Molecular Genetic Causes of Columnar Growth in Apple (*Malus* x *domestica*)

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

(Dr. rer. nat.)

am Fachbereich Biologie der Johannes Gutenberg-Universität Mainz

vorgelegt von

Romina Petersen

geboren am 20.04.1986 in Wiesbaden

Mainz, 2014

Dekan:

1. Berichterstatter:

2. Berichterstatter:

Tag der mündlichen Prüfung:

There is nothing like looking if you want to find something […] You certainly usually find something, if you look, but it is not always quite the something you were after.

JRR Tolkien, The Hobbit

Cumulative Statement

This thesis is based on the following seven research articles, which are presented as attachments and are referred to in the introduction, results summary and discussion by their respective numbers:

Paper 1:
Petersen R., Krost C. (2013) Review: 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

Paper 2:
Otto D., Petersen R., Krost C., Schmidt E.R., Brandl R., Brauksiepe B., Braun P. (in press) Molecular Characterization of the *Co* Gene Region in *Malus* x *domestica*. *Acta Hort*

Paper 3:
Otto D.*, Petersen R.*, Brauksiepe B., Braun P., Schmidt E.R. (2014) The columnar mutation ("*Co* gene") of apple (*Malus* x *domestica*) is associated with an integration of a *Gypsy*-like retrotransposon. *Mol Breeding* **33**: 863-880 *: joint first authorship

Paper 4:
Krost C., Petersen R., Schmidt E.R. (2012) The transcriptomes of columnar and standard type apple trees (*Malus* x *domestica*) – a comparative study. *Gene* **498**(2): 223-230

Paper 5:
Krost C., Petersen R., Braun P., Schmidt E.R. (in press) Deep sequencing of the shoot apical meristem transcriptome of columnar apple trees (*Malus* x *domestica*). *Acta Hort*

Paper 6:
Krost C., Petersen R., Lokan S., Brauksiepe B., Braun P., Schmidt, E.R. (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

Paper 7:
Petersen R., Djozgic H., Rieger B., Rapp S., Schmidt E.R. (under review) Columnar apple primary roots share some features of the columnar-specific gene expression profile of aerial plant parts as evidenced by comparative RNA-Seq analysis.

Table of Contents

Figures and Tables

Figures

Tables

List of Abbreviations

A14	‘A14-190-93K’
A73	‘A73-19-97K’
AFLP	amplified fragment length polymorphism
BAC	bacterial artificial chromosome
bp	base pair(s)
cDNA	complementary DNA
Co	*Columnar*
DNA	deoxyribonucleic acid
EN	endonuclease
ENV	envelope
EST	expressed sequence tag
FAIRE	formaldehyde-assisted isolation of regulatory elements
GAG	group-specific antigen
IAA	indole-3 acetic acid
ChIP	chromatin immunoprecipitation
INT	integrase
LINE	long interspersed nuclear element
LTR	long terminal repeat
Mb	megabase(s)
MDP	*Malus* x *domestica* protein
miRNA	microRNA
MITE	miniature inverted repeat transposable element
MULE	mutator-like element
NGS	next generation sequencing
ORF	open reading frame
P28	‘Procats 28’
PA	palindromic region
PBS	primer binding site
POL	polymerase
PPT	polypurine tract
PR	primary root
qRT-PCR	quantitative real-time polymerase chain reaction
REP	replicator protein
RH	RNaseH
RNA	ribonucleic acid
RT	reverse transcriptase
SAGE	serial analysis of gene expression
SAM	shoot apical meristem
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
TAIR	the *Arabidopsis* information resource
tb1	*teosinte branched 1*
TE	transposable element
TIR	terminal inverted repeat
tRNA	transfer RNA
TSD	target site duplication
vgt1	*vegetative to generative transition 1*
Wijcik	‘McIntosh Wijcik’

Abstract

The columnar growth habit of apple is interesting from an economic point of view as the pillar-like trees require little space and labor. Genetic engineering could be used to speed up breeding for columnar trees with high fruit quality and disease resistance. For this purpose, this study dealt with the molecular causes of this interesting phenotype. The original bud sport mutation that led to the columnar growth habit was found to be a novel nested insertion of a *Gypsy*-44 LTR retrotransposon on chromosome 10 at 18.79 Mb. This subsequently causes tissue-specific differential expression of nearby downstream genes, particularly of a gene encoding a 2OG-Fe(II) oxygenase of unknown function (*dmr6-like*) that is strongly upregulated in developing aerial tissues of columnar trees. The tissue-specificity of the differential expression suggests involvement of cis-regulatory regions and/or tissue-specific epigenetic markers whose influence on gene expression is altered due to the retrotransposon insertion. This eventually leads to changes in genes associated with stress and defense reactions, cell wall and cell membrane metabolism as well as phytohormone biosynthesis and signaling, which act together to cause the typical phenotype characteristics of columnar trees such as short internodes and the absence of long lateral branches. In future, transformation experiments introducing *Gypsy*-44 into non-columnar varieties or excising *Gypsy*-44 from columnar varieties would provide proof for our hypotheses. However, since site-specific transformation of a nested retrotransposon is a (too) ambitious objective, silencing of the *Gypsy*-44 transcripts or the nearby genes would also provide helpful clues.

Zusammenfassung

Das Kolumnarwachstum von Apfelbäumen ist eine wirtschaftlich interessante Wuchsform, da die säulenförmig wachsenden Bäume wenig Platz und Arbeitseinsatz erfordern. Gentechnologische Ansätze könnten zukünftig dazu verwendet werden, die Züchtung kolumnarer Bäume mit hoher Fruchtqualität und Krankheitsresistenz zu beschleunigen. Zu diesem Zweck beschäftigt sich die vorliegende Arbeit mit den molekulargenetischen Ursachen dieses interessanten Phänotyps. Die ursprüngliche Sprossmutation, welche in den 1960er Jahren zum Kolumnarwachstum führte, wurde als neue Insertion des LTR-Retrotransposons *Gypsy*-44 in ein anderes LTR-Retrotransposon auf Chromosom 10 an Position 18,79 Mb identifiziert. Diese führt zur gewebespezifischen differentiellen Expression benachbarter, *downstream* liegender Gene, insbesondere eines für eine 2OG-Fe(II)-Oxygenase unbekannter Funktion kodierenden Gens (*dmr6-like*), das in oberirdischen, sich entwickelnden Geweben kolumnarer Bäume stark hochreguliert ist. Die Gewebespezifität der differentiellen Genexpression deutet auf den Einfluss *cis*-regulatorischer Regionen bzw. gewebespezifischer epigenetischer Marken hin, deren Effekt auf die Genexpression aufgrund der Retrotransposon-Insertion verändert ist. Dies führt im Folgenden zu Veränderungen der Aktivität von Genen, die an Stress- und Abwehrreaktionen, dem Metabolismus von Zellwand- und Zellmembran-Komponenten sowie der Biosynthese und den Signalwegen von Phytohormonen beteiligt sind. Gemeinsam verursachen diese Veränderungen die Ausbildung der typischen phänotypischen Charakteristika kolumnarer Apfelbäume wie verkürzte Internodien und das Fehlen langer Seitenäste. In Zukunft könnten Transformationsexperimente, in denen *Gypsy*-44 in nicht-kolumnare Apfelsorten eingefügt oder aus kolumnaren Apfelsorten herausgeschnitten wird, weitere Beweise für unsere Hypothesen erbringen. Da die ortsspezifische Transformation eines Restrotransposons vermutlich ein (zu) ehrgeiziges Ziel ist, könnte auch der Knock-down der Transkripte von *Gypsy*-44 oder der benachbarten Gene hilfreiche Hinweise liefern.

General Introduction

The Domesticated Apple, *Malus* x *domestica*

"The apple does not fall far from the tree" ("Der Apfel fällt nicht weit vom Stamm") is probably the oldest and most common genetic "axiom" (albeit not always true). It also demonstrates that apples are deeply rooted in our everyday life as they are used in idioms and phrases in many European languages. In several cultures and religions apples are a symbol of sexuality and fertility, life, insight and decision, but they can also represent sin, e.g. as the forbidden fruit that had Adam and Eve expelled from paradise. The broad distribution of apples in our language and culture is most likely due to the fact that apples are the third most popular fruit in the world (after watermelons and bananas) with a production of more than 76 million tons worldwide in 2012 (www.faostat.fao.org). As part of the recommended five portions of fruits and vegetables a day, apples confer health benefits such as a reduced risk for cancer, cardiovascular disease, obesity, type II diabetes and asthma thanks to their high content of polyphenols, antioxidants and fibers (reviewed in Boyer and Liu 2004; Hyson 2011). Fortunately, apples provide easy access to healthy nutrition since they can be consumed raw as well as in processed forms such as apple juice, apple pie and apple sauce.

The domesticated apple (*Malus* x *domestica*) belongs to the family of Rosaceae, which is the third most agronomically important plant family in temperate regions (Dirlewanger et al., 2002), comprising fruit species such as pear (genus *Pyrus*), peach (*Prunus persica*), cherry (*Prunus avium*) and strawberry (genus *Fragaria*). *Malus* x *domestica* is part of the subfamily Spiraeoideae, tribe Pyreae, subtribe Pyrinae. The Pyreae lineage arose in the Eocene (55.5 – 33.7 million years ago) and one of its dominant characteristics is the pome fruit, which is a fleshy, indehiscent false fruit created by an enlarged floral tube, a fleshy receptacle, or both (Juniper and Mabberley, 2006). The genus *Malus* consists of 30 – 50 species, the exact number being debatable according to the characteristics applied for species delimitation, as *Malus* species are highly diverse and prone to hybridization, polyploidization and apomixis (Forsline et al., 2003; Janick et al., 1996; Luby, 2003; Phipps et al., 1991; Robinson, 2001).

Even though the apple tree now seems to be a native plant to us, it is of exotic origin: its center of highest genetic diversity is the Tian Shan region in Kazakhstan near Almaty (formerly Alma-Ata, meaning "father of apples") (Janick et al., 1996). In this region in Central Asia, the wild progenitor of the domesticated apple, *Malus sieversii*, can still be found

in forest regions (Forsline et al., 2003). Based on molecular analyses, *Malus sieversii* and *Malus* x *domestica* have been found to be so closely related that they can be considered the same species, for which the name *Malus pumila* has been proposed (Juniper and Mabberley, 2006; Mabberley et al., 2001; Velasco et al., 2010). However, in this thesis the name *Malus* x *domestica* (Korban and Skirvin, 1994) is maintained because it better reflects its interspecific origin due to hybridization of *Malus sieversii* with *Malus asiatica*, *Malus orientalis*, *Malus baccata*, *Malus mandshurica*, *Malus prunifolia* and especially with the European crabapple, *Malus sylvestris* (Cornille et al., 2012; Janick et al., 1996; Pereira-Lorenzo et al., 2009). Nowadays apples are common in temperate regions around the world and have adapted to colder climates with longer winters as well as to a wide range of altitudes (Kellerhals, 2009). They are monoecious, deciduous trees with a height of up to 10 m and a longevity of up to 100 years (Pereira-Lorenzo et al., 2009).

