Corso di dottorato di ricerca in:

"Scienze e Biotecnologie Agrarie"

in convenzione con Fondazione E. Mach

Ciclo XXXIII

Titolo della tesi

"Investigating the molecular targets of Apple Proliferation Phytoplasma effector SAP11$_{CaPM}$ in apple"

in co-tutela con Fondazione Edmund Mach

Dottorando
Mattia Tabarelli

Supervisore
Dr. Mickael Malnoy

Co-supervisore
Dr. Katrin Janik

Anno 2021

TABLE OF CONTENTS

ABSTRACT

Apple Proliferation (AP) is a phytoplasma-related associated with *'Candidatus* Phytoplasma mali' ('*Ca*. P. mali'). Trees affected by the disease display several symptoms associated with plant shape and development, ultimately leading to the generation of non-marketable fruits. The disease is widespread in many apple cultivated areas of Europe, where outbreaks in recent years caused huge economic losses. There is no cure or treatment for phytoplasmoses but to control the vector insects, the uprooting of affected trees and the use of healthy propagation material. In the last years, extensive research led to the identification of AP-tolerant experimental rootstock, but the molecular basis of this inheritable trait is unknown. Phytoplasmas secrete pathogenic proteins, called effectors, that affect the host physiology and determine a fitness advantage for the bacteria. Among these, the best characterized in '*Ca*. P. mali' is $SAP11_{CaPM}$, which was observed to target and deactivate two *M.* × *domestica* transcription factors of the TCP family, i.e., *MdTCP4a* and *MdTCP13a*. This work aims to investigate the molecular targets of $SAP11_{CaPM}$ to shed light on the molecular basis of disease development and the phenomenon of tolerance.

The second chapter consists of a revision of the *MdTCP* gene family set by performing a re-identification on the latest high-quality *M.* × *domestica* genome assembly. Of the 52 *MdTCP* sequences previously described, 15 were discarded because of redundancy with other *MdTCP*s (six) or the absence of a complete open reading frame (nine). Of the 37 remaining *MdTCP* sequences, 30 were observed to be included in the list of functionally annotated genes automatically computed. Furthermore, three novel *MdTCP* sequences have been identified, for a total of 40 sequences constituting the revised set of the *MdTCP* gene family. Analysis of the intergenic identity, combined with the synteny analysis previously performed on the whole *M.* × *domestica* genome, allowed the identification of 15 pairs of genes that likely originated through a recent whole-genome duplication event. Finally, the 40 *MdTCP* sequences identified were named according to the homology with *A. thaliana TCP*.

In chapter 3, the sequences of the two $SAP11_{CaPM}$-target genes have been analyzed in susceptible and tolerant genotypes to screen for potential differences that could explain

the tolerant behaviour displayed by some plants. Sequence analyses led to the identification of two rare amino acid substitutions in MdTCP4a, of which one, in particular, was not found in any of the TCP sequences published to date. The interaction between the bacterial effector and the protein alternatives from susceptible and tolerant genotypes was measured both in a heterologous system (Y2H) and *in planta* (BiFC). The interaction assays showed no difference in the interaction strength between the different alternatives, suggesting that this mechanism cannot explain the tolerant phenotype. Nonetheless, a good degree of correlation between the presence of the two amino acid substitutions and tolerance to AP was found, indicating a possible role of these sequences as genetic markers of tolerance.

Finally, chapter 4 describes the generation and characterization of *M.* × *domestica* transgenic plants overexpressing the two *MdTCP* genes via At-mediated transformation. While plants overexpressing *MdTCP13a* died shortly after the regeneration, suggesting that overexpression of this gene can hinder the plant development, several independent *MdTCP4a*-overexpressing transgenic lines could be generated. Transgenic lines kept *in vitro* showed a low level of overexpression and no differences in phenotype compared to non-transformed plants. Following the soil acclimatization, the transgenic plants demonstrated an increase in the overexpression level and displayed some peculiar phenotype characteristics, such as smaller and crinkled leaves, loss of apical dominance, and generation of more shoots than non-transformed. *In vitro* transgenic plants were infected with '*Ca*. P. mali' using the micrografting technique and were screened for the presence of phytoplasma with a qPCR assay. Results show that two out of three transgenic lines tested display a phytoplasma concentration significantly lower than non-transformed, which is commonly associated with less severity or complete absence of symptoms. Furthermore, the transcriptomes of three soil acclimatized lines have been analyzed via RNAseq, and preliminary analyses show an up-regulation of several genes associated with plant development and abiotic stress responses.

CHAPTER 1
GENERAL INTRODUCTION

THE DOMESTICATED APPLE (*Malus* × *domestica*)

Economic importance

The domesticated apple (*M.* × *domestica*) is a member of the Rosaceae family, which includes more than 100 genera and 3000 species distributed worldwide. *M*. x *domestica* is among the most cultivated fruit crops, primarily in the temperate regions of both hemispheres (Velasco *et al*., 2010). The latest data indicate apple as the third most valuable fruit plant and the 16th most valuable cultivated crop globally (FAOSTAT, 2019). The annual apple production consists of 87 million tonnes (2019), with an increase of nearly 50% since the 59 million tonnes produced in 2000. Globally, apple harvest area occupies nearly 5 million hectares and saw a decrease of more than 5% in respect to 2015. However, both indicators appear to follow a relatively stable trend over the last few years (FAOSTAT, 2020).

China is the largest apple producer globally, constituting approximately half of the global production alone, followed by United States, Turkey, Poland, and Italy (2015-2019 average; FAOSTAT, 2020). Europe accounts for approximately 20% of the global production, of which 15% come from the Trentino-South Tyrol region, Europe's largest interconnected apple growing area.

Ecology and biology

M. × *domestica* originated 4000-10000 years ago in Kazakhstan from the wild species *Malus sieversii*. Genomic analyses revealed that additional wild species constitute *M*. × *domestica* major genomic contributors, mainly from Europe (*Malus sylvestris*) and Caucasus (*Malus orientalis*), through introgression and hybridization events that likely took place during the dispersion towards West Europe (Cornille *et al*., 2012; Cornille *et al*., 2014; Duan *et al*., 2017; Sun *et al*., 2020). Introgression events have been so extensive that the modern apple appears more genetically similar to *M. sylvestris* than the original progenitor *M. sieversii* based on microsatellite analysis (Cornille *et al*., 2012).

Apples are self-incompatible trees and rely on vegetative propagation by grafting. Moreover, apple cultivars are grafted onto dwarfed apple trees used as rootstocks. Rootstocks provide essential characteristics to the scions, such as determining the tree's height, degree of resistance towards biotic (cold, drought), or even abiotic stresses (fire blight, apple replant disease). Throughout its centuries-long history of cultivation, several thousand cultivars of the domesticated apple have been developed and documented (Dell, Morgan, and Richards 2003; Qian *et al.*, 2010). Despite the diversity of available cultivars, apple production worldwide is now primarily based on the cultivation of a few dozen cultivars and a series of clonal rootstocks with high levels of chemical inputs (Cornille *et al*., 2014).

The genome of *M. × domestica* was published in 2010 by Velasco *et al*., based on the diploid cultivar (cv.) 'Golden delicious'. The apple genome is constituted of 17 chromosomes, an estimated total size of roughly 750 Mb, and approximately 57.000 genes. The genome shows hints of a recent (40-60 mya) genome-wide duplication event and remnants of ancient duplications. According to chromosome homologies, four pairs of chromosomes are principally derived from one ancestor each, with minor interchromosomal rearrangements. In contrast, the duplications of two and three ancient chromosomes, followed by translocation and deletion events, generated the remaining four and five modern chromosomes, respectively (Velasco *et al*., 2010). In 2017, a *de novo* high-quality *M.* × *domestica* assembly was carried out on double haploid cv. 'Golden delicious' (Daccord *et al.,* 2017), thus overcoming the technical issues associated with sequencing and assembly of highly heterozygous genomes (Zharkikh *et al*., 2008). The genome size estimation was lowered to 650 Mb and 42.000 genes approximately, closer to the gene number reported for pear (Wu *et al*., 2013). However, the authors point out that some minor parts of the genome could have been lost in the assembly process (Daccord *et al*., 2017). Several other *Malus* genomes have been published in the last years (Zhang *et al.,* 2019; Chen *et al.*, 2020), but the previously mentioned assembly performed in 2017 by Daccord and colleagues is commonly considered the reference genome.

To date, apple production keeps facing new challenges, such as improving marketability characteristics (e.g. taste, storage, texture) and resistance to stress conditions. Besides environmental factors, many pathogens cause diseases in apple trees, posing a severe risk to their cultivation.

PHYTOPLASMAS: YELLOWS-ASSOCIATED BACTERIA

Biology and ecology of phytoplasmas

Phytoplasmas consist of a large group of specialized plant-pathogenic bacteria associated with yellow diseases spreading worldwide. The first evidence of phytoplasmas existence was given by Doi *et al.* (1967), who described wall-less pleomorphic bodies in the phloem sieve tube elements of infected plants and identified them as mycoplasma-like organisms (MLOs). Their association with yellow diseases was then confirmed taking advantage of phytoplasmas' high sensitivity to tetracycline (Hull *et al*., 1969; Granados, Maramorosch and Shikata, 1968). In the 90s, the phylogenetic classification of MLOs was improved by a 16S rDNA PCR amplification technique followed by restriction fragment length polymorphism (RFLP) analysis (Schneider *et al*., 1993; Namba, 1993), which led to the introduction of the term 'phytoplasma' (International Committee on Systematic Bacteriology Subcommittee on the Taxonomy of Mollicutes, 1993, 1997) and the provisional status '*Candidatus* Phytoplasma' ('*Ca.* P.') (Firrao *et al*., 2005).

Many plant traits can be affected as a consequence of phytoplasma infection. Typically detected symptoms are: stunting; yellowing, probably caused by high consumption of plant carbon sources by phytoplasmas (Oshima *et al*., 2007); witches' broom, consisting of many tiny shoots originating from a single point; phyllody, which consists in the formation of leaf-like tissues instead of flowers; floral virescence; abnormal proliferation; and purple top, observed as reddening of the upper leaves and apical part, caused by anthocyanin biosynthetic pathway activation (Maejima, Oshima *et al*., 2014; Himeno *et al*., 2014). Additionally, phytoplasma-infected plants show inhibition of phloem transport leading to an imbalanced accumulation of carbohydrates in mature leaves compared to storage organs (Maust *et al*., 2003; Lepka *et al*., 1999). Consequently, many primary physiological processes such as photosynthesis,

transpiration, phytohormone balances, root respiration and many others are disturbed (a complete list is available in Marcone, 2014).

Phytoplasmas usually reside in plant phloem sieve tube elements, and their transmission is mediated from plant to plant by phloem-feeding homopteran insects such as leafhoppers, planthoppers, and psyllids (Weintraub and Beanland, 2006). Once in phloem sieve tubes, phytoplasmas spread throughout the plant via the day/night bipolar movement of phloem flux (Wei *et al*., 2004). An alteration in the expression of approximately one-third of phytoplasmas' genes, mediated by transcription factors, allows the host switching between plants and insects. In particular, a strong upregulation of the large-conductance mechanosensitive channel (MscL) gene represents the high capacity of phytoplasma adaptation from the insect to the plant-cell osmotic pressure (Oshima *et al*., 2011). Some phytoplasmas can be transmitted by several insect vectors more than others, and, at the same time, many insect vectors can transmit more than one phytoplasma. However, a highly specific interaction is needed for the phytoplasma to cross the vector's gut and salivary glands' barriers (Namba, 2019). Three types of phytoplasma immunodominant membrane proteins play a key role in these interaction mechanisms: Imp, Amp (antigenic membrane protein), and IdpA (immunodominant membrane protein A) (Kakizawa *et al*., 2006). Amp is the most detected and it can form a complex with insect host microfilaments, while Imp interacts with plant actin. Since phytoplasmas lack genes coding for movements, the binding to the host microfilaments could allow their movement in the phloem and, thus, the success of their colonization (Boonrod *et al*., 2012).

Like many other obligate parasites, phytoplasma genome sizes are representative of the significant losses of their ancestors' chromosome portions (Razin *et al*., 1998). The size generally ranges from 530 to 1350 kb for most species and does not exceed 1000 kb for apple proliferation (AP) group members (Marcone *et al*., 1999). The small size of the genome is due to the loss of many metabolic pathways (Oshima *et al*., 2013), such as amino acid and fatty acid biosynthesis, tricarboxylic acid cycle, electron transport/oxidative phosphorylation, phosphotransferase transport system, pentose phosphate pathway, and even adenosine triphosphate (ATP) synthase. It is supposed

that the ATP synthesis in phytoplasma takes place through the glycolytic and malate pathways (Fraser *et al*., 1995; Oshima *et al*., 2007; Saigo *et al*., 2014). To date, no association of genome size with virulence is known. Phytoplasma genome is characterized by the occurrence of many potential mobile units (PMUs), whose clusters are repeated in the genome with a size up to 20 kb. The PMUs can act as potential independent replication units and form extrachromosomal units, which can integrate again in the genome. While several PMUs appear to be truncated and are considered remnants of insertion sequences, others also contain protein-encoding genes. Moreover, they are horizontally transmitted among different phytoplasmas (Toruño *et al*., 2010; Ku *et al*., 2013; Bai *et al*., 2006; Hogenhout *et al*., 2008).

Effector proteins

Phytoplasmas secrete effector proteins that interact with plant cell processes and are predicted to be key virulence factors. The release of these proteins into the host cell cytoplasm is probably due to a SecA- dependent protein translocation system (Kakizawa *et al*., 2004). The three best-characterized effector proteins are SAP11, SAP54 and TENGU.

The candidate effector protein SAP11 was identified in aster yellows-associated phytoplasma '*Ca.* P. asteris', strain AY-WB. Once released inside the phloem cells by phytoplasmas, this small peptide reaches its nuclear localization in the host plant cell (Bai *et al*., 2009), where it interacts directly with transcription factors, influencing the Jasmonic acid (JA) signalling pathway and plant morphology. In particular, SAP11 interacts with a subclass of plant Teosinte-branched 1, Cycloidea, Proliferating (TCP) transcription factor family, namely the Class II Cincinnata (CIN) subgroup, which control cell maturation and senescence (Sugio *et al.,* 2011a; Sugio *et al*., 2011b).

The effector protein SAP54, identified in AY-WB, promotes the degradation of members of the MADS-domain transcription factor (MTF) family via the ubiquitin-proteasome pathway. MTFs are involved in the regulation of floral development, and their deactivation induces phyllody. The SAP54-induced floral phenotype is thought to represent a dissemination strategy for phytoplasmas since these leaflike flowers are more attractive for insect vectors (MacLean *et al*., 2014; Maejima, Iwai, *et al*., 2014).

Finally, TENGU candidate effector protein was first identified in the onion yellows phytoplasma (OY). Expression of this peptide in *Nicotiana benthamiana* and *Arabidopsis thaliana* induced the down-regulation of many auxin-responsive genes and JA synthesis and the generation of witches' brooms and dwarfism (Hoshi *et al*., 2009; Minato *et al*., 2014). Despite being expressed only in the phloem, TENGU appears to be transported from into stem and shoot apical meristems, where the symptoms are induced (Hoshi *et al*., 2009; Sugawara *et al*., 2013).

APPLE PROLIFERATION DISEASE

Distribution, transmission and symptoms

Apple proliferation (AP) is one of the most important phytoplasmoses in Europe, spread across major apple growing areas in many European countries (Tedeschi *et al.*, 2013), with a few occurrences also recorded in Northern Africa and America (CABI/EPPO 2020). The disease has been described for the first time in the 1950s in the Italian regions of Veneto and Trentino (Refatti and Ciferri, 1954) and reported in Southern Germany (Kunze 1989) and South Tyrol (Österreicher and Thomann, 2003) soon after. Several outbreaks occurred in the last years in various parts of Europe, causing considerable economic damages. In Italy, the Trentino region was hit by an outbreak in the 90s (Vindimian *et al.,* 2002). In the same years, AP epidemics also occurred in North-Western Italy in Piedmont and Aosta Valley. In South Tyrol, the disease peaked severely in 2006 and 2013, leading to a loss of nearly 50 million € (Österreicher and Thomann 2015). In 2006, a sanitation program was implemented thanks to a Ministerial decree (MIPAAF 2006), which imposed the regular inspection of trees and removal of infected ones (or the whole orchard in case of infection higher than 25%). Consequently, the spread of AP declined, although monitoring activities and treatments to control the vectors are constantly being carried out (Janik *et al*., 2020).

Since phytoplasmas reside in the phloem, they can be transmitted by phloem-feeding insects. Although they can be ingested during phloem sap feeding, phytoplasmas do not necessarily multiply in the insect. To complete the replication process, the phytoplasma must pass from the gut epithelium to the salivary glands, which requires high adaptation to the vector insect (Hogenhout *et al.,* 2008; Alma *et al*., 2015). Therefore, the detection

of a phytoplasma in a phloem-feeding insect is not enough to demonstrate that the respective insect is able to transmit the pathogen, and only transmission trials can prove the effective transmission capacity. So far, only the two psyllids *Cacopsylla picta* and *Cacopsylla melanoneura* and the leafhopper *Fieberiella florii* have been proven to be AP vectors, although several phloem-feeding hemipterans were found to be occasional carriers of '*Ca*. P. mali' (Frisinghelli *et al.,* 2000; Tedeschi *et al.*, 2002; Jarausch *et al*., 2003; Tedeschi and Alma 2006; Carraro *et al.*, 2008; Alma *et al.*, 2015; Oppedisano *et al.,* 2019). Laboratory transmission trials showed that the phytoplasma transmission occurs in a persistent-propagative manner (Weintraub and Beanland 2006), which means that the pathogen multiplies in insects and that the ability to inoculate the pathogen persists for the whole life of the insect (Fletcher *et al.,* 1998). While the transmission through insect vectors is considered the most relevant, several examples of '*Ca*. P. mali' transmission by grafting, natural root bridges and dodder (Cuscuta sp., a parasitic plant that forms phloem connections between different plants) have been reported (Carraro *et al.*, 1988; Seemüller *et al.,* 2011; Luge *et al.,* 2014).

AP induces a broad range of specific and aspecific symptoms in wild and commercial *Malus* species (Bovey 1963; Blattny *et al.,* 1963; Refatti and Ciferri 1954; Morvan and Castelain 1975; Kartte and Seemüller 1988), yet the symptom expression can vary enormously (Schmid 1975). AP-specific symptoms, which means unequivocally related to a '*Ca*. P. mali' infection, consist in the formation of witches' brooms and enlarged stipules (Seemüller 1990; Jarausch 2007; Seemüller *et al*., 2011). AP induces then a wide range of non-exclusive, aspecific symptoms, whose severity often depends on the apple variety. These may indeed be induced by many other biotic or abiotic factors, such as fungal infections, mechanical damage, nutritional deficiencies, or hormonal treatments (Schmid 1975; Mattedi *et al.*, 2008a; 2008b) and include foliar reddening or chlorosis, stunted branches, rosette formation of apical leaves, formation of shoots from auxiliary buds and late flowering (Bovey 1963; Zawadzka 1976; Kartte and Seemüller 1988; Mattedi *et al.*, 2008b). As a final point, the fruits appear small, taste- and colourless (Blattny *et al.*, 1963; Zawadzka 1976; Seidl 1980; Schmidt *et al.*, 2009; Seemüller *et al*. 2011), causing the impossibility to put them on the market and, ultimately, the

economic impact of AP. Examples of symptoms affecting leaves and branches are displayed in figure 1.

Although the presence of specific symptoms undoubtedly indicates an infection, it is worth noting that the absence of any of the symptoms cannot exclude an ongoing infection. At the same time, the contemporaneous expression of two or more non-AP-specific symptoms is a reliable indication of an AP infection (Thomann and Tumler 2000; Mattedi *et al.*, 2008a; Öttl *et al.*, 2008). It is still not clear which factors influence the irregularity of the pattern of symptom development, that may include strain pathogenicity, the sensitivity of the cultivar, colonization behaviour of the phytoplasma, or other environmental factors (Seemüller *et al.*, 1984a; Seemüller 2002; Carraro *et al.*, 2004; Seemüller and Schneider 2007; Herzog *et al.*, 2010; Baric *et al.*, 2011), although a high phytoplasma concentration in the aerial parts of the trees was shown to be necessary for the development of severe symptoms. Typically, AP phytoplasma is eliminated from the aerial parts of the tree during winter due to sieve tube degeneration, persisting in the roots and recolonizing the above-ground parts of the trees in spring (Schaper and Seemüller 1984; Seemüller *et al.*, 1984a; 1984b; Bisognin *et al.*, 2008).

Figure 1: Examples of apple branches displaying the typical symptoms of AP infection (B, C, D) compared to a healthy shoot (A). Copyright Laimburg Research Centre.

Several studies aimed to clarify the molecular mechanisms that underlie the development of the symptoms. It was proposed that stunting and chlorosis may be caused by an alteration of carbohydrate metabolism (Lepka *et al.*, 1999), further confirmed by the observation that key genes of the Calvin cycle and photosynthesis

pathways are altered in AP infected plants (Aldaghi *et al.*, 2012; Luge *et al.*, 2014). Many studies suggested that symptom development could result from an effect of 'Ca. P. mali' on plant hormonal regulation (Luge *et al.*, 2014; Zimmermann *et al.*, 2015; Tan *et al.,* 2016; Janik *et al.,* 2017). A three-group clustering of apple genes affected by AP was proposed: a) genes involved in photosynthesis pathways, which appear deregulated being thus important for the development of symptoms; b) genes that affect auxin accumulation and senescence, that could cause the inhibition of apical dominance (stunting) and affect the biosynthesis of flavonoids, leading to leaf reddening; c) genes that regulate the plant defence driving the reduction of H_2O_2, increasing the plant susceptibility and favouring the phytoplasma multiplication (Aldaghi *et al.*, 2012).

The first AP effector protein, which shares homology with AY-WB effector SAP11, has been identified in previous studies (Siewert *et al.*, 2014; Janik *et al.,* 2017). SAP11-like of '*Ca*. P. mali' ($SAP11_{CaPM}$) binds two members of the subgroup CIN-Class II of TCP transcription factor family, *MdTCP24* and *MdTCP25* (Sugio *et al.*, 2011a; Janik *et al.*, 2017), which regulate different growth, defence, and developmental plant processes (Cubas *et al.*, 1999; Lopez *et al.*, 2015; Ikeda and Ohme-Takagi 2014). Moreover, *Nicotiana benthamiana* expressing $SAP11_{CaPM}$ showed how this effector could change the volatile profile of the plant, induce defects in trichome development and suppress Jasmonic acid responses (Tan *et al.*, 2016).

Recovery and tolerance to AP

Recovery is defined as the spontaneous remission of symptoms (Osler *et al.*, 2000), which does not imply a complete recovery from the infection since bacteria can still be detected in the plant, even if only in the roots. Eventually, bacteria can recolonize the upper part of the tree and trigger the reappearance of symptoms (Carraro *et al.*, 2004). Interestingly, it was observed that recovered trees are less prone to be reinfected than healthy trees are, thus indicating the possible occurrence of some form of induced resistance (Osler *et al.,* 2000).

Analyses of recovered, symptomless trees showed increased H_2O_2 and Ca^{2+} concentration, callose deposition and protein accumulation in the leaf phloem (Musetti *et al.*, 2004; Musetti *et al.*, 2010). It was proposed that the H_2O_2 accumulation might

counteract AP symptom development, while the callose deposition could induce a plugging of the phloem, physically blocking the distribution of the bacteria or the effector proteins translocation and thus leading to a loss of symptoms (Musetti *et al.*, 2011; Guerriero *et al.*, 2012). However, studies performed on *A. thaliana* infected with '*Ca*. P. asteris' showed no correlation between the phloem flow limitation and phytoplasma concentration (Pagliari *et al.*, 2017). Another hypothesis implies the role of salicylic acid (SA)- and Jasmonic acid (JA)- dependent defences, which leads to the accumulation of H_2O_2 and, consequently, to an increased level of SA (Musetti *et al.*, 2004; 2013). Moreover, accumulation of JA and a decrease of SA levels in recovered trees were also reported (Patui *et al*., 2013), further highlighting a reciprocal antagonism between the two pathways. The authors further suggested that the observed H_2O_2 accumulation might be caused by a decrease in the reactive oxygen species scavenging activity combined with peroxidase and oxidase. Ultimately, it is not clear whether the accumulation of H_2O_2 is a cause or consequence. Nonetheless, the concurrent activation of SA and JA pathways can likely improve plant defence responses (van Wees *et al.*, 2000).

While many *M. × domestica* cultivars have been observed to show different degrees of tolerance based on symptom occurrence in the field (Nemeth 1986; Thakur and Handa 1999; Richter 2003), an extensive search for AP natural genetic resistance has been carried out in the last years (Kartte and Seemüller 1991; Seemüller *et al.*, 1992). Here, hundreds of rootstocks and genotypes have been tested by graft-mediated infection trials and long-term monitoring. This led to the identification of the so-called inheritable 'resistance', defined in literature as the absence or the temporal development of mild symptoms and low phytoplasma titer. 'Resistant' trees consist of experimental rootstocks derived from a specific accession of the wild crabapple *Malus sieboldii.* However, scions grafted on these promising *M. sieboldii*-based rootstocks were observed to develop vigorous and low productive trees and, moreover, appear more susceptible to latent apple viruses (Kartte and Seemüller 1991; Seemüller *et al.*, 1992; Seemuller *et al.*, 2008; Bisognin *et al.,* 2008). To improve the agronomic value of these rootstocks, the genotypes were crossed with the standard rootstock M9 and screened

using the *in vitro* micrografting technique (Jarausch *et al.*, 1999; Bisognin *et al.,* 2009; Seemüller *et al.,* 2010; Seemüller and Harries 2010; Jarausch *et al.*, 2010; Jarausch *et al.,* 2011). After long-term field trials, 'resistant' genotypes with adequate pomological properties and tolerance to apple viruses were finally identified and are now entering the agronomic field studies phase (Seemüller *et al.*, 2018).

This definition of AP resistance is empirical and characterized by trees showing mild to no impact of phytoplasma infection. Moreover, the physiological and molecular bases of this phenomenon are, to date, unknown. Disease resistance requires the involvement of several processes of pathogen detection and immune system activation (Andersen *et al*., 2018), which could not be identified for AP resistance so far. For this reason, in the present work, we will refer to this phenomenon as 'tolerance'.

THE *TCP* TRANSCRIPTION FACTOR GENE FAMILY

The *TCP* (TEOSINTE BRANCHED 1, CYCLOIDEA, PCF) gene family consists of a large group of genes encoding for plant-specific transcription factor proteins (TF). The name 'TCP' is based on the first four members described, which play critical roles in the development of plant morphology: TEOSINTE BRANCHED 1 (TB1) from maize (*Zea mays*) (Doebley, Stec, and Hubbard 1997), CYCLOIDEA (CYC) from snapdragon (*Antirrhinum majus*) (Luo *et al*., 1996), and PROLIFERATING CELL FACTORS 1 and 2 (PCF1 and PCF2) from rice (*Oryza sativa*) (Kosugi and Ohashi 1997; Cubas *et al*., 1999). TCPs are found in all the land plants, from mosses to eudicots, and their presence was also observed in green algae. Sequence analyses show a rapid expansion from lower to higher plants, characterized by several duplication events (Navaud *et al*., 2007; Martín-Trillo and Cubas, 2010; Liu *et al*., 2019). A relatively low number of TCP genes have been found in the basal groups of land plants, e.g., six in the moss *Physcomitrella patens* and the lycophyte *Selaginella moellendorffii*. The number of TCPs increases to approximately 20-30 in monocots and eudicots (26 in rice, 24 in *A. thaliana*) with several species showing a larger number, like the 46 and 52 TCP genes respectively identified in maize and *M.* × *domestica* (Martín-Trillo and Cubas, 2010*;* Li, 2015; Ding *et al*., 2019; Xu *et al*., 2014).

All the TCP proteins share an N-terminal non-canonical basic-helix-loop-helix (bHLH) motif of 59 amino acids known as the TCP domain. This domain mediates nuclear

localization, DNA binding and protein-protein interactions. Interestingly, the TCP domain shows little sequence similarity to the canonical bHLH and binds to different DNA elements (Murre *et al.*, 1994; Cubas *et al*., 1999; Kosugi and Ohashi, 2002), thus representing a DNA binding domain uniquely found in the plant kingdom (Aggarwal, Gupta, *et al*. 2010).

Based on sequence homology, two distinct classes of TCP can be identified: Class I, which contains PCF1 and PCF2 (PCF-class, TCP-P) and Class II, which contains CYC and TB1 (TCP-C). The most evident difference between the two groups is a conserved deletion of four amino acids in the TCP domain of members of Class I. Additionally, Class II can be further divided into CINCINNATA (CIN) and CYC/TB1 subclasses. A subgroup of CIN clade presents a C-terminal recognition binding for a microRNA (miR319), which post-transcriptionally regulates their activity (Palatnik *et al.*, 2003).

The DNA binding sequences of the two TCP classes differ slightly, reflecting similar but distinct specificities between members of the two classes (Kosugi and Ohashi, 2002; Nicolas and Cubas, 2016). At the same time, the motifs are not mutually exclusive, suggesting that target genes may be potentially targeted by both Class I and Class II TCP proteins (Danisman, 2016), although an antagonistic relationship between the two classes was only described in one case (Danisman *et al*., 2012). Like many other transcription factor families, TCP proteins require dimerization to bind DNA (Trémousaygue *et al*., 2003), forming homo- and heterodimers preferably with members of the same class (Danisman *et al*., 2013).

TCP proteins participate in the regulation of a large number of biological processes related to plant growth, development, and response to various stimuli.

The regulation of branching by controlling the axillary meristematic activity is probably one of the most evolutionary conserved roles of TCPs being described in a bryophyte, where the knockout of a *TCP* resulted in an increase in the number of sporangia attached to a single seta, resembling the branching phenotypes of tcp mutants in monocots and eudicots (Ortiz-Ramírez *et al*., 2016; Takeda *et al.*, 2003; Aguilar-Martínez *et al.*, 2007; Braun *et al.*, 2012; Danisman, 2016).

Another effect of TCPs, Class II in particular, is related to floral development as shown in a wide range of species such as *A. thaliana* (Costa *et al.*, 2005), *Oryza sativa* (Yuan *et al*., 2009), and *Pisum sativum* (Wang *et al.*, 2008). For example, in *A. thaliana* the expression of a miR319-resistant form of *TCP4* in floral organs leads to smaller flowers missing petals and stamens (Nag *et al.*, 2009). Moreover, the alleviation of the normal transcription repression of three *TCP*s during flower development leads to an aberrant petal development caused by anomalies in the mitotic cell cycle (Huang and Irish, 2015). Several TCPs have been reported to play a role in regulating the cell cycle also during leaf development. In snapdragon, class II *TCP* mutant displays crinkly leaves (Nath *et al.*, 2003), similarly to *A. thaliana* jaw-D (overexpressor of miR319) mutant, showing delayed leaf development and senescence and serrated leaves (Palatnik *et al.*, 2003). Overexpression of miR319 leads to wider, thicker leaves in the monocot *Agrostis stolonifera* and to leaves growing three months longer than wild type in tomato (Ori *et al.*, 2007; Efroni *et al.*, 2008; Zhou *et al.*, 2013). Also, two *A. thaliana* Class I TCPs were shown to redundantly affect cell proliferation during leaf and stem development (Kieffer *et al.*, 2011).

Data indicate additional potential roles of TCP proteins in organ identity determination (Wang *et al.*, 2015; Mizuno *et al.*, 2015), stem development (Hao *et al.*, 2012), root development (Mukhopadhyay *et al.*, 2015) and affect growth by regulating hormone synthesis (Danisman, 2012; Danisman, 2016), circadian clock (Wu *et al.*, 2016) and the mitochondrial oxidative phosphorylation machinery (Giraud *et al.*, 2010).

Finally, the interest in potential roles of TCP proteins also in plant defence mechanisms is increasing after the observation that several pathogens have been observed to target TCP genes, like *Pseudomonas syringae*, *Hyaloperonospora arabidopsidis* (Mukhtar *et al.*, 2011), and phytoplasma in *A. thaliana* (Sugio *et al.*, 2011) as well as in *M*. × *domestica* (Janik *et al.*, 2017). Interestingly, an *A. thaliana* triple-TCP mutant infected with *P. syringae* showed an increasing growth rate of the pathogen compared to the wild-type plants (Kim *et al.*, 2014).

