D, F) of fold change plots of the three subgroups of genes enriched and depleted for H3K9ac. A, B) Distribution of genes enriched 3x and depleted 0.99; C, D) Distribution of genes enriched 3x and depleted 0.66; E, F) Distribution of genes enriched 3x and depleted 0.33

Class	f cum% 101.14_3x_ M4_ ≤0.99	f cum% M4_3x_ 101.14_≤0.99	f % 101.14_3x_ M4_≤0.99	f % M4_3x_ 101.14_≤0.99	f cum% 101.14_3x_ M4_≤0.66	f cum% M4_3x_ 101.14_≤0.66	f % 101.14_3x_ M4_≤0.66	f % M4_3x_ 101.14_≤0.66	f cum% 101.14_3x_ M4_≤0.33	f cum% M4_3x_ 101.14_≤0.33	f % 101.14_3x_ M4_≤0.33	f % M4_3x_ 101.14_≤0.33
-12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-11.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-10.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-9.5	0.0	0.3	0.0	0.3	0.0	0.4	0.0	0.4	0.0	0.0	0.0	0.0
-9.0	0.0	0.6	0.0	0.3	0.0	0.7	0.0	0.4	0.0	0.5	0.0	0.5
-8.5	0.0	1.2	0.0	0.6	0.0	1.4	0.0	0.7	0.0	1.4	0.0	0.9
-8.0	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-7.5	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-7.0	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-6.5	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-6.0	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-5.5	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-5.0	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-4.5	0.0	1.2	0.0	0.0	0.0	1.4	0.0	0.0	0.0	1.4	0.0	0.0
-4.0	0.4	1.2	0.4	0.0	0.7	1.4	0.7	0.0	1.3	1.4	1.3	0.0
-3.5	1.2	1.5	0.8	0.3	1.4	1.8	0.7	0.4	1.3	1.4	0.0	0.0
-3.0	1.2	2.1	0.0	0.6	1.4	2.5	0.0	0.7	1.3	2.3	0.0	0.9
-2.5	1.6	2.9	0.4	0.9	2.0	3.2	0.7	0.7	1.3	2.7	0.0	0.5