Apple breeding probably has its origins in the selection and propagation of wild apple fruits before 6500 BC (Pereira-Lorenzo et al., 2009); however, the first real evidence for apple cultivation found in Anatolia and Mesopotamia dates back to the second millennium BC (Luby, 2003). The discovery of vegetative propagation via grafting 3800 years ago greatly facilitated apple cultivation because it enabled the maintenance and distribution of a specific cultivar (Harris et al., 2002). Since the 16th century, dwarfing rootstocks have been used to control tree height (Pereira-Lorenzo et al., 2009). The first targeted crosses were performed by Thomas Knight in 1806 (Kellerhals, 2009). Since then, breeders have always sought cultivars with an improved fruit quality and storage ability, and an increased pest and disease resistance, which could be produced at a lower labor and energy cost (Peace and Norelli, 2009). Today, more than 20,000 highly diverse apple cultivars are known. However, only few of them have reached economic importance. 'Delicious', 'Golden Delicious', 'McIntosh', 'Braeburn', 'Gala', 'Granny Smith' and 'Fuji' dominate the apple market, and these are mostly chance seedlings that did not originate from controlled crossings (Janick et al., 1996). The reasons for the sparse success of apple breeding strategies are the long juvenile phase of apple trees (3 – 10 years depending on the genotype and cultivation practices) (Hackett, 1985; Janick et al., 1996) during which no offspring can be produced, self-incompatibility, which hinders backcrossing (Halász et al., 2006; Hegedüs, 2006), and a high level of heterozygosity complicating the introduction of a specific trait into a stable genetic background (Pereira-Lorenzo et al., 2009; Velasco et al., 2010). Therefore, producing a marketable new apple variety can take more than 50 years (Schouten et al., 2006). This makes apple a popular target for genetic engineering, which could alleviate some of the problems of traditional breeding. First gene technological approaches have already enabled shortening of the juvenile phase to about one year

(Flachowsky et al., 2011, 2007). Furthermore, marker-assisted selection and the development of single nucleotide polymorphism (SNP) microarrays (Chagné et al., 2012) nowadays facilitate the selection of the desired genetic foreground and background. However, these applications require solid knowledge of the genome and transcriptome of the apple, which is why a lot of research effort has been and is being put into these fields.

Apple Genomics

Compared with the study of animal genomes, plant genome sequencing and analysis are still in their infancies. This is due to a high number of plant characteristics hampering genome sequencing and assembly. First, it can be daunting to isolate high-quality DNA from plant tissues owing to their high content of carbohydrates and secondary metabolites such as polyphenols, which interfere with isolation techniques or downstream applications (Lee and Nicholson, 1997; Varma et al., 2007). Second, despite gene numbers of similar orders of magnitude, plant genomes have a wide range of sizes (from the 82 Mb of *Utricularia gibba* (Ibarra-Laclette et al., 2013) to the 110,000 Mb of the lily *Fritillaria assyriaca* (Arumuganathan and Earle, 1991)) and many economically interesting plants have very large genomes. This is mostly due to the large amount of repetitive elements (reaching 80 % in maize) (SanMiguel et al., 1998, 1996; Vitte et al., 2007) and/or a high level of polyploidy, since up to 70 % of higher plants have undergone at least one event of polyploidization (Levin, 2002). Third, plant species are often highly heterozygous with SNP frequencies of up to 1 SNP/61 bp in maize (Jones et al., 2009). Hence, so far only the rather small genomes of *Arabidopsis* (125 Mb) (The Arabidopsis Genome Initiative, 2000) and rice (389 Mb) (International Rice Genome Sequencing Project, 2005) have been completely sequenced and assembled. This was achieved by a BAC-by-BAC (Bacterial Artificial Chromosome) procedure and Sanger sequencing, requiring a large consortium of researchers, a long time frame and significant financial investment. All other plant genomes have since been sequenced based partially or entirely on a less laborious whole genome shotgun approach (Shangguan et al., 2013). With the advent of Next Generation Sequencing (NGS) technologies such as 454 pyrosequencing or Solexa/Illumina, generating raw genomic shotgun sequences became much faster, easier and cheaper. However, NGS delivers millions of unordered short reads and therefore provides tremendous challenges for assembling, especially when the content of repetitive elements is high (Imelfort and Edwards, 2009). Thus, about 30 plant genomes are now in a draft stage of differing quality (Shangguan et al., 2013). Fortunately, the Rosaceae family has been the subject of several genome sequencing efforts so that high-

quality drafts are already available for apple (Velasco et al., 2010), strawberry (Shulaev et al., 2011), pear (Wu et al., 2013) and peach (The International Peach Genome Initiative 2013), while those of other species such as raspberry are underway (Jung and Main, 2013).

Malus x *domestica* is diploid, with the exception of some triploid cultivars such as 'Jonagold' (Giovannoni, 2010). As opposed to most Rosaceae, which have a haploid chromosome set of x=8 or x=9, apple and its close relative pear have 17 chromosomes, suggesting a duplication event in the Pyreae lineage (Juniper and Mabberley, 2006). This might have happened either by allopolyploidization of an Amygdaloideae ancestor (x=8) and a Spiraeoideae ancestor (x=9) (Chevreau et al., 1985; Sax, 1933) or by autopolyploidization of a Spiraeoideae ancestor (x=9) and the subsequent loss of a chromosome (Campbell et al., 1995; Morgan et al., 1994). Newer molecular genetic evidence supports the spiraeoid hypothesis and points towards a *Gillenia*-like progenitor of apple (Evans and Campbell, 2002; Velasco et al., 2010).

The genome of Golden Delicious has a size of 742.3 Mb and has been deciphered by Velasco et al. (2010) by a combination of BAC sequencing and 454 sequencing. About 81 % of the genome were assembled into 1,629 metacontigs using a gene-centric approach. Metacontigs were anchored to linkage groups based on 1,643 markers, but could not be merged to individual chromosome sequences due to the low sequence identity of the two different alleles. The average SNP frequency was detected to be 1 SNP/227 bp (Velasco et al., 2010), while subsequent studies determined the SNP frequency in the apple germplasm to be 1 SNP/52 bp and the insertion and deletion rate 1 Indel/333 bp (Micheletti et al., 2011; Troggio et al., 2012). High colinearity between large segments of chromosomes 5 and 10, 3 and 11, 9 and 17 as well as 13 and 16 and additional shorter segments of different chromosomes are the result of the recent whole-genome duplication in the Pyreae lineage about 60 – 65 million years ago and remnants of paleohexaploidization, which occurred in most eudicot lineages about 140 million years ago (Tang et al., 2008; Van de Peer et al., 2009; Wu et al., 2013). As for most plants, repetitive sequences, mostly transposable elements (TEs), take up the biggest portion of the genome (67 %). Because of their significance for plant genome structure and evolution, TEs will be described in detail in the following section. Besides the nuclear genome, the chloroplast (160,068 bp) and mitochondrial (396,947 bp) genome sequences were sequenced and assembled (Velasco et al., 2010).

Unfortunately, despite high accuracy of the apple genome sequence (i.e. nearly all expressed sequence tags (ESTs) from databases can be found as gene annotations on the genomic sequence) it has a very low integrity, indicating incorrect assembly (Shangguan et

al., 2013) and/or mis-anchoring of contigs (Khan et al., 2012). Therefore, particularly regions containing a high amount of repetitive DNA are in need of reassembly.

Transposable Elements

In plants, even individuals of the same species or of two closely related sister species can show large differences in genome content and structure (Bennetzen, 2005; Morgante et al., 2007; Vitte and Panaud, 2005). The main reason for this is the difference in the amount of TEs. For example, two maize inbred lines have been shown to contain about 50 % of different genomic sequences mainly due to long terminal repeat (LTR) retrotransposon insertions (Brunner et al., 2005). TEs were first discovered by Barbara McClintock in maize (McClintock, 1951) and can be divided into two classes: class I TEs (Fig 1A) transpose via an RNA intermediate by a ”copy-and-paste” mechanism, whereas class II TEs (Fig. 1B) usually transpose in their DNA form by a “cut-and-paste” mechanism (Bennetzen, 1996).

Class II TEs (DNA transposons) encode a transposase mediating excision and integration and have terminal inverted repeats, which are the transposase recognition sites (Fig. 1B). When a DNA transposon integrates into a DNA stretch, transposase makes a staggered cut causing short, single-stranded nucleotide overhangs. After insertion, the gaps opposite the overhangs are filled with the corresponding nucleotides so that a short target site duplication (TSD) is created. When the TE excises, the TSD and sometimes also small parts of the TE itself remain as footprints. Due to their cut-and-paste mechanism, the copy number of specific class II TEs within a genome is usually low and they do not significantly influence genome size (Kumar and Bennetzen, 1999).