Given the large array of biological processes in which TCP genes are involved, it is likely to believe that many more will be discovered in the near future. At the same time, a

deeper understanding of the interaction networks involving TCP genes appears increasingly necessary to fully understand many aspects of plant biology.

AIM OF THE THESIS

The present work focuses on the study of the molecular targets of $SAP11_{CaPM}$, two members of the TCP transcription factor family, to get insights regarding the role of the two *MdTCPs* in the development of Apple Proliferation disease.

The second chapter consists of a review of the *MdTCP* gene family based on the latest high-quality apple genome assembly, followed by analyses of the intergenic differences and the similarity of *A. thaliana TCP.*

The third chapter focuses on the sequences of the two *MdTCPs* of interest, focusing on tolerant and susceptible genotypes, to find potential hints explaining the tolerant behaviour. Following extensive analyses of the amino acid substitutions found in the two *TCP* genes of tolerant genotypes, the interaction strength between bacterial effector and the protein alternatives was tested both in a heterologous system and *in planta*.

The fourth and last chapter involves the generation and characterization of transgenic plants overexpressing an *MdTCP* gene. Following the analyses of their expression and phenotypes, plants were infected with the phytoplasma and subsequently analyzed. Ultimately, some preliminary results related to an analysis of the transcriptome of transgenic plants are reported.

REFERENCES

Aggarwal P., Gupta M. Das, Joseph A. P., Chatterjee N., Sinivasan N. & Nath, U. (2010). Identification of specific DNA binding residues in the TCP family of transcription factors in arabidopsis. *Plant Cell*. 22 1174–1189 https://doi.org/10.1105/tpc.109.066647

Aguilar-Martínez J. A., Poza-Carrión C., Cubas P. (2007). Arabidopsis BRANCHED1 acts as an integrator of branching signals within axillary buds. *Plant Cell* 19 458–472. 10.1105/tpc.106.048934

Aldaghi M., Bertaccini A., Lepoivre P. (2012). cDNA-AFLP analysis of gene expression changes in apple trees induced by phytoplasma infection during compatible interaction. *European Journal of Plant Pathology* 134 (1): 117-130.

Alma A., Tedeschi R., Lessio F., Picciau L., Gonella E., Ferracini C. (2015). Insect vectors of plant pathogenic Mollicutes in the Euro-Mediterranean region. *Phytopathogenic Mollicutes* 5 (2): 53-73.

Andersen, E. J., Ali, S., Byamukama, E., Yen, Y., & Nepal, M. P. (2018). Disease resistance mechanisms in plants. *In Genes*. https://doi.org/10.3390/genes9070339

Bai, X., Correa, V.R., Toruño, T.Y., Ammar, E.D., Kamoun, S. and Hogenhout, S.A. (2009) AY-WB phytoplasma secretes a protein that targets plant cell nuclei. *Mol. Plant-Microbe Interact.*, 22, 18–30.

Bai, X., Zhang, J., Ewing, A., et al. (2006) Living with genome instability: The adaptation of phytoplasmas to diverse environments of their insect and plant hosts. *J. Bacteriol.*, 188, 3682–3696.

Baric S., Berger J., Cainelli C., Kerschbamer C., Letschka T., Dalla Via J. (2011b). Seasonal colonization of apple trees by '*Candidatus* Phytoplasma mali' revealed by a new quantitative TaqMan real-time PCR approach. *European Journal of Plant Pathology* 129: 455-467.

Bisognin C., Schneider B., Salm H., Grando M. S., Jarausch W., Moll E., Seemüller E. (2008b). Apple proliferation resistance in apomictic rootstocks and its relationship to phytoplasma concentration and simple sequence repeat genotypes. *Phytopathology* 98 (2): 153-158.

Blattny J. C., Seidl V., Erbenova M. (1963). The apple proliferation of various sorts and the possible strain differentiation of the virus. *Phytopathologia Mediterranea* 2 (3): 119-123.

Boonrod, K., Munteanu, B., Jarausch, B., Jarausch, W. and Krczal, G. (2012) An immunodominant membrane protein (Imp) of 'Candidatus Phytoplasma mali' binds to plant actin. *Mol. Plant-Microbe Interact.*, 25, 889–895.

Bovey R. (1963). Observations and experiments on apple proliferation disease. *Phytopathologia Mediterranea* 2 (3): 111-114.

Braun N., Germain A., de S., Pillot J.-P., Boutet-Mercey S., Dalmais M., et al. (2012). The Pea TCP transcription factor PsBRC1 acts downstream of strigolactones to control shoot branching. *Plant Physiol*. 158 225–238. 10.1104/pp.111.182725

Carraro L., Ermacora P., Loi N., Osler R. (2004). The Recovery Phenomenon in Apple Proliferation-Infected Apple Trees. *Journal of Plant Pathology* 86 (2): 141-146.

Carraro L., Ferrini F., Ermacora P., Loi N. (2008). Infectivity *of Cacopsylla picta (Syn. Cacopsylla costalis),* Vector of *'Candidatus Phytoplasma mali'* in North East Italy. *Acta Horticolturae* 781: 403-408.

Carraro L., Osler R., Refatti E., Poggi P. C. (1988). Transmission of the possible agent of apple proliferation to *Vinca rosea* by dodder. *Rivista di Patologia Vegetale* 24 (4): 43-52.

Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *In Trends in Genetics*. https://doi.org/10.1016/j.tig.2013.10.002

Cornille, A., Gladieux, P., Smulders, M. J. M., Roldán-Ruiz, I., Laurens, F., Le Cam, B., Nersesyan, A., Clavel, J., Olonova, M., Feugey, L., Gabrielyan, I., Zhang, X. G., Tenaillon, M. I., & Giraud, T. (2012). New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1002703

Costa M. M. R., Fox S., Hanna A. I., Baxter C., Coen E. (2005). Evolution of regulatory interactions controlling floral asymmetry. *Development* 132 5093–5101. 10.1242/dev.02085

Cubas, P., Lauter, N., Doebley, J., & Coen, E. (1999). The TCP domain: A motif found in proteins regulating plant growth and development. *Plant Journal*. https://doi.org/10.1046/j.1365-313X.1999.00444.x

Daccord, N., Celton, J. M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., Van De Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E. A., Gouzy, J., Rees, D. J. G., Guérif, P., Muranty, H., Durel, C. E., Laurens, F., Lespinasse, Y., Gaillard, S., … Bucher, E. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*. https://doi.org/10.1038/ng.3886

Danisman S., Wal F., van der Dhondt S., Waites R., Folter S., de Bimbo A., et al. (2012). Arabidopsis Class I and Class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. *Plant Physiol*. 159 1511–1523. 10.1104/pp.112.200303

Danisman, S. (2016). TCP transcription factors at the interface between environmental challenges and the plant's growth responses. *In Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2016.01930

Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., van Dijk, A. J., Muino, J. M., Cutri, L., Dornelas, M. C., Angenent, G. C., & Immink, R. G. H. (2012). Arabidopsis class I and class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. *Plant Physiology*. https://doi.org/10.1104/pp.112.200303

Danisman S., Van Dijk A. D. J., Bimbo A., Van Der Wal F., Hennig L., De Folter S., Angenent G. C., & Immink R. G. H. (2013). Analysis of functional redundancies within the Arabidopsis TCP transcription factor family. *Journal of Experimental Botany*- 64 5673–5685. https://doi.org/10.1093/jxb/ert337

Dell, T. La, Morgan, J., & Richards, A. (2003). The New Book of Apples. *Garden History*. https://doi.org/10.2307/1587407

Ding, S., Cai, Z., Du, H., & Wang, H. (2019). Genome-wide analysis of TCP family genes in Zea mays l. Identified a role for ZmTCP42 in drought tolerance. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms20112762

Doebley, J., Stec, A. and Hubbard, L. (1997). The evolution of apical dominance in maize. *Nature*. https://doi.org/10.1038/386485a0

Doebley, J., Stec, A. and Gustus, C. (1995). Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* 141, 333–346.

Doi, Y., Teranaka, M., Yora, K., and Asuyama, H. (1967). Mycoplasma- or PLT group-like microorganisms found in the phloem elements of plants infected with mulberry dwarf, potato witches' broom, aster yellows or paulownia witches' broom. *Ann. Phytopathol. Soc. Jpn*. 33:259–266. doi: 10.3186/jjphytopath.33.259

Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., Wang, X., Jiao, C., Legall, N., Mao, L., Wan, S., Wang, K., He, T., Feng, S., Zhang, Z., Mao, Z., Shen, X., Chen, X., Jiang, Y., … Chen, X. (2017). Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications*, 8(1). https://doi.org/10.1038/s41467-017-00336-7

Efroni I., Blum E., Goldshmidt A., Eshed Y. (2008). A protracted and dynamic maturation schedule underlies Arabidopsis leaf development. *Plant Cell* 20 2293–2306. 10.1105/tpc.107.057521

Firrao, G., Gibb, K. and Streten, C. (2005) Short taxonomic guide to the genus 'Candidatus Phytoplasma'. *J. Plant Pathol.*, 87, 249–263.

Fletcher J., Wayadande A., Melcher U., Ye F. (1998). The phytopathogenic mollicute-insect vector interface: a closer look. *Phytopathology* 88 (12): 1351-1358.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Fritchman, J.L. (1995) The Minimal Gene Complement of Mycoplasma genitalium. *Science* (80-.)., 270, 397–404.

Frisinghelli C., Delaiti M., Grando S., Forti D., Vindimian E. (2000). Cacopsylla costalis (Flor 1861) as a Vector of Apple Proliferation in Trentino. *Journal of Phytopathology* 148:425-431.

Giraud E., Ng S., Carrie C., Duncan O., Low J., Lee C. P., et al. (2010). TCP transcription factors link the regulation of genes encoding mitochondrial proteins with the circadian clock in Arabidopsis thaliana. *Plant Cell* 22 3921–3934. 10.1105/tpc.110.074518

Granados, R.R., Maramorosch, K., Shikata, E. (1968) Mycoplasma: Suspected etiologic agent of corn stunt. *Proc. Natl. Acad. Sci.*, 60, 841–844.

Guerriero G., Giorno F., Ciccotti A. M., Schmidt S., Baric S. (2012a). A gene expression analysis of cell wall biosynthetic genes in *Malus* x *domestica* infected by '*Candidatus* Phytoplasma mali'. *Tree Physiology* 32 (11): 1365-1377.

Hao J., Tu L., Hu H., Tan J., Deng F., Tang W., et al. (2012). GbTCP, a cotton TCP transcription factor, confers fibre elongation and root hair development by a complex regulating system. *J. Exp. Bot*. 63 6267–6281. 10.1093/jxb/ers278

Herzog U., Wiedemann W., Trapp A. (2010). Apfeltriebsucht in Sachsen. *Schriftenreihe des LfULG* (19).

Himeno, M., Kitazawa, Y., Yoshida, T., Maejima, K., Yamaji, Y., Oshima, K., Namba, S. (2014) Purple top symptoms are associated with reduction of leaf cell death in phytoplasma-infected plants. *Sci. Rep*., 4, 1–7.

Hogenhout S. A., Oshima K., Ammar E. D., Kakizawa S., Kingdom H. N., Namba S. (2008). Phytoplasmas: bacteria that manipulate plants and insects. *Molecular Plant Pathology* 9 (4): 403-423.

Hoshi, A., Oshima, K., Kakizawa, S., et al. (2009) A unique virulence factor for proliferation and dwarfism in plants identified from a phytopathogenic bacterium. *Proc. Natl. Acad. Sci. U. S. A.*, 106, 6416–6421.

Huang T., Irish V. F. (2015). Temporal control of plant organ growth by TCP transcription factors. *Curr. Biol.* 25 1765–1770. 10.1016/j.cub.2015.05.024

Hull, R., Horne, R.W. and Nayar, R.M. (1969) Mycoplasma-like bodies associated with sandal spike disease [29]. *Nature* 224, 1121–1122.

Ikeda M., Ohme-Takagi M. (2014). TCPs, WUSs, and WINDs: families of transcription factors that regulate shoot meristem formation, stem cell maintenance, and somatic cell differentiation. *Frontiers in Plant Science* 5:427.

Janik K., Mithöfer A., Raffeiner M., Stellmach H., Hause B., Schlink K. (2017). An effector of apple proliferation phytoplasma targets TCP transcription factors - a generalized virulence strategy of phytoplasma? *Molecular Plant Pathology* 18 (3): 435-443.

Janik, K., Barthel, D., Oppedisano, T., Weil, T. (2020). *Apple Proliferation: a Joint Review*

(K. Janik & G. Anfora (eds.)). Laimburg Research Centre; Edmund Mach Foundation.

Jarausch B., Fuchs A., Schwind N., Krczal G., Jarausch W. (2007). Cacopsylla picta as most important vector for '*Candidatus Phytoplasma mali*' in Germany and neighbouring regions. *Bulletin of Insectology* 60: 189-190.

Jarausch B., Jarausch W. (2010). Psyllid vectors and their control. In: Weintraub P. G., Jones P. (Eds.): Phytoplasmas - Genomes, Plant Hosts and Vectors. Oxfordshire, UK: CABI: 250-271.

Jarausch B., Schwind N., Jarausch W., Krczal G., Seemüller E., Dickler E. (2003). First report of Cacopsylla picta as a vector for apple proliferation phytoplasma in Germany. *Plant Disease* 87 (1): 101.

Jarausch W., Bisognin C., Schneider B., Grando S., Velasco R., Seemüller E. (2011). Breeding apple proliferation-resistant rootstocks: durability of resistance and pomological evaluation. *Bulletin of Insectology* 64 (Supplement): S275-S276.

Kakizawa, S., Oshima, K., Nishigawa, H., *et al*. (2004) Secretion of immunodominant membrane protein from onion yellows phytoplasma through the Sec protein-translocation system in *Escherichia coli*. *Microbiology*, 150: 135–142.

Kakizawa, S., Oshima, K., Namba, S. (2006) Diversity and functional importance of phytoplasma membrane proteins. *Trends Microbiol*., 14: 254-256. https://doi.org/10.1016/j.tim.2006.04.008

Kartte S., Seemüller E. (1988). Variable response within the genus *Malus* to the apple proliferation disease. *Journal of Plant Diseases and Protection* 95 (1): 25-34.

Kartte S., Seemüller E. (1991). Susceptibility of grafted *Malus* taxa and hybrids to apple proliferation disease. *Journal of Phytopathology* 131: 137-148.

Kieffer M., Master V., Waites R., Davies B. (2011). TCP14 and TCP15 affect internode length and leaf shape in Arabidopsis. *Plant J*. 68: 147–158.

Kim S. H., Son G. H., Bhattacharjee S., Kim H. J., Nam J. C., Nguyen P. D. T., et al. (2014). The Arabidopsis immune adaptor SRFR1 interacts with TCP transcription factors that redundantly contribute to effector-triggered immunity. *Plant J*. 78: 978–989. 10.1111/tpj.12527

Kosugi, S., and Ohashi, Y. (1997). PCF1 and PCF2 specifically bind to cis elements in the

rice proliferating cell nuclear antigen gene. *Plant Cell* 9, 1607–1619.

Kosugi, S., & Ohashi, Y. (2002). DNA binding and dimerization specificity and potential targets for the TCP protein family. *Plant Journal*.

Ku, C., Lo, W.-S. and Kuo, C.-H. (2013) Horizontal transfer of potential mobile units in phytoplasmas. *Mob. Genet. Elements*, 3, e26145.

Kunze L. (1989). Apple proliferation. In: *Virus and Virus-like Diseases of Pome Fruits and Simulating Noninfectious Disorders*, Cooperative Extension College of Agriculture and Home Economics. Pullmann, WA. Washington State University: 99-113.

Lepka P. Stitt M., Moll E., Seemüller E. (1999). Effect of phytoplasmal infection on concentration and translocation of carbohydrates and amino acids in periwinkle and tobacco. *Physiological and Molecular Plant Pathology* 55: 59-68.

Li, S. (2015). The Arabidopsis thaliana TCP transcription factors: A broadening horizon beyond development. *Plant Signaling and Behavior*. https://doi.org/10.1080/15592324.2015.1044192

Liu, M. M., Wang, M. M., Yang, J., Wen, J., Guo, P. C., Wu, Y. W., Ke, Y. Z., Li, P. F., Li, J. N., & Du, H. (2019). Evolutionary and comparative expression analyses of TCP transcription factor gene family in land plants. *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms20143591

Lopez J. A., Sun Y., Blair P. B., Mukhtar M. S. (2015). TCP three-way handshake: linking developmental processes with plant immunity. *Trends in Plant Science* 4: 238-245.

Luge T., Kube M., Freiwald A., Meierhofer D., Seemüller E., Sauer S. (2014). Transcriptomics assisted proteomic analysis of *Nicotiana occidentalis* infected by '*Candidatus* Phytoplasma mali' strain AT. *Proteomics* 14 (16): 1882-1889.

Luo, D., Carpenter, R., Vincent, C., Copsey, L., Coen, E. (1996). Origin of floral asymmetry in Antirrhinum. *Nature* 383: 794–799. https://doi.org/10.1038/383794a0

MacLean, A.M., Orlovskis, Z., Kowitwanich, K., Zdziarska, A.M., Angenent, G.C., Immink, R.G.H. and Hogenhout, S.A. (2014) Phytoplasma Effector SAP54 Hijacks Plant Reproduction by Degrading MADS-box Proteins and Promotes Insect Colonization in a RAD23-Dependent Manner. *PLoS Biol*., 12.

Maejima, K., Iwai, R., Himeno, M., et al. (2014) Recognition of floral homeotic MADS

domain transcription factors by a phytoplasmal effector, phyllogen, induces phyllody. *Plant J.*, 78, 541–554.

Maejima, K., Oshima, K. and Namba, S. (2014) Exploring the phytoplasmas, plant pathogenic bacteria. *J. Gen. Plant Pathol.*, 80, 210–221.

Marcone, C. (2014) Molecular biology and pathogenicity of phytoplasmas. *Ann. Appl. Biol.*, 165, 199–221.

Marcone, C., Neimark, H., Ragozzino, A., Lauer, U. and Seemüller, E. (1999) Chromosome sizes of phytoplasmas composing major phylogenetic groups and subgroups. *Phytopathology*, 89, 805–810.

Martín-Trillo, M., & Cubas, P. (2010). TCP genes: a family snapshot ten years later. *Trends in Plant Science*, 15(1), 31–39. https://doi.org/10.1016/j.tplants.2009.11.003

Mattedi L., Ciccotti A. M., Bianchedi P. L., Bragagna P., Deromedi M., Filippi M., Forno F., Pedrazzoli F. (2008a). Trasmissione di Apple proliferation attraverso anastomosi radicali. In: Ioriatti C., Jarausch W. (Eds.): *Scopazzi del melo - Apple proliferation*. San Michele all'Adige (TN): Fondazione Edmund Mach: 76-91.

Mattedi L., Forno F., Branz A., Bragagna P., Battocletti I., Gualandri V., Pedrazzoli F., Bianchedi P. L., Deromedi M., Filippi M., Dallabetta N., Varner M., Ciccotti A. M. (2008b). Come riconoscere la malattia in campo: novità sulla sintomatologia. In: Ioriatti C., Ja-rausch W. (Eds.): *Scopazzi del melo - Apple proliferation.* San Michele all'Adige (TN): Fondazione Edmund Mach: 41-50.Mattedi L., Forno F., Branz A., Piffer I., G

Maust, B.E., Espadas, F., Talavera, C., Aguilar, M., Santamaría, J.M. and Oropeza, C. (2003) Changes in carbohydrate metabolism in coconut palms infected with the lethal yellowing phytoplasma. *Phytopathology*, 93, 976–981.

Minato, N., Himeno, M., Hoshi, A., et al. (2014) The phytoplasmal virulence factor TENGU causes plant sterility by downregulating of the jasmonic acid and auxin pathways. *Sci. Rep*., 4, 1–7.

Ministero delle Politiche Agricole e Forestali (23.02.2006). Misure per la lotta obbligatoria contro il fitoplasma *Apple Proliferation Phytoplasma*.

Mizuno S., Sonoda M., Tamura Y., Nishino E., Suzuki H., Sato T., et al. (2015). Chiba

Tendril-Less locus determines tendril organ identity in melon (Cucumis melo L.) and potentially encodes a tendril-specific TCP homolog. *J. Plant Res.* 128 941–951. 10.1007/s10265-015-0747-2

Morvan G., Castelain C. (1975). Nouvelles observations sur la sensibilitè de *Malus* x *dawsoniana* Rehd. à la maladie de la proliferation du pommier et sur son utilisation comme indicateur. *Acta Horticolturae* (44): 175-180.

Mukhopadhyay P., Tyagi A. K. (2015). OsTCP19 influences developmental and abiotic stress signaling by modulating ABI4-mediated pathways. *Sci. Rep.* 5:9998 10.1038/srep09998

Mukhtar M. S., Carvunis A.-R., Dreze M., Epple P., Steinbrenner J., Moore J., *et al.* (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333 596–601. https://doi.org/10.1126/science.1203659

Murre, C., Bain, G., van Dijk, M. A., Engel, I., Furnari, B. A., Massari, M. E., Matthews, J. R., Quong, M. W., Rivera, R. R., Stuiver, M. H. (1994). Structure and function of helix-loop-helix proteins. In BBA - Gene Structure and Expression. https://doi.org/10.1016/0167-4781(94)90001-9

Musetti R., De Marco F., Farhan K., Polizzotto R., Santi S., Ermacora P., Osler R. (2011). Phloem-specific protein expression patterns in apple and grapevine during phytoplasma infection and recovery. *Bulletin of Insectology 64* (Supplement): S211-S212.

Musetti R., Farhan K., De Marco F., Polizzotto R., Paolacci A., Ciaffi M., Ermacora P., Grisan S., Santi S., Osler R. (2013). Differentially-regulated defence genes in Malusdomestica during phytoplasma infection and recovery. *European Journal of Plant Pathology* 136: 13-19.

Musetti R., Paolacci A., Ciaffi M., Tanzarella O. A., Polizzotto R., Tubaro F., Mizzau M., Ermacora P., Badiani M., Osler R. (2010). Phloem cytochemical modification and gene expression following the recovery of apple plants from apple proliferation disease. *Phytopathology* 100 (4): 390-399.

Musetti R., Sanità di Toppi L., Ermacora P., Favali M. A. (2004). Recovery in apple trees infected with the apple proliferation phytoplasma: an ultrastructural and

biochemical study. *Phytopathology* 94 (2): 203-208.

Nag A., King S., Jack T. (2009). miR319a targeting of TCP4 is critical for petal growth and development in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 106 22534–22539. 10.1073/pnas.0908718106

Namba, S. (1993) Detection and Differentiation of Plant-Pathogenic Mycoplasmalike Organisms Using Polymerase Chain Reaction. *Phytopathology*, 83, 786.

Namba, S. (2019) Molecular and biological properties of phytoplasmas. *Proc. Japan Acad. Ser. B Phys. Biol. Sci.*, 95, 401–418.

Nath U., Crawford B. C. W., Carpenter R., Coen E. (2003). Genetic control of surface curvature. *Science* 299 1404–1407. 10.1126/science.1079354

Navaud, O., Dabos, P., Carnus, E., Tremousaygue, D., Hervé, C. (2007). TCP transcription factors predate the emergence of land plants. *Journal of Molecular Evolution*. https://doi.org/10.1007/s00239-006-0174-z

Németh M. V. (1986). Virus, Mycoplasma and Rickettsia Diseases of Fruit Trees. Lancaster, Boston, USA/Dordrecht, Netherlands: M. Nijhoff Publishers.

Nicolas, M., Cubas, P. (2016). TCP factors: New kids on the signaling block. *In Current Opinion in Plant Biology*. 33:33–41 https://doi.org/10.1016/j.pbi.2016.05.006

Oppedisano T., Panassiti B., Pedrazzoli F., Mittelberger C., Bianchedi P. L., Angeli G., De Cristofaro A., Janik K., Anfora G., Ioriatti C. (2019). Importance of psyllids' life stage in the epidemiology of apple proliferation phytoplasma. *Journal of Pest Science* 93: 49-61.

Ori N., Cohen A. R., Etzioni A., Brand A., Yanai O., Shleizer S., et al. (2007). Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. *Nat. Genet.* 39 787–791. 10.1038/ng2036

Ortiz-Ramírez C., Hernandez-Coronado M., Thamm A., Catarino B., Wang M., Dolan L., et al. (2016). A transcriptome atlas of physcomitrella patens provides insights into the evolution and development of land plants. *Mol. Plant.* 9 205–220. 10.1016/j.molp.2015.12.002

Oshima, K., Ishii, Y., Kakizawa, S., et al. (2011) Dramatic transcriptional changes in an intracellular parasite enable host switching between plant and insect. *PLoS One*, 6.

Oshima, K., Kakizawa, S., Arashida, R., Ishii, Y., Hoshi, A., Hayashi, Y., Kagiwada, S. and Namba, S. (2007) Presence of two glycolytic gene clusters in a severe pathogenic line of Candidatus Phytoplasma asteris. *Mol. Plant Pathol.*, 8, 481–489.

Oshima, K., Maejima, K. and Namba, S. (2013) Genomic and evolutionary aspects of phytoplasmas. Front. *Microbiol*., 4, 1–8.

Osler R., Loi N., Carraro L., Ermacora P., Refatti E. (2000). Recovery in plants affected by phytoplasmas. Proceedings of the 5th Congress of the European Foundation for Plant Pathology, Taormina, Italy: 589-592.

Österreicher J., Thomann M. (2003). Apfeltriebsucht in Südtirol. *Obstbau Weinbau* (11): 305-307.

Öttl S., Baric S., Dalla Via J. (2008). Teilweise Rotfärbung weist nicht auf Apfeltriebsucht hin. *Obstbau Weinbau* (2): 58-59.

Pagliari L., Buoso S., Santi S., Furch A. C. U., Martini M., Degola F., Loschi A., van Bel A. J.E., Musetti R. (2017). Filamentous sieve element proteins are able to limit phloem mass flow, but not phytoplasma spread. *Journal of Experimental Botany* 68 (13): 3673-3688.

Palatnik J. F., Allen E., Wu X., Schommer C., Schwab R., Carrington J. C., et al. (2003). Control of leaf morphogenesis by microRNAs. *Nature* 425 257–263. 10.1038/nature01958

Patui S., Bertolini A., Clincon L., Ermacora P., Braidot E., Vianello A., Zancani M. (2013). Involvement of plasma membrane peroxidases and oxylipin pathway in the recovery from phytoplasma disease in apple (*Malus domestica*). *Physiologia Plantarum* 148 (2): 200-213.

Qian, G. Z., Liu, L. F., & Tang, G. G. (2010). (1933) Proposal to conserve the name Malus domestica against M. pumila, M. communis, M. frutescens, and Pyrus dioica (Rosaceae). *Taxon*. https://doi.org/10.1002/tax.592038

Razin, S., Yogev, david and Naot, Y. (1998) Molecular Biology and Pathogenicity of Mycoplasmas. *Microbiol. Mol. Biol. Rev.*, 62, 1094–1156.

Refatti E., Osler R., Loi N., Roggero P. (1986). Research on transmission of apple proliferation. *Acta Horticolturae* 193: 345-350.

Richter S. (2003). Achtung auf Apfeltriebsucht and Flachästigkeit. *Besseres Obst* 12: 4-6.

Saigo, M., Golic, A., Alvarez, C.E., Andreo, C.S., Hogenhout, S.A., Mussi, M.A. and Drincovich, M.F. (2014) Metabolic regulation of phytoplasma malic enzyme and phosphotransacetylase supports the use of malate as an energy source in these plant pathogens. *Microbiol*. (United Kingdom), 160, 2794–2806.

Schaper U., Seemüller E. (1984). Recolonization of the stem of apple proliferation and pear decline-diseased trees by the causal organisms in spring. *Zeitschrift Für Pflanzenkrankheiten Und Pflanzenschutz / Journal of Plant Diseases and Protection* 91: 608-613.

Schmid G. (1975). Prolonged observations on spread and behaviour of proliferation disease in apple orchards. *Acta Horticolturae* (44): 183-192.

Schmidt S., Vanas V., Schweigkofler W., Öttl S., Berger J., Cainelli C., Baric S. (2009). Das Schwerpunktprojekt Apfeltriebsucht am Versuchszentrum Laimburg. *Obstbau Weinbau* (7): 272-277.

Schneider, B., Ahrens, U., Kirkpatrick, B.C. and Seemuller, E. (1993) Classification of plant-apthogenic mycoplasma-like organism using restriction-site analysis of PCR-amplified 16S rDNA. J. *Gen. Microbiol.*, 139, 519–527.

Seemüller E. (1990). Apple proliferation. Jones, A. L., Aldwinkel, H. S. (Eds.): Compendium of apple and pear diseases. St. Paul, MN, USA: APS Press: 67-68.

Seemüller E. (2002). Apple proliferation: Etiology, epidemiology and detection. In: Brunelli A., Canova A. (Eds.): Proceedings Giornate Fitopatologiche, Vol. 1. Baselga di Piné (TN): 3-6.

Seemüller E., Carraro L., Jarausch W., Schneider B. (2011). Apple Proliferation Phytoplasma. In: Hadidi A., Barba M., Candresse T., Jelkmann W. (Eds): Virus and Virus-like diseases or Pome and Stone fruits: The American Phytopathological Society: 67-73.

Seemüller E., Gallinger J., Jelkmann W., Jarausch W. (2018). Inheritance of apple proliferation resistance by parental lines of apomictic *Malus sieboldii* as donor of resistance in rootstock breeding. *European Journal of Plant Pathology* 151, (3): 767-779.

Seemüller E., Kartte S., Kunze L. (1992). Resistance in established and experimental apple rootstocks to apple proliferation disease. *Acta Horticolturae* 309: 245-251.

Seemüller E., Kunze L., Schaper U. (1984a). Colonisation behavior of MLO, and symptom expression of proliferation-diseased apple trees and decline-diseased pear trees over a period of several years. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz / Journal of Plant Diseases and Protection* 91: 525-532.

Seemüller E., Moll E., Schneider B. (2008). Apple proliferation resistance of *Malus sieboldii*-based rootstocks in comparison to rootstocks derived from other *Malus* species. *European Journal of Plant Pathology* 121 (2): 109-119.

Seemüller E., Schaper U., Zimbelmann F. (1984b). Seasonal variation in the colonization patterns of mycoplasmalike organisms associated with apple proliferation and pear decline. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz / Journal of Plant Diseases and Protection* 91: 371-382.

Seemüller E., Schneider B. (2007). Differences in Virulence and Genomic Features of Strains of '*Candidatus* Phytoplasma mali', the Apple Proliferation Agent. *Phytopathology* 97 (8): 964-970.

Seidl V. (1980). Some results of several years' on apple proliferation disease. *Acta Phytopathologica Academiae Scientiarum Hungaricae* 15: 241-245.

Siewert C., Luge T., Duduk B., Seemüller E., Büttner C., Sauer S., Kube M. (2014). Analysis of expressed genes of the bacterium '*Candidatus* Phytoplasma mali' highlights key features of virulence and metabolism. *PLoS ONE* 9 (4), art. e94391.

Sugawara, K., Honma, Y., Komatsu, K., Himeno, M., Oshima, K. and Namba, S. (2013) The Alteration of plant morphology by small peptides released from the proteolytic processing of the bacterial peptide TENGU. *Plant Physiol*., 162, 2005–2014.

Sugio A., Kingdom H. N., MacLean A. M., Grieve V. M., Hogenhout S. A. (2011a). Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proceedings of the National Academy of Sciences USA* 108 (48): E1254-E1263.

Sugio, A., MacLean, A.M., Kingdom, H.N., Grieve, V.M., Manimekalai, R. and Hogenhout, S.A. (2011b) Diverse targets of phytoplasma effectors: From plant development to

defense against insects. *Annu. Rev. Phytopathol*., 49, 175–195.

Sun, X., Jiao, C., Schwaninger, H., Chao, C. T., Ma, Y., Duan, N., Khan, A., Ban, S., Xu, K., Cheng, L., Zhong, G. Y., & Fei, Z. (2020). Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nature Genetics*. https://doi.org/10.1038/s41588-020-00723-9

Takeda T., Suwa Y., Suzuki M., Kitano H., Ueguchi-Tanaka M., Ashikari M., et al. (2003). The OsTB1 gene negatively regulates lateral branching in rice. *Plant J.* 33 513–520. 10.1046/j.1365-313X.2003.01648.x

Tan C. M., Li C. H., Tsao N. W., Su L. W., Lu Y. T., Chang S. H., Lin Y. Y., Liou J. C., Hsieh L-C., Yu J. Z., Sheue C. R., Wang S. Y., Lee C. F, Yang J. Y. (2016). Phytoplasma SAP11 alters 3-isobutyl-2-methoxypyrazine biosynthesis in *Nicotiana benthamiana* by suppressing *NbOMT1*. *Journal of Experimental Botany* 67 (14): 4415-4425.