-2.0	2.7	5.9	1.2	2.9	3.4	6.8	1.4	3.6	1.3	6.3	0.0	3.6
-1.5	3.9	8.2	1.2	2.4	4.7	9.3	1.4	2.5	2.5	8.6	1.3	2.3
-1.0	7.0	13.2	3.1	5.0	7.4	14.6	2.7	5.4	3.8	14.9	1.3	6.3
-0.5	13.6	23.8	6.6	10.6	12.8	26.4	5.4	11.8	10.0	28.1	6.3	13.1
0.0	**49.2**	**52.1**	**35.7**	**28.2**	**48.0**	**53.2**	**35.1**	**26.8**	**50.0**	**53.8**	**40.0**	**25.8**
0.5	79.1	80.3	29.8	28.2	75.0	79.6	27.0	26.4	76.3	78.7	26.3	24.9
1.0	89.1	88.2	10.1	7.9	85.8	87.5	10.8	7.9	86.3	87.8	10.0	9.0
1.5	94.2	93.2	5.0	5.0	92.6	92.9	6.8	5.4	95.0	93.2	8.8	5.4
2.0	95.3	95.6	1.2	2.4	94.6	95.0	2.0	2.1	96.3	94.1	1.3	0.9
2.5	95.7	97.4	0.4	1.8	94.6	97.1	0.0	2.1	96.3	96.8	0.0	2.7
3.0	97.7	97.6	1.9	0.3	97.3	97.5	2.7	0.4	97.5	97.3	1.3	0.5
3.5	98.1	99.1	0.4	1.5	97.3	98.9	0.0	1.4	97.5	98.6	0.0	1.4
4.0	98.8	99.4	0.8	0.3	98.6	99.3	1.4	0.4	98.8	99.1	1.3	0.5
4.5	98.8	99.4	0.0	0.0	98.6	99.3	0.0	0.0	98.8	99.1	0.0	0.0
5.0	99.6	99.4	0.8	0.0	99.3	99.3	0.7	0.0	98.8	99.1	0.0	0.0
5.5	99.6	99.4	0.0	0.0	99.3	99.3	0.0	0.0	98.8	99.1	0.0	0.0
6.0	100.0	99.4	0.4	0.0	100.0	99.3	0.7	0.0	100.0	99.1	1.3	0.0
6.5		99.4	0.0	0.0		99.3	0.0	0.0		99.1	0.0	0.0
7.0		99.4	0.0	0.0		99.3	0.0	0.0		99.1	0.0	0.0
7.5		99.7	0.0	0.3		99.6	0.0	0.4		99.5	0.0	0.5
8.0		99.7	0.0	0.0		99.6	0.0	0.0		99.5	0.0	0.0
8.5		99.7	0.0	0.0		99.6	0.0	0.0		99.5	0.0	0.0
9.0		99.7	0.0	0.0		99.6	0.0	0.0		99.5	0.0	0.0
9.5		99.7	0.0	0.0		99.6	0.0	0.0		99.5	0.0	0.0
10.0		99.7	0.0	0.0		99.6	0.0	0.0		99.5	0.0	0.0
10.5		99.7	0.0	0.0		99.6	0.0	0.0		99.5	0.0	0.0
11.0		100.0	0.0	0.3		100.0	0.0	0.4		100.0	0.0	0.5
11.5			0.0	0.0			0.0	0.0			0.0	0.0
12.0			0.0	0.0			0.0	0.0			0.0	0.0