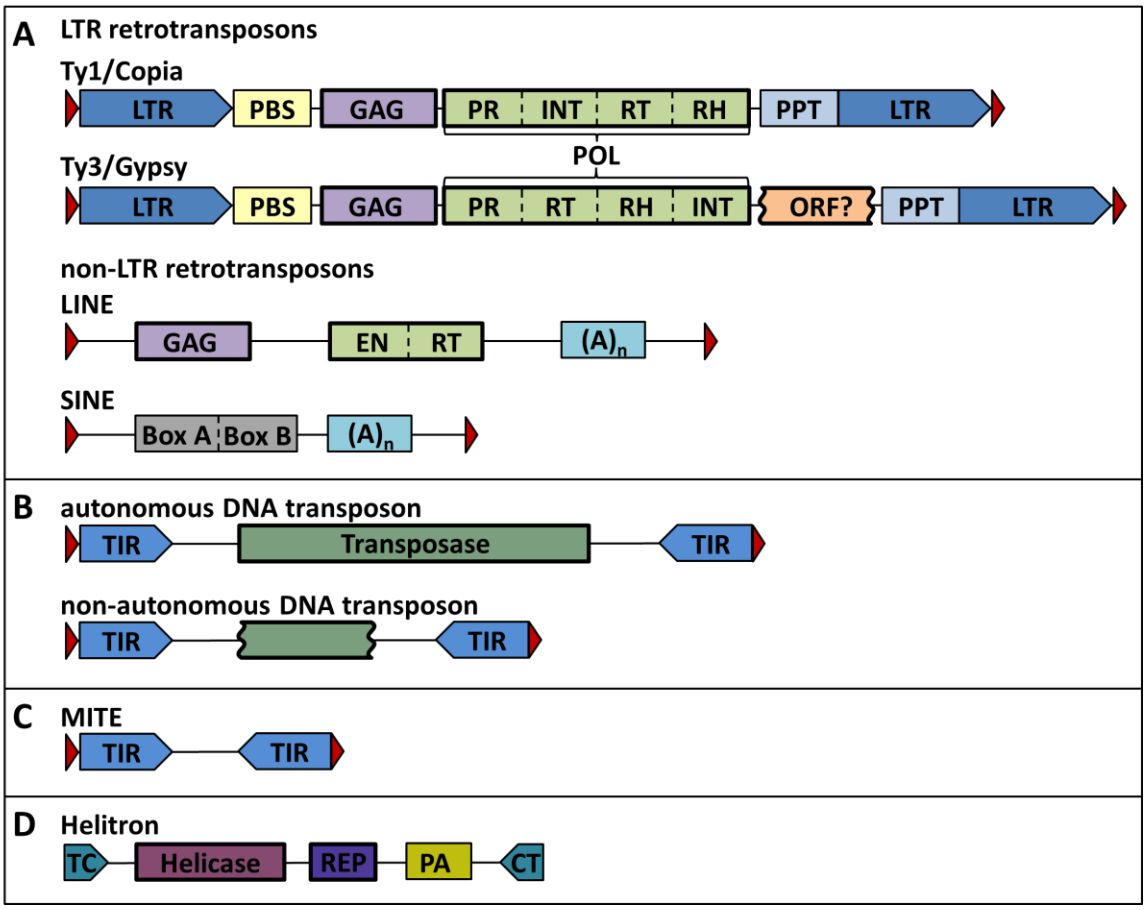


Figure 5: Classification of transposable elements.
Transposable elements (TEs) can be subdivided into class I elements (A) transposing via an RNA intermediate, class II elements (B) transposing in their DNA form, miniature inverted-repeat elements (MITEs) (C), and Helitrons (D) using a rolling-circle mechanism for transposition. Class I elements are further classified into LTR retrotransposons of the *Ty1/Copia* and the *Ty3/Gypsy* family differing by the order of their polymerase domains, and non-LTR retrotransposons of the autonomous LINE and non-autonomous SINE types. Class II elements comprise autonomous DNA transposons harboring a functional transposase and non-autonomous DNA transposon with a truncated transposase. Red triangles represent target sites. $(A)_n$: poly-A stretch, EN: endonuclease, GAG: group-specific antigen, INT: integrase, LTR: long terminal repeat, ORF: open reading frame, PA: palindromic region, PBS: primer binding site, POL: polymerase, PPT: polypurine tract, PR: protease, REP: replication initiator protein, RH: RNase H, RT: reverse transcriptase, TIR: terminal inverted repeat. This figure was inspired by the publications of Kumar and Bennetzen (1999), Casacuberta and Santiago (2003), Wicker et al. (2007), and Otto (2013).

By contrast, class I TEs can be present in high copy numbers within the ten thousands (SanMiguel et al., 1996). Class I TEs comprise LTR retrotransposons and non-LTR retrotransposons (retroposons) (Fig. 1A). LTR retrotransposons are closely related to retroviruses and therefore have a virus-like organization and life cycle. They can reach sizes of up to 18 kb (Vitte and Panaud, 2005) and their eponymous characteristic are the LTRs, direct repeats at the TE ends that can span up to a few thousand base pairs (Wright and Voytas, 2002). As active TE they possess at least two open reading frames (ORFs) encoding a

group-specific antigen (*gag*) and a polymerase (*pol*). Some LTR retrotransposons contain an additional envelope (*env*)-like ORF (Laten et al., 2003; Wright and Voytas, 2002), or ORFs that were sequestered from other genes (Laten and Gaston, 2012). *pol* encodes a polyprotein composed from four domains fulfilling different functions in the transposon life cycle: protease, integrase, reverse transcriptase and RNaseH. After transcription by RNA polymerase II and translation, protease cleaves the polyprotein into its individual components. Typically two RNA strands are packaged into virus-like particles composed of GAG proteins (Sabot and Schulman, 2006). Within these particles, reverse transcriptase produces transposon cDNA using a host tRNA molecule as a primer, and RNaseH degrades the RNA strand of the intermediate RNA-DNA hybrid. Subsequently integrase mediates cDNA integration at a new chromosomal location, producing a TSD. With regard to the order of the individual domains within the *pol* ORF, two families of LTR retrotransposons are distinguished, named after their representatives in yeast and *Drosophila*: *Ty1/Copia*, in which the order is protease – integrase – reverse transcriptase – RNaseH, and *Ty3/Gypsy*, whose integrase domain is arranged downstream of the reverse transcriptase and RNaseH domains (Xiong and Eickbush, 1990). *Ty1/Copia* elements are usually found in euchromatic regions, whereas *Ty3/Gypsy* TEs more frequently cluster in heterochromatic regions around the centromeres and telomeres (Kumar and Bennetzen, 1999). Both types often occur nested (inserted into each other) at transposon-rich regions (Kalendar et al., 1999; Ramsay et al., 1999; SanMiguel et al., 1996).

Non-LTR retrotransposons can be classified into Long Interspersed Nuclear Elements (LINEs) and Short Interspersed Nuclear Elements (SINEs). Similar to LTR retrotransposons, LINEs possess a *gag* ORF and an ORF encoding an endonuclease-reverse transcriptase polyprotein mediating TE transcription and integration. However, LINEs have no LTRs and contain a poly-A stretch at their 5' end (Feschotte et al., 2002). SINEs comprise promoter elements for RNA polymerase III as well as a poly-A stretch, but lack complete ORFs and can therefore only be transposed by a second, autonomous element (Kajikawa and Okada, 2002).

Miniature Inverted-Repeat Transposable Elements (MITEs) and Helitrons are often considered as distinct groups of TEs. MITEs are structurally reminiscent of non-autonomous class II elements due to their terminal inverted repeats (Fig. 1C), but have the high copy numbers of class I elements (Bureau and Wessler, 1994, 1992). Helitrons are only flanked by a TC on the 5' and a CTRR motif on the 3' end and do not produce TSDs (Fig. 1D). They transpose via a unique rolling-circle mechanism most likely involving a helicase and a replication initiator protein (Dong et al., 2011; Kapitonov and Jurka, 2001). Helitrons have been tightly associated with gene capture and exon shuffling (Morgante et al., 2005).

Most of the TEs within a plant genome have acquired stop codons, frameshifts and insertions/deletions, rendering them non-functional and thus non-autonomous (Sun et al., 2008; Vitte and Panaud, 2005). However, after activation, often by stress (Casacuberta and Santiago, 2003; Kumar and Bennetzen, 1999; Wessler, 1996), an autonomous element can either transpose itself or mobilize a related non-autonomous element, which can induce mutations within the plant. Since these mutations can be harmful to the host, plants try to silence TEs and eliminate TE sequences. Silencing of TEs can occur transcriptionally via hypermethylation of cytosines within the transposon sequences, which in addition to weakening gene expression favors C-T transitions leading to an accumulation of deleterious mutations within the TE (Lisch, 2009; Rabinowicz et al., 2003; SanMiguel et al., 1998, 1996). An alternative mechanism is posttranscriptional silencing via RNA degradation mediated by antisense RNA (Feschotte et al., 2002; Okamoto and Hirochika, 2001). Elimination of TE sequences is mostly achieved through unequal homologous recombination, which in case of retrotransposons can take place between the two LTRs of the same TE or between LTRs of two related TEs (Devos et al., 2002; Ma et al., 2004). The former mechanism creates a solo-LTR fringed by TSDs, the latter results in a solo-LTR flanked by two different target sites and the deletion of a genomic region. Furthermore, illegitimate recombination leads to small-scale deletions within TE sequences and thus combats “genome obesity” (Bennetzen and Kellogg, 1997; Devos et al., 2002; Ma et al., 2004).

On the other hand, from an evolutionary point of view TEs can be advantageous to plants because they promote genetic diversity and thus facilitate adaptation to environmental changes, which is especially important for plants as sessile organisms (Nakayashiki, 2011). This is mediated via changes in gene expression or gene function, which can occur in many different ways (reviewed in Lisch (2013)). The simplest possibility is the introduction of null mutations when TEs insert into the coding region of a gene (Grandbastien et al., 1989; Vignols et al., 1995). TE insertions might also destroy or provide promoters, enhancers or silencers (Kobayashi et al., 2004; Studer et al., 2011; Yao et al., 2001) or cause changes in the methylation pattern of the nearby chromosomal region (Fujimoto et al., 2008; Kinoshita et al., 2007). In case of TE integration into introns, tissue-specific alternative splicing may be induced (Leprince et al., 2001; Varagona et al., 1992). Furthermore, TEs can provide new genes or exons themselves; e.g. a specific family of class II Mutator-Like Elements (MULEs), the so-called Pack-MULEs, has been shown to frequently capture parts of genes (Jiang et al., 2004). Large-scale insertions/deletions and rearrangements induced by the recombination between two transposon copies can lead to alterations in gene structure or gene expression (C. Yu et al., 2011). Finally, after TE “domestication” a sequence derived from TEs can fulfill roles that are advantageous to the host (reviewed in Volff (2006)). For example, in

Drosophila, HeT-A and TART TEs preferentially jump to the ends of the chromosomes, alleviating the end replication problem (Pardue and DeBaryshe, 2003). In *Arabidopsis*, the hAT-like transposase gene *daysleeper* regulates growth and development (Volff, 2006).

In apple, the number of TEs reaches almost 500,000, adding up to a length of more than 300 Mb (Velasco et al., 2010). *Ty3/Gypsy* retrotransposons are the biggest family of TEs with about 200,000 copies, representing about 38 % of the whole apple genome. By contrast, all class II TEs taken together only comprise about 50,000 copies (Sun et al., 2008; Velasco et al., 2010). Approximately 50 % of the LTR retrotransposons have aquired premature stop codons and frameshifts, and no transcription of LTR retrotransposon reverse transcriptase was detected in somatic tissues of 'Fuji' *in vitro* cultures under non-stressful conditions (Sun et al., 2008).

Apple Transcriptomics

Transcriptome analyses provide an overview of the transcripts that are present at a given time point within a certain tissue and can be used to compare gene expression between samples. For several years, in model organisms the field was dominated by microarray experiments, which can efficiently detect the expression of a known set of genes per se as well as in comparison between two different individuals, tissues, time points or conditions (Schena et al., 1995). Later, serial analysis of gene expression (SAGE) and related tag-based strategies conferred the advantage of generating digital gene expression counts and partial mRNA sequences (Velculescu et al., 1995), the latter of which could also be obtained by cDNA amplification length polymorphism (cDNA-AFLP) (Breyne et al., 2003; Vuylsteke et al., 2007). Furthermore, whole genome tiling arrays enabled the detection of new genes or exons (Kapranov et al., 2002; Yamada et al., 2003). For non-model organisms or when longer mRNA sequences were desired, expressed sequence tag (EST) cDNA libraries were the method of choice (Adams et al., 1991). However, their generation is time- and cost-intensive. The "gold standard" for the determination of absolute and relative gene expression is quantitative real-time polymerase chain reaction (qRT-PCR) (Canales et al., 2006), but it also requires at least a partial knowledge of the target gene sequence and is rather expensive and labor-intensive so that it cannot be applied to a comprehensive analysis of an entire transcriptome.