Tedeschi R., Alma A. (2006). Fieberiella florii (Homoptera: Auchenorrhyncha) as a Vector of "*Candidatus Phytoplasma mali*". *Plant Disease* 90 (3): 284-290.

Tedeschi R., Bosco D., Alma A. (2002). Population dynamics of *Cacopsylla melanoneura* (Homoptera: Psyllidae), a vector of apple proliferation phytoplasma in northwestern Italy. *Journal of Economic Entomology* 95 (3): 544-551.

Thakur P., Handa A. (1999). Phytoplasma Diseases of Temperate Fruits. In: Verma L. R., Sharma R. C. (Eds.) Diseases of Horticultural Crops: Fruits. New Dehli: Indus: 291-315.

Thomann M., Tumler E. (2000). Der Besenwuchs des Apfels. *Obstbau Weinbau* (10): 293-294.

Toruño, T.Y., Seruga Musić, M., Simi, S., Nicolaisen, M. and Hogenhout, S.A. (2010) Phytoplasma PMU1 exists as linear chromosomal and circular extrachromosomal elements and has enhanced expression in insect vectors compared with plant hosts. *Mol. Microbiol*., 77, 1406–1415.

Trémousaygue D., Garnier L., Bardet C., Dabos P., Hervé C., Lescure B. (2003). Internal telomeric repeats and "TCP domain" protein-binding sites co-operate to regulate gene expression in Arabidopsis thaliana cycling cells. *Plant Journal*. 33 957–966 https://doi.org/10.1046/j.1365-313X.2003.01682.x

van Wees S.C., de Swart E.A., van Pelt J. A., van Loon L. C., Pieterse C.M. (2000). Enhancement of induced disease resistance by simultaneous activation of salicylate- and jasmonate- dependent defense pathways in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences USA* 97 (15): 8711-8716.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, Viola, R. *et al.* (2010). The genome of the domesticated apple (Malus × domestica Borkh.). *Nature Genetics*. https://doi.org/10.1038/ng.654

Vindimian M. E., Ciccotti A. M., Filippi M., Springhetti M., Deromedi M. (2002). Trasmissione di Apple Proliferation (AP) tramite anastomosi radicale. *Petria 12* (3): 375.

Wang S., Yang X., Xu M., Lin X., Lin T., Qi J., et al. (2015). A rare SNP identified a TCP transcription factor essential for tendril development in cucumber. *Mol. Plant* 8 1795–1808. 10.1016/j.molp.2015.10.005

Wang X., Gao J., Zhu Z., Dong X., Wang X., Ren G., et al. (2015). TCP transcription factors are critical for the coordinated regulation of ISOCHORISMATE SYNTHASE 1 expression in Arabidopsis thaliana. *Plant J.* 82 151–162. 10.1111/tpj.12803

Wang Z., Luo Y., Li X., Wang L., Xu S., Yang J., et al. (2008). Genetic control of floral zygomorphy in pea (Pisum sativum L.). *Proc. Natl. Acad. Sci. U.S.A.* 105 10414–10419. 10.1073/pnas.0803291105

Wei, W., Kakizawa, S., Suzuki, S., et al. (2004) In planta dynamic analysis of onion yellows phytoplasma using localized inoculation by insect transmission. *Phytopathology*, 94, 244–250.

Weintraub P. G., Beanland L. (2006). Insect vectors of phytoplasmas. *Annual Review of Entomology* 51: 91-111.

Wu J.-F., Tsai H.-L., Joanito I., Wu Y.-C., Chang C.-W., Li Y.-H., et al. (2016). LWD–TCP complex activates the morning gene CCA1 in Arabidopsis. *Nat. Commun*. 7:13181 10.1038/ncomms13181

Wu, J., Wang, Z., Shi, Z., Zhang, S., *et al.* (2013). The genome of the pear (Pyrus bretschneideri Rehd.). *Genome Research*. https://doi.org/10.1101/gr.144311.112

Xua, R., Sunb, P., Jiab, F., Luc, L., Lib, Y., Zhang, S., & Huang, J. (2014). Genome-wide

Analysis of TCP Transcription Factor Gene Family in Malus Domestica. *Ias.Ac.In*, 93(3), 733–746. http://www.ias.ac.in/jgenet/forthcoming/unedited/jgen-14-101-ue.pdf

Yuan Z., Gao S., Xue D.-W., Luo D., Li L.-T., Ding S.-Y., et al. (2009). RETARDED PALEA1 controls palea development and floral zygomorphy in rice. *Plant Physiol*. 149 235–244. 10.1104/pp.108.128231

Zawadzka B. J. (1976). Reaction of apple cultivars to infection by apple proliferation disease. *Acta Horticolturae* (67): 113-120.

Zharkikh, A., Troggio, M., Pruss, D., Cestaro, A., Eldrdge, G., Pindo, M., Mitchell, J. T., Vezzulli, S., Bhatnagar, S., Fontana, P., Viola, R., Gutin, A., Salamini, F., Skolnick, M., & Velasco, R. (2008). Sequencing and assembly of highly heterozygous genome of Vitis vinifera L. cv Pinot Noir: Problems and solutions. *Journal of Biotechnology*. https://doi.org/10.1016/j.jbiotec.2008.04.013

Zhou M., Li D., Li Z., Hu Q., Yang C., Zhu L., et al. (2013). Constitutive expression of a miR319 gene alters plant development and enhances salt and drought tolerance in transgenic creeping bentgrass. *Plant Physiol.* 161 1375–1391. 10.1104/pp.112.208702

Zimmermann M. R., Schneider B., Mithöfer A., Reichelt M., Seemüller E., Furch C. U. (2015). Implications of '*Candidatus* Phytoplasma mali' infection on phloem function of apple trees. *Endocytobiosis and Cell Research* 26: 67-75.

CHAPTER 2

RE-IDENTIFICATION OF *MALUS* × *DOMESTICA TCPs* WITHIN THE GDDH13v1.1 GENOME ASSEMBLY

INTRODUCTION

The TCP Classification

The huge advances in genome sequencing technologies in the last few years and the drastic decrease in sequencing costs have led to the generation of enormous amounts of primary and derived genomic data. The first complete genome of a plant, the model organism *Arabidopsis thaliana*, was published 20 years ago (The Arabidopsis Genome Initiative, 2000). The Malus × domestica genome was released ten years later by Velasco and co-authors (2010), being the tenth plant genome sequenced. Since then, genome sequences of many important plant organisms have been published, and more than 100 novel plant genomes have been published in the last two years only (https://www.plabipd.de/index.ep). The availability and ease of access to vast amounts of genomic sequences allowed the identification of several gene families for a broad range of plant species, including members of the TEOSINTE-BRANCHED1/ CYCLOIDEA/PROLIFERATING CELL FACTORS (TCP) transcription factor gene family. *TCP* genes have been identified for dozens of plant species, including *A. thaliana* (Riechmann *et al.,* 2000), *Oryza sativa* (Cubas *et al*., 1999), *Zea mays* (Doebley *et al*., 1997), and many Rosaceae, like *Fragaria vesca* (Wei *et al*., 2016), *Prunus persica* (Guo *et al*., 2018) and *M.* × *domestica* (Xu *et al*., 2014).

Usually, when describing *TCPs* in a species for the first time, a numbering system based on the homology of the sequences with the corresponding *A. thaliana TCP* (*AtTCP*) is adopted*,* but often *TCPs* are named after the relative order of the sequences on the chromosomes. The adoption of two different standards to classify newly discovered *TCP* sequences can lead to ambiguities and disorder: emblematic is the example of *Vitis vinifera*, for which two works aiming to identify *TCP* genes were published almost simultaneously by two different groups (Leng *et al*., 2019; Jiu *et al*., 2019). The two

nearly identical sets of genes were numbered using two different naming systems and, as a result, the same gene is indicated with two different numbers in the two articles. The same degree of ambiguity affects the gene annotation of *TCP*s in public databases, where identical nucleotide sequences are often indicated with two or even three different numbers. This ambiguity constitutes a source of confusion when comparing and contextualizing scientific results on the sequence and functional level. An effort must be made to establish a common standard to overcome equivocality and ambiguity, especially in light of the large number of TCP sequences that will be identified in new species in the future.

The *TCP* gene family in *M.* × *domestica*

In *M.* × *domestica,* the identification of *TCP* sequences (*MdTCP*s) dates back to a study from 2014 performed by Xu and collaborators. The authors distinguished 52 *MdTCPs* on the *M.* × *domestica* genome v1.0 (Velasco *et al*., 2010; http:/www.rosaceae.org), designating them from *MdTCP1* to *MdTCP52* based on their chromosomal locations. The 52 genes showed homology to 15 *AtTCPs* and were subdivided as follows: 22 in Class I, 26 in Class II-CINCINNATA (CIN) and 4 in Class II – CYCLOIDEA/TEOSINTE BRANCHED1 (CYC/TB1) subgroups. Furthermore, 12 sister pairs of paralogous *MdTCP* genes were identified, of which seven pairs linked to potential chromosomal duplications associated with a genome-wide duplication and two pairs tightly collocated in the apple genome. Therefore, it was proposed that the evolution of the *TCP* gene family in *M.* × *domestica* was shaped by segmental duplication and transposition events (Xu *et al*., 2014).

In 2017, a new assembly of *M.* × *domestica* genome, GDDH13, was released (Daccord *et al*., 2017), which serves now as the reference genome (Peace *et al*., 2019). This assembly is characterised by an overall higher quality, rendering it the most complete among apple genome sequences published until now, despite having the lowest number of predicted genes compared to previous assemblies.

This study aimed to perform a novel identification of *MdTCP* genes based on the latest published genome assembly GDDH13v1.1, followed by an analysis of the gene sequences to identify genes that originated from genome duplication events. Moreover,

a nomenclature system based on homology with *AtTCP* is proposed for the identified *MdTCP* genes.

RESULTS AND DISCUSSION

Revision and de novo identification of the *MdTCP* gene family

To ensure that the starting sequences retrieved from the work of Xu and colleagues (2014) were representative of the whole set of *MdTCP* genes, the 52 MdTCP sequences were aligned with all the putative TCP-containing sequences of *M.* × *domestica* collected from three databases: NCBI Reference database RefSeq (O' Leary *et al*., 2016), UniProt-based database Pfam (El-Gebali *et al*., 2019) and PlantTFDB (Jin *et al*., 2017). The alignment showed that the sequences retrieved from databases either correspond to the published *MdTCP* gene set (from Xu *et al*., 2014), are fragments of the same genes, or do not contain a TCP domain. This suggests that, to our knowledge, no novel unique *Malus* TCP-containing sequence has been identified outside the original set, which can be therefore considered as representative of the whole *MdTCP* gene family.

The 52 MdTCP sequences were then manually evaluated and aligned using MEGA version X (Kumar *et al*., 2018) and BioEdit software (Hall, 1999), highlighting that four couples (*MdTCP5*/*MdTCP6*, *MdTCP13*/*MdTCP14*, *MdTCP19*/*MdTCP20*, *MdTCP43*/*MdTCP44*) and one trio (*MdTCP8*/*MdTCP9*/*MdTCP10*) exhibit a reciprocal 100% identity. Interestingly, all those identical sequences have sequential names that reflect physical proximity on the chromosomes (Xu *et al*., 2014), suggesting that these sequences could originate from segmental duplication, as proposed by Xu and co-authors. However, the detection of complete identities could also indicate the generation of artefacts during the genome assembly process. The accessions of the 11 identical sequences were aligned to the *M.* × *domestica* double haploid genome GDDH13v1.1 (Daccord *et al*., 2017) (https://iris.angers.inra.fr/gddh13/) and each pair and trio resulted as reciprocally overlapping. This means that the segmental duplications that led to the identification of the 11 identical MdTCP sequences are missing in the most recent genome assembly GDDH13v1.1. The redundant sequences, one for each pair and two for the trio, were thus discarded (details of discarded sequence are reported in Table 1).

Another pair (*MdTCP3*/*MdTCP4*) differs for the first 153 and 174 nucleotides, respectively, but the two sequences were provisionally kept in the dataset for further analyses, given the inability to determine a complete identity.

The 46 non-redundant *MdTCP* sequences were aligned over the *M. × domestica* genome GDDH13v1.1 to obtain the corresponding accessions. A total of 33 of the starting MdTCP sequences appeared to match with predicted protein-coding genes in the GDDH13v1.1 genome, while 13 mapped onto intergenic regions or were partially overlapping non-TCP predicted gene regions. The corresponding nucleotide sequences of the 13 non-TCP mapped genes were therefore retrieved and verified for the presence of the TCP domain by screening against the Pfam (http://pfam.sanger.ac.uk/), and InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) databases: five deduced amino acid sequences (*MdTCP19*, *MdTCP36*, *MdTCP37*, *MdTCP41*, *MdTCP50*) display premature stop codons, two (*MdTCP7*, *MdTCP42*) constitute fragments of non-TCP gene sequences and were thus discarded. The two sequences previously observed to be nearly identical, *MdTCP3* and *MdTCP4*, resulted in being constituted of two exons each, of which the second completely overlapped among the two genes. Both deduced nucleotide sequences showed the presence of premature stop codons; thus, both were discarded. On the contrary, the remaining four sequences (*MdTCP2*, *MdTCP17*, *MdTCP23*, *MdTCP52*) contain a full-open reading frame (ORF) including a TCP domain and were included in the set. Overall, 15 out of the 52 starting sequences were discarded during the revision process. Details regarding discarded and confirmed sequences, including previous and updated accessions, are resumed in Table 1.

Table 1: *MdTCP* genes identified by Xu *et al.* (2014) and relative accession numbers determined on the 2010 (Velasco *et al.*, 2010) and most recent genome assembly GDDH13v1.1 (Daccord *et al.*, 2017). Six sequences were observed to be redundant and were thus discarded (indicated in gray). A new GDDH13v1.1 accessions was determined for 33 *MdTCPs*, while the 13 remaining sequences did not show a corresponding accession. Discarded sequences are indicated in red.

GENE NUMBER (Xu *et al.*, 2014)	ACCESSION NUMBER (Velasco *et al.*, 2010)	ACCESSION NUMBER (GDDH13 - Daccord *et al.*, 2017)
MdTCP1	MDP0000123919	MD01G1194100
MdTCP2	MDP0000594000	/
MdTCP3	MDP0000681033	/
MdTCP4	MDP0000393985	/
MdTCP5	MDP0000182310	MD05G1305100
MdTCP6	MDP0000259723	/
MdTCP7	MDP0000534647	/
MdTCP8	MDP0000927314	MD10G1259500
MdTCP9	MDP0000920127	/
MdTCP10	MDP0000763497	/
MdTCP11	MDP0000287069	MD05G1281100
MdTCP12	MDP0000280252	MD06G1070100
MdTCP13	MDP0000877369	/
MdTCP14	MDP0000531313	MD06G1191800
MdTCP15	MDP0000120671	MD06G1226300
MdTCP16	MDP0000219838	MD06G1211100
MdTCP17	MDP0000199422	/
MdTCP18	MDP0000260056	MD00G1004900
MdTCP19	MDP0000617746	/
MdTCP20	MDP0000253526	/
MdTCP21	MDP0000319266	MD08G1240000
MdTCP22	MDP0000523096	MD09G1008300
MdTCP23	MDP0000130524	/
MdTCP24	MDP0000692406	MD09G1068200
MdTCP25	MDP0000442611	MD09G1232700
MdTCP26	MDP0000264920	MD04G1069300
MdTCP27	MDP0000184743	MD03G1239100
MdTCP28	MDP0000238683	MD13G1238400

MdTCP29	MDP0000189749	MD10G1284400
MdTCP30	MDP0000243495	MD11G1258900
MdTCP31	MDP0000139807	MD12G1115100
MdTCP32	MDP0000535805	MD02G1196100
MdTCP33	MDP0000173048	MD13G1047200
MdTCP34	MDP0000242185	MD13G1073500
MdTCP35	MDP0000374900	MD13G1122900
MdTCP36	MDP0000202241	/
MdTCP37	MDP0000693146	/
MdTCP38	MDP0000210785	MD14G1198200
MdTCP39	MDP0000155433	MD14G1213400
MdTCP40	MDP0000224810	MD14G1221800
MdTCP41	MDP0000617459	/
MdTCP42	MDP0000247249	/
MdTCP43	MDP0000515080	MD15G1431700
MdTCP44	MDP0000608645	/
MdTCP45	MDP0000272980	MD16G1049000
MdTCP46	MDP0000319941	MD16G1074800
MdTCP47	MDP0000915616	MD17G1002500
MdTCP48	MDP0000320363	MD17G1061100
MdTCP49	MDP0000916623	MD17G1233500
MdTCP50	MDP0000149841	/
MdTCP51	MDP0000851695	MD16G1243300
MdTCP52	MDP0000373350	/

Furthermore, the GDDH13v1.1 functional annotated gene list was scanned to confirm the presence of all the *MdTCP* genes and to check for novel sequences, and a total of 33 was found. Of the 33 predicted *MdTCP* genes, 30 correspond to sequences already included in the previously revised MdTCP set, while three appear to be novel sequences. The three novel sequences were manually confirmed to be carrying a start and stop codon, and a TCP-containing deduced coding sequence. The sequences were provisionally named *MdTCPnc1*, *MdTCPnc2* and *MdTCPnc3*, and added to the set. The alignment of the deduced amino acid sequences of the three novel *MdTCPs*, displayed

in Figure 1, shows that the three sequences are highly similar. In particular, *MdTCPnc2 and MdTCPnc3* are shorter and similar to *MdTCPnc1*, which is the longest. *MdTCPnc2* aligns with the 5' of *MdTCPnc1* (positions 1-265), while *MdTCPnc3* aligns with the 3' end of *MdTCPnc1* (position 140-393). Consequently, the two shorter sequences share similarity across 120 amino acids, which constitute the longer sequence's central region. Moreover, the two shorter sequences are mapped on physically close regions of the same chromosome, indicating a possible origin by segmental duplication as previously proposed for other *MdTCP* genes (Xu *et al*., 2014).

Three MdTCP sequences (*MdTCP33*, *MdTCP40*, *MdTCP45*) that were previously observed to match with protein coding genes are not present in the mRNA set, despite being functionally annotated as TCP domain-containing proteins.

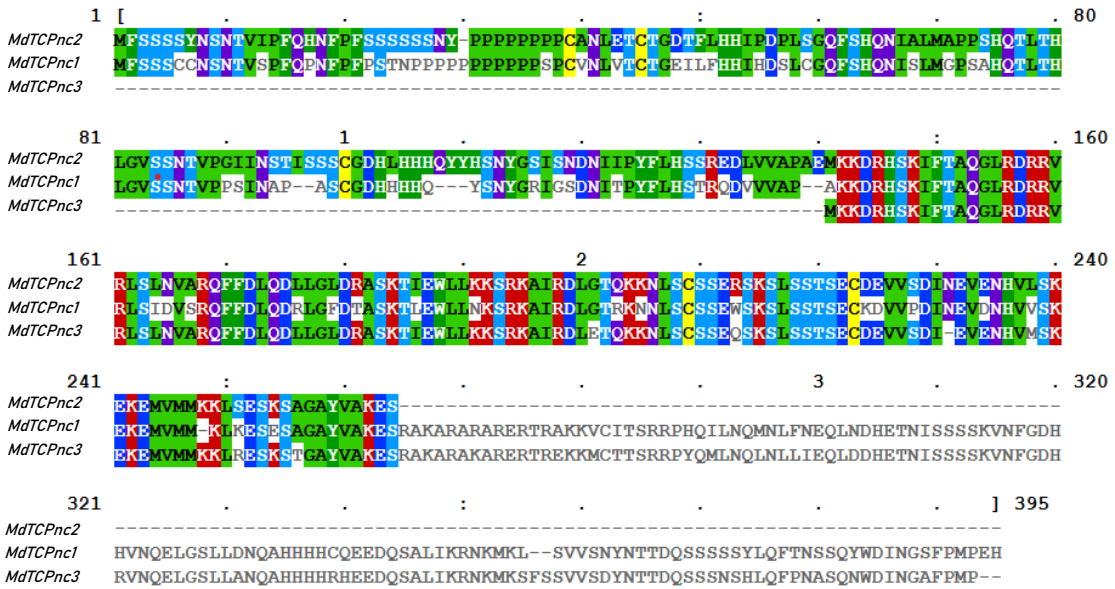

Figure 1 – Alignment of the deduced amino acid sequences of three novel MdTCP sequences, provisionally named *MdTCPnc1*, *MdTCPnc2* and *MdTCPnc3*. The alignment shows that the three sequences share a portion of 125 AA displaying high similarity (positions 140 to 265, highlighted with colours). Moreover, *MdTCPnc2* and *MdTCPnc3* align on the 5' and 3' fragments of *MdTCPnc1,* respectively. Alignments were performed and visualized with Clustal Omega Multiple Sequence Alignment tool (Madeira *et al*., 2019).

To further confirm the absence of supplementary non-identified TCP-containing sequences in the GDDH13v1.1 genome, all the mRNA sequences identified in GDDH13v1.1 displaying similarity to *AtTCP* genes were retrieved from the list 'Malus_x_domestica_GDDH13_v1.1_vs_TAIR.xlsx' (tp.bioinfo.wsu.edu). This list includes the whole set of GDDH13v1.1 mRNA sequences and the relative homologous genes of *A. thaliana*, deduced as the best hit of a BLAST (Basic Local Alignment Search Tool) query of each *M.* × *domestica* sequence over the *A. thaliana* gene database "The Arabidopsis Information Resource" (TAIR; www.arabidopsis.org/aboutarabidopsis). Overall, 14 *AtTCP* genes resulted as the best hits for the 33 previously identified *MdTCP* genes, while no *M.* × *domestica* mRNA matched with the remaining 10 *AtTCP*s. No additional mRNA sequences were found to be displaying similarity with any of the *AtTCP* genes, thus confirming that no other putative *MdTCP* was identified outside our dataset.

Identification of potential sister genes among the *MdTCP* gene family

To identify the potential sister paralogous genes in the MdTCP family, the reciprocal identity values were calculated for the whole set by counting the proportion of invariant sites for each pair of deduced amino acid sequences. To determine the threshold to define two genes as highly similar, and thus as potential sister genes, the intrinsic identity among AtTCPs was observed. Most of the AtTCPs have a low reciprocal percentage of identity (mean = 12%), while eight couples of AtTCPs display a consistent identity (> 33%), with a mean of 47% and the highest observed value of 62%. These two values were thus selected as thresholds to define the MdTCP sequences displaying moderate (47%<x≤62%) and high (>62%) identity.

The identity matrix calculated on the deduced amino acid sequences of the 40 MdTCPs using BioEdit (Hall, 1999) is graphically represented in figure 2, where identities lower than the threshold are indicated as blue squares, moderate as yellow, and high as red squares, respectively. Results indicate that 14 pairs of MdTCPs share a high identity rate, and one pair shows a discrete degree of identity. One trio exhibits a moderate reciprocal identity: the three sequences are those retrieved from the GDDH13 mRNA sequences list, in which *MdTCPnc2* and *MdTCPnc3* appear partially overlapping. These two sequences do not show a reciprocal identity percentage higher than the threshold, but

both show a very similar identity value with the third sequence, *MdTCPnc1*. Four sequences (*MdTCP2*, *MdTCP17*, *MdTCP23* and *MdTCP52*) appear to constitute a larger group with a moderate-high degree of similarity (with a mean identity of 62%) in which one sequence displays high identity with two others, making the classification of this cluster challenging to disentangle. Finally, three sequences do not display a moderate or high level of identity with any other *MdTCP*.

These results suggest that, overall, 15 pairs of *MdTCPs* can be considered potential sister genes originated through the whole genome duplication event. Moreover, hints of a recent duplication appear also in the trio *MdTCPnc1*, *MdTCPnc2* and *MdTCPnc3* and the cluster of four (*MdTCP2*, *MdTCP17*, *MdTCP23* and *MdTCP52*).

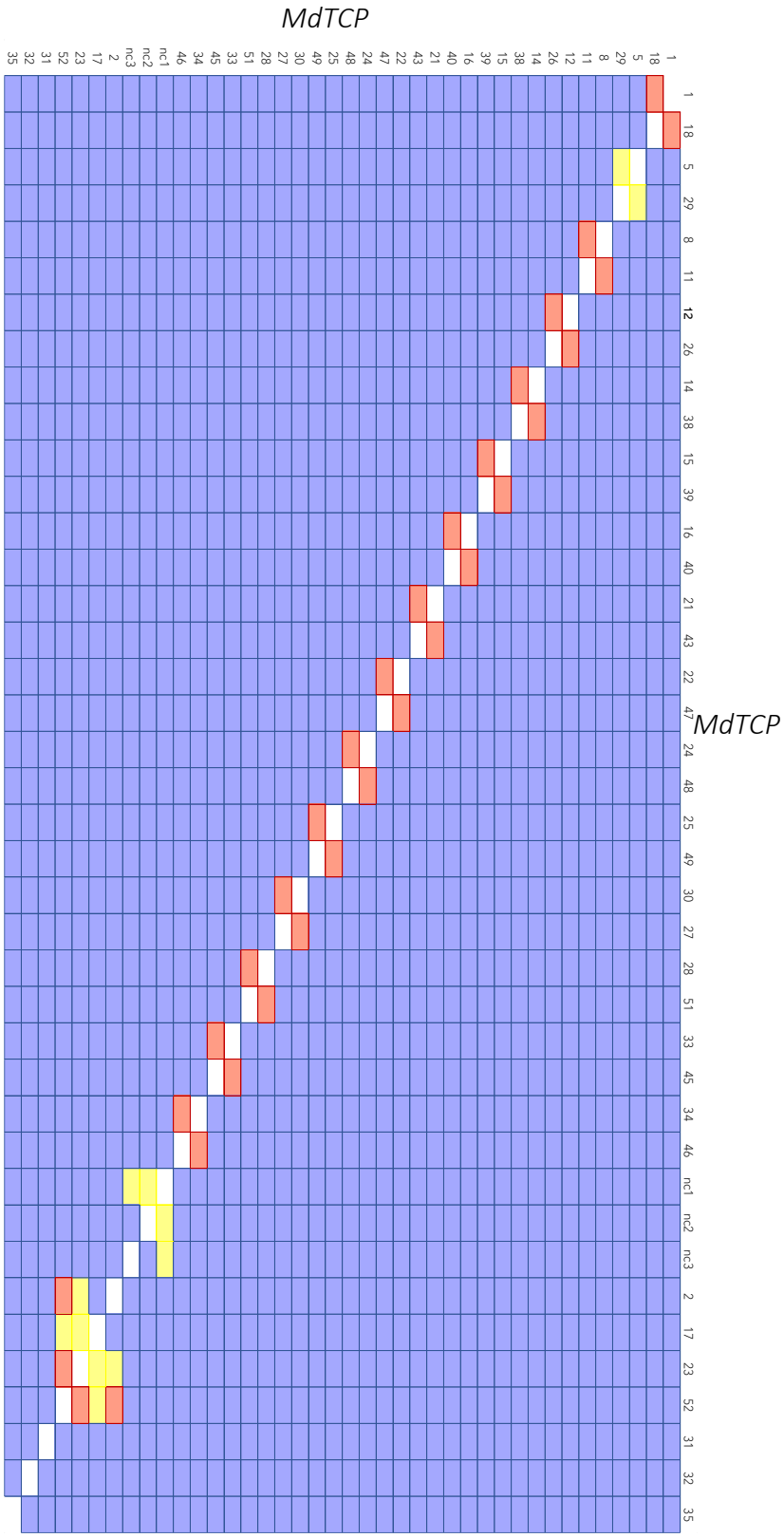

Figure 2 – Graphical representation of the identity matrix calculated on the amino acid sequences of the 40 *MdTCPs* identified. Numbers identifying each *MdTCP* are reported to the left of each row and on top of each column. The color of the square indicates the percentage of identity calculated for each pair: blue (0% – 46,9%) indicates low identity, yellow (47% - 62%) moderate and red (62,1% - 100%) high identity, respectively.

Data regarding each potential duplicated pair has been compared with intragenomic synteny determined on the GDDH13 assembly (Daccord *et al.*, 2017) to confirm the hypothesis that the *MdTCPs* originated from whole-genome duplication events. Data were computed on Galaxy servers (Afgan *et al*., 2016) and visualized using the Circos tool (Krzywinski *et al.*, 2009), as shown in figure 3. In panel A the link between two potential duplicated genes is represented as a line connecting the physical localization of the two genes on the chromosomes. In panel B, the pattern of apple genome duplication as determined on GDDH13 assembly by Daccord *et al.* (2017) is shown as coloured ribbons connecting the homologous regions between the chromosomes. Two genes (*MdTCP2* and *MdTCP18*) are mapped on the pseudochromosome 0, meaning they were detected on contigs not located on any of the 17 chromosomes. Thus, the links involving two genes could not be represented.

Figure 3 shows that the hypothesis of duplication of the 14 pairs of genes represented is consistent with the results of the synteny analysis. Regarding the trio, the two sequences *MdTCPnc2* and *MdTCPnc3* are located on the same chromosome (9), and both links with *MdTCPnc1* are consistent with the duplication pattern (chromosome 9 – chromosome 17). Given the pattern observed, it is possible to hypothesize that one of the two genes on chromosome 9 originated via segmental duplication after the whole genome duplication event, with secondary losses of the 3' and 5' ends in the two genes. As expected, the definition of the evolutionary relationship in the cluster of four genes is elusive and furthermore complicated by the fact that one of the genes is mapped on the pseudochromosome 0. The link between *MdTCP23* and *MdTCP52* (chromosome 12 – chromosome 13) is supported, while the two links involving *MdTCP17* (chromosome 7 – chromosome 12 / chromosome 7 – chromosome 13) are not (data not shown). Given the high similarity observed between the four sequences and the ambiguity of these results, the four genes *MdTCP2*, *MdTCP17*, *MdTCP23* and *MdTCP52* will not be considered sister genes.

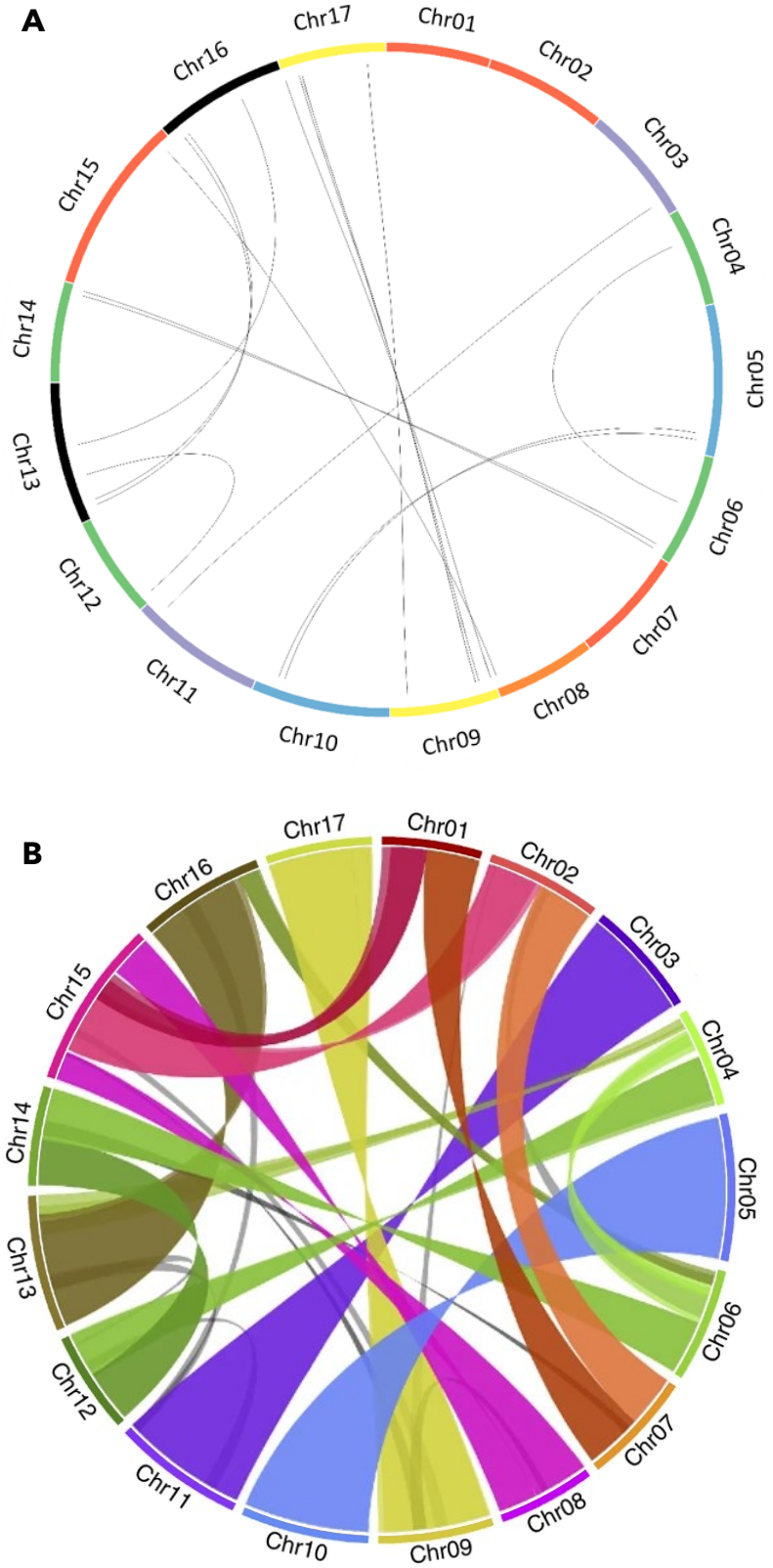

Figure 3 – Graphical representation of the duplication pattern observed in MdTCP genes (A) **compared to pattern of chromosome duplication observed in Apple genome** (B**,** modified from Daccord *et al*., 2017**).** The 17 chromosomes are indicated as colored boxes and the link between potential duplicated regions are indicated as black lines (*MdTCP*, A) or colored ribbons (whole genomic duplication pattern, B).