Table 4.15: Cumulative distribution and frequency function of Fold Change values in the three subgroups of genes enriched and depleted for H3K9ac

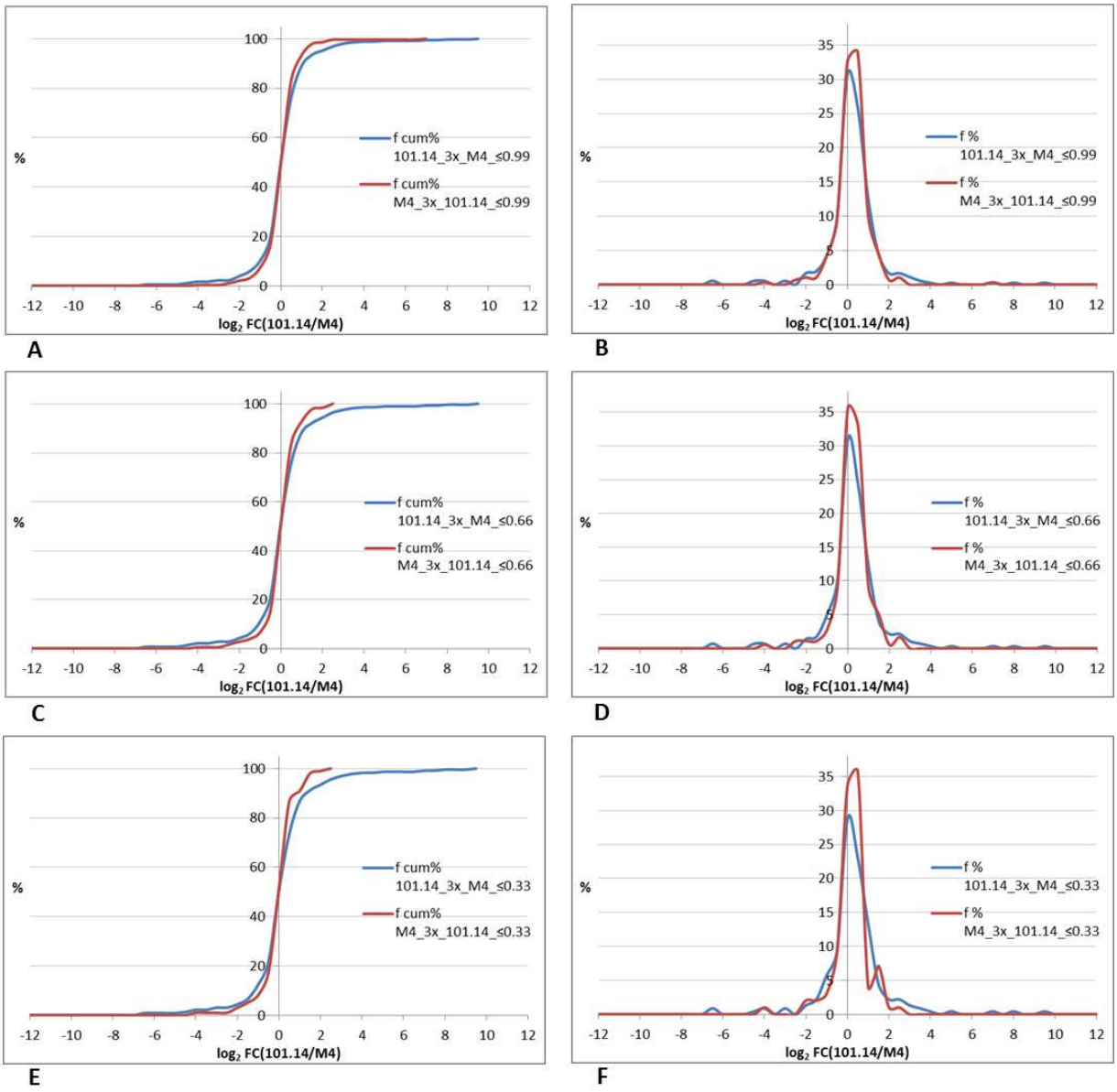


Figure 4.16: Cumulative distribution (A, C, E) and frequency function (B, D, F) of fold change values in the three subgroups of genes enriched and depleted for H3K4me3

