Molecular aspects of columnar growth in apple

Pieter Jacobus Wolters

Thesis committee

Promotor
Prof. Dr Evert Jacobsen
Professor of Plant Breeding (Genetical Variation and Reproduction)
Wageningen University

Co-promotors
Dr Paolo Baldi
Senior Scientist, Genomics and Biology of Fruit Crops
Fondazione Edmund Mach, Michele All'adige, Italy

Dr Henk J. Schouten
Senior Scientist, Plant Research International
Wageningen UR Plant Breeding

Other members

Prof. Dr Harro J. Bouwmeester, Wageningen University
Prof. Dr Dolf Weijers, Wageningen University
Dr René Geurts, Wageningen University
Dr Jeremy Harbinson, Wageningen University

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences

Molecular aspects of columnar growth in apple

Pieter Jacobus Wolters

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 7 October 2014
at 4 p.m. in the Aula.

Pieter Jacobus Wolters
Molecular aspects of columnar growth in apple,
168 pages.

PhD thesis, Wageningen University, Wageningen, NL (2014)
With references, with summaries in Dutch and English

ISBN 978-94-6257-084-9

Table of contents

CHAPTER 1	**General introduction**	**7**
CHAPTER 2	**Genetic and physical characterisation of the locus controlling columnar habit in apple (*Malus x domestica* Borkh.)**	**25**
CHAPTER 3	**Evidence for regulation of columnar habit in apple by a putative 2OG-Fe(II) oxygenase**	**51**
CHAPTER 4	**Genomic characterisation of the 'Wijcik' mutation**	**81**
CHAPTER 5	**Functional analysis of *MdCo31***	**105**
CHAPTER 6	**General discussion**	**133**
	Summary	**151**
	Samenvatting	**155**
	Acknowledgements	**159**
	About the author	**161**
	List of publications	**163**
	Education statement of the Graduate School of Experimental Plant Sciences	

Chapter 1

General introduction

Introduction

As rich sources of flavonoids, carotenoids, vitamin C, folic acid and dietary fibre, fruits and vegetables can help prevent cardiovascular diseases and cancers (Amine et al. 2002). According to the world health report that was published by the WHO (2002), 2.7 million deaths per year can be attributed to low fruit and vegetable intake.

Apple is the most common fruit crop grown in temperate regions (FAOSTAT 2012) and apples are affordable and widely available in most parts of the world. Apple and apple-derived products have been subject of many studies that have linked their intake to a beneficial effect on many diseases (Hyson 2011). Several studies have indicated a positive effect of phenolic compounds, which are abundantly present in apple, on preventing coronary heart disease and cancer (Hertog et al. 1993; Keli et al. 1996; Gerhauser 2008). The dietary fibres that are present in fruits like apple seem to decrease the risk for developing coronary heart disease, stroke, hypertension, diabetes, obesity and some gastrointestinal diseases (Anderson et al. 2009). Apples have shown to have high anti-oxidant activity and the intake of apple is associated with a reduced risk for heart disease, lung cancer, type 2 diabetes, and asthma (Boyer and Liu 2004), substantiating the beneficial effects of apple consumption on human health.

Quality traits of apple

Besides the health-beneficial properties, a number of other quality traits should be considered when breeding apple trees. First, traits that influence the appeal to the consumer are important. These traits concern properties of the fruit, like appearance, flavour, fruit composition, texture and shelf life. Secondly, for apple growers, the characteristics of the tree are also important. These are traits like tree architecture, alternate bearing, plant vigour and resistance to diseases and abiotic stresses.

Tree architecture is a trait that influences many aspects of apple production. The right tree shape is necessary to optimise light capture on the scale of the apple orchard, but also to facilitate the light exposure of the apple fruits. Maintaining the right tree shape will reduce fruit variation and direct the conversion of energy into fruits rather than into vegetative structures (Costes et al. 2006).

Moreover, tree architecture has a direct influence on planting density, fruit yield and pruning requirements (Kenis and Keulemans 2007).

Use of rootstocks in apple cultivation

During the domestication of apple, quality traits were initially selected for in seedlings. More than 3800 years ago, grafting was discovered and it became an important technique in the domestication of apple (Cornille et al. 2012). Using this propagation technique, it is a lot easier to select for and to control desired traits, as the genotypes can be preserved despite self-incompatibility (Harris et al. 2002). Nowadays, all commercial apple cultivars are propagated vegetatively and grafted onto rootstocks. Only new apple cultivars originate from crossings or are derived from mutants.

Besides facilitating asexual reproduction, rootstocks are also being used to directly influence properties of the scion. Rootstocks can influence the size of the tree, winter hardiness and resistance to pathogens and abiotic stresses (Pereira-Lorenzo et al. 2009). One of the most commonly used apple rootstocks is Malling 9 (‘M9’) (Zhu et al. 2001). Scions grafted onto this rootstock are 25-35% reduced in size, giving rise to dwarf trees that are well-suited for high density apple orchards (Zhu et al. 2001).

‘Wijcik’ apple and apple cultivation

An interesting alternative way to control the growth of apple trees was proposed by Tobutt in 1985 and makes use of the columnar growth habit of a mutant of the apple cultivar ‘McIntosh’ called ‘Wijcik’ (Tobutt 1985). ‘Wijcik’ is a somatic mutant that was identified by the apple grower Anthony Wijcik in 1963 as a shoot on top of a normal ‘McIntosh’ tree (Fisher 1969; Petersen and Krost 2013).

‘Wijcik’ trees show very little lateral branching, as their axillary buds develop into spurs (that bear fruits) rather than lateral branches. The trees have short internodes and a thick stem (Fisher 1969; Tobutt 1985). Because of the distinct compact nature of the trees, ´columnar´ was distinguished as one of the four fruiting types of apple that were proposed by Lespinasse (Lespinasse 1992). Columnar trees need very little pruning and could be planted close together, resulting in a potential yield increase (Tobutt 1985). The columnar trait was shown

to be caused by a single dominant locus (*Co*) (Lapins 1976) that maps on chromosome 10 (Conner et al. 1997).

Despite the potential benefits of using columnar apple trees in high density orchards, their use has so far largely been restricted to ornamental purposes. There are a number of difficulties related to apple breeding in general that have complicated the introduction of the columnar trait in commercial apple cultivars, hindering their evaluation in these commercial systems. These difficulties are discussed below.

Apple breeding challenges

Apple is highly heterozygous (Velasco et al. 2010), making the outcome of a cross between two apple cultivars unpredictable. Incompatibility and the limited number of S-alleles available in commercial apple cultivars reduce the number of compatible parents (Sakurai et al. 2000).

Apple breeding is further complicated by the long generation time and the space required for maintaining progeny trees. This makes the introduction of new traits in commercial apple cultivars challenging, even more so when wild apple species are involved in the cross. Due to genetic drag, it is necessary to perform many crosses to get rid of the unwanted alleles from the wild species that are introduced in the progeny. An illustrative example is the introgression of the dominant gene for scab resistance from wild *M. floribunda* (*HcrVf*), that took over 80 years (Joshi et al. 2009).

Trees derived from ´Wijcik´ suffer from the genetic drag of unwanted alleles as well, leading to undesirable characteristics such as biannual bearing, low sugar content, soft fruit flesh and susceptibility to apple scab. Some of these traits appear to be tightly linked to the *Co* locus (Tobutt 1994; Kenis et al. 2008; Moriya et al. 2009).

Marker assisted selection

In order to facilitate apple breeding, molecular techniques have been introduced in apple breeding programs. The first detailed genetic map comprising all 17 chromosomes of apple was constructed by Maliepaard et al. in 1998 (Maliepaard et al. 1998). Since then, many more genetic maps have been

constructed, which enabled researchers to map agronomically important traits to the apple genome (Keller-Przybyłkowicz and Korbin 2013). These mapping experiments have resulted in the identification of genetic loci responsible for diverse traits, such as disease resistance, self-incompatibility, root suckers, fruit skin colour, fruit flesh colour, pH, fruit allergens and ethylene production (Marić et al. 2010).

The first molecular marker for the columnar trait was the SSR marker (SSR^{Co}) developed by Hemmat at al. in 1997 (Hemmat et al. 1997). Nine additional markers were identified by Tian et al., including two SSRs markers (CH3d11 and COL) and two SCAR markers (SCAR216 and SCAR682) that delimit the *Co* locus. The closest marker to *Co* that was found is the RAPD marker S14412600, at a distance of 1.9 cM (Tian et al. 2005). Other markers developed are $WB82_{670}$ (Kim et al. 2003), Hi01a03 (Silfverberg-Dilworth et al. 2006; Moriya et al. 2009) and UBC8181000 (Zhu et al. 2007).

The extensive genetic maps that are now available for apple make it possible to map quantitative trait loci (QTLs) as well. Complex traits, such as fruit texture physiology (Longhi et al. 2012) and phenolic compound content (Khan et al. 2012), have been mapped in QTL studies in apple.

The ultimate aim of these genetic mapping studies is to identify the causative genes underlying phenotypes. With the completion of the apple genome sequence in 2010 (Velasco et al. 2010), a powerful new tool became available for developing new molecular markers and to find putative candidate genes for the studied traits. This tool has proven to be essential for the fine mapping studies targeting the *Co* region that were performed in subsequent years by Bai et al., Moriya et al. and Baldi et al. (Bai et al. 2012; Moriya et al. 2012; Baldi et al. 2013)

In the last years, the knowledge about the genetic basis for agronomically important traits is steadily increasing. This genetic information can be applied in breeding programs, for example using marker assisted selection (MAS), to greatly increase the precision and efficiency of these programs (Dirlewanger et al. 2004). MAS could help breeders to pick the right parents for crossings, to plan the subsequent backcrosses and to select promising progeny plants at an early stage, based on genetic rather than phenotypic evaluation (Collard and Mackill 2008).

MAS is currently being applied in many apple breeding programs worldwide (Zhu and Barritt 2008; Folta and Gardiner 2009; Luby et al. 2009; Patocchi et al. 2009).

Although MAS is a promising tool for speeding up apple breeding, the long juvenile period of apple remains a limiting factor. The period from sowing to seeds can take from 5 to 12 years (Fischer 1994). Reducing the juvenile phase of apple would result in a shortened generation cycle.

Flachowsky et al. showed that overexpression of the *bpMADS4* gene from silver birch in apple resulted in plants that flowered within a few months instead of a couple of years (Flachowsky et al. 2007). Also, transgenic apple trees overexpressing *MdFT1* from apple showed an early flowering phenotype (Tränkner et al. 2010). In combination with marker assisted selection, these approaches could significantly speed up apple breeding programs (Flachowsky et al. 2011).

Genetic engineering

If the genes underlying the desired traits are known, an alternative to classical breeding, or breeding using MAS, would be to make use of genetic engineering. Using this technique, genes of interest can be transferred directly from one organism to another by transferring the genes causing the traits, without the need for performing crosses. Because only these preferred genes are transferred, the genetic drag of unwanted alleles is avoided. In this way, specific traits can be targeted while leaving the rest of the plant unchanged.

Transgenic apples have been developed for research purpose since the development of the first transgenic apple in 1989 by James et al. (James et al. 1989), but transgenic apples have not yet been a commercial success. Recently, ‘Arctic apples’ were developed by Okanagan Specialty Fruits Inc. In these apples, the polyphenol oxidase (*PPO*) gene is silenced, using RNA interference, using a form of genetic engineering (Saurabh et al. 2014). The silencing of *PPO* leads to a reduction of bruising and browning of the fruits (Armstrong and Lane 2013). ‘Arctic apples’ are currently under review by the U.S. Department of Agriculture (USDA) to get approval to enter the US market. However, there are some controversies surrounding genetic engineering, with consumers raising concerns about food safety, environmental effects and ethical implications of using this technique.

Nielsen proposed to distinguish between different forms of transgenic organisms, based on the source of the introduced DNA. He introduced the term 'intragenic' and 'famigenic' for genetically engineered organisms that carry DNA from sexually compatible species, so DNA that could have been introduced to the organism through conventional breeding as well (Nielsen 2003). According to Nielsen, distinguishing between different categories of genetically modified organisms (GMOs) could improve the opinion of consumers about GMO products, as many concerns about GMOs have to do with 'foreign' DNA being introduced in the organism.

Following a similar line of thought, the term cisgenic was introduced by Jochemsen and Schouten (Jochemsen 2000). Cisgenesis was defined as "*the genetic modification of a recipient plant with a natural gene from a crossable-sexually compatible-plant*". "*The gene includes its introns and is flanked by its native promoter and terminator in the normal sense orientation"*. No 'foreign' gene is present in a cisgenic plant (Schouten et al. 2006a). Schouten et al. argue that the risks connected to the cultivation and breeding of cisgenic plants are not greater than the risks related to conventionally bred plants. Therefore, cisgenic crops should not be treated differently from these conventionally bred crops (Schouten et al. 2006a; Schouten et al. 2006b), a view that was adopted by The European Food Safety Authority (EFSA) in a recent report that was published in the EFSA journal (Andersson et al. 2012).

Columnar-like phenotypes in diverse plant species

Most traits related to tree architecture are under polygenic control and are therefore investigated in QTL studies (Folta and Gardiner 2009). However, some genes and major QTLs controlling plant architecture have been identified in fruit trees and other plants species. Some examples are given below.

The major locus controlling dwarf growth habit of rootstock 'M9' of apple was mapped on chromosome 5 (Pilcher et al. 2008), that has homology with chromosome 10 that also contains *Co*. However, the phenotypes from 'M9' and 'Wijcik' are not the same and inheritance of the dwarfing phenotype is also different (Pilcher et al. 2008).

Overexpression of the LEAFY (*LFY*) gene from Arabidopsis leads to precocious flowering and the determination of lateral meristems into floral meristems in diverse species, including the aspen tree (Weigel and Nilsson 1995). In an attempt to break the juvenile stage of apple, Flachowsky et al. transformed apple cv. 'Pinova' with the *LFY* gene from Arabidopsis under control of the 35S promoter. Instead of observing precocious flowering, they saw that overexpression of *LFY* leads to a columnar-like phenotype (Flachowsky et al. 2010), suggesting a possible role for *LFY* in controlling columnar growth of apple.

Silencing the chalcone synthase (*CHS*) genes from apple resulted in a significant reduction in phenylpropanoid levels. Besides having an effect on the colouring of flowers, fruit and stems, some major changes in plant development were reported as well. The apple lines in which the *CHS* genes were silenced had shortened internodes and smaller leaves. This effect of *CHS* silencing on the growth rate of the plants was explained by the inhibitory effect of flavonoids on auxin transport (Dare et al. 2013).

Columnar-like phenotypes have been described in other (fruit) trees besides apple. Examples are the 'pillar' trait of peach and *stumpy* poplar. The pillar trait of peach is characterised compact trees with narrow branch angles that can be used in high-density production systems (Scorza et al. 2002), similar to *Co*. This type of growth was shown to be caused by the gene *PpeTAC1*, which has homologs in diverse plant species including Arabidopsis and apple. *PpeTAC1* is located on linkage group 2 of peach, which has no synteny with chromosome 10 from apple (Dardick et al. 2013). Also, the upright phenotype of pillar peach is a recessive trait and affects branch angle instead of branch length, distinguishing it from the columnar trait of 'Wijcik'.

Stumpy poplar emerged from an activation tagging screen performed by Busov et al. in 2003. The trees have smaller internode lengths, resulting in a dwarfed plant that bears similarity to 'Wijcik'. Moreover, they show reduced branching and reduced branch length. The phenotype of *stumpy* poplar was shown to be caused by the over expression of a gibberellin-2 oxidase, an enzyme that is involved in the deactivation of gibberellin. Decreased levels of active gibberellin were found in *stumpy* and application of active gibberellic acid 3 (GA_3) could

reverse the dwarfing phenotype of the s*tumpy* mutant (Busov et al. 2003). Similarly, apple transformed with a construct to overexpress gibberellic acid insensitive (*gai*), a repressor of the gibberellin response from Arabidopsis (Peng et al. 1997), resulted in compact apple trees (Zhu et al. 2008).

Hormonal control of branching

The first hormone to be identified to have a role in the regulation of shoot branching is auxin. Auxin is required for apical dominance. When the shoot apex from a plant is removed by decapitation, plants start to branch. When auxin is applied on the top of the plant (mimicking auxin production by the shoot apical meristem), axillary bud outgrowth is repressed (Skoog and Thimann 1933). Auxin is produced in active meristems and transported basipetally, from the shoots to the root (Blakeslee et al. 2005), and is required for cell division and cell expansion (Raven et al. 1999).

A more recently identified plant hormone that has an important role in the regulation of branching is strigolactone. Strigolactone was first identified in root exudates, as a signal molecule for communication between plants and mycorrhizal fungi (Akiyama et al. 2005). Mutants of pea and rice that showed more axillary growth were found to have lower levels of strigolactones. Application of strigolactones restores their phenotype (Gomez-Roldan et al. 2008; Umehara et al. 2008).

Auxin and strigolactone seem to work together to control shoot branching, as auxin positively regulates strigolactone biosynthesis genes and strigolactone levels are reduced upon auxin depletion (Hayward et al. 2009). Application of strigolactone to axillary buds prevents their outgrowth upon decapitation (Brewer et al. 2009). It was shown that strigolactone acts as an inhibitor of auxin transport, by regulating the expression of PIN auxin efflux carriers (Bennet et al. 2006).

Cytokinins are involved in the same network to control shoot branching (Ongaro and Leyser 2008). They are produced in the shoot and stem of a plant, move upwards through the plant (Nordstrom et al. 2004) and are stimulators of bud outgrowth (Sachs and Thimann 1967).

As is clear from the poplar mutant *stumpy*, gibberellin is another hormone that can affect plant architecture. In fact, mutations leading in genes homologous to

gai from Arabidopsis that lead to decreased sensitivity to gibberellin are responsible for the compact phenotype and increased yields of modern cereals that were bred during the green revolution in the sixties (Peng et al. 1999). Another example of how a mutation in this gene can lead to altered plant architecture is the grapevine dwarf mutant described by Boss and Thomas. This mutant has a dwarf appearance with short internodes and the tendrils that are normally formed along the length of the shoot are converted into inflorescences (Boss and Thomas 2002), bearing resemblance to 'Wijcik'.

Transcription factors controlling meristem identity

Plant architecture depends on the activation of axillary buds through the control of plant hormones, but the shape of the plant is also determined by the type of branches that are formed. The lateral branches of 'Wijcik' are not only shorter, but develop mostly into fruit bearing spurs instead of vegetative branches. In Arabidopsis, several factors that control the switch from vegetative to floral meristem identity have been identified.

As already discussed before, LFY is an important transcription factor that promotes floral fate in Arabidopsis (Weigel et al. 1992). The fact that apple that over express *LFY* displayed a columnar-like phenotype (Flachowsky et al. 2010), suggests that the columnar phenotype of 'Wijcik' could indeed be explained by a change in identity of the axillary buds that leads to the development of spurs instead of branches.

Other floral meristem identity genes from Arabidopsis are APETALA1 (*AP1*) CAULIFLOWER (*CAL*) and FRUITFULL (*FUL*). The expression of these genes is regulated through the floral integrators SUPPRESSOR OF OVEREXPRESSION OF CO 1 (*SOC1*) and Flowering Locus T (*FT*), whose expression depend on the perception on environmental and developmental signals (temperature, light, gibberellins and age) through five distinct genetic pathways. (Liu et al. 2009; Srikanth and Schmid 2011). The genetic basis for floral meristem initiation seems to be rather well conserved among plants (Hanke et al. 2007), meaning that homologs of the floral meristem identity genes in Arabidopsis could be affected in columnar apple.

Efforts to identify *Co*

Many studies have been carried out in order to help the identification of the genetic basis for columnar growth in apple. The two fine mapping studies that were recently carried out by Moriya et al. and Bai et al. resulted in the identification of a region corresponding to respectively 193 and 196 kb in the homologous region from 'Golden Delicious' (Bai et al. 2012; Moriya et al. 2012). The overlap between these two regions is only around 50 kb.

Additionally, three transcriptome studies have been carried out in order to better understand the mechanism that leads to columnar growth (Krost et al. 2012; Zhang et al. 2012; Krost et al. 2013). These studies showed that plant hormones are likely to be involved in the phenotype of columnar trees. Zhang et al. highlighted the differential expression of a number of genes coding for DELLA proteins, which are involved in the response to gibberellin (Zhang et al., 2012), whereas Krost et al. proposed a link between increased levels of auxin and cytokinins and the columnar phenotype. Although these studies have been a great help to focus the search for the *Co* gene, they have not resulted in the identification of a clear *Co* gene candidate.

An interesting property of 'Wijcik' is the dominant inheritance of the columnar trait. There are several mechanisms that could explain the dominant effect of a mutation (recently described for dominant mutations in Arabidopsis, (Meinke 2013)), but the most commonly described dominant mutation leads to increased expression of a gene. Increased expression of the *teosinte branched1* (*tb1*) gene (caused by the insertion of an upstream transposon insertion (Studer et al. 2011)) from maize was shown to be responsible for the dramatic increase in apical dominance of cultivated maize (Doebley et al. 1997). A good example of a dominant mutation in apple, is the rearrangement in the upstream regulatory region of the transcription factor MYB10. Increased expression of this transcription factor led to increased levels of anthocyanin, resulting in red-fleshed apple (Espley et al. 2009).

Outline of the thesis

The goal of this thesis is to identify the genetic basis for columnar growth in apple. Identification of the mutation that led to columnar growth will enable us to

develop a powerful molecular marker that can be used in MAS. The availability of the *Co* gene would enable us to directly manipulate (fruit) tree architecture of varieties that exist already, using gene technology. Moreover, the work described in this thesis will increase our understanding of the physiological background of columnar growth

In chapter 2 of this thesis, we will describe the fine mapping of the *Co* region. Using a total of over 1500 progeny trees, which are derived from 'Wijcik' and that are segregating for the columnar phenotype, we were able to reduce the *Co* region to 393 kb in the homologous region from 'Golden Delicious'.

In chapter 3 of this thesis, we analyse two BAC libraries that were prepared from genomic DNA from both 'Wijcik' and 'McIntosh'. This sequence comparison resulted in the identification of a single mutation in 'Wijcik' and a subsequent gene expression analysis of genes present in the *Co* region resulted in the identification of *MdCo31*. We provide evidence for the role of *MdCo31* in causing the columnar phenotype of 'Wijcik'.

In chapter 4, the insertion in 'Wijcik' and the region containing the mutation are characterised in more detail. We compare the results from our study with the findings of Otto et al., who, very recently, published a study where they describe the identification of a 8.2 kb retrotransposon in 'Wijcik' (Otto et al. 2013). Also, we provide a model for the effect of the 'Wijcik' insertion on the expression of *MdCo31*.

In Chapter 5, *MdCo31* is investigated more extensively. We show that constitutive expression of *MdCo31* causes a columnar-like phenotype in diverse plant species like Arabidopsis and tomato and investigate the biochemical function of the gene, proposing a possible role for *MdCo31* in flavonoid or gibberellin biosynthesis.

The results described in the separate chapters of this thesis will be discussed in the general discussion at the end of this thesis. In this general discussion, we compare our results with studies performed by others and suggest additional experiments that should be performed in order to better understand the cause for columnar growth.

References

Akiyama, K., K. Matsuzaki, et al. (2005). "Plant sesquiterpenes induce hyphal branching in arbuscular mucorrhizal fungi." Nature **435**: 824-827.

Amine, E., N. Baba, et al. (2002). Diet, nutrition and the prevention of chronic diseases: report of a Joint WHO/FAO Expert Consultation, World Health Organization.

Anderson, J. W., P. Baird, et al. (2009). "Health benefits of dietary fiber." Nutrition reviews **67**(4): 188-205.

Andersson, H., S. Arpaia, et al. (2012). "Scientific opinion addressing the safety assessment of plants developed through cisgenesis and intragenesis." EFSA J **10**(2): 2561.

Armstrong, J. and W. D. Lane. Genetically modified reduced-browning fruit-producing plant and produced fruit thereof, and method of obtaining such. U.S. Patent No. 8,563,805, filed 22 Oct. 2013.

Bai, T., Y. Zhu, et al. (2012). "Fine genetic mapping of the Co locus controlling columnar growth habit in apple." Molecular Genetics and Genomics **287**(5): 437-450.

Baldi, P., P. J. Wolters, et al. (2013). "Genetic and physical characterisation of the locus controlling columnar habit in apple (Malus × domestica Borkh.)." Molecular Breeding **31**(2): 429-440.

Bennet, T., T. Sieberer, et al. (2006). "The *Arabidopsis MAX* pathway controls shoot branching by regulating auxin transport." Current Biology **16**: 553-563.

Blakeslee, J. J., W. A. Peer, et al. (2005). "Auxin transport." Current Opinion in Plant Biology **8**: 494-500.

Boss, P. K. and M. R. Thomas (2002). "Association of dwarfism and floral induction with a grape 'green revolution'mutation." Nature **416**(6883): 847-850.

Boyer, J. and R. H. Liu (2004). "Apple phytochemicals and their health benefits." Nutr J **3**(5): 12.

Brewer, P. B., E. A. Dun, et al. (2009). "Strigolactone acts downstream of auxin to regulate bud outgrowth in pea and *Arabidopsis*." Plant Physiology **150**: 482-493.

Busov, V. B., R. Meilan, et al. (2003). "Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature." Plant Physiology **132**(3): 1283-1291.

Collard, B. C. and D. J. Mackill (2008). "Marker-assisted selection: an approach for precision plant breeding in the twenty-first century." Philosophical Transactions of the Royal Society B: Biological Sciences **363**(1491): 557-572.

Conner, P. J., S. K. Brown, et al. (1997). "Randomly amplified polymorphic DNA-based genetic linkage map of three apple cultivars." Journal of American Society for Horticultural Science **122**(3): 350-359.

Cornille, A., P. Gladieux, et al. (2012). "New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties." PLoS genetics **8**(5): e1002703.

Costes, E., P. Lauri, et al. (2006). "Analyzing fruit tree architecture: implications for tree management and fruit production." Horticultural reviews **32**: 1-61.

Dardick, C., A. Callahan, et al. (2013). "PpeTAC1 promotes the horizontal growth of branches in peach trees and is a member of a functionally conserved gene family found in diverse plants species." Plant J **75**(4): 618-630.

Dare, A. P., S. Tomes, et al. (2013). "Phenotypic changes associated with RNA interference silencing of chalcone synthase in apple (Malus× domestica)." The Plant Journal **74**(3): 398-410.

Dirlewanger, E., E. Graziano, et al. (2004). "Comparative mapping and marker-assisted selection in Rosaceae fruit crops." Proc Natl Acad Sci U S A **101**(26): 9891-9896.

Doebley, J., A. Stec, et al. (1997). "The evolution of apical dominance in maize." Nature **386**(6624): 485-488.

Espley, R. V., C. Brendollse, et al. (2009). "Multiple repeats of a promoter segement causes transcription factor autoregulation in red apples." The Plant Cell **21**: 168-183.

Fischer, C. (1994). Shortening of the juvenile period in apple breeding. Progress in Temperate Fruit Breeding, Springer**:** 161-164.

Fisher, D. V. (1969). "Spur-type strains of McIntosh for high density planting." British Columbia Fruit Grower's Association Quart. Rep. **14**: 3-10.

Flachowsky, H., C. Hattasch, et al. (2010). "Overexpression of LEAFY in apple leads to a columnar phenotype with shorter internodes." Planta **231**(2): 251-263.

Flachowsky, H., P. M. Le Roux, et al. (2011). "Application of a high-speed breeding technology to apple (Malus× domestica) based on transgenic early flowering plants and marker-assisted selection." New Phytologist **192**(2): 364-377.

Flachowsky, H., A. Peil, et al. (2007). "Overexpression of BpMADS4 from silver birch (Betula pendula Roth.) induces early-flowering in apple (Malus× domestica Borkh.)." Plant Breeding **126**(2): 137-145.

Folta, K. M. and S. E. Gardiner (2009). Genetics and Genomics of Rosaceae, Springer.

Gerhauser, C. (2008). "Cancer chemopreventive potential of apples, apple juice, and apple components." Energy (kcal/kJ) **54**: 227.

Gomez-Roldan, V., S. Fermas, et al. (2008). "Strigolactone inhibition of shoot branching." Nature **455**: 189-194.

Hanke, M.-V., H. Flachowsky, et al. (2007). "No flower no fruit–genetic potentials to trigger flowering in fruit trees." Genes Genomes Genomics **1**(1): 1-20.

Harris, S. A., J. P. Robinson, et al. (2002). "Genetic clues to the origin of the apple." TRENDS in Genetics **18**(8): 426-430.

Hayward, A., P. Stirnberg, et al. (2009). "Interactions between auxin and strigolactone in shoot branching control." Plant Physiology **151**(1): 400-412.

Hemmat, M., N. F. Weeden, et al. (1997). "A DNA marker for columnar growth habitat in apple contains a simple sequence repeat." Journal of American Society for Horticultural Science **122**: 347-349.

Hertog, M. G., E. J. Feskens, et al. (1993). "Dietary antioxidant flavonoids and risk of coronary heart disease: the Zutphen Elderly Study." The Lancet **342**(8878): 1007-1011.

Hyson, D. A. (2011). "A comprehensive review of apples and apple components and their relationship to human health." Advances in Nutrition: An International Review Journal **2**(5): 408-420.

James, D. J., A. J. Passey, et al. (1989). "Genetic transformation of apple (Malus pumila Mill.) using a disarmed Ti-binary vector." Plant cell reports **7**(8): 658-661.

Jochemsen, H. (2000). "Toetsen en begrenzen." Een ethische en politieke beoordeling van de moderne biotechnologie, Wetenschappelijke Studiecentra van RPF en GPV (ChristenUnie).

Joshi, S., J. M. Soriano, et al. (2009). "Approaches for development of cisgenic apples." Transgenic Plant Journal **3**: 40-46.

Keli, S. O., M. G. Hertog, et al. (1996). "Dietary flavonoids, antioxidant vitamins, and incidence of stroke: the Zutphen study." Archives of Internal medicine **156**(6): 637.

Keller-Przybyłkowicz, S. and M. U. Korbin (2013). "The history of mapping the apple genome." Folia Horticulturae **161**: 168.

Kenis, K. and J. Keulemans (2007). "Study of tree architecture of apple (*Malus* x *domestica* Borkh.) by QTL analysis of growth traits." Molecular Breeding **19**(3): 193-208.

Kenis, K., J. Keulemans, et al. (2008). "Identification and stability of QTLs for fruit quality traits in apple." Tree genetics & genomes **4**(4): 647-661.

Khan, S. A., P.-Y. Chibon, et al. (2012). "Genetic analysis of metabolites in apple fruits indicates an mQTL hotspot for phenolic compounds on linkage group 16." Journal of Experimental Botany **63**(8): 2895-2908.

Kim, M. Y., K. J. Song, et al. (2003). "Development of RAPD and SCAR markers linked to the Co gene conferring columnar growth habit in apple (Malus pumila Mill.)." The Journal of Horticultural Science & Biotechnology **78**(4): 512-517.

Krost, C., R. Petersen, et al. (2013). "Evaluation of the hormonal state of columnar apple trees (Malus x domestica) based on high throughput gene expression studies." Plant Mol Biol **81**(3): 211-220.

Krost, C., R. Petersen, et al. (2012). "The transcriptomes of columnar and standard type apple trees (Malus x domestica) - A comparative study." Gene **498**(2): 223-230.

Lapins, K. O. (1976). "Inheritance of compact growth type in apple." Journal of American Society for Horticultural Science **101**: 133-135.

Lespinasse, D. (1992). Le pommier. Amélioration des espèces végétales cultivées: objectifs et critères de sélection. A. Gallais and H. Bannerot. Paris, Editions Quae**:** 579–594.

Liu, C., Z. Zonghui, et al. (2009). "Coming into bloom: the specification of floral meristems." Development **136**: 3379-3391.

Longhi, S., M. Moretto, et al. (2012). "Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (Malus x domestica Borkh.)." Journal of Experimental Botany **63**(3): 1107-1121.

Luby, J., E. van de Weg, et al. (2009). RosBREED: Enabling marker-assisted breeding in Rosaceae. International Symposium on Molecular Markers in Horticulture 859.

Maliepaard, C., F. H. Alston, et al. (1998). "Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers." Theoretical and Applied Genetics **97**(1-2): 60-73.

Marić, S., M. Lukić, et al. (2010). "Application of molecular markers in apple breeding." Genetika **42**(2): 359-375.

Meinke, D. W. (2013). "A survey of dominant mutations in Arabidopsis thaliana." TRENDS in Plant Science **18**(2): 84-91.

Moriya, S., H. Iwanami, et al. (2009). "Development of a marker-assisted selection system for columnar growth habit in apple breeding." Journal of Japanese Society for Horticultural Science **78**: 279-287.

Moriya, S., K. Okada, et al. (2012). "Fine mapping of Co, a gene controlling columnar growth habit located on apple (Malus×domestica Borkh.) linkage group 10." Plant Breeding **131**(5): 437-450.

Nielsen, K. M. (2003). "Transgenic organisms—time for conceptual diversification?" Nature Biotechnology **21**(3): 227-228.

Nordstrom, A., P. Tarkowski, et al. (2004). "Auxin regulation of cytokinin biosynthesis in *Arabidopsis thaliana*: A factor of potential importance for auxin-cytokinin-regulated development." Proceedings of the National Academy of Science **101**: 8039-8044.

Ongaro, V. and O. Leyser (2008). "Hormonal control of shoot branching." Journal of Experimental Botany **59**(1): 67-74.

The World Health Organization (2002). The world health report 2002: reducing risks, promoting healthy life.

Otto, D., R. Petersen, et al. (2013). "The columnar mutation ("Co gene") of apple (Malus × domestica) is associated with an integration of a Gypsy-like retrotransposon." Molecular Breeding: 1-18.

Patocchi, A., A. Frei, et al. (2009). "Towards improvement of marker assisted selection of apple scab resistant cultivars: Venturia inaequalis virulence surveys and standardization of molecular marker alleles associated with resistance genes." Molecular Breeding **24**(4): 337-347.

Peng, J., P. Carol, et al. (1997). "The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses." Genes & Development **11**(23): 3194-3205.

Peng, J., D. E. Richards, et al. (1999). "'Green revolution' genes encode mutant gibberellin response modulators." Nature **400**(6741): 256-261.

Pereira-Lorenzo, S., A. M. Ramos-Cabrer, et al. (2009). Breeding Apple (Malus x Domestica Borkh). Breeding Plantation Tree Crops: Temperate Species. P. M. Priyadarshan and S. M. Jain. New York, Springer**:** 33-81.

Petersen, R. and C. Krost (2013). "Tracing a key player in the regulation of plant architecture: The columnar growth habit of apple trees (Malus x domestica)." Planta **238**(1): 1-22.

Pilcher, R. R., J. Celton, et al. (2008). "Genetic markers linked to the dwarfing trait of apple rootstock 'Malling 9'." Journal of the American Society for Horticultural Science **133**(1): 100-106.

Raven, P. H., R. F. Evert, et al. (1999). Biology of plants, W.H. Freeman and Company.

Sachs, T. and K. V. Thimann (1967). "The role of auxins and cytokinins in the release of buds from dominance." American Journal of Botany **54**: 136-144.

Sakurai, K., S. K. Brown, et al. (2000). "Self-incompatibility alleles of apple cultivars and advanced selections." HortScience **35**(1): 116-119.

Saurabh, S., A. S. Vidyarthi, et al. (2014). "RNA interference: concept to reality in crop improvement." Planta: 1-22.

Schouten, H. J., F. A. Krens, et al. (2006a). "Cisgenic plants are similar ro traditionally bred plants." EMBO reports **7**: 750-753.

Schouten, H. J., F. A. Krens, et al. (2006b). "Do cisgenic plants warrant less stringent oversight?" Nature Biotechnology **24**(7): 753-753.

Scorza, R., D. Bassi, et al. (2002). "Genetic interactions of pillar (columnar), compact, and dwarf peach tree genotypes." Journal of the American Society for Horticultural Science **127**(2): 254-261.

Silfverberg-Dilworth, E., C. L. Matasci, et al. (2006). "Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome." Tree genetics & genomes **2**(4): 202 - 224.

Skoog, F. and K. V. Thimann (1933). "Further experiments on the inhibition of the development of lateral buds by growth hormone." Proceedings of the National Academy of Science **20**: 480-485.

Srikanth, A. and M. Schmid (2011). "Regulation of flowering time: all roads lead to Rome." Cellular and Molecular Life Sciences **68**(12): 2013-2037.

Studer, A., Q. Zhao, et al. (2011). "Identification of a functional transposon insertion in the maize domestication gene tb1." Nat Genet **43**(11): 1160-1163.

Tian, Y., C. Wang, et al. (2005). "Mapping *Co*, a gene controlling the columnar phenotype of apple, with molecular markers." Euphytica **145**: 181-188.

Tobutt, K. R. (1985). "Breeding columnar apples at East Malling." Acta Horticultura **159**: 63-68.

Tobutt, K. R. (1994). "Combining apetalous parthenocarpy with columnar growth habit in apple." Euphytica **77**(1-2): 51-54.

Tränkner, C., S. Lehmann, et al. (2010). "Over-expression of an FT-homologous gene of apple induces early flowering in annual and perennial plants." Planta **232**(6): 1309-1324.

Umehara, M., A. Hanada, et al. (2008). "Inhibition of shoot branching by new terpenoid plant hormones." Nature(195-200).

Velasco, R., A. Zharkikh, et al. (2010). "The genome of the domesticated apple (Malus x domestica Borkh.)." Nature Genetics **42**(10): 833-839.

Weigel, D., J. Alvarez, et al. (1992). "LEAFY controls floral meristem identity in Arabidopsis." Cell **69**(5): 843-859.

Weigel, D. and O. Nilsson (1995). "A developmental switch sufficient for flower initiation in diverse plants." Nature **377**: 495-500.

Zhang, Y., J. Zhu, et al. (2012). "Characterization of transcriptional differences between columnar and standard apple trees using RNA-Seq." Plant Molecular Biology Reporter **30**(4): 957-965.

Zhu, L.-H., A. Holefors, et al. (2001). "Transformation of the apple rootstock M.9/29 with the *rol*B gene and its influence on rooting and growth." Plant Science **160**(3): 433-439.

Zhu, L., X. Li, et al. (2008). "Overexpression of the Arabidopsis gai gene in apple significantly reduces plant size." Plant cell reports **27**(2): 289-296.

Zhu, Y. and B. H. Barritt (2008). "Md-ACS1 and Md-ACO1 genotyping of apple (Malus x domestica Borkh.) breeding parents and suitability for marker-assisted selection." Tree genetics & genomes **4**(3): 555-562.

Zhu, Y. D., W. Zhang, et al. (2007). "Evaluation of inter-simple sequence repeat analysis for mapping the Co gene in apple (Malus pumila Mill.)." The Journal of Horticultural Science & Biotechnology **82**(3): 371-375.

Chapter 2

Genetic and physical characterisation of the locus controlling columnar habit in apple (*Malus x domestica* Borkh.)

Molecular Breeding **31** (2013, p. 429-440)

Paolo Baldi, Pieter J. Wolters, Matteo Komjanc, Roberto Viola, Riccardo Velasco and Silvio Salvi

Abstract

Increasing our understanding of the genetic control of tree architecture would potentially allow us to better tailor newly bred apple cultivars in relation to aspects of field management such as planting density, pruning, pest control and disease protection, also having an indirect impact on yield and fruit quality. The *Columnar (Co)* locus strongly suppresses lateral branch elongation and is the most important genetic locus influencing tree architecture in apple. *Co* has previously been mapped on apple linkage group 10. In order to obtain fine mapping of *Co,* genetically and physically, three adult segregating experimental populations, with a total of 301 F_1 plants, and one substantial 3-year old population of 1,250 F_1 plants were phenotypically analyzed and screened with newly developed SSR markers, based on the 'Golden Delicious' apple genome sequence available. *Co* was found to co-segregate with SSR marker Co04R12 and was confined in a region of 0.56 cM between SSR markers Co04R11 and Co04R13, corresponding to 393 kb on the 'Golden Delicious' genome sequence. In this region, 36 genes were predicted, including at least seven sequences potentially belonging to genes that could be considered candidates involved in control of shoot development. Our results provide highly reliable, virtually co-segregating markers to assist in apple breeding to obtain modified tree habit and lay the foundations for cloning *Co*.

Keywords: **Tree architecture; Fine mapping; *Co* locus; Growth habit**

Introduction

Apple (*Malus x domestica* Borkh.) is one of the most widely cultivated fruit crops around the world. In 2010, worldwide apple production was estimated to be 69.5 million tons (FAOSTAT, Food and Agriculture Organisation of the United Nations; http://faostat.fao.org). It is therefore not surprising that many breeding programs are in place in order to improve agronomically important traits such as productivity, crop resistance and tolerance to biotic and abiotic stresses and fruit quality.

Apple tree architecture is an important trait affecting plant density, pruning requirements, harvest efficiency and fruit yield and quality (Kenis and Keulemans 2004), however its genetic control is still poorly understood. Considerable variability has been observed among apple cultivars, which can be classified into several architectural types, going from dwarf to spur and columnar, according to tree growth, length and distribution of branches and fruit position (Lespinasse and Delort 1986; Lespinasse 1992; De Wit et al. 2004).

Improved knowledge of basic plant biological processes (including plant development) and the impact of genomics is already making it possible to accelerate the selection process in many species, including apple (Troggio et al. 2012), especially for traits difficult or expensive to select for (Cook and Varshney 2010),. This is particularly valuable for a species such as apple, characterised by a very long juvenile phase. Specifically, in apple breeding programs, the availability of molecular markers closely linked to plant architecture genes and alleles could be exploited by marker-assisted selection, while the availability of the genes and their regulatory sequences would provide the breeder with the opportunity to use genetic engineering.

The columnar tree habit, corresponding to the Type 1 habit as described by Lespinasse (1992), shows specific characteristics such as short internodes and a low number of lateral shoots with a high percentage of short spurs (Lapins 1976). An apple mutant showing a columnar habit was discovered as a spontaneous bud sport of 'McIntosh' (Fisher 1970). This mutant was called 'Wijcik' or 'Wijcik McIntosh'. Genetic control of the columnar habit in 'Wijcik' is due to a single dominant gene (*Co*) (Lapins and Watkins 1973; Lapins 1974; Kelsey and Brown 1992), which is thought to interact with several modifiers, as shown by the segregation ratio of some crossings (Lapins 1976) and the intermediate-type phenotype described by some authors (Hemmat et al., 1997; Kim et al., 2003). Furthermore, a number of columnar-associated phenotypic effects have been described at genetic, physiological, morphological and microscopic levels (Kenis and Keulemans 2007; Zhang and Dai 2011). A Type 1 habit which resembles the 'Wijcik' mutation was also induced by overexpression of the Arabidopsis *LFY* gene into apple and regeneration of plants (Flachowsky et al. 2010).