Much like genome sequencing, the analysis of transcriptomes has been revolutionized with the introduction of NGS. NGS transcriptome sequencing (RNA-Seq) combines the generation of up to full-length mRNA sequences (after assembly) of known and unknown

transcripts with the benefits of digital read counts (Wang et al., 2009). Compared with microarrays, there is no problem of cross-hybridization or background signals, no upper threshold for gene detection and the chance of detecting low abundance transcripts can easily be increased by sequencing to a higher depth. In addition, RNA-Seq facilitates the analysis of alternative splicing, trans-splicing, alternative polyadenylation, variants and intergenic transcriptionally active regions (Consortium, 2007; Mortazavi et al., 2008; Nagalakshmi et al., 2008; Wang et al., 2009). However, the large amounts of data necessitate computer aided sequence analysis and, in the case of differential gene expression analysis, additional statistical evaluation. There are two approaches to deal with RNA-Seq data: the assembly approach, which generates contigs ideally representing full-length mRNAs, and the mapping approach, which assigns the sequence reads to their appropriate positions on a reference genome or a reference transcriptome (Korf, 2013). Similar to genomics, assembling the large number of short reads still provides challenges (reviewed in Steijger et al. (2013)), even though transcriptome assemblies are less complex since mRNA sequences are much shorter and less repetitive than sequences of whole chromosomes. Mapping is essentially a standard technique now (reviewed in Engström et al. (2013)), but it requires access to the genome or transcriptome sequence of the species analyzed or a close relative.

To date, our knowledge of the apple genome and its transcriptome is still rather rudimental. Velasco et al. (2010) estimated the gene content of Golden Delicious to be 57,386 genes encoding *Malus* x *domestica* proteins (MDPs), plus 31,678 TE-related ORFs, which would be the highest gene content reported in plants so far. There is a high number of genes for transcription factors, and 11,444 genes were proposed to be apple-specific as they have not been found in any other species yet (Velasco et al., 2010). The GC content of apple genes was determined to be on average 44 % (Newcomb et al., 2006). However, pear only has about 42,000 genes, and when the apple genome is re-assembled filtering out overlapping genes that might be alleles rather than individual genes, the gene number drops down to 45,293 (Wu et al., 2013). This would be more consistent with gene numbers of other close relatives such as peach (27,852) (The International Peach Genome Initiative 2013) and strawberry (34,809) (Shulaev et al., 2011). Newcomb et al. (2006) conducted one of the first exhaustive EST analyses in apple and identified about 43,000 non-redundant sequences, which they thought to be approximately half of the apple genes. By contrast, the apple EST data set of 68,599 sequences from databases was considered an overestimate by Allan et al. (2009). On the other hand, based on EST analyses by Sanzol (2010), 68 % of apple genes cluster into families with a mean copy number of 4.6, and the members of one family can have highly similar sequences but still represent different genes rather than alleles. Hence, the exact gene number remains debatable.

Gene expression in apple has been analyzed in a variety of tissues and developmental stages and with many different methods. Microarrays have been used to depict gene expression in buds and fruits throughout the seasons and development (Costa et al., 2010; Janssen et al., 2008; Pichler et al., 2007; Vimolmangkang et al., 2014) and to find similarities in gene expression patterns between fruit abscission induced by shading or a synthetic auxin (Zhu et al., 2011). Fruitlet abscission (Dal Cin et al., 2009), rootstock effects (Jensen et al., 2003) and interaction with the causal agent of fire blight, *Erwinia amylovora* (Baldo et al., 2010), have been investigated using cDNA-AFLP, whereas suppression subtractive hybridization was used to identify differentially expressed transcripts after infection with *Venturia inaequalis*, which induces apple scab (Degenhardt et al., 2005). Heat shock transcription factors (Giorno et al., 2012), ion transporter and aquaporin genes (Liu et al., 2012) as well as cell cycle genes (Malladi and Johnson, 2011) were subjected to qRT-PCR analysis. Recently, RNA-Seq has been used to identify new candidate genes for resistance to apple scab (Gusberti et al., 2013) and it can be expected that new RNA-Seq studies will shed light on the apple transcriptomes in different tissues and developmental stages in the near future. Consequently, an NGS approach clearly is the method of choice to unravel the molecular basis of an interesting apple phenotype such as columnar growth.

Columnar Growth and Plant Growth Regulation

The columnar growth habit of apple is characterized by the formation of thick main stems with short internodes and the almost complete absence of long lateral branches, which are replaced by short fruit spurs (Tobutt, 1994, 1985). It has been reviewed in detail in Paper 1 and therefore only the key points will be briefly mentioned here. Columnar growth arose as a spontaneous somaclonal limb sport mutation of a ‘McIntosh’ tree in Canada in the 1960s, which was spotted by the apple grower Anthony Wijcik (Fisher, 1995, 1969). He grafted the columnar sport and thus created the first columnar apple variety, ‘McIntosh Wijcik’ (Wijcik). Subsequent crossings of Wijcik with non-columnar varieties yielded different columnar cultivars such as ‘Procats 28’ (P28), which was obtained from a cross of ‘Telamon’ and ‘Topaz’ at the Geisenheim University, the former being a direct columnar descendant of Wijcik and the latter a commercially available non-columnar variety. McIntosh, Wijcik, P28 and ‘A14-190-93K’ (A14), which is a non-columnar cultivar created at the Geisenheim University, are the main apple varieties used for genomic and transcriptomic analyses throughout this project.

The columnar phenotype might be advantageous to apple growers since trees can be planted at a high density and require no staking and little pruning (Tobutt, 1985). Unfortunately, columnar apple varieties usually have a low fruit quality and disease resistance and are therefore not yet market-ready (Gelvonauskienë et al., 2006; Lauri and Lespinasse, 1993; Meulenbroek et al., 1998). Genetic engineering might change this, provided that the molecular genetic basis of columnar growth is elucidated. Early crossings showed that the columnar phenotype is caused by one dominant allele of a gene, called "*Columnar*" (*Co*), for which Wijcik and all commercially available apple cultivars are heterozygous (Lapins and Watkins, 1973; Lapins, 1969; Tian et al., 2005). One homozygous individual, 'A73-19-97K' (A73), has recently been identified in the Geisenheim orchards (Paper 3). Several mapping studies located *Co* on chromosome 10 of the apple genome, and the target region has most recently been refined to about 18.5 – 19.0 Mb (Bai et al., 2012; Baldi et al., 2012; Moriya et al., 2012). However, the identity, function and gene product of *Co* have remained enigmatic up to now. In addition, the influence of one or several modifier genes has been proposed in order to explain the phenomenon that the percentage of columnar progeny in crossings of columnar and non-columnar varieties is lower than it would be expected from a strictly dominant-recessive Mendelian inheritance (Lapins 1976, Paper 1).

It seems obvious that columnar apple trees suffer changes in key plant growth regulation pathways. Therefore growth regulation of aerial organs is also covered in Paper 1. Phytohormones are the major players in plant architecture control and mostly target transcription factors that coordinate key modules of plant growth. Studies on phytohormone levels of columnar apple trees have yielded ambiguous results, but there is ample evidence that they have a higher auxin/cytokinin ratio and lower levels of gibberellins and abscisic acid than normal apple trees, which might explain some of their unique features such as the high apical dominance (Looney and Lane, 1984; Watanabe et al., 2004).

Aims

The aims of this project are to discover the original mutation that led to the columnar growth habit in apple and to identify the downstream genes and/or pathways that are affected by its presence. For the former, comparative genomic sequence analysis of A14 and P28 as well as of McIntosh and Wijcik will be applied to the *Co* target region based on genomic PCRs followed by Sanger and/or Illumina sequencing. A14 and P28 are selected since a progeny of A14 x P28 has previously been used to generate new molecular markers for the fine-mapping of *Co* and because P28 is the variety chosen for the construction of BAC libraries (Otto, 2013). McIntosh and Wijcik are chosen because they should be genetically identical except for the presence of the *Co* mutation. We will analyze not only the coding sequences as annotated by Velasco et al. (2010), but the whole region of interest. Genomic differences found between A14 and P28 may be associated with the columnar growth habit and genomic differences found between McIntosh and Wijcik should correspond to the original *Co* mutation.

For the identification of affected downstream pathways, RNA-Seq analyses based on Illumina sequencing of different tissues obtained from McIntosh and Wijcik as well as A14 and P28 trees will be conducted and the corresponding samples will be compared with regard to differential gene expression and other transcriptomic differences. The shoot apical meristem (SAM), leaves and primary roots are chosen for analysis as they represent the three main organs of a plant: shoot, leaf and root. Gene annotations in the *Co* region (Velasco et al., 2010) will be extended and refined and differential gene expression will be analyzed. In addition, differential expression of typical plant growth regulation pathway members in general such as genes involved in phytohormone synthesis and signal transduction are of major interest. RNA-Seq results on the structure and expression of relevant genes will be validated by PCRs using cDNA as a template and by qRT-PCRs.

This project will not only expand our knowledge of the apple genome and transcriptome but might also improve our understanding of plant growth regulation.

Results Summary

There Is No Mutation in a Protein Coding "Columnar Gene"

In order to identify the causal mutation that leads to the columnar growth habit, genomic sequences of A14 and P28 or McIntosh and Wijcik are compared with each other. The results of the comparative genomic examinations of the *Co* target region are described in detail in Paper 2 and Paper 3. An overview of the region(s) of interest, the genes annotated therein and the distinct analysis stages is given in Fig. 2.