Class	f cum% 101.14_3x_ M4_ ≤0.99	f cum% M4_3x_ 101.14_ ≤0.99	f % 101.14_3x_ M4_ ≤0.99	f % M4_3x_ 101.14_≤0.99	f cum% 101.14_3x_ M4_≤0.66	f cum% M4_3x_ 101.14_≤0.66	f % 101.14_3x_ M4_≤0.66	f % M4_3x_ 101.14_≤0.66	f cum% 101.14_3x_ M4_≤0.33	f cum% M4_3x_ 101.14_≤0.33	f % 101.14_3x_ M4_≤0.33	f % M4_3x_ 101.14_≤0.33
-12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-11.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-10.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-9.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-9.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-8.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-8.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-7.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-7.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-6.5	0.6	0.0	0.6	0.0	0.7	0.0	0.7	0.0	0.9	0.0	0.9	0.0
-6.0	0.6	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.9	0.0	0.0	0.0
-5.5	0.6	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.9	0.0	0.0	0.0
-5.0	0.6	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.9	0.0	0.0	0.0
-4.5	1.1	0.0	0.6	0.0	1.4	0.0	0.7	0.0	1.3	0.0	0.4	0.0
-4.0	1.7	0.3	0.6	0.3	2.1	0.6	0.7	0.6	2.2	1.0	0.9	1.0
-3.5	1.7	0.3	0.0	0.0	2.1	0.6	0.0	0.0	2.2	1.0	0.0	0.0
-3.0	2.3	0.3	0.6	0.0	2.8	0.6	0.7	0.0	3.1	1.0	0.9	0.0
-2.5	2.3	1.0	0.0	0.7	2.8	1.7	0.0	1.1	3.1	1.0	0.0	0.0
-2.0	3.9	2.1	1.7	1.0	4.3	2.8	1.4	1.1	4.4	3.1	1.3	2.0
-1.5	5.9	3.1	2.0	1.0	6.0	3.9	1.8	1.1	6.6	5.1	2.2	2.0
-1.0	10.1	7.3	4.2	4.2	11.0	6.7	5.0	2.8	12.4	8.2	5.8	3.1
-0.5	19.7	16.8	9.6	9.4	21.0	15.6	10.0	8.9	21.7	17.3	9.3	9.2
0.0	**50.4**	**49.3**	**30.7**	**32.5**	**52.0**	**51.1**	**31.0**	**35.6**	**50.4**	**51.0**	**28.8**	**33.7**
0.5	76.3	83.2	25.9	33.9	76.2	83.9	24.2	32.8	73.5	86.7	23.0	35.7
1.0	89.0	93.4	12.7	10.1	88.3	92.8	12.1	8.9	86.7	90.8	13.3	4.1
1.5	93.5	97.9	4.5	4.5	92.2	97.8	3.9	5.0	91.2	98.0	4.4	7.1
2.0	95.2	98.6	1.7	0.7	94.3	98.3	2.1	0.6	93.4	99.0	2.2	1.0
2.5	96.9	99.7	1.7	1.0	96.4	100.0	2.1	1.7	95.6	100.0	2.2	1.0
3.0	98.0	99.7	1.1	0.0	97.5		1.1	0.0	96.9		1.3	0.0
3.5	98.6	99.7	0.6	0.0	98.2		0.7	0.0	97.8		0.9	0.0
4.0	98.9	99.7	0.3	0.0	98.6		0.4	0.0	98.2		0.4	0.0
4.5	98.9	99.7	0.0	0.0	98.6		0.0	0.0	98.2		0.0	0.0
5.0	99.2	99.7	0.3	0.0	98.9		0.4	0.0	98.7		0.4	0.0
5.5	99.2	99.7	0.0	0.0	98.9		0.0	0.0	98.7		0.0	0.0
6.0	99.2	99.7	0.0	0.0	98.9		0.0	0.0	98.7		0.0	0.0
6.5	99.2	99.7	0.0	0.0	98.9		0.0	0.0	98.7		0.0	0.0
7.0	99.4	100.0	0.3	0.3	99.3		0.4	0.0	99.1		0.4	0.0
7.5	99.4		0.0	0.0	99.3		0.0	0.0	99.1		0.0	0.0
8.0	99.7		0.3	0.0	99.6		0.4	0.0	99.6		0.4	0.0
8.5	99.7		0.0	0.0	99.6		0.0	0.0	99.6		0.0	0.0
9.0	99.7		0.0	0.0	99.6		0.0	0.0	99.6		0.0	0.0
9.5	100.0		0.3	0.0	100.0		0.4	0.0	100.0		0.4	0.0
10.0			0.0	0.0			0.0	0.0			0.0	0.0
10.5			0.0	0.0			0.0	0.0			0.0	0.0
11.0			0.0	0.0			0.0	0.0			0.0	0.0
11.5			0.0	0.0			0.0	0.0			0.0	0.0
12.0			0.0	0.0			0.0	0.0			0.0	0.0

Table 4.16: Cumulative distribution and frequency function of Fold Change values in the three subgroups of genes enriched and depleted for H3K4me3