It has been clearly recognised that the columnar habit could be profitably used in breeding programs to obtain compact cultivars requiring little or no staking and pruning, while the absence of long lateral shoots could allow high density planting (Janick et al. 1996). However, the same habit has also been associated with negative traits, such as poor fruit quality and shape and pronounced biennial bearing (Lauri and Lespinasse 1993; Tobutt 1985).

The *Co* gene was located on linkage group 10 and several groups developed molecular markers in the surroundings of the gene. A random amplified polymorphic DNA (RAPD) marker ($OA11_{1000}$) was found to map at 6 cM from the *Co* gene (Conner et al. 1997) and subsequently transformed into a simple sequence repeat (SSR) marker (SSR^{Co}) (Hemmat et a. 1997). A sequence characterised amplified region (SCAR) marker ($SCB82_{670}$) was obtained from a RAPD marker previously found ($WB82_{670}$) using the mapping population 'Fuji' x 'Tuscan' (Kim et al. 2003). Other SCAR markers ($SCAR_{682}$ and $SCAR_{216}$) were developed in a second study starting from a RAPD marker and an amplified fragment length polymorphism (AFLP) marker (Tian et al. 2005). Two inter simple sequence repeat markers (ISSR) were developed by Zhu et al. (2007) using a segregating population derived from the 'Fuji' x 'Telamon' cross. The *Co* locus was shown to co-localise with a major QTL for plant architecture on LG10, in crosses involving one parent carrying the *Co* mutation (Kenis and Keulemans 2007).

In the last decade several highly saturated apple genetic maps have been developed, containing highly informative SSR markers (Liebhard et al. 2003; Silfverberg-Dilworth et al. 2006) and SNPs (Velasco et al. 2010;). Although such maps were not developed using experimental crosses with *Co*-segregating parents, several markers can now be easily found in the vicinity of the target region by map comparisons, tested for polymorphism, and eventually used for linkage analysis. Finally, the recently completed apple genome sequence (Velasco et al. 2010) represents the ultimate resource for the development of new markers and to support *Co* mapping and cloning.

In this paper we describe the saturation of the *Co* region at sub-cM resolution with molecular markers using four different segregating populations. Reliable co-dominant SSR markers virtually co-segregating with *Co* are provided.

Finally, we anchor and confine the *Co* locus on the apple genomic sequence and provide an initial list of possible candidate genes.

Materials and methods

Plant material and phenotype analysis

Three mature segregating populations derived respectively from the crosses 'Golden Delicious' x 'Wijcik' (101 individuals; 6 year-old plants from the time of grafting), 'Goldrush' x 'Wijcik' (138 individuals; 4 year-old plants from the time of grafting) and 'Galaxy' x 'Wijcik' (62 individuals; 4 year-old plants from the time of grafting) were used in the first part of this work to delimit the *Co* target region on LG10. The three populations were grafted on 'M9' rootstock. Additionally, two young segregating populations of 374 and 876 (total of 1,250) individuals respectively, deriving from two independent 'Golden Delicious' x 'Wijcik' crosses performed in 2009, were used in the second part of the work in order to increase the genetic resolution of analysis and saturate the *Co* region. All plants were classified as columnar or standard during the dormant season through multiple visual inspections. The visual tree characteristics considered during visual inspection were the diameter of the main stem (columnars showing little difference between the top and base), the length of the internodes (columnars showing very short internodes) and the number and type of lateral shoots (columnars showing a considerable majority of spurs) (Hemmat et al. 1997). For the plants developed from the 2009 crosses, final phenotyping was carried out in 2011, since expression of the columnar trait is reduced in younger seedlings (See Results and Discussion and also Kelsey and Brown 1992).

DNA extraction

Genomic DNA was extracted from young leaves using the DNeasy 96 Plant Kit (Qiagen, Valencia, CA, USA). Four leaf discs of 1 cm diameter were harvested and lyophilised using a Scanvac CoolSafe lyophiliser (LaboGene, Lynge, DK). The dry material was then ground to a fine powder and used for the extraction procedure.

SSR marker analysis

Four SSR markers available in the literature (AU223548; CH03d11; Hi01a03; Hi01b01) were tested on three different segregating populations in order to delimit the *Co* region. The SSRs were selected from those developed by Liebherd et al. (2003) and Silfverberg-Dilworth et al. (2006) and mapping in the vicinity of *Co*. According to the HiDRAS (High-quality Disease Resistant Apples for a Sustainable Agriculture) web site (http://users.unimi.it/hidras/) SSR markers Hi01a03 and Hi01b01 contain the same repeat sequence and identical flanking regions, therefore they were considered to characterise the same locus and only Hi01b01 was used for genetic analysis.

SSR analysis was carried out using the direct fluorescent primer method (Schuelke et al. 2000). For each primer set, a forward primer with a fluorescent dye at the 5' end was synthesised (Sigma-Aldrich, St. Louis, MO, USA). PCR amplification was carried out in a final volume of 15 µl with 1.5 µl of 10x buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 1.5 µM of each primer and 0.2 U of Taq polymerase (AmpliTaq Gold; Applied Biosystems, Foster City, CA, USA). Amplification consisted of a denaturing step at 94 °C for 10 min, followed by 35 cycles at 94 °C for 30 s, 60 °C for 30 s, 72 °C for 60 s and a final extension of 5 min at 72 °C. PCR products were separated and detected using a 3730 DNA analyzer (Applied Biosystems, Foster City, CA, USA). The size of the different alleles was determined using an internal standard DNA (MassRuler Low Range DNA Ladder; Fermentas, Glen Burnie, MD, USA).

For the development of new SSR markers at the *Co* region, the genomic sequence of 'Golden Delicious' was taken and exploited as a source of putative microsatellite sequences (Velasco et al. 2010). SSR sequences were searched using the IMEx software (Suresh and Hampapathalu 2007) and primer pairs were designed in the flanking region of each sequence. PCR amplification and amplicon separation were carried out as described above. Marker distances and genetic maps were computed using the software JoinMap4 (Van Ooijen and Voorrips 2001), using default parameter settings.

Candidate gene analysis at the *Co* region

When the 'Golden Delicious' genomic region putatively containing *Co* was examined looking for ORFs, only 17 transcripts with a predicted function were found in the gene set in the 'Golden Delicious' Genome Browser (http://genomics.research.iasma.it). Moreover the length of the predicted ORFs was highly variable, going from 120 bp to 7218 bp. Therefore, in order to integrate the available ORF predictions, the 'Golden Delicious' genome sequence corresponding to the putative *Co* region was also analyzed using two different gene predictors, AUGUSTUS (Stanke et al. 2004) and GENSCAN (Burge and Karlin 1997).

BAC library construction and screening

The BAC library was constructed at the Amplicon Express Inc. starting from frozen 'Wijcik' leaf material as described by Tao et al. (2002). The restriction enzyme used for partial digestion was HindIII. The genomic fragments were cloned into pCC1BAC Epicentre vector in Invitrogen DH10b Phage Resistant competent cells. The average insert size estimated by Amplicon Express Inc. was 145 kb. The storage media used was LB Lennox, 36 mM K_2HPO_4, 1.7 mM Sodium citrate, 0.4 mM $MgSO_4$, 6.8 mM $(NH_4)_2SO_4$,, 4.4% Glycerol, 12.5 µg/ml Chloramphenicol in 384-well plates (50 µl/well). For the screening with SSR markers, BAC DNA was extracted using a modified alkaline lysis method. 1,5 ml of culture was used for each DNA extraction. The cultured cells were spinned down and the pellet were resuspended in 150 µl of resuspension buffer (50 mM Tris-HCl, 10 mM EDTA, pH=8,0). 150 µl of lysis buffer (0,2 M NaOH, 1% SDS) was added to the suspensions. After 5 minutes incubation, 150 µl of cold neutralisation buffer (3M K-acetate, pH=5,5) was added and the suspension was filtered using a UNIFILTER microplate (Whatman), by centrifugation. The flowthrough was collected in a UNIPLATE (Whatman) and DNA was precipitated by adding 0,7 volume of isopropanol. The DNA pellet was collected by centrifugation, washed with 70% ethanol and dissolved in 20 µl of H_2O. PCR amplification of SSRs was performed in a volume of 10 µl in 1x PCR buffer containing 1 µl of DNA template, 0.2 mM dNTPs, 3 mM MgCl2, 0.5 units of AmpliTaq Gold DNA polymerase (Applied Biosystems) and 0.1 mM of each primer (of which one is 6-FAM labeled). The

cycling conditions include an initial denaturation step of 10 minutes at 95 °C followed by 35 cycles of 30 s at 95 °C, 1 min at 60 °C, 30 s at 72 °C and a final extension phase of 10 min at 72 °C. Sequencing of the BAC ends was performed using M13 forward and M13 reverse primers.

Results

Phenotypic evaluation of plants

Three mature, well established populations were used to delimit the exact *Co* region. 101 individuals, belonging to a 6 year-old segregating population, derived from the 'Golden Delicious' x 'Wijcik' cross were categorised as columnar or standard (Table 1) according to the growing pattern, without ambiguous phenotypes or missing data. The segregation was 39.6% columnar and 60.4% standard, significantly different from the expected 1:1 ratio ($P<0.05$. Table 1). 63 plants from the 'Galaxy' x 'Wijcik' population and 141 plants from the 'Goldrush' x 'Wijcik' population were also evaluated. Probably because these populations were only 4 years old at the time of phenotypic evaluation, some individuals (three individuals from 'Goldrush' x 'Wijcik' and one from 'Galaxy' x 'Wijcik') showed an intermediate phenotype and could not be categorised as columnar or standard and were therefore excluded from the analysis. For the 'Goldrush' x 'Wijcik' cross the segregation was 39.1% columnar and 60.9% standard ($P<0.05$), while for the 'Galaxy' x 'Wijcik' cross the Co-WT segregation did not differ statistically from the expected 1:1 ratio (Table 1).

In order to achieve a higher mapping resolution suitable for positional cloning, two new segregating populations were obtained in 2009 from two independent 'Golden Delicious' x 'Wijcik' crosses. The populations originating from these crosses originally consisted of 401 individuals (population A) and 998 individuals (population B) respectively. Due to the impossibility of assigning a reliable columnar score to plants during the first year of growth, two further evaluations were performed in 2010 and 2011, with plants in their second and third year of growth. Nonetheless, at the end of the third year, 25 plants in population A and 22 plants in population B still showed ambiguous phenotypes and could not be classified as columnar or standard. They were therefore not used for genetic

mapping. In these latter populations, the *Co* segregation frequencies did not statistically differ from the expected 1:1 ratio (Table 1).

Table 1. Segregating populations used and phenotypic evaluation of growing pattern.

Maternal parent	Paternal parent	Age at evaluation	Root system	Number of plants	Columnar	Standard	P-value
'Golden'	'Wijcik'	6	Grafted ('M9')	101	40	61	0.037*
'Galaxy'	'Wijcik'	4	Grafted ('M9')	62	28	34	0.446
'Goldrush'	'Wijcik'	4	Grafted ('M9')	38	54	84	0.011*
'Golden'	'Wijcik'	2 and 3	Self-rooted	374	199	175	0.215
'Golden'	'Wijcik'	2 and 3	Self - rooted	876	442	434	0.787

* Statistically significant difference from 1:1 segregation ratio

Identification of the *Co* region

Genetic mapping of *Co* was carried out in two steps. In the first step, existing and newly developed SSR markers were used to score three pre-existing adult plant populations in order to clearly delimit *Co* on LG 10, at cM-range resolution. In the second step, by taking advantage of the larger new population, *Co* was genetically fine mapped and physically mapped on the 'Golden Delicious' genome sequence.

Published markers Hi01b01, Ch03d11 and AU223548 (Liebhard et al. 2003; Silfverberg-Dilworth et al. 2006) were mapped on the three adult populations. Hi01b01 was located at 4.2 cM from *Co* in 'Golden Delicious' x 'Wijcik', at 5.2 cM in 'Galaxy' x 'Wijcik', and at only 1.5 cM from *Co* in 'Goldrush' x 'Wijcik' (Fig. 1A-C). When CH03d11 was first tested, unclear results were obtained after amplification and electrophoresis, leading to a possible loss of information. Since in a previous work CH03d11 was shown to map very close to *Co* (Moriya et al. 2009), a new primer pair was designed in order to improve the reliability of marker analysis. The new SSR marker was called Co02R10 (Table 2) and, after analysis, was found to co-segregate with *Co* (Fig. 1A-C). Finally, AU223548 was found to map too distant from *Co* (data not shown) on the opposite side of *Co* as compared to Hi01b01. In order to define the portion of LG10 containing *Co* on the side of AU223548 more precisely, new SSR markers were developed by exploiting

the genomic sequence of 'Golden Delicious' as a source of putative SSR markers. Using this approach, two new SSR markers, Co01L02 and Co01L03 (Table 2), were developed and mapped on the distal side of *Co* in all three mapping populations, at genetic distances varying from 1.5 to 4.8 cM, with Co01L03 mapping further away in two out of three populations (Fig. 1A-C).

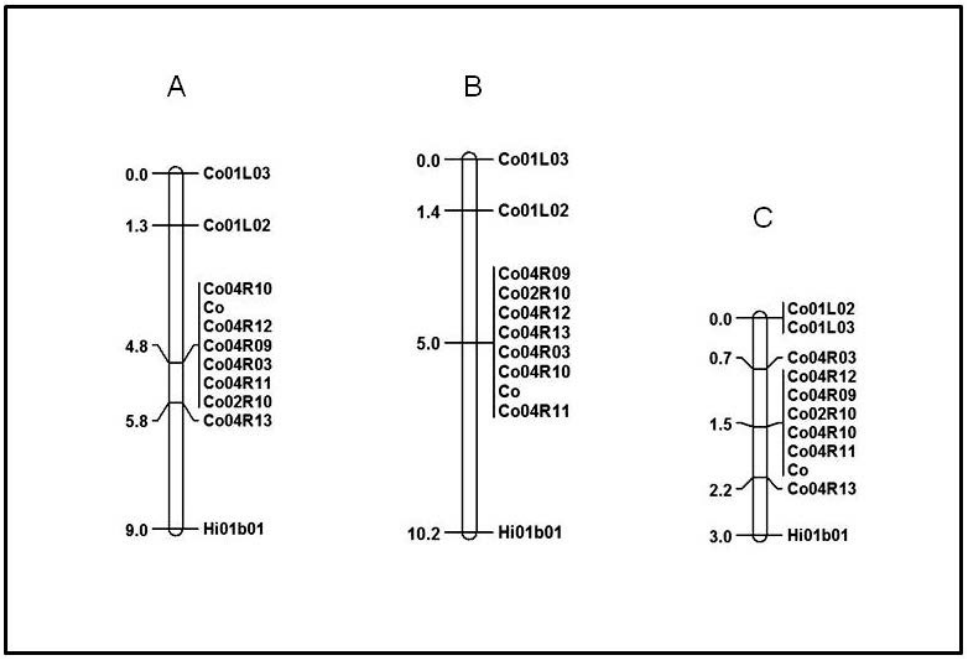

Figure 1. Linkage maps of the region surrounding *Co* in the three segregating populations used. **A.** 'Golden Delicious' x 'Wijcik'; **B.** 'Galaxy' x 'Wijcik'; **C.** 'Goldrush' x 'Wijcik'. Maps were aligned using SSR markers successfully mapped on all the populations. Map distances are expressed in centimorgans (cM) and are shown on the left.

One additional round of microsatellite searching on the 'Golden Delicious' genome sequence in the region between Co01L02 and Hi01b01 enabled us to design and map six additional SSRs (Co04R03, Co04R09, Co04R10, Co04R11, Co04R12 and Co04R13. Table 2). Four of these markers (Co04R09, Co04R10, Co04R11, Co04R12) were found to co-segregate with *Co*, as no recombinants were found in any population (Fig. 1A-C). On the other hand, a single recombinant was found between *Co* and Co04R03 in 'Goldrush' x 'Wijcik', placing the marker at 0.8 cM proximal of *Co* (Fig.1C). Similarly, Co04R13 was mapped 1.0 or 0.7 cM

distal of *Co* in 'Golden Delicious' x 'Wijcik' and 'Goldrush' x 'Wijcik' populations respectively. Therefore, according to the results of the first experimental step, the genomic region putatively containing the gene(s) responsible for *Co* was reliably comprised between Co04R03 and Co04R13, separated by approximately 1.5 cM.

Table 2. Markers mapped in the *Co* region. Underlined alleles were found associated to the columnar phenotype

Marker	Primer sequences	Repeat sequence	'Golden' alleles	'Goldrush' alleles	'Galaxy' alleles	'Wijcik' alleles
Hi01b01	F: 5'-GCTACAGGCTTGTTGATAACGC-3'	GAA	190	166-190	190	154-190
	R: 5'-ACGAATGAAATGTCTAAACAGGC-3'					
Co01L02	F: 5'-AGCCACAAACGGTACATAGC-3'	AT	143-153	153	153-157	133-139
	R: 5'-ATTATTGTTATTTCTCGTTGACG-3'					
Co01L03	F: 5'-ACATCCAATTCCTATTCGGC-3'	AT	213-243	241-243	219-243	203-215
	R: 5'-TGCTGGATTATATAGTATGGC-3'					
Co04R03	F: 5'-GTTTGCTCTTTTGACTGACGC-3'	GA	205-209	205	205-209	209-213
	R: 5'-CTCAGCTTTTCAGCCATTTCC-3'					
Co02R10	F: 5'-ATCATGGAGGGTCTACTTCG-3'	CT	243	243-247	243	251-299
	F: 5'-GAGATTAAGAAAGCGCGAACC-3'					
Co04R09	F: 5'-TAGTGACATATACATGGTGCG-3'	AT	188-192	176-188	188-192	156-188
	R: 5'-GTTGGAGAATGAGTGACGGC-3'					
Co04R10	F: 5'-ACCTGGTTCCGGTACATAGC-3'	TC	193	183-193	193	185-203
	R: 5'-AACCTTCCATGGCAGCAATC-3'					
Co04R11	F: 5'-ACATCATGGTATGACAGAGGTG-3'	CT	214	190-214	214	172-184
	R: 5'-TCTAAGCCTGTCAAGATGGC-3'					
Co04R12	F: 5'-TTTATCTGACTAAGGGGAAGG-3'	GA	208	208	208	194-222
	R: 5'-ATGGACTTGTATTCCTTAGGG-3'					
Co04R13	F: 5'-ATTTTCCCTCTCTTCTGTTGC-3'	AT	247	241-247	247-249	235-237
	R: 5'-TCTTGGAAAGACGTGGCACG-3'					

Fine mapping of the *Co* region and ORF analysis

In the second experimental step, the newly developed SSRs were mapped in reference to *Co* using the two new large experimental populations (Table 1) developed for the purpose. Although the two populations originated from distinct crosses and were made up of 374 (pop. A) and 876 (pop. B) individuals respectively, they were treated as a single segregating population of 1,250 individuals for genetic mapping. The genetic map surrounding the *Co* locus after mapping the two markers (Co04R03 and Co04R13) flanking *Co*, and the five SSRs (Co02R10, Co04R09, Co04R10, Co04R11 and Co04R12) previously co-

segregating with Co, is shown in Fig. 2A. Based on these results, *Co* was mapped between two co-mapping SSRs (Co04R10 and Co04R11) and Co04R13, in a region of 0.56 cM. The SSR marker Co04R12 was found to co-map with *Co* even after this high-resolution analysis (Fig. 2A).

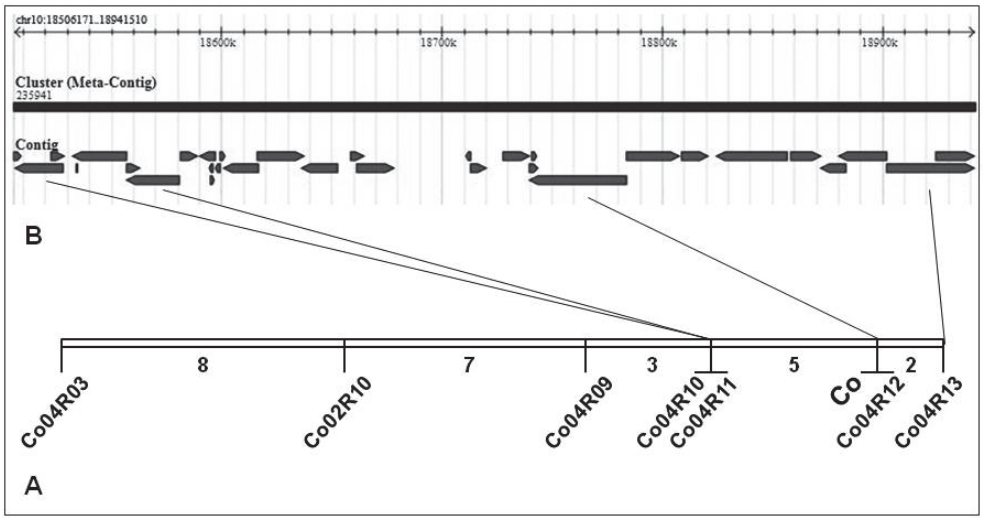

Figure 2. Schematic representation of the region surrounding *Co*. A. Genetic map. Numbers indicate the recombinants found in a segregating population of 1,250 individuals derived from two crossing between 'Golden Delicious' and 'Wijcik'. **B.** Representation of genomic region spanning the *Co* locus in 'Golden Delicious' as available from the apple genomic sequencing project (http://genomics.research.iasma.it)

In order to estimate the physical size of the *Co* region delimited after the fine mapping results, the genetic map was projected on the genomic sequence of 'Golden Delicious' as available at http://genomics.research.iasma.it and the position of Co04R10, Co04R11, Co04R12 and Co04R13 checked (Fig. 2B). Based on this projection, Co04R10 resulted 64 kb more distal than Co04R11. The actual genomic region between Co04R10 and Co04R13 was estimated to be 393 kb. If Co04R11 was considered as the marker distally delimiting the *Co* region, the size of the genomic portion putatively containing the gene decreased to 329 kb.

However, it should be noted that this last observation (the relative linear order of Co04R10 and Co04R11 on the chromosome) should be treated with caution, as the physical order of the genes on the current apple genomic sequence

assembly is prone to local errors. Therefore, for the purposes of prevention, the whole region between Co04R10 and Co04R13 should still be considered as the putative *Co* region.

The *Co*-containing genomic sequence was examined by searching for open reading frames (ORFs). A total of 36 putative ORFs were found (see supporting material, Table S1). Among the gene functions coded by the putative ORFs we identified at least one MYB transcription factor, several AP2/ERF-like transcription factors, a possible NAM transcription factor, an auxin-induced gene and a basic helix-loop-helix (bHLH) protein.

In order to physically isolate the actual region putatively containing *Co*, a 'Wijcik' bacterial artificial chromosome (BAC) library was constructed and screened using all the SSR markers available. In all, 10 BAC clones were identified for the allele in coupling with *Co*. Sequencing of the BAC ends allowed the mapping and orientation of the contigs on the corresponding region of 'Golden Delicious' genome. According to the results obtained, a minimum of four BAC clones (W34F8; W32I3; W11K19; W61E8) were necessary to completely span the putative *Co* region (Fig. 3).

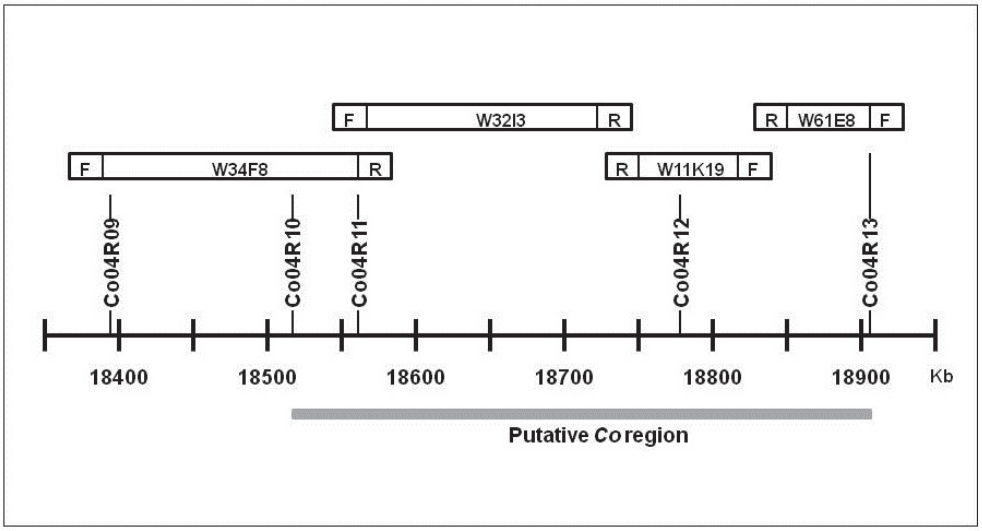

Figure 3. ´Wijcik´ bacterial artificial chromosome (BAC) contigs spanning the *Co* region. Position of the SSR markers on the corresponding region of the 'Golden Delicious' genome is indicated (black line). F: M13 forward BAC insert-end; R: M13 reverse BAC insert-end.

Discussion

Phenotypic analysis and segregation ratio

In this work, different populations segregating for the *Co* locus were used. In order to accurately delimit the position of the *Co* gene on LG10, in the first part of this study, three pre-existing internally available populations of adult plants were utilised. The availability of adult plants was crucial to correctly phenotype the plants and build a reliable framework genetic map at the *Co* locus, since the columnar habit is fully phenotypically expressed only after several years of plant growth (Kim et al. 2003; Tian et al. 2005; Zhu et al. 2007).

For two of the populations ('Golden Delicious' x 'Wijcik' and 'Goldrush' x 'Wijcik') a statistically significant ($P<0.05$) deviation from the expected 1:1 segregation ratio (between columnar and standard plants) was found. In both cases, the standard plants outnumbered the columnar plants. A deviation from the expected segregation rate was previously observed at the *Co* locus in different mapping populations and attributed to the segregation of unlinked modifier genes (Lapins 1976; Moriya et al. 2009). In our case, it is also noticeable that two out of the three grafted populations showed a deviated segregation ratio, whereas the larger 'Golden Delicious' x Wijcik' populations utilised for fine mapping were not grafted and did not show any deviation. Two possible explanations for this observation are i) interaction (epistasis) of the *Co* gene with one or more unknown loci; ii) incomplete penetrance of the Columnar phenotype caused by a masking effect or interaction due to the grafting process and/or to the rootstock. However, both explanations do not seem to fit our results since, in both cases (epistasis and grafting-induced incomplete penetrance), the molecular genotyping of the mapping populations should have revealed plants with 'McIntosh Wijcik' haplotype at the *Co* chromosome region, and with 'standard' phenotype. Such plants were not observed. One alternative more likely explanation is that a loss of Columnar plants occurred during the processes of grafting and/or the early phase of field cultivation because of their weak or abnormal growth, possibly enhanced by the reduced vigor of the utilised rootstock ('M9').

A clear indication of the influence of plant age on columnar phenotype was observed when two new 'Golden Delicious' x 'Wijcik' segregating populations were

obtained. During the first and second year of growth, only some of the individuals could be unambiguously categorised as columnar or standard, as also previously observed by others (Hemmat et al. 1997; Kim et al. 2003; Moriya et al. 2009). This difficulty in classifying plants into one of the two phenotypic classes was confirmed in our study, even in the third year of evaluation, when a total of 47 individuals were excluded from further analysis because they still showed an intermediate phenotype. However, the remaining individuals showed a clear 1:1 columnar vs. standard segregation ratio.

Co fine and physical mapping

The genetic characterisation of the *Co* locus presented here has several potentially interesting implications. First, the development of molecular markers virtually cosegregating to the most important apple tree architecture locus, fully expressed only at the adult stage of development, will make it possible to apply fully efficient marker-assisted selection (MAS) at the seedling stage. This should relevantly reduce the costs of breeding programs designed to introduce the columnar trait into new cultivars (Khan et al. 2007; Chagné et al. 2007). Secondly, mapping the *Co* locus to a physically delimited chromosome region is the starting point for its positional cloning. Cloning *Co* would provide invaluable insight into the molecular, cellular and physiological control and expression of apple tree architecture and would eventually make it possible to manipulate the trait using genetic engineering.

Although previous works have made it possible to map several molecular markers in the surroundings of *Co* (Kim et al. 2003; Tian et al. 2005; Zhu et al. 2007; Moriya et al. 2009), none reached a resolution suitable for its positional cloning. This notwithstanding, in this work we took advantage of markers available in the literature in order to proceed towards fine and physical mapping. It was decided to consider the SSR marker type alone, since it is one of the most informative and effective types of markers. In order to delimit the genomic region containing *Co*, molecular markers available in the literature on both sides of *Co* were initially searched. According to our data, of the markers tested, only Hi01b01 was mapped on one side of *Co* at less than 5 cM from the gene. Indeed, AU223548 was found to map on the opposite side to Hi01b01, but the distance

from *Co* was considered excessive, while CH03d11, which was transformed into Co02R10, was shown to co-map with *Co* in all the populations examined.

The development of two new SSR markers (Co01L02 and Co01L03) on the same side as AU223548 but closer to *Co*, enabled the construction of genetic maps encompassing the *Co* region in all three mapping populations (Fig. 1). The size of these partial maps varied according to the population used. In 'Golden Delicious' x 'Wijcik' (101 individuals), the distance between Co01L02 and Hi01b01 was estimated to be 7.7 cM. A similar result was found in 'Galaxy' x 'Wijcik' (62 individuals), where the distance was 8.8 cM. However, the distance was only 3 cM in 'Goldrush' x 'Wijcik' (138 individuals). Differences in the length of genetic maps observed for the same chromosome region in different segregating populations are relatively common and can probably be ascribed to sampling and genetic factors controlling recombination frequencies both locally and genome-wide (Doligez et al. 2006; Vezzulli et al. 2008; Moriya et al. 2009).

In order to reduce the putative *Co* region and saturate it with SSR markers, the availability of the 'Golden Delicious' genomic sequence (Velasco et al. 2010) proved to be a powerful tool, as it could be directly searched for putative repeat sequences. In this way, six new SSR markers were developed and mapped. Moreover, it was possible to separate and orient SSR markers (Co04R10 and Co04R11. Fig. 2B) that had remained linked even after analysis of the 1,250-plant population. Here it should be noted that two more works describing the fine genetic mapping of the *Co* locus were published during the preparation of this paper (Bai et al. 2012; Moriya et al. 2012). The two groups identified two different, not overlapping, genomic regions as containing *Co*, therefore leaving a small degree of uncertainty about the actual position of the gene. The data presented in our work are well in agreement with the data presented by Moriya et al. When a marker described by Bai et al. as cosegregating with *Co* (C18470-25831) was tested on our populations, three recombinants were found (data not shown), placing the marker outside the *Co* region on the same side of *Co* as Co04R13.

The alignment of the genetic and physical maps at the *Co* region indicated that a chromosome region of 0.56 cM corresponded to 393 kb (physical/genetic distance ratio = 702 kb/cM). This is similar to what has previously been observed in

apple at the scab resistance *Vf* locus (423 kb/cM to 857 kb/cM (Patocchi et al. 1999). A lower and more favorable ratio was found in a similar work characterizing the scab resistance *Vr2* locus (Galli et al. 2010). In the latter case the observed physical/genetic distance ratio was 91 kb/cM.

ORF analysis

When the putative ORFs present in the physical genomic region completely associated with *Co* were searched and analyzed (see supporting material, Table S1), several potentially interesting *Co* candidate genes were found. Putative transcription factors were identified belonging to four different classes, MYB, bHLH, AP2/ERF and NAM. Plant MYB transcription factors are one of the best known transcription factor families (Feller et al. 2011) and are characterised by the presence of one to four, or more, imperfect MYB repeats that can function in DNA binding and protein-protein interaction. MYB transcription factors are thought to be involved in a number of physiological processes such as flavonoid biosynthesis (Czemmel et al. 2009), inflorescence development (Zhang et al. 2009), shoot morphogenesis and leaf patterning (Guo et al. 2008). bHLH proteins are another widely studied group of transcription factors that in plants have been found to regulate important developmental and physiological processes such as gibberellin response (Lee et al. 2006), anther development (Zhang et al. 2006) and fruit development (Tani et al. 2011). The AP2/ERF transcription factor family has mainly been involved in plant stress response (Xu et al. 2011) but in some cases members of this group have been related to gibberellin biosynthesis (Yano et al. 2009), seed germination (Yamagishi et al. 2009) and root development (Kitomi et al. 2011). NAM proteins are plant-specific transcription factors that have been found to control several developmental processes, such as shoot apical meristem development (Souer et al. 1996; Aida et al. 1997), cell expansion in flower organs (Sablowski and Meyerowitz 1998) and auxin-dependent formation of lateral roots (Xie et al. 2000). Therefore, genes belonging to one of any of the above mentioned transcription factors families could be considered putative candidate genes controlling the *Co* habit in apple. Moreover, a gene showing a high similarity to the 5NG4 protein-coding gene family was predicted inside the *Co* region. 5NG4-like proteins are widely conserved among plant species, are auxin-inducible and have

been found to be involved in root formation (Busov et al. 2004). Therefore, a possible role of such proteins in controlling tree architecture cannot be ruled out.

It should be noted that the physical characterisation of the *Co* locus and the resulting list of coding sequences is based exclusively on the 'Golden Delicious' genetic background and haplotypes, and not on those of 'McIntosh' or 'McIntosh Wijcik'. Therefore, considering the relatively high rate of genetic diversity present in apple, including a supposedly high level of presence/absence gene variation between different haplotypes (Velasco et al. 2010), the physical structure of the locus and its coding content described here should be treated with caution. This will only be clarified by sequencing the *Co*-mutated haplotype in 'Wijcik'. As a matter of fact, the construction of a BAC library starting from 'Wijcik' leaf material allowed the identification of four BAC clones completely spanning the putative *Co* region. This can be considered a good indication about the feasibility of the subsequent characterisation step. According to the preliminary results presented here, it seems that the structure of 'Wijcik' *Co* region could be quite similar to the known one of 'Golden Delicious', as all the BAC ends could be correctly found and positioned on the 'Golden Delicious' reference genome. Further indications about the length and the gene content of the 'Wijcik' *Co* region will come only after the complete sequencing of the BAC clones.

In conclusion, the data obtained in this work allowed us to identify a region of 393 kb on apple LG10 putatively containing the main gene responsible for the columnar habit tree trait. According to the genomic sequence of 'Golden Delicious', 36 protein-coding genes were predicted and annotated in the region, including at least seven different genes coding for protein functions which can be related to plant development. We consider these results to be a valuable starting point for future positional cloning steps as indicated also by obtaining the physical coverage of the 'Wijcik' *Co* region. In order to actually identify the mutation responsible for the columnar habit and produce a reliable candidate gene, a more extensive analysis is still required, possibly including the comparison with the corresponding genomic region of standard 'McIntosh'.

Acknowledgments

The authors would like to thank Dr. Pierluigi Magnago for his essential contribution in developing the segregating populations and Luca Pinelli for his technical help.

References

Aida, M., T. Ishida, et al. (1997). "Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant." The Plant Cell Online **9**(6): 841-857.

Bai, T., Y. Zhu, et al. (2012). "Fine genetic mapping of the Co locus controlling columnar growth habit in apple." Molecular Genetics and Genomics **287**(5): 437-450.

Burge, C. and S. Karlin (1997). "Prediction of complete gene structures in human genomic DNA." Journal of molecular biology **268**(1): 78-94.

Busov, V. B., R. Meilan, et al. (2003). "Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature." Plant Physiology **132**(3): 1283-1291.

Chagné, D., C. M. Carlisle, et al. (2007). "Mapping a candidate gene (MdMYB10) for red flesh and foliage colour in apple." BMC genomics **8**(1): 212.

Conner, P. J., S. K. Brown, et al. (1997). "Randomly amplified polymorphic DNA-based genetic linkage map of three apple cultivars." Journal of American Society for Horticultural Science **122**(3): 350-359.

Cook, D. R. and R. K. Varshney (2010). "From genome studies to agricultural biotechnology: closing the gap between basic plant science and applied agriculture." Current Opinion in Plant Biology **13**(2): 115-118.

Czemmel, S., R. Stracke, et al. (2009). "The grapevine R2R3-MYB transcription factor VvMYBF1 regulates flavonol synthesis in developing grape berries." Plant Physiology **151**(3): 1513-1530.

De Wit, I., N. Cook, et al. (2003). Characterization of tree architecture in two-year-old apple seedling populations of different progenies with a common columnar gene parent. XI Eucarpia Symposium on Fruit Breeding and Genetics 663.

Doligez, A., A.-F. Adam-Blondon, et al. (2006). "An integrated SSR map of grapevine based on five mapping populations." Theoretical and Applied Genetics **113**(3): 369-382.

Feller, A., K. Machemer, et al. (2011). "Evolutionary and comparative analysis of MYB and bHLH plant transcription factors." The Plant Journal **66**(1): 94-116.

Flachowsky, H., C. Hattasch, et al. (2010). "Overexpression of LEAFY in apple leads to a columnar phenotype with shorter internodes." Planta **231**(2): 251-263.

Galli, P., G. A. L. Broggini, et al. (2010). "High-resolution genetic map of the Rvi15 (Vr2) apple scab resistance locus." Molecular Breeding **26**(4): 561-572.

Guo, M., J. Thomas, et al. (2008). "Direct repression of KNOX loci by the ASYMMETRIC LEAVES1 complex of Arabidopsis." The Plant Cell **20**(1): 48-58.

Hemmat, M., N. F. Weeden, et al. (1997). "A DNA marker for columnar growth habitat in apple contains a simple sequence repeat." Journal of American Society for Horticultural Science **122**: 347-349.

Janick, J. and J. N. Moore (1996). Fruit breeding, tree and tropical fruits, John Wiley & Sons.

Kelsey, D. and S. Brown (1992). "'McIntosh Wijcik': a columnar mutation of'McIntosh'apple proving useful in physiology and breeding research." Fruit varieties journal (USA).

Kenis, K. and J. Keulemans (2003). QTL analysis of growth characteristics in apple. XI Eucarpia Symposium on Fruit Breeding and Genetics 663.

Kenis, K. and J. Keulemans (2007). "Study of tree architecture of apple (*Malus* x *domestica* Borkh.) by QTL analysis of growth traits." Molecular Breeding **19**(3): 193-208.

Khan, M. A., C. E. Durel, et al. (2007). "Development of molecular markers linked to the 'Fiesta' linkage group 7 major QTL for fire blight resistance and their application for marker-assisted selection." Genome **50**(6): 568-577.

Kim, M. Y., K. J. Song, et al. (2003). "Development of RAPD and SCAR markers linked to the Co gene conferring columnar growth habit in apple (Malus pumila Mill.)." The Journal of Horticultural Science & Biotechnology **78**(4): 512-517.

Kitomi, Y., H. Ito, et al. (2011). "The auxin responsive AP2/ERF transcription factor CROWN ROOTLESS5 is involved in crown root initiation in rice through the induction of OsRR1, a type-A response regulator of cytokinin signaling." The Plant Journal **67**(3): 472-484.

Lapins, K. (1974). "Spur type growth habit in 60 apple progenies [Genetic transmission]." Journal American Society for Horticultural Science.

Lapins, K. and R. Watkins (1972). "Genetics of compact growth habit." Report of east malling research station for **136**.

Lapins, K. O. (1976). "Inheritance of compact growth type in apple." Journal of American Society for Horticultural Science **101**: 133-135.

Lauri, P. and J. Lespinnasse (1992). The relationship between cultivar fruiting-type and fruiting branch characteristics in apple trees. V International Symposium on Orchard and Plantation Systems 349.

Lee, S., S. Lee, et al. (2006). "Overexpression of PRE1 and its homologous genes activates gibberellin-dependent responses in Arabidopsis thaliana." Plant and cell physiology **47**(5): 591-600.

Lespinasse, D. (1992). Le pommier. Amélioration des espèces végétales cultivées: objectifs et critères de sélection. A. Gallais and H. Bannerot. Paris, Editions Quae**:** 579–594.

Lespinasse, J. and J. Delort (1984). Apple tree management in vertical axis: appraisal after ten years of experiments. III International Symposium on Research and Development on Orchard and Plantation Systems 160.

Liebhard, R., B. Koller, et al. (2003). "Creating a saturated reference map for the apple (Malus x domestica Borkh.) genome." Theor Appl Genet **106**(8): 1497-1508.

Moriya, S., H. Iwanami, et al. (2009). "Development of a marker-assisted selection system for columnar growth habit in apple breeding." Journal of Japanese Society for Horticultural Science **78**: 279-287.

Moriya, S., K. Okada, et al. (2012). "Fine mapping of Co, a gene controlling columnar growth habit located on apple (Malus×domestica Borkh.) linkage group 10." Plant Breeding **131**(5): 437-450.

Mudunuri, S. B. and H. A. Nagarajaram (2007). "IMEx: Imperfect Microsatellite Extractor." Bioinformatics **23**(10): 1181-1187.

Ooijen, J. W. and R. Voorrips (2002). JoinMap: version 3.0: software for the calculation of genetic linkage maps, University and Research Center.

Patocchi, A., B. Vinatzer, et al. (1999). "Construction of a 550 kb BAC contig spanning the genomic region containing the apple scab resistance gene Vf." Molecular and General Genetics MGG **262**(4-5): 884-891.

Sablowski, R. W. and E. M. Meyerowitz (1998). "A Homolog of NO APICAL MERISTEM Is an Immediate Target of the Floral Homeotic Genes APETALA3/PISTILLATA." Cell **92**(1): 93-103.

Schuelke, M. (2000). "An economic method for the fluorescent labeling of PCR fragments." Nature Biotechnology **18**(2): 233-234.

Segura, V., C. Cilas, et al. (2008). "Dissecting apple tree architecture into genetic, ontogenetic and environmental effects: mixed linear modelling of repeated spatial and temporal measures." New Phytologist **178**(2): 302-314.

Silfverberg-Dilworth, E., C. L. Matasci, et al. (2006). "Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome." Tree genetics & genomes **2**(4): 202 - 224.

Souer, E., A. van Houwelingen, et al. (1996). "The No Apical Meristem Gene of Petunia Is Required for Pattern Formation in Embryos and Flowers and Is Expressed at Meristem and Primordia Boundaries." Cell **85**(2): 159-170.

Stanke, M., R. Steinkamp, et al. (2004). "AUGUSTUS: a web server for gene finding in eukaryotes." Nucleic Acids Res **32**(suppl 2): W309-W312.

Tani, E., E. Flemetakis, et al. (2011). "study of a SPATULA-like bHLH transcription factor expressed during peach (Prunus persica) fruit development." Plant Physiol Biochem. **49**: 654-663.

Tao, Q., A. Wang, et al. (2002). "One large-insert plant-transformation-competent BIBAC library and three BAC libraries of Japonica rice for genome research in rice and other grasses." Theoretical and Applied Genetics **105**(6-7): 1058-1066.