A novel *MdTCP*s nomenclature system based on the homology with *AtTCP*s

The *A. thaliana* homolog for each sequence was estimated through a BLAST query of each nucleotide sequence over the *A. thaliana* gene database to determine the best hits for each gene. Results were then manually validated by an investigation of alignments, and the genes were named accordingly, appending the letter "a" and "b" to indicate sister genes in accordance with the previous analyses. In one case, three *MdTCP* sequences (*MdTCP12*, *MdTCP26,* and *MdTCP32*) showed homology to the same *A. thaliana* sequence (*AtTCP9*), although only two showed a high reciprocal identity and are therefore considered to be recently duplicated genes. To highlight the similarity with *AtTCP9* and, at the same time, to distinguish it from the two likely duplicated genes, the two sister genes were named *MdTCP9a*/*b* and the third *MdTCP9-like*. Similarly, the cluster of four genes *MdTCP2*, *MdTCP17*, *MdTCP23,* and *MdTCP52*, whose relationships are not exhaustively explained by whole-genome duplication origin, showed similarity with the same *A. thaliana* gene (*AtTCP17*): the four genes were therefore named *MdTCP17-like_a*/*b*/*c*/*d*. The three novel genes (*MdTCPnc1*, *MdTCPnc2*, *MdTCPnc3*) show homology to *AtTCP1* and constitute an identity trio for which the genome duplication hypothesis partially explains the relationship between a gene and the other two, thus were named *MdTCP1a*/*b*/*c*.

Nineteen *A*tTCP sequences in total were identified to be homologs for the whole set of 40 *MdTCPs*. A corresponding *Malus* homolog was not found for *AtTP11*, *AtTCP16*, *AtTCP22*, *AtTCP23* and *AtTCP24*. The complete revised set of *MdTCP* genes (details listed in table 2) comprises 40 unique sequences, of which 33 were deduced from the first assembly in 2010 (Velasco *et al.*, 2010) and the GDDH13v1.1 assembly, four deduced from the first assembly but not from the GDDH13v1.1 assembly, and three additional sequences that have not been predicted before. The division of *MdTCPs* in the two classes and subclasses is visible in the Neighbour Joining phylogenetic tree displayed in figure 4: 17 genes belong to TCP Class I (22 in the previous classification by Xu and co-authors), 16 to the subclass CIN of Class II (26 in the previous classification) and seven to the Class II subclass CYC/TB1 (four in the previous classification). In *A. thaliana,* Class I contains 13 genes, Class II-CIN eight and Class II-TB1/CYC three, respectively.

Interestingly, two homolog sister *MdTCP*s have been found for each Class II *AtTCP* except *AtTCP24*, while several members of Class I appear to have only one or no homolog in *Malus*. On the one hand, these results indicate a significant role of genome-wide duplication in the generation of the *MdTCP* set; on the other, the lack of some sister and *AtTCP* homologous genes may suggest that additional members of the *MdTCP* gene family, and especially Class I *MdTCP*s, have not been identified in the genome yet.

In the Maximum Likelihood dendrogram in figure 5, which includes *MdTCP* and *AtTCP* sequences, a general consistency with the *MdTCP* - *AtTCP* homologies determined can be observed. The *M.* × *domestica* and the corresponding *A. thaliana* homologs appear to group together. Even though the phylogenetic tree is unrooted, by implicitly rooting the tree at the split between Class I and Class II TCP, some inferences about the relationships among genes can be made. While Class I and subclass CIN of Class II appear to be monophyletic, the relationships among Class II subclass CYC/TB1 are not well defined. In particular, *AtTCP18*, *MdTCP18a* and *MdTCP18b* are clearly separated from the other Class II sequences, and, in general, the CYC/TB1 subgroup appears paraphyletic and characterized by the presence of many polytomies. This indicates that the alignment of the CYC/TB1 subclass sequences is too ambiguous to allow a confident reconstruction of the phylogenetic relationship in this group. In general, the tree is characterized by many short branches, especially inside subgroups (data not shown), reflecting the fact that closely related genes display a high degree of similarity.

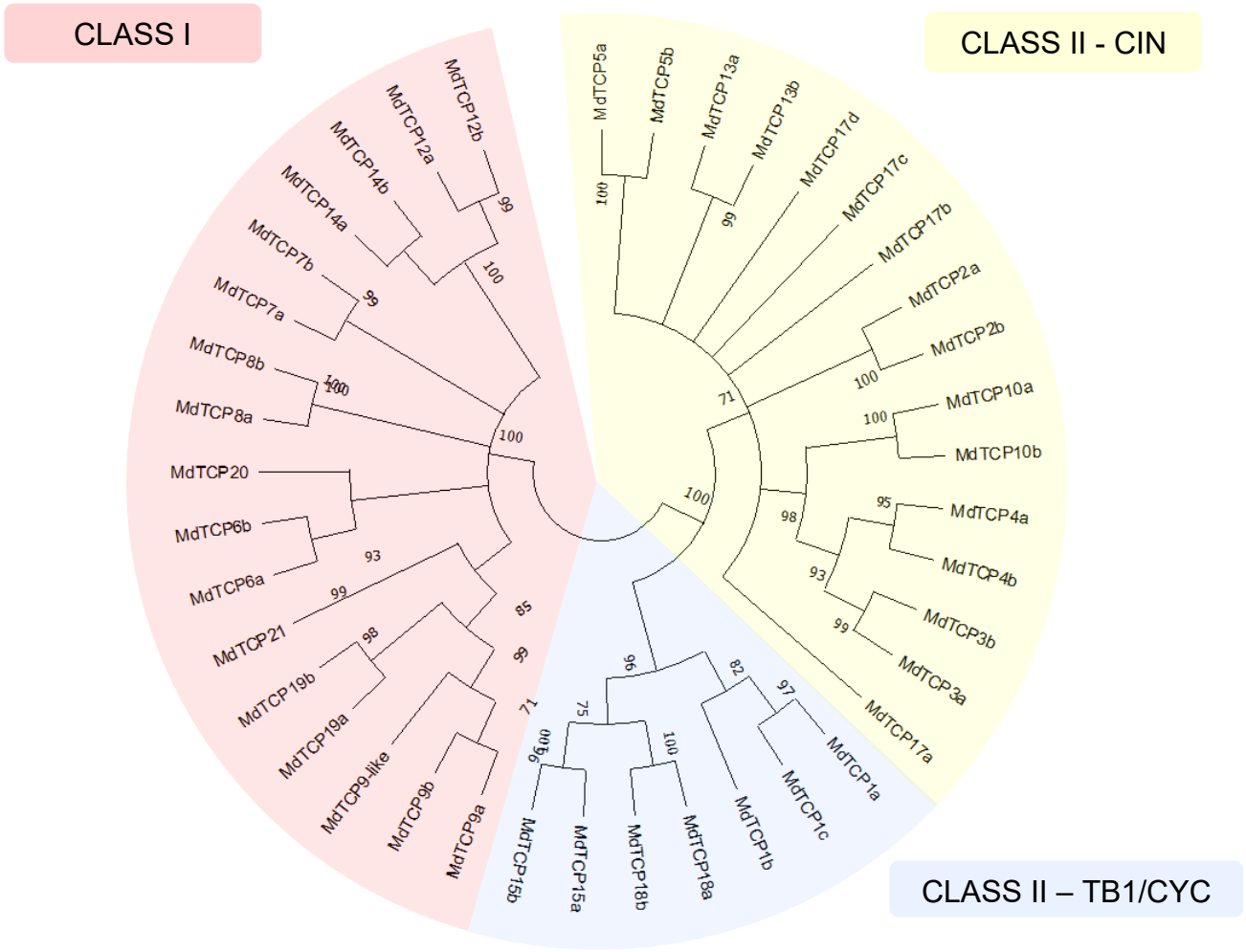

Figure 4 – Phylogenetic tree calculated on the *40 MdTCP* genes classified in the present chapter. The different classes of TCPs are indicated with red (Class I), yellow (Class II - CIN) and blue (Class II – TB1/CYC). The tree was inferred using the Neighbor-Joining method (Saitou and Nei, 1987), with 1000 bootstrap reiterations. Node support is indicated as bootstrap values and nodes with a support lower than 50 are collapsed. Evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution. Analyses were conducted in MEGA X Software (Kumar *et al.,* 2018*)*.

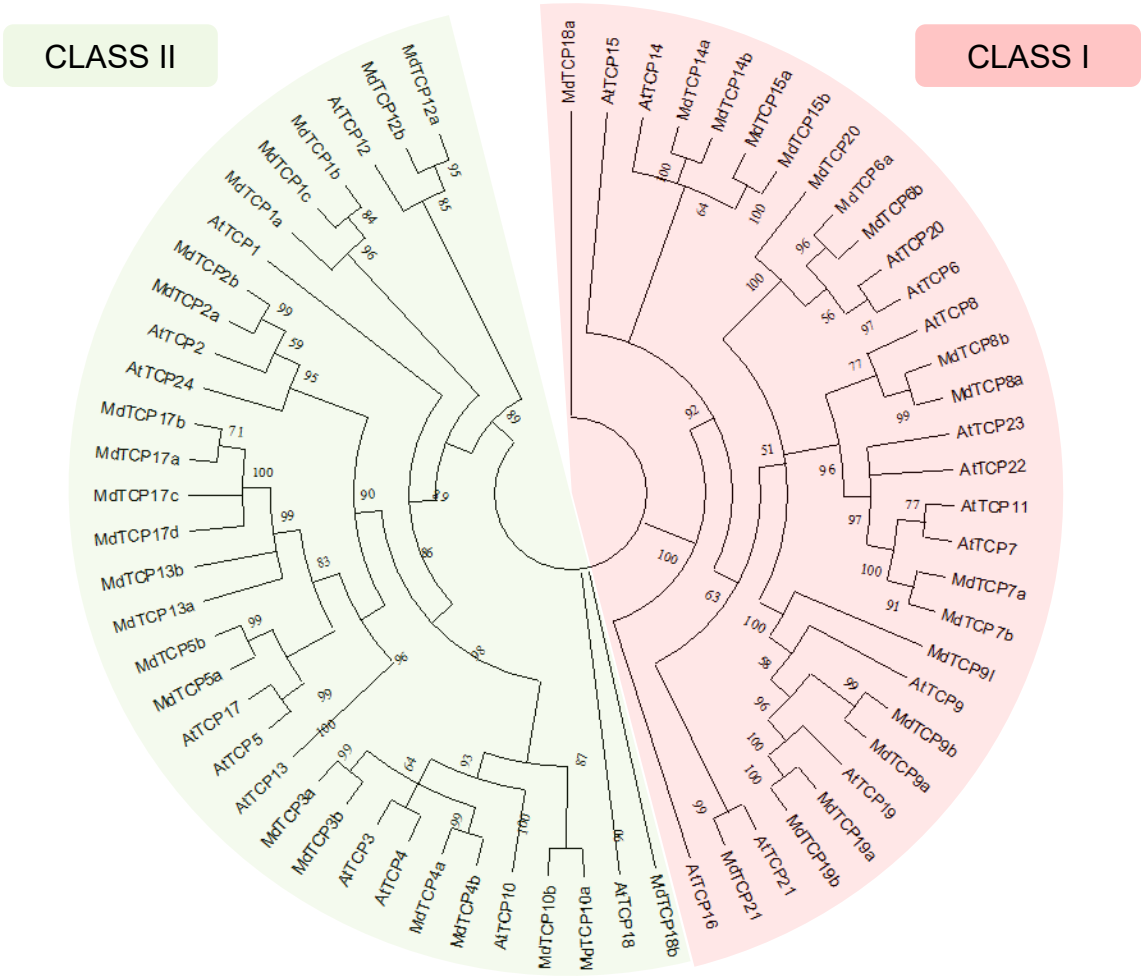

Figure 5 – Phylogenetic tree calculated on the amino acid sequences of the 40 *MdTCP* genes identified in the present chapter and 24 *AtTCPs*. Members of the Class I are indicated with a red background, while the Class II with a green background. The tree was inferred using the Maximum Likelihood method and JTT+G+I+F matrix-based model (Jones *et al*., 1992). Node support is indicated as bootstrap values, calculated on 100 reiterations. Nodes with a support lower than 50 are collapsed. The tree with the highest log likelihood (-38172.20945) is shown. Initial trees for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (four categories). Analyses were conducted in PhyML software (Guindon *et al*., 2010).

Table 2: Details of the 40 *MdTCP* sequences classified in the present work. The table lists the name of the genes determined on the homology with *AtTCPs* and the corresponding accession number and gene name as described by Xu and co-authors (2014). Information regarding chromosome location, numbers of exon and length of the coding sequence (CDS) for each gene are included

GENE NAME (AtTCP HOMOLOGY)	CHROMOSOME LOCATION	GDDH13v1.1 ACCESSION	GENE NAME (Xu *et al.*, 2014)
MdTCP1a	Chr17:5918494..5919692	MD17G1073600	/
MdTCP1b	Chr09:5829036..5829824	MD09G1083300	/
MdTCP1c	Chr09..5866332..5867581	MD09G1083500	/
MdTCP2a	Chr10:35357675..35359117	MD10G1259500	*MdTCP8*
MdTCP2b	Chr05:41511031..41512488	MD05G1281100	*MdTCP11*
MdTCP3a	Chr03:32436450..32434382	MD03G1239100	*MdTCP27*
MdTCP3b	Chr11:37242949..37239803	MD11G1258900	*MdTCP30*
MdTCP4a	Chr09:29081922..29080873	MD09G1232700	*MdTCP25*
MdTCP4b	Chr17:28210574..28211638	MD17G1233500	*MdTCP49*
MdTCP5a	Chr06:35690088..35691110	MD06G1226300	*MdTCP15*
MdTCP5b	Chr14:29783260..29784375	MD14G1213400	*MdTCP39*
MdTCP6a	Chr09: 581190..580225	MD09G1008300	*MdTCP22*
MdTCP6b	Chr17:174117..175064	MD17G1002500	*MdTCP47*
MdTCP7a	Chr13:24221252..24220386	MD13G1238400	*MdTCP28*
MdTCP7b	Chr16:26389521..26388679	MD16G1243300	*MdTCP51*
MdTCP8a	Chr08:30627261..30630168	MD08G1240000	*MdTCP21*
MdTCP8b	Chr15:53149356..53150897	MD15G1431700	*MdTCP43*
MdTCP9a	Chr06:16974531..16975661	MD06G1070100	*MdTCP12*
MdTCP9b	Chr04:9490696..9491838	MD04G1069300	*MdTCP26*
MdTCP9-like	Chr02:18841493..18842398	MD02G1196100	*MdTCP32*
MdTCP10a	Chr05:43753231..43753953	MD05G1305100	*MdTCP5*
MdTCP10b	Chr10:37444923..37446820	MD10G1284400	*MdTCP29*
MdTCP12a	Chr13:3317170..3318603	MD13G1047200	*MdTCP33*
MdTCP12b	Chr16:3451626..3453011	MD16G1049000	*MdTCP45*
MdTCP13a	Chr09:4658466..4659617	MD09G1068200	*MdTCP24*
MdTCP13b	Chr17:4957596..4958738	MD17G1061100	*MdTCP48*
MdTCP14a	Chr06:32754201..32752924	MD06G1191800	*MdTCP14*
MdTCP14b	Chr14:28828139..28826856	MD14G1198200	*MdTCP38*
MdTCP15a	Chr13:5176606..5177805	MD13G1073500	*MdTCP34*
MdTCP15b	Chr16:5227692..5228900	MD16G1074800	*MdTCP46*

MdTCP17-like_a	Chr13:13829138..13829581	/	*MdTCP23*
MdTCP17-like_b	Chr12: 2031509..2031066	/	*MdTCP52*
MdTCP17-like_c	Chr00:20520105..20525532	/	*MdTCP2*
MdTCP17-like_d	Chr07:8098897..8101110	/	*MdTCP17*
MdTCP18a	Chr06:34383370..34381464	MD06G1211100	*MdTCP16*
MdTCP18b	Chr14:30344698..30343026	MD14G1221800	*MdTCP40*
MdTCP19a	Chr01:29290244..29291342	MD01G1194100	*MdTCP1*
MdTCP19b	Chr00:735807..736895	MD00G1004900	*MdTCP18*
MdTCP20	Chr13:9108413..9109165	MD13G1122900	*MdTCP35*
MdTCP21	Chr12:18364567..18365235	MD12G1115100	*MdTCP31*

MATERIALS AND METHODS

Revision and *de novo* identification of the *MdTCP* gene family

Each of the 52 amino acid MdTCP sequences identified by Xu *et al*. (2014) was used to query a tblastn search on the RefSeq database (O' Leary *et al*., 2016). Hits were analysed to eliminate sequences already annotated as non-TCP and redundancies (sharing more than 99% identity). Additional sequences were obtained as TCP-containing proteins on Pfam (El-Gebali *et al*., 2019) and PlantTFDB (Jin *et al*., 2017) databases. The whole set of sequences was aligned on MEGA X software (Kumar *et al*., 2018) with the Clustal Omega Multiple Alignment Tool (Madeira *et al*., 2019), with the following parameters: a gap opening penalty of 5.00 and a gap extension penalty of 0.1 for the pairwise alignments; a gap opening penalty of 10.00 and a gap penalty of 0.05 for multiple alignments; BLOSUM substitution weight matrix; sequences with a per cent identity lower than 30% were aligned later (delayed divergent cutoff: 30%). Using each of the 52 sequences as indices of the alignment, sequences that shared more than 99% amino acids were discarded.

The 52 MdTCP GDR accessions, identified by Xu *et al.* (2014), were used as queries to identify the gene positions on the GDDH13v1.1 assembly (Daccord *et al*., 2017; https://iris.angers.inra.fr/gddh13/). Subsequently, each of the 46 non-redundant nucleotide sequences was used as a query for a Blast search on GDDH13v1.1 to confirm

the previous results. The GDDH13v1.1 nucleotide sequences of the 46 putative *MdTCP* genes were obtained and used for further analyses.

The MdTCPs mapping on non-TCP or intergenic regions were examined for the presence of an ORF and a TCP domain on the Pfam (http://pfam.sanger.ac.uk/), and InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/) databases.

Novel putative MdTCP sequences were obtained by scanning the GDDH13v1.1 mRNA list and characterised as previously described.

Identification of potential sister genes among the *MdTCP* gene family

The amino acid TCP sequences of *A. thaliana* and the putative MdTCP amino acid sequences previously obtained were aligned on MEGA X with Clustal Omega Tool with the previously described parameters. The alignment was then manually examined and edited if needed. Reciprocal identities were calculated on BioEdit software (Hall, 1999). To build the Circos Plot, a bed file containing the positions of the putative sister genes was manually produced, while the Karyotype file of the GDDH13v1.1 assembly was downloaded from the GDDH13 database. The two files were used as input for the Circos Plot Tool (Krzywinski *et al*., 2009) on Galaxy servers (Afgan *et al*., 2016; https://usegalaxy.eu/).

Identification of *AtTCP* homologs and construction of phylogenetic trees

The *A. thaliana* homolog for each of the putative *MdTCP* was inferred based on sequence similarity. Each MdTCP amino acid sequence was used as a query for a BLAST search restricted to *A. thaliana*. The query cutoffs were set as: query cover > 75% and Evalue < 1^ (-50). The same Blast search was performed using AtTCP sequences as queries, restricted to *M.* × *domestica.* The 40 MdTCP sequences were named according to the best Blast hit.

The MdTCP Neighbour-Joining (Saitou and Nei, 1987) phylogenetic tree was inferred on amino acid sequences with 1000 bootstrap reiterations. Evolutionary distances were computed using the Poisson correction method (Zuckerkandl and Pauling, 1965) in units of number of amino acid substitutions per site. The rate variation among sites was modelled with a gamma distribution, a Gamma parameter of 1.00 and a homogeneous pattern. All positions with less than 60% site coverage were eliminated, i.e., fewer than

40% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). Analyses were conducted in MEGA X software (Kumar *et al*., 2018).

To calculate the Maximum Likelihood tree on *A. thaliana* and *M.* × *domestica TCP*s, the amino acid sequences of 40 MdTCPs and 24 AtTCPs were aligned on MEGA X as described above. The alignment was manually inspected and edited. To calculate the appropriate evolution model, the Smart Model Selection algorithm (Lefort *et al*., 2017) with Akaike Information Criterion (AIC) was applied. The Maximum Likelihood phylogenetic tree was inferred in PhyML 3.0 software (Guindon et al., 2010) with Jones-Taylor-Thornton substitution model with a discrete Gamma, plus a proportion of invariable site, plus empirical amino acid frequencies count (JTT+G+I+F). Initial trees for the heuristic search were obtained by automatically applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model and then selecting the topology with the superior log-likelihood value.

CONCLUSIONS

In the present work, the *MdTCP* gene family has been identified from the latest *M.* × *domestica* genome assembly GDDH13v1.1. Compared to the previous classification, 15 sequences have been discarded (6 because of redundancy with other sequences, nine because of the absence of a TCP-containing ORF), while three novel sequences have been added, for a total of 40 sequences now included in the *MdTCP* gene family. By analysing the intragenic similarity and comparing it with the synteny observed within the *M.* × *domestica* genome, an origin of whole-genome duplication for 14 pairs of *MdTCP*s is proposed and is likely to explain the origin of three more pairs. This suggests that genome duplication had a significant role in the generation of the *TCP* set in *M.* × *domestica*, while segmental duplication had probably a minor effect.

Furthermore, in this work, 19 *AtTCP*s were identified as homologs for the 40 *MdTCP*s genes. For comparison, in the classification carried out on the previous version of the *M.* × *domestica* genome assembly, only 15 *AtTCP*s resulted as homologs for the whole set of *MdTCP* genes. Nonetheless, despite the higher number of *AtTCP* homolog genes found in the present work, no correspondent *MdTCP* was found for any of the five

remaining *AtTCP*s: this could suggest the presence of additional MdTCP sequences yet to be discovered in the future.

The *MdTCP* nomenclature was previously based on the relative position of the sequences on pseudochromosomes, while in the present study the identified genes have been renamed after the homology with the *AtTCP* equivalents. In our view, such classification will facilitate future research by reducing the ambiguities associated with a random numbering of TCPs in different species.

In the future, the presence of the genes identified in this work will need to be confirmed on the *M.* × *domestica* transcriptome, which could also shed light on the potential redundant roles of highly similar *MdTCPs*.

REFERENCES

Cubas, P., Lauter, N., Doebley, J., & Coen, E. (1999). The TCP domain: A motif found in proteins regulating plant growth and development. *Plant Journal*. https://doi.org/10.1046/j.1365-313X.1999.00444.x

Daccord, N., Celton, J. M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., Van De Geest, H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E. A., Gouzy, J., Rees, D. J. G., Guérif, P., Muranty, H., Durel, C. E., Laurens, F., Lespinasse, Y., Gaillard, S., … Bucher, E. (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*. https://doi.org/10.1038/ng.3886

Doebley, J., Stec, A., & Hubbard, L. (1997). The evolution of apical dominance in maize. *Nature*. https://doi.org/10.1038/386485a0

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky995

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3):307-21, 2010.

Guo, Z. H., Shu, W. S., Cheng, H. Y., Wang, G. M., Qi, K. J., Zhang, S. L., & Gu, C. (2018). Expression Analysis of TCP Genes in Peach Reveals an Involvement of PpTCP.A2 in Ethylene Biosynthesis During Fruit Ripening. *Plant Molecular Biology Reporter*. https://doi.org/10.1007/s11105-018-1105-z

Hall, T. A. (1999). BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. *Nucleic Acids Symposium Series*.

Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw982

Jiu, S., Xu, Y., Wang, J., Wang, L., Wang, S., Ma, C., Guan, L., Abdullah, M., Zhao, M., Xu, W., Ma, W., & Zhang, C. (2019). Genome-Wide Identification, Characterization, and Transcript Analysis of the TCP Transcription Factors in Vitis vinifera. *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2019.01276

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/8.3.275

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msy096

Lefort, V., Longueville, J.E., Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, 34(9):2422-2424

Leng, X., Wei, H., Xu, X., Ghuge, S. A., Jia, D., Liu, G., Wang, Y., & Yuan, Y. (2019). Genome-wide identification and transcript analysis of TCP transcription factors in grapevine. *BMC Genomics*. https://doi.org/10.1186/s12864-019-6159-2

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv1189

Peace, C. P., Bianco, L., Troggio, M., van de Weg, E., Howard, N. P., Cornille, A., Durel, C. E., Myles, S., Migicovsky, Z., Schaffer, R. J., Costes, E., Fazio, G., Yamane, H., van Nocker, S., Gottschalk, C., Costa, F., Chagné, D., Zhang, X., Patocchi, A., … Vanderzande, S. (2019). Apple whole genome sequences: recent advances and new prospects. In *Horticulture Research*. https://doi.org/10.1038/s41438-019-0141-7

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z.,

Ghandehari, D., Sherman, B. K., & Yu, G. L. (2000). Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*. https://doi.org/10.1126/science.290.5499.2105

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. https://doi.org/10.1093/oxfordjournals.molbev.a040454

The Arabidopsis Genome Initiative. (2000). Analysis of the genome sequence of Arabidopsis thaliana. *Nature*.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S. K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., … Viola, R. (2010). The genome of the domesticated apple (Malus × domestica Borkh.). *Nature Genetics*. https://doi.org/10.1038/ng.654

Wei, W., Hu, Y., Cui, M. Y., Han, Y. T., Gao, K., & Feng, J. Y. (2016). Identification and transcript analysis of the TCP transcription factors in the diploid woodland strawberry Fragaria vesca. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2016.01937

Xu, R., Sun, P., Jia, F., Lu, L., Li, Y., Zhang, S., & Huang, J. (2014). Genomewide analysis of TCP transcription factor gene family in Malus domestica. *Journal of Genetics*. https://doi.org/10.1007/s12041-014-0446-0

Zuckerkandl, E., & Pauling, L. (1965). Evolutionary Divergence and Convergence in Proteins. In *Evolving Genes and Proteins*. https://doi.org/10.1016/b978-1-4832-2734-4.50017-6

CHAPTER 3

THE MOLECULAR TARGETS OF '*Ca*. P. mali' IN APPLE PROLIFERATION SUSCEPTIBLE AND TOLERANT *MALUS* GENOTYPES

INTRODUCTION

Apple Proliferation (AP) is a severe disease widespread in apple-growing areas in Europe. Typical symptoms comprise foliar reddening, shoot proliferation, small leaves with altered shape, and undersized, tasteless and colourless fruits. The bacteria associated with the disease is the phytoplasma '*Candidatus* Phytoplasma mali' ('*Ca*. P. mali') (Seemüller and Schneider, 2004).

Phytoplasmas secrete small peptides, called effectors, in the host cells, altering the physiological processes and providing a fitness advantage to the bacteria (Sugio *et al*., 2011a). The first potential effectors were identified in Aster Yellows Phytoplasma (AY-WB) and thus named Secreted AY-WB Proteins (SAP) (Bai *et al.*, 2008).

SAP11 is one of the best characterized among phytoplasma effectors. It can target different TCP members (TEOSINTE-BRANCHED 1/CYCLOIDEA/PROLIFERATING CELL FACTORS) transcription factor family in many plant species, leading to the destabilization of Class II TCPs (Sugio *et al*., 2011b). However, a direct link between the deactivation of the transcription factors and symptoms development is yet to be demonstrated. The silencing of Class II TCP genes by stably overexpressing their regulator, the small RNA miR319, induces crinkly leaves in *Arabidopsis thaliana* (Schommer et al. 2014). Interestingly, a similar phenotype was observed to be induced by SAP11 expression (Lu *et al*., 2014; Chang *et al.*, 2018), indicating that SAP11 affects TCP-regulated downstream events similar to miR319.

The genome analysis of different strains of '*Ca*. P. mali' revealed the presence of a homolog of SAP11, known as ATP00189 or $SAP11_{CaPM}$ (Siewert et al. 2014). The effector was observed to target two apple Class II TCP transcription factors: *MdTCP13a* (ex *MdTCP24*) and *MdTCP4a* (ex *MdTCP25*) (Janik *et al.*, 2017). These two transcription

factors are likely to be involved in developmental processes and defence responses, such as leaf and root growth, phosphate tolerance and the Jasmonic acid biosynthesis pathway (Martín-Trillo and Cubas, 2010; Xu *et al*., 2014; Janik *et al*., 2017). However, there is a lack of in-depth research regarding the specific functions of TCP proteins in *Malus*.

Inheritable tolerance to AP was empirically observed in some experimental rootstock selections, mainly *M. sieboldii*-based rootstocks. After the infection, in some of these plants the phytoplasma titre was low or not detectable, and they either never developed symptoms or recovered within a few years (Bisognin *et al*., 2008; Seemüller *et al*., 2010). The physiological and molecular bases of this phenomenon are nonetheless mostly unknown to date.

The hypothesis investigated in this chapter is that the reduced severity of Apple Proliferation symptoms displayed by tolerant plants is explained by the inability of the bacterial effector $SAP11_{CaPM}$ to deactivate its molecular targets. Thus, the sequences of *MdTCP4a* and *MdTCP13a* were examined to screen for nucleotide polymorphisms in tolerant genotypes that could lead to structural changes in the proteins. Alternative variants of the protein were then tested to investigate if the amino acid changes reflect on the interaction strength with the effector.

To test for potential differences in the affinity for the bacterial effector displayed by susceptible and tolerant protein variants, the interactions were measured by applying two different protein-protein interaction assays, namely Yeast-2-hybrid (Y2H) and Bimolecular Fluorescence Complementation (BiFC) systems.

The Yeast-2-Hybrid is a technique first described in 1989 by Fields and Song, based on the modular properties of the yeast transcriptional activator Gal4, in which the N-terminal fragment binds to the DNA (DNA-binding domain - DBD) and the C-terminal fragment mediates the activation of the transcription (Activation Domain - AD). The principle of the Y2H reporter is that the transcription factor Gal4 is separated into two functional domains. Each domain is fused to one of the two potentially interacting proteins to be tested, called bait and prey: interaction between the two proteins reconstitutes the functional transcription factor Gal4, leading to the expression of a

reporter gene. In the assay presented in this chapter, the bacterial effector SAP11$_{CaPM}$ (bait) is genetically fused to the DBD of the *E. coli* repressor protein LexA, and the reporter genes are auxotrophic markers LEU2 and HIS3, that allow growth by positive clones on minimal media lacking leucine and histidine. To increase the stringency of the assay and identify cells with a high expression of the reporter gene, implying high binding affinity of the proteins tested, in the second part of the assay a competitive inhibitor of HIS3 reporter gene product, 3-aminotriazole (3-AT), was included in the media.