	log2(101.14/M4)		
H3K9ac ∩H3K4me3 101.14_3x_M4≤0.99	Expression *in vitro*	H3K9ac	H3K4me3
VIT_00s0207g00100	0.070	2.784	2.345
VIT_00s0207g00130	2.548	6.192	7.392
VIT_00s0978g00010	0.452	1.907	5.120
VIT_00s1339g00010	-0.211	11.275	9.049
VIT_00s1569g00020	0.076	5.154	7.479
VIT_00s1847g00010	-0.055	13.323	11.953
VIT_00s2077g00020	-0.350	11.183	4.362
VIT_00s2769g00010	0.510	15.233	13.882
VIT_01s0010g00880	0.407	9.418	11.535
VIT_01s0150g00630	0.419	10.660	12.317
VIT_02s0012g00030	0.231	19.956	18.703
VIT_02s0012g00830	2.311	3.147	4.026
VIT_02s0025g00200	0.359	2.507	1.975
VIT_02s0025g02170	-2.100	5.429	14.839
VIT_02s0033g00050	0.297	10.585	11.350
VIT_02s0154g00300	-1.852	8.336	10.626
VIT_03s0038g01090	0.608	2.667	1.971
VIT_03s0091g00220	-0.313	4.991	2.663
VIT_03s0097g00210	0.509	13.563	12.746
VIT_03s0180g00180	0.273	13.735	6.395
VIT_04s0008g01030	-0.275	3.738	3.816
VIT_04s0023g02460	0.131	2.564	2.491
VIT_05s0020g01180	-0.327	4.981	4.976
VIT_05s0020g02680	-0.323	10.808	11.203
VIT_05s0020g04600	0.080	3.479	2.402
VIT_05s0051g00410	0.696	6.967	13.473
VIT_05s0051g00850	0.008	9.725	8.587
VIT_05s0077g01940	0.298	1.975	2.436
VIT_05s0124g00340	-0.178	1.931	3.267
VIT_06s0004g05770	-0.964	10.930	10.194
VIT_07s0031g00740	0.818	18.672	17.001
VIT_08s0040g00160	-0.393	13.412	20.880
VIT_09s0002g04190	-0.409	2.339	3.609
VIT_09s0054g01590	0.526	2.095	12.206
VIT_10s0003g03260	-0.067	2.558	3.287
VIT_10s0003g04870	2.944	3.456	2.509
VIT_10s0116g00630	0.814	7.986	3.667
VIT_11s0016g02440	0.895	9.781	19.073
VIT_11s0052g00130	0.131	12.476	19.626
VIT_11s0118g00620	-0.211	3.013	1.934
VIT_12s0057g00950	1.308	2.768	4.313
VIT_12s0059g01000	0.215	12.665	11.952
VIT_13s0064g00270	0.202	6.137	4.880
VIT_13s0064g01770	-0.419	2.951	3.817
VIT_14s0030g00140	2.935	19.070	19.967
VIT_14s0036g01360	3.589	14.965	5.924
VIT_14s0060g00970	0.130	19.853	19.666
VIT_14s0060g02110	-0.349	11.607	20.505
VIT_14s0066g01410	0.122	2.348	2.242
VIT_14s0068g00120	-0.462	17.246	19.159
VIT_14s0068g00140	4.825	11.983	12.293
VIT_14s0068g00410	-0.300	11.950	20.992
VIT_15s0021g00850	-2.120	2.825	11.664
VIT_15s0046g01060	1.355	12.672	2.553
VIT_16s0039g00300	-0.244	5.267	4.340
VIT_16s0050g01260	0.505	12.053	22.731
VIT_17s0000g02730	-0.305	22.029	12.725
VIT_17s0000g07030	-1.404	18.196	16.825
VIT_18s0001g05630	0.220	5.305	14.035
VIT_18s0001g07470	-0.075	1.701	2.768
VIT_18s0001g07980	-0.406	12.306	8.663
VIT_18s0001g10850	-4.088	20.284	17.687
VIT_18s0001g12190	1.189	13.598	14.163
VIT_18s0086g00690	0.376	11.406	11.647
VIT_19s0014g00180	-0.360	11.948	11.425
VIT_19s0014g04440	0.490	17.324	17.859
VIT_19s0090g01760	-0.473	14.260	14.855