Tobutt, K. R. (1985). "Breeding columnar apples at East Malling." Acta Horticultura **159**: 63-68.

Troggio, M., A. Gleave, et al. (2012). "Apple, from genome to breeding." Tree genetics & genomes **8**(3): 509-529.

Velasco, R., A. Zharkikh, et al. (2010). "The genome of the domesticated apple (Malus x domestica Borkh.)." Nature Genetics **42**(10): 833-839.

Vezzulli, S., M. Troggio, et al. (2008). "A reference integrated map for cultivated grapevine (Vitis vinifera L.) from three crosses, based on 283 SSR and 501 SNP-based markers." Theoretical and Applied Genetics **117**(4): 499-511.

Xie, Q., G. Frugis, et al. (2000). "Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development." Genes & Development **14**(23): 3024-3036.

Yamagishi, K., K. Tatematsu, et al. (2009). "CHOTTO1, a double AP2 domain protein of Arabidopsis thaliana, regulates germination and seedling growth under excess supply of glucose and nitrate." Plant and cell physiology **50**(2): 330-340.

Yano, R., Y. Kanno, et al. (2009). "CHOTTO1, a putative double APETALA2 repeat transcription factor, is involved in abscisic acid-mediated repression of gibberellin biosynthesis during seed germination in Arabidopsis." Plant Physiology **151**(2): 641-654.

Zhang, W., Y. Sun, et al. (2006). "Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor." Development **133**(16): 3085-3095.

Zhang, Y., G. Cao, et al. (2009). "Characterization of Arabidopsis MYB transcription factor gene AtMYB17 and its possible regulation by LEAFY and AGL15." Journal of Genetics and Genomics **36**(2): 99-107.

Zhang, Y. and H. Dai (2011). "Comparison of photosynthetic and morphological characteristics, and microstructure of roots and shoots, between columnar apple and standard apple trees of hybrid seedlings." Phyton-Revista Internacional de Botanica Experimental: 119.

Zhu, Y. D., W. Zhang, et al. (2007). "Evaluation of inter-simple sequence repeat analysis for mapping the Co gene in apple (Malus pumila Mill.)." The Journal of Horticultural Science & Biotechnology **82**(3): 371-375

Supporting material

Table S1. Putative ORFs predicted inside the *Co* region. In bold the transcripts belonging to possible candidate genes.

Sequence name	Genome Browser	ORF (bp)	AUGUSTUS	ORF (bp)	GENSCAN	ORF (bp)
			integrase	1348	integrase	660
MDP0000897594	**MYB transcription factor**	3234	**putative MYB transcription factor**	690	**putative MYB transcription factor**	936
MDP0000326311	mRNA splicing factor	123	pre-mRNA-splicing factor SLU7	1623	pre-mRNA-splicing factor SLU7-A-like	1656
MDP0000524262	PABP-interacting motif	395				
			putative COBL7	432	putative COBL7	927
MDP0000136858	DNA mismatch repair protein	432	DNA mismatch repair protein Msh2	789	DNA mismatch repair protein Msh2	1215
MDP0000367163	phosphate transporter	120	phosphate transporter, putative	627	inorganic phosphate transporter, putative	627
MDP0000284965	actin depolymerizing factor	7218	actin depolymerizing factor	441	actin depolymerizing factor	399
			UPF0420 protein C16orf58 homolog	1380	UPF0420 protein C16orf58 homolog	1434
MDP0000508371	phosphate transporter	1616	phosphate transporter, putative	1617	phosphate transporter	864
					putative reverse transcriptase	1353
MDP0000186457	maintenance of chromosome 1	3957				
MDP0000855671	**AP2 transcription factor**	752	**AP2/ERF transcription factor**	759	**AP2/ERF transcription factor**	864
			AP2/ERF transcription factor	759	**AP2/ERF transcription factor**	1527
			putative reverse transcriptase	489		
			gag-pol polyprotein	1926	Putative gag-pol polyprotein	1130
			putative polyprotein	2205		
					putative Zinc knuckle containing protein	2061
MDP0000187369	**AP2 transcription factor**	655	**eth-responsive transcription factor**	438	**ethylene-responsive transcription factor**	453
MDP0000286915	**AP2 transcription factor**	5151	**eth-responsive transcription factor**	720	**ethylene-responsive transcription factor**	498
			ribonuclease E	1650	ribonuclease E	1671
			RING-H2 finger protein ATL79	501		
			putative COBL7	546	putative COBL7	735

			hyoscyamine 6-dioxygenase-like	1020		
			putative reverse transcriptase	1275	putative reverse transcriptase	2076
			putative COBL7	537	putative COBL7	1581
MDP0000187760	protein autophagy 9	1496	autophagy-related protein 9-like	2562	autophagy-related protein 9-like	2460
MDP0000927091	homology to kinesin light chain	4840	kinesin light chain, putative	1515	kinesin light chain, putative	864
MDP0000287209	ring finger protein	3363	RING-H2 finger protein ATL1-like	723	ring finger protein, putative	1905
			putative reverse transcriptase	2029	putative reverse transcriptase	1086
			auxin-induced protein 5NG4	1050	**auxin-induced protein 5NG4**	639
MDP0000934866	gal-1-phosphate uridylyltransferase	2457	gal-1-phosphate uridyltransferase	995		
			formate-tetrahydrofolate ligase	780	formate-tetrahydrofolate ligase, putative	1443
MDP0000139773	**NAM transcription factor**	1044				
MDP0000934869	**basic helix-loop-helix protein**	535	**basic helix-loop-helix protein**	2211	**basic helix-loop-helix-containing protein**	2160
					gal-1-phosphate uridyl transferase	768

Chapter 3

Evidence for regulation of columnar habit in apple by a putative 2OG-Fe(II) oxygenase

New Phytologist **200 (4)** (2013, p. 993-999)

Pieter J. Wolters, Henk J. Schouten, Riccardo Velasco, Azeddine Si-Ammour, Paolo Baldi

Summary

- **Understanding the genetic mechanisms controlling columnar-type growth in the apple mutant 'Wijcik' will provide insights on how tree architecture and growth are regulated in fruit trees.**
- **In apple, columnar-type growth is controlled by a single major gene at the *Columnar* (*Co*) locus. By comparing the genomic sequence of the *Co* region of 'Wijcik' with its wild-type 'McIntosh', a novel non-coding DNA element of 1956 bp specific to Pyreae was found to be inserted in an intergenic region of 'Wijcik'.**
- **Expression analysis of selected genes located in the vicinity of the insertion revealed the upregulation of the *MdCo31* gene encoding a putative 2OG-Fe(II) oxygenase in axillary buds of 'Wijcik'.**
- **Constitutive expression of *MdCo31* in *Arabidopsis thaliana* resulted in compact plants with shortened floral internodes, a phenotype reminiscent of the one observed in columnar apple trees. We conclude that *MdCo31* is a strong candidate gene for the control of columnar growth in 'Wijcik'.**

Keywords: Tree architecture, columnar gene (*Co*), *Malus x domestica* (apple), 'Wijcik', 2OG-Fe(II) oxygenase.

Introduction

Tree architecture exerts a considerable influence on fruit yield and directly affects the amount of work required for orchard maintenance, including pruning, training and tree support systems which represent a major cost for apple growers (Rom and Barritt 1990; Barritt 1992; Lespinasse and Delort 1993; Wünsche and Lakso 2000). Columnar trees such as the apple mutant 'Wijcik' have been proposed as an interesting solution for creating high-density orchards, despite the columnar habit being linked to undesirable traits such as alternate bearing (Looney and Lane 1984; Davenport 2000; Kenis and Keulemans 2007). 'Wijcik' was first identified in the 1960s as a shoot of a normal 'McIntosh' tree and exhibits a pronounced columnar phenotype with short internodes and a thick stem. Axillary buds usually develop into spurs rather than lateral branches, resulting in a tree bearing fruits close to the stem (Fisher 1969; Tobutt 1985). Nevertheless, the development of long side shoots from spurs also occurs in columnar trees, for instance when the central leader is cut or damaged. In this particular case, some

spurs near the top can grow and give rise to shoots exhibiting a columnar phenotype, suggesting that lateral buds are under tight apical control (Petersen and Krost 2013).

The columnar phenotype was shown to be caused by a single, dominant, allele named *Co*, mapping on chromosome 10 (Conner et al. 1997). Different studies helped to narrow down the genetic window for the *Co* region (Bai et al. 2012; Moriya et al. 2012; Baldi et al. 2013). Baldi et al. (2013) identified a candidate region delimited by the SSR markers Co04R10 and Co04R13, which spans 393 kb in the homologous region of the 'Golden Delicious' genome sequence (Velasco et al. 2010) and overlaps with the 196 kb region identified by Moriya et al. (2012).

In parallel, others attempted to identify the *Co* gene using gene expression profiling with RNAseq, which resulted in the identification of several candidates for the *Co* gene with roles in plant hormone signalling, in particular DELLA proteins involved in the regulation of the gibberellin pathway (Zhang et al. 2012). A similar approach was used to propose a correlation between cytokinins and IAA (indole-3-acetic acid) and the columnar phenotype (Krost et al. 2013; Petersen and Krost 2013).

Understanding the genetic mechanisms controlling columnar growth habit will help design novel strategies to develop new apple cultivars for high-density planting while avoiding the introduction of undesirable characters linked to the columnar locus.

Thus, in this investigation we compared the sequences of a BAC library containing the *Co* region of the mutant 'Wijcik' with the one prepared from the corresponding wild-type cultivar 'McIntosh' and an insertion of 1956 bp was identified in the 'Wijcik' genome which was absent in the wild-type.

We analysed the expression patterns of the six genes identified within the 50 kb region surrounding the insertion and found only one gene (*MdCo31*) encoding a putative 2OG-Fe(II) oxygenase which was differentially expressed between 'McIntosh' and 'Wijcik' buds. We then investigated the biological function of this gene by expressing it constitutively in *Arabidopsis thaliana* and observed that transformed plants exhibited phenotypes reminiscent of the one observed in

'Wijcik, providing an evidence for the role of *MdCo31* in regulating the columnar-type growth in apple.

Materials and methods

Plant material and growth conditions

Apple (*Malus x domestica* Borkh.) trees were grown and maintained at the orchard "Giaroni" belonging to the Fondazione Edmund Mach (FEM) at the Istituto Agrario di San Michele all'Adige (IASMA) located in Italy (latitude 46.181539°, longitude 11.119877°). 'McIntosh' and 'Wijcik' trees (Fisher 1969) were obtained from the department of fruit trees and woody plant species at the University of Bologna (DCA-UNIBO, Italy) and grafted on 'M9' rootstock. The cross between 'Wijcik' and 'Golden Delicious' was performed using 'Wijcik' pollen, giving rise to a segregating population of 103 plants (**Fig. S1**). Axillary buds from six-year old 'Wijcik' x 'Golden Delicious' trees and two-year old 'McIntosh' and 'Wijcik' trees were harvested at the end of March and leaves from the same trees in May. Arabidopsis plants Col-0 used in this study were grown in GS90 soil (Manna Italia) and raised at 22 °C (16 h of 100 µmol m^{-2} s^{-1} light, 8 h of dark).

BAC library construction and sequencing

The 'McIntosh' BAC library was prepared by Amplicon express (Pullman, WA, USA) from leaf material and using the same method as described for the 'Wijcik' BAC library (Baldi et al. 2013). The SSR markers used in Baldi et al. (2013) and novel ones indicated in **Table S1** were used to screen the BAC libraries. BAC plasmids were purified using the Plasmid Midi kit (QIAGEN) and sequenced on a GS FLX Titanium platform (Roche). The reads were assembled using Newbler Assembler software (Roche) and MIRA assembler (Chevreux et al. 1999), using the 'Golden Delicious' genome as a reference. The resulting contigs were joined together using Sanger sequencing. The assembled sequences of 'McIntosh' and 'Wijcik' (NCBI Accession number KF530875 and KF530876) were compared by alignment using SSAHA2 (Ning et al. 2001) and visualised using IGV (Robinson et al. 2011).

GlimmerHMM was used to identify ORFs in the 'Wijcik' *Co* region sequence (Majoros et al. 2004). Gene predictions and full-length cDNA sequences from the

'Golden Delicious' *Co* region were used as a reference. Predicted genes in the *Co* region were annotated by performing a BLASTx search against the Arabidopsis protein database (Rhee et al. 2003) and by carrying out a conserved protein domain search using NCBI's Conserved Domain Database (Marchler-Bauer et al. 2011). A BLAST search in GeneBank databases for the 1956 bp insertion sequence was performed. Matches were found only on the apple and pear genome (Wu et al. 2013). Apple and pear genome contigs containing the insertion were aligned using MEGA 5.1 (Tamura et al. 2011).

RNA extraction and qRT-PCR analysis

Total RNA was isolated from apple buds and leaves using the Plant spectrum kit (Sigma-Aldrich) and from Arabidopsis leaves using TRI reagent (Sigma-Aldrich). cDNAs were synthesised from 1 µg of total RNA, previously subjected to DNAse treatment, using the superscript VILO cDNA synthesis kit according to the manufacturer's instructions (Life Technologies). Real-Time PCR reactions were performed using SYBR green chemistry (Platinum SYBR Green qPCR SuperMix-UDG, Life Technologies) with the primer pairs listed in **Table S1.** Transcript levels were normalised to the reference genes Actin (Li and Yuan 2008) and *Md_4592:1:a* (Botton et al. 2011) for apple, and *TIP41-like* (Kutter et al. 2007) for Arabidopsis. Data were analysed using the comparative Ct method (Pfaffl 2001).

Binary vector construction and plant transformation

MdCo31 was amplified from total RNA of 'Wijcik' using the SuperScript One-Step RT-PCR System (Life Technologies) and the primers indicated in **Table S1** and cloned into the pENTR/D vector using the pENTR-D/TOPO cloning kit (Life Technologies). pENTR/D-MdCo31 was recombined with the pH2GW7 vector (Karimi et al. 2002) using the Gateway LR Clonase II kit (Life Technologies) to give rise to pPro35S:MdCo31. Arabidopsis transformants were produced by the floral dip method using *Agrobacterium tumefaciens* strain GV3101 carrying the T-DNA binary vector pPro35S:MdCo31 (Clough and Bent 1998).

Development and testing of SSR markers

The 'Golden Delicious' genome was used to search for potential SSRs using IMEx software (Mudunuri & Nagarajaram, 2007). Potential SSRs were tested on genomic DNA of 96 columnar and standard progeny trees deriving from a 'Wijcik' x 'Golden Delicious' cross performed in 2006 in order to confirm co-segregation of the marker with the columnar phenotype. PCR amplification of SSRs was performed in a volume of 10 µl in 1x PCR buffer containing 1 µl of DNA template, 0.2 mM dNTPs, 3mM MgCl2, 0.5 units of AmpliTaq Gold DNA polymerase (Life Technologies) and 0.1 mM of each primer (the forward primer was 6-FAM labelled at the 5' end). The cycling conditions include an initial denaturation step of 10 min at 95 °C followed by 35 cycles of 30 s at 95 °C, 1 min at 60 °C, 30 s at 72 °C and a final extension phase of 10 min at 72 °C. PCR cycling was carried out using an Applied Biosystems 2720 Thermal Cycler. PCR products were separated and detected using a 3730 DNA analyzer (Applied Biosystems). The size of the PCR products were determined by comparison with the GeneScan-500 LIZ Size Standard (Applied Biosystems) using GeneMapper v4.0 software (Applied Biosystems).

Creation of BAC contigs covering the *Co* region

Plate pools were prepared from all BAC plates of the 'McIntosh' and 'Wijcik' BAC libraries. DNA was isolated using a modified alkaline lysis protocol. Briefly, after resuspending the bacterial pellet of 1.5 ml culture in 150 µl of resuspension buffer (50 mM Tris-HCl, 10 mM EDTA, pH=8.0), 150 µl of lysis buffer (0,2M NaOH, 1% SDS) was added. After an incubation of 5 min, 150 µl of cold neutralisation buffer (3M K-acetate, pH=5.5) was added and the suspension filtered by centrifugation using a UNIFILTER microplate (Whatman) on top of a UNIPLATE (Whatman). The flow-through was collected in the UNIPLATE and DNA was precipitated by adding 0.7 volume of isopropanol. The DNA pellet was collected by centrifugation, washed with 70% ethanol, and dissolved in 20 µl of H_2O. The plate pools were screened using SSR markers developed for the *Co* region. Row and column pools were prepared from the positive plates and screened in order to identify the positive clones. The PCR conditions were the same as described above for the SSR analysis. BAC ends of the positive clones were sequenced

using M13 forward and M13 reverse primers. The BAC end sequences and SSR marker information was used to anchor the BAC clones on the 'Golden Delicious' genome.

Sequencing and assembly of BAC insert sequences

BAC plasmids were purified from 250 ml cultures using the Plasmid Midi kit (QIAGEN). One µg of each BAC plasmid was used for sequencing on a 454 GS FLX Titanium sequencer (Roche). After cleaning and trimming of the reads, a hybrid assembly was performed, combining the results of the Newbler Assembler software (Roche) and MIRA assembler (Chevreux et al., 1999). The resulting contigs were ordered using the 'Golden Delicious' genome as a reference. In order to join the contigs, primers were designed on the contig ends (**Table S1**) and used for Sanger sequencing, thereby filling the gaps between the contigs. When first compared, the assembled *Co* sequence of 'McIntosh' and 'Wijcik' revealed a small number of differences, mainly SNPs. By visually comparing the sequences with the underlying 454 reads, almost all differences were eliminated. Other differences that were difficult to assess using 454 reads were checked by Sanger sequencing and corrected manually. After this last correction, the 1956 bp insertion in 'Wijcik' remained as the only difference compared to McIntosh.

RNA analysis and qRT-PCR

The integrity of the RNA was assessed by loading 1µL of RNA on an agarose gel and RNA concentrations were measured using the Nanodrop 8000 instrument (Thermo Scientific). cDNA was synthesised from 1µg of RNA, using the VILO cDNA synthesis kit (Invitrogen). The reaction was first incubated at room temperature for 10 min, then for 1 h at 42 °C and the reaction was terminated by an incubation for 5 min at 85 °C.

The real-time PCR reactions were carried out on a ViiA 7 real-time PCR system (Life Technologies), using SYBR green chemistry (Platinum SYBR Green qPCR SuperMix-UDG, Life Technologies). 1 µL of diluted cDNA (1:5) was used in each PCR reaction. Reaction conditions included an incubation for 2 min at 50 °C, followed by 2 min at 95 °C and 40 cycles of an incubation at 95 °C for 15 s followed by 30 s at 60 °C. Ct values were calculated by the ViiA 7 software based on the Ct

values obtained from three technical replicates per sample. Specific amplification of the template was confirmed by including a non-template control for each primer pair and by including a melt-curve analysis at the end of the real-time PCR reaction. For the analysis comparing the expression levels of candidate *Co* genes between wild-type 'McIntosh' and 'Wijcik', three independent biological replicates were used to calculate the expression values. For the gene expression analysis concerning the segregating population, the three independent biological replicates are displayed separately. The cDNA from each sample comes from a mix of plant material derived from 5-8 intact axillary buds at the green tip stage collected from different branches of the same tree. A number of reference genes from literature were tested as house-keeping genes. Actin (Li & Yuan, 2008) and *Md_4592:1:a* (Botton et al., 2011) were found to be the most stably expressed in the studied experimental conditions and were therefore used as reference genes. The results were analysed using the comparative Ct method (Pfaffl, 2001), taking into account the primer efficiencies of all the genes tested, which were calculated using dilution curves. The relative *MdCo31* expression levels in the Pro35S:MdCo31 Arabidopsis lines were normalised to Col-0 using TIP41-like as a reference gene (Kutter et al., 2007).

Results and discussion

BAC clones covering the genomic *Co* region of 'Wijcik' and 'McIntosh' were identified by screening the respective BAC libraries with specific markers indicated in **Table S1**. The BAC ends of all positive clones in coupling phase with *Co* were sequenced and used to anchor the corresponding BACs to the 'Golden Delicious' genome creating a minimum tiling path of BAC clones spanning the complete *Co* region, except for a small portion of 13 kb that was not covered by the 'McIntosh' BAC library (**Fig. 1a**). The selected BAC clones were sequenced using 454 with an average coverage of 100x, resulting in a first assembly composed of 25 contigs for the complete *Co* region of 'Wijcik' and only six for the corresponding region in 'McIntosh'. The gaps between contigs and the 13 kb segment not covered by the 'McIntosh' BAC clones were sequenced by Sanger sequencing. Thus, the sequences of the 'Wijcik' and 'McIntosh' *Co* region were covered by two contigs each, named WiC1/WiC2 and McC1/McC2 respectively (**Fig. 1a**). In both

genotypes, the two contigs could not be joined as they were separated by a long sequence of repetitive DNA (**Fig. 1a**).

The sequences of WiC1/WiC2 and McC1/McC2 were compared. 43 genes, named *MdCo1* to *MdCo43*, were predicted in the *Co* region encoding proteins and enzymes with diverse functions (**Table S2**). As 'Wijcik' originated from a somatic mutation in 'McIntosh' (Fisher 1969), we expected the two sequences to be almost identical, apart from the mutation that led to columnar growth. Indeed, after correction of sequencing errors, the only difference found was an insertion of 1956 bp in 'Wijcik' (**Fig. 1a**). This insertion falls within the smaller putative *Co* region described by Moriya et al. (2012), who used segregating populations from different parents, in agreement with our results.

In order to confirm the presence of the insertion in the 'Wijcik' genome and to exclude possible artefacts that could have occurred during BAC library construction or during the assembly of 454 reads, selective amplifications were performed on genomic DNA of 'McIntosh' and 'Wijcik' using specific primers indicated in **Table S1**. When primers spanning the genomic flanking regions and the insert sequence were designed at the 5' and 3' ends of the insertion, specific PCR products of 586 bp (5' CR) and of 319 bp (TC/CR 3') respectively, were obtained only in 'Wijcik' (**Fig. 1b**). When amplification was performed using primers designed in the core region (CR/TC), a PCR product of 931 bp was obtained in 'McIntosh' as well, suggesting the presence of similar elements at other locations in the apple genome (**Fig. 1b**). A BLAST search against the apple reference genome confirmed that more than 250 sequences with similarities to the 'Wijcik' insertion are distributed throughout the *Malus* chromosomes. Comparison of these sequences by alignment revealed that the 'Wijcik' insertion has a conserved sequence of ca. 800 bp at the 5' end, followed by a less conserved TC-rich region and a highly conserved stretch of 160 bp at the 3' end (**Fig. 1a,** Supporting information**, Fig. S2**). Therefore, the 'Wijcik' insertion might be a novel type of mobile DNA element, lacking the terminal inverted repeats (TIR) that are normally present in transposons (Bennetzen 2000). Moreover, it seems to be conserved only in Pyreae since similar sequences were found in *Pyrus.*

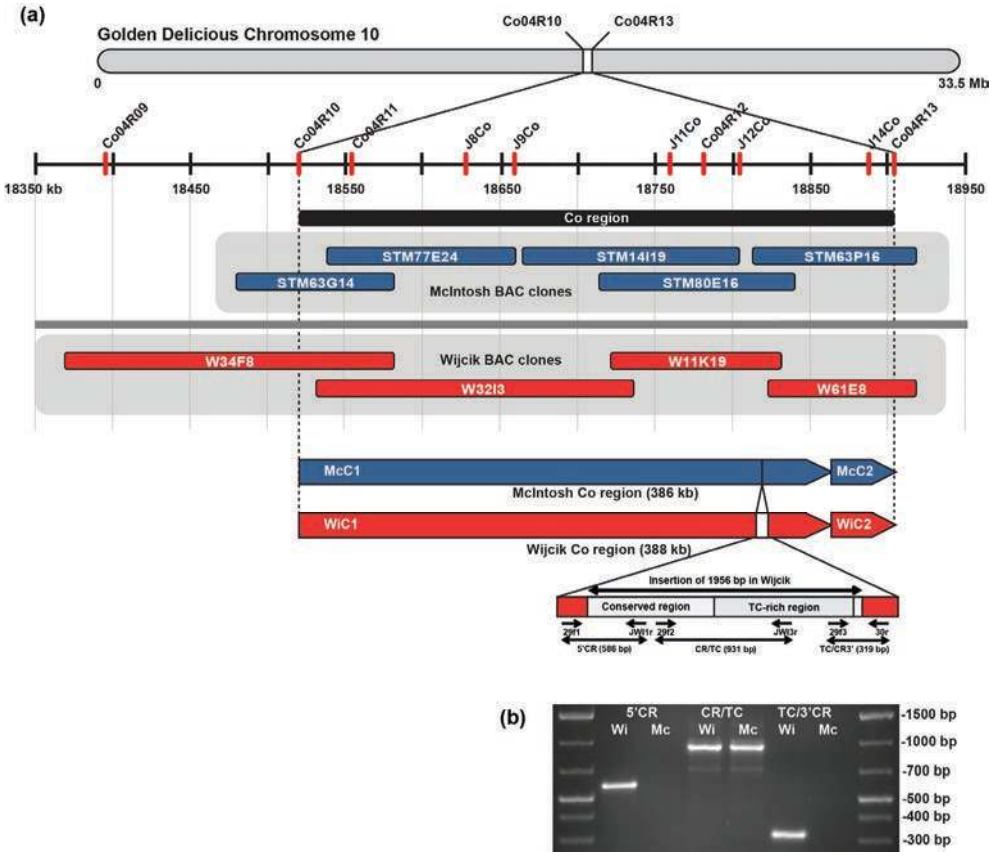

Figure 1. BAC sequence assembly and identification of 'Wijcik' insertion. A. The location of the *Co* region (delimited by SSR markers Co04R10 and Co04R13) on chromosome 10 of Golden Delicious is shown at the top of the figure. Just below, an enlargement of the *Co* region is shown, indicating the position (on the Golden Delicious genome) of the SSR markers used to screen the 'McIntosh' and 'Wijcik' BAC library. The assembly of the BAC clones covering the *Co* region resulted in two contigs both in 'McIntosh' ('McC1' and 'McC2', displayed in blue) and 'Wijcik' ('WiC1' and 'WiC2', displayed in red). The location of the insertion in 'Wijcik' is indicated on McC1 and on WiC1. Primers used to detect the presence of the insertion in 'Wijcik' and 'McIntosh' and the names of the corresponding amplification products are indicated on the 'Wijcik' insertion. **B.** Specific amplification of the insert sequence from genomic DNA of 'Wijcik' ('Wi') and 'McIntosh' ('Mc'). The different amplification products are indicated above each lane. When one primer was designed outside and one inside the insert sequence, amplification products (fragments '5'CR' and 'TC/CR3'') were only detected in Wijcik. When both primers were designed within the insert sequence, an amplification product (fragment 'CR/TC') was detected in 'Wijcik' as well as in McIntosh.

bretschneideri Rehd., but not in other Rosaceae genera for which whole genome sequences exist (**Fig. S2**). The fact that both ends are highly conserved suggests that the insertion might occur through a yet unknown transposition-based mechanism.

The 'Wijcik' insertion occurred in an intergenic region and therefore did not disrupt the coding region of any gene. Moreover, the insertion itself does not contain any open reading frame, raising the question on how the insertion causes the columnar phenotype of 'Wijcik'. One possible explanation is that the insertion influences the expression of neighbouring genes. To verify this hypothesis, we tested by qRT-PCR the expression of all genes located 25 kb upstream and downstream of the insertion. Unlike standard trees, most of the buds present on columnar stems develop into short spurs instead of lateral branches, suggesting that regulation of the columnar phenotype takes place at an early stage of bud development. For this reason, young buds at the green tip stage were selected for expression analysis and leaf tissue was used as a control. Among the six genes located within the 50 kb region containing the insertion (**Fig. 2a**), only *MdCo31* (NCBI Accession number KF562006) was found to be differentially expressed, showing a 14-fold induction in 'Wijcik' buds compared to 'McIntosh', whereas the mRNA levels in leaves of both genotypes were negligible (**Fig. 2b**). This suggests an involvement of *MdCo31* in controlling the columnar phenotype. In addition, we checked the expression of all transcription factors predicted inside the complete 386 kb *Co* region (*MdCo4, MdCo16, MdCo25* and *MdCo26*) (**Table S2**), as they could have a potential role in plant development. No differences between 'McIntosh' and 'Wijcik' trees were found (data not shown), therefore excluding a role for such genes in controlling columnar phenotype.

The increased expression of *MdCo31* in axillary buds of 'Wijcik' is in agreement with the dominant inheritance of the columnar phenotype. Moreover, the fact that the expression of *MdCo31* seems to be bud-specific is in accordance with the supposed function of controlling shoot development. Additionally, a strict correlation between the presence of the insertion and an increased expression of *MdCo31* in columnar trees compared to standard trees of a segregating population derived from a 'Golden Delicious' x 'Wijcik' cross confirmed that high expression of *MdCo31* in buds is a general feature of columnar trees and not a peculiar feature of 'Wijcik' (**Fig. 3**).

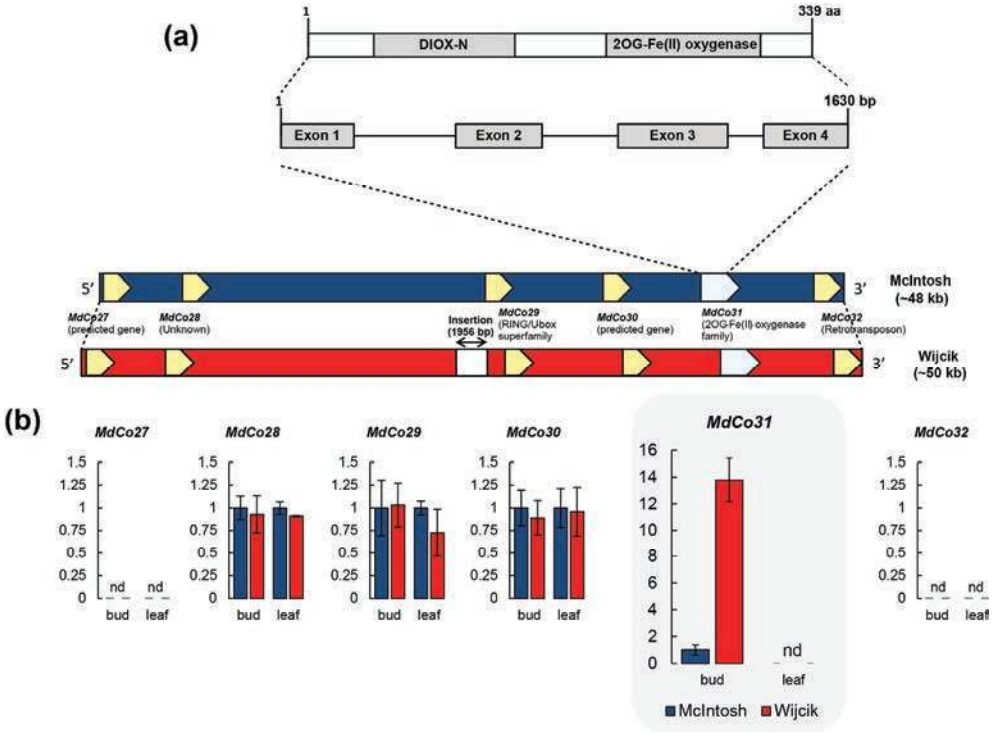

Figure 2. Annotation of insert region and qPCR results. A. 'Wijcik' 50 Kb region (red bar) containing the insertion, the homologous region from 'McIntosh' (blue bar) and structure of *MdCo31*. The conserved DIOX_N like and 2OG-FeII_oxygenase like domains together with the predicted exons are shown (grey boxes). The position of the predicted genes within the candidate region is indicated by yellow arrows. *MdCo31* is depicted in light blue. **B.** qPCR analysis of the predicted genes within the candidate region in bud and leaf material from 'McIntosh' and Wijcik. Expression levels are calculated based on three biological replicates and the standard deviations are indicated by error bars. Gene expression in 'Wijcik' (red bar) is normalised against the expression of the corresponding gene in 'McIntosh' (blue bar). Samples for which the expression level is below the limit for reliable quantification are marked 'nd' (not detectable).

The 5' CR and TC/CR 3' PCR products could be used as molecular markers in marker-assisted selection for columnar-type growth in apple. Such a marker could be used to predict columnar growth in segregating progenies with absolute accuracy, and would enable apple breeders to identify plants carrying the *Co* locus from 'Wijcik' even when offspring of 'McIntosh' were involved in the cross.

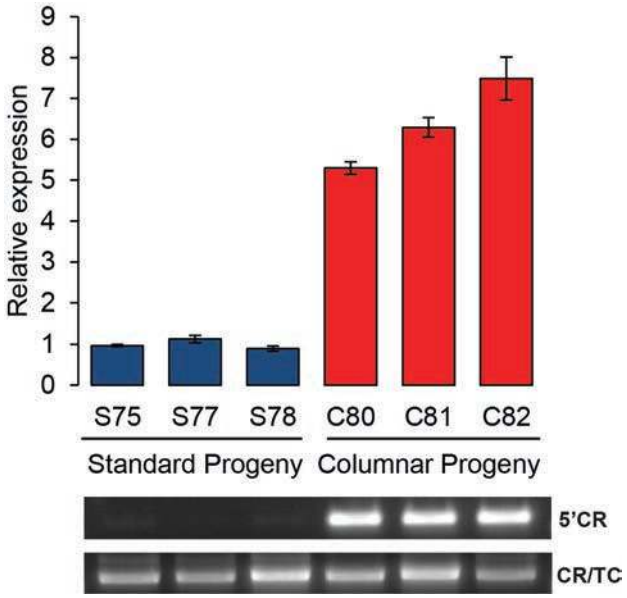

Figure 3. *MdCo31* expression and insert amplification in standard and columnar progeny trees. A. Expression of *MdCo31* in buds of standard (blue bars) and columnar trees (red bars). Expression levels are calculated based on three biological replicates and the standard deviations are indicated by error bars. **B.** Specific amplification of insert sequence produced a PCR product for columnar progeny trees only. The primers used for amplification were 29f1 and JWI1r.

A search for domains contained in the *MdCo31* protein revealed two significant matches: a non-haem dioxygenase in morphine synthesis N-terminal (DIOX-N) motif and a 2-oxoglutarate and Fe(II)-dependent oxygenase (2OG-Fe(II) oxygenase) domain (**Fig. 2a**). Members of the gene family containing these domains catalyse the oxidation of organic substrates and were shown to be involved in the biosynthesis of ethylene, flavonoids, gibberellins, and defence against downy mildew (Prescott and John 1996; van Damme et al. 2008). Interestingly, a poplar tree with a columnar-like habit called *stumpy* was identified during an activation tagging screen that caused a constitutive expression of a GA2-oxidase involved in gibberellin catabolism, resulting in reduced levels of gibberellin (Busov et al. 2003). Alternatively, flavonoids have been shown to inhibit auxin transport, a mechanism that is thought to be required for axillary bud outgrowth and involved in the high apical dominance typical of columnar trees (Brown et al. 2001; Prusinkiewicz et al. 2009; Petersen and Krost 2013). Some insights towards the possible role of gibberellin in conferring a columnar-like phenotype were already

proposed by two recent transcriptome studies, however, none of the described candidate genes corresponds to *MdCo31* (Zhang et al. 2012; Krost et al. 2013).

Although a role of *MdCo31* in gibberellin metabolism or flavonoid biosynthesis could possibly explain the phenotype of 'Wijcik', it does not provide in itself a causal link between the overexpression of *MdCo31* and the columnar habit. Transformation of apple, and fruit trees in general, requires a long regeneration time. Moreover, phenotypes related to tree architecture are visible only after several years. Therefore, we chose to use Arabidopsis in order to characterise the role of *MdCo31* in plant development. The coding sequence of *MdCo31* was expressed in *Arabidopsis thaliana* in a constitutive manner. 34 hygromycin-resistant transformants were obtained and three independent T3 segregating lines named Pro35S:MdCo31#1 to #3 were selected for further analysis (**Fig. S3**). All three lines expressed *MdCo31* RNA at high levels and showed a rather normal leaf phenotype but later, at maturity, displayed short inflorescence stems (**Fig. 4a to 4c**). In all three lines, the observed marked reduction in overall height was primarily due to a ca. three-fold shortening ($P<0.01$) in floral internodes, a phenotype reminiscent of the one observed in columnar plants (**Fig. 4d**). Overall, our results strongly support the hypothesis that *MdCo31* is responsible for the columnar habit.

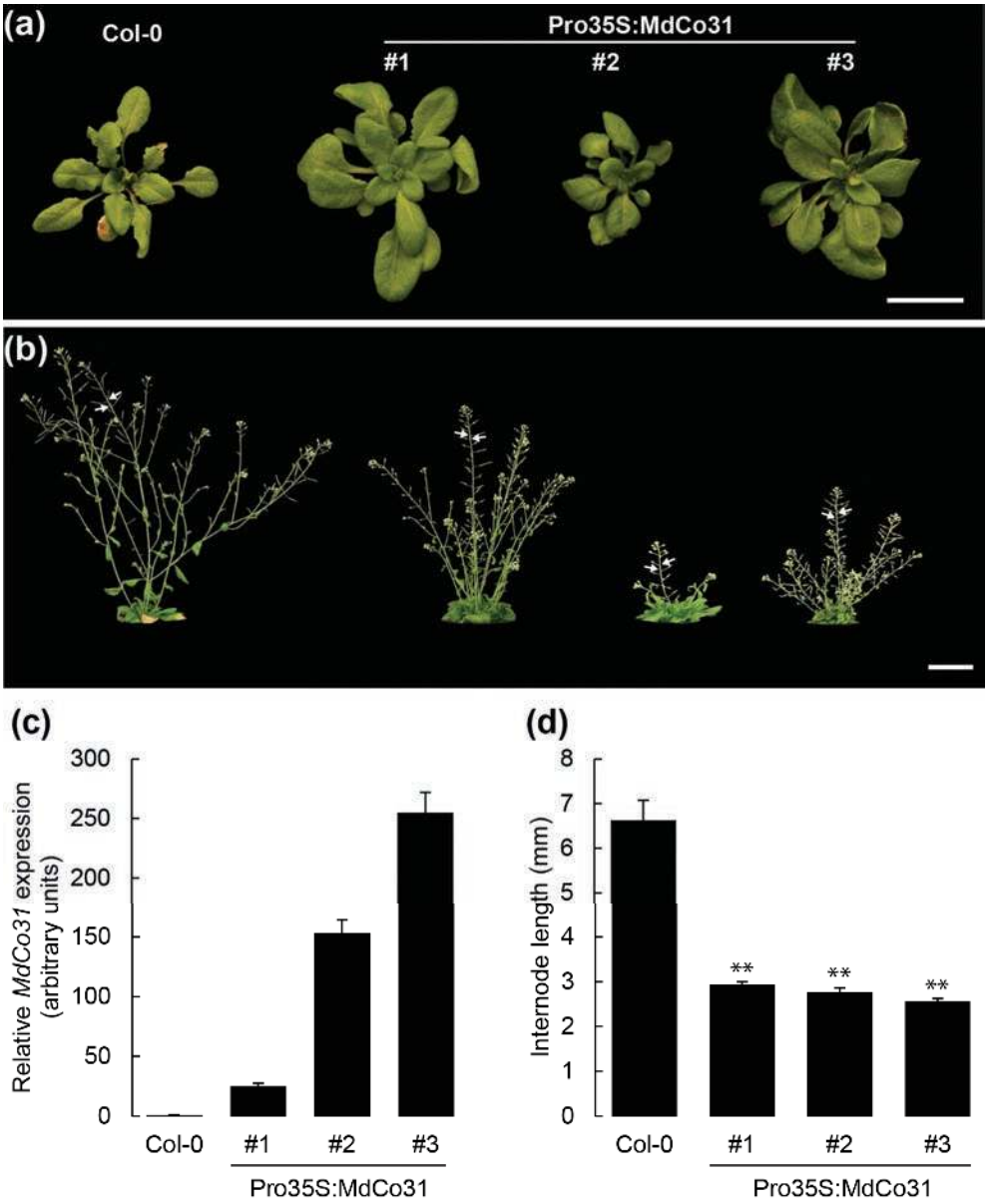

Figure 4. Phenotypes and *MdCo31* expression in *Arabidopsis* plants overexpressing *MdCo31*. A. Top view of rosettes of 4-week old *Arabidopsis* Col-0 and three independent transformants overexpressing *MdCo31*. The white bar corresponds to 2.5 cm. **B.** Side view of the same *Arabidopsis* plants 3 weeks later. The white bar corresponds to 5 cm. **C.** Expression level of *MdCo31* in the three *Arabidopsis* lines overexpressing *MdCo31*, compared to Col-0. Error bars indicate the standard deviation. **D.** Internode length of the inflorescences. The mean internode distance was calculated for 10 inflorescences per *Arabidopsis* line. Error bars indicate the standard error of the mean. Asterisks indicate a significant difference from Col-0 (Student's T test, $P<0.05$).

In summary, our work describes the identification of a 1956 bp insertion in the *Co* region of 'Wijcik' that correlates with the overexpression of *MdCo31*, coding for a putative 2OG-Fe(II) oxygenase. The overexpression of *MdCo31* in 'Wijcik' buds provides a plausible explanation for the dominant gain-of-function of the *Co* gene. However, how the insertion influences the expression level of *MdCo31* remains unclear and needs to be investigated in the future. Nevertheless, a further line of evidence supporting the involvement of *MdCo31* in controlling plant internode length was provided by transformation experiments using Arabidopsis. Taken together, the results presented in this paper let us conclude that *MdCo31* is a very reliable candidate gene for controlling the columnar habit in apple.

Acknowledgements

We thank Pierluigi Magnago for providing and maintaining the 'McIntosh' and 'Wijcik' trees as well as the segregating population. We also thank the FEM sequencing platform at CRI for sequencing the BAC clones. We thank the Autonomous Province of Trento (PAT) for financial support to A.S.A (GBPF GF) and the GMPF PhD school for funding the scholarship of P.J.W. We also thank the Technological Top Institute Green Genetics in collaboration with Inova Fruit BV for funding H.J.S. We thank Daniel James Sargent, and Nada Šurbanovski for their critical comments. All authors participated to the design of the study, P.J.W performed the experiments, A.S.A., P.B., and P.J.W wrote the paper.

References

Bai, T., Y. Zhu, et al. (2012). "Fine genetic mapping of the Co locus controlling columnar growth habit in apple." Molecular Genetics and Genomics **287**(5): 437-450.

Baldi, P., P. J. Wolters, et al. (2013). "Genetic and physical characterisation of the locus controlling columnar habit in apple (Malus × domestica Borkh.)." Molecular Breeding **31**(2): 429-440.

Barritt, B. H. (1992). Intensive orchard management. Washington USA, Good Fruit Growers.