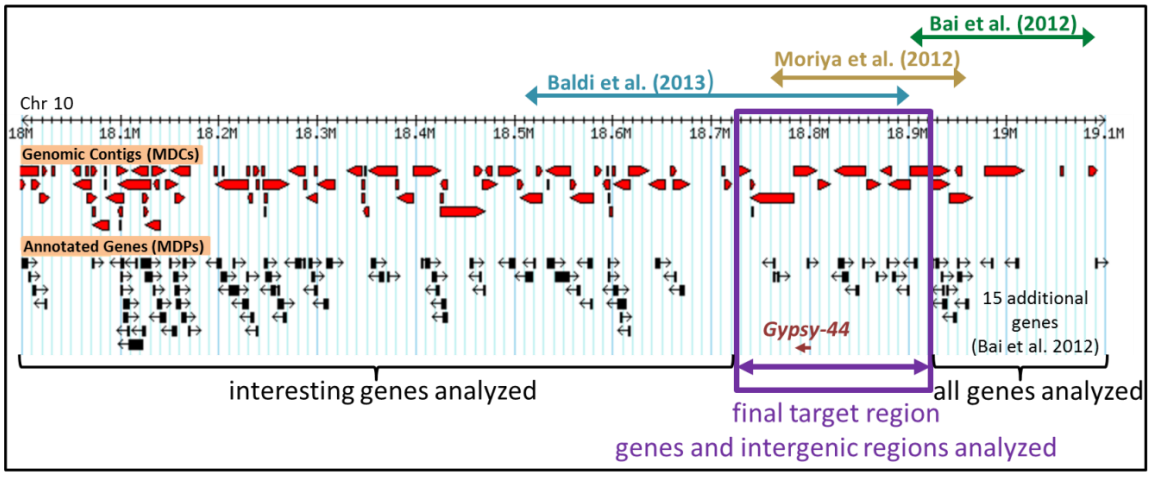


Figure 6: Overview of the *Co* target region(s).
The putative localization of the *Co* gene was first determined to be chromosome 10, 18.0 – 19.0 Mb, and was further delimited by marker studies of Bai et al. (2012), Moriya et al. (2012) and Baldi et al. (2013). In a view of the apple genome browser, genomic contigs (red arrows) and genes (black arrows) that have been annotated within this region are shown. Interesting genes were sequenced within the region 18.0 – 18.73 Mb, all protein coding genes were compared within the region 18.92 – 19.09 Mb, and the final target region of 18.73 – 18.92 Mb was exhaustively scanned for differences in genes as well as intergenic regions.

Paper 2 describes the analysis of the annotated gene regions. First the rough chromosomal localization of *Co* was appointed to about 18 – 19 Mb on chromosome 10 based on the publicly available markers $SCAR_{216}$ and $SCAR_{682}$ (Tian et al., 2005). BAC libraries of this region were generated within the scope of Dominik Otto's doctoral thesis (Otto, 2013). Following the hypothesis that a single dominant gene is responsible for columnar growth, the structures and sequences of all 98 annotated protein coding genes within this region as of April 1, 2011 (Velasco et al., 2010) were extracted, and BLAST (Basic Local Alignment Search Tool) (Altschul et al., 1990) searches against the SwissProt and The

Arabidopsis Information Resource (TAIR) (Lamesch et al., 2012) databases were conducted in order to determine their possible function. 61 genes matched to homologs that could play a role in plant growth regulation with regard to their function in transcriptional regulation, replication, chromatin modification, RNA or protein degradation, transport, membrane modification, protein-protein interactions, protein regulation or phytohormone biosynthesis and transport. Some of them did not yield any hits in the BLAST searches and were therefore chosen due to their unknown function. Primers were designed to span all annotated exons of these genes as well as 500 bp of upstream and downstream sequence in order to obtain coding regions and some regulatory sequences. Subsequently, PCRs were conducted using genomic DNA of A14 and P28 as a template. PCR products were either pooled and sequenced by Illumina or they were individually sequenced by the Sanger technology, and sequences of A14 and P28 were assembled and compared against each other and against the Golden Delicious reference genome. Since the apple genome is not well annotated, in an additional approach cDNA was generated from RNA of A14 and P28 *in vitro* cultures, and intron-exon structures were verified or revised based on PCRs to ensure correct interpretation of genomic SNPs, e.g. as missense mutations or mutations affecting splice sites.

Of the 61 genes investigated, 55 genes show differences in their nucleotide sequence between A14 and P28. Most variations are synonymous SNPs that most likely are not responsible for the development of the columnar phenotype. No nonsense mutations are present, which would have the most severe effect on the structure and function of the corresponding protein. However, 24 genes carry up to seven non-synonymous substitutions occurring heterozygously in P28, whereas A14 and Golden Delicious are homozygous. This would be in line with the genotype of these cultivars regarding the *Co* gene. Since non-synonymous SNPs might alter the protein structure and thus its function, these 24 genes were seen as potential candidates for *Co*.

The regions containing non-synonymous SNPs were re-sequenced in non-columnar McIntosh and its heterozygous columnar bud sport Wijcik because these two varieties should be genetically identical except for the *Co* mutation. For each region investigated, McIntosh and Wijcik carry the same alleles, and one of them represents the variation that is found heterozygously in P28 while being absent from A14 and Golden Delicious. This means that this is the allele present on the columnar Wijcik chromosome and it was handed down from Wijcik to Telamon and then to P28. It most likely segregates with *Co* due to its proximity to the *Co* mutation. However, because it is also present in non-columnar McIntosh, it cannot be associated with columnar growth.

At the time of these studies, two new marker papers tracking the *Co* target region were published (Bai et al., 2012; Moriya et al., 2012). Bai et al. (2012) delimited the putative *Co* location to 18.90 – 19.09 Mb on chromosome 10 and additionally assigned two contigs that had been unanchored by Velasco et al. (2010) to a gap between 19.0 and 19.1 Mb. Moriya et al. (2012) suggested 18.76 – 18.96 Mb to be the most likely localization for *Co*. All protein coding sequences annotated within the region 18.76 – 19.09 Mb that had not been included in our initial set of 61 genes were extracted along with the coding sequences of the two unanchored contigs, together totaling 24 genes. Exons and 500 bp of upstream and downstream sequences were sequenced in McIntosh and Wijcik and the intron-exon structure was determined in A14 and P28. The genomic sequences of all 24 genes were found to be 100 % identical in McIntosh and Wijcik. Therefore we concluded that the columnar growth habit did not originate from a mutation within a protein coding region or a regulatory region close to a protein coding region.

The *Co* Mutation Is a *Gypsy*-44 LTR Retrotransposon Insertion

Paper 3 describes our final steps towards the identification of the *Co* mutation. In this study we analyzed the whole genomic sequence rather than focusing on genes. The *Co* target region of a third new marker study by Baldi et al. (2012), 18.51 – 18.90 Mb, overlaps with the putative *Co* locus of Moriya et al. (2012) in the range of 18.76 – 18.90 Mb, which is therefore considered the most likely target region for *Co*. We used the newly generated P28 BAC metacontig (Otto, 2013) assigned to 18.73 – 18.92 Mb as a reference due to its higher integrity compared with the Golden Delicious genome sequence. 100 bp paired end Illumina sequencing runs of DNA extracted from McIntosh and Wijcik were conducted, and the reads were mapped against the BAC metacontig sequence. Subsequently, SNP detections, DIP (deletion-insertion-polymorphism) detections and structural variation detections were performed on these mappings and all variants were analyzed in McIntosh and Wijcik. Where necessary, e.g. due to low coverage of a region in at least one of the mappings, investigations were supported by genomic PCRs.

The mappings indicate no SNPs or DIPs that only occur in one variety and not in the other. However, there is one striking structural difference between McIntosh and Wijcik: a heterozygous 8200 bp insertion is present at position 18.79 Mb in Wijcik that does not occur in McIntosh (position indicated in Fig. 2). Sequence analyses and database searches against CENSOR (Kohany et al., 2006) identify it as a *Gypsy*-44 LTR retrotransposon containing identical LTRs with a length of 1951 bp, a conserved primer binding site and a

polypurine tract. It is flanked by a 5 bp TSD and inserted in the 3' to 5' orientation with regard to the genomic contig. Genomic PCRs using primers spanning either the left border or the right border or the whole retrotransposon show that this insertion occurs heterozygously in 13 heterozygous columnar cultivars tested, homozygously in one homozygous columnar cultivar examined, and is absent from seven non-columnar cultivars analyzed. Of particular strength are the results obtained with the cultivar 'Topaz'. 'Topaz' is phenotypically non-columnar, but shows the molecular marker genotype that is typical for the heterozygous columnar genotype when tested with several important published molecular markers that have been shown to be linked to the *Co* gene. This means that the non-columnar 'Topaz' has a genotype that is very similar or even identical to the non-columnar McIntosh in the *Co* target region. Only with the primers spanning the left and right borders of the *Gypsy*-44 insertion we were able to obtain the non-columnar pattern of bands for 'Topaz', which is in agreement with its non-columnar phenotype. They are therefore the only markers so far that can reliably detect the genotype with regard to the presence of *Co* independent of the cultivar used for investigation. Based on these results and the fact that it is the only genomic difference between McIntosh and Wijcik in the putative *Co* target region, this *Gypsy*-44 insertion is considered the original *Co* mutation.

Gypsy-44 is inserted into the 5' LTR of another LTR retrotransposon, *Gypsy*-33, in a gene-poor genomic region. The nested *Gypsy*-44-*Gypsy*-33 complex is not inserted in any known gene and thus it does not cause a simple "insertional inactivation" of an active gene. *Gypsy*-44 is transcribed, but it does not contain any transposon-specific ORFs and is thus probably a non-autonomous TE copy. However, several short ORFs of unknown function are present. The longest one (609 bp) is located in the antisense direction relative to *Gypsy*-44 and is transcribed at high levels. As genomic data do not provide a simple explanation why the *Gypsy*-44 insertion leads to the columnar phenotype, transcriptomic analyses were conducted on several tissues, and gene expression levels were compared in columnar and non-columnar varieties in order to detect any possible gene regulatory effects of *Gypsy*-44.

Tissue-specific Differential Gene Regulation in the *Gypsy*-44 Region

Within the scope of this thesis together with the doctoral thesis of Clemens Krost (Krost, 2012), 18 different RNA-Seq Illumina datasets were generated and analyzed with a focus on differential gene expression due to the *Gypsy*-44 insertion. Genes involved in plant growth regulatory pathways were of special interest. Table 1 shows a summary of the transcriptomic experiments, the material used and the publications describing their

analysis. Papers 4 – 6 deal with gene expression in SAMs of P28 compared with A14, with a special focus on phytohormone-associated genes in Paper 6. Paper 3 describes the overall differential gene expression in leaves of McIntosh and Wijcik. Paper 7 covers comparative transcriptome analyses in non-columnar, heterozygous columnar and homozygous columnar primary roots obtained from seeds extracted from P28 apples, which were fertilized by open pollination. In addition, Paper 7 compares gene expression within the *Gypsy*-44 vicinity across the different tissues.

Within our BAC metacontig containing the *Co* target region, eight genes were found to be expressed at levels high enough in at least one of the tissues to reliably define their intron-exon structure based on transcriptomic Illumina data and/or PCRs conducted using cDNA of A14 and P28 *in vitro* cultures as a template (Fig 3). For the analysis of differential expression, genes were annotated on the BAC metacontig genomic sequence, transcriptomic Illumina reads were mapped to the annotated genes and fold changes were calculated based on normalized read counts. In addition, expression levels were verified by qRT-PCR analysis.