A

	log2(101.14/M4)		
H3K9ac ∩H3K4me3 M4_3x_101.14≤0.99	Expression *in vitro*	H3K9ac	H3K4me3
VIT_00s0163g00050	-1.729	-3.237	-13.745
VIT_00s0246g00190	0.207	-6.616	-7.855
VIT_00s0265g00130	-0.837	-2.901	-4.709
VIT_00s0361g00100	0.122	-2.356	-2.995
VIT_00s0769g00010	-0.246	-16.475	-17.648
VIT_00s1764g00020	-1.028	-17.935	-9.974
VIT_00s2608g00010	0.237	-18.784	-19.683
VIT_01s0011g02530	0.139	-5.021	-4.057
VIT_01s0011g03160	-0.103	-4.465	-3.307
VIT_01s0011g03890	0.615	-15.543	-6.062
VIT_01s0137g00010	-0.273	-8.345	-5.221
VIT_01s0137g00690	0.016	-5.297	-3.379
VIT_02s0033g01000	0.154	-9.591	-11.937
VIT_02s0087g00140	0.210	-3.465	-2.706
VIT_03s0180g00100	0.036	-7.147	-3.811
VIT_04s0008g04550	-0.067	-12.425	-1.975
VIT_05s0062g01270	-0.307	-12.001	-4.777
VIT_05s0124g00030	-0.152	-11.271	-12.536
VIT_06s0004g00390	-0.095	-3.270	-2.810
VIT_06s0009g01120	-0.214	-2.871	-2.933
VIT_06s0009g03610	-0.200	-10.282	-16.276
VIT_06s0061g00270	-0.191	-14.648	-5.629
VIT_07s0031g00260	-1.076	-17.140	-9.911
VIT_07s0031g00650	-0.155	-2.856	-5.632
VIT_07s0141g00240	-0.221	-14.506	-13.238
VIT_08s0007g01740	-0.032	-21.025	-12.884
VIT_08s0007g08420	0.392	-2.860	-2.942
VIT_08s0058g00040	1.366	-12.527	-2.408
VIT_09s0002g02360	0.935	-4.306	-2.296
VIT_09s0002g06530	-0.182	-9.943	-11.314
VIT_09s0018g01180	-0.412	-5.326	-11.167
VIT_09s0054g00010	0.039	-5.342	-3.538
VIT_11s0037g00290	-1.341	-11.796	-2.248
VIT_12s0028g02100	-0.090	-3.613	-3.964
VIT_12s0035g00740	0.462	-12.500	-19.740
VIT_12s0035g00860	1.755	-3.193	-13.032
VIT_13s0019g03910	0.147	-3.639	-4.127
VIT_13s0175g00180	0.108	-18.512	-20.036
VIT_14s0083g00920	1.283	-19.256	-18.126
VIT_16s0013g00210	-2.022	-4.587	-5.097
VIT_16s0050g01430	-1.019	-19.496	-10.466
VIT_16s0098g01780	0.380	-4.046	-2.793
VIT_17s0000g07090	-0.621	-10.485	-3.698
VIT_18s0001g02640	0.135	-3.124	-3.466
VIT_18s0001g03730	0.253	-11.818	-3.766
VIT_18s0001g11250	-0.382	-10.016	-10.962
VIT_18s0001g11570	-2.313	-12.331	-18.551
VIT_18s0001g12780	-0.412	-21.439	-12.469
VIT_18s0041g00230	0.956	-16.536	-12.127
VIT_18s0041g01640	-0.495	-9.318	-9.911
VIT_18s0117g00150	-0.080	-20.363	-9.772
VIT_19s0177g00120	-0.485	-18.500	-18.491

B

Table 4.17: Fold change and histone modifications values for gene3x_≤0.99 enriched in 101.14 (A) and in M4 (B)