In Bimolecular Fluorescence Complementation (Hu *et al.*, 2002), the two proteins of interest are fused to two nonfluorescent halves of a fluorophore and expressed in living cells. In case of interaction between the two proteins, the two halves will complement, determining the emission of a fluorescent signal. In the assay shown in this study, the 2in1 BiFC system vector is used (Grefen and Blatt, 2012), which allows the subcloning and expression of both proteins in a single vector, thus ensuring equal gene dosage. The proteins are fused to the two halves of an enhanced yellow fluorescent protein (EYFP), and the same vector carries a monomeric red fluorescent protein (mRFP1) that serves as an internal transformation control.

RESULTS

Analysis of *MdTCP4a* and *MdTCP13a* in AP tolerant and susceptible genotypes

Two susceptible *Malus* × *domestica* plants (cv. Golden delicious, apple rootstock M9) and three AP tolerant (*Malus sieboldii* and its hybrids 4608 and D2218) were initially used for the analysis of potential differences in the sequences of the target genes of the AP effector SAP11$_{CaPM}$. The full-length *MdTCP13a* and *MdTCP4a* genes were *de novo* amplified from the DNA of the five plants and sequenced. The obtained nucleotide sequences were aligned to detect differences in the amino acid level between the three tolerant and the two susceptible plants for both genes. The analysis highlighted the presence of a total of eight and five non-synonymous Single Nucleotide Variants (SNVs) between the two susceptible and the three tolerant genotypes for *MdTCP4a* and *MdTCP13a*, respectively, as displayed in the alignments in figure 1.

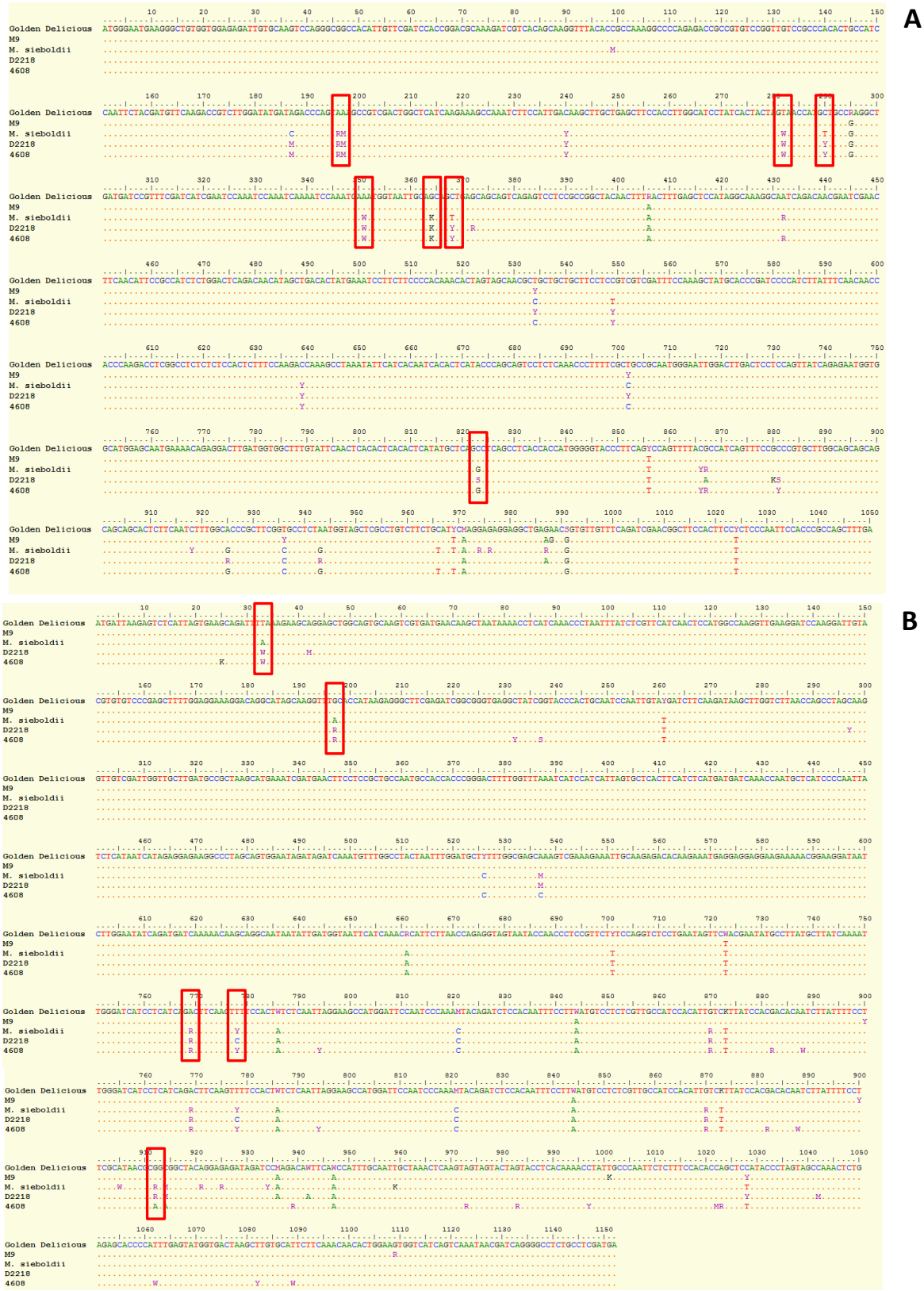

Figure 1 - Alignment of full-length *MdTCP4a* (A) *MdTCP13a* (B) of two susceptible (*M.* × *domestica* 'Golden Delicious' and M9) and three AP tolerant (*M. sieboldii*, 4608 and D2218) genotypes determined via Sanger sequencing. Identical residues to M. × domestica 'Golden delicious' are represented as dots. The non-synonymous SNVs displayed by the three tolerant genotypes are boxed in red. The alignments are visualized as BioEdit graphical output.

Since many *Malus* genomes are characterized by polyploidy, the Sanger sequencing cannot provide exhaustive information about the nucleotide substitutions but instead provide double or triple peaks for the polymorphic positions. To acquire a more explicit definition of the SNVs and unambiguously define the differences between AP susceptible and tolerant genotypes, the full-length genes were subcloned into a plasmid and re-sequenced using a primer specific for the vector backbone, 50 bp upstream of the amplicon insertion site.

The alignment of the *MdTCP13a* sequences revealed that the previously observed differences are equally distributed among the five genotypes, rejecting the hypothesis of the presence of unanimous tolerant-only mutations for this gene. Contrarily, based on the deduced amino acid substitutions, it was possible to deduce four distinct MdTCP4a protein variants. The four deduced variants were named after the colours attributed to the mutated amino acids in the software MEGA X (Molecular Evolutionary Genetics Analysis (S. Kumar *et al*., 2018)), used for the analysis: “Green”, “Red”, “Blue” and “Yellow”. In all cases, differences are described compared to the variant “Green”, the only of the four ubiquitously found in the five genotypes and thus identified as the wild type isoform. The two deduced protein variants observed in tolerant genotypes are characterized by two and one deduced amino acid substitutions, respectively. In the “Yellow” variant, the first amino acid substitution is deduced from two nucleotide substitutions in the first and second codon positions (AAA>GCA), and the second by a third codon position substitution (GCA>TCA). The “Blue” protein variant is characterized instead by a third codon position substitution (AGT>AGA). Details of the substitutions are listed in Table 1.

Table 1 – Position and type of non-synonymous SNVs observed in *TCP4* homologs of tolerant and susceptible species and varieties and related deduced amino acid substitutions. The nomenclature used to indicate the four potential protein variants is reported in the left column. SNV refer to the most commonly observed *MdTCP4a* sequence, named "Green".

VARIANT NAME	***SUBSTITUTION (cDNA)***	***DEDUCED SUBSTITUTION (AA)***	***GENOTYPE***
Green	…	…	'Golden delicious', 'M9', *M. sieboldii*, '4608', 'D2218'
Red	c.295G>A c.406A>G	p.E99K p.N136D	'Golden delicious', 'M9'
Yellow	c.196A>G; c.197A>C c.364G>T	p.K66A p.A122S	*M. sieboldii*, '4608', 'D2218'
Blue	c.282T>A	p. S94R	*M. sieboldii*, '4608', 'D2218'

The *MdTCP4a* sequences of 139 *Malus* genotypes were collected from data published by Duan and collaborators (2017) and analysed to get an insight into the frequency of the four *MdTCP4a* "Blue" and "Yellow" SNVs in *Malus* populations and species. The complete list of genotypes comprises 97 varieties of *M.* × *domestica,* 30 *Malus sieversii (M. sieversii),* 10 *Malus sylvestris (M. sylvestris),* one *Malus micromalus (M. micromalus)* and one *Malus fusca (M. fusca)*.

The results of the alignment are summarised in Table 2: the "S94R" deduced amino acid substitution ("Blue" variant) is present in eight genotypes: four *M.* × *domestica*, two *M. sieversii* and two *M. sylvestris*. While in three cases the third position of the codon, responsible for the amino acid change, is called as 'W', suggesting the presence of a heterozygous A-T peak as expected, in the remaining five cases the nucleotide is called as 'N', that may indicate a polymorphic position or a miscall.

Regarding the "Yellow" SNVs, no genotype displays both the c.196A>G; c.197C>A SNVs observed in the previously analysed genotypes, but six (five *M.* × *domestica,* one *M.*

micromalus) display instead a change in the second codon position only (c.197A>C) that would result in a p.K66T substitution. Three genotypes display as well the second amino acid change in position 122 that characterizes the deduced "Yellow" protein variant. The nucleotide changes appear to be polymorphic in *M.* × *domestica* F22682922, while *M. micromalus* displays monomorphism for both changes. In the remaining four cases, the positions are called as 'N'.

Finally, around half of the genotypes included in the analysis display the polymorphisms found in the susceptible genotypes M9 and 'Golden delicious', identifying "Red" as a common allelic variant diffused in the population.

The analysis results were used to identify interesting genotypes, that were cross-correlated with the plant material available at the Fondazione Edmund Mach (FEM) collection. The correspondent plant material was then sampled and further investigated: of the 13 genotypes from the previous analysis observed to carry at least one of the SNVs, only the three *Malus* × *domestica* varieties 'Yellow Transparent', 'Honeycrisp' and 'Chisel Jersey' were available. Although not the same accessions, individuals of the same species were available for *M. micromalus*, *M. sylvestris* and *M. sieversii*. Additionally, several more *M. sieboldii* were added to the analysis, together with 24 experimental varieties from FEM breeding programs, of which several have been used in preliminary screening tests for AP tolerance. The complete list of the genotypes included in the analysis is displayed in Table 3.

Table 2: List of the 13 genotypes (Duan *et al.*, 2017) displaying one of the two uncommon SNV in *MdTCP4a* previously observed in plants tolerant to AP. Details of the SNVs and the relative deduced translations are shown for each genotype. Each SNV is associated with one of the previously characterized protein variants. Nucleotide substitutions that differ from the canonical four bases are indicated according to the IUPAC code: M: A or C; K: G or T; W: A or T; N: any nucleotide.

SPECIES	VARIETY	SUBSTITUTION (cDNA)	DEDUCED AA SUBSTITUTION	VARIANT
M. × *domestica*	Yellow Transparent	c.197A>N c.364G>N	K66T/R/I A122S/T/P	"Yellow"
M. × *domestica*	Honeycrisp	c.197A>N	K66T/R/I	"Yellow"
M. × *domestica*	Chisel Jersey	c.197A>N	K66T/R/I	"Yellow"
M. × *domestica*	X0421	c.197A>N	K66T/R/I	"Yellow"
M. × *domestica*	F22682922	c.197A>M c.364G>K	K66T A122S	"Yellow"
M. micromalus	/	c.197A>C c.364G>T	K66T A122S	"Yellow"
M. × *domestica*	Budagowsky	c.282T>W	S94R	"Blue"
M. × *domestica*	Novosibirski	c.282T>W	S94R	"Blue"
M. × *domestica*	R11	c.282T>W	S94R	"Blue"
M. × *domestica*	Yellow Transparent	c.282T>N	S94R	"Blue"
M. sylvestris	A57	c.282T>N	S94R	"Blue"
M. sylvestris	A59	c.282T>N	S94R	"Blue"
M. sieversii	A37	c.282T>N	S94R	"Blue"
M. sieversii	A39	c.282T>N	S94R	"Blue"

Table 3: List of the genotypes further analysed for the resequencing of a region of *MdTCP4a*. On the left, genotypes selected based on preliminary analyses on data published by Duan *et al*. (2017). On the right column, experimental varieties generated at FEM, of which 17 were previously tested for AP tolerance.

ACCESSION	SPECIES/VARIETY	EXPERIMENTAL VARIETIES
PI 588859 16	Yellow transparent (Rus)	CC23
PI 588806 15	Chisel jersey (Uk)	CC82
/	Honeycrisp (Usa)	CC86
PI 392303	Gala	D45
PI 633815 AC1	*M. sieboldii*	D52
PI 590051 05	*M. sieboldii*	H29
PI 589958 14	*M. sieboldii*	H38
PI 633814 G1	*M. sieboldii*	W66
PI 594094 87	*M. sieboldii*	W92
PI 589749 10	*M. sieboldii*	AA43
PI 589957 05	*M. sieboldii*	H5
PI 613858 A1	*M. sieboldii*	L33
PI 594092 A12	*M. micromalus*	P16
PI 589955 12	*M. micromalus*	M26
PI 589753 11	*M. micromalus*	W471
PI 316974 06	*M. micromalus*	W12
PI 588976 08	*M. micromalus*	W355
PI 619190 A2	*M. micromalus*	W567
PI 594096 A9	*M. micromalus*	E1
GMAL 3608 H	*M. sieversii A39*	H118
GMAL 3608 N	*M. sieversii A39*	CC38
PI 392302 06	*M. sylvestris*	D49
PI 589382 11	*M. sylvestris*	W600
		M7

New primers for *MdTCP4a* were designed to amplify 293 bases covering the region of interest (nucleotide positions 136 - 429). DNA was extracted and purified from leaves of the 47 plants, amplified and sequenced with Next-Generation Sequencing techniques Illumina® (Illumina Inc., San Diego, California, USA). The unique sequences in total reads were identified as an operational taxonomic unit (OTU) to associate the single nucleotide variants with the different genotypes, and each read classified accordingly. The reads for each OTU per genotype were then counted to obtain the fraction of each deduced protein variant for each individual analysed. Results are displayed in Figure 2, where the four colours green, blue, yellow and red represent the corresponding deduced protein variant and asterisks the AP-tolerant plant. The experimental varieties from the FEM breeding program are indicated as "FEM 1" to "FEM 24" since data regarding AP tolerance of these plants cannot be disclosed yet. The ratio of each colour to the total is directly correlated with the fraction of each OTU on the total reads for each genotype. If known, the raw fraction values calculated on Illumina reads were normalized on the ploidy number of the corresponding genotype.

The three *M. domestica* varieties 'Yellow Transparent', 'Chisel Jersey' and 'Honeycrisp', which show the presence of "Yellow" or "Blue" protein variants in the preliminary analyses, were not confirmed to carry any of the two. The "Yellow" variant is displayed by all the *M. sieboldii* included in the analysis and nearly all *M. micromalus* accessions investigated, and none displays the hypothetical "K66T" substitution instead of "K66A", suggesting that the mutation AAA>ACA observed previously in place of AAA>GCA determined via Sanger sequencing is likely to be an artefact.

Of the 24 experimental varieties analysed, 16 carry one or both "Blue" or "Yellow" variants and, among these, 15 were tolerant in the screening tests, while among the eight not displaying the "Blue" nor the "Yellow" substitutions, only two resulted as AP tolerant.

Additional *MdTCP4a* sequences were obtained from a SNP calling analysis performed by Qiangmei *et al.* (data not published) by resequencing 308 *Malus* genotypes. Data were elaborated to identify the genotypes displaying the *MdTCP4a* SNVs observed in AP-tolerant genotypes: among a total of 20 SNPs in the 308 *TCP4a* sequences, ten appear

to be synonymous (four in the DBD) and ten non-synonymous (one in the DBD). Interestingly, but unfortunately, none of the 308 genotypes displays the SNVs characterizing the “Yellow”, “Blue”, or even “Red” protein variants.

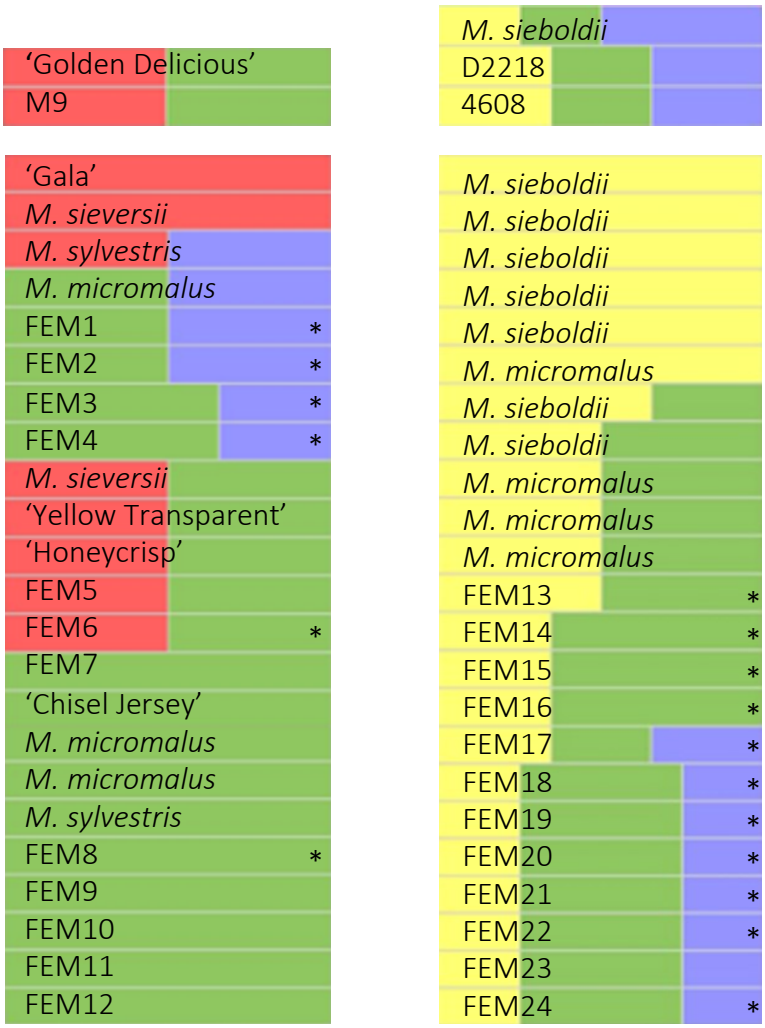

Figure 2: Graphical representation of the sequencing of 293 nt of *MdTCP4a* compared with AP screening tests*:* colours indicate the fraction of each of the four *MdTCP4a* alternative variants for each genotype,. The accessions indicated as “FEM” represent exprerimental varieties developed in the context of breeding programs at FEM. The asterisk indicates genotypes tolerant to AP. Of the 17 tolerant plants, 15 carry the “Yellow” or the “Blue” variant (four carry only “Blue”, four only “Yellow”), while two plants do not. The two susceptible and three tolerant starting plants are reported on top if the two columns.

In silico analysis of the potential impact of A66K and S94R amino acid substitutions on the function of the protein MdTCP4a

To investigate the potential impact on the protein function of the five deduced amino acid substitutions that characterize the three alternative variants, an *in silico* analysis has been conducted by interrogating two online predictors of the effect of amino acid substitutions on the protein structure: SIFT (Kumar *et al*., 2009; https://sift.bii.a-star.edu.sg/) and PROVEAN (Choi and Chan, 2015; http://provean.jcvi.org). Both software packages require the reference amino acid sequence and the positions and types of the amino acid substitutions as input, providing as output a score, which is associated with the probability of the substitution to be tolerated by the protein structure and function or not. The SIFT score varies between 0 and 1, where values lower than 0.05 mean the mutation is not tolerated, while the PROVEAN score usually ranges between +1.000 and -6.000 and, as default, values lower than -2.500 mean that the substitutions as a high impact on protein functionality.

The deduced amino acid substitution that characterizes the "Blue" variant, S94R, was predicted to be neutral/tolerated by both software packages (PROVEAN score -1.306, SIFT score 0.57), as well as both of "Red" deduced amino acid substitutions (E99K: PROVEAN score -0.499; SIFT score 0.45; N136D: PROVEAN +0.267, SIFT 0.75). For the "Yellow" variant, the substitution A122S was called neutral (PROVEAN +0.322, SIFT 0.46), while K66A was the only substitution predicted to have a strong/deleterious effect on the protein by both, with a PROVEAN score of -6.000 and a SIFT score of 0.00.

The examination of MdTCP4a homologs confirmed these results: alignments of *AtTCP4* homologs of different species, of *MdTCP*-Class II and Class II *TCP* from multiple species are displayed as Hidden Markov Model (HMM) logos in figure 3, which is a graphical representation of the conservation of sites in which the height of the letters is correlated with the frequency of that particular amino acid or nucleotide. The Lys-66 site is boxed to highlight that the site is conserved in the amino acid sequences with a frequency near 100%, as well as in the corresponding nucleotide sequences, with the first position of the codon displaying a 100% rate of conservation and the second being slightly lower. However, while the TCP DNA-Binding Domain (DBD) is highly conserved, said

conservancy drastically decreases outside the DBD, causing the reliability of the alignments to be very low even for close homolog sequences and close to 0 for orthologs. For this reason, only a tiny portion of the sequences flanking the TCP domain is displayed in the HMM; thus, the amino acid substitutions outside the DBD are not displayed.

The non-*Malus* sequences included in two alignments are listed in tables 4 and 5 (materials and methods), while members of the *MdTCP* Class II are specified in chapter 2.

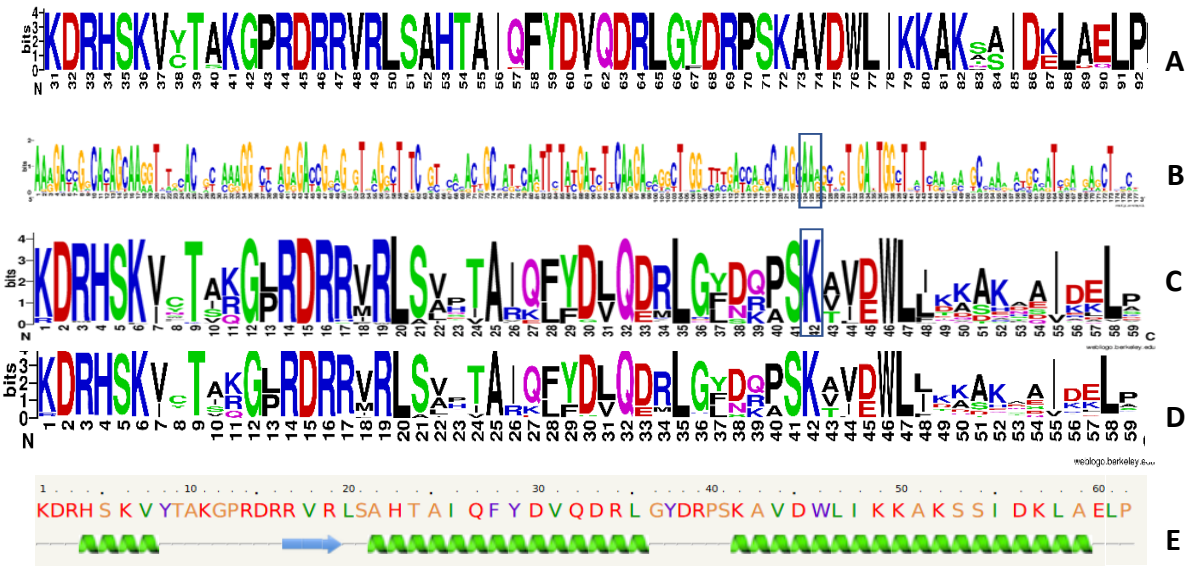

Figure 3 – HMM logos generated on alignments of several TCP sequences: in A), TCP4 homologs are displayed, in B) and C) nt and AA sequences of Class II - MdTCP, respectively; in D) Class II - TCPs of many species. The corresponding secondary structure prediction according to Phyre2 is displayed in E). The complete list of the sequences used to generate the alignments is reported in tables 4 and 5 of materials and methods.

The MdTCP4a Lys-66 amino acid is located in a strongly conserved position at the interface between the loop and the II helix that constitutes the basic Helix-Loop-Helix (bHLH) structural motif. Ser-94 is instead located ten amino acids downstream the Helix II end, a much less conserved site. The secondary structure prediction of MdTCP4a, calculated with Phyre2 (Protein Homology/analogy Recognition Engine V2.0) *(Kelley et al.,* 2015), is displayed in Figure 3, panel E. Outside the DBD of the TCP, the secondary

structure prediction has a low level of confidence, and the protein is predicted to be disordered. Despite being only ten amino acids downstream of the core domain, little can be said about the secondary and tertiary structures of the surroundings of Ser-94 position, as well as for the other amino acid substitutions previously identified.

Testing the interaction between "Yellow" and "Blue" deduced MdTCP4a protein variants and the bacterial effector $SAP11_{CaPM}$ through functional studies

To test the hypothesis that the "Yellow" and "Blue" deduced protein variants of tolerant genotypes might display less interaction affinity for $SAP11_{CaPM}$, the interaction strength between the effector and the targets was evaluated. Assays were performed both in a Yeast-2-Hybrid (Y2H) heterologous system and *in planta* with Bimolecular Fluorescence complementation (BiFC).

The full-length *MdTCP4a* genes from tolerant and susceptible plants were cloned into the yeast expression vector pGAD-HA-ccdb and co-transformed with the $SAP11_{CaPM}$-expressing bait vector pLexA (generated by Janik *et al*., 2017) in *Saccharomyces cerevisiae* NMY51, a Y2H reporter strain. Positive colonies expressing *MdTCP4a* and $SAP11_{CaPM}$ grew on selective media for all four variants tested, and no significant differences in the number or dimension of the colonies were detected.

Positive clones were transferred to a selective medium containing 10-fold increasing concentrations of 3-amino-1,2,4-triazole (3-AT) (1mM - 150 mM) to reinforce the stringency of the selection and, thus, to identify possible stronger and weaker interactions. The growth was measured after three and five days of incubation at 30°C. No differences were detected in growth on 3-AT-containing selective medium, with all clones growing at similar rates on media with 3-AT concentrations lower than 100 mM, while no growth was observed with 3-AT concentrations higher than 100 mM for any of the four variants.

Similar results were observed with co-expression in an *in planta* system, determined via BiFC. Full-length *MdTCP4a* genes corresponding to "Green", "Blue" and "Yellow" natural variants were *de novo* amplified from *M.* × *domestica* cv. 'Golden Delicious' and *M. sieboldii* DNA and subcloned together with the effector $SAP11_{CaPM}$ sequence without signal peptide in a BiFC plasmid. *Nicotiana benthamiana* protoplasts were transfected

with the three respective plasmids for transient expression of bacterial effector and *MdTCP4a* coupled with two fragments of the reporter fluorescent protein YFP. No significant difference in the proportion of positive cells was observed after 8-12 hours from transfection, with a percentage of 36-45%, 34-38%, and 37-39% positive protoplasts measured for “Green”, “Yellow” and “Blue” protein variants, respectively. Several attempts to obtain the purified MdTCP4a protein variants have been made to carry on additional functional studies, but unfortunately, with no success.

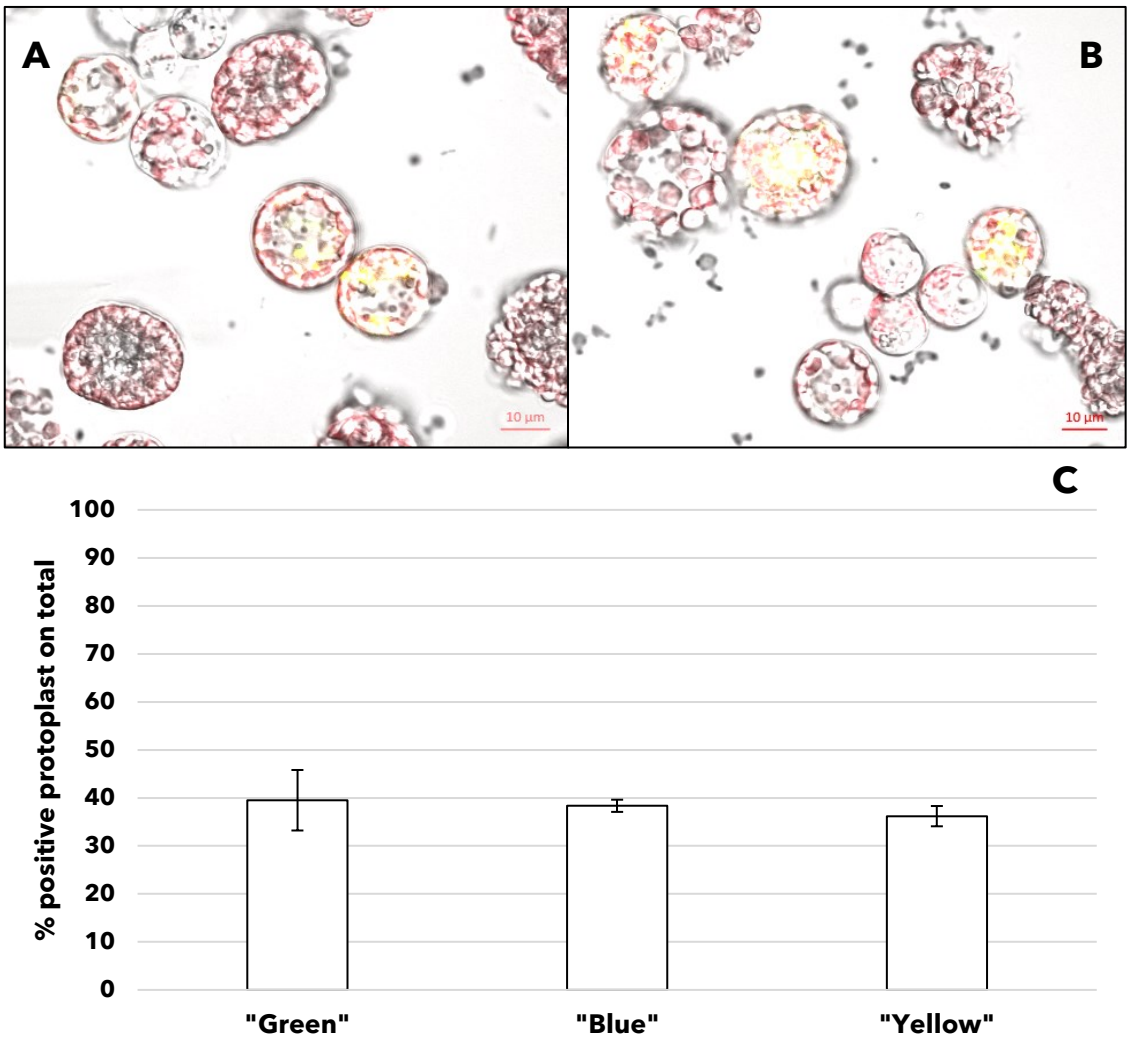

Figure 4 – *In planta* interaction of SAP11$_{CaPM}$ with MdTCP4a protein alternatives “Green”, “Yellow” and “Blue”: *N. benthamiana* protoplasts were transformed with pBiFCt-2in1 plasmids carrying the three MdTCP4a protein alternatives and SAP11$_{CaPM}$. Positive interaction leads to the reconstitution of YFP, as visible in the example in panels A (“Green” variant) and B (“Yellow” variant). In panel C, the bar chart reporting the percentages of positive protoplasts for the three variants is shown, displaying no differences in the percentage of positive in the three variants. Values were calculated on a minimum of 100 and a maximum of 150 randomly chosen protoplasts per each of the three protein alternatives, per replicate. The transfection and counts were *de novo* performed in triplicate.

Protein was detected by SDS-Page and Western Blot following the lysis of *Saccharomyces cerevisiae* (*S. cerevisiae*) strain NMY51 expressing HA-*MdTCP4a* from the pGAD-HA plasmid previously mentioned but was not detected after the purification steps. The full-length genes were thus *de novo* cloned in an *S. cerevisiae* expression plasmid with a C-terminal HIS-tag, but protein product was not detected following the induction in *S. cerevisiae*.

DISCUSSION

Plants displaying tolerance to Apple Proliferation have been observed and described in the past, primarily as a result of breeding of the resistant species *M. sieboldii* (Bisognin et al. 2008), classifying it as an inheritable phenotype. Nonetheless, despite being described for the first time in the late 80s, very little is known about this phenomenon's physiological and molecular bases.