Bennetzen, J. L. (2000). "Transposable element contributions to plant gene and genome evolution." Plant Molecular Biology **42**(1): 251-269.

Botton, A., G. Eccher, et al. (2011). "Signaling pathways mediating the induction of apple fruitlet abscission." Plant Physiology **155**(1): 185-208.

Brown, D. E., A. M. Rashotte, et al. (2001). "Flavonoids act as negative regulators of auxin transport in vivo in arabidopsis." Plant Physiology **126**(2): 524-535.

Busov, V. B., R. Meilan, et al. (2003). "Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature." Plant Physiology **132**(3): 1283-1291.

Chevreux, B., T. Wetter, et al. (1999). Genome sequence assembly using trace signals and additional sequence information. German Conference on Bioinformatics Hannover**:** 45-56.

Clough, S. J. and A. F. Bent (1998). "Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana." Plant J **16**(6): 735-743.

Conner, P. J., S. K. Brown, et al. (1997). "Randomly amplified polymorphic DNA-based genetic linkage map of three apple cultivars." Journal of American Society for Horticultural Science **122**(3): 350-359.

Davenport, T. L. (2000). "Processes influencing floral initiation and bloom: the role of phytohormones in a conceptual flowering model." HortTechnology **10**(4): 733-739.

Fisher, D. V. (1969). "Spur-type strains of McIntosh for high density planting." British Columbia Fruit Grower's Association Quart. Rep. **14**: 3-10.

Karimi, M., D. Inze, et al. (2002). "GATEWAY vectors for Agrobacterium-mediated plant transformation." TRENDS in Plant Science **7**(5): 193-195.

Kenis, K. and J. Keulemans (2007). "Study of tree architecture of apple (*Malus* x *domestica* Borkh.) by QTL analysis of growth traits." Molecular Breeding **19**(3): 193-208.

Krost, C., R. Petersen, et al. (2013). "Evaluation of the hormonal state of columnar apple trees (Malus x domestica) based on high throughput gene expression studies." Plant Mol Biol **81**(3): 211-220.

Kutter, C., H. Schöb, et al. (2007). "MicroRNA-mediated regulation of stomatal development in Arabidopsis." Plant Cell **19**(8): 2417-2429.

Lespinasse, J. M. and J. F. Delort (1993). "Regulation of fruiting in apple role of the course and crowned brindles." Acta Horticultura **349**: 239-245.

Li, J. and R. Yuan (2008). "NAA and ethylene regulate expression of genes related to ethylene biosynthesis, perception, and cell wall degradation during fruit

abscission and ripening in 'Delicious' apples." Journal of Plant Growth Regulation **27**(3): 283-295.

Looney, N. E. and W. D. Lane (1984). "Spur-type growth mutants of McIntosh apple: A review of their genetics, physiology and field performance." Acta Horticultura **146**: 31-46.

Majoros, W. H., M. Pertea, et al. (2004). "TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders." Bioinformatics **20**(16): 2878-2879.

Marchler-Bauer, A., S. Lu, et al. (2011). "CDD: a Conserved Domain Database for the functional annotation of proteins." Nucleic Acids Research **39**(Database Issue): D225-229.

Moriya, S., K. Okada, et al. (2012). "Fine mapping of Co, a gene controlling columnar growth habit located on apple (Malus×domestica Borkh.) linkage group 10." Plant Breeding **131**(5): 437-450.

Mudunuri, S. B. and H. A. Nagarajaram (2007). "IMEx: Imperfect Microsatellite Extractor." Bioinformatics **23**(10): 1181-1187

Ning, Z., A. J. Cox, et al. (2001). "SSAHA: a fast search method for large DNA databases." Genome Research **11**(10): 1725-1729.

Petersen, R. and C. Krost (2013). "Tracing a key player in the regulation of plant architecture: The columnar growth habit of apple trees (Malus x domestica)." Planta **238**(1): 1-22.

Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Research **29**(9): 2002-2007.

Prescott, A. G. and P. John (1996). "DIOXYGENASES: Molecular structure and role in plant metabolism." Annual Review of Plant Physiology and Plant Molecular Biology **47**: 245-271.

Prusinkiewicz, P., S. Crawford, et al. (2009). "Control of bud activation by an auxin transport switch." Proceedings of the National Academy of Sciences **106**(41): 17431-17436.

Rhee, S. Y., W. Beavis, et al. (2003). "The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community." Nucleic Acids Research **31**(1): 224-228.

Robinson, J. T., H. Thorvaldsdottir, et al. (2011). "Integrative genomics viewer." Nature Biotechnology **29**(1): 24-26.

Rom, C. R. and B. Barritt (1990). "Spur development of `Delicious' apple as influenced by position, wood age, strain, and pruning." HortScience **25**(12): 1578-1581.

Tamura, K., D. Peterson, et al. (2011). "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods." Molecular biology and evolution **28**(10): 2731-2739.

Tobutt, K. R. (1985). "Breeding columnar apples at East Malling." Acta Horticultura **159**: 63-68.

van Damme, M., R. P. Huibers, et al. (2008). "Arabidopsis DMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew." Plant J **54**(5): 785-793.

Velasco, R., A. Zharkikh, et al. (2010). "The genome of the domesticated apple (Malus x domestica Borkh.)." Nature Genetics **42**(10): 833-839.

Wu, J., Z. Wang, et al. (2013). "The genome of the pear (Pyrus bretschneideri Rehd.)." Genome Res **23**(2): 396-408.

Wünsche, J. N. and A. N. Lakso (2000). "The relationship between leaf area and light interception by spur and extension shoot leaves and apple orchard productivity." HortScience **35**(7): 1202-1206.

Zhang, Y., J. Zhu, et al. (2012). "Characterization of transcriptional differences between columnar and standard apple trees using RNA-Seq." Plant Molecular Biology Reporter **30**(4): 957-965.

Supporting material

Table S1. Sequences of primers used in this study

Primer name	Primer sequence (5'-3')
Primers used to complete the 'McIntosh' and 'Wijcik' BAC assembly	
JseqSTMC1f	CCGACCACAGCTACGAAAAT
JseqSTMC2r	AAAAACTAAAAACGTTTAACCAAACA
JseqSTMC3f	AAAGAACATGGAAGTTGATAATTGG
JseqSTMC5f	AAAGGCACTGCATAGGAAGAA
JseqSTMC6f	CACACTTAAGGGGGAATGTTG
JseqSTMC6r	GGGGACACTCTCTTGATCGT
JseqSTMC7r	CAGAAGGATTAATGTCACAACAATTT
JseqWiC1f	TCATTCTTTTTGCTACTTCGTTGA
JseqWiC2f	AGCTCGGGGATCATATAGGC
JseqWiC2r	GTCCCTTAACTTTGTACACCTCA
JseqWiC3f	AAACCTTGTAATTTCATGCATTTTT
JseqWiC4f	GTACCTGGAGCTTTTCGTTAAA
JseqWiC5f	TGCCCAAAATACCCCTATTG
JseqWiC6r	TCTTCTTGTTATGGGTAAATTATTCTC
JseqWiC7f2	TGGTTCATCCAAACCTTCAA
JseqWiC8r2	CTTCCAAGTCTTCGGTCGAG
JseqWiC9r	TAGGCCAAACTGCGTCCT
JseqWiC10f	AAAACCATTCCTTCAAAAGTGAT
JseqWiC10r	GTTACTCAGGGGTGGCTCTG
JseqWiC12f	CAATTTGCTCTAGCCCATGT
JseqWiC14f	AGCCCAAAACTCCTCATTCA
JseqWiC15f	CGCCGTAAGAGGGAATAGAC
JseqWiC16f	GGAAAACAGGATGGCCACTA
JseqWiC17f	CATGATTCACAATATGGTTTTTGA
JseqWiC18f	GCAAATAATTTACGTACACTTTGAAT
JseqWiC19f2	CGGAGATGGTGTGCCTATTC
JseqWiC19r	TTACTCCAGCGGCAACTTCT
JseqWiC20f	CTCGTCTCCAATCAAATTGTGT
JseqWiC20r	GCATTTCCATATGCCTAACCA
JseqWiC22f	TCAAATTATTTTGCCTAATGAGAA
JseqWiC22r	AAATTTAGTTCAAACATTTCGCAGT
JseqWiC23f	GGCAAGATAGACTTTATGAACCAAA

JseqWiC24f	AGACACCTCATGTGGGAAAAA
JseqWiC24r	CCTGCATAACCTTTCTTCATCA
JseqWiC25f	GACGAAAATTGACTGCACGA
JseqWiC25f2	GGATTACGCAATGTAATCATAGTTCT
JseqWiC26r	CTTCACTCACGCACGTCCTA
JseqWiC27f	CATCTTTAGGGATGTGTTTTACG
JseqWiC28f	TCAACGTTGCAAAATCTATCAAA
JseqWiC28r	GCATCCTACGTTCACCTTCG
JseqWiC29f	CCTGGATCTAAAAGCCCCATA
JseqWiC29f2	GCTAGAGGGTTCTTCTCCACA
JseqWiC30f	AAAAACCACCCTCAAATTTCC
JseqWiC30r	GCTTTGAGAAGGGGTGAGAA
JseqWiC31f3	CAACAATGTGCCTCGACACT
JseqWiC32f	CAAGATTCGAAGAACGATGC
JseqWiC32r2	CATCACCTTTCATCACATTTCA
JseqWiC33f	TTTTTCGTCTCAGAGTCATACCT
JseqWiC33f2	TGGAAGACATTAAATAAAAGGGTAAAA
JseqWiC33r	CATAAATTTTGAAGGTCTAGCTACCA
JseqWiC34f	ACTGCATATTTGCCACGTCA
JseqWiC34r3	TTCCCTCGGCTTTTTATTGA
JseqWiC35f	TTGGTTTGAACATCCCCTTT
JseqWiGAPC1f	TGGTCTTTTCCTTGGAGGAGT
JseqWiGAPC1r	TGACAGCCTTGTTTAAACGGTA
JseqWiGAPC2f	AGTGGCTGTGCTCAGTGATG
JseqWiGAPC2r	GCATTGGAGGAGGAGGGTAA

SSR markers used for screening the BAC libraries

		SSR size (bp) columnar/standard	**Coordinate start on chromosome 10**[d]
Co04R09[a]	TAGTGACATATACATGGTGCG GTTGGAGAATGAGTGACGGC	**156**/189	18406760
Co04R10[a]	ACCTGGTTCCGGTACATAGC AACCTTCCATGGCAGCAATC	**185**/203	18522961
Co04R11[a]	ACATCATGGTATGACAGAGGTG TCTAAGCCTGTCAAGATGGC	**184**/172	18562094
J8Co	TCGTTGCTCAATTACTGGCTA AGGAATGCACAAATCCTCCA	**245**/232	18637186
J9Co[b]	TGTTGAGGGATTTTTCAGACG GCACTCCCACCCCACTACTA	**203**/201	18665642
J11Co[b]	ACCCGAGACCTACCGCTTAT TTTGAATGTGGTCCCTAGCC	**151**/156	18757238

Co04R12[a]	TTTATCTGACTAAGGGGAAGG ATGGACTTGTATTCCTTAGGG	**195**/223	18774919
J12Co[b]	GCACCCCCAGATTTCAACT GCTACTGCTGATGGCCTCTC	**218**/216	18797594
J14Co[b]	CTGAACATAACACCCTATACTTTGAA GACTTCAAACCTGAAGGGATAAAA	**192**/139	18883408
Co04R13[ac]	ATTTTCCCTCTCTTCTGTTGC TCTTGGAAAGACGTGGCACG	**237**/235	18905432

Primers used to sequence the part of the 'McIntosh' *Co* region that was not covered by the BAC library

JseqSTMG1f	AACTGTCAGGAATCGGAATAGG
JseqSTMG1r	TCAACATCATTCTTGCCATCTT
JASSTMG2f	TGTAAACGTGCTTTCAGCAAA
JASSTMg1r	AGAAGAGTTCCGGTGAGCAA
JseqSTMG2f	CACGTGTCAATGTCAGAATGAA
JseqSTMG2r	CGTCCACAGAGTAGAGTTCTCA
JASSTMg3f	TCAGGAAACAATGGGGTTCT
JASSTMg2r	TGGTGAAAATGATGGGGTTT
JseqSTMg3f	TTGGCCCAAAATCCAAATAA
JseqSTMg3r	CTTCAAGTTTTCATGGGTGATG
JASSTMg4f	TCCATTGGAGGAAACAAAGC
JASSTMg3r	TTGTGGTCTCCAGTTCAACG
JseqSTMg4f	TGCATCGAATCCTCAAATCA
JseqSTMg4r	ATGGTCCGTATGGAAGGTGA
JASSTMg5f	GGCGTCAATATTGGTGATCC
JASSTMg4r	AAACTAGCATAATAGATCGAAATCTCA
JseqSTMg5f	TGCTCTCAATTCGACTGTGG
JseqSTMg5r	GTTCTCGCGATCTCTTCAGG
JASSTMg6f	ACGCATCATGCATCAAACAG
JASSTMg5r	CCAATATGGGTGCCTCGTTA
JseqSTMg6f	GACTTGCTGCTTCGTGCATA
JseqSTMg6f2	ACGTCCACGGAATAGAGGAG
JseqSTMg6r	TTTTAAACCCAAGACCTCCTCA
JASSTMg7f	GACTCAACAACAACAACAACAACA
JASSTMg6r	GGAAGTCAAGCCGATACCAC
JseqSTMg7f	AGGCCGACTGAAGAATTTGA
JseqSTMg7r	CTTACCCTGGCATTGCTCTC
JASSTMg7r2	TCTAACAAAGTTGAGACTACTTTTCCA

Primers used for qRT-PCRs and PCRs on genomic DNA

MdCo27	CTCCTAGAGGCCCGAGATG
	TGCCGTGAGTTTCCTATTCC
MdCo28	ACTGTGGAGGTCGGAAACC
	AAGGCTCGTAAATGCCACAG
MdCo29	GGATGGGGAAGGGATTAGAG
	GGAGACTGAGGAAGGCACAT
MdCo30	CCGTCTGTGGAACGACACTT
	ATGCGTGTCATTTTGTGTGC
MdCo31	CATAAAATGCCCCGAAAAGA
	CAGAAGAATGAGCAGGGTGAG
MdCo32	CCCCGAAGAGACAAGGAAAC
	CCAGAATGACAGACCCCAAT
MdActin	TGACCGAATGAGCAAGGAAATTACT
	TACTCAGCTTTGGCAATCCACATC
Md_4592:1:a	GTCGAAATGGTCAGCGGTAG
	GCAATGGCAAACTCCACCTT
AtTIP41-like	GTGAAAACTGTTGGAGAGAAGCAA
	TCAACTGGATACCCTTTCGCA
AtActin-2	TCCCTCAGCACATTCCAGCAG
	AACGATTCCTGGACCTGCCTCATC

Primers used for 'Wijcik' insert amplification

		Coordinate on chromosome 10[d]
Cr29f1	CCTGGATCTAAAAGCCCCATA	18796670
JWI1r	AACCAAACACCCACCCATTA	
29f2	GCTAGAGGGTTCTTCTCCACA	
JWI3r	GTAGGAGAGTCCGGGGAAAG	
29f3	TGATCTACACAGACAGTTTTGACG	
30r	GCTTTGAGAAGGGGTGAGAA	

Primers used for cloning *MdCo31*

caccMdCo31f	CACCATGGAGACATTAGATCAG
MdCo31r	TTAGCTACTCGAGAGGCTTAAAACC

[a] SSR marker was used in Baldi et al (2012)
[b] Homologue of SSR marker was described by Moriya et al (2012)
[c] Homologue of SSR marker was described by Bai et al (2012)
[d] Coordinate start of the forward primer on chromosome 10 of the Golden Delicious genome (http://genomics.research.iasma.it/)

Table S2. Gene predictions from the *Co* region and their putative function

Prediction	Contig	Start Position	Description	Gene ID[a]
MdCo1	WiC1	4885	Chromosome segregation ATPase	***MDP0000143705***
MdCo2	WiC1	16370	translocase inner membrane subunit	***MDP0000143705***
MdCo3	WiC1	27606	predicted protein	*MDP0000254095*
MdCo4	WiC1	43209	myb domain protein	***MDP0000897594***
MdCo5	WiC1	60715	DNA mismatch repair protein	***MDP0000524262***
MdCo6	WiC1	72499	Protein of unknown function, DUF647	***MDP0000284965***
MdCo7	WiC1	76626	Protein of unknown function	*MDP0000090747*
MdCo8	WiC1	79902	actin depolymerizing factor	***MDP0000329966***
MdCo9	WiC1	83253	phosphate transporter	***MDP0000942873***
MdCo10	WiC1	92605	RNAseH-like protein	*MDP0000508375*
MdCo11	WiC1	95489	predicted protein	*MDP0000265238*
MdCo12	WiC1	106812	predicted protein	*MDP0000852896*
MdCo13	WiC1	109222	predicted protein	***MDP0000186457***
MdCo14	WiC1	115595	predicted protein	*MDP0000852890*
MdCo15	WiC1	122386	predicted protein	**n.f**
MdCo16	WiC1	131567	AP2 transcription factor	***MDP0000855671***
MdCo17	WiC1	139082	Retrotransposon protein	*MDP0000440636*
MdCo18	WiC1	151296	Retrotransposon protein	**n.f**
MdCo19	WiC1	181283	Retrotransposon protein	**n.f**
MdCo20	WiC1	191207	predicted protein	**n.f**
MdCo21	WiC1	192786	predicted protein	**n.f**
MdCo22	WiC1	200156	predicted protein	**n.f**
MdCo23	WiC1	204431	predicted protein	*MDP0000776816*
MdCo24	WiC1	223258	cysteine-rich RLK	**n.f**
MdCo25	WiC1	236340	AP2 transcription factor	***MDP0000286915***
MdCo26	WiC1	242114	AP2 transcription factor	***MDP0000187369***
MdCo27	WiC1	246782	predicted protein	**n.f**
MdCo28	WiC1	258318	Protein of unknown function	**n.f**
MdCo29	WiC1	281163	RING/U-box superfamily protein	***MDP0000766466***
MdCo30	WiC1	286736	predicted protein	**n.f**
MdCo31	WiC1	297570	2OG-Fe(II) oxygenase family protein	*MDP0000687812*
MdCo32	WiC1	301585	Retrotransposon protein	*MDP0000687812*
MdCo33	WiC1	307277	RING/U-box superfamily protein	***MDP0000927098***
MdCo34	WiC1	310294	Vacuolar protein sorting protein	*MDP0000927097*
MdCo35	WiC1	313533	Tetratricopeptide repeat-like superfamily protein	***MDP0000187760***

MdCo36	WiC1	323260	Autophagy protein 9	***MDP0000927091***
MdCo37	WiC1	335884	predicted protein	*MDP0000832682*
MdCo38	WiC1	339324	predicted protein	**n.f**
MdCo39	WiC1	344289	nodulin MtN21 /EamA-like transporter family protein	*MDP0000912170*
MdCo40	WiC1	348746	predicted protein	***MDP0000912172***
MdCo41	WiC2	3305	formyltetrahydrofolate synthetase	***MDP0000139773***
MdCo42	WiC2	8297	predicted protein	*MDP0000598647*
MdCo43	WiC2	27835	cysteine-rich RLK	**n.f**

[a] Gene ID numbers were retrieved from http://genomics.research.iasma.it (predictions from gene set in bold). n.f.: not found.

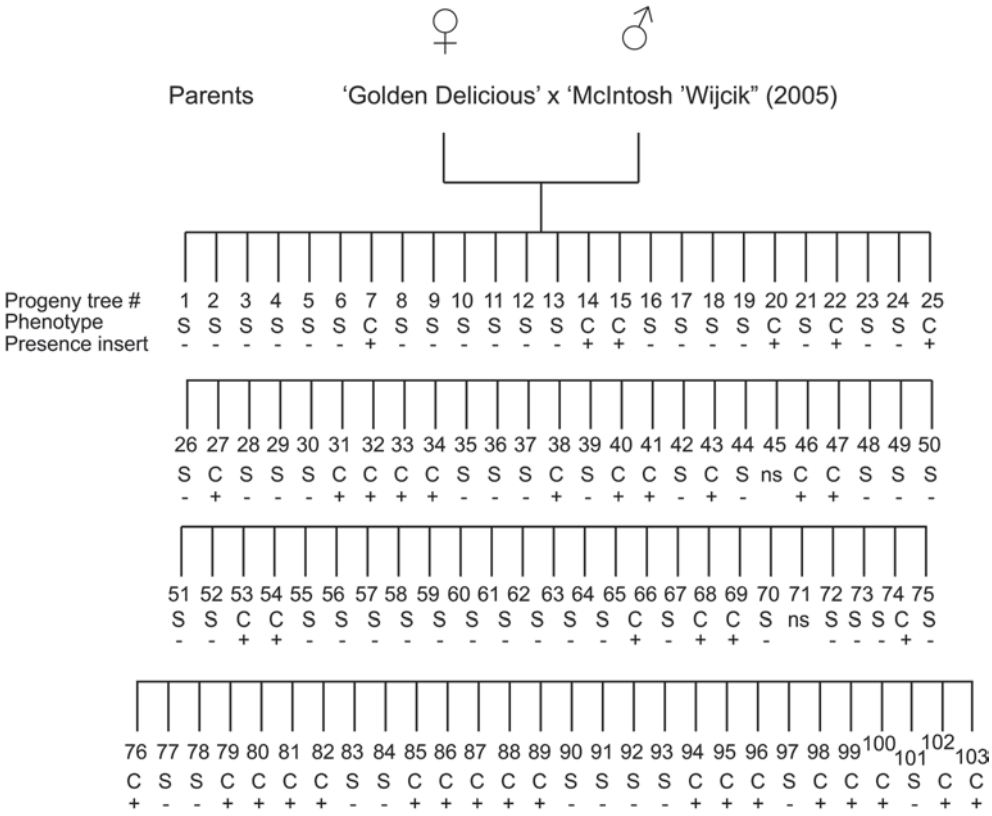

Figure S1. Overview of the progeny population derived from the cross between 'Golden Delicious' and 'Wijcik' that was performed in 2005. This population was used to test the Co-segregation of the newly developed SSR markers described in this paper. The phenotype (S: standard, C: columnar) and the presence of the insert is shown below each progeny tree. ns indicates a tree that has not been scored.

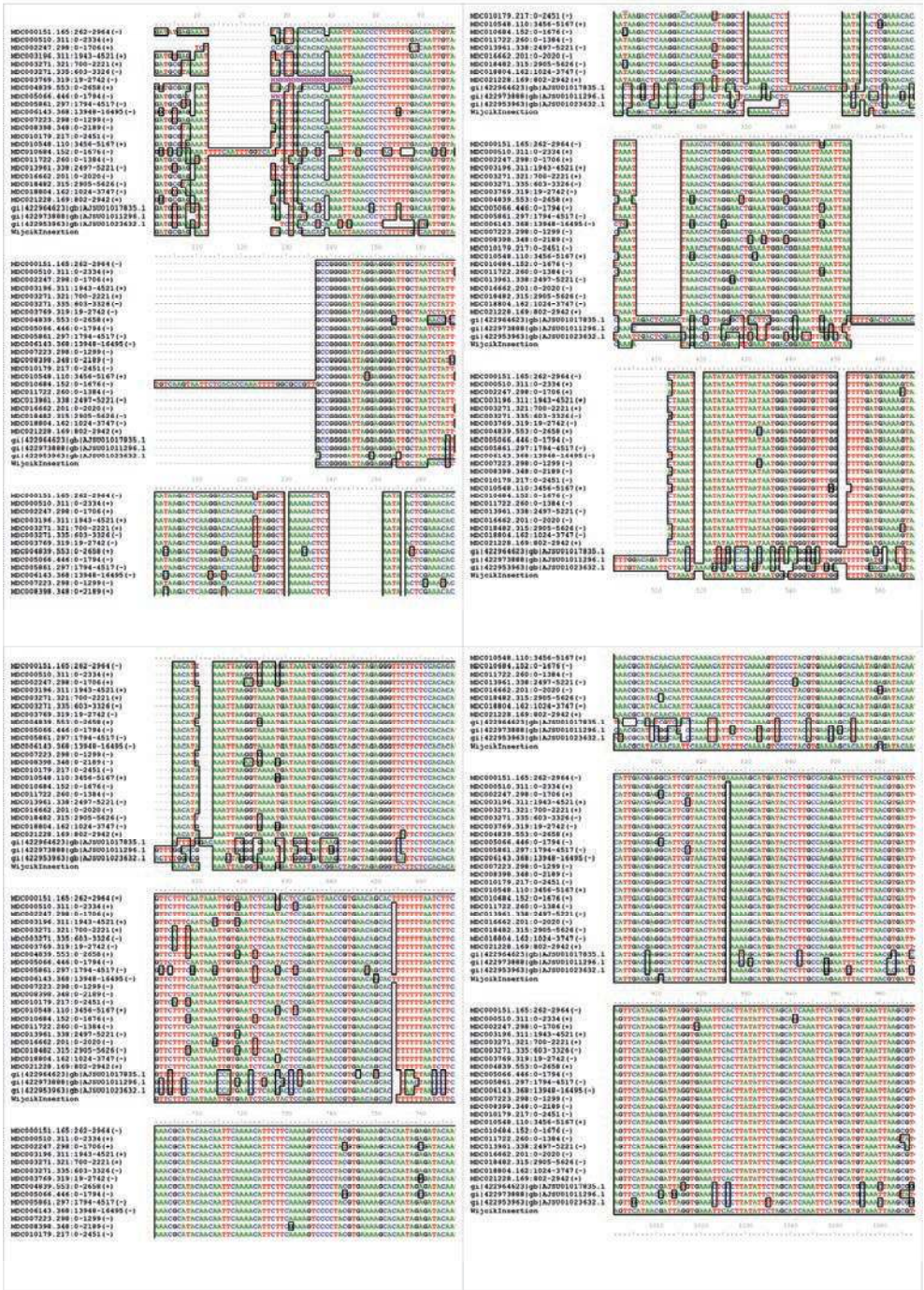

Figure S2. Alignment of contigs containing the 1956 bp insertion. This figure is continued on the next two pages

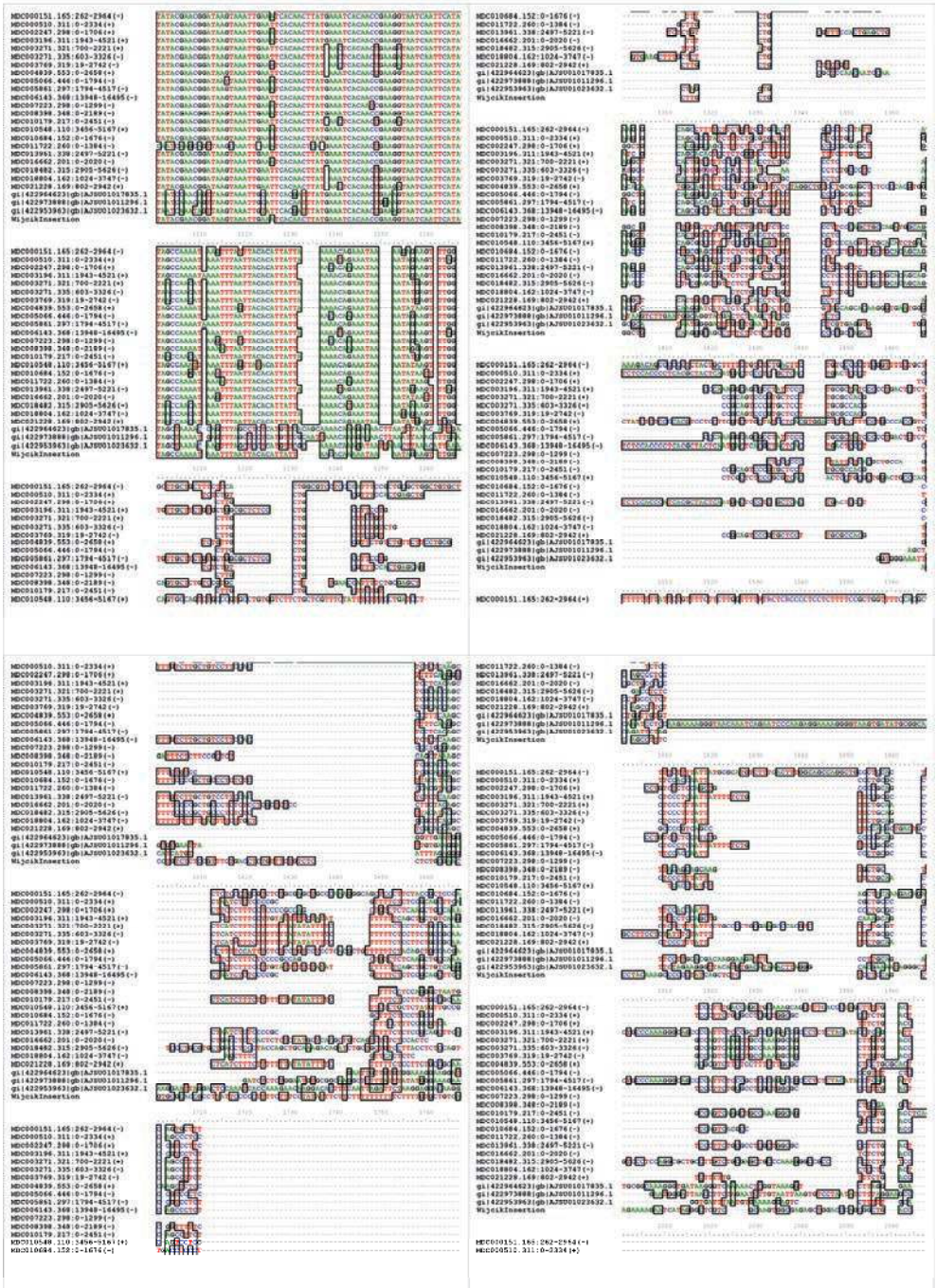

Figure S2. continued

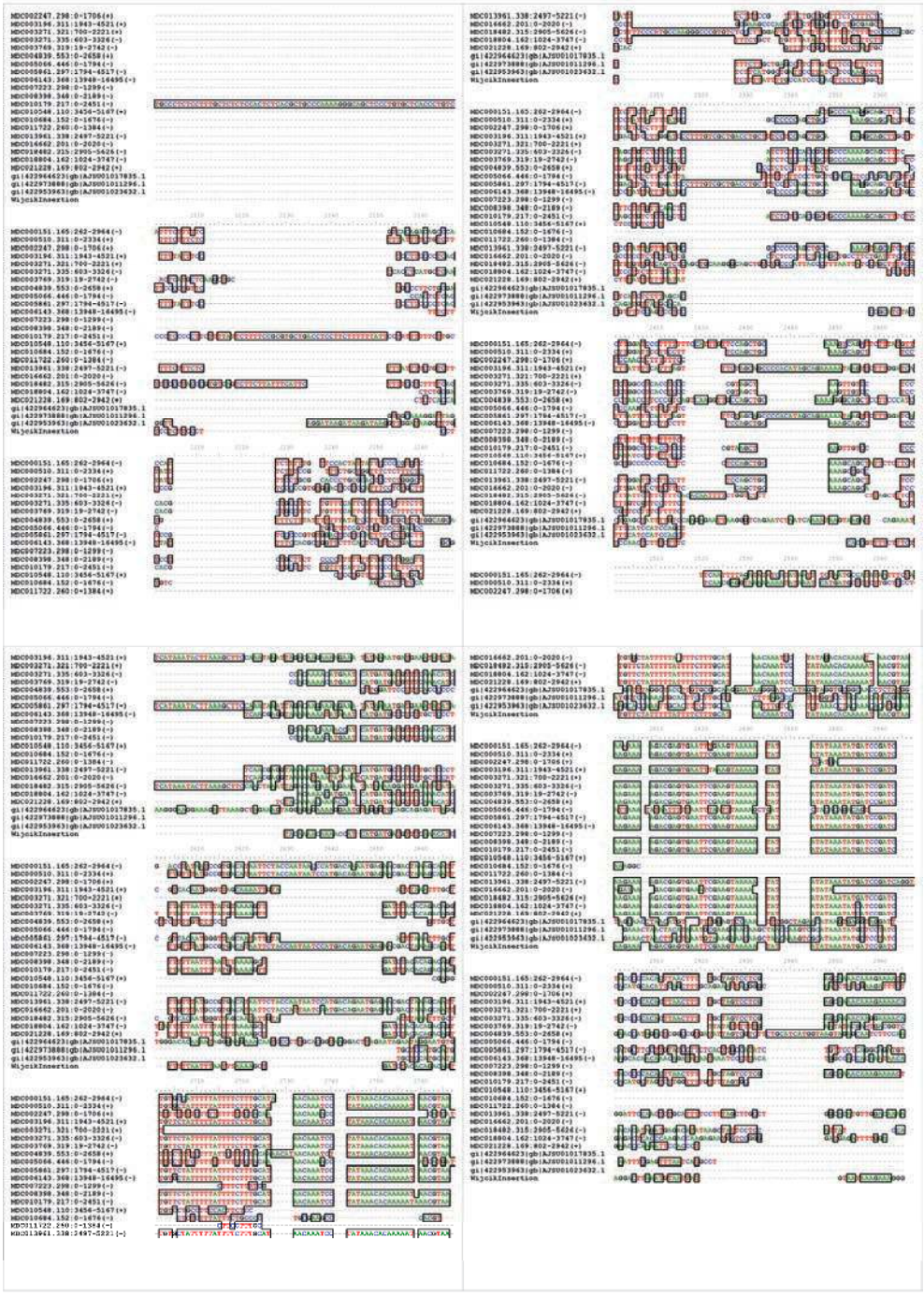

Figure S2. Alignment of contigs containing the 1956 bp insertion. An alignment of the genomic region of 'Wijcik' containing the insertion with the 13 closest matches from 'Golden Delicious' (starting with MDC...) and three matches from *Pyrus bretschneideri* Rehd.(gi...)

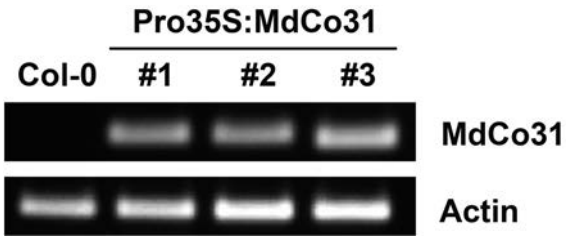

Figure S3. Genomic PCR on Col-0 and the three Pro35S:MdCo31 Arabidopsis lines. The primers described for qRT-PCR quantification of *MdCo31* were also used to detect the presence of *MdCo31* in the Arabidopsis plants by PCR. Actin-2 was included as a control using the primers indicated in **Table S1**.

Chapter 4

Genomic characterisation of the 'Wijcik' mutation

Pieter J. Wolters, Henk J. Schouten, Azeddine Si-Ammour, Paolo Baldi

Abstract

The dominantly inherited columnar trait of 'Wijcik' apple is caused by a somatic mutation in apple cv. 'McIntosh'. By comparing the genomic sequences of 'McIntosh' and 'Wijcik', an insertion was found in the *Co* region that was mapped in 'Wijcik'. This insertion was shown to be associated with increased expression of a gene, *MdCo31*, located more than 15 kb downstream. A discrepancy was found between the size of the insertion as described by Wolters et al. in 2013 and as described almost two months later by Otto et al. (2013). This discrepancy is solved in this chapter and the mutation that is involved in the columnar phenotype of 'Wijcik' is characterised in more detail. The 'Wijcik' insertion is classified as a non-autonomous retrotransposon of 8.2 kb, which inserted into the 5' LTR of another retrotransposon of 4.0 kb that was already present in the 'McIntosh' genome. Using the LTR sequences of the retrotransposons, copies of both elements are identified throughout the 'Golden Delicious' genome. Elements similar to the retrotransposon that was found in 'McIntosh' show a preference for inserting in gene rich regions, while the element that was found inserted in 'Wijcik' does not show such a preference. The 'Wijcik' mutation is another example of how transposons can affect specific properties of their host, but the mechanism by which the 'Wijcik' transposon leads to increased expression of *MdCo31* is still unknown. Addressing this question would be a very interesting topic for further studies.

Introduction

Although genome sizes vary greatly among different plant species (from the ~125-megabase genome of Arabidopsis to the largest sequenced plant genome so far, the 2.3-gigabase genome of maize) (The Arabidopsis Genome Initiative 2000; Schnable et al. 2009), the total gene number is rather constant among different plant species. Only a relatively small part of the Arabidopsis genome is formed by repetitive DNA, whereas repetitive DNA represents the major part of the maize genome (Meyers et al. 2001).

The non-coding part of the genome was long regarded as junk DNA, but recent research has showed that this non-coding part can have an important function. Although only 3% of the human genome consists of genes, an estimated 80% of that genome is believed to perform a functional role (ENCODE Project Consortium 2012).

Two major classes of repetitive elements exist. The first class comprises the elements that are generated by the host itself, through recombination or the replication machinery. These elements include terminal repeats and tandem repeats. The second class is formed by transposable elements (TEs), which can replicate independent of the host replication machinery (Jurka et al. 2005). The expansion of TEs is responsible for the main difference in size between plant genomes (Bennetzen et al. 2005).

Apple has an estimated complete genome size of 742.3 megabases (Mb) and consists for 67% of repetitive DNA. The major part of this repetitive DNA, approximately 42.4% of the apple genome, is formed by TEs (Velasco et al. 2010).

Two classes of TEs can be recognised: retrotransposons, which replicate through an RNA intermediate, and DNA transposons, which replicate through a DNA intermediate (Finnegan 1989). Wicker et al. proposed a classification system in which a distinction is made between retrotransposons and DNA transposons and two further subclasses of DNA transposons: those that copy themselves for insertion and those that leave the original insertion site to move to another place in the genome. These classes are further divided in different orders, based on the insertion mechanism of the elements (Wicker et al. 2007).

On a more detailed level, TEs are grouped into superfamilies, according to large-scale features such as the structure of protein or non-coding domains and size of the target site duplication (TSD). Superfamilies are subsequently divided into families, based on DNA sequence conservation which is generally high in the coding parts (the open reading frames, ORFs) of TEs, such as the integrase domain, reverse transcriptase, or capsid protein. Non-autonomous elements have lost some or all of the ORFs found in autonomous elements and depend on their autonomous partners for transposition. As the termini of the elements are required for transposition, these parts are generally well conserved (Wicker et al. 2007).

Movement of TEs has greatly affected the content of plant genomes, making up the majority of plant DNA, and have likely played an important role in plant genome evolution. Not only do TEs provide additional DNA to the genome, or do they disrupt existing genes; they are also able to affect the expression of neighbouring genes (Lisch 2013). They can mediate genome rearrangements and move (parts of) genes to new genomic contexts (Lisch 2013). TE activity is under epigenetic control and this control can affect host gene expression as well. An interesting hypothesis states that TEs were the original targets of epigenetic regulation and that this mechanism was later adapted to control the expression of host genes (Slotkin and Martienssen 2007). Some examples of these different effects of TE insertions on the host will be given below.

The insertion of a TE close to a gene can disrupt the promoter of the gene or affect repressor or enhancer elements. Such a mutation was found in maize, where the insertion of a Mutator (Mu) TE in a conserved sequence in the intron of 1 *knotted1* led to overexpression of that gene (Greene et al. 1994).

Naito et al. (2006) showed a big expansion of the DNA transposon mPing in some rice strains and, in a subsequent study, showed that the mPing element has a preference for inserting into the first 1-5 kilobases (kb) upstream of the transcription start site of genes. The mPing insertion affects the expression of those genes, making them stress-inducible (Naito et al. 2009). A preference for inserting in the 5' flanking region of genes has been described for other TEs as well, for example for the Mu and Mutator–like TEs (Pack-MULEs) (Liu et al. 2009; Jiang et al. 2011), suggesting that these elements also target the 5' region of genes and are able to affect the expression of these genes.

Capturing of gene fragments by TEs and the movement of these fragments to new places in the genome was reported for helitrons in maize and Mutator like elements (MULEs) in rice (Juretic et al. 2005; Morgante et al. 2005). In theory, this movement could lead to new combinations of ORFs. When the *SUN* gene got copied by a retrotransposon to a new genomic context in tomato, this resulted in increased expression of that gene and in an elongated fruit shape (Xiao et al. 2008).

Activity of TEs is under epigenetic control by post-transcriptional silencing and chromatin modifications, mediated through small interfering RNAs (siRNAs) (Slotkin and Martienssen 2007). A number of studies have demonstrated that this epigenetic control of TEs can alter the expression of genes close by. An example is the post-transcriptional silencing of FLOWERING LOCUS C (*FLC*) in the early flowering Landsberg erecta (Ler) accession of Arabidopsis, that is due to the insertion of a MULE in an intron of that gene (Liu et al. 2004). Spreading of methylation (to control TE activity) from TE insertions to the surrounding area is another mechanism that can lead to changes in expression of genes around TEs (Martienssen and Baron 1994; Iida et al. 2004).

The insertion of TEs can result in alterations in gene expression and possibly in novel coding sequences, but the most likely effect is deleterious. Although TE movement is normally prevented by the host, it is clear from the many TEs present in plant genomes that they do manage to escape this control occasionally. Several studies showed that the control of TE movement is released under stress, a mechanism that is hypothesised to be employed by the host to generate diversity to adapt to new conditions (McClintock 1984; Levin and Moran 2011; Shapiro 2011).

The 'Wijcik' mutation was already briefly described in chapter 3 of this thesis. In that chapter, we performed a sequence comparison between the 'Wijcik' and the 'McIntosh' *Co* region and identified a single difference between the two cultivars: a 1956 bp insertion in 'Wijcik'. We also showed that many copies of similar sequences are present in the 'Golden Delicious' genome.

Otto et al. (2013) compared the *Co* region between 'McIntosh' and 'Wijcik' as well and described an insertion in 'McIntosh' at the exact same position as the insertion that we identified in our study (Wolters et al., 2013). They, too, concluded that this insertion in 'Wijcik' is the only genomic difference between the 'McIntosh' and 'Wijcik' *Co* region. However, the insertion described by Otto et al. is 8.2 kb and not 1956 bp, as was reported by us. From both studies, though, it is clear that a TE-like sequence is responsible for the columnar growth of 'Wijcik'.

The precise conditions that led to the TE insertion in 'Wijcik' are unknown, but the mutation was reported to occur spontaneously. 'Wijcik' originated as a

single upright shoot at the top of a 50-year-old 'McIntosh' tree (Fisher 1969). About 50% of the progeny trees derived from a cross from the 'Wijcik' mutant with 'Golden Delicious' displayed a similar growth habit, indicating its monogenic, dominant character (Fisher 1969). The past 50 years have shown that columnar growth is a very stable trait. Grafts from the original 'Wijcik' shoot that are maintained at the germplasm collection of Fondazione Edmund Mach, and at other places, still show the same phenotype that was originally described by Fisher (1969). This stability already suggested that the *Co* mutation is a genetic and not only an epigenetic mutation, although the genetic mutation may have induced epigenetic changes.