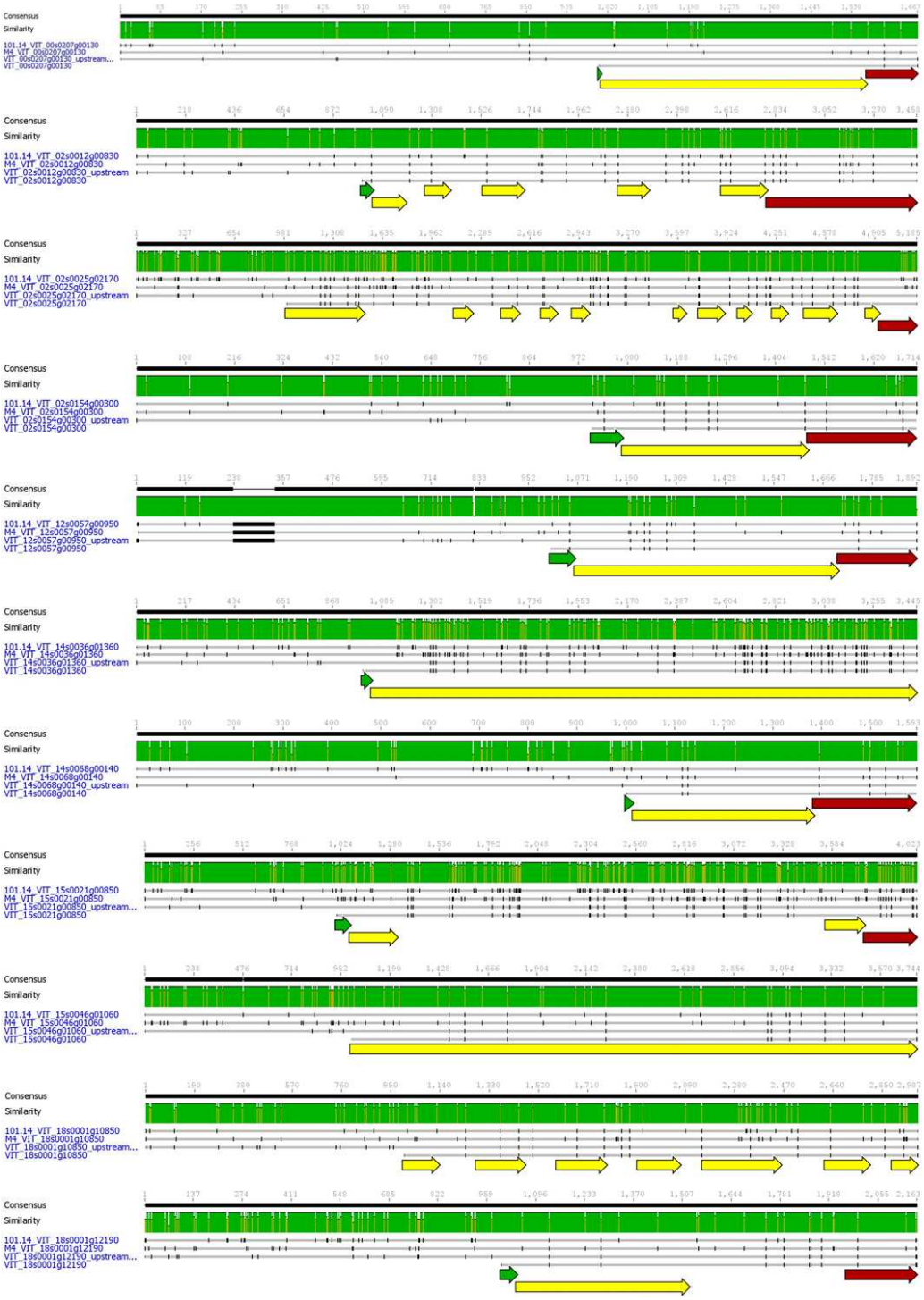


Figure 4.17: Alignment snapshots of 101.14 3x_ M4≤0.99 DEGs. Genes and their upstream regions were reported for 101.14, M4 and reference genome PN40024. The green arrow represent 5'UTRs, the yellow ones the CDSs and the red ones 3'UTRs