In this chapter, sequences of *MdTCP4a* and *MdTCP13a*, the only known molecular targets of the bacterial effector $SAP11_{CaPM}$ in *Malus*, have been sequenced and analysed in both tolerant and susceptible genotypes, leading to the identification of two seemingly tolerant-exclusive deduced *MdTCP4a* protein alternatives. The two alternative sequences are characterized by three uncommon SNVs that lead to two potential amino acid substitutions: K66A in “Yellow” and S94R in “Blue” variants, respectively. The first deduced amino acid substitution is located at the end of the loop between the two helices that constitute the bHLH domain, while the latter ten amino acids downstream of the end of the Helix II, outside the core domain.

Several more *Malus* genotypes sequences were investigated, including experimental hybrids that did not display symptoms after the infection with Apple Proliferation phytoplasma. The analyses revealed that most of the tolerant varieties carry one or both previously described SNVs, highlighting at the same time the scarcity of *Malus domestica* common varieties displaying the nucleotide substitutions.

These results suggest that the two or three SNVs could potentially serve as a marker for tolerance to Apple Proliferation, but more data are required to assess the sensitivity and sensibility of the two mutations for the correct identification of tolerant plants.

Concurrently, these findings raised questions regarding the potential direct or indirect role of the mutations in the phenomenon of tolerance itself and were thus investigated. All the available *Malus* sequences deduced from a SNP calling of 308 *Malus* genotypes were analysed, suggesting no genotype carries any of the previously observed SNVs. These findings support the idea that substitutions are rare, although the sensitivity of a genome-scale SNP calling for such a high number of genotypes may not be high enough to correctly identify the totality of such uncommon polymorphisms. It is indeed possible that the number of SNPs observed in those data may be an underestimation of the real variability.

The search was thus expanded to include all the TCP4 homologs known to date: while the Ser-94 position displays a certain degree of variability among different species, it is interesting to note that no member of the *AtTCP4* homolog in any of the species displays an amino acid other than a Lysine residue in the 66 relative position. The same strong degree of conservation was detected when considering ortholog sequences, like all *Malus* Class II *TCP* genes, where the Lys-66 residue was observed to be conserved among genes at a 100% level and also in Class II *TCP*s of phylogenetically distant groups.

Studies performed on *AtTCP4* showed that the Lys-87 residue of *AtTCP4*, the ortholog of *MdTCP4a* Lys-66, seems not to be directly involved in the interaction with the DNA, although the two residues downstream, Ala-88 and Val-89, are physically contacting the DNA double helix during the interaction (Aggarwal *et al.,* 2010). Moreover, the residue appears to be functionally conserved even in non-TCP bHLH Transcription Factors like the animal proteins Upstream Transcription Factor 1 or Myoblast Determination Protein 1 (Aggarwal *et al*., 2010). These observations suggest that the Lysine residue between the two helices may have an essential functional role in the secondary or tertiary structures of the protein.

Most recent data indicate that $SAP11_{CaPM}$ can bind all members of *A. thaliana* Class II TCPs and, in addition, some members of Class I in Y2H assays (Strohmayer *et al., 2020* [preprint]), implying that the SAP11 binding site may be identified in the bHLH TCP domain, it being the only conserved region among different TCP proteins.

To assess the possible impact of the mutation on the protein functionality, automatic predictor software were interrogated, confirming the possible strong/deleterious effect of the K66A substitution on the protein structure, while the "Blue", S94R, appears instead to be tolerated. Although expected, these results may be biased by the lack of similar sequences that display enough variability. In fact, despite being slightly different to one another, the algorithms driving these software are based on the same simple principle. After collecting all the sequences that share a certain degree of identity with the query and discarding the ones that are too similar, they finally calculate the frequency of observing a mutation in the site of interest in similar sequences, providing an index that reflects the freedom of change of that particular site. Although the indices are influenced by an estimation of the tolerability of a substitution by the surrounding amino acids, the prediction is heavily driven by the availability of sequences that display the same or similar amino acid substitutions. Given the rate of conservation typical of the TCP DBD and of the residue Lys-66 in particular, the results of the predictions are thus not surprising. Ultimately, automatized software will not provide much more additional information than the Lys-66 position is strongly conserved across orthologues and paralogues.

A non-functional analysis of how a mutation could affect the structure and function of a protein should consider at least three aspects: the conservation among orthologs and paralogs, the location and exposure level of the residue and the potential role in protein structure or function (Betts & Russell, 2003). As demonstrated before, most of these aspects regarding Lys-66 amino acid residue suggest an essential role of the amino acid in the protein function. However, this kind of *in silico* analyses cannot overcome a degree of uncertainty and are prone to limitations that can only lead to make a reasonable guess, making the functional analyses to test these hypotheses a *sine qua non* condition.

Functional studies aimed to compare the level of affinity of the different protein variants of *MdTCP4a* for the phytoplasma effector SAP11$_{CaPM}$ were hence carried out, both in a heterologous system (Yeast-2-hybrid) and *in planta* (BiFC). The results of both assays indicated no difference in the interaction affinity between the bacterial effector and the

protein alternatives. Although the data may be confirmed by further analysis, these results suggest that the interaction strength itself is not involved in the molecular mechanisms conferring tolerance to Apple Proliferation. In biological conditions, the two interactor concentrations are probably several orders of magnitude lower than *in vitro* experiments, and the original conditions may determine a different outcome, but these conditions can hardly be reproduced.

The hypothesis that the mutations have an impact on the protein function or structure cannot be rejected: the deduced changes in the amino acid sequences may reflect directly on the affinity of the protein with the DNA, or with other members of the *TCP* family, and these hypotheses have never been explored. Unfortunately, all the attempts to obtain the purified proteins for further functional studies were unsuccessful.

Besides the functional elucidation, the potential role of *MdTCP4a* as a marker of tolerance is promising and needs to be considered for additional studies.

MATERIALS AND METHODS

Malus DNA extraction, amplification and sequencing of *MdTCP4a* and *MdTCP13a*

Malus in vitro plants were provided by P.L. Bianchedi (FEM): *M.* × *domestica* cv. 'Golden Delicious'; *M.* × *domestica*-based rootstock M9; *M. sieboldii; M. sieboldii* F1 hybrid 4608; *M. sieboldii* F2 hybrid D2218.

200 µg of fresh leaf tissue was collected for each plant and subjected to DNA extraction with the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions with elution in 50 µL DNase-Free Water. DNA samples were quantified via Nanodrop and diluted to a final concentration of 5 ng/µL.

MdTCP4a and *MdTCP13a* full-length genes were amplified from 10 ng of starting *Malus* DNA using the primer combinations MdTCP4a_for/MdTCP4a_rev and MdTCP13a_for/MdTCP13a_rev (Table 6) respectively (0,5 µM each), iProof™ High-fidelity polymerase (Bio-Rad, Hercules, California, USA) in 1 X HF-buffer (Bio-Rad, Hercules, California, USA), 400 µM dNTPs, with the following cycling conditions: 30'' initial denaturation at 98°C; 35 cycles of 10" at 98°C, 30" at 59,5°C, 90" at 72°C; final elongation step at 72°C for 5'. Amplicons were purified, checked on agarose gel 1,5 %

and sequenced by GATC Biotech (Constance, Germany). Sequencing data were analysed using MEGA X Software.

MdTCP4a and *MdTCP13a* genes were de novo amplified with SfiI restriction sites - containing primers MdTCP4a_SfiIA_forw/MdTCP4a_SfiIB_rev and MdTCP13a_SfiIA_forw/ MdTCP13a_SfiIB_rev (Table 6) respectively (0,5 µM each) as described above and the following cycling conditions: 3' of initial denaturation at 98°C, 35 cycles of 30" at 98°C, 30" at 62°C (MdTCP4a_SfiI primer pair) or 60°C (MdTCP13a_SfiI primer pair), 90" at 72°C, and a final elongation step at 72°C for 4'. Amplicons were purified, digested with SfiI (Roche, Switzerland), subcloned into SfiI-linearized pGAD-HA-ccdB (ClonTech, Mountain View, California, USA) via T4-ligation and transformed into *E. coli* MegaX DH10B™ Electrocomp™ Cells (Invitrogen™, Thermo Fisher Scientific, Waltham, Massachusetts, United States). Positive clones were propagated and eight clones per genotype per gene isolated and sequenced by GATC Biotech (Ebersberg, Germany) with forward primer GAL4-AD-SEQ. Sequencing data were analysed with Mega X Software.

MiSeq Illumina

The three/four youngest leaves were collected in the field in April for all the plants included in the analysis, and flash-frozen in liquid nitrogen. DNA was extracted from 1 mg of leaf material with the DNeasy Kit Mini following the manufacturer's protocol, eluted in 50 µL of DNase-free water, quantified via Nanodrop 8000 and subsequently diluted with water to a concentration of 10 ng/µL each. 10 ng of DNA were used to amplify the region of interest of *MdTCP4a* (136 to 429 bp) with the primer pair MdTCP_AS_Ill (Table 6) and overhang Illumina adapters to generate the Illumina library amplicons and sequenced on an Illumina MiSeq (PE300) platform (MiSeq Control Software 2.0.5 and Real-Time Analysis Software 1.16.18) as reported by Quail *et al*. (2012). The Micca tool (www.micca.org) was used to elaborate fastq data using a quality filtering cut-off of 0.75 and identify unique reads.

TCP sequence analysis

The annotated TCP sequences used for the alignments were retrieved from GenBank® and UniProt®. Unknown potential TCP proteins were obtained via BLAST searches

(Johnson *et al.* 2008), using known TCP sequences as queries. The sequences are listed in Tables 4 and 5, while the *MdTCP* sequences are reported in the previous chapter.

The analysis was performed on MEGA X® Software and BioEdit®. Sequences were aligned with ClustalW using different parameters. The HMM logos displayed were obtained with the Weblogo tool (Crooks *et al*., 2004) (weblogo.berkeley.edu).

The SIFT and PROVEAN software queries had the following parameters: the database used for the SIFT query is UniProt-SwissProt 2010_09, the median conservation of sequences 3.00, and sequences found to be more than 90 % identical to query were removed from the set. The database used for the Provean query was NCBI nr, September 2012, and the prediction cut-off was set at 2.5000.

Table 4: List of *TCP4* homolog sequences included in the alignment, used to determine the conservation rate of *TCP4* DBD by constructing the HMM Logo in Figure 3, panel A

SPECIES	ACCESSION NUMBER
Aquilegia coerulea	AKA88916.1
Arabidopsis lyrata	XP_020887933.1
Arabidopsis thaliana	BAA97066.1
Arachis duranensis	XP_015936317.1
Arachis ipaensis	XP_016171331.1
Brassica napus	XP_013644742.1
Brassica oleracea	XP_013587208.1
Brassica rapa	XP_009146296.1
Cajanus cajan	XP_020229472.1
Capsella rubella	XP_006297817.1
Cephalotus follicularis	GAV75987.1
Cicer arietinum	XP_004487176.1
Citrus clementina	XP_006446753.1
Citrus sinensis	XP_006469046.1
Citrus unshiu	GAY66791.1
Coffea canephora	CDP15905.1
Eucalyptus grandis	XP_010035834.1
Fragaria vesca	XP_004294116.1

Glycine max	XP_003540389.1
Glycine soja	KHN27924.1
Gossypium arboreum	XP_017642085.1
Gossypium hirsutum	XP_016702515.1
Gossypium raimondii	XP_012457642.1
Herrania umbratica	XP_021280352.1
Hevea brasiliensis	XP_021685492.1
Ipomoea nil	XP_019159423.1
Jatropha curcas	XP_012079039.1
Juglans regia	XP_018824268.1
Lupinus angustifolius	XP_019449056.1
Macleaya cordata	OVA14499.1
Manihot esculenta	XP_021631691.1
Medicago truncatula	XP_013465197.1
Morus notabilis	XP_010087306.1
Nelumbo nucifera	XP_010277724.1
Parasponia andersonii	PON41948.1
Phaseolus vulgaris	XP_007149767.1
Populus euphratica	XP_011000248.1
Populus trichocarpa	XP_002317368.1
Prunus avium	XP_021826431.1
Prunus persica	XP_007211426.1
Pyrus × bretschneideri	XP_009357156.1
Quercus suber	XP_023928745.1
Ricinus communis	XP_002525138.1
Rosa chinensis	XP_024183758.1
Theobroma cacao	XP_017980250.1
Trema orientalis	PON81604.1
Trifolium pratense	PNX75974.1
Trifolium subterraneum	GAU27089.1
Vigna angularis	XP_017425730.1
Vitis vinifera	XP_010657020.1

Table 5: List of Class II *TCP* homolog sequences included in the alignment, used to determine the conservation rate of *TCP4* DBD by constructing the HMM Logo in Figure 3, panel D. Accessions are referred to Refseq database unless indicated [[1] TAIR REFERENCE [2]PLANT TFDB [3]TF ID]

SPECIES	ANNOTATION	ACCESSION NUMBER [123]
Amborella trichopoda	TCP2	XM_011622476.2
	TCP4	XM_011627971.2
	TCP5	XP_011623969.1
Arabidopsis thaliana	TCP1	AT1G67260.1 [1]
	TCP10	AT2G31070.1 [1]
	TCP12	AT1G68800.1 [1]
	TCP13	AT3G02150.1 [1]
	TCP17	AT5G08070.1 [1]
	TCP18	AT3G18550.1 [1]
	TCP2	AT4G18390.1 [1]
	TCP24	AT1G30210.1 [1]
	TCP3	AT1G53230.1 [1]
	TCP4	AT3G15030.1 [1]
	TCP5	AT5G60970.1 [1]
Marchantia polymorpha	TCP1	Mapoly0001s0298.1.p [3]
Nymphaea colorata	TCP13	XP_031491538.1
	TCP24	XP_004175578.1
	TCP4	XP_031498212.1
	TCP5	XP_031489432.1
Populus euphratica	TCP1	CCG000586.1 [2]
	TCP10	CCG007701.1 [2]
	TCP11	CCG007964.1 [2]
	TCP14	CCG011685.1 [2]
	TCP15	CCG012390.1 [2]
	TCP16	CCG012899.1 [2]
	TCP18	CCG015462.1 [2]
	TCP2	CCG001135.1 [2]
	TCP23	CCG021224.1 [2]
	TCP26	CCG025130.1 [2]
	TCP27	CCG026334.1 [2]
	TCP29	CCG026590.1 [2]
	TCP3	CCG002481.1 [2]
	TCP31	CCG028420.1 [2]
	TCP4	CCG002729.1 [2]
Physcomitrella patens	TCP3	Pp3c7_5230V3.2.p [3]
	TCP7	Pp3c11_26670V3.1.p [3]
Prunus persica	Cycloidea TF	XP_020419090.1
	Dichotoma TF	XP_007216681.2
	TCP13	XP_007215550.1

	TCP18	XP_020409504.1
	TCP2	XP_007211837.1
	TCP4	XP_020418233.1
	TCP4_2	XP_007214360.1
	TCP4_3	XP_020416703.1
	TCP5	XP_007209843.1
Rosa chinensis	PCF7-like	XP_024160085.1
	TCP1	XP_024184504.1
	TCP10-like	XP_024171607.1
	TCP13	XP_024184148.1
	TCP18	XP_024173309.1
	TCP4	XP_024157688.1
	TCP5	XP_024173744.1
	uncharacterized	XP_024197562.1
Selaginella moellendorffii	TCP1	89227 [3]
	TCP2	29162 [3]
	TCP6	28794 [3]
Vitis vinifera	TCP1	XP_010660504.1
	TCP10	RVW29667.1
	TCP11	CBI15411.3
	TCP12	XP_010647995.1
	TCP13	XP_002277074.1
	TCP18	XP_019072942.1
	TCP2	XP_002271548.1
	TCP4	RVW42962.1
	TCP8	XP_010657020.1

Yeast-2-Hybrid assay

MdTCP4a alternative sequences were cloned into pGAD-HA plasmids as described above and used to transform *S. cerevisiae* strain NMY51 carrying the bait vector pLexA_SAP11$_{CaPM}$ as described by Janik *et al*. (2017). The transformation was evaluated after incubation at 30°C for 72 - 96 hours on selective plates. After the transformation, three positive clones for each isoform were mixed and suspended in 3 mL liquid selective media and incubated for four hours on a shaking incubator at 30°C. After the incubation 0,25 OD_{600} units of yeast suspension were centrifuged and resuspended in 100 µL 0,8% NaCl solution. 1 µL yeast solution was then pipetted on solid selective media containing 3-AT at the following concentrations: 1 mM, 10 mM, 20 mM, 30 mM, 40 mM, 50 mM, 60 mM, 70 mM, 80 mM, 100 mM, 110 mM, 125 mM, 150 mM. Plates were incubated at 30°C and evaluated after 72 and 120 hours.

Bimolecular Fluorescence Complementation

MdTCP4a full-length genes were *de novo* amplified from *M.* × *domestica* cv. 'Golden Delicious' and *M. sieboldii* DNA with primers containing attB1 and attB4 overhangs, following the same protocol and cycling conditions as described above. Amplicons were cloned into Gateway™ pDONR™221_attB1-4 vector (Invitrogen™, Thermo Fisher Scientific, Waltham, Massachusetts, United States) via Gateway™ BP reaction following the manufacturer's protocol, transformed into *E. coli* MegaX DH10B™ Electrocomp™ Cells, isolated and purified. The four entry plasmids were then subcloned together with previously generated $SAP11_{CaPM}$-containing pDONR®221_attB3-2 via Gateway™ LR reaction in the pBiFCt-2in1-NN plasmid (Addgene plasmid 105111; http://n2t.net/addgene:105111) (Grefen and Blatt, 2012) for *in planta* transient expression of N-terminal cYFP and nYFP fused proteins and transformed in *E. coli* MegaX DH10B™ Electrocomp™ Cells. *Nicotiana benthamiana* protoplasts were isolated and transfected with 5 ug of the respective purified plasmid as described by Waadt and Kudla (2008).

After 72-96 hours post-transfection, protoplasts were observed by confocal laser scanning microscopy using an LSM800/Axio Observer Z1 (Carl Zeiss GmbH, Jena, Germany), with the following parameters: for DsRed (RFP) excitation wavelength at 561 nm, detection at 540-639nm; for YFP excitation wavelength at 488, detection at 410-540nm. Autofluorescence of chlorophyll was detected at 656-700nm. Brightfield: ESID-detector.

Protein expression and SDS Page/WB

S. cerevisiae, strain NMY51, previously transformed with the pGAD-HA plasmid containing the four *MdTCP4a* alternative sequences "Green", "Yellow", "Red" and "Blue" was incubated in SD minimal selective media at 30°C for 16 hours on a shaking incubator, then pelleted and subjected to whole-cell lysis with Y-PER™ Yeast Protein Extraction Reagent (Thermo Fisher Scientific, Waltham, Massachusetts, USA) following the manufacturer's protocol.

To detect the protein in a Western blot, samples were first separated on 12% SDS-PAGE gels and subsequently blotted onto a PVDF membrane. The membrane was probed with anti-HA mouse primary antibody (Santa Cruz Biotechnology, Dallas, Texas, USA) in Milk-PBS 5% overnight at 4°C, followed by incubation with Goat anti-mouse-IgG-HRP conjugated (Thermo Fisher Scientific, Waltham, Massachusetts, United States) secondary antibody. Proteins were detected by the HRP activity using ChemiDoc™ Imaging System (Bio-Rad, Hercules, California, USA) with Pierce ECL Western Blot Substrate (Thermo Fisher Scientific, Waltham, Massachusetts, United States).

MdTCP4a alternative sequences were *de novo* sequenced with a reverse primer lacking the stop codon and cloned into Gateway™ pENTR/D-TOPO vectors via TOPO cloning (Thermo Fisher Scientific, Waltham, Massachusetts, United States) and subsequently subcloned in pYES2-DEST52 Gateway™ destination vector for yeast expression with a HIS-Tag. Electro-competent *S. cerevisiae* strain INVSc1 was transformed with the four destination vectors and the expression induced following the manufacturer's protocol. Induced yeast was subjected to protein extraction by following the procedures described by von den Haar (2007). Lysate samples were detected in SDS Page/Western Blot as described above, with a HIS-Tag Rabbit Antibody (GenScript® Biotech, Piscataway, New Jersey, USA) as primary and Goat Anti-Rabbit IgG HRP conjugate (Millipore®, Burlington, Massachusetts, USA).

Table 6 – List of primers used in the present chapter, with the corresponding nucleotide sequences.

NAME	SEQUENCE (5' – 3')
MdTCP4a_fw	caccATGGGAATGAAGGGCTG
MdTCP4a_rev	TCAAAGCTGGCGGGTG
MdTCP13a_for	caccATGATTAAGAGTCTCATTAG
MdTCP13a_rev	TCATCGAGGCAGAGGC
MdTCP4a_SfiIA_forw	ccccccggccattacggccATGGGAATGAAGGGCTG
MdTCP4a_SfiIB_rev	ccccccggccgaggcggccTCAAAGCTGGCGGGTG
MdTCP13a_SfiIA_forw	ccccccggccattacggccATGATTAAGAGTCTCATTAG

MdTCP13a_SfiIB_rev	ccccccggccgaggcggccTCATCGAGGCAGAGGC
GAL4-AD-SEQ	AATACCACTACAATGGAT
MdTCP4a_AS_III_fw	CCCACACTGCCATCCAATTC
MdTCP4a_AS_III_rev	CCTTTGCCTATGGAGCTCAA
MdTCP4a_attB1-fw	ggggacaagtttgtacaaaaaagcaggcttaATGGGAATGAAGGGCTG
MdTCP4a_attB4-rev	ggggacaactttgtatagaaaagttgggtgTCAAAGCTGGCGGGTG
MdTCP4a_rev-nostop	AAGCTGGCGGGTG

REFERENCES

Aggarwal, P., Das Gupta, M., Joseph, A. P., Chatterjee, N., Srinivasan, N., & Nath, U. (2010). Identification of Specific DNA Binding Residues in the TCP Family of Transcription Factors in *Arabidopsis*. *The Plant Cell*, *22*(4), 1174–1189. https://doi.org/10.1105/tpc.109.066647

Bai, X., Correa, V. R., Toruño, T. Y., Ammar, E.-D., Kamoun, S., & Hogenhout, S. A. (2008). AY-WB Phytoplasma Secretes a Protein That Targets Plant Cell Nuclei. *Molecular Plant-Microbe Interactions*, *22*(1), 18–30. https://doi.org/10.1094/mpmi-22-1-0018

Betts, M. J., & Russell, R. B. (2003). Amino Acid Properties and Consequences of Substitutions. In M.R. Barnes and I.C. Gray (Ed.), *Bioinformatics for Geneticists* (pp. 289–316). https://doi.org/10.1002/0470867302.ch14

Bisognin, C., Schneider, B., Salm, H., Grando, M. S., Jarausch, W., Moll, E., & Seemüller, E. (2008). Apple proliferation resistance in apomictic rootstocks and its relationship to phytoplasma concentration and simple sequence repeat genotypes. *Phytopathology*. https://doi.org/10.1094/PHYTO-98-2-0153

Chang, S. H., Tan, C. M., Wu, C. T., Lin, T. H., Jiang, S. Y., Liu, R. C., Tsai, M. C., Su, L. W., & Yang, J. Y. (2018). Alterations of plant architecture and phase transition by the phytoplasma virulence factor SAP11. *Journal of Experimental Botany*. https://doi.org/10.1093/jxb/ery318

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btv195

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*. https://doi.org/10.1101/gr.849004

Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., Wang, X., Jiao, C., Legall, N., Mao, L., Wan, S., Wang, K., He, T., Feng, S., Zhang, Z., Mao, Z., Shen, X., Chen, X., Jiang, Y., … Chen, X. (2017). Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications*, *8*(1).

https://doi.org/10.1038/s41467-017-00336-7

Fields, S., & Song, O. K. (1989). A novel genetic system to detect protein-protein interactions. *Nature*. https://doi.org/10.1038/340245a0

Grefen, C., & Blatt, M. R. (2012). A 2in1 cloning system enables ratiometric bimolecular fluorescence complementation (rBiFC). *BioTechniques*. https://doi.org/10.2144/000113941

Hu, C. D., Chinenov, Y., & Kerppola, T. K. (2002). Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Molecular Cell*. https://doi.org/10.1016/S1097-2765(02)00496-3

Janik, K., Mithöfer, A., Raffeiner, M., Stellmach, H., Hause, B., & Schlink, K. (2017). An effector of apple proliferation phytoplasma targets TCP transcription factors—a generalized virulence strategy of phytoplasma? *Molecular Plant Pathology*, *18*(3), 435–442. https://doi.org/10.1111/mpp.12409

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkn201

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*. https://doi.org/10.1038/nprot.2015.053

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. https://doi.org/10.1038/nprot.2009.86

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msy096

Lu, Y., Li, M., Cheng, K., Tan, C. M., Su, L., & Lin, W. (2014). *Transgenic Plants That Express the Phytoplasma Effector SAP11 Show Altered Phosphate Starvation and*. *164*(March), 1456–1469. https://doi.org/10.1104/pp.113.229740

Martín-Trillo, M., & Cubas, P. (2010). TCP genes: a family snapshot ten years later. *Trends in Plant Science*, *15*(1), 31–39. https://doi.org/10.1016/j.tplants.2009.11.003

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., & Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*. https://doi.org/10.1186/1471-2164-13-341

Schommer, C., Debernardi, J. M., Bresso, E. G., Rodriguez, R. E., & Palatnik, J. F. (2014). Repression of cell proliferation by miR319-regulated TCP4. *Molecular Plant*, *7*(10), 1533–1544. https://doi.org/10.1093/mp/ssu084

Seemüller, E., Bisognin, Grando, Schneider, Velasco, & Jarausch, W. (2010). *Breeding of rootstocks resistant to apple proliferation disease*. 2009–2011.

Seemüller, E., & Schneider, B. (2004). "Candidatus Phytoplasma mali", "Candidatus Phytoplasma pyri" and Candidatus Phytoplasma prunorum', the causal agents of apple proliferation, pear decline and European stone fruit yellows, respectively. *International Journal of Systematic and Evolutionary Microbiology*. https://doi.org/10.1099/ijs.0.02823-0

Siewert, C., Luge, T., Duduk, B., Seemüller, E., Büttner, C., Sauer, S., & Kube, M. (2014). Analysis of expressed genes of the bacterium "candidatus phytoplasma mali" highlights key features of virulence and metabolism. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0094391

Strohmayer, A., Schwarz, T., Braun, M., Krczal, G., & Boonrod, K. (2020). The anticipated potential nuclear localization sequence of ' Candidatus Phytoplasma mali' SAP11-like protein is required for TCP binding but not for transport into the nucleus. *BioRxiv*. https://doi.org/10.1101/2020.05.13.094045

Sugio, A., Kingdom, H. N., MacLean, A. M., Grieve, V. M., & Hogenhout, S. A. (2011). Phytoplasma protein effector SAP11 enhances insect vector reproduction by manipulating plant development and defense hormone biosynthesis. *Proceedings of the National Academy of Sciences*, *108*(48), E1254–E1263.

https://doi.org/10.1073/pnas.1105664108

Sugio, Akiko, MacLean, A. M., Kingdom, H. N., Grieve, V. M., Manimekalai, R., & Hogenhout, S. A. (2011). Diverse Targets of Phytoplasma Effectors: From Plant Development to Defense Against Insects. *Annual Review of Phytopathology*, *49*(1), 175–195. https://doi.org/10.1146/annurev-phyto-072910-095323

von der Haar, T. (2007). Optimized protein extraction for quantitative proteomics of yeasts. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0001078

Waadt, R., & Kudla, J. (2008). In plant visualization of protein interactions using bimolecular fluorescence complementation (BiFC). *Cold Spring Harbor Protocols*. https://doi.org/10.1101/pdb.prot4995

Xua, R., Sunb, P., Jiab, F., Luc, L., Lib, Y., Zhang, S., & Huang, J. (2014). Genome-wide Analysis of TCP Transcription Factor Gene Family in Malus Domestica. *Ias.Ac.In*, *93*(3), 733–746. http://www.ias.ac.in/jgenet/forthcoming/unedited/jgen-14-101-ue.pdf

CHAPTER 4

OVEREXPRESSION OF MOLECULAR TARGETS OF THE APPLE PROLIFERATION PHYTOPLASMA EFFECTOR $SAP11_{CaPM}$ IN APPLE

INTRODUCTION

Significant progress has been made in unravelling the functions of the TCP transcription factors in a wide array of different species, and their crucial role in regulating plant morphogenesis and their effect on flowering (Li *et al*. 2019), leaf development (Sarvepalli and Nath, 2018) and axillary bud growth (Aguilar-Martínez *et al.*, 2007) has been described. Of particular interest is their role in the responses to abiotic and biotic stresses, including cold-related responses, shade avoidance and trichome regulation (Zhou *et al*., 2018; Vadde *et al*., 2018; Fan *et al*., 2020; Fang *et al.,* 2020), but also during effector-triggered immunity (Mukhtar *et al*., 2011; S. Li, 2015; Lopez *et al*., 2015). Attention on these genes was further intensified by the observation that one of the principal effector proteins of phytoplasma, SAP11, binds and deactivates members of this group (Sugio *et al.*, 2011). In fact, many of the typical symptoms of phytoplasma related diseases, like shoot proliferation and abnormal growth of leaves and flowers, may be induced by the deactivation of TCPs, although direct molecular evidence is lacking.

Despite the agronomic importance of apple and the considerable negative impact of Apple Proliferation disease on the economic yield, studies focusing on the roles of the TCP proteins in *Malus* are very rare and, therefore, there is limited knowledge on the specific functions of these transcription factors in this organism, with information mainly being borrowed from other species.

Two members of the *M.* × *domestica* TCP family targeted by $SAP11_{CaPM}$, *MdTCP4a* and *MdTCP13a*, are members of Class II, CINCINNATA (CIN)-like subgroup. This subclass comprises the vast majority of Class II TCP genes (Martín-Trillo and Cubas, 2010),

involved primarily in negatively regulating cell division to control the morphology of leaves and shoot organs (Palatnik *et al.* 2003; Koyama *et al*. 2007).

As mentioned before, most data regarding the function of these transcription factors derive from studies on model plant species, such as *Arabidopsis thaliana*. Specifically, *A. thaliana TCP13* (*AtTCP13*), besides being involved in plant morphogenesis, was observed to bind and activate the transcription of the chloroplast light-responsive promoter psbD (Thum *et al.,* 2001). Interestingly, *AtTCP13* was shown to be bound by effectors of the Gram-negative plant-pathogenic bacterium *Pseudomonas syringae* and the obligate biotrophic oomycete *Hyaloperonospora arabidopsidis*. *A. thaliana tcp13* single mutants exhibit enhanced susceptibility to the disease (Mukhtar *et al.*, 2011), suggesting a pivotal role of this transcription factor in the immune system functioning.

AtTCP4 was shown to activate the transcription of several genes related to morphogenesis, being involved in mitotic cell differentiation signalling (Challa *et al.*, 2019), promotion of cell elongation in hypocotyls (Challa *et al.*, 2016) and suppression of trichome initiation (Vadde *et al.* 2019). This transcription factor also participates in ovule development (Wei *et al.* 2015) and plays a key role in jasmonic acid biosynthesis (Schommer *et al.* 2008; Danisman *et al*. 2012).

Several CIN-Class II genes, including *AtTCP4*, but not *AtTCP13*, have a binding domain for a microRNA, namely miR319 (Palatnik *et al.,* 2003). It was shown that this miRNA tightly regulates the activity of CIN-Class II TCPs, and mutants with point mutations in the TCP miRNA recognition regions display dramatic changes in the phenotype, suggesting that the post-transcriptional regulation plays a significant role in determining the activity of this set of transcription factors (Palatnik *et al*., 2007).

These mutants have been used to determine many of the roles of the TCPs. It was shown that TCP single-mutant often display mild or no effects on the phenotype, while the simultaneous suppression of many TCP genes has a more substantial impact (Schommer et al. 2008; Efroni *et al.* 2008). These results indicate that *TCP* genes possess a high potential for functional redundancy. Together with the capacity of TCPs to bind a broad range of different proteins, this aspect has hindered efforts in defining precisely the variety of functions exhibited by each member of this transcription factor family.

In this chapter, *M.* × *domestica* plants have been modified to stably overexpress the two targets of Apple Proliferation phytoplasma effector $SAP11_{CaPM}$, i.e., the Class II CIN-like TCP transcription factors *MdTCP4a* and *MdTCP13a*. Such transgenic plants have been analysed and characterized, both *in vitro* and *ex vitro*, including a preliminary analysis of the transcriptome (still in progress) to shed light on the role of the protein in apple. Moreover, *MdTCP4a* overexpressing plants have been infected with the Apple Proliferation phytoplasma to test the hypothesis that the deactivation of TCP transcription factors by the bacterial effector SAP11 is a crucial step for the assessment of a successful infection.