The presence of the 'Wijcik' insertion was correlated with an increase in transcription of *MdCo31*, a gene 15.6 kb downstream of the insertion. The increased transcription of *MdCo31* in columnar trees provides a good explanation for the dominant character of the columnar trait. No increased transcription was found for the predicted genes between the insertion and *MdCo31* (Wolters et al. 2013). We expected to find an effect on a gene close to the mutation, but, as also shown in the ENCODE project, regulatory elements are not necessarily close to the affected gene (Sanyal et al. 2012). An interesting example, with some parallels to the 'Wijcik' insertion, is the retrotransposon *Hopscotch* in maize that inserted around 60 kb upstream of the teosinte branched1 gene. The retrotransposon insertion caused increased expression of that gene, resulting in increased apical dominance and a reduction in branching (Studer et al. 2011).

In this chapter, we compared the results of the two studies performed by Wolters et al. (2013) and Otto et al. (2013), which both show that the 'Wijcik' *Co* region contains an insertion that is not present in the *Co* region of 'McIntosh'. We characterised this insertion in more detail and looked at homologous sequences from 'Golden Delicious'. We also had a closer look at the region surrounding the 'Wijcik' insertion and showed that this region contains another retrotransposon that has homologs in the 'Golden Delicious' genome. In addition to characterising the 'Wijcik' mutation, possible explanations for the effect of the 'Wijcik' insertion on the expression of *MdCo31* are discussed, together with experiments that could be performed to test these hypotheses.

Materials and methods

PCR

PCRs were performed on an Applied Biosystems 2720 Thermal Cycler using 5 PRIME Taq DNA Polymerase in a total volume of 20 µl. For amplification of 'McIntosh' DNA, BAC clone STM80E16 was used, and for amplification of 'Wijcik' DNA, BAC clone W11K19 was used (both are described in chapter 3 of this thesis). Cycling conditions included an initial denaturation step of 5 min at 95 °C followed by 32 cycles of 30 s at 95 °C, 1 min at 60 °C and 2 min at 72 °C, terminating in a final extension step of 5 min at 72 °C. Primers PPTr (5'-GAGGGTCACATAGGGGGAAG-3') and PBSf (5'-TCCCTCAGCTCCATTTGAAC-3') were designed on the PPT and PBS of the 8.2. kb retrotransposon described by Otto et al. (2013).

Analysis of homologous retrotransposons from 'Golden Delicious' and gene density

The consensus long terminal repeat (LTR) sequences of the two nested retrotransposons from 'Wijcik' were blasted in the genome of 'Golden Delicious' (Velasco et al. 2010). The locations of the LTR matches were stored as tabular blast output files and a python script was developed to search for LTR matches on the same contig, within 10 kb from each other. Sequences containing 2 LTRs and varying in length between 3 and 10 kb were considered putative retrotransposons and were extracted from the 'Golden Delicious' genome using the BEDtools suite (Quinlan and Hall 2010).

A similar approach was used to look at the gene density around the complete retrotransposons that were identified in the previous step, the LTRs of the 'Wijcik' insertion and 500 randomly picked locations on the 'Golden Delicious' genome. Gene locations were taken from the GFF-file published with the 'Golden Delicious' genome (Velasco et al. 2010).

Results

The 'Wijcik' insertion

In the previous chapter, we stated that the 'Wijcik' insertion has a length of 1956 bp, whereas Otto et al. (2013) described an 8.2 kb retrotransposon insertion.

In Fig. 1a, both results are compared next to each other. Strikingly, the size of the LTRs of the retrotransposon that was described by Otto et al. (1951 bp) is very similar to the size of the 1956 insertion that was described by us. When comparing both sequences, it turns out that the 1951 bp LTR that was described by Otto et al. is completely contained within the sequence of the 1956 insertion. The remaining 5 bp (AGGAC) preceding the insertion matches the last 5 bp of the 1956 bp insertion and is identical to the TSD described by Otto et al. This observation led us to question our previous findings.

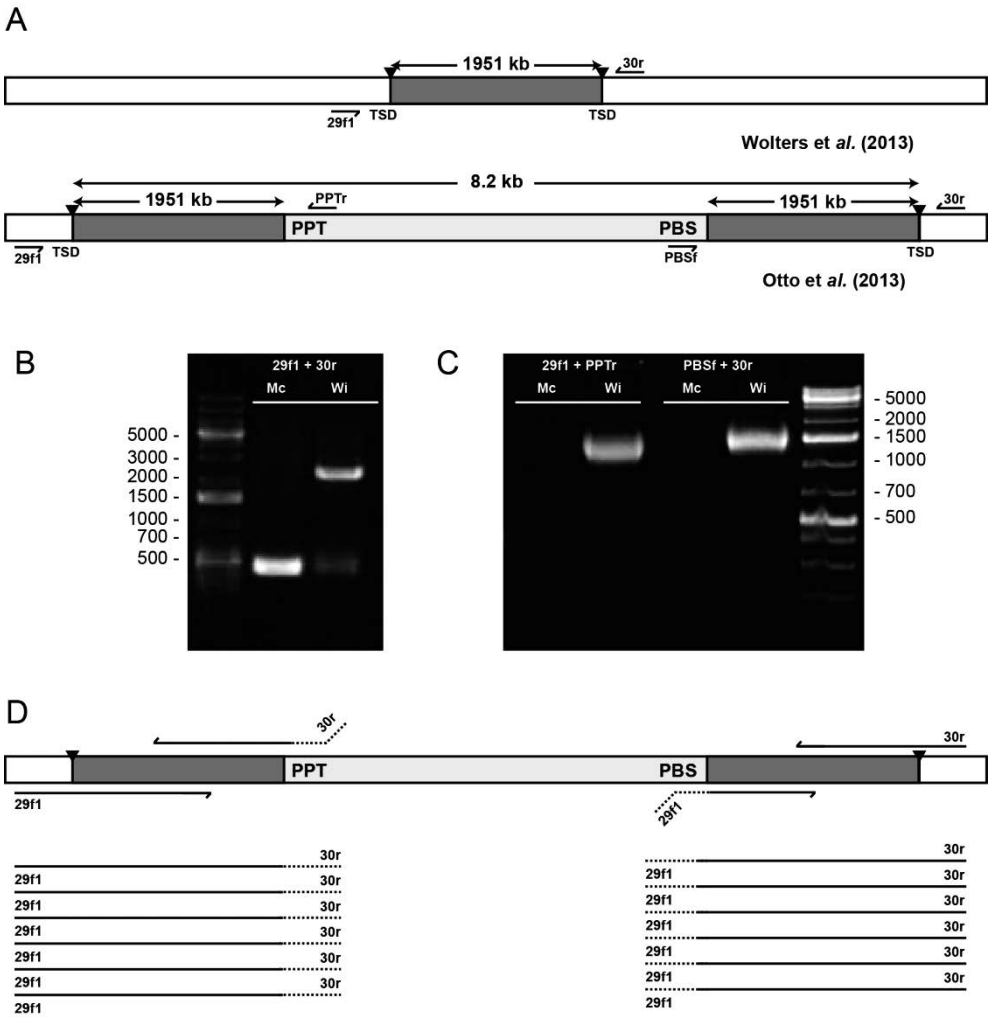

Figure 1. Investigating the correct size of the 'Wijcik' insertion. **A.** Comparison of the 'Wijcik' insertion described by Wolters et al. (2013) (top) and the insertion described by Otto et al. (2013) (bottom). The size of the LTR of the retrotransposon that was described by Otto et al. is identical to the size of the insertion described by Wolters et al. (1951 bp, not counting the TSD, target site duplication). The location of the polypurine tract (PPT) and the primer binding site (PBS) are indicated on the retrotransposon described by Otto et al. The locations of the primers used in the PCR that is displayed in (B) and (C) are indicated on the two different depictions of the 'Wijcik' insertion. **B.** A PCR using primers 29f1 and 30r, designed on DNA flanking the insertion, resulted in a PCR product of around 400 bp when genomic DNA of 'McIntosh' (Mc) was used as a template and around 2300 bp when genomic DNA of 'Wijcik' (Wi) was used. This result suggests that the 'Wijcik' insertion is around 2 kb and not 8.2 kb. **C.** PCR products were obtained for Wi template when primers were used that should specifically amplify the insertion that was described by Otto et al. and that should not result in amplification products for the insertion described by Wolters et al. As expected, no PCR products were obtained for Mc template. These results indicate that Wolters et al. made a mistake in assembling the insertion and that the insertion was correctly described by Otto et al. **D.** Explanation for the conflicting results of the PCRs shown in (B) and (C). PCR fragments that were derived from the 5' long terminal repeat (LTR) could re-anneal to the 3' LTR and vice versa. In this way, new DNA template was formed that contained both primer sites flanking the LTR only, instead of the complete transposon. In subsequent cycles, this shorter fragment would have amplified better than the original template containing the 8.2 kb sequence.

After finding the 'Wijcik' insertion, we performed several control PCRs on the BAC plasmids containing the *Co* region of 'McIntosh' and 'Wijcik' and on genomic DNA of 'McIntosh', 'Wijcik' and columnar and standard trees derived from the 'Golden Delicious' x 'Wijcik' cross. These PCRs showed a clear correlation between presence of the insert and the columnar phenotype (Wolters et al. 2013). Interestingly, a PCR using primers designed on both ends flanking the 'Wijcik' insertion resulted in PCR products of a bit more that 2 kb (not published), suggesting that the insertion was correctly described by Wolters et al. (2013) and a lot smaller than the 8.2 kb that was reported by Otto et al. (2013).

To rule out all doubts about the size of the insertion, the PCR using the primers designed on both flanking parts of the 'Wijcik' insertion was repeated on BAC clones containing the *Co* region (Fig. 1b). This PCR should amplify the complete insertion. Additionally, new primers (PPTr and PBSf) were designed on the polypurine tract (PPT) and the primer binding site (PBS) of the retrotransposon described by Otto et al, which should not be present in the 1956 bp insertion (Fig. 1c).

As observed before, the PCR using primers designed on the regions flanking the 'Wijcik' insertion resulted in PCR products that were slightly larger than 2 kb (Fig. 1b), suggesting that the 'Wijcik' insertion is around 2 kb. Surprisingly, the PCRs using primers specific for the 8.2 kb retrotransposon also yielded the

products that would be expected if this element was present at the 'Wijcik' insert site (Fig. 1c). These results seem to conflict each other.

To better understand these contradicting results, we went back to the original 454 reads that were used to assemble the *Co* region of 'Wijcik'. Alignment of these reads to the 8.2 kb retrotransposon showed that this sequence was completely covered, indicating that there is a missing piece in the assembly of the 'Wijcik' insertion that was part of Wolters et al. (2013). Because the 5' LTR of the 8.2 kb retrotransposon is 100% identical to the 3' LTR (Otto et al. 2013) and because the LTRs are a lot longer (1951 bp) than the maximum length of the 454 reads (806 bp), it looks like the assembly incorrectly collapsed on the LTR of the retrotransposon.

The PCR product suggesting that the insert is 1956 bp instead of 8.2 kb in length could be an artefact. As we expected to amplify a 1956 bp insertion, the elongation phase of that PCR lasted only 2 min. In 2 min, it is not possible to amplify the full 8.2 kb retrotransposon, but many PCR fragments are produced from the terminal parts (the LTRs) of the retrotransposon. In the next round of PCR, fragments derived from the 5' LTR can re-anneal to the 3' LTR and vice versa. This way, new DNA template could be formed that contains both primer sites that flank a single LTR only, instead of the complete element. Because of the smaller size of this template, it would amplify better than the original template containing the 8.2 kb sequence in subsequent amplification cycles (summarised in Fig. 1d).

With this explanation for the 2 kb fragments obtained in the PCR displayed in Fig. 1b, the fact that the PCR displayed in Fig. 1c matches with the findings of Otto et al. (2013) and after the identification of 454 reads among the reads obtained from the BAC clone containing the 'Wijcik' insertion that cover the complete 8.2 kb retrotransposon, it can be concluded that the discrepancy between the results obtained in chapter 3 of this thesis and the results described by Otto et al. is explained by an assembly error and a PCR artefact. The correct size of the 'Wijcik' insertion is 8.2 kb.

The 8.2 kb insertion happened inside another piece of repetitive DNA of about 350 bp. A closer look at this 350 bp sequence, and its surrounding region,

revealed a second match of a very similar sequence around 3.2 kb downstream. The presence of a TSD at the 5' and the 3' of the two repetitive sequences indicates that this sequence is another retrotransposon with LTRs of 348 bp. A gene fragment is contained within this retrotransposon that was annotated before as *MDP0000766466* and named *MdCo29* in the previous chapter. The complete size of this 'McIntosh' retrotransposon is 3959 bp. The 8.2 kb 'Wijcik' retrotransposon inserted in the 5' LTR of this smaller retrotransposon, resulting in a nested TE in 'Wijcik', as presented in Fig. 2.

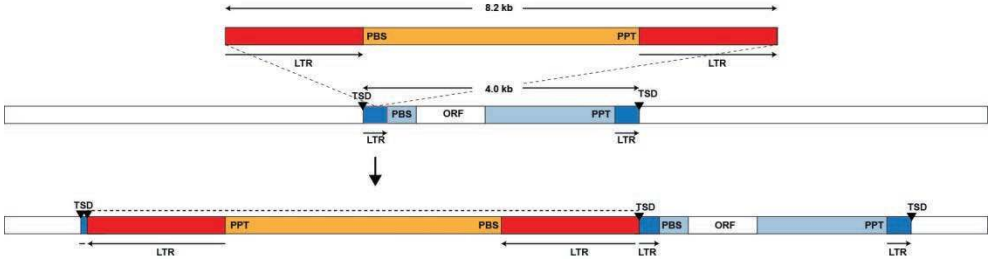

Figure 2. **Correct depiction of the 'Wijcik' mutation**. The mutation that resulted in columnar growth of 'Wijcik' (red) is an insertion of an 8.2 kb retrotransposon in the LTR of another, 4.0 kb, retrotransposon that was already present in 'McIntosh' (blue). Primer binding site (PBS), polypurine tract (PPT), long terminal repeats (LTRs) and target site duplications (TSDs) are indicated in the figure.

Similar TEs in 'Golden Delicious'

Both retrotransposons that were identified contain small ORFs, which seem to correspond to fragments of genes, but they do not contain the typical ORFs that are necessary for autonomous replication of the elements. They contain a PPT (a sequence rich in purines that is cleaved during reverse transcription and used as a primer to initiate syntheses of the plus strand of the retrotransposon DNA) (Petropoulos 1997; Otto et al. 2013) and a PBS (the binding site for a tRNA that functions as the primer for reverse transcriptase to initiate synthesis of the minus strand of the DNA for replication). The two elements can thus be classified as large retrotransposon derivatives (LARDs, (Kalendar et al. 2004)), or, according to the classification system that was proposed by Wicker et al. (2007), as class I elements belonging to the LTR order. Until the identification of their autonomous partners, it will not be possible to assign them to an existing superfamily or family. We will refer to the 8.2 kb retrotransposon as 'Wijcik' retrotransposon and to the smaller 4.0 kb retrotransposon as 'McIntosh' retrotransposon.

In an attempt to identify similar elements (and to possibly identify the autonomous partners), a BLAST search for LTRs of the two retrotransposons was performed against the 'Golden Delicious' genome. This BLAST search resulted in the identification of 137 LTR matches for the 'Wijcik' retrotransposon and 305 matches for the 'McIntosh' retrotransposon LTR sequence.

As complete retrotransposons are delimited by LTRs, a subsequent analysis is performed to identify sequences that are flanked by two LTRs within 10 kb from each other. The results of this analysis are displayed as a graph in Fig. 3. The distance (in kb) from the LTRs that were identified during the BLAST search is shown on the horizontal axis and the number of LTRs with a second similar LTR match at that location is displayed on the vertical axis. As can be concluded from this figure, many elements homologous to the 'McIntosh' retrotransposon are present in the 'Golden Delicious' genome, but there is only one element present that has two LTRs with homology to the 'Wijcik' retrotransposon. The size of this second element is only 4.8 kb, a lot less than the 8.2 kb of the 'Wijcik' insertion. As the typical ORFs for autonomous retrotransposons are also absent in this sequence, this cannot be the autonomous partner of the 'Wijcik' retrotransposon.

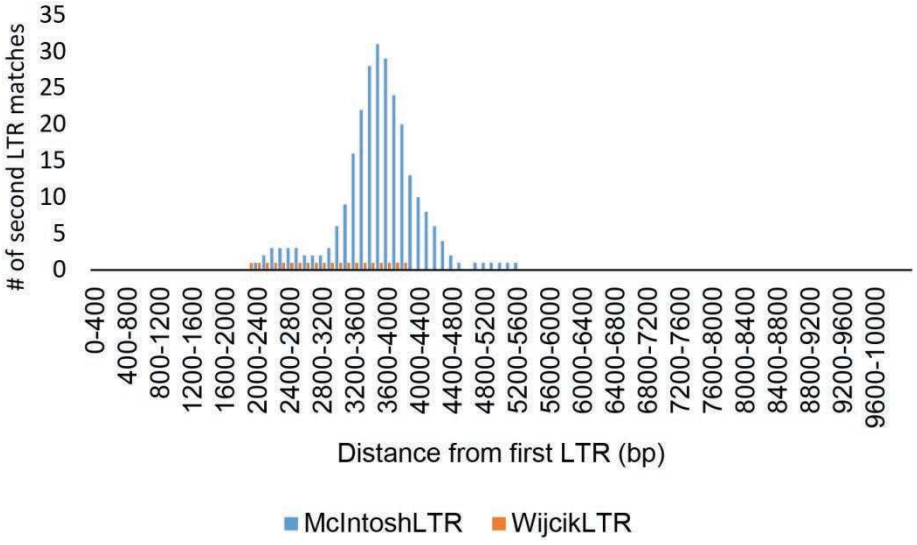

Figure 3. Second LTR matches around LTR locations. Graph depicting the number of LTRs that have a second LTR match at a distance from 0 to 10 kb from the 'Wijcik' retrotransposon-like LTR (red) or the 'McIntosh'-like retrotransposon LTR (blue).

A total of 40 ‘McIntosh’-like retrotransposon sequences were detected in the ‘Golden Delicious’ genome, which range in length from around 3-4 kb (Fig. 4, see Table 1 for an overview of the locations of these elements and find a summary of the locations of both elements in Fig. 5). A BLASTX search was performed against these 40 elements, in order to find candidate autonomous partners of the ‘McIntosh’ retrotransposon. Around 75% of the 40 elements have good matches to fragments of proteins from the Uniref90 protein database (Suzek et al. 2007), but no complete genes were detected inside the ‘McIntosh’ retrotransposon and no retrotransposon domains were identified in any of them.

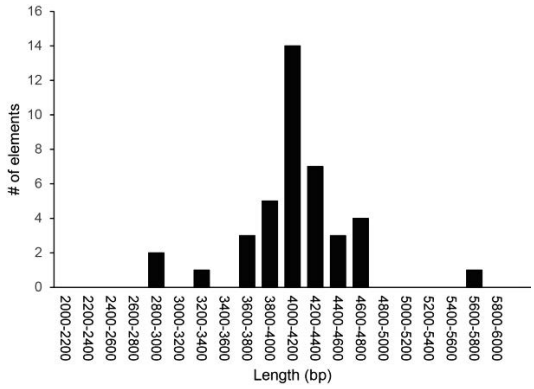

Figure 4. ‘McIntosh’-like element lengths. Size distribution of the 40 ‘McIntosh’ retrotransposon-like sequences that were found in the ‘Golden Delicious’ genome.

TE insert sites

The ‘Wijcik’ retrotransposon moved only about 50 years ago. To see if the ‘McIntosh’ and ‘Wijcik’ retrotransposon-like sequences are still active, the insert sites of the elements identified in the ‘Golden Delicious’ genome were compared to ‘McIntosh’, using Illumina reads derived from genomic DNA of ‘McIntosh’ that were produced as part of a bigger project to identify single nucleotide polymorphisms in diverse apple cultivars. Thus, the presence of the elements identified in ‘Golden Delicious’ was verified in ‘McIntosh’. Additionally, the corresponding ‘vacant’ sites (before insertion of the retrotransposons) were searched for in ‘McIntosh’. This analysis did not result in the identification of any differences between ‘Golden Delicious’ and ‘McIntosh’.

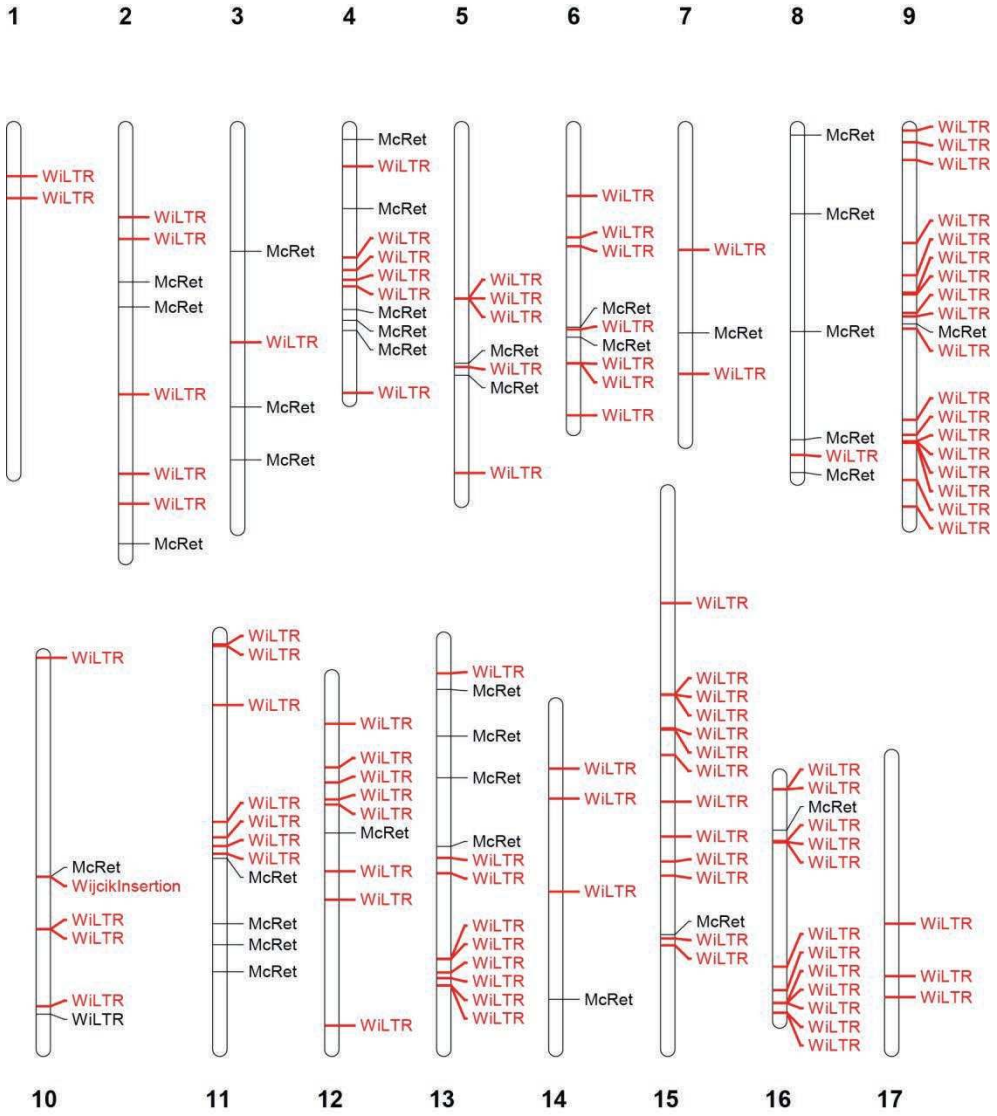

Figure 5. Overview of the distribution of the 'McIntosh'-like retrotransposons and 'Wijcik' retrotransposon-like LTRs over the 17 chromosomes of 'Golden Delicious'. 'McIntosh'-like retrotransposons are indicated in black, 'Wijcik' retrotransposon-like LTRs in red.

As explained in the introduction, TEs can affect the expression of genes in their vicinity and some TEs show a preference for inserting close to genes (Liu et al. 2009; Naito et al. 2009; Jiang et al. 2011). The insertion of the ´Wijcik´ retrotransposon was shown to be correlated with increased expression of *MdCo31* in chapter 3. To see if the TEs identified in 'Golden Delicious' could be potential

gene regulators, similar to the mPing element that was identified in rice (Naito et al. 2009), a gene density analysis was performed around these elements. 500 locations that were picked randomly in the 'Golden Delicious' genome were included in the analysis to show the average gene density in the genome. The gene density was examined in blocks of 5 kb in a total region of 50 kb around the LTRs that show similarity to the ´Wijcik´ and 'McIntosh'-like retrotransposon and the 500 randomly picked locations in the 'Golden Delicious' genome. The results are displayed in Fig. 6.

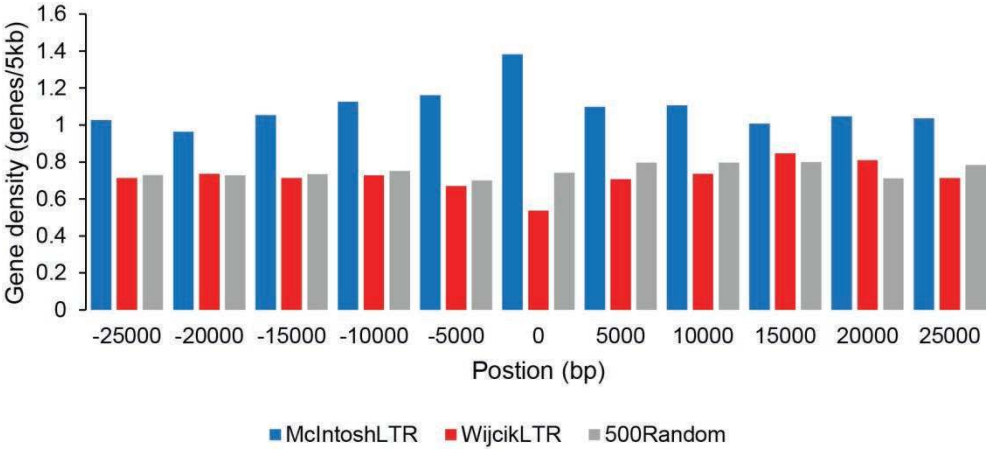

Figure 6. Gene density around 'Golden Delicious' genome locations. The average number of genes in intervals of 5 kb around 'McIntosh'-like retrotransposon LTRs (blue), 'Wijcik' retrotransposon-like LTRs (red) and randomly picked positions (grey) in the 'Golden Delicious' genome.

Around the 1950 bp LTRs that are similar to the LTRs in the 'Wijcik' retrotransposon, the gene density is around 0.72 genes/5 kb. Around the randomly picked locations, the gene density is about 0.75, very similar to the number for the 'Wijcik' retrotransposon-like LTR locations, so it seems that this element has no preference for inserting close to genes. The ´McIntosh´-like retrotransposon sequences, however, do seem to have a preference for inserting into gene rich regions. The average gene density around the LTRs of these retrotransposons is about 1.10 genes/5 kb. The peak at the location of the retrotransposon itself can be explained by the presence of ORFs of gene fragments inside the retrotransposon, which are often annotated as genes in the 'Golden Delicious' gene predictions.

Table 1. Overview of the locations of the 40 retrotransposons that bear similarity to the 'McIntosh' retrotransposon that were identified in the 'Golden Delicious' genome

Name	GD contig	start	end	orientation	chromosome	start	end
McIntosh1	MDC007740.273	30764	34589	-	-	-	
McIntosh2	MDC013825.249	4139	7976	+	-	-	
McIntosh3	MDC004651.427	2790	6747	-	-	-	
McIntosh4	MDC008041.193	5165	9627	+	-	-	
McIntosh5	MDC011766.735	2239	4935	-	-	-	
McIntosh6	MDC025235.40	8634	12601	-	chr2	13044369	13048336
McIntosh7	MDC021955.80	6696	10543	+	chr2	15181238	15185085
McIntosh8	MDC004448.281	2322	6771	+	chr2	35342756	35347205
McIntosh9	MDC011766.720	345	2978	-	chr3	10444238	10446871
McIntosh10	MDC021353.148	2686	6423	+	chr3	23690989	23694726
McIntosh11	MDC001196.307	5456	9365	+	chr3	28197997	28201906
McIntosh12	MDC029659.16	1929	5437	-	chr4	921402	924910
McIntosh13	MDC007239.507	5554	9599	-	chr4	6810708	6814753
McIntosh14	MDC011832.263	1733	5906	+	chr4	15407720	15411893
McIntosh15	MDC011287.339	16219	20758	-	chr4	16315500	16320039
McIntosh16	MDC012792.364	5895	11305	-	chr4	17193419	17198829
McIntosh17	MDC013219.318	5836	9510	+	chr5	19960689	19964363
McIntosh18	MDC011373.251	7237	11053	+	chr5	21005990	21009806
McIntosh19	MDC013503.290	9698	12807	+	chr6	16891710	16894819
McIntosh20	MDC002608.621	3593	7922	+	chr7	17393595	17397924
McIntosh21	MDC021393.441	16572	20677	-	chr6	17752338	17756443
McIntosh22	MDC014206.69	8023	12019	+	chr8	547004	551000
McIntosh23	MDC006328.490	2782	6302	+	chr8	7266438	7269958
McIntosh24	MDC021955.78	3934	7772	+	chr8	17276132	17279970
McIntosh25	MDC022206.381	14362	18118	-	chr8	26474169	26477925
McIntosh26	MDC008101.327	4562	8293	+	chr8	29280369	29284100
McIntosh27	MDC008995.306	10728	14783	-	chr9	16625859	16629914
McIntosh28	MDC009292.256	13322	17271	+	chr10	18796841	18800790
McIntosh29	MDC008101.323	1573	5569	+	chr11	19066775	19070771
McIntosh30	MDC037014.7	18093	21910	-	chr11	24626736	24630553
McIntosh31	MDC013297.192	22015	26073	+	chr11	26411660	26415718
McIntosh32	MDC012795.433	5547	9843	+	chr11	28712042	28716338
McIntosh33	MDC017110.186	24472	28592	-	chr12	13288971	13293091

McIntosh34	MDC009334.243	21650	25863	-	chr13	4299583	4303796
McIntosh35	MDC021754.182	14694	18564	+	chr13	8267463	8271333
McIntosh36	MDC017873.442	7636	11346	+	chr13	11807546	11811256
McIntosh37	MDC011267.253	33551	38138	+	chr13	17657023	17661610
McIntosh38	MDC021221.287	3429	6943	+	chr14	25074401	25077915
McIntosh39	MDC008573.596	34910	38939	+	chr15	37696716	37700745
McIntosh40	MDC015244.92	4119	8087	-	chr16	4609410	4613378

Discussion

The results described in this chapter show that there was a mistake in the assembly of the insertion that is responsible for the columnar phenotype of 'Wijcik' apple earlier described by Wolters et al. (2013). After comparing the results from the previous chapter with the findings of Otto et al. (2013) and after a review of the analysis that lead to the mistake in the assembly, a correct depiction of the 'Wijcik' mutation has been presented in Fig. 2.

The assembly error can be explained by the size of the LTRs of the 'Wijcik' retrotransposon. The LTR sequences are almost 2 kb. Because the 454 reads are too short to span the complete LTR, it was possible for the assembly to collapse on this sequence. Also, the PCR that was designed to amplify the complete 'Wijcik' insertion produced PCR fragments that could re-anneal to both sides of the retrotransposon, leading to an incorrect interpretation of the insert size in a control PCR. This could not have happened for the shorter LTRs of the 'McIntosh' element.

Looking at the 'Golden Delicious' genome, it seems that the assembly error is rather common in the genome assembly, as many single LTRs can be found that are flanked by TSDs. In fact, although many LTRs homologous to the 'Wijcik' retrotransposon are present in the 'Golden Delicious' genome, only one sequence is found that is flanked by two LTRs like the 'Wijcik' retrotransposon. A second explanation for the low number of elements homologous to the 'Wijcik' insertion found in the 'Golden Delicious' genome is the fact that more than 70% of the contigs containing a similar LTR sequence is smaller than 10 kb, indicating that many contigs probably contain only a part of the complete retrotransposon. The 'McIntosh'-like retrotransposons are smaller in size and have shorter LTRs, making

them easier to assemble. This is illustrated by the fact that many homologous elements could be identified in the apple genome.

The LTRs of the ‘Wijcik’ retrotransposon start with 5′-TG-3′ and end with 5′-CA-3′, whereas the LTRs of the ‘McIntosh’-like retrotransposon start with 5′-TG-3′ and end with 5′-CAA-3′. They both have a 4-5 bp TSD and a recognizable PBS and PPT, looking like typical LTR retrotransposons (Wicker et al. 2007). However, no ORFs could be identified in the retrotransposons that could enable them to replicate, meaning that they are non-autonomous elements. The ‘Wijcik’ retrotransposon would have needed the help of an autonomous partner when it moved 50 years ago.

It proved to be difficult to identify such an autonomous partners of the 'McIntosh' and the 'Wijcik' retrotransposon in the ‘Golden Delicious’ genome. An alternative to the *in silico* approach that was carried out here, a PCR-based approach could help to elucidate the autonomous partners. By using combinations of PCR primers designed on the LTR sequences of the retrotransposons and primers designed on the conserved sequences of ORFs that would be expected to be present in autonomous elements, it might be possible to amplify and identify the sequences of the autonomous partners.

Although no ORFs specific for retrotransposons could be detected in the two nested elements, or in the homologous elements from ‘Golden Delicious’, many ORFs were found in the ‘McIntosh’-like retrotransposon from ‘Golden Delicious’. Those ORFs corresponded to gene fragments, but it is unclear where these fragments originated from. They could have been captured from other places in the host genome, like was found for the Pack-MULEs in rice (Juretic et al. 2005) and helitrons in maize (Morgante et al. 2005), but a BLAST search in the apple gene set did not result in the identification of the original genes from which the fragments could have been derived.

The LTRs of the ´Wijcik´ retrotransposon are 100% identical, supporting its recent integration in the ‘McIntosh’ genome. The 350 bp LTRs of the ‘McIntosh’ share 95% identity, demonstrating that this element inserted at an earlier time point in the ‘McIntosh’ genome. An attempt was made to compare the retrotransposon

content of ‘Golden Delicious’ and ‘McIntosh’, but no differences were found, suggesting that movement of these elements is not that common. It would be interesting to perform a more extensive analysis, looking at a bigger number of cultivars and focussing on the integration sites of the retrotransposons identified in this study, to see if differences can be observed between different cultivars. If such differences could be identified, these results could also be used to assess the effect of the integration of these retrotransposons on the expression of surrounding genes.

The effect of the 'Wijcik' insertion on the expression of *MdCo31*, and its subsequent effect on plant development, is another example of how TEs can contribute to phenotypic changes in their host. Many of such examples were already described in the introduction of this chapter. How the ‘Wijcik’ insertion affects the expression of *MdCo31* is still unclear. The effect could be explained by a stimulating effect of the presence of the ‘Wijcik’ retrotransposon on the expression of *MdCo31*, or by the disrupting effect of the ‘Wijcik’ insertion on the ‘McIntosh’ retrotransposon. In the first scenario, an enhancing element is present in the ‘Wijcik’ retrotransposon, whereas the second scenario can be explained by the presence of an inhibitor in the ‘McIntosh’ retrotransposon. The gene density analysis that was carried out in this chapter shows no preference for inserting close to gene-rich areas for the ‘Wijcik’ transposon, but the ‘McIntosh’ retrotransposon does seem to have a preference for inserting in such areas.

Whether the effect of the mutation in ‘Wijcik’ and the expression of *MdCo31* is due to the insertion of the ‘Wijcik’ retrotransposon or due to the disruption of the ‘McIntosh’ retrotransposon, the mutation site and *MdCo31* have to be in close proximity to each other for the effect to occur (schematically displayed in Fig. 7). In an alternative explanation, *MdCo31* is under control of an enhancing or repressing element at a different location of the ‘McIntosh’ genome. The insertion in ‘Wijcik’ could have changed the DNA organisation, leading to a change in the interaction between *MdCo31* and this controlling element. In all explanations, the conformation of the chromatin seems to be important to explain the effect of the ‘Wijcik’ insertion on the expression of *MdCo31*.

A technique that could be used to see if the insertion in 'Wijcik' led to a change in chromatin conformation is Chromosome Conformation Capture (3C) (Dekker et al. 2002). The first step in 3C is to crosslink DNA and proteins using formaldehyde. Next, the DNA is fragmented by digesting it with a restriction enzyme, after which the DNA molecules are religated. This ligation step results in many different ligation products, but fragments that are physically close (due to DNA conformation mediated through protein interactions) are more likely to ligate to each other. The potential interactions can be investigated using PCR based assays on the resulting 3C templates using primers designed on restriction site junctions. In plants, 3C has been used to investigate the chromatin looping in the booster 1 (*b1*) locus of maize (Louwers et al. 2009). A detailed protocol for performing 3C experiments and experiments based on 3C technologies was recently published (Hövel et al. 2012).

Although we showed a sequence difference between 'McIntosh' and 'Wijcik', and although the stability of the columnar phenotype already suggested a genomic mutation rather than an epigenetic mutation only, it would still be interesting to investigate the epigenetic landscape around the 'Wijcik' retrotransposon and *MdCo31*. The 'Wijcik' insertion could have had an effect on the epigenetic marks of the *Co* region. A simple way to test this would be to make use of isoschizomers of restriction enzymes that are sensitive and non-sensitive to DNA methylation (for example MspI/HpaII). After separate digestion with the methyl sensitive and the insensitive enzyme, PCRs on the restriction sites within the region of interest would show whether these sites were methylated. Alternatively, bisulphite sequencing or chromatin immunoprecipitation experiments could be carried out to investigate and compare DNA and histone methylation between 'McIntosh' and 'Wijcik'.

A suitable restriction enzyme for the 3C assay could be EcoRI, which performs well in 3C experiments (Simonis et al. 2007) and which has many restriction sites in the 'Wijcik' region. The MspI/HpaII combination could be used for the methylation sensitive digestion. The restriction sites of these enzymes in a region of 60 kb surrounding the ´Wijcik´ insertion are displayed in Fig. 7. Using both methods, candidate loci could be identified that are responsible for the regulation of

MdCo31 expression. Once these loci have been identified, their effect on the expression of *MdCo31* could be tested using reporter gene assays.

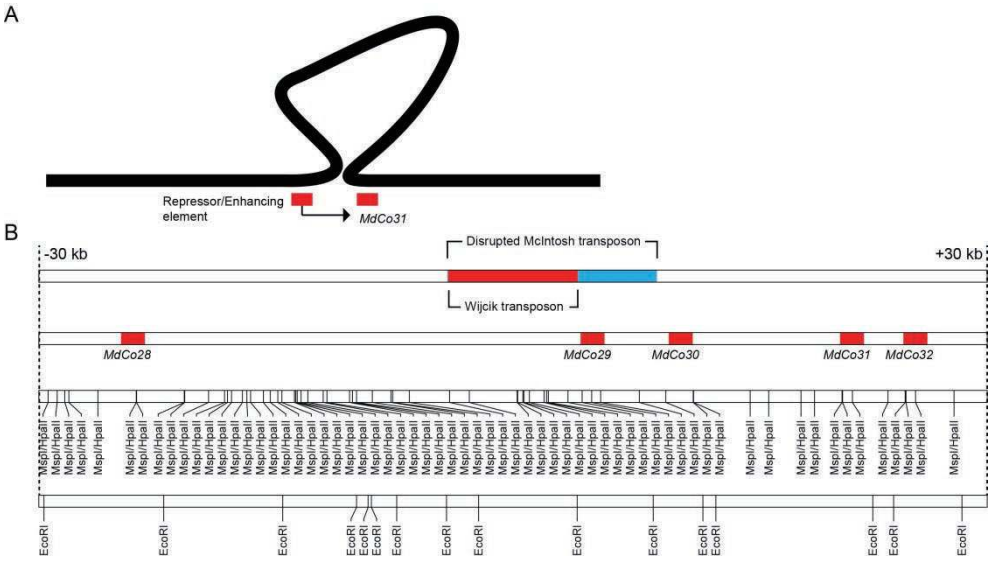

Figure 7. Organisation of the region around the 'Wijcik' insertion. If the 'Wijcik' insertion contains an enhancer, or if the disrupted 'McIntosh' motif contained a repressor that lost its functionality due to the 'Wijcik' insertion, chromatin folding could be responsible for bringing the regulatory element and *MdCo31* close to each other, as schematically drawn in (A). **B.** Different views of a region of 60 kb around the 'Wijcik' insertion are shown. At the top, the location of the 'Wijcik' retrotransposon (red) and the disrupted 'McIntosh' retrotransposon (blue) are indicated. Below, the locations of the gene predictions in the *Co* region (chapter 2) are depicted in red. The locations of MspI/HpaII and EcoRI restriction sites in the 60 kb region are shown at the bottom. These restriction sites could be used to investigate DNA methylation of the region, or to perform a 3C assay.

References

Bennetzen, J. L., J. Ma, et al. (2005). "Mechanisms of recent genome size variation in flowering plants." Annals of botany **95**(1): 127-132.

ENCODE Project Consortium. (2012). "An integrated encyclopedia of DNA elements in the human genome." Nature **489**(7414): 57-74.

Dekker, J., K. Rippe, et al. (2002). "Capturing chromosome conformation." Science **295**(5558): 1306-1311.

Finnegan, D. J. (1989). "Eukaryotic transposable elements and genome evolution." TRENDS in Genetics **5**: 103-107.

Fisher, D. V. (1969). "Spur-type strains of McIntosh for high density planting." British Columbia Fruit Grower's Association Quart. Rep. **14**: 3-10.

Greene, B., R. Walko, et al. (1994). "Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations." Genetics **138**(4): 1275-1285.

Hövel, I., M. Louwers, et al. (2012). "3C Technologies in plants." Methods **58**(3): 204-211.

Iida, S., Y. Morita, et al. (2004). "Genetics and epigenetics in flower pigmentation associated with transposable elements in morning glories." Advances in biophysics **38**: 141-159.

Initiative, A. G. (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana." Nature **408**(6814): 796.

Jiang, N., A. A. Ferguson, et al. (2011). "Pack-Mutator–like transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition." Proceedings of the National Academy of Sciences **108**(4): 1537-1542.

Juretic, N., D. R. Hoen, et al. (2005). "The evolutionary fate of MULE-mediated duplications of host gene fragments in rice." Genome Res **15**(9): 1292-1297.

Jurka, J., V. V. Kapitonov, et al. (2005). "Repbase Update, a database of eukaryotic repetitive elements." Cytogenetic and genome research **110**(1-4): 462-467.