The gene structure of three out of the eight genes, *MDP0000927098* (*ATL5K-like*), *MDP0000912172* (*PP2C15-like*) and *MDP0000163720* (*ACC1-like*), is in agreement with the annotation in the Golden Delicious genome. Another two genes, *MDP0000927091* (*Autophagy9-like*) and *MDP0000934866* (*At1g06150-like*), have an intron-exon structure that is different from the annotation as published for Golden Delicious. Three other genes have so far not been annotated in the reference genome and are therefore named after their homolog in the SwissProt/UniProtKB database: *At1g08530-like*, *dmr6-like* and *5NG4-like*. Two of the eight genes, *5NG4-like* and *At1g06150-like*, are present in two different isoforms, but these isoforms occur in columnar as well as in non-columnar samples.

Table 3: Summary of transcriptomic analyses.
RNA-Seq studies were conducted on shoot apical meristems (SAMs), leaves and primary roots (PRs). For the SAM data, linearly amplified mRNA was used (Krost 2012); for all other data, total RNA was sequenced. The datasets for Leaf 2 were obtained from an Illumina HiSeq2000 51 bp single-end sequencing run within the scope of an F1 practical and were later excluded from detailed analyses as explained in the text. All other data were generated by Illumina GA_{IIx}, HiSeq 2000 or HiSeq 2500 95 – 101 bp paired-end runs.

Dataset	Material of Origin	No. of Reads	Publication(s)	EBI SRA Accession No.
SAM 1 A14	shoot tips collected on May 22, 2009	80,289,084	Papers 4 – 7, Krost 2012	ERP000629
SAM 1 P28	shoot tips collected on September 29, 2009	76,177,216	Papers 4 – 7, Krost 2012	ERP000629
SAM 2 A14	shoot tips collected on May 22, 2009	44,134,596	Papers 5 – 7, Krost 2012	ERP000629
SAM 2 P28	shoot tips collected on May 22, 2009	67,515,702	Papers 5 – 7, Krost 2012	ERP000629
SAM 3 A14	shoot tips collected on July 20, 2010	58,616,798	Papers 5 – 7, Krost 2012	ERP000629
SAM 3 P28	shoot tips collected on July 20, 2010	162,540,416	Papers 5 – 7, Krost 2012	ERP000629
Leaf 1 McIntosh	young, fully developed leaf collected on May 22, 2012	177,766,172	Paper 3, Paper 7	ERP002576
Leaf 1 Wijcik	young, fully developed leaf collected on May 6, 2012	169,700,162	Paper 3, Paper 7	ERP002576
Leaf 2 McIntosh	adult leaf collected at unknown time	67,290,413	n/a	n/a
Leaf 2 Wijcik	adult leaf collected at unknown time	55,713,531	n/a	n/a
Leaf 3 McIntosh	young, fully developed leaf collected on May 22, 2012	49,524,567	Paper 7	ERP002576
Leaf 3 Wijcik	young, fully developed leaf collected on May 6, 2012	66,932,678	Paper 7	ERP002576
PR 1 non-columnar	one non-columnar radicle (about 3 cm) obtained from seeds of P28 apples	118,400,555	Paper 7	PRJEB6212
PR 1 heterozygous	one heterozygous radicle (about 3 cm) obtained from seeds of P28 apples	104,464,241	Paper 7	PRJEB6212
PR 1 homozygous columnar	one homozygous columnar radicle (about 3 cm) obtained from seeds of P28 apples	125,731,385	Paper 7	PRJEB6212
PRs 2 non-columnar	three pooled non-columnar radicles (about 3 cm) obtained from seeds of P28 apples	40,261,320	Paper 7	PRJEB6212
PRs 2 heterozygous	three pooled heterozygous radicles (about 3 cm) obtained from seeds of P28 apples	27,708,416	Paper 7	PRJEB6212
PRs 2 homozygous columnar	three pooled homozygous columnar radicles (about 3 cm) obtained from seeds of P28 apples	67,322,394	Paper 7	PRJEB6212

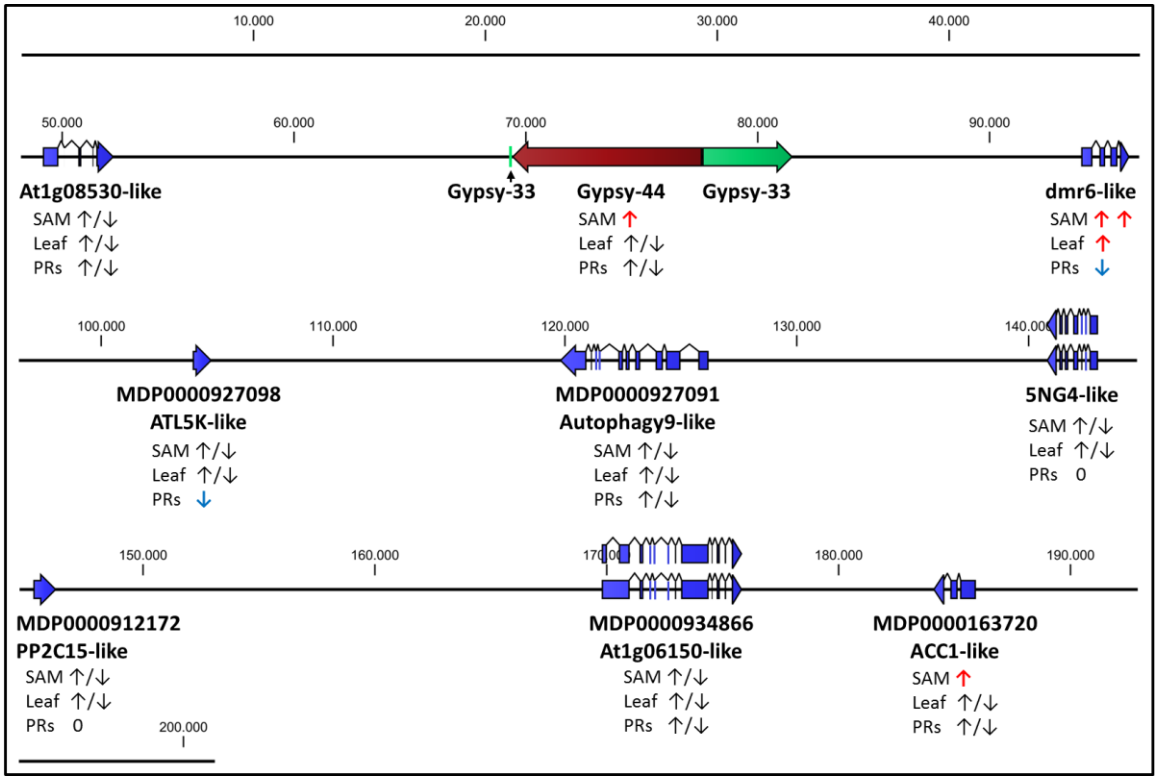


Figure 7: Gene structure and expression analysis in the final *Co* target region.
Within the region of the *Gypsy*-44 transposon insertion (nested into *Gypsy*-33) eight genes were found to be expressed and were reliably annotated on the BAC metacontig sequence. Intergenic regions are packed with TEs and TE fragments (Otto 2012), which have been omitted for reasons of clarity. Two genes, *5NG4-like* and *MDP0000934866* (*At1g06150-like*), are present in two different isoforms. Differential regulation of *Gypsy*-44 and the eight genes in columnar shoot apical meristems (SAM), leaves and primary roots (PRs) compared with non-columnar tissue as detected in the RNA-Seq and qRT-PCR experiments is depicted by arrows: ↑ = upregulation, ↑↑ = strong upregulation, ↓ = downregulation, ↑/↓ = no differential regulation, 0 = no expression.

Gypsy-44 itself is expressed at slightly higher levels in the SAMs of columnar trees than in those of non-columnar trees. While *At1g08530-like*, the gene upstream of *Gypsy*-44, shows no significant differential expression in any of the tissues analyzed, induction or repression of some of the genes downstream of *Gypsy*-44 does occur. The most striking example is the direct downstream "neighbor" of *Gypsy*-44, *dmr6-like*. *dmr6-like* is expressed at medium levels in primary roots and is repressed in heterozygous columnar primary roots and some samples of the homozygous columnar primary roots compared with non-columnar samples. In the fully developed leaves that were used for Illumina sequencing, *dmr6-like* is not transcribed, but in the very young apical leaves analyzed by qRT-PCR it does show low expression in non-columnar McIntosh and about 10-fold higher expression in columnar Wijcik. In the SAM, almost no *dmr6-like* transcripts are detected in non-columnar A14, whereas medium expression is found in columnar P28, leading to changes in the range of 30 – 200 fold across the different biological replicates of P28 compared with A14. The second-

next gene downstream of *Gypsy*-44, *MDP0000927098* (*ATL5K-like*), is also downregulated in columnar compared with non-columnar primary root RNA-Seq data. A third gene, *MDP0000163720* (*ACC1-like*), is upregulated in the columnar SAM. The other five genes within the *Co* target region have fold changes around 1 and/or show variation across biological replicates, suggesting that they are not significantly up- or downregulated in a specific columnar tissue. In conclusion, the presence of *Gypsy*-44 most likely influences the expression of *dmr6-like* and possibly *MDP0000927098* (*ATL5K-like*) and *MDP0000163720* (*ACC1-like*). The strongest effect is the induction of *dmr6-like* in SAMs. The tissue-specificity of the differential gene expression suggests involvement of one (or several) cis-regulatory regions that are bound by transcription factors only occurring in some tissues.

The Overall Gene Expression Pattern of Columnar Trees

For an overall comparison of transcriptomic data, Table 2 illustrates Pearson correlation coefficients of all libraries based on read counts obtained by RNA-Seq mappings against the Golden Delicious reference genome. In general, the read counts show a higher correlation within a given tissue than across tissues, and the expression patterns of SAMs and primary roots are more similar to each other than either of them is to leaves. The datasets for Leaf 2 (especially the Wijcik data) display a lower correlation with the other leaf datasets than the biological replicates of the other tissues display across each other. Due to this low correlation in addition to the short reads, an unclear material collection date and a high GC content pointing to high rRNA contamination, the datasets for Leaf 2 were omitted from downstream analyses. The SAM A14_3 dataset also shows low correlation with the other SAM datasets, possibly due to beginning degradation of the RNA sequenced (Paper 6).

Whole-transcriptome comparisons were conducted either based on BLAST searches against MDPs followed by collapsing to SwissProt descriptions (Papers 4 – 6) or based on mappings against the annotated Golden Delicious sequence (Papers 3 and 7). For statistical evaluation of differential gene expression, either the R_j value (Stekel et al., 2000) was used (Papers 4 and 3) or DESeq (Anders and Huber, 2010) was applied with default parameters (Papers 5 – 7). Visualization of differential gene expression was achieved in MapMan (Thimm et al., 2004), which assigns genes to functional groups (BINs) and depicts them in different colors depending on whether they are up- or downregulated.