RESULTS

Malus × *domestica* transformation

The full-length coding sequences of the *MdTCP4a* and *MdTCP13a* genes of *M.* × *domestica* (cv 'Golden Delicious') were subcloned in a binary vector for *in planta* stable expression under the control of the Cauliflower Mosaic Virus (CaMV) 35S promoter and used for transformation of *Agrobacterium tumefaciens*. For each of the two genes, three infection events with *A. tumefaciens* on approximately 1.000 leaf explants of *M.* × *domestica* cv 'Gala' were carried out.

After on average six months, a total of 11 regenerants for *MdTCP4a* overexpression were obtained and successfully micropropagated, while no living plants were obtained for the putative CaMV35S::*MdTCP13a*, where six plantlets emerged from callus but aborted the growth and died after approximately 20 days.

The 11 regenerant lines are named with a roman numeral, corresponding to the transformation event (I, III or V) and a capital letter that indicates an independent transgenic line. For two lines (IA, IB), multiple shoots originated from the same callus and a number of those was thus separately analysed as an independent transgenic line: a dot followed by a number indicates these.

DNA was extracted from the leaves of the putative transgenic plants and amplified using primers specific for the gene of interest and the CaMV 35S promoter region to test for the correct insertion of the fragment. To exclude the possibility of a false positive due to an *A. tumefaciens* contamination, an additional PCR on the bacterial gene *virG* was

performed on the same samples. All 11 the CaMV35S::*MdTCP4a* regenerants tested positive for the amplification of the CaMV 35s promoter – *MdTCP4a* region, and no *A. tumefaciens* contamination was detected, resulting in a transformation efficiency of 1.1%. These transgenic plants were thus further analysed.

Copy number determination

To determine the T-DNA copy number (CN), that is, the number of copies of 35S::*MdTCP4a* that were successfully integrated into the plant's genome, the protocol developed by Dalla Costa and colleagues (2019) was applied by performing a qPCR targeting the Kanamycin resistance marker gene *nptII* and using *MdTOPO6* as an endogenous reference gene. The results are displayed in Table 1. Eight out of 11 lines carry two copies of the gene, while the remaining five show a copy number of three or higher.

Table 1: Copy number determination, as described by Dalla Costa *et al*. (2019), of the 11 transgenic lines originated from the At-mediated transformation with a binary vector containing *MdTCP4a* full-length gene under the control of 35SCaMV.

IA.1	IA.2	IB.2	IB.3	IC	ID	IE	IIIA	IIIB	IIIC	IIID	VA	VB
2.2	2.1	2.2	2.7	2.1	2.3	3.5	3.3	3.2	2.4	3.4	2.3	1.9

Analysis of *MdTCP4a* relative expression *in vitro* and *ex vitro*

For further analyses, six lines representative of the observed variability in the copy number have been selected: the two showing extreme values (IE, VB), the two closest to the group mean (IB.3, IIIC) and the two closest to the first (IA.2) and third (IIIA) quartiles.

The relative expression of *MdTCP4a* of the selected transgenic plants was estimated via a quantitative PCR assay (qPCR): RNA was extracted from the leaves of the six lines growing *in vitro* and the greenhouse and used as the template for cDNA synthesis. The analysis was performed using two pairs of *MdTCP4a* specific primers (MdTCP4a_RT1/MdTCP4a_RT3) and two housekeeping genes, actin and ubiquitin, as reference (primers listed in Table 2). Expression of *MdTCP4a* in the transgenic lines

growing *in vitro* are displayed in the graph in Figure 1, normalized on non-transformed - 'Gala' expression: lines IA.2 and IE show an approximately two-fold overexpression compared to 'Gala', while IB.3 has nearly three times the expression of non-transformed plants. The remaining three lines display a similar (IIIC, VB) or lower (IIIA) expression compared to non-transformed 'Gala'. No correlation between the copy number and the relative expression levels was observed, as showed by the Pearson Correlation Coefficient calculated on the sample (R=0,1402).

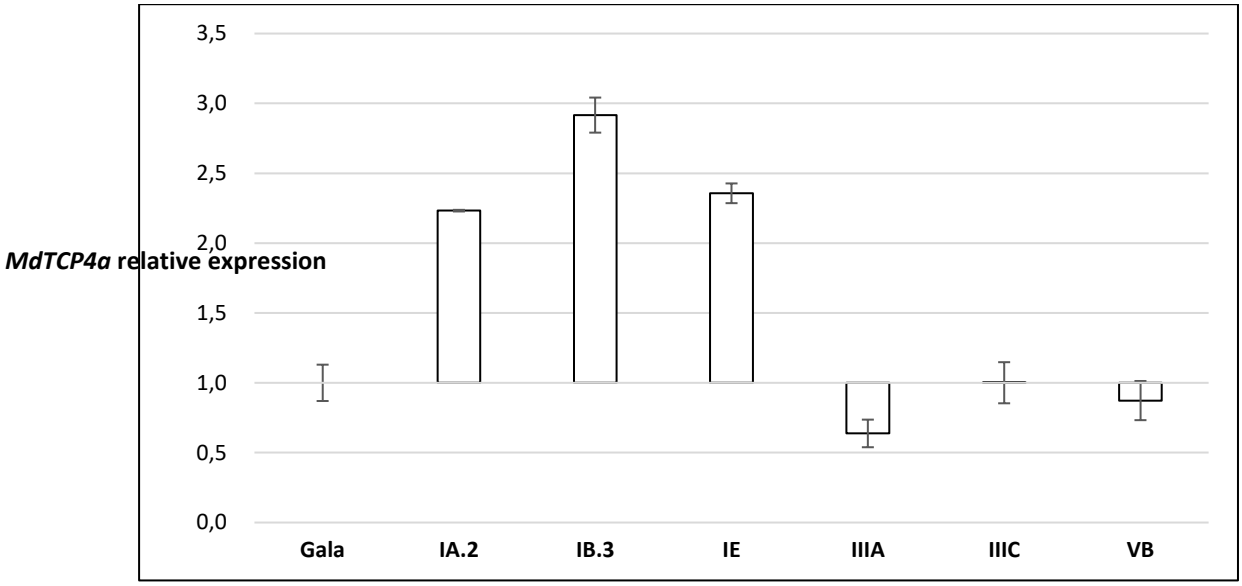

Figure 1: Relative expression of *MdTCP4a* in six independent 35S::*MdTCP4a* in vitro transgenic lines. Data are normalized on the expression of two endogenous genes (*MdActin* and *MdUbiquitin*) and relative to the average expression of a non-transformed 'Gala'. Values are calculated on technical triplicates and biological duplicates. Error bars indicate the standard deviation calculated accordingly.

The three lines displaying higher *MdTCP4a* expression than non-transformed 'Gala' (IA.2, IB.3, IE) were selected for further analyses. In addition, line IIIA, displaying a lower expression than non-transformed control, was included.

The four lines were thus acclimatized in soil, together with a non-transformed 'Gala' as control. One month after the complete acclimatisation to greenhouse conditions, the two youngest leaves of each line were collected and subjected to RNA extraction. A fraction of the total RNA was used as the template for reverse transcription to cDNA for an additional analysis of the *MdTCP4a* expression, with the same primers and setup as described above. The results are displayed in Figure 2, normalized on 'Gala' average expression measured.

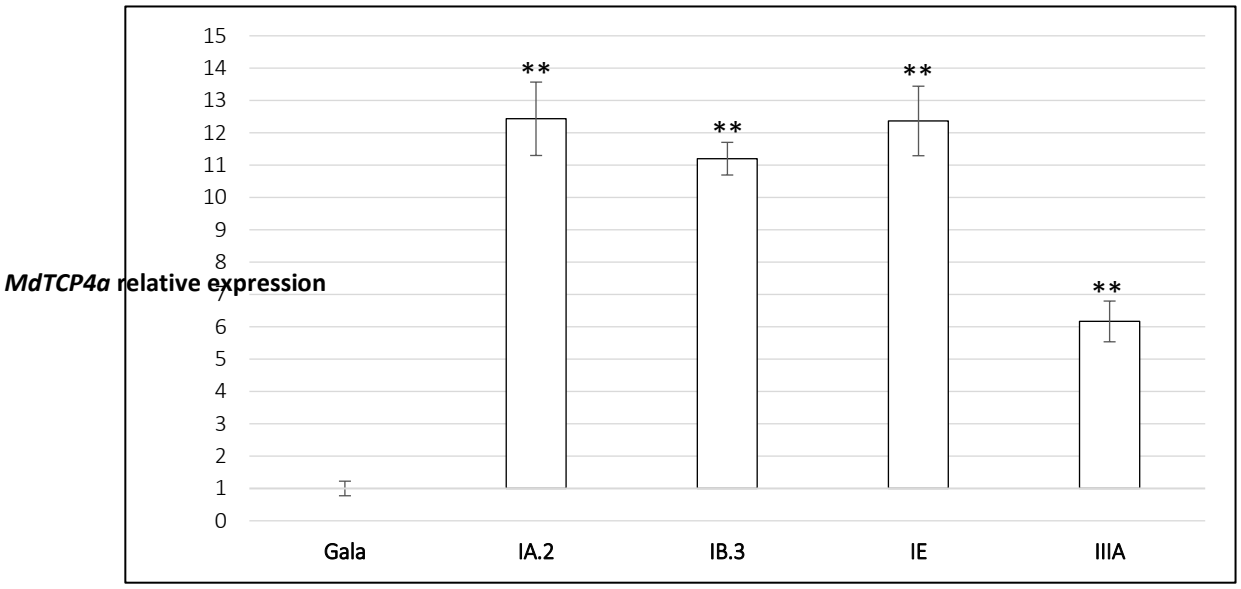

Figure 2: Relative expression of *MdTCP4a* in four independent 35S::*MdTCP4a* soil-acclimatised transgenic lines. Quantities are normalized on the average of non-transformed 'Gala' expression using two housekeeping genes (*MdUbi, MdAct*) as normalization factor, calculated on three technical replicates. Error bars indicate the standard error calculated on biological triplicates. Double asterisks indicate statistical significance ($p < .01$) by applying non-parametric Mann-Whitney U-test (Glover and Mitchell, 2008).

The relative expression of the transgenic lines appears to be noticeably higher compared to that determined for the *in vitro* plants: IA, IB and IE show more than a 10-fold increase of the *MdTCP4a* transcript compare to the non-transformed plants, while the line IIIA has on average roughly six times the 'Gala' expression. The difference of expression between transformed and non-transformed proved statistically significant by applying the non-parametric Mann-Whitney U-test ($p < .01$)

These results reflect on the plants' phenotype as well: while the *in vitro* transgenic plants did not display any evident phenotype after more than a year of micropropagation, short after the soil acclimatisation, the transgenic lines displayed a loss of apical dominance, smaller leaves and shorter stems compared to the non-transformed, as evidenced by the pictures in Figure 3. Interestingly, the manifestation of these phenotypic characteristics disappeared approximately six months after the first appearance.

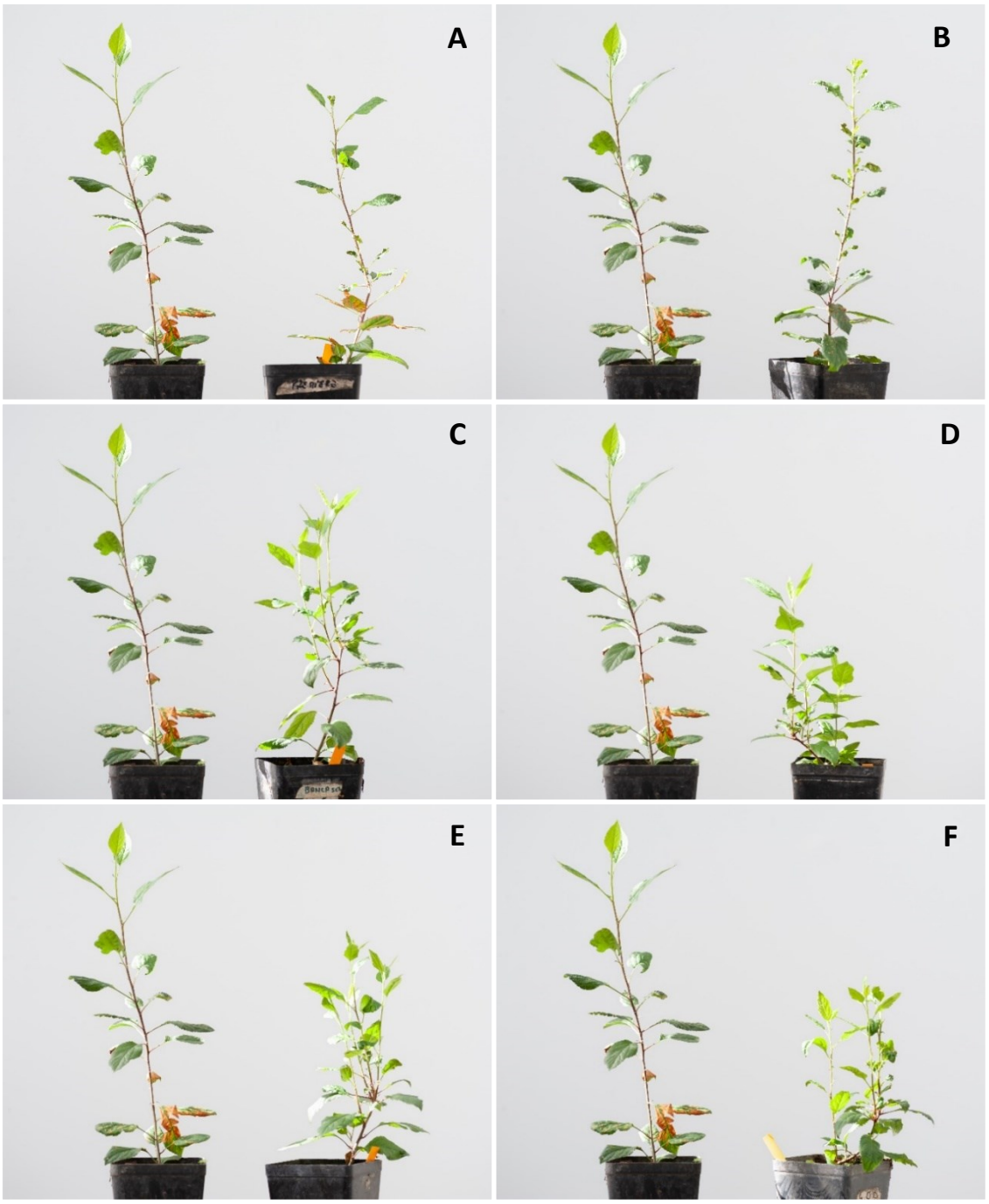

Figure 3: Phenotype characteristics of transgenic apple plants (cv 'Gala') overexpressing *MdTCP4a*. Pictures were taken two months after the soil acclimatisation. In each picture, a non-transformed 'Gala' plant is shown for comparison (left). A-B: line IA.2; C – D: line IB.3, E: line IE, F: line IIIA. Line IA.2 (A, B) shows smaller and crinkled leaves compared to non-transformed, while lines IB.3, IE, IIIA (C, D, E, F) show a loss of apical dominance and are smaller than the non-transformed.

Infection of transgenic lines with '*Ca.* P. mali '

The three lines IA.2, IB.3 and IIIA, and 'Gala' for comparison were infected with '*Ca*. P. mali' by *in vitro* micrografting with infected material of *M*. × *domestica*, cv 'Golden delicious', plants as described by Jarausch *et al*. (1999). A minimum of ten replicates per line were infected. Healthy plantlets with a stem diameter of 1 mm or more and a length of at least 3 cm were selected as rootstocks, and infected material with comparable dimensions as scions. After 40 days of grafts contact, scions were separated from the rootstocks, and the success of the grafts was visually evaluated and recorded. Only strong grafts, displaying formation of callus at the graft union and exerting physical resistance to traction, were considered for further analyses.

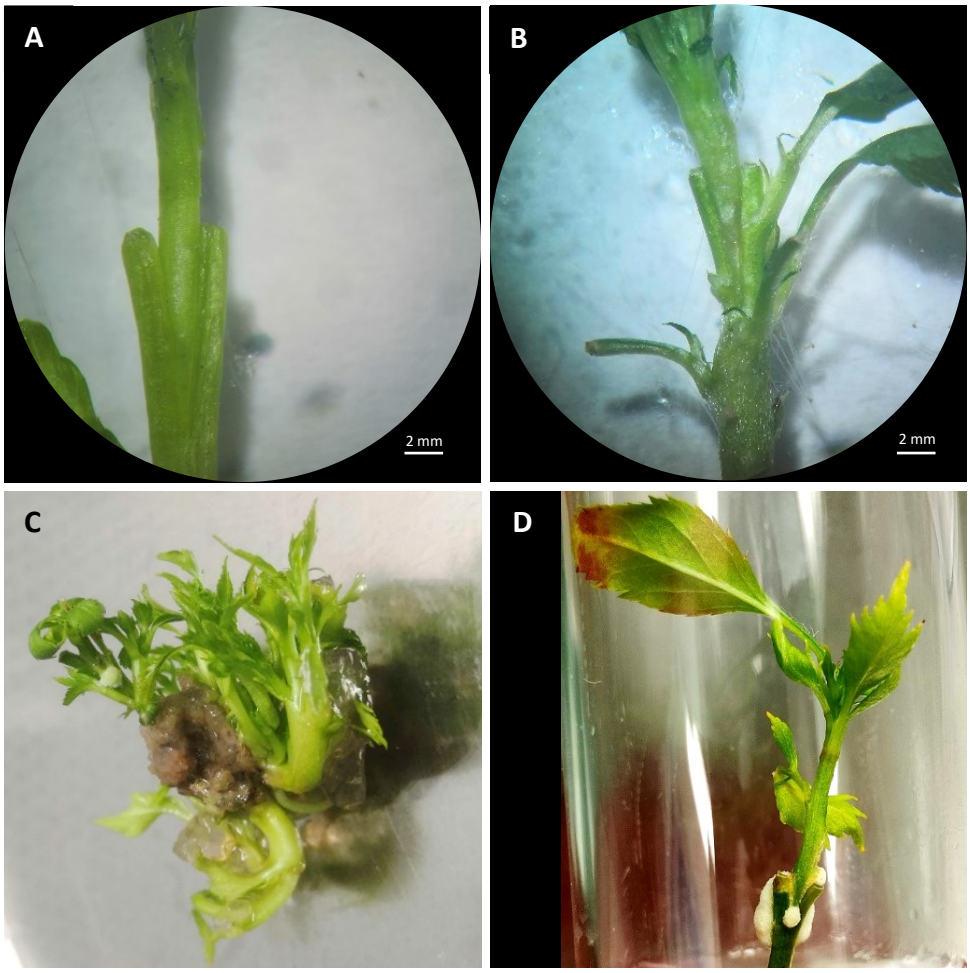

Figure 4: *In vitro* micrografting process. In A – B immediately after the micrografting, in C – D after 40 days. C: a whole pool, where the callus at the graft is visible. D: closeup of the graft point where the callus formation is evident.

Scions were analysed by quantitative PCR (qPCR) to confirm the presence of phytoplasma. The grafted plants generated from strong grafts were subsequently separated and partitioned into two/three subgroups each, and each subgroup analysed separately. All analysed pools tested positive for the presence of phytoplasma, resulting in an overall phytoplasma transmission efficiency of 25% (i.e. phytoplasma transmission was observed in 10 grafts out of 40 total grafts).

Finally, the two/three pools originating from two micrografting events per line were subcultured and propagated into single plants to generate a final set of 100 single infected plants (25 per line). After 30 additional days of culture (90 days overall after the *in vitro* micrografting was carried out), DNA was extracted from the aerial parts of the 100 plants and used in a qPCR assay to detect the presence of the phytoplasma.

To detect phytoplasma presence, a duplex qPCR assay was performed as described by Baric and Dalla Via (2004) by amplifying the phytoplasma 16S rRNA gene and the *M*. × *domestica* chloroplast gene coding for tRNA leucine. Of the 100 samples analysed, phytoplasma was detected in 95 samples. Of the five samples that tested negative for phytoplasma presence, three displayed a 16S signal later than the threshold cycle (set to 36th), one tested negative for the presence of plant DNA, and only one did not show any amplification of the phytoplasma gene.

The Cq values of target and reference, calculated on technical triplicate, were combined to calculate the ΔCq value for each plant and these values used to estimate the relative phytoplasma titer with the formula $x=2^{\Delta Cq}$ (Silver *et al*., 2006). The values obtained were then multiplied by 100000 to facilitate the data visualization. The final result for the three transgenic lines and the non-transformed control are illustrated in the box plot in Figure 5, calculated using average values of the pools as single points.

Results suggest that the transgenic line IA.2 does not show a significant difference in the phytoplasma concentration compared to the non-transformed, while on the contrary, lines IB.3 and IIIA display a lower quantity, which appears to be statistically significant.

A one-way ANOVA analysis with Tukey's Honest Significant Difference post hoc test was carried out to analyse differences between the bacterial load in the different groups. Differences between groups with $p < .05$ were considered significant.

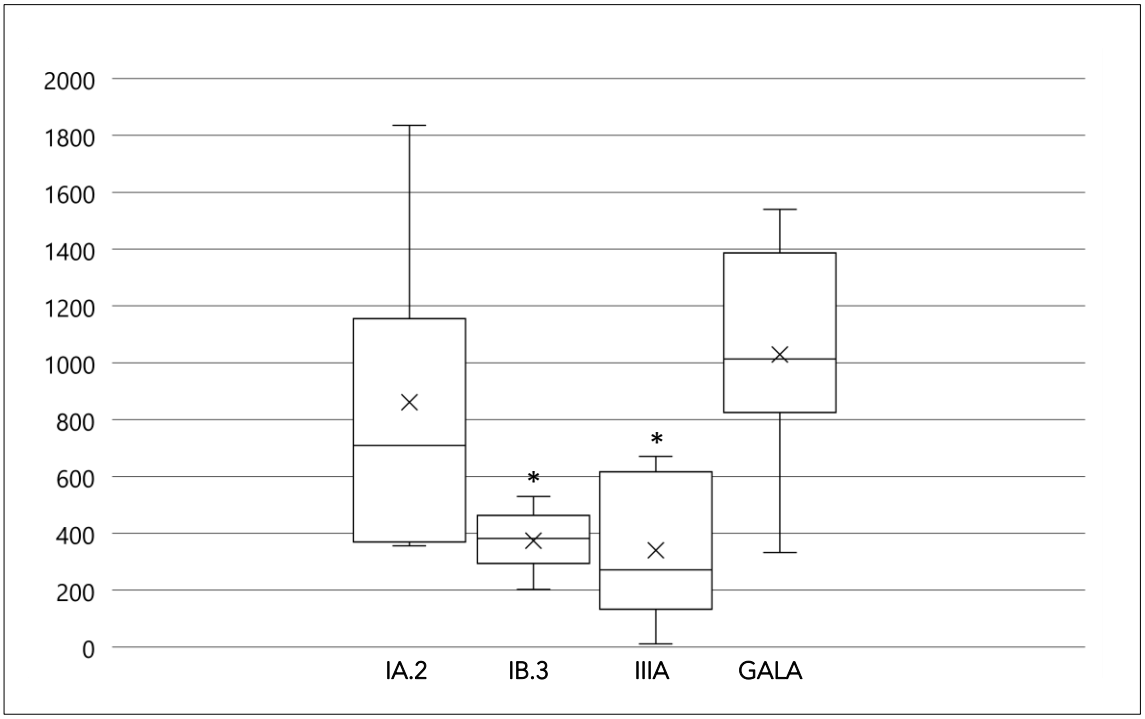

Figure 5: Box plot illustrating the '*Ca*. P. mali' titer measured for the three *MdTCP4a* overexpressing transgenic lines (IA.2; IB.3; IIIA) and the control (non-transformed 'Gala'). Phytoplasma detection was performed via qPCR three months after the micrografting. Boxes represent the distribution of first and third quartiles, while the error bars refer to the total distribution of non-outliers' samples. "X" and the horizontal lines inside the boxes indicate mean and median values, respectively. An asterisk indicates a statistically significant difference from the non-transformed as for one-way ANOVA analysis with Tukey's Honest Significant Difference post hoc test ($p < .05$)

To rule out that the normalization method with the multicopy chloroplast gene as a reference biased the phytoplasma quantification, an independent single-copy nuclear-encoded reference gene was used (Baric 2019). The ΔCq values of a 20 sample subset served as a proxy to determine the reliability of the normalization method. The aim was to verify that quantification with both methods gave the same results with a significant correlation of >95 %. In the scatterplot in Figure 6, the ΔCq values generated with the chloroplast-gene based quantification method (y-axis) are plotted against the respective ΔCq values generated with the primers amplifying the single-copy nuclear gene (x-axis).

The correlation between the ΔCq values generated with both quantification methods was determined by linear regression analysis (see trend line in figure 6). The correlation coefficient R^2 of 0,9795 (i.e. p<.05 for variants from the trend line) demonstrates that both methods provide similar results.

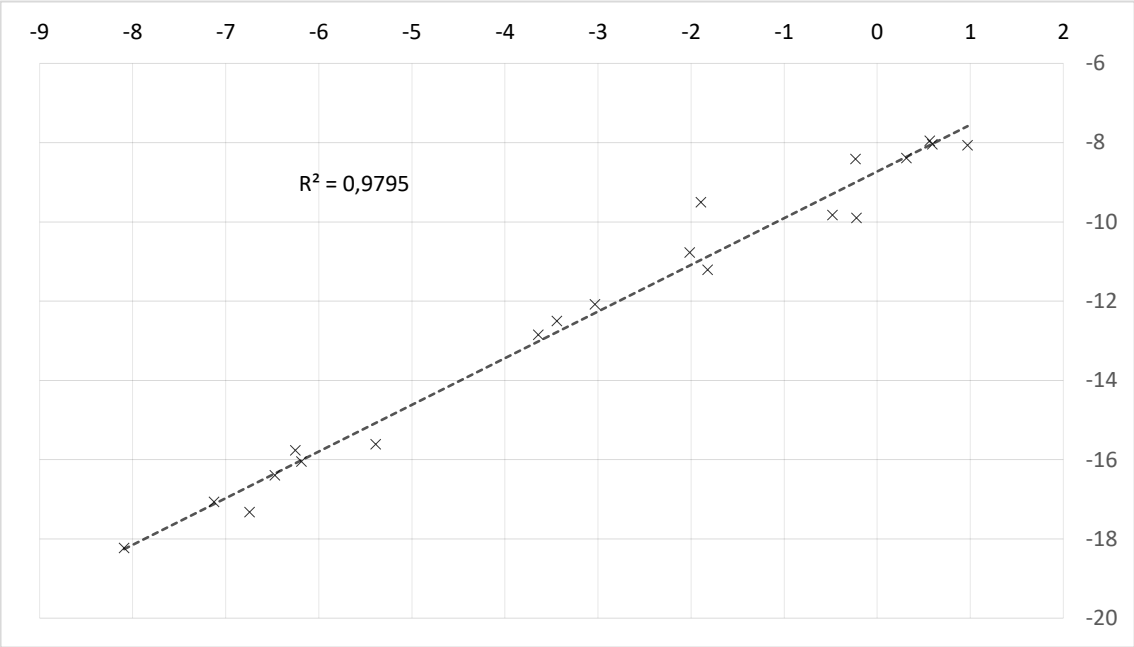

Figure 6: Scatterplot representing the correlation between relative phytoplasma quantity determined using a multicopy chloroplast - or single copy nuclear gene as reference. All the 20 samples are clearly following a linear pattern, consistent with the hypothesis that the use of either a nuclear or a chloroplast specific gene does not affect the final outcome of the analysis.

RNAseq: preliminary results

RNA extracted from the leaves of the three acclimatized transgenic lines IA.2, IE and IIIA and non-transformed 'Gala' was used to construct the RNA library and subsequent RNAseq analysis. After data filtration, it was determined that a total of 385 genes were up-regulated and 186 down-regulated in all three transgenic lines, whereas several others were up- or down-regulated only in one or between two transgenic lines (Figure 7).

Although data still need to be validated and thoroughly evaluated, some preliminary results relative to the plant metabolism are displayed in Figure 8. The three *MdTCP4a* overexpressing lines display up-regulation of several genes related to photosynthesis, photorespiration and antioxidant reactions compared to non-transformed 'Gala'. On the

contrary, pathways involving lipid degradation and secondary metabolism (relative to simple phenol, terpenes and phenylpropanoids) appear to be down-regulated in the transgenic lines. Moreover, many genes related to abiotic stress responses and redox state, as well as the stress-induced Ethylene Response Factor (ERF) transcription factors, display an up-regulation. Many other genes related to plant physiology and development, such as metabolism, ethylene signalling, biosynthesis and transcription factors, are instead down-regulated in *MdTCP4a* overexpressing plants (Figure 8).

Figure 7: Representation of the differentially expressed genes in the three transgenic lines compared to Gala resulting from the RNAseq analyses. In A the upregulated genes are reported; in B-downregulated genes. Numbers indicate the respective number of genes, indicated as differentially expressed by three lines, two, or one.

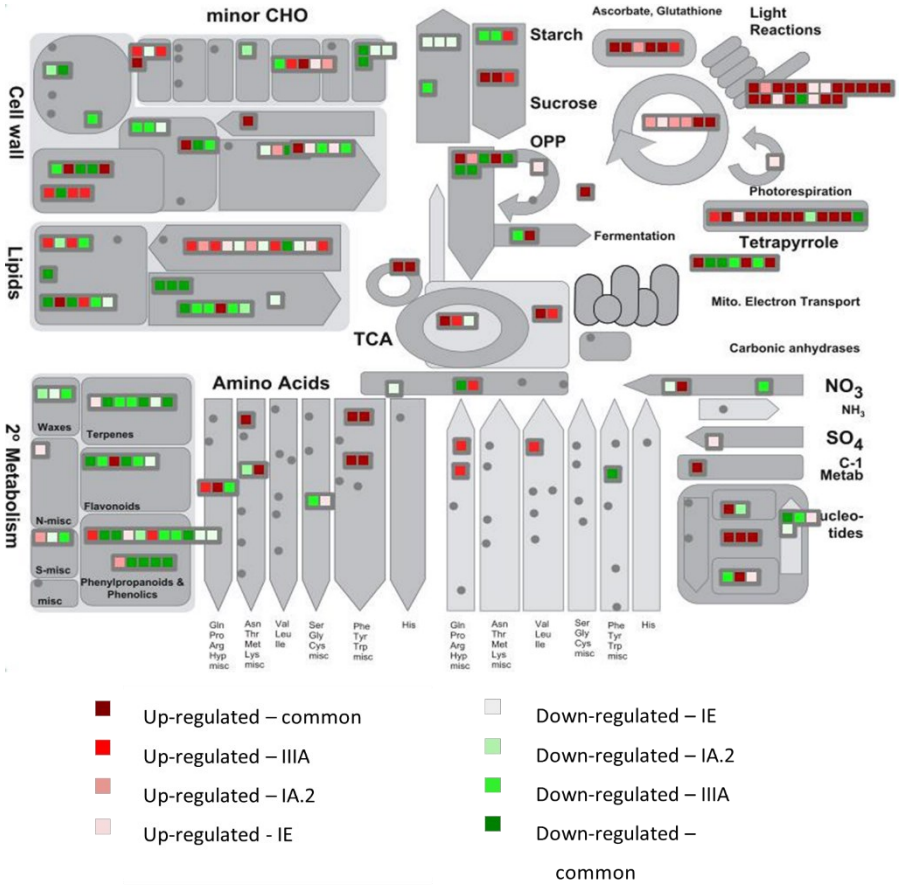

Figure 8: RNAseq preliminary data. Up- and down-regulation patterns of genes involved in plant metabolism pathways in transgenic 35S::*MdTCP4a* lines, compared to a non-transformed plant. Results are displayed as a MapGene output

DISCUSSION

Among the phytoplasma effector proteins, SAP11 is one of the best characterized. This effector, which targets Class II TCP transcription factors (Sugio *et al.*, 2011), is thought to induce many phenotypic changes that affect phytoplasma-infected plants, although direct molecular evidence is still lacking. The '*Ca*. P. mali' homolog of SAP11 binds two *M*. × *domestica* TCP transcription factors, *MdTCP4a* and *MdTCP13a* (Janik *et al*., 2017). Despite the central role of *TCP* genes in plant physiology and the economic importance of Apple Proliferation, little is known about the specific roles of these two transcription factors in apple. The present work aimed to generate transgenic apple plants overexpressing *MdTCP4a* and *MdTCP13a* to shed light on their role in apple physiology and disease development.