Kalendar, R., C. M. Vicient, et al. (2004). "Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes." Genetics **166**(3): 1437-1450.

Levin, H. L. and J. V. Moran (2011). "Dynamic interactions between transposable elements and their hosts." Nature Reviews Genetics **12**(9): 615-627.

Lisch, D. (2013). "How important are transposons for plant evolution?" Nat Rev Genet **14**(1): 49-61.

Liu, J., Y. He, et al. (2004). "siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis." Genes & Development **18**(23): 2873-2878.

Liu, S., C.-T. Yeh, et al. (2009). "Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome." PLoS genetics **5**(11): e1000733.

Louwers, M., R. Bader, et al. (2009). "Tissue-and expression level–specific chromatin looping at maize b1 epialleles." The Plant Cell Online **21**(3): 832-842.

Martienssen, R. and A. Baron (1994). "Coordinate suppression of mutations caused by Robertson's mutator transposons in maize." Genetics **136**(3): 1157.

McClintock, B. (1984). "The significance of responses of the genome to challenge." Science **226**(4676): 792-801.

Meyers, B. C., S. V. Tingey, et al. (2001). "Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome." Genome Res **11**(10): 1660-1676.

Morgante, M., S. Brunner, et al. (2005). "Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize." Nat Genet **37**(9): 997-1002.

Naito, K., E. Cho, et al. (2006). "Dramatic amplification of a rice transposable element during recent domestication." Proceedings of the National Academy of Sciences **103**(47): 17620-17625.

Naito, K., F. Zhang, et al. (2009). "Unexpected consequences of a sudden and massive transposon amplification on rice gene expression." Nature **461**(7267): 1130-1134.

Otto, D., R. Petersen, et al. (2013). "The columnar mutation ("Co gene") of apple (Malus × domestica) is associated with an integration of a Gypsy-like retrotransposon." Molecular Breeding: 1-18.

Petropoulos, C. (1997). Retroviral taxonomy, protein structures, sequences, and genetic maps.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Sanyal, A., B. R. Lajoie, et al. (2012). "The long-range interaction landscape of gene promoters." Nature **489**(7414): 109-113.

Schnable, P. S., D. Ware, et al. (2009). "The B73 Maize Genome: Complexity, Diversity, and Dynamics." Science **326**(5956): 1112-1115.

Shapiro, J. A. (2011). Evolution: a view from the 21st century, Pearson Education.

Simonis, M., J. Kooren, et al. (2007). "An evaluation of 3C-based methods to capture DNA interactions." Nature methods **4**(11): 895-901.

Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nature Reviews Genetics **8**(4): 272-285.

Studer, A., Q. Zhao, et al. (2011). "Identification of a functional transposon insertion in the maize domestication gene tb1." Nat Genet **43**(11): 1160-1163.

Suzek, B. E., H. Huang, et al. (2007). "UniRef: comprehensive and non-redundant UniProt reference clusters." Bioinformatics **23**(10): 1282-1288.

Velasco, R., A. Zharkikh, et al. (2010). "The genome of the domesticated apple (Malus x domestica Borkh.)." Nature Genetics **42**(10): 833-839.

Wicker, T., F. Sabot, et al. (2007). "A unified classification system for eukaryotic transposable elements." Nature Reviews Genetics **8**(12): 973-982.

Wolters, P. J., H. J. Schouten, et al. (2013). "Evidence for regulation of columnar habit in apple by a putative 2OG-Fe(II) oxygenase." New Phytologist **200**(4): 993-999.

Xiao, H., N. Jiang, et al. (2008). "A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit." Science **319**(5869): 1527-1530.

Chapter 5

Functional analysis of *MdCo31*

Pieter J. Wolters, Henk J. Schouten, Paolo Baldi, Azeddine Si-Ammour

Abstract

Earlier work showed that increased expression of a putative 2OG-Fe(II) oxygenase, *MdCo31*, is likely to be the underlying cause of columnar growth in apple, which occurs in a mutant of cv. 'McIntosh' called 'Wijcik'. Enzymes belonging to the 2-oxoglutarate-dependent dioxygenase (2OGD) family are involved in oxidation and hydroxylation reactions of organic substrates. The 2-OGD family is vast and contains members with important and well characterised functions in plant development, but also includes specialised enzymes and uncharacterised genes. In this chapter, we show that *MdCo31* expression has a comparable phenotypic effect on the development of diverse species, suggesting a conserved function of *MdCo31*. Nonetheless, no well-characterised homologs were found in these species. 2-OGDs can have roles in the pathways leading to biosynthesis of gibberellic acid (GA) and flavonoids and these compounds could be responsible for the phenotype resulting from *MdCo31* expression. Profiling of phenolic compounds in columnar apple trees and standard apple trees showed increased levels of rutin in columnar trees and elevated levels of luteolin were found in Arabidopsis plants that constitutively expressed *MdCo31*. However, no effect of these compounds was found when they were added to medium that was used to grow Arabidopsis seedlings that could explain how *MdCo31* affects plants development. Conversely, application of active GA on Arabidopsis plants constitutively expressing *MdCo31* almost completely abolished the effect of *MdCo31*, suggesting a role for *MdCo31* in GA de-activation. Whether *MdCo31* is involved in flavonoid biosynthesis, GA biosynthesis, or another process is still unknown and a number of experiments are discussed to further investigate this topic.

Introduction

MdCo31 putatively codes for a 2OG-FeII oxygenase (Wolters et al. 2013). The 2-oxoglutarate-dependent dioxygenase (2OGD) family is vast, with 130 members in Arabidopsis (Kawai et al. 2014) and 263 2OGD genes in apple according to the PLAZA 2.5 predictions (Van Bel et al. 2012).

2OGDs catalyse the oxidation and hydroxylation of organic substrates, often using 2-oxoglutarate as a co-substrate. In addition to 2-oxoglutarate, they require Fe^{2+} for optimal substrate conversion (Prescott and John 1996). The family includes well characterised genes with essential roles in the biosynthesis of flavonoids, gibberellic acid (GA), alkaloids and ethylene, but also includes many uncharacterised genes for which the biochemical function is unknown (Kawai et al. 2014).

Because of the many 2OGDs present in plant genomes, their functional diversity and their highly divergent amino acid sequences, the classification of the different enzymes is complicated. A phylogenetic classification of 2OGDs from diverse species ranging from green algae to angiosperms was published very recently, in which a distinction is made between 3 different classes of 2OGDs. The largest class, DOXC, consists of 57 clades, of which 31 were found in only one of the 6 species analysed (Kawai et al. 2014), suggesting specialised functions for many 2OGDs.

Regarding the columnar growth of 'Wijcik', the role of 2OGDs in flavonoid and GA biosynthesis is particularly interesting. Several examples exist of columnar-like phenotypes that are due to GAs or flavonoids. The most striking example is *stumpy* poplar, in which overexpression of a GA2-oxidase leads to a phenotype that is very similar to that of 'Wijcik', with reduced internode length, reduced branching and shorter branches (Busov et al. 2003). Support for the role of flavonoids in a columnar-like phenotype in apple came from a recent study where the chalcone synthase (*CHS*) genes were silenced (Dare et al. 2013). Besides having an effect on plant colouring, *CHS* silencing resulted in small plants, characterised by short internodes and smaller leaves. In the next section, we will describe the pathways leading to the biosynthesis of GAs and flavonoids in more detail, while focusing on the role of 2OGDs in these pathways.

GAs are well-known regulators of plant growth, and plants with reduced levels of active GAs or a reduced response to GA show dwarf phenotypes (Peng et al. 1997; Peng et al. 1999; Boss and Thomas 2002; Busov et al. 2003). GAs are produced through the diterpenoid pathway in plants. Three different classes of

enzymes are involved in this pathway: terpene synthases (TPSs), cytochrome P450 monooxygenases (P450s), and several types of 2OGDs.

The TPSs are responsible for the conversion of geranylgeranyl diphosphate to ent-kaurene. The subsequent conversion of ent-kaurene to GA_{12} is mediated through the P450s. The next steps in GA biosynthesis are carried out by 2OGDs mainly (Yamaguchi 2008).

Many different forms of GA exist in plants, of which most are precursors or deactivated forms of the principle active forms of GA, GA_1 and GA_4. GA_4 production starts directly from GA_{12}, whereas GA_1 is produced from GA_{53} that is produced through C13-hydroxylation of GA_{12}. The genes encoding GA13-oxidases have been identified very recently in rice and do not belong to the 2OGD class (Magome et al. 2013).

The later steps in the GA biosynthesis pathway are carried out by several 2OGD classes that include GA20-oxidases, GA3-oxidases and GA2-oxidases. GA_{12} and GA_{53} are processed through sequential oxidation of C-20 by GA20-oxidases while GA_1 and GA_4 are produced from GA_{20} and GA_9 by GA3-oxidases. The active forms of GA are deactivated by GA2-oxidases mainly (Hedden and Thomas 2012). A possible GA2-oxidase activity of *MdCo31* would result in decreased levels of active GAs in 'Wijcik', in which expression of this gene was found to be increased in axillary buds that were harvested in spring. A role for *MdCo31* in a different part of the GA pathway could possibly affect the balance between other forms of GA.

2OGDs also have roles in flavonoid biosynthesis. Flavonoids can negatively regulate auxin transport (Brown et al. 2001; Peer and Murphy 2007) and auxin transport is believed to be required for outgrowth of axillary buds (Prusinkiewicz et al. 2009), providing a possible mechanism for causing the columnar phenotype of 'Wijcik'. A role for auxin in columnar growth of apple was also hypothesised by Petersen et al. (2013), who suggested to compare the rate of auxin transport between columnar and standard trees to see if auxin transport rate differs between these plants (Petersen and Krost 2013).

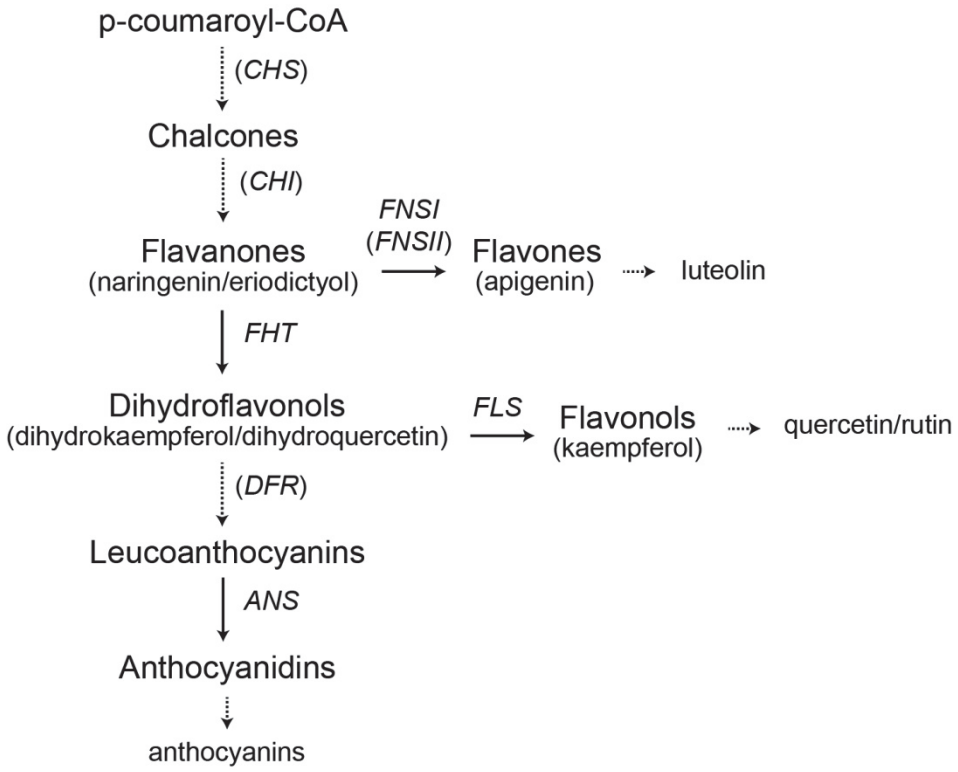

Figure 1. 2OGDs and the flavonoid pathway. Adapted from Martens et al. (2010) and the KEGG pathway. p-coumaroyl-CoA is converted to the chalcone backbone that forms the basis of flavonoid compounds. Solid arrows indicate conversion steps that can be mediated through 2OGD enzymes, whereas dotted arrows indicate conversion step that are mediated by enzymes belonging to other classes (between brackets). Luteolin is produced through the flavone branch of the flavonoid pathway, whereas rutin is produced in the flavonol branch. *CHS*: chalcone synthase, *CHI*: chalcone isomerase, *F3H*: flavanone 3-ß-hydroxylase, *FNSI/FNSII*: flavone synthase I/II, *FLS*: flavonol synthase, *DFR*: dihydroflavonol 4-reductase, *ANS*: anthocyanidin synthase.

The flavonoid biosynthesis pathway is part of the phenylpropanoid pathway, where phenylalanine is transformed in p-coumaroyl-CoA. This pathway, starting from p-coumaroyl-CoA and leading to production of the different classes of flavonoids, is summarised in Fig. 1. *CHS* is the first flavonoid specific enzyme and is responsible for the conversion of p-coumaroyl-CoA into the chalcone scaffold that forms the backbone of all flavonoids. Chalcones are converted to flavanones (such as naringenin and eriodictyol) through chalcone isomerase (*CHI)*. Different oxidation reactions, mediated by several isomerases, reductases, hydroxylases and including 2OGDs, lead to the different flavonoid subclasses (Ferreyra et al.

2011). Four different types of 2OGDs are active in the flavonoid pathway: flavanone 3β-hydroxylase (*FHT*, or *F3H*), flavonol synthases (*FLS*) flavone synthase 1 (*FNS1*) and leucoanthocyanidin dioxygenase (*LDOX* or *ANS*) (Cheng et al. 2014).

FHT is responsible for the conversion of flavanones into dihydroflavonols, that can be converted to flavonols through the action of *F3H*, or, alternatively, to anthocyanidins through *DFR* (dihydroflavonol 4-reductase) and *ANS* (Martens et al. 2010). The 2OGD *FNSI* mediates the conversion of flavanones into flavones in *Apiaceae* only (Gebhardt et al. 2005). This conversion is mediated by the P450 *FNSII* in other species (Lee et al. 2008).

The 2OGDs from the flavonoid pathway show overlapping substrate specificity and all 2OGDs involved in the pathway can accept naringenin as a substrate (Cheng et al. 2014). *FLS* and *ANS* have a broader substrate specificity and accept the flavanone substrates that are processed by *FHT* and *FNSI*. The substrate specificity of *FNSI* and *F3H* is more restricted (Martens et al. 2010). If *MdCo31* has a role in the flavonoid pathway, this could affect the levels of flavonoids that act as negative regulators of auxin transport.

The biochemical function of *MdCo31* was investigated in this chapter. Besides in apple, we studied the function of *MdCo31* in Arabidopsis and tomato. Arabidopsis plants were used to test the hypothesis that *MdCo31* is involved in the biosynthesis of GAs, using GA treatments. The alternative hypothesis, that auxin transport is affected due to the putative role of *MdCo31* in flavonoid biosynthesis, was investigated with a gravitropism assay on Arabidopsis seedlings.

Materials and methods

Plant material

The Arabidopsis lines used in this study were of Columbia ecotype and the tomato variety that was used is 'Moneymaker'. Plants were grown in soil in a growth chamber at 22 °C with a long photoperiod (16 h of 100 µmol m^{-2} s^{-1} light, 8 h of dark). Apple trees were grown and maintained at the orchard 'Giaroni' belonging to the Fondazione Edmund Mach (FEM) at the Istituto Agrario di San Michele all'Adige (IASMA) located in Italy (latitude 46.181539°, longitude 11.119877°).

Nucleic acids were extracted from ~100 mg plant tissue that was grinded to a fine powder in liquid nitrogen. Genomic DNA was isolated using the DNeasy Plant Mini Kit (QIAGEN).

Tomato transformation

pENTR/D-MdCo31 (Wolters et al. 2013) was recombined with the pK7WG2 vector (Karimi et al. 2002), using the Gateway LR Clonase II kit (Life Technologies), to give rise to pPro35S:MdCo31_K. Tomato plants were sown on germination medium (0.5x MS, 10% sucrose, 0.8% agar, pH=5.8).

Cotyledon pieces of the tomato seedlings were incubated with *Agrobacterium tumefaciens* strain GV3101 carrying the T-DNA binary vector pPro35S:MdCo31_K for 10-15 min and incubated for 48 h at 25 °C. Explants were transferred to GCF10 medium (1x MS, 108.7 mg/l Nitsch vitamins, 1.5 mg/l zeatine riboside and 0.2 mg/l IAA, 30% sucrose, 0.8% agar, pH=5.8) containing 300 mg/l timentin and 100 mg/l kanamycin and incubated at 25 °C for 3 weeks. After this period, the explants were transferred to fresh plates. Once formed, calli were excised and transferred to GCF11 medium (1x MS, 108.7 mg/l Nitsch vitamins, 1.9 mg/l zeatine riboside, 30% sucrose, 0.8% agar, pH=5.8) containing 300 mg/l timentin and 100 mg/l kanamycin. Once shoots were big enough, they were transferred to MS30B5 medium (1x MS, 112 mg/l vitamin B5, 30% sucrose, 0.8% agar, pH=5.8) containing 100 mg/l kanamycin. Plants with roots were transferred to soil and grown in a growth chamber under the same conditions as the Arabidopsis plants.

PCR

PCR amplifications were performed in a volume of 20 µl in 1x PCR buffer containing 1 µl of DNA template, 0.2 mM dNTPs, 3mM $MgCl_2$, 0.5 units of 5 PRIME Taq DNA polymerase and 0.1 mM of each primer. The cycling conditions included an initial denaturation step of 5 min at 95 °C followed by 35 cycles of 30 s. at 95 °C, 1 min at 60 °C, 1 min at 72 °C and a final extension phase of 5 min at 72 °C. PCR cycling was carried out using an Applied Biosystems 2720 Thermal Cycler.

The primers used for amplification of *MdCo31* were MdCo31f and MdCo31r (chapter 3). LeActinF (5'-CCAAAAGCCAATCGAGAGAA-3') and

LeActinR (5'-GGTACCACCACTGAGGACGA-3') were used to amplify the house keeping gene Actin from tomato.

Phylogenetic analysis

Tomato genes coding for putative 2OGDs were identified through phytozome (Goodstein et al. 2011) and Arabidopsis 2OGDs were taken from the PLAZA 2.5 database (Van Bel et al. 2012). Amino acid sequences were aligned by ClustalW (Larkin et al. 2007) and phylogenetic trees were constructed in MEGA (Tamura et al. 2011) using the neighbour joining method (Saitou and Nei 1987). A bootstrap test with 1000 replicates was performed to calculate confidence levels of the tree.

Flavonoid profiling

Bud material from 'Wijcik' and 'McIntosh' apple trees was harvested in spring in 2013. Arabidopsis plants were grown in soil in a growth chamber at 22 °C with a long photoperiod (16 h of 100 µmol m^{-2} s^{-1} light, 8 h of dark) and the complete rosettes were harvested for the analysis. Around 100 mg of plant tissue was used for both the apple and Arabidopsis samples, which was ground to a fine powder in liquid nitrogen using a pestle and mortar. Profiling of phenolic compounds was carried out using the method described by Vrhovsek et al. (Vrhovsek et al. 2012).

Auxin assay

Arabidopsis seeds were sterilised by incubating the seeds for 2 min in 70% ethanol, followed by an incubation of 10 min in bleach solution (2-3% sodium hypochlorite and 5 µL Tween 20 in 10 ml of H_2O). After this incubation step, the seeds were washed 5 times in water and resuspended in 0.1% agar solution.

The seeds were plated out on 0.5x MS medium (2.2 g MS, 0.5g MES, 0.1g myo-inositol and 10 g sucrose in 1 l of H_2O) containing 0.7% agar and different concentrations of active GA, luteolin, rutin, naringenin and 1-N-Naphthylphthalamic acid (NPA) and the plates were placed vertically in a growth chamber at 22 °C under long photoperiod. After 10 days, the plates were turned 90 degrees and the root length and root angle were measured.

GA treatment

Arabidopsis plants were grown under normal conditions. After the seeds had germinated, the plants were treated with active GA (~95%, ≥50% total GA_3, SIGMA-ALDRICH) by spraying the plants twice per week with a 100 µM GA solution. Control plants were sprayed with water. The internode length and final plant height of the GA treated and control plants were measured after flowering.

***E. coli* protein production and purification**

The pENTR/D-MdCo31 vector (Wolters et al. 2013) was recombined into pDEST15 (Life Technologies), introducing a glutathione S-transferase (GST) tag at the N-terminus of *MdCo31*, using the Gateway LR Clonase II kit (Life Technologies), resulting in pDEST15-MdCo31. pDEST15-MdCo31 was transformed to BL21 *E. coli* cells.

For protein production, BL21 was grown at 37 °C until an optical density of 0.3, after which protein expression was induced by addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to obtain a 1 mM IPTG concentration. After addition of IPTG, the culture was grown for 3 h at 28 °C.

Proteins were extracted from the *E. coli* pellets by resuspending the pellet of 500 µl of culture in 500 µl lysis buffer (50 mM KH_2PO_4, pH=7.8; 400 mM NaCl; 100 mM KCl; 10% glycerol; 0.5% Triton X-100; 10 mM imidazole) and by performing 3 subsequent freeze-thaw cycles in liquid nitrogen and a water bath at 42 °C.

Purification of *MdCo31* was performed making use of the GST tag. 100 µl of protein extract was incubated with 30 µl of GST beads for 3 h on a rotary shaker, after which the beads were spinned down and washed 3 times with PBS buffer (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 2.7 mM KCl, 140 mM NaCl, pH=7.3). Test samples were taken during the protein expression and purification experiment, which were dissolved in 2x Laemmli buffer (65.8 mM Tris-HCl, pH=6.8, 2.1% SDS, 26.3% (w/v) glycerol, 0.01% bromophenol blue) and loaded on 12% acrylamide gels for SDS PAGE, using the Mini-PROTEAN Tetra cell system (BIO-RAD).

Results

Effect of *MdCo31* in different plants

The phenotypes of young 'McIntosh' and 'Wijcik' plants are shown next to each other in Fig. 2a. The difference in branching is not so obvious yet, because young apple trees show little lateral branching in general, but the typical columnar phenotype is visible through the observed dwarfing phenotype. The 'Wijcik' trees have shorter internodes than the 'McIntosh' trees and leaves of 'Wijcik' trees are thicker and darker than leaves from 'McIntosh' trees.

A similar effect is observed in the Arabidopsis lines that constitutively express *MdCo31* (Pro35S:MdCo31). The phenotypes of the homozygous lines were more pronounced than the phenotypes of the hemizygous lines that were shown in chapter 3 and a clear difference with wildtype (WT) Arabidopsis is already evident at the rosette stage. Three plants of a representative transgenic line are compared to WT Arabidopsis in Fig. 2b. The transgenic lines showed a dwarf phenotype and have darker leaves than the WT plants. The Pro35S:MdCo31 plants are delayed in flowering and senescence (data not shown).

Tomato (cv. 'Moneymaker') plants transformed with a similar construct for overexpression of *MdCo31* also showed a compact phenotype and darker leaves, similar to the 'Wijcik' mutant and the transgenic Arabidopsis lines. Three independent transformants are displayed in Fig. 2c, which all contain the *MdCo31* gene (Fig. 2d), but that vary in their phenotype. All plants have a similar number of internodes (8-9 at the moment the picture displayed in Fig. 2c was taken), but the plants ranged in height from 12 cm (transgenic line #1) to 30 cm (WT). The difference in plant height was caused by a different internode length (Fig. 2e), similar to the columnar trees and transgenic Arabidopsis lines constitutively expressing *MdCo31*. The fact that overexpressing the *MdCo31* gene from apple has a comparable effect in diverse species, suggests a conserved function of *MdCo31*.

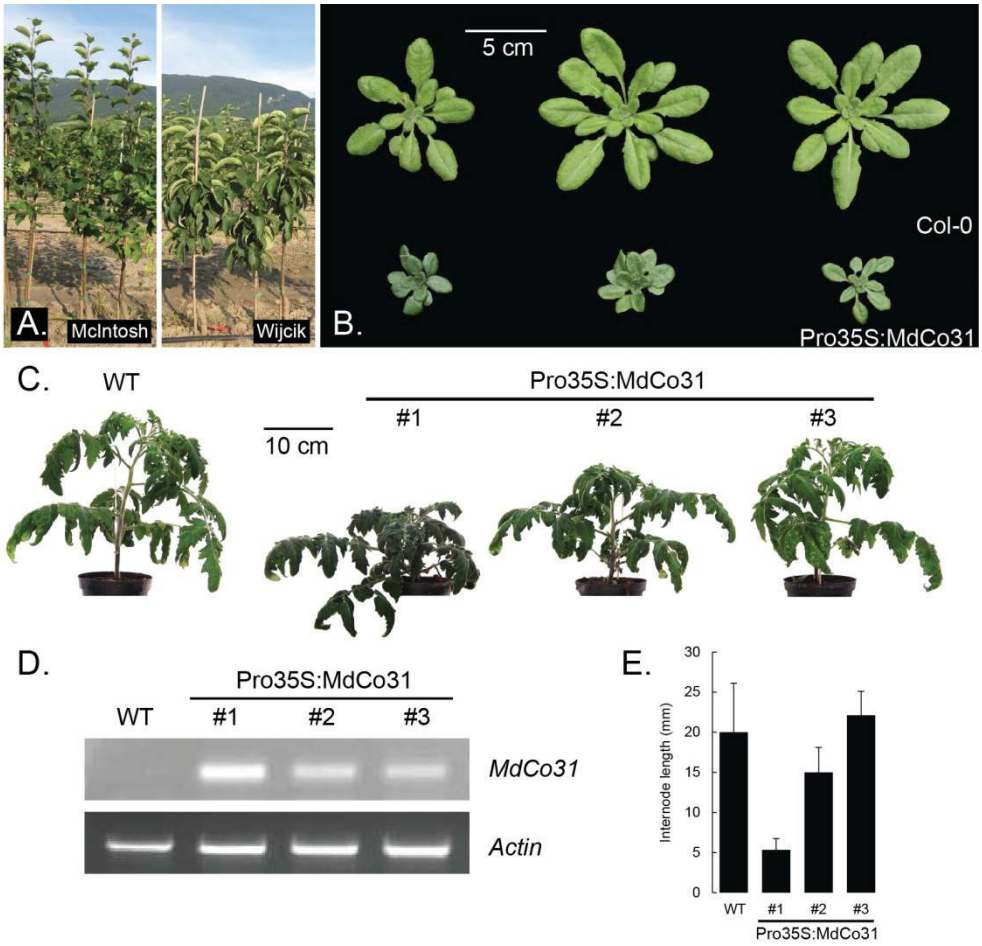

Figure 2. Effect of *MdCo31* expression on phenotypes of various plants. A. Young 'McIntosh' and 'Wijcik' trees. 'Wijcik' trees are dwarfed, have short internodes and darker leaves. **B.** Three plants of WT Arabidopsis (Col-0) and 3 plants of Pro35S:MdCo31 Arabidopsis are shown. The transgenic lines look dwarfed and have darker leaves than the WT plants. **C.** A similar effect is seen when WT Moneymaker tomato (left) is compared to three independent tomato lines constitutively expressing *MdCo31* (on the right). **D.** A PCR on genomic DNA of the plants shown in (C) shows that the transgenic lines contain the *MdCo31* gene (top). The gene coding for Actin is used as a PCR control (bottom). **E.** The compact phenotype of the transgenic tomato lines expressing *MdCo31* is explained by a difference in internode length.

Genes related to *MdCo31*

If *MdCo31* has such a conserved function, *MdCo31* might be involved in a basal process, making it likely that well conserved homologs of this gene are present in other plant genomes. A BLAST search for the *MdCo31* protein sequence in the Arabidopsis protein database (Rhee et al. 2003) or NCBI's refseq protein database (Pruitt et al. 2005) showed good matches for predicted genes with unknown functions from the refseq database, but did not result in the identification of a well characterised homolog. All good matches to *MdCo31* code for putative 2OGDs. The typical iron binding triad (HDH) (Roach et al. 1995) and the NyYPXCXXP motif that is thought to be involved in the binding of 2-oxoglutarate by 2OGDs (Xu et al. 1995) are conserved in *MdCo31* (Fig. 3)

.

>*MdCo31*

METLDQNLVSSWFDVQSVPQTFVHPPEKRPGNSIDVPPCKNIPVVDLGSRDRSHTIQQIS

KASQDFGFFQVFNHGVCKKLIDDIMSISKEFHKMPRKDKIIEGSKDPSGRCKFYTSSENY

ANEEVHYWRDALIHPAHSSENYMQFLPQKPTQYREVFKAYVDEVRNMGSMILEMLAEGLG

LSKEFFNGGLSENPTLLSNHYPPCPDPTLTLGLTKHRDPSLITILLQDSEGLQVFKDGNW

IGVEPISSGFVVNIGYVMQMISNSKFKGADHRVVTNSRAARTTIAYFIYPSNETVIEPAN

VLCNPPLYRSMKFTEFLQHFKSKAANDEEMSKVLSLSSS

Figure 3. Amino acid sequence of *MdCo31*. The location of the motif that is specific for DMR6/DLO is highlighted. This motif is not fully conserved in *MdCo31*. The NyYPXCXXP motif, that is thought to be involved in the binding of 2-oxoglutarate, is indicated with '+' and the iron binding triad (HDH) is indicated with asterisks.

In an attempt to better understand the function of *MdCo31*, a phylogenetic analysis was carried out using the predicted amino acid sequences of *MdCo31* and all 2OGD genes from *Arabidopsis* (90 genes according to the Plaza 2.5 predictions (Van Bel et al. 2012)) and tomato (169 genes, taken from the phytozome database (Goodstein et al. 2011)).

The phylogenetic tree displayed in Fig. 4 shows clear clustering of tomato and Arabidopsis 2OGD genes in clades of well characterised genes such as *FLS*,

LDOX, ethylene forming enzymes and the different GA oxidases. In addition to these clades that contain genes from both tomato and Arabidopsis, clades are visible that contain genes from a single species only. The genes from these species-specific clades may have more specialised functions. The *MdCo31* clade is an example of a clade that contains tomato genes only. It groups closely to another clade that includes genes from tomato together with downy mildew resistant 6 (*DMR6*) and the *DMR6*-like oxygenases (*DLO*s) from Arabidopsis that were described in the thesis of Zeilmaker (Zeilmaker 2012). *DLO1* has later been characterised as a salicylic acid (SA) 3-hydroxylase (*S3H*) (Zhang et al. 2013). The *DMR6*/*DLO* motif (WRD(F/Y)LR), which is conserved in all members of the *DMR6*/*DLO* protein family and which can be used to identify homologous genes with the same function of *DMR6*/*DLO* (Zeilmaker 2012), is not conserved in *MdCo31* (Fig. 3).

When comparing the results of this phylogenetic analysis with the study that was published recently by Kawai et al. in *the Plant Journal*, a similar picture emerges. The names of clades that were described by Kawai et al. that contain functionally characterised genes are indicated in Fig. 4. *MdCo31* falls in the DOXC41 clade from this classification, that contains hyoscyamine 6-hydroxylase (*H6H*) and other, functionally diverse, 2OGD genes that are involved in specialised metabolisms (Kawai et al. 2014). The closest related clade containing Arabidopsis genes that was described in this study is DOXC38 and contains *DMR6* and the *DLO*s (Kawai et al. 2014). *Solyc11g010400* and *Solyc11g0104010*, that cluster in the same clade as *MdCo31* in Fig. 4, are the closest homologs of the *H6H* gene that was originally cloned from *Hyoscyamus niger* (Matsuda et al. 1991) in tomato.

These phylogenetic analyses give some interesting insights in the organisation of 2OGDs in different species, but the function of *MdCo31* remains unknown. It is possible that *MdCo31* has a novel function that has not been characterised yet in other species. Although *MdCo31* is not clearly clustering with the 2OGDs involved in biosynthesis of GAs or flavonoids, a role for *MdCo31* in these pathways would be the easiest explanation for the phenotype of columnar trees derived from ‘Wijcik’.

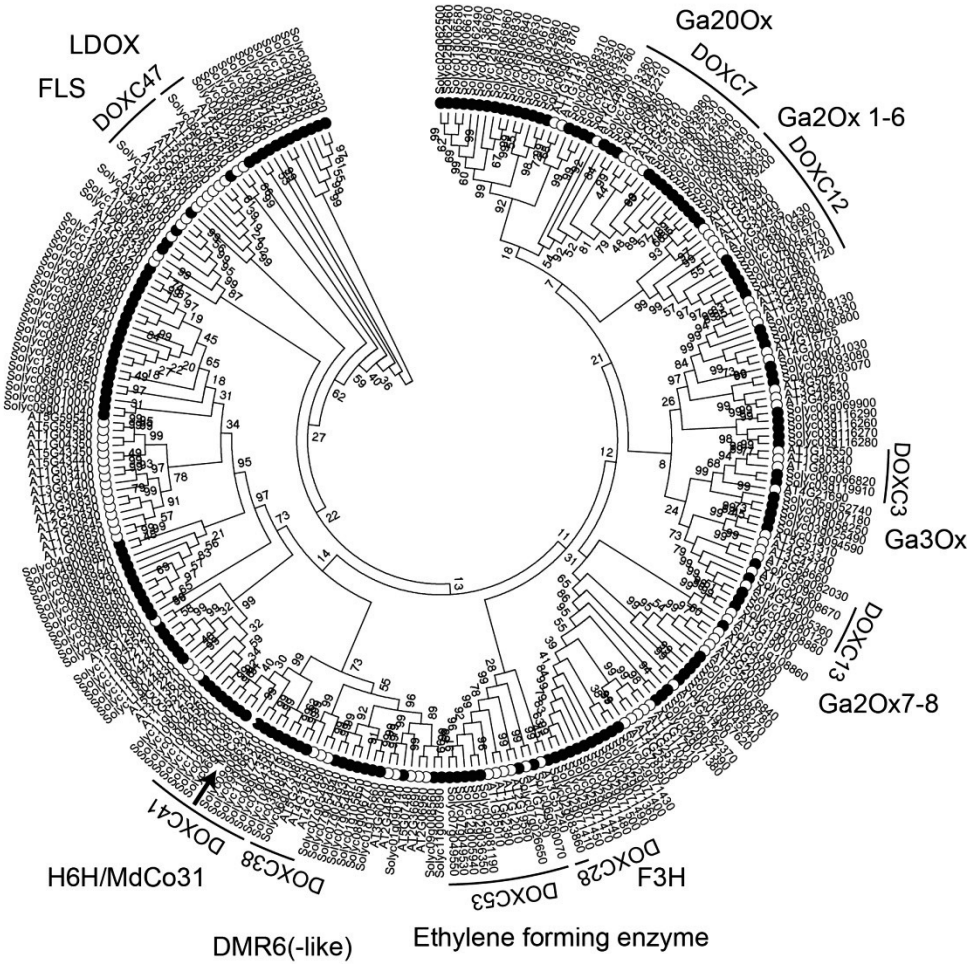

Figure 4. Phylogenetic tree containing *MdCo31* and all 2OGDs from Arabidopsis and tomato. The amino acid sequences were aligned using clustalW and the tree was constructed in MEGA using the Neighbor joining method. Bootstrap values are calculated bases on 1000 bootstraps. Tomato genes are marked with a filled circle and Arabidopsis genes are indicated by an empty circle. *MdCo31* is marked with an arrow. The clades containing *DMR6*(-likes), ethylene forming enzymes, *F3H*, *FLS*, *LDOX*, GA2-oxidases 1-6, GA2-oxidases 7-8, GA3-oxidases and GA20-oxidases are also indicated in the figure.

MdCo31 and flavonoid biosynthesis

In order to compare the flavonoid content between columnar and standard trees, a profiling of phenolic compounds was carried out on spur buds of standard and columnar apple trees. The bud material was harvested in spring and analysed according to the method published by Vrhovsek et al. (2012). The levels of the phenolic compounds that could be detected are summarised in Fig. 5. There is some variation between the samples, as only three replicates were used per sample, but the rutin levels seem to be a lot higher in columnar trees. A similar analysis was performed on tissue from the complete rosettes of 5-week old WT Arabidopsis plants and Pro35S:MdCo31 Arabidopsis plants (Fig. 6). No rutin was detected in the Arabidopsis assay, but luteolin levels were significantly increased in the Pro35S:MdCo31 plants ($P<0.01$) instead. Additionally, elevated levels of vanillic acid were found in Pro35S:MdCo31 plants ($P<0.01$).

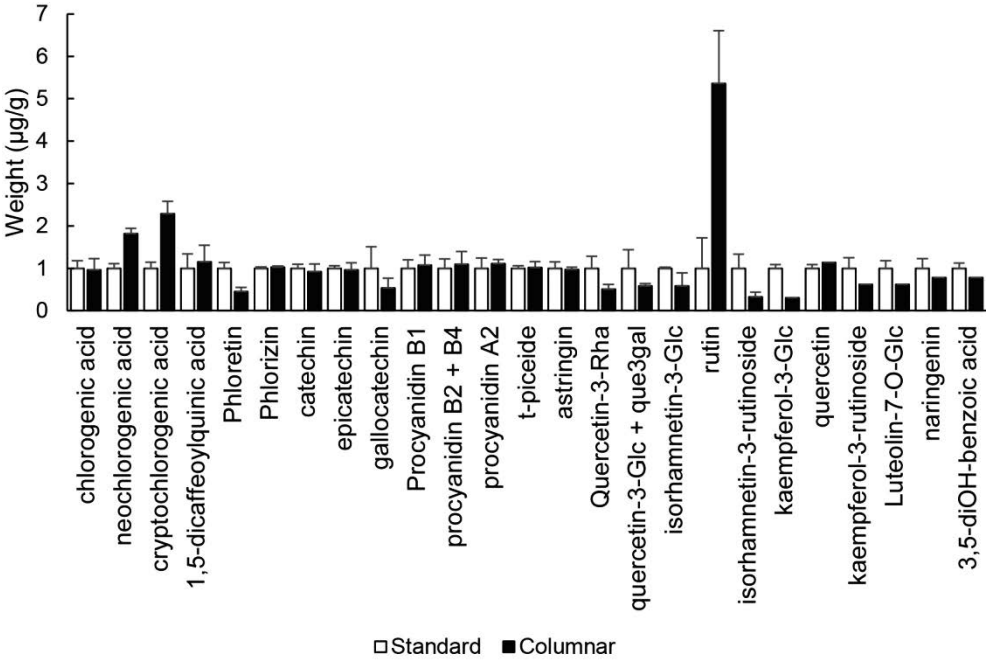

Figure 5. Profiling of phenolic compounds in apple. A phenolic compound profiling experiment on 3 bud samples of standard progeny trees (white bars) and 3 bud samples of columnar progeny trees from a segregating population revealed an increased level of rutin in columnar buds. Error bars indicate standard deviations of the mean.

As hypothesised before, if flavonoids are indeed responsible for the columnar phenotype of 'Wijcik', this would probably be through their effect on auxin transport. In order to see if rutin and luteolin can affect auxin transport, a gravitropism assay was performed on seedlings of WT Arabidopsis and Pro35S:MdCo31 Arabidopsis lines. Arabidopsis seedlings were grown *in vitro*, on media containing different concentrations of luteolin or rutin (the flavonoid compounds for which elevated levels were found in columnar trees and Arabidopsis plants constitutively expressing *MdCo31*) and on plates containing the auxin transport inhibitor NPA, or naringenin (a precursor for flavonoids that can inhibit auxin transport (Brown et al. 2001)). As auxin transport is necessary for a proper gravity response, the effect of these compounds on auxin transport was investigated through measuring the root angle after turning the assay plates 90 degrees. Additionally, root length was measured.

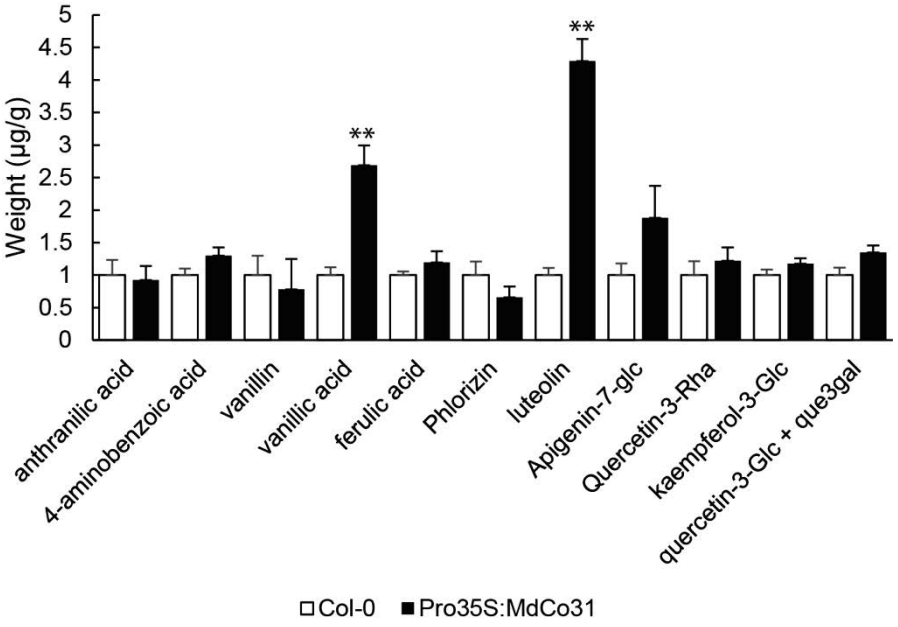

Figure 6. Profiling of phenolic compounds in Arabidopsis. The graph shows the result of a phenolic compound profiling experiment on 15 samples of Col-0 (white bars) and 15 samples of a transgenic Pro35S:MdCo31 line. The levels of vanillic acid and luteolin are significantly increased in the Arabidopsis plants constitutively expressing *MdCo31* (Student's *t* test, P<0.01, marked with asterisks). Error bars indicate standard deviations of the means.