Table4: Pearson correlation coefficients of Illumina datasets.
Pearson correlation coefficients were calculated based on read counts obtained from RNA-Seq mappings against the annotated Golden Delicious genome. Color shading indicates the degree of correlation (green = high, yellow = medium, red = low). nc: non-columnar, het: heterozygous columnar, hom: homozygous columnar

		SAM						Leaf						PRs					
		A14_1	A14_2	A14_3	P28_1	P28_2	P28_3	MCl_1	MCl_2	MCl_3	Wij_1	Wij_2	Wij_3	nc_1	nc_2	het_1	het_2	hom_1	hom_2
SAM	A14_1	1.00	0.83	0.29	0.78	0.80	0.70	0.15	0.04	0.11	0.15	0.00	0.13	0.46	0.51	0.46	0.51	0.52	0.44
	A14_2		1.00	0.22	0.72	0.84	0.69	0.13	0.04	0.09	0.14	0.00	0.11	0.44	0.49	0.44	0.48	0.50	0.42
	A14_3			1.00	0.16	0.23	0.38	0.03	0.01	0.02	0.03	0.00	0.02	0.13	0.10	0.09	0.09	0.11	0.10
	P28_1				1.00	0.74	0.67	0.11	0.04	0.09	0.13	0.00	0.10	0.46	0.50	0.45	0.50	0.51	0.44
	P28_2					1.00	0.79	0.15	0.04	0.11	0.15	0.01	0.13	0.49	0.55	0.50	0.55	0.56	0.49
	P28_3						1.00	0.14	0.04	0.10	0.14	0.01	0.12	0.44	0.45	0.40	0.44	0.45	0.42
Leaf	MCl_1							1.00	0.83	0.97	0.94	0.78	0.97	0.08	0.09	0.11	0.09	0.09	0.08
	MCl_2								1.00	0.94	0.86	0.98	0.88	0.06	0.08	0.13	0.08	0.08	0.08
	MCl_3									1.00	0.95	0.90	0.97	0.08	0.09	0.13	0.09	0.09	0.08
	Wij_1										1.00	0.80	0.95	0.12	0.11	0.13	0.12	0.12	0.11
	Wij_2											1.00	0.82	0.02	0.04	0.09	0.05	0.04	0.04
	Wij_3												1.00	0.09	0.10	0.13	0.10	0.10	0.09
PRs	nc_1													1.00	0.84	0.80	0.83	0.85	0.86
	nc_2														1.00	0.91	0.98	0.95	0.96
	het_1															1.00	0.93	0.93	0.89
	het_2																1.00	0.95	0.96
	hom_1																	1.00	0.92
	hom_2																		1.00

In Paper 4, the SAM 1 datasets are analyzed together with 454 data generated within the scope of Clemens Krost's thesis (Krost, 2012). In the Illumina data, genes playing a role in light reactions, mitochondrial electron transport, lipid metabolism, cell wall modification, chromatin structure, RNA processing and protein biosynthesis are downregulated in columnar P28 compared with non-columnar A14. Genes involved in cell wall modification, transport and protein modification as well as terpenoid and tryptophan biosynthesis are upregulated in P28. With regard to phytohormones, genes associated with the synthesis of precursors and signal transduction of auxin (indole-3 acetic acid, IAA) and jasmonates are upregulated, whereas genes associated with gibberellic acid signal transduction are downregulated in columnar SAMs. Expression levels of different cyclin genes suggest a cell cycle arrest in G2.

In Paper 5, the analysis is extended to three datasets each of A14 and P28 SAMs, which are considered biological replicates. Genes of the categories biotic stress, protein modification and degradation, cell wall degradation, amino acid metabolism, regulation of transcription, mitochondrial electron transport and transport proteins are differentially regulated in P28. Genes associated with IAA, cytokinin, jasmonates and brassinosteroids biosynthesis and signal transduction are upregulated in P28, whereas genes associated with gibberellic acid synthesis and signaling are downregulated.

Paper 6 focuses on RNA-Seq results of phytohormone-associated genes and validates their expression levels with microarray data of four collection dates and qRT-PCRs based on

RNA isolated from *in vitro* cultures of A14 and P28. Genes encoding enzymes that inactivate IAAs are downregulated in P28, whereas IAA transporters are upregulated, probably leading to increased auxin activity and transport. Genes encoding regulators and/or signal transduction components of phytohormones support the hypotheses of higher levels of cytokinins, brassinosteroids and jasmonates and lower levels of polar gibberellins in columnar SAMs. Furthermore, the detection of a statistically approved enrichment of differentially regulated genes on chromosome 10 is in line with the hypothesis that the *Gypsy*-44 insertion influences expression levels of nearby genes.

Paper 3 describes the pattern of differential gene expression in leaves of McIntosh and Wijcik. Genes involved in photosynthesis, protein biosynthesis and nucleotide metabolism are repressed, genes associated with lignin and terpenoid biosynthesis are strongly induced in Wijcik compared with McIntosh. Genes with a role in IAA, jasmonate and ethylene signaling and/or biosynthesis as well as almost all genes encoding proteins involved in defense or stress reactions such as pathogen recognition receptor and heat shock proteins are highly upregulated in Wijcik, whereas genes encoding enzymes for redox state regulation are downregulated.

In Paper 7, the subject of investigation is the third major organ of the plant – the root. Genes assigned to cell wall modification and degradation, cellulose synthesis and phenylpropanoid biosynthesis (the precursors of lignin) are upregulated in heterozygous and homozygous columnar primary roots compared with non-columnar primary roots and are more strongly induced in homozygous columnar than in heterozygous columnar primary roots. With regard to genes involved in stress reactions, ethylene and jasmonate-associated genes are upregulated in the columnar samples – an effect that is even stronger in homozygous columnar than in heterozygous columnar radicles. Very few genes show downregulation in columnar primary roots, and except for single genes involved in lipid degradation these are not differentially regulated in both heterozygous and homozygous columnar samples.

In conclusion, columnar trees show altered gene expression in all tissues investigated, and some features of their gene expression profiles are similar across different tissues, while others are tissue-specific. In columnar SAMs, changes in cell membrane and cell wall biosynthesis and in transport as well as elevated levels of IAA, cytokinins, brassinosteroids and jasmonates and reduced levels of gibberellins can be inferred from differential gene expression results. Leaves of columnar trees seem to suffer from stress and show alterations in photosynthesis. Heterozygous and homozygous primary roots share the alterations in cell wall metabolism with SAMs and the upregulation of jasmonate and ethylene associated genes with SAMs and leaves.

General Discussion

The Molecular Basis of Columnar Growth

The results obtained within the scope of this thesis have significantly contributed to the understanding of the molecular processes leading to the formation of the columnar growth habit in apple trees. First, the original *Co* mutation was identified as a *Gypsy*-44 LTR retrotransposon insertion on chromosome 10 at 18.79 Mb. Second, tissue-specific differential expression of the genes within the retrotransposon vicinity was discovered, pointing towards an influence of the *Gypsy*-44 insertion on its downstream genes. Third, members of some regulatory pathways such as IAA and jasmonate metabolism as well as stress-related signaling were found to have an altered expression pattern in columnar trees, most likely due to the differential gene activity in the *Co* region. The final proof for the event cascade elucidated here would be the site-specific integration of *Gypsy*-44 into non-columnar varieties or its excision in columnar varieties followed by phenotype, gene expression and phytohormone level analyses of the genetically modified plants. However, the transformation or excision of a nested retrotransposon provides great challenges or might even be impossible at this time. Verifying the influence of differential gene expression in the *Co* target region by overexpression or knock-out/knock-down of the *Gypsy*-44 and/or *dmr6-like* transcripts might be feasible though.

At the time of our studies, Wolters et al. (2013) published similar results, which they had obtained by constructing BAC libraries of McIntosh and Wijcik and comparing the BAC metacontig sequences. They found an insertion at 18.79 Mb on chromosome 10 in Wijcik, which they did not identify as the *Gypsy*-44 solo-LTR with identical TSDs, to be the only genomic difference between McIntosh and Wijcik within their target region (18.51 – 18.90 Mb). The position they identified corresponds to the locus at which we detected the whole *Gypsy*-44 retrotransposon. It is therefore most likely that recombination between the *Gypsy*-44 LTRs occurred either in their BAC clones or in the plant material they used. Wolters et al. (2013) investigated the expression levels of genes in the vicinity of the insertion by qRT-PCR in axillary meristems of McIntosh and Wijcik, which are responsible for the formation of fruit spurs instead of long lateral branches in columnar trees. They found strong upregulation of *dmr6-like* in Wijcik and the columnar progeny of a cross Golden Delicious x Wijcik, but not in the non-columnar progeny. Overexpression of *dmr6-like* in *Arabidopsis* resulted in plants with shorter internodes and smaller branches. Results of its overexpression in apple can be expected in the near future. Based on the exceptionally

strong induction of *dmr6-like* in SAMs and young leaves of columnar trees observed in our studies, combined with the observations of Wolters et al. (2013), it can be concluded that whereas the *Gypsy*-44 insertion is the *Co* mutation, *dmr6-like* probably is the most important causal gene of columnar growth having a dominant effect due to its higher expression in columnar trees. In addition, *MDP0000927098* (*ATL5K-like*) and *MDP0000163720* (*ACC1-like*), whose expression levels also seem to be changed due to the retrotransposon insertion, might represent two of the modifier genes hypothesized by Lapins (1976). Despite these advances, two questions still remain: how does the *Gypsy*-44 insertion induce changes in the expression of *dmr6-like* and other nearby genes? And what is the link between differential expression of genes in the *Co* region and the observed downstream effects on the transcriptome profile?