Apple transformation with the two *MdTCP* genes induced callus regeneration in both cases. However, no *MdTCP13a* overexpressing plantlet grew enough to be subcultured and died shortly after regeneration. It has recently been reported that transgenic *A. thaliana* overexpressing *AtTCP13* displays shorter rosette leaves. It was shown that *AtTCP13* is not expressed in actively dividing 10-days old leaves, and adults show a significant reduction in leaf cell size, especially at the cell expansion phase. Data indicate that *AtTCP13* inhibits leaf growth by acting as the upstream regulator of the *ATHB12* gene (Hur *et al.,* 2019), which promotes cell expansion of leaves (Hur *et al*., 2015). High activity of *AtTCP13*, particularly in the first phases of leaf growth, would prevent cell expansion and impede leaf development. An analogous process would be consistent with the early abortion of development and the subsequent death observed in *MdTCP13a* overexpressing plants.

A similar behaviour should be expected for *MdTCP4a* overexpressing plants since *AtTCP4* and *AtTCP13* belong to the same TCP subclass, whose members have similar, and sometimes redundant functions (Martin-Trillo and Cubas, 2010), nevertheless *MdTCP4a* overexpressing plants survived both *in vitro* and *ex vitro*. Even though *AtTCP4*, like *AtTCP13*, is thought to negatively regulate leaf growth, it was observed that the precocious activation of *AtTCP4* induces the generation of miniature leaves, while a delayed activation induces expanded leaves (Efroni *et al.* 2008), suggesting that the leaf size may be determined by spatial and temporal activation of *TCP*s (Chen *et al.*, 2020). These findings are consistent with the idea that, despite having similar functions, the precocious activation of *AtTCP4* and *AtTCP13*, and similarly of *M*. × *domestica* homologs, may have an opposite effect on the growth of young leaves and plantlets. Moreover, contrarily to *AtTCP13*, *AtTCP4* belongs to the group of CIN-Class II TCP tightly regulated by miR319. Experimental data relative to the overexpression of wild-type and microRNA-resistant *AtTCP* variants suggested that mRNA cleavage by means of microRNA is sufficient to restrict *AtTCP* function to their normal domain of activity (Palatnik et al. 2003). Despite never being experimentally demonstrated, an analogous process likely occurs in *M.* × *domestica* since both the *M.* × *domestica* miR319-homolog and the *MdTCP4a*-microRNA-binding domain show high nucleotide sequence

conservancy to *A. thaliana*. Unfortunately, due to the unavailability of *MdTCP4a*-specific antibodies, it was not possible to acquire information at the protein level. Similarly, the microRNA expression levels in transgenic lines were not investigated in this work.

Half of the 35S::*MdTCP4a* independent transgenic lines analysed showed a partial increase of the gene expression *in vitro* (2- to 3-fold increase compared to non-transformed), whereas the other half showed an expression comparable to the control or slightly lower (figure 1). None of the plants displayed any distinguishable phenotype change for the whole period of in vitro culture (approximately 18 months) compared to non-transformed. Following the *ex vitro* acclimatisation, *MdTCP4a* expression levels of transgenic lines showed instead a significant increase (figure 2), ranging from approximately 6- to 12-fold that of non-transformed expression. Soil-acclimatized transgenic plants also showed dramatic phenotypic changes during the first three months (figure 3), including smaller stems, loss of apical dominance, and small, crinkled leaves. Interestingly, these phenotype characteristics resemble some of the typical symptoms of Apple Proliferation, which include generation of small leaves, development of shoots from axillary buds, which give rise to secondary shoots that originate witches' brooms and, often, stunting (Schmid 1975; Seemüller 1990). The phenotype characteristics disappeared three months after the acclimatisation, indicating the establishment of a physiological condition after the disequilibrium induced by the acclimatisation process. This phenomenon is consistent with the hypothesis of a post-transcriptional tight regulation of *TCP* operated by miR319.

Interestingly, two out of three *in vitro* transgenic lines infected with '*Ca*. P. mali' showed a statistically significant lower concentration of the phytoplasma in the plant's aerial parts than non-transformed 'Gala'. The concentration of phytoplasma in the aerial parts of the plant is strongly correlated with the severity of the symptoms displayed by infected plants (Carraro *et al*., 2004). Additionally, genotypes are considered AP-tolerant in *in vitro* screening tests if characterized by a low concentration of bacteria (Jarausch *et al.*, 2008). Nonetheless, these promising results will need to be confirmed by

increasing the sample size and performing *ex vitro* infection screening, thus allowing the phenotype characterization of infected plants.

The preliminary analyses performed on RNAseq data indicate a pattern of differential expression of many pathways associated with plant development, such as WRKY transcription factors, and of many genes associated with abiotic stress responses, like AP2/ERF transcription factors. Data will be exhaustively analysed in the coming months to clarify the molecular basis of the phenotype observed in the transgenic plants and, possibly, the roles of *MdTCP4a* in plant physiology.

MATERIALS AND METHODS

Plant material

For *A. tumefaciens*-mediated transformation, *in vitro* shoots of *M.* × *domestica,* cv 'Gala' were used. Baby jars containing apple plantlets in shoot propagation medium (Pessina *et al.*, 2016) were maintained in a growth chamber at 24 ± 1°C with a 16h/8h light/dark period (100 mmol/m2/s). Regenerants were kept in the same conditions, and media was renewed every 30 days.

For acclimatisation, 2-week old *in vitro* shoots of the selected transgenic lines and non-transformed 'Gala' were initially transferred from propagation medium to a Murashige and Skoog medium supplemented with indole-3-butyric acid (Pessina *et al.*, 2016) to promote rooting and maintained in a growth chamber in the dark for 2 weeks, followed by 16h/8h light/dark conditions until the complete formation of roots. Rooted plants were acclimatized in soil by progressively reducing humidity for 3 weeks and subsequently maintained at greenhouse conditions: 16h/8h light/dark period at 24 ± 1°C, with a relative humidity of 70% ± 5%.

For the micrografting experiment, 4-week old shoots of the three selected transgenic lines and 'Gala' were grafted with 4-week old shoots of '*Ca*. P. mali' (strain AP5) infected *M.* × *domestica* cv 'Golden Delicious', following the procedure described by Jarausch *et al.* (2000). Before and after the separation from the scion, grafted plants were maintained in a growth chamber in the same conditions described above.

Plasmid preparation and At-mediated transformation:

Gateway™ pEntry™ vectors (Invitrogen™, Thermo Fisher Scientific, Waltham, Massachusetts, United States) containing the full-length sequences of *MdTCP4a* and *MdTCP13a* were subcloned in the Gateway™ binary vector pK7WG2.D via Gateway™ LR reaction and transformed in electrocompetent *E. coli* MegaX DH10B™ Electrocomp™ Cells following the manufacturer's protocol. Purified plasmids were then confirmed by Sanger sequencing with primers annealing on the CaMV 35S Promoter and Terminator regions (35SProm_fw/35Ster_rev, table 2) flanking the gene of interest and used to transform electrocompetent cells of *Agrobacterium tumefaciens* strain EHA105 (Hood *et al.*, 1993).

Plantlets of *Malus x domestica*, cv 'Gala', were transformed with *A. tumefaciens*, as described by Joshi *et al.* (2011). The transformation was performed in triplicates (I, III, V for *MdTCP4a*; II, IV, VI for *MdTCP13a*) using a total of approximately 1000 leaf explants for each replicate.

Regenerated plants, obtained on average 6 months after co-culture with *A. tumefaciens*, were screened to detect the presence of the T-DNA. Genomic DNA was extracted from approximately 20 mg of leaf material using the DNeasy™ Plant Mini Kit (Qiagen, Hilden, Germany), quantified on the NanoDrop 8000 Spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, United States), diluted to 5 ng/µL and used in the PCR reaction on the Thermocycler 3000 (Biometra, Göttingen, Germany), diluted in Green Master Mix GoTaq™ 1X (Promega, Madison, Wisconsin, United States) and primers MdTCP4a_fw/35STer_rev, 35SProm_fw/MdTCP4a_rev and VirG_fw/rev (0.4 µM), listed in Table 2.

Copy number determination

The determination of copy number of marker gene Neomycin phosphotransferase II (*nptII*), which confers kanamycin resistance to apple transformants, was evaluated to infer the number of T-DNA insertion events in the transgenic plants generated by At-mediated transformation using *MdTOPO6* as reference. Genomic DNA was extracted from 50 mg of fresh leaf tissue as described above, and the experimental procedure was conducted according to the TaqMan real-time PCR method described by Dalla Costa *et*

al. (2019). The qPCR reaction was performed in a 96 wells plate on a C1000 thermal cycler (Bio-Rad, Hercules, California, USA) equipped with CFX96 real--time PCR detection system (Bio-Rad, Hercules, California, USA), in 10 µl final volume containing 1 x SsoAdvanced™ Universal Probes Supermix (Bio-Rad, Hercules, USA), 15 ng of genomic DNA, 0.3 µM primers (Sigma, Haverhill, UK) and a 0.2 µM specific Taqman probe (Sigma, Haverhill, UK) with the following thermal protocol: 3' at 95°C, followed by 40 cycles of 10'' at 95 °C, 5'' at 58 °C, 5'' at 60 °C, 30'' at 72 °C. *MdTOPO6* and *nptII* primers and TaqMan probes used are listed in table 2. The standard curves for nptII and MdTOPO6 were built with MdTOPO6:nptII plasmid calibrator using five serial dilutions of 1:5 starting from 4 x 10^5 plasmid molecules. The total copies of nptII and *MdTOPO6* were calculated as mean values of the Cq of three technical replicates, and the nptII CN was calculated by the formula: $\frac{\text{nptII total copies}}{\text{MdTOPO6 total copies}} \times 2$

MdTCP4a relative expression determination

For the determination of the relative *MdTCP4a* expression of transgenic *in vitro* and soil acclimatised plants, the same protocol was applied: 50 mg of leaf material was collected and ground with liquid nitrogen in 1.5 mL microtubes and pestle, and the resulting powder was subjected to total RNA extraction using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, Missouri, United States) and eluted in DEPC-treated water. Total RNA was quantified with the Spectrophotometer NanoDrop 8000 (Thermo Fisher Scientific, Waltham, Massachusetts, United States) and treated with DNase I (Sigma-Aldrich, St. Louis, Missouri, United States) to remove DNA contamination. A total of 1 µg of RNA was then used as the template for cDNA synthesis with the Superscript III RT kit (Invitrogen™, Thermo Fisher Scientific, Waltham, Massachusetts, United States) with random primers following the manufacturer's instructions.

The resulting cDNA was diluted 1:10 in water, and 1 µL of each diluted sample was used to generate a total cDNA pool used to estimate the primers' efficiencies. The real-time reactions were conducted in a 96-well plate with 5 ng of cDNA diluted in SsoAdvancedTM Universal SYBR® Green Supermix (Bio-Rad, Hercules, USA) and the primers couples MdTCP4a_RT1 and MdTCP4a_RT2 for the gene of interest, MdUbiquitin and MdActin for housekeeping reference (primers listed in Table 2).

RNAseq

The raw data set from Illumina sequencing (averagely 320 M reads/sample) were analysed by the following pipeline: adapter sequences and low-quality reads (2-3%) were removed using Trimmomatic software (Bolger *et al*., 2014). Filtered reads were then mapped on the reference *M*. × *domestica* genome cv. 'Golden Delicious' GDDH13v1.1 with Hisat2 software (Kim *et al*., 2015). In total, between 88% and 90% of the reads could be mapped to the 'Golden Delicious' genome. The differentially expressed genes in the three transgenic lines against non-transformed were determined with R package EdgeR, with the following thresholds: FDR<0.05 and logFC>1 (up-regulation) or <-1 (down-regulation). Finally, differentially expressed genes were analysed with the Mapman software (Thimm *et al.,* 2004)

Analysis of phytoplasma-infected *in vitro* plants

For the analysis of phytoplasma infected plants, DNA was extracted from approximately 50 mg of stem material using DNeasy™ Plant Mini Kit (Qiagen, Hilden, Germany) as described above. DNA samples were analysed by applying a duplex qPCR protocol as described by Baric and Dalla-Via (2004), with AP16S (phytoplasma specific) and qMd-cpLeu (*M.* × *domestica* specific) primer pairs. Analysis was performed in technical triplicates in a 384-wells plate on a C1000 thermal cycler (BioRad, Hercules, California, USA) equipped with CFX96 real-time PCR detection system (Bio-Rad, Hercules, California, USA) and the data analysis software CFX Maestro (Bio-Rad, Hercules, California, USA).

Table 2: List of primers and probes used in the present chapter

PRIMER NAME	SEQUENCE (5' – 3')
MdTCP4a_fw	caccATGGGAATGAAGGGCTG
MdTCP4a_rev	TCAAAGCTGGCGGGTG
35S Prom_fw	GCTATCGTTCAAGATGCCTCT
35S Ter_rev	CACATGAGCGAAACCCTATAAG
VirG_fw	GCCGGGGCGAGACCATAGG
VirG_rev	CGCACGCGCAAGGCAACC
MdTOPO6_fw	TGTGGAAGGAGATCAAAGCGCA

MdTOPO6_rev	CGCGTTGCTTCTTTGCTGCA
MdTOPO6_Probe	*FAM*-ACATGCCAACAGGAACAATCACA-*TAMRA*
nptII_fw	CTTGCCGAATATCATGGTGGAA
nptII_rev	GGTAGCCAACGCTATGTCCTGA
nptII_Probe	*FAM*-TTCTGGATTCATCGACTGTGGC-*TAMRA*
MdTCP4a_RT1_fw	CTATGCACCCGATCCCCATC
MdTCP4a_RT1_rev	TCATTGCTCCATGCCACCAT
MdTCP4a_RT3_fw	ATGGTGGCATGGAGCAATGA
MdTCP4a_RT3rev	AACAACACCGTTCTCAGCCT
MdUbiquitin_fw	CATCCCCCCAGACCAGCAGA
MdUbiquitin_rev	ACCACGGAGACGCAACACCAA
MdActin_fw	TGACCGAATGAGCAAGGAAATTACT
MdActin_rev	TACTCAGCTTTGGCAATCCACATC
qAP-16S-F	CGAACGGGTGAGTAACACGTAA
qAP-16S-R	CCAGTCTTAGCAGTCGTTTCCA
qAP-16S	*FAM*-TAACCTGCCTCTTAGACG
qMd-cpLeu-F	CCTTCATCCTTTCTGAAGTTT
qMd-cpLeu-R	AACAAATGGAGTTGGCTGCAT
qMd-cpLeu	*VIC*-TGGAAGGATTCCTTTACTAAC
qMd-ACO-F	CCAGAATGTCGATAGCCTCGTT
qMd-ACO-R	GGTGCTGGGCTGATGAATG
qMd-ACO	*VIC*–TACAACCCAGGCAACG

REFERENCES

Aguilar-Martínez, J. A., Poza-Carrión, C., & Cubas, P. (2007). Arabidopsis Branched1 acts as an integrator of branching signals within axillary buds. *Plant Cell*. https://doi.org/10.1105/tpc.106.048934

Baric, S. (2019). Duplex TaqMan real-time PCR for rapid quantitative analysis of a phytoplasma in its host plant without external standard curves. In *Methods in Molecular Biology*. https://doi.org/10.1007/978-1-4939-8837-2_10

Baric, S., & Dalla-Via, J. (2004). A new approach to apple proliferation detection: A highly sensitive real-time PCR assay. *Journal of Microbiological Methods*. https://doi.org/10.1016/j.mimet.2003.12.009

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btu170

Carraro, L., Ermacora, P., Loi, N., & Osler, R. (2004). The recovery phenomenon in apple proliferation-infected apple trees. *Journal of Plant Pathology*. https://doi.org/10.4454/jpp.v86i2.949

Challa, K. R., Aggarwal, P., & Nath, U. (2016). Activation of YUCCA5 by the transcription factor TCP4 integrates developmental and environmental signals to promote hypocotyl elongation in arabidopsis. *Plant Cell*. https://doi.org/10.1105/tpc.16.00360

Challa, K. R., Rath, M., & Nath, U. (2019). The CIN-TCP transcription factors promote commitment to differentiation in Arabidopsis leaf pavement cells via both auxin-dependent and independent pathways. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1007988

Chen, H. W., Lee, P. L., Wang, C. N., Hsu, H. J., & Chen, J. C. (2020). Silencing of PhLA, a CIN-TCP gene, causes defected petal conical epidermal cell formation and results in reflexed corolla lobes in petunia. *Botanical Studies*. https://doi.org/10.1186/s40529-020-00300-7

Dalla Costa, L., Bozzoli, M., Pompili, V., Piazza, S., Broggini, G. A. L., Patocchi, A., & Malnoy,

M. (2019). Development of a Taqman real-time PCR method to quantify nptII in apple lines obtained with 'established' or 'new breeding' techniques of genetic modification. *European Food Research and Technology*. https://doi.org/10.1007/s00217-018-3187-0

Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., van Dijk, A. J., Muino, J. M., Cutri, L., Dornelas, M. C., Angenent, G. C., & Immink, R. G. H. (2012). Arabidopsis class I and class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. *Plant Physiology*. https://doi.org/10.1104/pp.112.200303

Efroni, I., Blum, E., Goldshmidt, A., & Eshed, Y. (2008). A protracted and dynamic maturation schedule underlies Arabidopsis leaf development. *Plant Cell*. https://doi.org/10.1105/tpc.107.057521

Fan, D., Ran, L., Hu, J., Ye, X., Xu, D., Li, J., Su, H., Wang, X., Ren, S., & Luo, K. (2020). miR319a/TCP module and DELLA protein regulate trichome initiation synergistically and improve insect defenses in Populus tomentosa. *New Phytologist*. https://doi.org/10.1111/nph.16585

Fang, Y., Zheng, Y., Lu, W., Li, J., Duan, Y., Zhang, S., & Wang, Y. (2020). Roles of miR319-regulated TCPs in plant development and response to abiotic stress. In *Crop Journal*. https://doi.org/10.1016/j.cj.2020.07.007

Hur, Y. S., Kim, J., Kim, S., Son, O., Kim, W. Y., Kim, G. T., Ohme-Takagi, M., & Cheon, C. I. (2019). Identification of TCP13 as an upstream regulator of ATHB12 during leaf development. *Genes*. https://doi.org/10.3390/genes10090644

Hur, Y. S., Um, J. H., Kim, S., Kim, K., Park, H. J., Lim, J. S., Kim, W. Y., Jun, S. E., Yoon, E. K., Lim, J., Ohme-Takagi, M., Kim, D., Park, J., Kim, G. T., & Cheon, C. I. (2015). Arabidopsis thaliana homeobox 12 (ATHB12), a homeodomain-leucine zipper protein, regulates leaf growth by promoting cell expansion and endoreduplication. *New Phytologist*. https://doi.org/10.1111/nph.12998

Janik, K., Mithöfer, A., Raffeiner, M., Stellmach, H., Hause, B., & Schlink, K. (2017). An

effector of apple proliferation phytoplasma targets TCP transcription factors—a generalized virulence strategy of phytoplasma? *Molecular Plant Pathology*, *18*(3), 435–442. https://doi.org/10.1111/mpp.12409

Jarausch, W., Bisognin, C., Peccerella, T., Schneider, B., & Seemüller, E. (2008). Development of resistant rootstocks for the control of apple proliferation disease. *Acta Horticulturae*. https://doi.org/10.17660/actahortic.2008.781.54

Jarausch, W., Lansac, M., Bliot, C., & Dosba, F. (1999). Phytoplasma transmission by in vitro graft inoculation as a basis for a preliminary screening method for resistance in fruit trees. *Plant Pathology*. https://doi.org/10.1046/j.1365-3059.1999.00326.x

Joshi, S. G., Schaart, J. G., Groenwold, R., Jacobsen, E., Schouten, H. J., & Krens, F. A. (2011). Functional analysis and expression profiling of HcrVf1 and HcrVf2 for development of scab resistant cisgenic and intragenic apples. *Plant Molecular Biology*. https://doi.org/10.1007/s11103-011-9749-1

Koyama, T., Furutani, M., Tasaka, M., & Ohme-Takagi, M. (2007). TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in Arabidopsis. *Plant Cell*. https://doi.org/10.1105/tpc.106.044792

Koyama, T., Sato, F., & Ohme-Takagi, M. (2017). Roles of miR319 and TCP transcription factors in leaf development. *Plant Physiology*. https://doi.org/10.1104/pp.17.00732

Li, D., Zhang, H., Mou, M., Chen, Y., Xiang, S., Chen, L., & Yu, D. (2019). Arabidopsis class II TCP transcription factors integrate with the FT-FD module to control FLoWERIng. *Plant Physiology*, *181*(1), 97–111. https://doi.org/10.1104/pp.19.00252

Li, S. (2015). The Arabidopsis thaliana TCP transcription factors: A broadening horizon beyond development. *Plant Signaling and Behavior*. https://doi.org/10.1080/15592324.2015.1044192

Lopez, J. A., Sun, Y., Blair, P. B., & Mukhtar, M. S. (2015). TCP three-way handshake: Linking developmental processes with plant immunity. In *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2015.01.005

Martín-Trillo, M., & Cubas, P. (2010). TCP genes: a family snapshot ten years later. *Trends in Plant Science*, *15*(1), 31–39. https://doi.org/10.1016/j.tplants.2009.11.003

Mukhtar, M. S., Carvunis, A. R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M. T., Pevzner, S. J., Donovan, S. E., Ghamsari, L., Santhanam, B., Romero, V., Poulin, M. M., Gebreab, F., Gutierrez, B. J., Tam, S., … Payne, T. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science*. https://doi.org/10.1126/science.1203659

Palatnik, J. F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J. C., & Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. *Nature*. https://doi.org/10.1038/nature01958

Palatnik, J.F., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., Carrington, J.C. & Weigel, D. (2007). Sequence and expression differences underlie functional specialization of Arabidopsis microRNAs miR159 and miR319. *Dev Cell*. 13(1):115-25. https://doi:10.1016/j.devcel.2007.04.012

Pessina, S., Angeli, D., Martens, S., Visser, R. G. F., Bai, Y., Salamini, F., Velasco, R., Schouten, H. J., & Malnoy, M. (2016). The knock-down of the expression of MdMLO19 reduces susceptibility to powdery mildew (Podosphaera leucotricha) in apple (Malus domestica). *Plant Biotechnology Journal*. https://doi.org/10.1111/pbi.12562

Sarvepalli, K., & Nath, U. (2018). CIN-TCP transcription factors: Transiting cell proliferation in plants. In *IUBMB Life*. https://doi.org/10.1002/iub.1874

Schmid, G. (1975). PROLONGED OBSERVATIONS ON SPREAD AND BEHAVIOUR OF PROLIFERATION DISEASE IN APPLE ORCHARDS. *Acta Horticulturae*. https://doi.org/10.17660/actahortic.1975.44.30

Schommer, C., Palatnik, J. F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E. E., Nath, U., & Weigel, D. (2008). Control of jasmonate biosynthesis and senescence by miR319 targets. *PLoS Biology*. https://doi.org/10.1371/journal.pbio.0060230

Seemüller, E. (1990). Apple proliferation. In A. L. Jones (Ed.), *Compendium of apple and pear diseases* (pp. 67–68).

Silver, N., Best, S., Jiang, J., & Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology*. https://doi.org/10.1186/1471-2199-7-33

Sugio, A., MacLean, A. M., Kingdom, H. N., Grieve, V. M., Manimekalai, R., & Hogenhout, S. A. (2011). Diverse Targets of Phytoplasma Effectors: From Plant Development to Defense Against Insects. *Annual Review of Phytopathology*, *49*(1), 175–195. https://doi.org/10.1146/annurev-phyto-072910-095323

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L. A., Rhee, S. Y., & Stitt, M. (2004). MAPMAN: A user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal*. https://doi.org/10.1111/j.1365-313X.2004.02016.x

Thum, K. E., Kim, M., Christopher, D. A., & Mullet, J. E. (2001). Cryptochrome 1, Cryptochrome 2, and Phytochrome A Co-Activate the Chloroplast psbD Blue Light–Responsive Promoter. *The Plant Cell*. https://doi.org/10.1105/tpc.010345

Vadde, B. V. L., Challa, K. R., & Nath, U. (2018). The TCP4 transcription factor regulates trichome cell differentiation by directly activating GLABROUS INFLORESCENCE STEMS in Arabidopsis thaliana. *Plant Journal*. https://doi.org/10.1111/tpj.13772

Vadde, B. V. L., Challa, K. R., Sunkara, P., Hegde, A. S., & Nath, U. (2019). The TCp4 transcription factor directly activates Trichomeless1 and 2 and suppresses TRICHOME INITiation1[Open]. *Plant Physiology*. https://doi.org/10.1104/pp.19.00197

Wei, B., Zhang, J., Pang, C., Yu, H., Guo, D., Jiang, H., Ding, M., Chen, Z., Tao, Q., Gu, H., Qu, L. J., & Qin, G. (2015). The molecular mechanism of SPOROCYTELESS/NOZZLE in controlling Arabidopsis ovule development. *Cell Research*. https://doi.org/10.1038/cr.2014.145

Zhou, Y., Zhang, D., An, J., Yin, H., Fang, S., Chu, J., Zhao, Y., & Li, J. (2018). TCP

transcription factors regulate shade avoidance via directly mediating the expression of both PHYTOCHROME INTERACTING FACTORs and auxin biosynthetic genes. *Plant Physiology*. https://doi.org/10.1104/pp.17.01566

GENERAL CONCLUSIONS AND FUTURE PERSPECTIVES

In this work, two *Malus* × *domestica* TCP transcription factors (*MdTCP*) were analyzed in detail regarding the sequence and functional differences and their role during apple proliferation (AP) disease. This disease is associated with the bacterium '*Candidatus* Phytoplasma mali', and the two TCP transcription factors are the principal molecular targets of the phytoplasma effector protein $SAP11_{CaPM}$.

In the second chapter, the TCP gene family of *Malus* × *domestica* has been updated based on the latest high-quality genome assembly. This analysis made it possible to identify three novel *MdTCP* genes and discard several redundant or non-TCP sequences. This study highlights that many MdTCPs share a very high degree of reciprocal similarity and likely originated from a recent genome duplication, underlining the importance of this evolutionary phenomenon in the generation of *TCP*s. Moreover, the *MdTCP* nomenclature has been revised, and an *A. thaliana TCP*-homology based nomenclature is proposed. Thus, this study encourages adopting a homology-based nomenclature for the TCPs and thus avoids confusion when discovering novel *TCP* genes in other species in future. Given the increasing amount of genomic data and the growing interest in *TCP* genes, creating an unambiguous and non-redundant database of TCP sequences is of high importance for future research in this sector.

The third chapter aims to shed light on the molecular basis of AP-tolerance, observed in several *Malus sieboldii*-based experimental rootstocks. Thus, the sequences of the two target TCPs in tolerant genotypes have been analysed, and several non-synonymous nucleotide substitutions have been found. In particular, *MdTCP4a* of tolerant genotypes exhibits three single nucleotide variations translating into two rare amino acid substitutions. The analyses of several other tolerant varieties showed a good degree of correlation between the presence of amino acid substitution and the tolerant behaviour. The assays set up to test the interaction strength between the bacterial effector $SAP11_{CaPM}$ and the MdTCP4a protein variant of tolerant *Malus* accessions show no detectable differences. Thus, the results do not indicate that AP tolerance is based

on a decreased affinity of the bacterial effector for the host target proteins. However, it cannot be excluded that downstream effects following the SAP11$_{CaPM}$-MdTCP4a interaction are altered and involved in the differential susceptibility versus tolerance phenotype. The extensive research carried out in the last years to develop and identify AP-tolerant rootstocks has been hindered by several challenges, such as the impossibility of defining simple sequence repeat markers due to the high level of apomixis (i.e., production of offspring genetically identical to the mother) of *M. sieboldii*. Thus, the identification of a reliable marker would significantly reduce the amount of work and time required to develop AP-tolerant plants. In this sense, the two *MdTCP4a* amino acid substitutions identified in the present work are promising, but their reliability as markers of AP tolerance needs to be tested in future, mainly by expanding the number of genotypes analysed.

The fourth chapter describes the generation and subsequent characterization of transgenic *M.* × *domestica* plants overexpressing the two *MdTCPs* of interest via *Agrobacterium tumefaciens*-mediated transformation. While no *MdTCP13a*-overexpressing transgenic plants could be obtained, several independent lines of 35S::*MdTCP4a* have been successfully generated. Interestingly, in soil acclimatised transgenic plants displayed a peculiar phenotype, characterized by loss of apical dominance and increase of shoot branching, together with smaller leaves compared to non-transformed. This phenotype resembles some of the symptoms displayed by AP-infected apple plants and suggests that some of the visible symptoms of the phytoplasma infection are likely to be caused by TCPs misfunctioning.

'*Ca*. P. mali' infection tests, performed by grafting the transgenic plants with infected material *in vitro*, showed that transgenic plants could be infected, although the concentration of the bacteria appears to be lower than in the non-transformed control plants. These interesting results need to be further confirmed by sampling different time points and increasing the number of replicates, which will be done in the near future. The preliminary results of the transcriptome analyses performed on the transgenic plants appear consistent with the phenotype observed, showing a differential expression of several pathways associated with the plant architecture development.

Moreover, there seems to be a different expression of pathways involved in abiotic stresses response and redox reactions. It is worth noting that plants recovered from AP show accumulation of H_2O_2; thus, this aspect will be further investigated. The following steps will include an extensive evaluation of the RNAseq data and, consequently, the validation of selected gene data by qPCR assay.

The transgenic plants presented in this work were generated for research purposes to shed light on the roles of *MdTCP4a* in healthy plants and the AP infection process. The infection tests that were conducted provided interesting but ultimately inconclusive results. Although these plants could be better characterised in the future, the research was not focused on a possible field application, also in light of the strict European legislation regarding GMOs.

ACKNOWLEDGEMENTS

The last three years have been intense and rich in satisfaction, but also difficulties and, sometimes, frustration. I am happy and proud to have completed this journey, but I am well aware that I could never have done it alone. My doctorate, which culminates in this manuscript, required the contribution of many people to whom I am sincerely grateful.

Starting with the more "technical" aspects, I would like to thank Dr Claudio Varotto and Dr Mingai Li for RNA library construction, Dr Matteo Buti and Dr Federico Martinelli for the RNAseq data analysis. Prof. Qingmei Guan and Zhonging Li for allowing me to access unpublished SNP-calling data.

Time for FEM people: firstly, I would like to thank my supervisor at FEM, Dr Mickael Malnoy, for allowing me to live this extraordinary experience. Thanks to Dr Pier Luigi Bianchedi and Dr Mirko Moser for providing the in vitro plant material and helping me with the micrografting process but, above all, for being constantly available whenever I needed advice or help. Dr Diego Micheletti, for the patience in helping me with and, often, teaching me the rudiments of bioinformatics. Also, thanks to Dr Massimiliano Trenti, who worked hard during pandemic times to obtain many successful grafting. I would like to thank my FEM lab "superiors", Dr Umberto Salvagnin, Dr Stefano Piazza, Dr Lorenza Dalla Costa and Dr Valentino Giarola, for the enormous amount of patience, support and laughs. Then, thank to my PhD and master student mates: Valerio Pompili, Giulia Pasqualetto, Carlotta Pirrello, Assunta Russo, Ayesha Yousaf, Michele Pedrotti, Riccardo Bottiroli (et al.) for all the great memories I will carry with me forever.

Moving on to Laimburg, I want to thank Christine Kerschbamer, who has been kind and extremely helpful on countless occasions, and all the current and older Molecular Biology staff members for creating a stimulating, motivating and ultimately happy work environment: Dr Thomas Letschka, Dr Elena Zini, Dr Valentina Cova; Evi Klammsteiner, Cecilia Mittelberger, the already mentioned Dr Massimiliano Trenti, Dana Barthel, Dr Hannes Schuler, Dr Bernd Panassiti. Finally, my supervisor at Laimburg, Dr Katrin Janik, who has been the best guide and example I could have hoped for.

I would like to end with the people who allowed me to cheerfully live outside the laboratory, starting with my dearest friends - in a strict alphabetical order - Luca, Mattia, Michael, and Ronny.

My mother, Daniela, without whom I would never have been able to pursue this career.

My big sister, Jessica, and my big brother, Ivan, for successfully managing never to make me feel alone.

Finally, I would like to thank Marika, who keeps holding my hand on the brightest day and the darkest night.