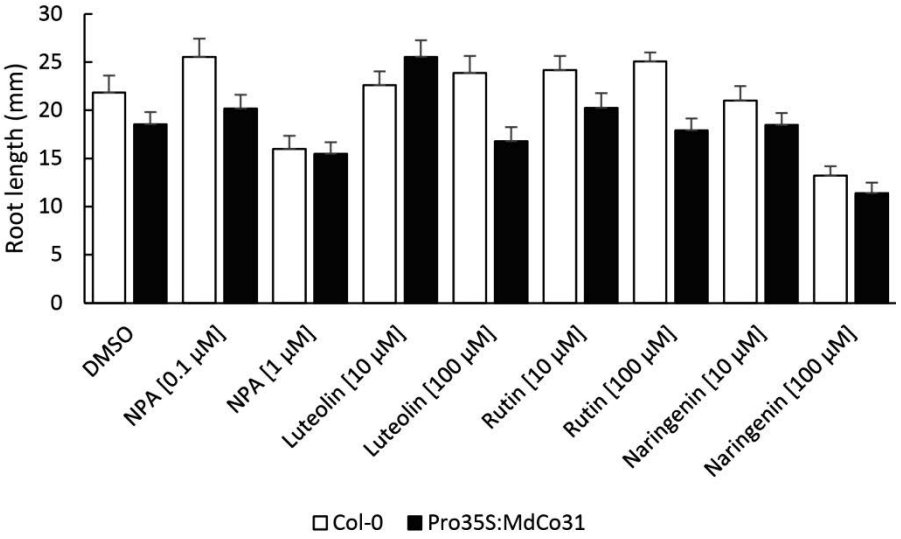

Figure 7. Root length of Arabidopsis seedlings. WT and Pro35S:MdCo31 Arabidopsis seedlings are grown on 0.5x MS medium containing different concentrations of 1-N-Naphthylphthalamic acid (NPA), luteolin, rutin and naringenin. Root length is measured after 10 days. Higher combinations of NPA and naringenin in the growth medium negatively affect the root length, but luteolin and rutin do not seem to affect root length. Pro35S:MdCo31 plants have shorter roots than WT plants in most cases. Error bars indicate errors of the mean.

Increased concentrations of NPA and naringenin negatively affected the root length of the Arabidopsis seedlings (Fig. 7), as would be expected when auxin transport is inhibited. The effect on the Pro35S:MdCo31 lines seems slightly larger. No effect of luteolin or rutin on root length is observed. A similar effect is observed for root angle (Fig. 8). The response to gravity is impaired in the seedlings that were grown on plates containing 1 µM NPA and 100 µM naringenin, but not in seedlings that were grown on media containing luteolin or rutin. The root length of Pro35S:MdCo31 lines is smaller than the root length of WT Arabidopsis in most cases, but there does not seem to be a clear difference in root angle between WT Arabidopsis and Arabidopsis plants constitutively expressing *MdCo31*.

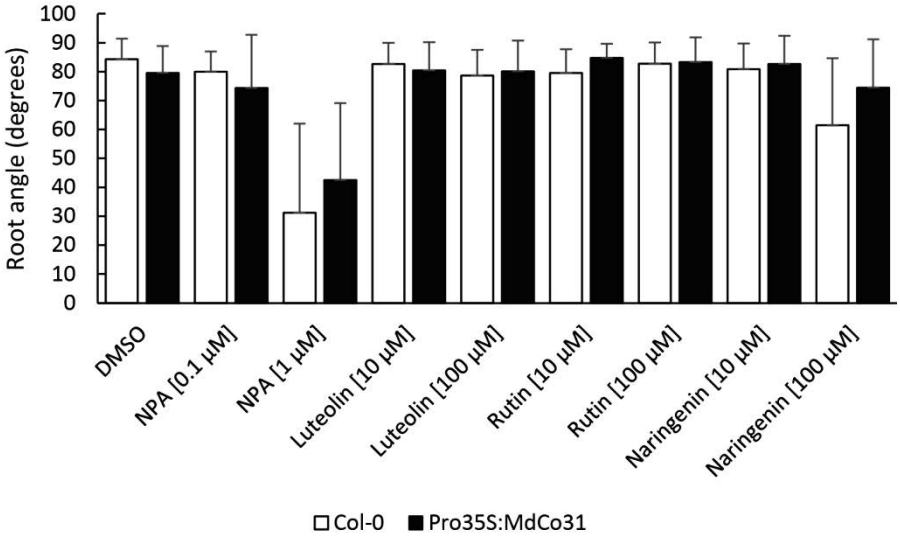

Figure 8. Gravity response of Arabidopsis seedlings. Col-0 and Pro35S:MdCo31 Arabidopsis seedlings are grown on 0.5x MS medium containing different concentrations of 1-N-Naphthylphthalamic acid (NPA), luteolin, rutin and naringenin on square petri dishes that are placed vertically. After 10 days, the plates are turned 90 degrees. The gravity response of the roots (measured by the angle of the root to the original growth direction) is measured after 3 days. NPA, and to a lesser extent naringenin, diminish the gravity response of the Arabidopsis seedlings, but no clear difference between Col-0 and Pro35S:MdCo31 Arabidopsis plants is observed. Error bars indicate errors of the mean.

MdCo31 and GA biosynthesis

If *MdCo31* is involved in GA biosynthesis, the most likely explanation for the columnar phenotype would be a shortage of active GAs. WT Arabidopsis plants and Pro35S:MdCo31 Arabidopsis plants were sprayed with a solution of 100 µM active GA twice per week throughout their development and compared to untreated plants.

The experiment shows that GA treated plants are almost indistinguishable from untreated WT plants. The untreated Pro35S:MdCo31 still show a compact phenotype and look very different from WT (Fig. 9). These results suggest that there is a GA deficiency in the Pro35S:MdCo31 plants that can be solved by GA treatments.

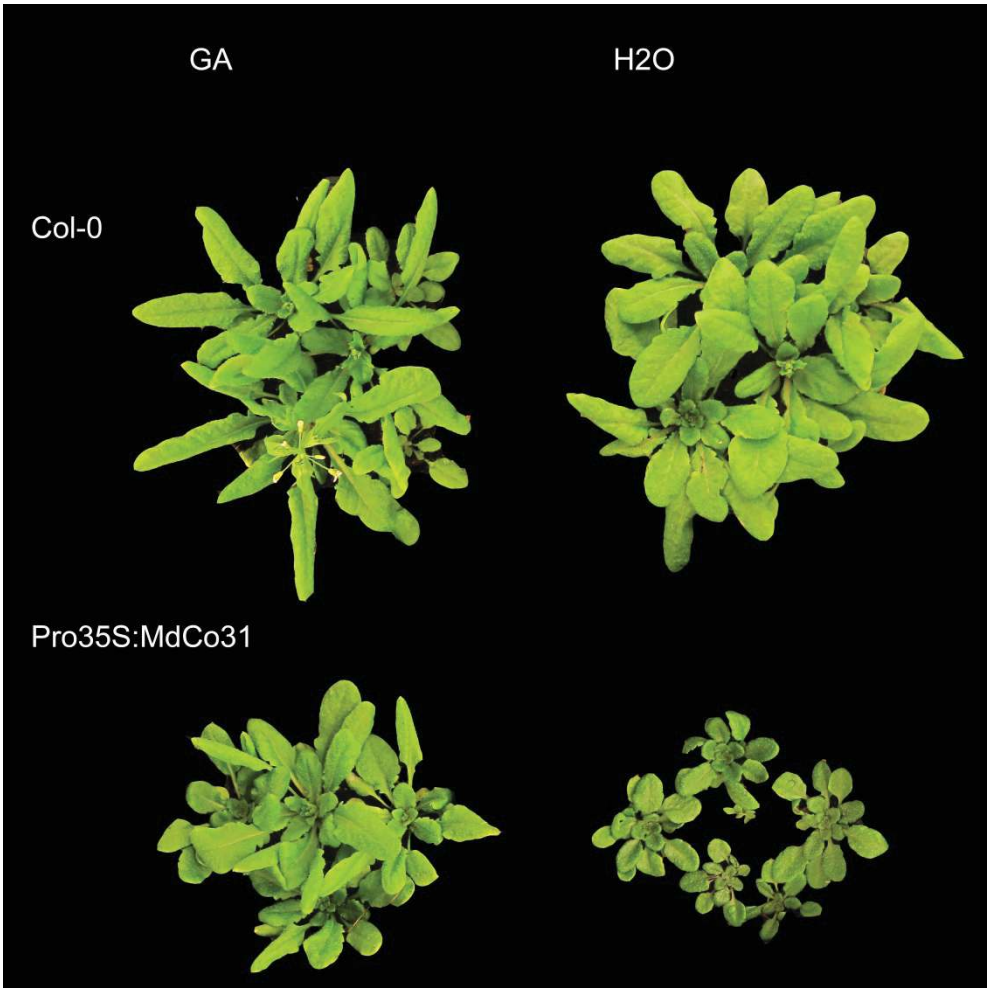

Figure 9. GA treatment on Col-0 and Pro35S:MdCo31 plants. Col-0 (top) and Pro35S:MdCo31 (bottom) plants are treated with 100 µM gibberellin solution (GA, left) or water (H_2O, right). Untreated Pro35S:MdCo31 plants look dwarfed compared to untreated Col-0 plants. GA treatment seems to slightly accelerate flowering in WT-Arabidopsis, but restores the phenotype of the Pro35S:MdCo31 plants to normal.

Producing the *MdCo31* protein

Both the flavonoid profiling experiment and the GA treatment on Arabidopsis plants give some support to the hypothesis that GAs or flavonoids could be involved in the columnar phenotype of 'Wijcik' apple. More evidence is needed to establish the function of *MdCo31*.

Availability of the MdCo31 protein would allow us to test the activity of the protein on different substrates and to identify the converted products. To achieve

this goal, *MdCo31* was cloned in the pD15 vector, introducing an N-terminal GST tag to the protein. The construct was transformed to *E. coli* and a protein purification experiment was performed to see if the protein could be produced. The first results show that MdCo31 protein is produced upon IPTG induction and that it is possible to purify the protein using glutathione agarose beads (Fig. 10). We have not been able yet to confirm a function for *MdCo31* in the biosynthesis of flavonoids, GAs, or other compounds, but these experiments are planned in the future.

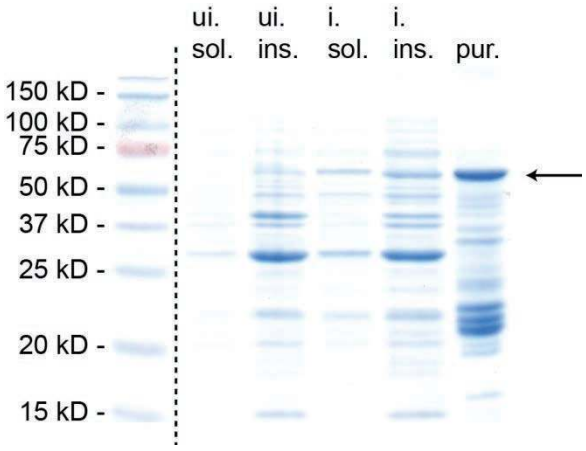

Figure 10. *MdCo31* protein expression and purification in *E. coli*. The marker lane and the sizes of the bands are indicated on the left. In the first lane, the soluble (sol.) proteins from an uninduced (ui.) culture are loaded. The next lane shows the insoluble (ins.) protein fraction of an uninduced culture. Then, the soluble fraction of an induced (i.) culture is loaded, followed by the insoluble fraction of the same culture. Finally, the proteins resulting from the protein purification experiment using GST agarose beads are loaded (pur.). Upon induction with IPTG, a band appears around 60 kilodalton (kDa), matching the size of MdCo31 (38.4 kDa) fused to the GST-tag (26 kDa). The GST purification step results in an increased intensity of this band (indicated with an arrow), demonstrating that MdCo31 is correctly produced in *E. coli*.

Discussion

MdCo31 similarly affects the development of diverse plant species, suggesting a role for this gene in a conserved process. The phylogenetic analysis performed in this chapter contradicts this supposed conserved function of *MdCo31*, placing it in a clade that contains genes with supposed specialised functions. One of the genes clustering with *MdCo31* is *H6H*, that acts in the biosynthesis pathway for tropane alkaloids in *Solanaceae*, via hydroxylation of hyoscyamine to produce

scopolamine (Matsuda et al. 1991). Hyoscyamine and scopolamine are tropane alkaloids specific for *Solanaceae* (Griffin and Lin 2000), making it unlikely that the *MdCo31* gene from apple is involved in the biosynthesis pathway of such compounds. *MdCo31* might have a similar origin as *H6H*, but must have adopted a different function in apple.

The closest clade containing 2OGDs that are conserved between species includes *DMR6* and *DLO1/2* from Arabidopsis. *DMR6* is responsible for susceptibility of Arabidopsis to downy mildew (van Damme et al. 2008), through a salicylic acid dependent mechanism (Zeilmaker 2012). *DLO1* and *DLO2* can complement the *dmr6* Arabidopsis mutant. The substrate for *DMR6* is unknown, but it was shown that *DLO1*, or called *S3H* in that study, is associated to senescence and involved in the catabolism of salicylic acid (Zhang et al. 2013). Functional homologs of *DMR6* were identified in several crop species, including tomato, which can complement *DMR6*-mediated resistance in the Arabidopsis *dmr6* mutants. (Zeilmaker 2012).

The delayed senescence phenotype and the increased chlorophyll content in the Arabidopsis plants overexpressing *S3H* (Zhang et al. 2013) show some interesting similarity to the Arabidopsis plants overexpressing *MdCo31*. However, no effect of *S3H* on plant architecture was reported by Zhang et al. (2013) and van Damme et al. also state that plant development is not affected in *dmr6* Arabidopsis mutants (van Damme et al. 2008). Zeilmaker proposed a 6-amino acid motif in his thesis, that shows high conservation among homologs of *DMR6* and *DLOs* in other species and that could be used to identify genes with a similar function (Zeilmaker 2012). This 6-amino acid motif is not well conserved in the predicted MdCo31 protein. Together with the reported role of *DMR6* in the loss of downy mildew resistance in Arabidopsis plants, these data suggest a different function for *MdCo31*.

There are multiple possible explanations for the dark colour of 'Wijcik' and of the tomato and Arabidopsis plants constitutively expressing *MdCo31*. It is possible that the darker leaf colour is caused by an increased chlorophyll content, like was found for the Arabidopsis plants overexpressing *S3H* (Zhang et al. 2013). If decreased cell expansion is responsible for the compact phenotype of the plants

that constitutively express *MdCo31,* this might result in increased concentrations of chlorophyll in the cells. Alternatively, the levels of anthocyanins, that are produced as part of the flavonoid pathway (Cheng et al. 2014), could be elevated in tissues expressing *MdCo31*. Additionally, dark green foliage is commonly observed in gibberellin deficient plants (Sun and Kamiya 1994; Helliwell et al. 1998; Busov et al. 2003). Although dark foliage was observed in 'Wijcik' and columnar trees derived from 'Wijcik', we did not find elevated expression of *MdCo31* in the leaves of these trees.

No direct evidence for the function of *MdCo31* was provided in this study, but we showed that the increased expression of *MdCo31* is correlated to increased levels of specific flavonoids in apple and Arabidopsis. Higher levels of rutin were found in columnar trees than in standard trees, but rutin could not be detected in Arabidopsis. The levels of the flavone luteolin were about 4-fold increased in Pro35S:MdCo31 plants compared to WT. Rutin is produced from quercetin in the flavonol part of the flavonoid biosynthesis pathway, whereas luteolin is derived from apigenin in the flavone part of this pathway. Both branches depend on naringenin and eriodictyol at the start of the flavonoid biosynthesis pathway (Martens et al. 2010). The different genetic backgrounds in Arabidopsis and apple could be responsible for the fact that increased expression of *MdCo31* leads to higher levels of rutin in apple, but higher levels of luteolin in Arabidopsis.

In the introduction, we hypothesised that increased levels of flavonoids could result in a columnar phenotype through their inhibitory effect on auxin transport. The effect of the flavonoids rutin and luteolin on auxin transport was investigated in a gravitropism assay, using WT Arabidopsis seedlings and Arabidopsis seedlings that constitutively express *MdCo31*. No effect of these flavonoids on the gravity response of roots of these plants was demonstrated. Also, no difference between the gravity responses was observed when comparing WT Arabidopsis with Pro35S:MdCo31 plants, undermining the hypothesis that the effect of *MdCo31* on plant development is explained through inhibition of auxin transport.

Alternatively, the effect of *MdCo31* could be explained through a role in the GA biosynthesis pathway. The increased expression of a GA2-oxidase is

responsible for the columnar-like phenotype of *stumpy* poplar (Busov et al. 2003) and plants with a reduced response to GA show compact phenotypes (Peng et al. 1997; Peng et al. 1999; Boss and Thomas 2002). Deactivation of GAs is mainly mediated by GA2-oxidases. To test this hypothesis, WT Arabidopsis plants and Pro35S:MdCo31 Arabidopsis plants were treated with active GA. The results of this experiment strongly support the hypothesis that *MdCo31* is involved in the deactivation of GA. Untreated Pro35S:MdCo31 plants showed a clear dwarf phenotype, but plants that were treated with active GA were almost indistinguishable from WT Arabidopsis. However, the conserved domains that are normally present in GA2-oxidases (or other GA oxidases) are not present in the amino acid sequence of MdCo31. If *MdCo31* is indeed involved in GA biosynthesis, it has to belong to a previously uncharacterised class.

The GA experiment is no direct proof for a role of *MdCo31* in GA biosynthesis. It is possible that *MdCo31* has a different function, but that GA deficiency is a downstream effect of the action of *MdCo31*. To prove that *MdCo31* is involved in GA catabolism, the enzyme that was produced in *E. coli* could be purified and tested on GA substrates, similar to the assay that was performed by Giacomelli et al. to characterise GA oxidases from grape (Giacomelli et al. 2013). Such an assay could also be used to test other potential substrates of MdCo31 (for example flavonoid compounds).

Because the amino acid sequence of MdCo31 does not look like a typical GA oxidase, it is well possible that MdCo31 belongs to a different class of enzymes. To investigate alternative roles of *MdCo31*, it would be interesting to perform an untargeted metabolite profiling experiment on tissue from columnar and standard apple, or from the wildtype tomato or Arabidopsis plants and the Pro35S:MdCo31 transformants. Additionally, an RNA-seq or microarray experiment could be carried out, to look at transcriptomic changes that are due to *MdCo31* expression, to get more clues about the function of the gene.

A number of RNA-seq experiments have already been carried out by others in order to try to understand the cause for columnar growth in apple (Krost et al. 2012; Zhang et al. 2012; Krost et al. 2013), but these experiments were always based on the comparison between standard trees and columnar trees

derived from ‘Wijcik’ and not on the direct comparison between 'Wijcik' and 'McIntosh'. These studies have not resulted in the identification of the cause for columnar growth, but showed that several hormones are likely involved in the phenotype of columnar trees. It is not surprising to find differences for multiple hormones. Because hormones are part of complex and interconnected pathways, changes in the level of a certain hormone will likely affect the level of other hormones as well.

To identify transcriptomic changes that are solely due to the mutation in ‘Wijcik’, and not due to other genetic differences, and to identify subsequent downstream effects, it would be better to perform a direct comparison between ‘McIntosh’ and ‘Wijcik’. A transcriptome study would probably be easier to perform in Arabidopsis because of its small and well annotated genome. The Arabidopsis plants could be grown *in vitro* to reduce environmental variation. The analysis of Arabidopsis lines transformed with an alcohol inducible construct for expression of *MdCo31* is currently in progress. By using an inducible promoter, the expression of *MdCo31* can be induced shortly before taking samples for the analysis in order to look at direct transcriptomic changes of *MdCo31* expression and to reduce secondary effects. Further proof for the role of hormones or mobile signals identified in these studies could be provided by performing grafting experiments (Turnbull et al. 2002) or complementation studies in Arabidopsis.

Besides being responsible for columnar growth in columnar trees derived from 'Wijcik', *MdCo31* should have a biological function in standard apple trees as well. We were unable to find expression of *MdCo31* in diverse apple tissues at different time points, so the normal function of *MdCo31* remains unknown. It would be interesting to silence the *MdCo31* gene in normal apple trees to see if the phenotype of those trees can help understand the function of *MdCo31*.

Combining these different approaches it should be possible to unravel the mechanism that leads to columnar growth in apple.

References

Boss, P. K. and M. R. Thomas (2002). "Association of dwarfism and floral induction with a grape 'green revolution'mutation." Nature **416**(6883): 847-850.

Brown, D. E., A. M. Rashotte, et al. (2001). "Flavonoids act as negative regulators of auxin transport in vivo in arabidopsis." Plant Physiology **126**(2): 524-535.

Busov, V. B., R. Meilan, et al. (2003). "Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature." Plant Physiology **132**(3): 1283-1291.

Cheng, A.-X., X.-J. Han, et al. (2014). "The Function and Catalysis of 2-Oxoglutarate-Dependent Oxygenases Involved in Plant Flavonoid Biosynthesis." International journal of molecular sciences **15**(1): 1080-1095.

Dare, A. P., S. Tomes, et al. (2013). "Phenotypic changes associated with RNA interference silencing of chalcone synthase in apple (Malus× domestica)." The Plant Journal **74**(3): 398-410.

Gebhardt, Y., S. Witte, et al. (2005). "Molecular evolution of flavonoid dioxygenases in the family Apiaceae." Phytochemistry **66**(11): 1273-1284.

Giacomelli, L., O. Rota-Stabelli, et al. (2013). "Gibberellin metabolism in Vitis vinifera L. during bloom and fruit-set: functional characterization and evolution of grapevine gibberellin oxidases." Journal of Experimental Botany **64**(14): 4403-4419.

Goodstein, D. M., S. Shu, et al. (2011). "Phytozome: a comparative platform for green plant genomics." Nucleic Acids Res.

Griffin, W. J. and G. D. Lin (2000). "Chemotaxonomy and geographical distribution of tropane alkaloids." Phytochemistry **53**(6): 623-637.

Hedden, P. and S. G. Thomas (2012). "Gibberellin biosynthesis and its regulation." Biochemical Journal **444**(1): 11-25.

Helliwell, C. A., C. C. Sheldon, et al. (1998). "Cloning of the Arabidopsis ent-kaurene oxidase gene GA3." Proceedings of the National Academy of Sciences **95**(15): 9019-9024.

Karimi, M., D. Inze, et al. (2002). "GATEWAY vectors for Agrobacterium-mediated plant transformation." TRENDS in Plant Science **7**(5): 193-195.

Kawai, Y., E. Ono, et al. (2014). "Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants." The Plant Journal.

Krost, C., R. Petersen, et al. (2013). "Evaluation of the hormonal state of columnar apple trees (Malus x domestica) based on high throughput gene expression studies." Plant Mol Biol **81**(3): 211-220.

Krost, C., R. Petersen, et al. (2012). "The transcriptomes of columnar and standard type apple trees (Malus x domestica) - A comparative study." Gene **498**(2): 223-230.

Larkin, M., G. Blackshields, et al. (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Lee, Y. J., J. H. Kim, et al. (2008). "Characterization of flavone synthase I from rice." BMB Rep **41**(1): 68-71.

Magome, H., T. Nomura, et al. (2013). "CYP714B1 and CYP714B2 encode gibberellin 13-oxidases that reduce gibberellin activity in rice." Proceedings of the National Academy of Sciences **110**(5): 1947-1952.

Martens, S., A. Preuß, et al. (2010). "Multifunctional flavonoid dioxygenases: Flavonol and anthocyanin biosynthesis in *Arabidopsis thaliana*." Phytochemistry **71**(10): 1040-1049.

Matsuda, J., S. Okabe, et al. (1991). "Molecular cloning of hyoscyamine 6 beta-hydroxylase, a 2-oxoglutarate-dependent dioxygenase, from cultured roots of Hyoscyamus niger." Journal of Biological Chemistry **266**(15): 9460-9464.

Ferreyra, M. L. F., S. P. Rius, et al. (2011). "Flavonoids: biosynthesis, biological functions, and biotechnological applications." Frontiers in plant science **3**: 222-222.

Peer, W. A. and A. S. Murphy (2007). "Flavonoids and auxin transport: modulators or regulators?" TRENDS in Plant Science **12**(12): 556-563.

Peng, J., P. Carol, et al. (1997). "The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses." Genes & Development **11**(23): 3194-3205.

Peng, J., D. E. Richards, et al. (1999). "'Green revolution' genes encode mutant gibberellin response modulators." Nature **400**(6741): 256-261.

Petersen, R. and C. Krost (2013). "Tracing a key player in the regulation of plant architecture: The columnar growth habit of apple trees (Malus x domestica)." Planta **238**(1): 1-22.

Prescott, A. G. and P. John (1996). "DIOXYGENASES: Molecular structure and role in plant metabolism." Annual Review of Plant Physiology and Plant Molecular Biology **47**: 245-271.

Pruitt, K. D., T. Tatusova, et al. (2005). "NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins." Nucleic Acids Res **33**(Database issue): D501-504.

Prusinkiewicz, P., S. Crawford, et al. (2009). "Control of bud activation by an auxin transport switch." Proceedings of the National Academy of Sciences **106**(41): 17431-17436.

Rhee, S. Y., W. Beavis, et al. (2003). "The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community." Nucleic Acids Research **31**(1): 224-228.

Roach, P. L., I. J. Clifton, et al. (1995). "Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes." Nature **375**(6533): 700-704.

Saitou, N. and M. Nei (1987). "The neighbor-joining method: a new method for reconstructing phylogenetic trees." Molecular biology and evolution **4**(4): 406-425.

Sun, T. and Y. Kamiya (1994). "The Arabidopsis GA1 locus encodes the cyclase ent-kaurene synthetase A of gibberellin biosynthesis." The Plant Cell Online **6**(10): 1509-1518.

Tamura, K., D. Peterson, et al. (2011). "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods." Molecular biology and evolution **28**(10): 2731-2739.

Turnbull, C. G., J. P. Booker, et al. (2002). "Micrografting techniques for testing long-distance signalling in Arabidopsis." The Plant Journal **32**(2): 255-262.

Van Bel, M., S. Proost, et al. (2012). "Dissecting plant genomes with the PLAZA comparative genomics platform." Plant Physiology **158**(2): 590-600.

van Damme, M., R. P. Huibers, et al. (2008). "Arabidopsis DMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew." Plant J **54**(5): 785-793.

Vrhovsek, U., D. Masuero, et al. (2012). "A versatile targeted metabolomics method for the rapid quantification of multiple classes of phenolics in fruits and beverages." Journal of agricultural and food chemistry **60**(36): 8831-8840.

Wolters, P. J., H. J. Schouten, et al. (2013). "Evidence for regulation of columnar habit in apple by a putative 2OG-Fe(II) oxygenase." New Phytologist **200**(4): 993-999.

Xu, Y. L., L. Li, et al. (1995). "The GA5 locus of Arabidopsis thaliana encodes a multifunctional gibberellin 20-oxidase: molecular cloning and functional expression." Proceedings of the National Academy of Sciences **92**(14): 6640-6644.

Yamaguchi, S. (2008). "Gibberellin metabolism and its regulation." Annu. Rev. Plant Biol. **59**: 225-251.

Zeilmaker, T. (2012). Functional and applied aspects of the DOWNY MILDEW RESISTANT 1 and 6 genes in Arabidopsis.

Zhang, K., R. Halitschke, et al. (2013). "Salicylic acid 3-hydroxylase regulates Arabidopsis leaf longevity by mediating salicylic acid catabolism." Proc Natl Acad Sci U S A **110**(36): 14807-14812.

Zhang, Y., J. Zhu, et al. (2012). "Characterization of transcriptional differences between columnar and standard apple trees using RNA-Seq." Plant Molecular Biology Reporter **30**(4): 957-965.

Chapter 6

General discussion

Columnar growth in apple

The columnar trait, which was discovered in a mutant of the apple cv. 'McIntosh' called 'Wijcik', has many properties that make it an interesting study topic. When the 'Wijcik' mutant was first discovered in 1964, initial studies discussed the use for columnar trees in commercial growing systems and aimed to characterise the 'Wijcik' mutant (Fisher 1969; Looney and Lane 1984; Tobutt 1985). These studies recognised the very sturdy and compact growth of 'Wijcik' and observed that spurs (that bear fruits) are often formed instead of branches. The 'Wijcik' mutant was regarded to be a promising genotype for improving the efficiency and yield of apple orchards in high density planting systems (Tobutt 1985).

Later studies analysed progeny populations, derived from 'Wijcik', for the development of molecular markers and to study the heritability of the columnar trait (Hemmat et al. 1997; Kim et al. 2003; Tian et al. 2005; Silfverberg-Dilworth et al. 2006; Zhu et al. 2007; Moriya et al. 2009). These studies showed that columnar growth is caused by a single, dominant locus that is mapping on chromosome 10 of apple. The columnar trait is a stable trait, as no revertants of columnar trees have been identified until today. It seems likely that the 'Wijcik' mutation is genetic mutation rather than an epigenetic mutation.

All these properties make the *Co* trait a remarkable trait, with a high potential in apple breeding programs. The molecular markers that are available for the *Co* locus could be used to specifically target the columnar trait, while helping to avoid the introduction of undesirable alleles when breeding for columnar trees. Despite their potential benefits, no commercial columnar apple trees have been developed yet. Therefore, it still remains to be seen if columnar trees can meet up with the high expectations on the scale of a commercial apple orchard.

Columnar growth is not only interesting from a commercial point of view, it is also surprising to see such a dramatic effect from a somatic mutation. Most mutations lead to a loss of gene function and result in recessive inheritance. The dominant character of the columnar trait suggests a gain-of-function mutation in 'Wijcik'. Such gain of function mutations are often caused by increased expression of a gene, like was found for the *MYB10* gene, that was due to a rearrangement in

its upstream regulatory region and that resulted in red-fleshed apple (Espley et al. 2009). In 'Wijcik', we expected to find a similar mutation in a regulatory region that would lead to increased expression of the gene responsible for columnar growth in apple.

Fine mapping *Co*

The 'Golden Delicious' genome, which was published in 2010 (Velasco et al. 2010), has been a great resource to speed up the search for the molecular basis for columnar growth. The results that are presented in this study highlight the relevance of genome sequencing projects in studies of agronomically important traits in non-model species. Because the development of molecular markers was greatly facilitated after the 'Golden Delicious' genome became available, it became possible to saturate the *Co* region with markers and to screen large apple populations segregating for the columnar trait, in order to reduce the *Co* region. This resulted in three fine mapping studies, which were published shortly after each other (Bai et al. 2012; Moriya et al. 2012; Baldi et al. 2013).

Figure 1. Overview of the different regions in the fine mapping studies published by Baldi et al. (2013), Moriya et al. (2013) and Bai et al. (2012). A region of 700 kb (from 18500 - 19200 kb) of chromosome 10 from 'Golden Delicious' is shown at the top of the figure, together with the positions of the markers that delimit the different *Co* regions in these studies. The corresponding *Co* regions are displayed below. The location of marker C7629-22009 (italics) is an estimation, because this marker was found in an unanchored contig of the 'Golden Delicious' genome assembly.

The three different fine mapping studies point to a narrow location for the *Co* region on chromosome 10, but the different regions are not exactly coinciding. The region that was described by Baldi et al. (2013) overlaps with the majority of the region that was described by Moriya et al. (2012), but the region that is

described by Bai et al. (2012) lies just outside that region. The results of these different studies are compared in Fig. 1.

Bai et al. (2012) used four segregating populations and three collections of columnar trees for their analysis. These columnar collections include trees with diverse genetic backgrounds that are not directly descendant from 'Wijcik', but have inherited the columnar trait from 'Wijcik' in crosses performed up to two generations earlier. The trees that were analysed vary in age and are located at different geographical locations in the US, UK and China. Many different populations were used in the study that was published by Moriya et al. (2012).

All the trees used in our analysis had a clear genetic background, as they were all direct descendants of 'Wijcik'. We made use of three older populations for initial mapping and two large progeny populations for the fine mapping experiment. Especially young plants can be difficult to phenotype (plants with intermediate phenotypes are common in young populations), but the phenotypes of the older progeny trees are very reliable. We tested the markers that were used by Bai et al in their study on the older segregating populations that were developed by us (that had very clear phenotypes) and found three different recombinants that support our previous findings and that contradict the findings of Bai et al. We conclude that the results published by us and Moriya et al. are the most reliable, a view that is shared by Otto et al. in their publication studying the difference between 'McIntosh' and 'Wijcik' (Otto et al. 2013).

Candidate Co genes

In chapter 2 of this thesis, we described a number of potential candidate genes for *Co*, which all code for transcription factors with putative roles in plant development. Other candidate genes have been proposed in the fine mapping study performed by Bai et al. and in the studies where the transcriptomes of columnar and standard apple were compared (Krost et al. 2012; Zhang et al. 2012; Krost et al. 2013). Zhang et al. highlighted the potential role for GRAS (the name is derived from the *GAI*, *RGA* and *SCR* gene members) transcription factors, which are involved in the response to gibberellic acid (GA), in columnar growth, whereas Krost et al. identified many differentially expressed genes, including several genes with roles in hormonal pathways. *MdCo31* was never marked as a candidate gene

in these studies. An explanation for this is the fact that *MdCo31* was not annotated in the consensus gene set of apple. Studies that only took these gene annotations in consideration (Bai et al. 2012; Krost et al. 2013) could therefore not have resulted in the identification of *MdCo31*. *MdCo31* was absent in the GenScan gene predictions presented in chapter 2 and annotated as a hyoscyamine 6-dioxygenase-like gene in the predictions from the Augustus gene prediction program in that chapter.

Identification of differences between 'McIntosh' and 'Wijcik'

Although the fine mapping studies have not resulted in the identification of the *Co* gene, they have been a great step forward in identifying the cause for columnar growth in apple. In order to identify the mutation that led to columnar growth of 'Wijcik', the most straightforward way is to make a direct comparison between 'McIntosh' and 'Wijcik'. Because the *Co* region was sufficiently narrowed down during the fine mapping studies, it became possible to make such a comparison. The markers that were developed in these studies are not only useful for application in MAS, they could also be used to screen BAC libraries prepared from genomic DNA of 'McIntosh' and 'Wijcik', to isolate and sequence the *Co* region of 'McIntosh' and 'Wijcik' and perform a direct sequence comparison. This sequence comparison was described in chapter 3 of this thesis.

The direct comparison of the *Co* region between 'McIntosh' and 'Wijcik' resulted in the identification of a single genomic difference, an insertion of 1956 bp in 'Wijcik'. Because the identification of this insertion does not directly explain the change in tree architecture of the 'Wijcik' mutant, the expression of genes predicted in a region of 50 kb around the insert site was compared between 'McIntosh' and 'Wijcik' in leaf and bud tissue. This expression analysis resulted in the identification of a putative 2-oxoglutarate-dependent dioxygenase (2OGD), *MdCo31*, whose expression is specifically up regulated in bud tissue of all columnar trees tested and not in buds of standard trees. The insertion of 'Wijcik' may be the direct cause for columnar growth in apple through increased expression of *MdCo31*, as presence of the insertion was found to be associated with increased expression of *MdCo31* and columnar growth in all progeny trees derived from 'Wijcik' that were tested. The observed increased expression of *MdCo31* supports our earlier

hypothesis that increased expression of a gene is the cause for the dominant inheritance of the columnar trait.

The 'Wijcik' insertion was initially described as a 1956 bp sequence in this thesis, which lacks the elements that are common for mobile elements. We compared our findings with the results that were reported later by Otto et al. (2013) and concluded that the 'Wijcik' insertion was assembled incorrectly by Wolters et al. (2013). The 1956 bp sequence corresponds to the long terminal repeat (LTR) and target site duplication (TSD) of a bigger retrotransposon with a complete length of 8.2 kb. The big length of the LTRs of this retrotransposon made it possible for the assembly to collapse on this LTR, explaining why the internal part of the retrotransposon was missing in the assembly and was also responsible for the PCR artefact that led to amplification of the LTR only. This explains why the size of the insertion was initially misinterpreted.

More detailed analysis of the 'Wijcik' mutation

Analysis of the 'Wijcik' insert region revealed that the 'Wijcik' retrotransposon inserted into the 5' LTR of another retrotransposon. Both elements look like non-autonomous versions of LTR retrotransposons because they contain the typical components that characterise such elements (primer binding site, polypurine tract, LTRs and TSD (Wicker et al. 2007)), but they do not contain functional ORFs. Transposable elements (TEs) are commonly present in plant genomes, and TE movement has led to many changes in plant gene expression and function (Lisch 2013), but such effects are often caused by TEs that insert close to, or even into, genes (Greene et al. 1994; Naito et al. 2009).

Although nested TEs are not uncommon (Kronmiller and Wise 2008), it is surprising to find that the insertion of a TE into another TE can lead to such dramatic effects as observed in the 'Wijcik' mutant. This observation shows once more that the classification of TEs as 'junk' DNA is highly out-dated.

Many similar copies of both retrotransposons are present in the 'Golden Delicious' genome, but we were unable to identify their autonomous partners, so it is still unclear how they are moving around in the apple genome. To be able to properly characterise the elements and to propose a correct name, it would be necessary to find out from which autonomous elements they are derived. Because

the autonomous partners were not found in the apple genome and as the 'Wijcik'-like retrotransposons were also often misassembled, a PCR based approach could be useful to identify these elements. Using one primer designed on the LTR and another primer designed on the conserved sequences of ORFs that would be expected to be present in autonomous partners, it might be possible to amplify and sequence the autonomous retrotransposons from which the 'Wijcik' and 'McIntosh' retrotransposons are derived.

It is not clear if the effect of the insertion on the expression of *MdCo31* is directly caused by the 'Wijcik' retrotransposon or a consequence of the disruption of the retrotransposon that was already present in the *Co* region in 'McIntosh'. Reporter assays, where different parts of the two retrotransposons are cloned in front of a reporter construct, could help to clarify the contribution of the two elements to the columnar phenotype of 'Wijcik'. This is highly relevant, as we were unable to find examples in literature of how a TE insertion into another TE can lead to such dramatic effects as observed in 'Wijcik'.

Because it can be difficult to work with long DNA fragments (the complete disrupted retrotransposon from 'Wijcik' is more than 10 kb in length), it could be convenient to employ a recombination system (Yu et al. 2000). BAC plasmids containing more than 100 kb of *Co* region, including the 'Wijcik' retrotransposon and *MdCo31*, are already available. Different modifications could be made to the insert site and *MdCo31* could be replaced with a reporter gene. BAC clones can be converted to binary vectors for transient expression assays (Takken et al. 2004). Modifying BAC clones with large inserts of *Co* region would be especially useful because it is not known if a larger genomic context around the insert site is required for the effect of the 'Wijcik' insertion. Such BAC clones could also be used for transformation experiments, to introduce the columnar trait to other apple cultivars or to study the apple *Co* region in a model species. It is important to realise that DNA methylation is lost in BAC clones, though, in case epigenetic effects play a role in columnar growth in apple.

Relationship between the 'Wijcik' insertion and *MdCo31* expression

We did not expect to find that the mutation that was identified in 'Wijcik' would influence the expression of a gene at a distance of more than 15 kb, while

leaving the expression of neighbouring gene unaffected. However, examples of enhancers that can affect the expression of genes at much greater distances can be found in literature. An example of such a distal enhancer is found in mouse for the sonic hedgehog gene, at a distance of more than 1 Mb (Lettice et al. 2003). Until the ENCODE project, not much attention was given to such long-range interactions, but the importance of long-range interaction and the three-dimensional context of genes and regulatory elements starts to be appreciated now (Sanyal et al. 2012).

It is possible that the change in expression of *MdCo31*, from repression in buds of standard trees to increased expression of *MdCo31* in buds of columnar trees, is due to the change in genomic context of *MdCo31,* which could be caused by the retrotransposon insertion in 'Wijcik'. For example if the expression of *MdCo31* is normally repressed in standard trees. In such a scenario, integration of the *MdCo31* gene at a different genomic location could lead to a similar change in expression of *MdCo31* as is observed in 'Wijcik' and would possibly also result in columnar growth. This situation would be ideal for cisgenesis, because the native promoter of *MdCo31* can be maintained in this approach.

Alternatively, modern genome editing techniques, for example using engineered zinc finger nucleases or the CRISPR system (Urnov et al. 2010; Pennisi 2013), could be used to reproduce the effect of the 'Wijcik' retrotransposon in apple trees to obtain columnar trees. Such experiments could also be used to clarify the relation between the retrotransposon insertion in 'Wijcik' and the increased expression of *MdCo31*.

It would be of interest to compare the retrotransposon content of different apple cultivars, to see if differences can be found for the two elements that were identified in this thesis. This would not only give information about the evolution and the history of these elements, but these results could also be used to look at the effect of both elements on the expression of neighbouring genes. By comparing the expression of genes in a region surrounding the retrotransposons, the presence/absence of such elements can be correlated to the expression of these genes to assess their effect on gene expression.

To gain further insights into the organisation of the region surrounding the 'Wijcik' mutation, we proposed to investigate the chromosomal conformation of the *Co* region in apple using 3C (-derived) assays. In a preliminary analysis, using the EcoRI restriction enzyme to digest cross-linked genomic DNA of buds of columnar and standard trees, the ligation frequency of putative interaction sites in the *Co* region was tested (see Fig. 2 for the steps for template preparation), but further optimisation is required to obtain results for this experiment.

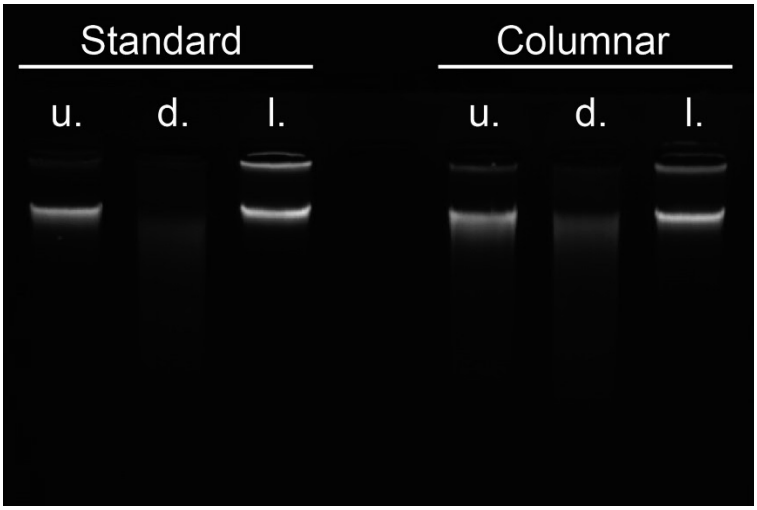

Figure 2. Steps required for 3C template preparation. Undigested (u.), cross-linked genomic DNA isolated from buds of standard and columnar trees is visible on the gel as a distinct band of high molecular weight on an agarose gel. Digestion of this DNA with the EcoRI restriction enzyme results in many small DNA fragment, as is clear from the DNA smear in the digested (d.) lane. After religation of these restriction fragments, a clear band is visible again (l., religated restriction fragments).

Although we would expect to find an interaction between the insertion site and *MdCo31*, it would also be interesting to see if other locations in the *Co* region show an interaction with *MdCo31*. In addition, the effect of the 'Wijcik' retrotransposon insertion on the expression of *MdCo31* could also be due to epigenetic changes in the *Co* region that are induced by the retrotransposon insertion. Information about chromatin methylation of the *Co* region could be investigated through chromatin precipitation experiments and DNA methylation could be investigated by performing digestions on genomic DNA of 'Wijcik' and 'McIntosh' using methylation sensitive/insensitive restriction enzymes followed by

PCR analyses using primers flanking the corresponding restriction sites (Wong 2006).