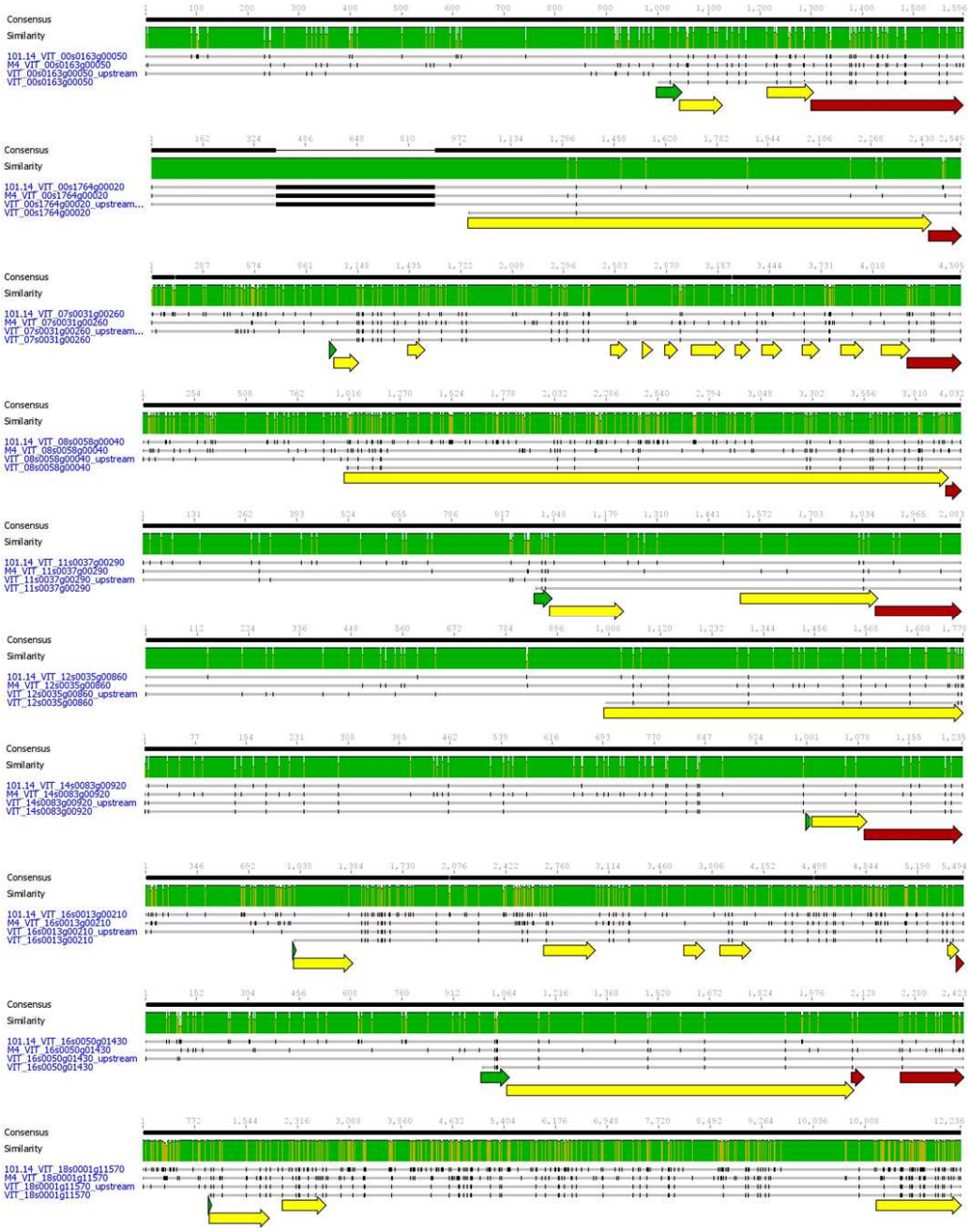


Figure 4.18: Alignment snapshots of M4 3x_ 101.14≤0.99 DEGs. Genes and their upstream regions were reported for 101.14, M4 and reference genome PN40024. The green arrow represent 5'UTRs, the yellow ones the CDSs and the red ones 3'UTRs

Ringraziamenti

Un doveroso ringraziamento va alla professoressa Margherita Lucchin per avermi dato la possibilità di intraprendere questo dottorato di ricerca e per il supporto datomi durante questi tre anni. Un particolare ringraziamento va alla professoressa Serena Varotto e al professor Claudio Bonghi per avermi aiutato con continui spunti intellettuali nonché per l'aiuto datomi nell'elaborazione ed interpretazione dei dati. Vorrei ringraziare inoltre Vincenzo Rossi per i preziosi consigli sulla ChIP e Slobodanka Radovic per i sequenziamenti.

I miei ringraziamenti vanno inoltre a tutti i colleghi ed amici che in questi tre anni mi hanno supportato e incoraggiato nell'attività di ricerca. Un particolare ringraziamento va a Silvia, Cristian ed Alessandro.