Conserved function of *MdCo31*?

The function of *MdCo31* needs further investigation. Transformation experiments that resulted in constitutive expression of *MdCo31* in tomato and Arabidopsis support the role of *MdCo31* in plant development, but the underlying mechanism is still unknown. The comparable effect of *MdCo31* that was observed in Arabidopsis, tomato and apple (short internodes and dark foliage), suggests a role for *MdCo31* in a conserved process, but the phylogenetic analysis that was carried out in chapter 5 suggests that *MdCo31* has a more specialised function.

When interpreting the phenotypes of the Arabidopsis and tomato plants constitutively expressing *MdCo31*, it is important to keep in mind that they are different from apple. Dwarfing is a commonly observed phenotype in mutant screens, so it might not be such a specific phenotype. No structures homologous to apple spurs are present in Arabidopsis and tomato, making it impossible to assess the effect of *MdCo31* overexpression on the development of spurs in these species. Also, the constitutive expression of *MdCo31* in Arabidopsis and tomato differs from the bud specific expression that is observed for that gene in columnar apple. It would be interesting to see what the developmental effects are of transforming tomato and Arabidopsis with a construct containing *MdCo31* under control of its native promoter.

No Arabidopsis genes are present in the clade containing *MdCo31* and the closest Arabidopsis genes are involved in susceptibility to downy mildew and salicylic acid catabolism (van Damme et al. 2008; Zhang et al. 2013). The *MdCo31* clade from the phylogenetic tree displayed in Chapter 5 contains several tomato genes. It would be interesting to silence these genes in tomato, to see what the developmental effects are and to investigate if there is a homolog of *MdCo31* in tomato.

The classification of *MdCo31* as a 2OGD suggests that *MdCo31* has a role in a metabolic pathway. The 2OGD gene family is large and contains well characterised genes with conserved functions as well as specialised members of which the function is often unknown (Kawai et al. 2014). Of the different processes

in which 2OGDs are involved, the pathways leading to the biosynthesis of GAs and flavonoids appear the most interesting concerning the function of *MdCo31*. Both possibilities are explored in chapter 5 and discussed below.

Potential role of *MdCo31* in flavonoid biosynthesis

Flavonoids can negatively regulate auxin transport (Brown et al. 2001; Peer and Murphy 2007) and the short internodes that were found in apple trees after silencing the chalcone synthase genes were explained by the inhibitory effect of flavonoids on auxin transport (Dare et al. 2013), providing a link between flavonoid content and plant architecture.

In order to investigate the effect of *MdCo31* expression on flavonoid content, a phenolic profiling experiment was carried out on Arabidopsis plants and on columnar and standard apple. This analysis showed higher levels of the flavonoids luteolin and rutin in plants with increased expression of *MdCo31*. However, we were unable to find an inhibitory effect of luteolin and rutin on auxin transport in Arabidopsis seedlings using a gravitropism assay. Moreover, no difference in gravity response was observed when WT Arabidopsis was compared to Arabidopsis plants that constitutively express *MdCo31*, undermining the hypothesis that auxin transport is affected in plants with increased levels of *MdCo31* transcription.

Potential role of *MdCo31* in GA biosynthesis

Plants with decreased active GA content or decreased GA sensitivity show dwarf- or columnar-like phenotypes (Peng et al. 1999; Busov et al. 2003; Zhu et al. 2008), suggesting a possible role for GAs in columnar growth. To see if application of GAs can reduce the effects of *MdCo31* expression in Arabidopsis, we sprayed Arabidopsis plants that constitutively express *MdCo31* with a GA solution. This GA treatment resulted in plants that were almost indistinguishable from WT plants, whereas untreated plants showed a clear compact phenotype. This effect suggests that plants that express *MdCo31* are indeed deficient in GA.

Columnar trees could similarly be treated with GA, to see if long branches are formed instead of the short spurs that are normally produced by such trees. The trees would probably needed to be treated for an extended time, as apple

trees develop a lot slower than Arabidopsis plants. The developmental fate of apple buds might be decided already during their formation, so it is not clear if long shoots can still be formed from axillary buds that are already present on the tree.

The effect of the GA treatment on the development of the Arabidopsis plants constitutively expressing *MdCo31* looks convincing, but it is not a proof that *MdCo31* is involved in GA catabolism. It would be useful to measure GAs in Arabidopsis and apple, but both the different levels of flavonoids that were found in plants with increased levels of *MdCo31* expression and the apparent GA deficiency in Arabidopsis plants expressing *MdCo31* can be downstream effects of another unknown mechanism that is affected by *MdCo31.*

Further experiments

Apart from being responsible for columnar growth in apple, *MdCo31* must be involved in a biological process in standard trees as well. However, no expression of *MdCo31* was detected during different conditions in different kinds of tissue collected from standard trees, failing to give us a clue about the function of *MdCo31* in normal plants. It would be interesting to see if silencing of *MdCo31* has an effect on the development of normal apple trees. Such an experiment would give information about the biological role of *MdCo31* in normal apple trees. Silencing *MdCo31* in 'Wijcik' trees should result in loss of columnar growth and would be a good test to prove that *MdCo31* is indeed the *Co* gene.

It would be interesting to perform a more extensive analysis to see if there are other changes in plants expressing *MdCo31* besides flavonoid levels and the observed effect of GA application. For example by performing a metabolite profiling targeting plant hormones, or by performing an untargeted metabolite profiling experiment. Combined with a transcriptome analysis, such an approach would give a better understanding of all the changes in the plant due to *MdCo31* expression and might provide information about the role of *MdCo31* even if the gene codes for a protein with a novel function.

Ideally, a direct comparison between 'McIntosh' and 'Wijcik' should be made, instead of comparing genetically diverse columnar and standard trees as was done in the transcriptome studies that were published previously (Krost et al. 2012; Zhang et al. 2012; Krost et al. 2013). Alternatively, the function of *MdCo31*

could be studied in a transcriptome study using Arabidopsis. The analysis of transcriptome data in Arabidopsis would be easier due to its small and well-annotated genome

Grafting experiments should give information about the type of hormones that are involved in columnar growth, as the production site and direction of transport is different for different hormones. Grafting is easy in apple, but it might take a while to obtain conclusive results because of its slow growth. Although a bit more challenging, such experiments could also be carried out in Arabidopsis (Turnbull et al. 2002). As 'Wijcik' was found on top of a 'McIntosh' tree, and as columnar trees derived from 'Wijcik' are often grafted onto rootstocks, we already know that grafting columnar trees onto standard trees does not affect the columnar trees. The effect on the lower part of the tree and the opposite graft (a standard tree on top of a columnar tree) is not yet investigated.

If the results from grafting experiments would enable us to develop a model for explaining the effect of *MdCo31* on plant architecture, this model could be further tested by crossing the Pro35S:MdCo31 line with Arabidopsis mutants for the appropriate genes, to see if *MdCo31* can complement the mutation or to see if the *MdCo31* effect is still present in the mutant background. Finally, *MdCo31* can be produced and purified in *E coli* to test the activity of the *MdCo31* protein on substrates of interest to provide additional proof for *MdCo31* function.

Conclusion

A large number of trees directly descendant from ‘Wijcik’ and segregating for the dominant columnar trait were analysed in this thesis. New SSR markers were developed for the columnar trait and the *Co* region was sufficiently reduced to enable a direct sequence comparison between the *Co* region of ‘McIntosh’ and ‘Wijcik’. With the use of two BAC libraries that were prepared from genomic DNA of both genotypes, a single difference was identified, an insertion in ‘Wijcik’.

The discovery of the insertion in ‘Wijcik’ enabled us to develop a convenient presence/absence marker that is 100% linked to *Co*. Considering that it takes a few years before tree architecture can be reliably scored, such a marker will be very useful for selecting trees in an early stage when breeding columnar apple trees.

The 'Wijcik' insertion corresponds to an 8.2 kb retrotransposon in the 5´ LTR of another retrotransposon of 4.0 kb that was already present at that position in 'McIntosh' apple. This was a surprising observation, as we expected to find a mutation in, or close to, a gene. An expression analysis of all genes predicted in a region of 50 kb surrounding this insertion revealed the upregulation of a single gene, *MdCo31*, in bud tissue of columnar trees. Remarkably, the expression of other genes, closer to the insertion of 'Wijcik' than *MdCo31*, is not affected.

Constitutive expression of *MdCo31* in Arabidopsis and tomato results in compact plants that bear resemblance to the 'Wijcik' mutant, supporting the role of *MdCo31* in the columnar phenotype of 'Wijcik'. Identification of *MdCo31* as a *Co* gene candidate opens possibilities for introducing the columnar trait to apple trees, or other plant species, without the need for performing crossing experiments, using gene technology. A transformation experiment with a construct for constitutive expression of *MdCo31* is currently in progress for apple, but it is not yet possible to evaluate the effect of constitutive expression of *MdCo31* on the development of apple. The bud-specific expression of *MdCo31* that was observed in columnar trees might in fact be essential for the desired phenotype, so constitutive expression of *MdCo31* could have other developmental effects in apple. If introduction of *MdCo31* in a different context would result in a similar increase in expression as is observed in columnar trees, the columnar trait could be introduced to any apple cultivar using cisgenesis. Introducing the gene in other species, or constitutive expression of the gene in apple, would require a transgenic approach.

The relationship between the retrotransposon insertion in 'Wijcik' and the increased expression of *MdCo31* needs further investigation. The conformation of the chromatin could play a role, or the retrotransposon insertion may have an effect on the epigenetic marks of the *Co* locus. Considering the recent attention for the functional role of the non-coding part of the genome, this is an interesting topic for further research. We proposed a number of experiments to investigate this in more detail.

MdCo31 codes for a putative 2OGD with a function that is currently unknown. The results presented in this thesis support a role for *MdCo31* in GA or flavonoid biosynthesis, but it is also possible that *MdCo31* is involved in a pathway

that has yet to be discovered. A transcriptome or metabolite comparison between standard plants and plants expressing *MdCo31* would help identify the function of *MdCo31*. The Arabidopsis plants constitutively expressing *MdCo31* could be a valuable resource for these analyses.

All in all, a lot of progress has been made in our understanding of the cause for the mutant phenotype of 'Wijcik'. We identified a single genomic difference between 'McIntosh' and 'Wijcik' and isolated a strong candidate gene. To understand the full mechanism leading to columnar growth in apple, more research is still needed.

References

Bai, T., Y. Zhu, et al. (2012). "Fine genetic mapping of the Co locus controlling columnar growth habit in apple." Molecular Genetics and Genomics **287**(5): 437-450.

Baldi, P., P. J. Wolters, et al. (2013). "Genetic and physical characterisation of the locus controlling columnar habit in apple (Malus × domestica Borkh.)." Molecular Breeding **31**(2): 429-440.

Brown, D. E., A. M. Rashotte, et al. (2001). "Flavonoids act as negative regulators of auxin transport in vivo in arabidopsis." Plant Physiology **126**(2): 524-535.

Busov, V. B., R. Meilan, et al. (2003). "Activation tagging of a dominant gibberellin catabolism gene (GA 2-oxidase) from poplar that regulates tree stature." Plant Physiology **132**(3): 1283-1291.

Dare, A. P., S. Tomes, et al. (2013). "Phenotypic changes associated with RNA interference silencing of chalcone synthase in apple (Malus× domestica)." The Plant Journal **74**(3): 398-410.

Espley, R. V., C. Brendollse, et al. (2009). "Multiple repeats of a promoter segement causes transcription factor autoregulation in red apples." The Plant Cell **21**: 168-183.

Fisher, D. V. (1969). "Spur-type strains of McIntosh for high density planting." British Columbia Fruit Grower's Association Quart. Rep. **14**: 3-10.

Greene, B., R. Walko, et al. (1994). "Mutator insertions in an intron of the maize knotted1 gene result in dominant suppressible mutations." Genetics **138**(4): 1275-1285.

Hemmat, M., N. F. Weeden, et al. (1997). "A DNA marker for columnar growth habitat in apple contains a simple sequence repeat." Journal of American Society for Horticultural Science **122**: 347-349.

Kawai, Y., E. Ono, et al. (2014). "Evolution and diversity of the 2-oxoglutarate-dependent dioxygenase superfamily in plants." The Plant Journal.

Kim, M. Y., K. J. Song, et al. (2003). "Development of RAPD and SCAR markers linked to the Co gene conferring columnar growth habit in apple (Malus pumila Mill.)." The Journal of Horticultural Science & Biotechnology **78**(4): 512-517.

Kronmiller, B. A. and R. P. Wise (2008). "TEnest: automated chronological annotation and visualization of nested plant transposable elements." Plant Physiology **146**(1): 45-59.

Krost, C., R. Petersen, et al. (2013). "Evaluation of the hormonal state of columnar apple trees (Malus x domestica) based on high throughput gene expression studies." Plant Mol Biol **81**(3): 211-220.

Krost, C., R. Petersen, et al. (2012). "The transcriptomes of columnar and standard type apple trees (Malus x domestica) - A comparative study." Gene **498**(2): 223-230.

Lettice, L. A., S. J. Heaney, et al. (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." Human molecular genetics **12**(14): 1725-1735.

Lisch, D. (2013). "How important are transposons for plant evolution?" Nat Rev Genet **14**(1): 49-61.

Looney, N. E. and W. D. Lane (1984). "Spur-type growth mutants of McIntosh apple: A review of their genetics, physiology and field performance." Acta Horticultura **146**: 31-46.

Moriya, S., H. Iwanami, et al. (2009). "Development of a marker-assisted selection system for columnar growth habit in apple breeding." Journal of Japanese Society for Horticultural Science **78**: 279-287.

Moriya, S., K. Okada, et al. (2012). "Fine mapping of Co, a gene controlling columnar growth habit located on apple (Malus×domestica Borkh.) linkage group 10." Plant Breeding **131**(5): 437-450.

Naito, K., F. Zhang, et al. (2009). "Unexpected consequences of a sudden and massive transposon amplification on rice gene expression." Nature **461**(7267): 1130-1134.

Otto, D., R. Petersen, et al. (2013). "The columnar mutation ("Co gene") of apple (Malus × domestica) is associated with an integration of a Gypsy-like retrotransposon." Molecular Breeding: 1-18.

Peer, W. A. and A. S. Murphy (2007). "Flavonoids and auxin transport: modulators or regulators?" TRENDS in Plant Science **12**(12): 556-563.

Peng, J., D. E. Richards, et al. (1999). "'Green revolution' genes encode mutant gibberellin response modulators." Nature **400**(6741): 256-261.

Pennisi, E. (2013). "The CRISPR Craze." Science **341**(6148): 833-836.

Sanyal, A., B. R. Lajoie, et al. (2012). "The long-range interaction landscape of gene promoters." Nature **489**(7414): 109-113.

Silfverberg-Dilworth, E., C. L. Matasci, et al. (2006). "Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome." Tree genetics & genomes **2**(4): 202 - 224.

Takken, F. L., R. Van Wijk, et al. (2004). "A one-step method to convert vectors into binary vectors suited for Agrobacterium-mediated transformation." Curr Genet **45**(4): 242-248.

Tian, Y., C. Wang, et al. (2005). "Mapping *Co*, a gene controlling the columnar phenotype of apple, with molecular markers." Euphytica **145**: 181-188.

Tobutt, K. R. (1985). "Breeding columnar apples at East Malling." Acta Horticultura **159**: 63-68.

Turnbull, C. G., J. P. Booker, et al. (2002). "Micrografting techniques for testing long-distance signalling in Arabidopsis." The Plant Journal **32**(2): 255-262.

Urnov, F. D., E. J. Rebar, et al. (2010). "Genome editing with engineered zinc finger nucleases." Nature Reviews Genetics **11**(9): 636-646.

van Damme, M., R. P. Huibers, et al. (2008). "Arabidopsis DMR6 encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew." Plant J **54**(5): 785-793.

Velasco, R., A. Zharkikh, et al. (2010). "The genome of the domesticated apple (Malus x domestica Borkh.)." Nature Genetics **42**(10): 833-839.

Wicker, T., F. Sabot, et al. (2007). "A unified classification system for eukaryotic transposable elements." Nature Reviews Genetics **8**(12): 973-982.

Wong, I. H. (2006). Qualitative and quantitative polymerase chain reaction-based methods for DNA methylation analyses. Clinical Applications of PCR, Springer**:** 33-43.

Yu, D., H. M. Ellis, et al. (2000). "An efficient recombination system for chromosome engineering in Escherichia coli." Proceedings of the National Academy of Sciences **97**(11): 5978-5983.

Zhang, K., R. Halitschke, et al. (2013). "Salicylic acid 3-hydroxylase regulates Arabidopsis leaf longevity by mediating salicylic acid catabolism." Proc Natl Acad Sci U S A **110**(36): 14807-14812.

Zhang, Y., J. Zhu, et al. (2012). "Characterization of transcriptional differences between columnar and standard apple trees using RNA-Seq." Plant Molecular Biology Reporter **30**(4): 957-965.

Zhu, L., X. Li, et al. (2008). "Overexpression of the Arabidopsis gai gene in apple significantly reduces plant size." Plant cell reports **27**(2): 289-296.

Zhu, Y. D., W. Zhang, et al. (2007). "Evaluation of inter-simple sequence repeat analysis for mapping the Co gene in apple (Malus pumila Mill.)." The Journal of Horticultural Science & Biotechnology **82**(3): 371-375.

Summary

Trees in modern apple orchards are grafted onto growth controlling rootstocks. The trees are planted close to each other and pruned extensively in order to obtain high density apple orchards containing compact trees in which apple production is highly optimised. Apple cultivars that naturally show such compact growth are of interest, because those trees would need less maintenance from apple growers.

In 1964, a compact shoot of the apple cultivar 'McIntosh' was discovered in an apple orchard in Canada, which was named after its discoverer Anthony Wijcik. This 'Wijcik' shoot was further propagated and showed a very interesting ‘columnar’ growth. 'Wijcik' trees have a thick stem, short internodes and they show very little branching. Instead of branches, short, fruit-bearing ‘spurs’ are formed, resulting in a sturdy, compact and fructuous tree.

Because of its promising characteristics, 'Wijcik' trees, and columnar trees derived from ‘Wijcik’, have been subject to many studies since their discovery. When columnar trees are crossed with normal growing apple trees, half of the progeny is also columnar, meaning that columnar growth is a dominant trait. The columnar locus (*Co*) was mapped on chromosome 10.

Due to the traits of the parental cultivar ‘McIntosh’, ‘Wijcik has undesirable traits, such as biannual bearing, low sugar content, soft fruit flesh and susceptibility to apple scab. When using conventional crossing to introduce the columnar trait into new apple cultivars, both these linked and unlinked negative properties will be transferred to the progeny as well. Molecular breeding techniques, such as marker assisted selection (MAS) and genetic engineering, could help to deal with this problem.

Using molecular markers for the *Co* locus, the precision and efficiency of apple breeding programs aiming to breed columnar trees could be greatly improved. Using MAS, it would be easier to pick the right parents for performing crosses. Promising progeny plants could be selected at an early stage already, based on their genotype. This is especially useful in case of breeding columnar apple, because the reliable evaluation of tree architecture and fruit quality can take several years after sowing.

Identifying the mutation in 'Wijcik', or identifying the gene responsible for the phenotype of columnar trees, would enable the development of columnar apple varieties using genetic engineering. As columnar growth is a trait that is found in a mutant of apple, we expected that it would be possible to use cisgenesis. Using this technique, genes of interest can directly be transferred to the apple variety of choice. In this way, a specific dominant trait like columnar growth could be directly targeted, while avoiding the introduction of undesirable traits due to genetic drag.

The goal of this thesis is to identify the mutation in 'Wijcik' and to isolate the gene responsible for the phenotype of columnar trees. These results will help the development of columnar apple varieties through MAS or through genetic engineering. Moreover, identifying the *Co* gene will improve our understanding of the physiological background of columnar growth.

In **chapter 2** of this thesis, we described the fine mapping of the *Co* region in apple. We used 5 progeny populations (in total more than 1500 individuals) segregating for the columnar trait, which are directly descendant from 'Wijcik', to reduce the *Co* region to 393 kb in the homologous region from 'Golden Delicious'. Several new SSR markers were developed in the *Co* region, including an SSR marker that co-segregates with *Co* in all individuals from our segregating populations. This marker could facilitate breeding of future columnar varieties and, in addition, the high density map for this trait is an essential condition for cloning of the *Co* gene. Although a number of interesting transcription factors is present in the homologous *Co* region in 'Golden Delicious', it was not yet possible to identify a single *Co* gene candidate.

In **chapter 3**, the SSR markers that were described in chapter 2 were used to screen BAC libraries of 'McIntosh' and 'Wijcik'. BAC clones covering the *Co* region of 'McIntosh' and 'Wijcik' were sequenced and the resulting sequences were compared. A novel non-coding DNA element of 1956 bp was reported in this chapter, which was found to be inserted in an intergenic region of 'Wijcik'. This single genomic difference was used to develop a new presence/absence marker that is 100% linked to *Co*. Because the presence of the insertion in 'Wijcik' is no direct explanation for its columnar phenotype, the genes in a region of 50 kb surrounding the insertion were compared in leaf and bud material of 'McIntosh' and

'Wijcik'. This expression analysis revealed the upregulation of a single gene, *MdCo31*, in axillary buds of 'Wijcik'. *MdCo31* codes for a putative 2OG-Fe(II) oxygenase and is located at 15.6 kb downstream of the insertion in 'Wijcik'. Surprisingly, the expression of other genes surrounding the 'Wijcik' insertion was found to be unaltered in 'Wijcik' trees. Transformation of Arabidopsis plants with a construct for constitutive expression of *MdCo31* provided additional evidence for the role of *MdCo31* in columnar growth of apple.

In **chapter 4**, the 'Wijcik' mutation is characterised in further detail**.** Our findings were compared with a second study that was recently published by another group, who identified the insertion of a retrotransposon at the exact same location as was reported by us, but who stated that this retrotransposon measured 8.2 kb in size instead of the 1956 bp mentioned in chapter 3. After a review of our previous analyses, we found that the sequence of the insertion that was identified by us is in fact the LTR of a bigger retrotransposon of 8.2 kb. Due to the long LTRs, our assembly collapsed on this LTR sequence and the PCR that should have amplified the full insertion produced a product for the LTR only, as a PCR artefact. We concluded that the mutation in 'Wijcik' is an 8.2 kb retrotransposon that inserted into the 5' LTR of another retrotransposon that was already present in 'McIntosh', resulting in a nested transposon in 'Wijcik'. Many copies of similar transposons were found to be present throughout the 'Golden Delicious' genome, and nested transposons are not unusual, but it is striking to find such a dramatic phenotypic effect of the insertion of a retrotransposon into another retrotransposon.

In **chapter 5**, we showed that *MdCo31* has a comparable effect in diverse plant species, suggesting a conserved function for *MdCo31*. The phylogenetic analysis that was carried out in this chapter points to a more specialised function of *MdCo31*. Because flavonoids and gibberellins are both interesting compounds that could possibly be involved in the columnar phenotype of 'Wijcik', we investigated a potential role for *MdCo31* in the biosynthesis pathways for these compounds. The levels of specific flavonoid compounds were found to be elevated in columnar trees and in Arabidopsis plants that constitutively express *MdCo31*, but we were unable to show an effect of these compounds on the development of Arabidopsis seedlings. Conversely, application of gibberellin on Arabidopsis plants that

constitutively expressed *MdCo31* abolished the phenotypic effect from *MdCo31*, leading to plants that were almost indistinguishable from wildtype Arabidopsis. This suggests a role for gibberellins in explaining the effect of *MdCo31* expression.

In **Chapter 6** of this thesis, we discuss the results presented in this thesis. The effect of the insertion of a transposon into another transposon on the expression of *MdCo31* is intriguing and the function of *MdCo31* is still unknown. We propose a number of experiments to address these questions.

The results described in this thesis are a big step forward in the research of columnar growth in apple. We identified the mutation leading to columnar growth in 'Wijcik' and we were able to develop a new molecular marker that is 100% linked to *Co*. A very likely *Co* gene candidate was also identified, which may be used to improve the architecture of apple trees and other plant species. The biochemical function of *MdCo31* and the relation between the 'Wijcik' retrotransposon insertion and the expression of *MdCo31* are both very interesting topics for further investigation

Samenvatting

De appelbomen in moderne appelboomgaarden worden geënt op onderstammen die de groei van de ent onder controle houden. De bomen worden dicht op elkaar geplant en flink gesnoeid, zodat een boomgaard ontstaat met een hoge dichtheid aan appelbomen, waarin de productie van appels gemaximaliseerd is. Appelbomen die uit zichzelf zo'n compacte groeiwijze hebben zijn interessant, omdat dit soort bomen minder onderhoud nodig hebben dan normale appelbomen.

In 1964 werd een compacte scheut van het appelras 'McIntosh' ontdekt, die naar de ontdekker Anthony Wijcik werd genoemd. Deze 'Wijcik' scheut werd verder vermeerderd en bleek een interessante 'kolomachtige' groeiwijze te hebben. 'Wijcik' appelbomen hebben een dikke stam, een kleine afstand tussen de internodiën en ze vertakken heel weinig. In plaats van laterale takken worden korte uitlopers gevormd die vrucht dragen. Op deze wijze ontstaat een stevige, compacte, boom die veel vrucht draagt.

Vanwege deze interessante eigenschappen was 'Wijcik' het onderwerp van vele studies sinds de mutant voor het eerst ontdekt werd. De groeiwijze van 'Wijcik' vererft dominant, wat betekent dat de helft van de nakomelingen van een kruising tussen 'Wijcik' en een normale appelboom op dezelfde wijze groeit als 'Wijcik'. Onderzoek heeft verder aangetoond dat de oorzaak voor de groeiwijze van 'Wijcik' op chromosoom 10 ligt, in het 'columnar' (*Co*) locus.

Ongewenste eigenschappen van 'McIntosh', zoals onregelmatige vruchtzetting, laag suikergehalte, zacht vruchtvlees en vatbaarheid voor appelschurft, zijn ook terug te vinden in 'Wijcik'. Wanneer 'Wijcik' gekruist wordt met andere appelrassen, dan worden, naast de groeiwijze, ook deze eigenschappen overgebracht naar de nakomelingen. Moleculaire veredelingstechnieken, zoals marker assisted selection (MAS) en genetische modificatie kunnen helpen om dit probleem te omzeilen.

De precisie en efficiëntie van veredelingsprogramma's om appelbomen met kolomgroei te ontwikkelen kan enorm verbeterd worden door gebruik te maken van moleculaire merkers. MAS kan gebruikt worden om de juiste ouders te kiezen voor de kruisingen. Bovendien kunnen nakomelingen van de kruising al in een vroeg stadium geselecteerd worden, op basis van hun genotype. Dit is erg nuttig,

aangezien het jaren kan duren voordat eigenschappen als vruchtkwaliteit en groeiwijze van de boom op een betrouwbare manier beoordeeld kunnen worden.

De ontwikkeling van appelbomen met kolomgroei zou met behulp van genetische modificatie kunnen gebeuren wanneer de mutatie in 'Wijcik', of het gen dat verantwoordelijk is voor de speciale groeiwijze, geïdentificeerd zou zijn. Aangezien kolomgroei voorkomt in een mutant van appel, verwachtten we dat dit via cisgenese zou kunnen. Specifieke eigenschappen, zoals kolomgroei, kunnen met cisgenese doelgericht overgebracht worden, terwijl de overbrenging van ongewenste eigenschappen wordt vermeden.

Het doel van dit promotieonderzoek is om de mutatie in 'Wijcik' op te sporen en om het gen dat verantwoordelijk is voor kolomgroei te identificeren. Deze resultaten zullen de ontwikkeling van appelbomen met kolomgroei vergemakkelijken. Daarnaast zal de identificatie van het *Co* gen ons begrip van het mechanisme dat tot kolomgroei in appel leidt verbeteren.

Het *fine mappen* van het *Co* locus werd beschreven in **hoofdstuk 2** van dit proefschrift. Door gebruik te maken van 5 nakomelingenpopulaties (in totaal meer dan 1500 individuen), die direct afstammen van 'Wijcik', konden we het genomisch gebied voor *Co* verkleinen tot 393 kb in de homologe regio van 'Golden Delicious'. Nieuwe SSR merkers werden ontwikkeld in dit *Co* locus, waaronder één merker die altijd gekoppeld bleek aan kolomgroei in de individuen uit onze nakomelingenpopulaties. Deze merker zou goed gebruikt kunnen worden in veredelingsprogramma's voor het ontwikkelen van appelbomen met kolomgroei en de genetische kaart die ontwikkeld werd in dit hoofdstuk is noodzakelijk als uitgangspositie voor het kloneren van he *Co* gen. Een aantal interessante transcriptiefactoren bleken aanwezig te zijn in het *Co* gebied van 'Golden Delicious', maar het was nog niet mogelijk om één enkel kandidaatgen aan te wijzen.

De SSR merkers die in hoofdstuk 2 werden beschreven, zijn in **hoofdstuk 3** gebruikt om een BAC library van 'McIntosh' en een BAC library van 'Wijcik' te analyseren. BAC klonen die het *Co* gebied van 'McIntosh' en 'Wijcik' beslaan werden gesequenct en de sequenties werden vergeleken. Op deze manier werd een nieuw, niet-coderend, DNA element van 1956 bp beschreven, die geïnserteerd

bleek in een intergenisch gebied van 'Wijcik'. Dit genomische verschil werd gebruikt om een nieuwe merker te ontwikkelen die 100% gekoppeld is aan het *Co* gen. Omdat de insertie geen directe verklaring is voor het fenotype van 'Wijcik', werd de expressie van alle genen in een gebied van 50 kb om de insertie vergeleken tussen 'McIntosh' en 'Wijcik', in knop- en bladmateriaal. Deze analyse leidde tot de identificatie van één enkel gen, *MdCo31*, waarvan de expressie specifiek verhoogd was in knoppen van 'Wijcik'. *MdCo31* codeert waarschijnlijk voor een 2OG-Fe(II) oxygenase en bevindt zich op een afstand van 15,6 kb van de insertie in 'Wijcik'. De expressie van andere genen in de buurt van de insertie bleek verrassend genoeg onveranderd te zijn. Aanvullend bewijs voor de rol van *MdCo31* in kolomgroei van appel kwam van een transformatie-experiment, waarbij *MdCo31* uit appel tot expressie werd gebracht in Arabidopsis.

De 'Wijcik' mutatie werd verder gekarakteriseerd in **hoofdstuk 4**. Onze eerdere bevindingen werden vergeleken met de resultaten van een onderzoek dat recent werd gepubliceerd door een andere onderzoeksgroep. Deze onderzoeksgroep vond een insertie op exact dezelfde plek als de locatie die door ons werd gerapporteerd, maar de insertie werd beschreven als een retrotransposon met een lengte van 8.2 kb, in plaats van de 1956 bp die beschreven werd in hoofdstuk 3. Na herziening van onze eerdere analyse, concludeerden we dat de insertie die door ons werd beschreven overeenkomt met de LTR van een groter retrotransposon van 8.2 kb. Door de grote lengte van de LTRs werd het middelste gedeelte van het retrotransposon gemist in de beschrijving van de insertie in 'Wijcik', en de PCR die het complete retrotransposon had moeten laten zien produceerde slechts de LTR als artefact. De insertie bleek dus inderdaad 8.2 kb in lengte te zijn en bleek bovendien geïnserteerd in de 5' LTR van een tweede retrotransposon die al eerder aanwezig was in het genoom van 'McIntosh'. In het genoom van 'Golden Delicious' bleken veel kopieën van beide elementen aanwezig te zijn en het gebeurt wel vaker dat een retrotransposon in een ander transposon terecht komt, maar het is verrassend om hiervan zo'n groot effect op de ontwikkeling van een plant te vinden.

In **hoofdstuk 5** lieten we zien dat *MdCo31* een verglijkbaar effect heeft op de ontwikkeling van verschillende plantensoorten, wat mogelijk op een basale rol

voor *MdCo31* duidt. De fylogenetische boom die in dit hoofdstuk gepresenteerd werd lijkt echter op een meer gespecialiseerde functie voor *MdCo31* te wijzen. In dit hoofdstuk gingen we verder in op een mogelijke rol voor *MdCo31* in de productie van flavonoïden, of gibberellines. We vonden een verhoogd niveau van bepaalde flavonoïden in appelbomen met kolomgroei en in Arabidopsis planten waarbij *MdCo31* tot overexpressie werd gebracht, maar we zagen geen effect van deze flavonoïden op de ontwikkeling van zaailingen van Arabidopsis. We zagen wel een effect van gibberelline op de ontwikkeling van deze Arabidopsis zaailingen. Het toedienen van gibberelline aan Arabidopsis planten waarin *MdCo31* tot expressie wordt gebracht, zorgt ervoor dat het effect van *MdCo31* nagenoeg teniet gedaan werd. Mogelijk zijn gibberellines dus betrokken bij het effect van de expressie van *MdCo31*.

De resultaten die beschreven worden in dit proefschrift worden verder bediscussieerd in **hoofdstuk 6**. Het effect van de insertie van een transposon in een ander transposon is interessant en de functie van *MdCo31* is nog steeds niet opgehelderd. We stellen een aantal experimenten voor om deze zaken verder uit te zoeken.

Het onderzoek dat beschreven is in dit proefschrift is een grote stap voorwaarts in het onderzoek naar kolomgroei in appel. We beschreven de mutatie die verantwoordelijk is voor kolomgroei in appel en konden een nieuwe merker ontwikkelen die 100% gekoppeld is aan het *Co* gen. We identificeerden een zeer aannemelijke kandidaat voor het *Co* gen, die gebruikt zou kunnen worden om de groeiwijze van appelbomen, en andere plantensoorten, te beïnvloeden. Het zou erg interessant zijn om de biochemische functie van *MdCo31* en het verband tussen de insertie in 'Wijcik' en de expressie van *MdCo31* verder uit te zoeken in de toekomst.

Acknowledgements

After spending four years in Italy, my PhD thesis is finished! There were many people involved in my thesis research, who I would all like to thank for their help.

First, I would like to thank Henk Schouten, who was involved in the project proposal for my PhD research and who introduced me to the GMPF school of FEM in San Michele all'Adige. I am thankful for your support when I applied for the PhD fellowship and it was great to have you as a supervisor. I appreciate your engagement in the project and your critical attitude has been very important in carrying out my thesis.

Riccardo Velasco, Silvio Salvi and Paolo Baldi were also involved in my PhD research from the beginning. Without their contribution, the project would not have existed. I am grateful to Paolo Baldi, who was my direct supervisor once I started to work in Italy. Under stressful conditions you kept your head cool and your confidence in my work has been a great encouragement.

I would like to thank my promotor, Evert Jacobsen, for the discussions and his suggestions, especially at the later stage of my PhD when I was writing my thesis. I am grateful to Inova Fruit b.v., Technological Topinstitute Green Genetics, the Autonomous Province of Trento and the GMPF PhD school at Fondazione Edmund Mach, who provided the necessary funds for carrying out my work.

Shortly after I started the PhD work in Italy, Paolo and I became part of the Functional Genomics group, headed by Azeddine Si-Ammour. I would like to thank all other current and past members of this group (Emanuele De Paoli, Marina Cavaiuolo, Giulia Miolli, Elisa Asquini, Matteo Brilli, Mirko Moser and Nada Šurbanovski) for their help, their company and their suggestions, which were all valued very much. I enjoyed to work in the group of Azeddine, whose enthusiasm resulted in many ideas for experiments. I really appreciated your help and the time you put in writing the paper for New Phytologist.

I am thankful for the work that was done by the members of the sequencing platform (Massimo Pindo, Simone Larger, Erika Stefani and Daniela Nicolini), who helped to genotype the segregating apple populations, screen the

BAC libraries of 'McIntosh' and 'Wijcik' and who have performed many sequencing reactions for me. I would like to thank the bioinformaticians working at FEM, especially Alessandro Cestaro, Emilio Potenza, Marco Moretto and Luca Bianco for their advice on bioinformatics and for their help with analyses.

Other people I would like to acknowledge are Pierluigi Magnago, for maintaining the trees in the apple orchard, Jan Schaart, who performed the transformation of tomato plants with the *MdCo31* gene, Lisa Giacomelli, for the talks we had about gibberellins and for her help with protein work, Dolf Weijers, for the useful discussions we had and for your fresh ideas, Harro Bouwmeester, for sharing his expertise on plant hormones, Dan Sargent, for his interest in my work and for giving me the chance to present it at the PAG conference in San Diego, Pier Luigi Bianchedi and Massimo Confalonieri for their help with the work we attempted to carry out in poplar, Alan Schulman and Ruslan Kalendar for their advice on transposable elements, Narda van der Krogt for her help with the design of the cover of this thesis, Stefan Martens for his advice about flavonoids and Urska Vrhovsek for the profiling of phenolic compounds in apple and Arabidopsis.

Besides these people that were directly involved in my work, I would like to thank my friends and family and all the amazing people I met during these four years. You made the time that I spent in Italy a wonderful experience that I will never forget.

Many thanks to all of you!!

Jaap

About the Author

Pieter Jacobus (Jaap) Wolters was born on the 19th of October, 1986, in Enschede (the Netherlands). After finishing his pre-university education, he started his bachelor's studies in Biology at Wageningen University in 2004. After graduating in 2007, he continued his master's studies in Biology (with the specialisation 'Cell Biology') at the same University. His master thesis research was carried out at the laboratory of phytopathology at Wageningen University, under supervision of Dr. Wladimir Tameling and Dr. Matthieu Joosten, where he investigated the role of the NB-LRR NRC1 in Solanaceae. For his internship, he spent 6 months in Cologne (Germany), as a member of the plant chemetics group at the Max Planck Institute for Plant Breeding Research. Under supervision of Dr. Renier van der Hoorn and Dr. Takayuki Shindo, he investigated the papain-like cysteine protease RD21 in Arabidopsis and *N. benthamiana*. One year after obtaining his master degree, Jaap moved to Trento (Italy), where he got a scholarship to carry out his PhD research at the Fondazione Edmund Mach. His PhD project aimed to characterise the molecular basis of columnar growth in apple. This project was supervised by Dr. Henk Schouten and Dr. Paolo Baldi and the results of his study are described in this thesis

List of publications

Baldi, P., P. J. Wolters, M. Komjanc, R. Viola, R. Velasco and S. Salvi (2013). "Genetic and physical characterisation of the locus controlling columnar habit in apple (Malus × domestica Borkh.)." Molecular Breeding **31**(2): 429-440.

Wolters, P. J., H. J. Schouten, R. Velasco, A. Si-Ammour and P. Baldi (2013). "Evidence for regulation of columnar habit in apple by a putative 2OG-Fe(II) oxygenase." New Phytologist **200**(4): 993-999.

Wolters, P. J., H. J. Schouten, R. Velasco, S. Salvi, M. Komjanc, A. Si-Ammour and P. Baldi (submitted). "Towards identification of the gene responsible for the columnar phenotype of McIntosh 'Wijcik' apple." Acta Horticultura.

Education Statement of the Graduate School
Experimental Plant Sciences

The Graduate School EXPERIMENTAL PLANT SCIENCES

Issued to: Pieter Jacobus Wolters
Date: 7 October 2014
Group: Plant Breeding
University: Wageningen University & Research Centre

1) Start-up phase	date
► **First presentation of your project**	
Isolation and functional analysis of the Co-gene for a columnar architecture of apple trees	Jun 25, 2010
► **Writing or rewriting a project proposal**	
Isolation and functional analysis of the Co-gene for a columnar architecture of apple trees	Jun 2010
► **Writing a review or book chapter**	
► **MSc courses**	
► **Laboratory use of isotopes**	
Subtotal Start-up Phase	*7.5 credits**

2) Scientific Exposure	date
► **EPS PhD student days**	
PhD day FIRST PhD school Fondazione Edmund Mach (FEM)	Oct 23, 2013
EPS PhD student day, Leiden University	Nov 29, 2013
► **EPS theme symposia**	
EPS theme 4 symposium 'Genome Biology', Wageningen University	Dec 13, 2013
EPS theme 4 symposium 'Genome Biology', Wageningen University	Dec 05, 2014
► **NWO Lunteren days and other National Platforms**	
TTI Green Genetics day	Sep 22, 2010
► **Seminars (series), workshops and symposia**	
Workshop 'Floral Biology and S-incompatibility in fruit species' (FEM)	Jun 22-25, 2011
Workshop 'Introduction to Bioinformatics' (FEM)	Jun 28-Jul 01, 2011
ENGAGE/GEUVADIS RNA-seq Workshop at EMBL-EBI (Hinxton, UK)	Dec 04-06, 2012
Lecture series James A. Shapiro, 'On genomics and evolution' (FEM)	Nov 12-15, 2013
► **International symposia and congresses**	
6th international Rosaceae genomics conference (Mezzocorona, Italy)	Sep 30-Oct 04, 2012
Molecular Markers in Horticulture conference (Riva del Garda, Italy)	Sep 25-27, 2013
Plant and Animal Genome conference (San Diego, USA)	Jan 11-15, 2014
► **Presentations**	
Poster: TTI Green Genetics	Sep 22, 2010
Oral: TTI Green Genetics	Sep 22, 2010
Poster: TTI Green Genetics	Sep 21, 2011
Oral: Molecular Markers in Horticulture conference (Riva del Garda, Italy)	Sep 27, 2013
Oral: FEM PhD day	Oct 23, 2013
Oral: James A. Shapiro lecture series(sensing TE, epigenetic control of TE transcription)	Nov 13, 2013
Oral: Plant and Animal Genome conference (San Diego, USA)	Jan 11-15, 2014
► **IAB interview**	
Meeting with a member of the International Advisory Board of EPS	Nov 05, 2012
► **Excursions**	
Subtotal Scientific Exposure	*17.1 credits**

3) In-Depth Studies	date
► **EPS courses or other PhD courses**	
Course: 'Introduction to Pland Metabolomics' (FEM)	Jul 04-08, 2011
Summer School 'Gene Expression and Pathway Reconstruction' (FEM)	Jun 26-29, 2012
Summer School 'Population and quantitative Genetics' (FEM)	Jul 02-06, 2012
► **Journal club**	
Member of literature discussion group at FEM	2010-2011
► **Individual research training**	
Subtotal In-Depth Studies	*5.7 credits**

4) Personal development	date
► **Skill training courses**	
The Essentials of Scientific Writing and Presenting	Dec 02,03,09,10, 2013
Course '6.00x: Introduction to Computer Science and Programming' (edX| MITx)	Feb 04-Jun 10, 2013
► **Organisation of PhD students day, course or conference**	
► **Membership of Board, Committee or PhD council**	
Subtotal Personal Development	*6.7 credits**

TOTAL NUMBER OF CREDIT POINTS*	**37.0**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study.*