Taken together, the genomic and transcriptomic data allow sketching a speculative model to answer the first question, assuming the presence of two cis-regulatory regions upstream of *dmr6-like*, which can bind transcriptional activators (Fig. 4A). The proximal cis-regulatory region for *dmr6-like* is located at least 17 kb upstream of the transcriptional start site, whereas the distal cis-regulatory region is positioned even further upstream and too far away to influence the promoter in non-columnar apple trees. In developing aerial tissues such as the SAM, axillary meristems and young leaves of non-columnar apple trees, the proximal cis-regulatory region is not bound by any transcription factors. A transcriptional activator specifically present in these organs does bind to the distal regulatory region, but its location is too far away to enable interaction with the general transcription factors at the *dmr6-like* promoter. Therefore, no transcription or only leaky transcription of *dmr6-like* occurs in developing aerial tissue of non-columnar apple trees. After the insertion of the *Gypsy*-44 retrotransposon between *dmr6-like* and its proximal regulatory region, on the columnar chromosome the distal regulatory region is located at spatial proximity of *dmr6-like*, e.g. via looping of the *Gypsy*-44 section. Therefore the transcriptional activator bound to it can interact with the *dmr6-like* promoter, enabling expression of *dmr6-like* in SAMs and young developing leaves of columnar trees. Since *MDP0000163720 (ACC1-like)* also experiences upregulation, the transcriptional activator present at the distal regulatory region might work over long distances. In primary roots, the proximal regulatory region of *dmr6-like* is bound by a root-specific transcriptional activator and therefore *dmr6-like* is expressed at high levels in non-columnar primary roots. This induction is more efficient than the induction by a transcriptional activator at the proximal regulatory region, so that the expression level of *dmr6-like* in non-columnar primary roots is higher than in columnar SAMs and columnar young leaves. When *Gypsy*-44 is inserted between *dmr6-like* and the proximal regulatory region, the root-specific transcriptional

activator is moved further away from the promoter, thus weakening its effect and subsequently reducing the expression level of *dmr6-like* in columnar primary roots. This effect might also be exerted upon *MDP0000927098* (*ATL5K-like*). In old leaves of non-columnar varieties, no expression or very low expression of *dmr6-like* takes place, which does not change upon insertion of *Gypsy*-44. Furthermore, there are almost no differences regarding the expression levels of genes in the *Gypsy*-44 vicinity between columnar and non-columnar old leaves. This implies that in old leaves, probably neither of the regulatory regions is occupied by any transcription factors at all.

The model described here would work almost as well assuming the binding of transcriptional repressors instead of transcriptional activators to the cis-regulatory regions (Fig. 4B). In this case, a transcriptional repressor present at the proximal regulatory region in aerial developing tissues would hinder *dmr6-like* expression in non-columnar varieties, and its effect would be reduced in columnar varieties due to its greater distance to the promoter. In primary roots, a repressor binding to the distal regulatory region would allow *dmr6-like* transcription in non-columnar roots, but not in columnar radicles, where it would be able to interact with the promoter owing to DNA looping. However, in this scenario it would be difficult to explain the reaction in leaves, where almost no expression is detected in columnar as well as non-columnar varieties.

A third model (Fig. 4C) would require a tissue-specific role of one of the regulatory regions. For example, the proximal region could function as a silencer in aerial developing tissues, with the repressor binding to it having a weaker effect at a greater distance from the gene in columnar varieties, leading to derepression of *dmr6-like* in developing columnar organs. By contrast, the proximal region would work as an enhancer in primary roots, and the activator bound to it would have a lesser effect at a greater distance in columnar trees, causing downregulation of *dmr6-like* in columnar primary roots. However, this option does not offer an explanation for the absence of *dmr6-like* transcription in old leaves of columnar as well as non-columnar trees, either. Therefore, the first model, which employs binding of transcriptional activators only, seems most plausible.

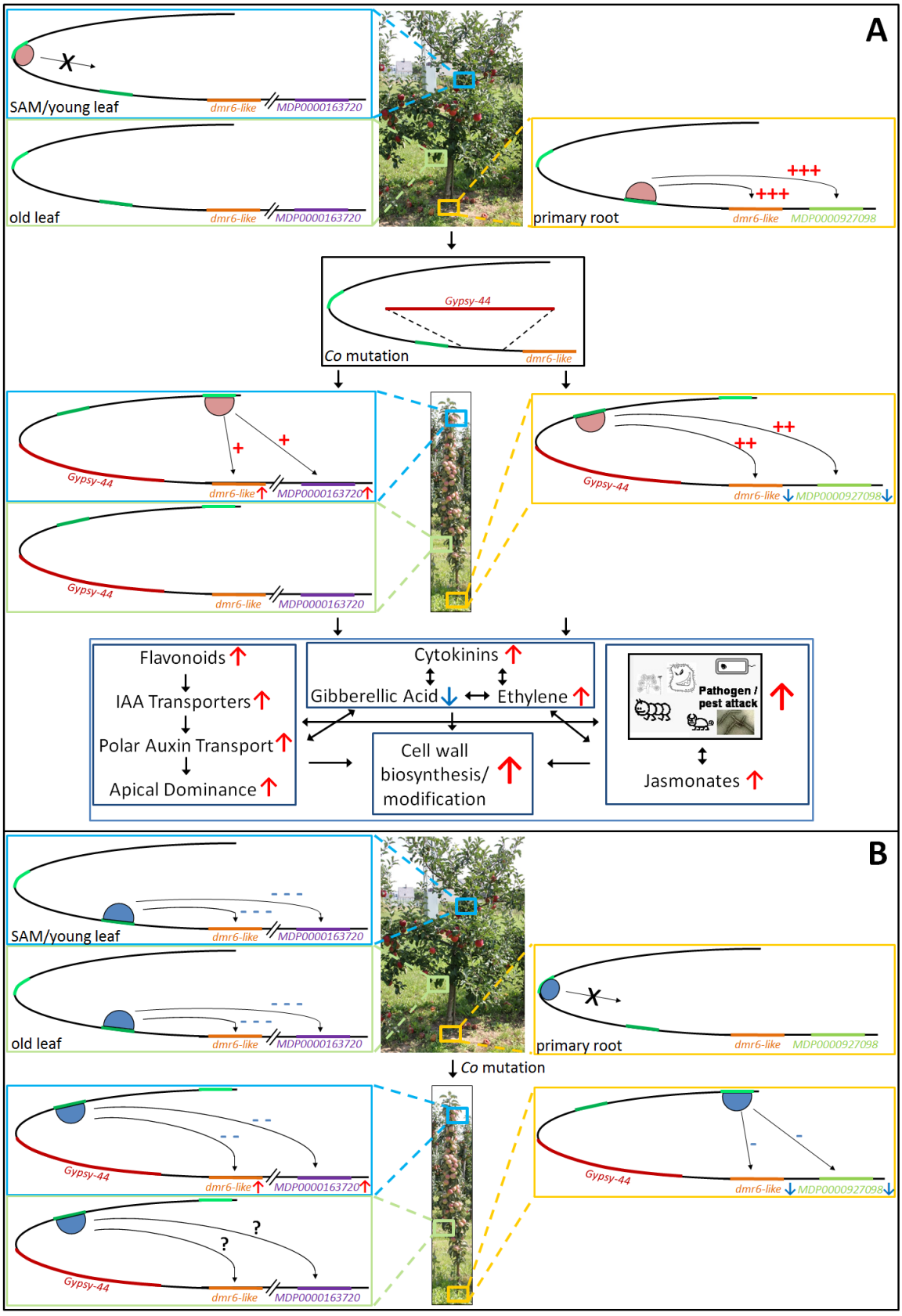


Figure 4: The event cascade leading to columnar growth in apple.

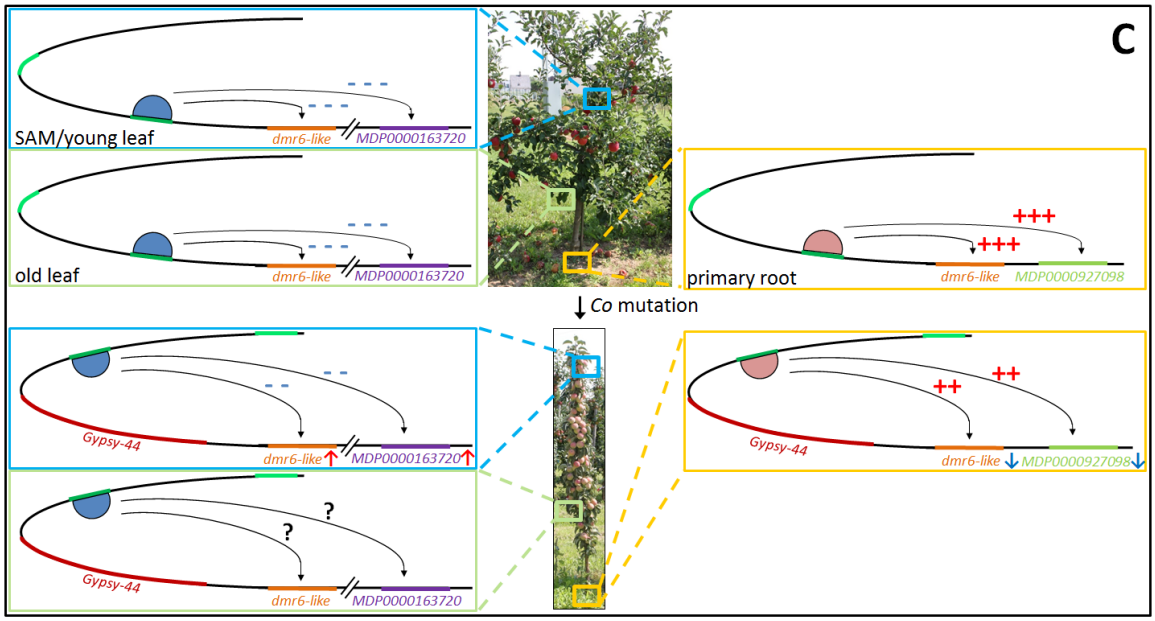


Figure 4 continued: The event cascade leading to columnar growth in apple.
Hypotheses of the molecular events leading to the columnar growth habit based on the presence of tissue-specific transcriptional activators (**A**), repressors (**B**) or both (**C**). The *Co* region harboring the genes *dmr6-like* (orange), *MDP0000163720* (violet) and *MDP0000927098* (olive), a proximal regulatory region (dark green) and a distal regulatory region (pale green) is shown as a DNA loop in the SAM (left, blue box), old leaves (left, olive box) and in the primary root (right, yellow box). Binding of tissue-specific transcriptional activators (pink half circles) has an activating effect (indicated by plus signs), binding of repressors (blue half circles) has an inhibiting effect (indicated by minus signs) on gene expression. After the insertion of *Gypsy*-44 (the *Co* mutation, red), on the columnar chromosome transcription factors are moved within reach or out of reach of the promoters leading to upregulation (↑) or downregulation (↓) of gene expression in columnar tissues. Secondary effects on gene expressions of phytohormone-associated, cell wall-associated and defense-related genes are only shown in (**A**). Detailed explanations of the individual models can be found in the text.

The idea of one or several regulatory regions playing a pivotal role in phenotype formation is supported by the discovery of a candidate cis-regulatory region 40 kb upstream of *dmr6-like* that is highly conserved across different Rosaceae species (Paper 7). Additional regulatory regions could be identified by DNase I hypersensitivity assays (Wu, 1980; Wu et al., 1979), DNase-Seq experiments (Crawford et al., 2006; Zhang et al., 2012) or formaldehyde-assisted isolation of regulatory elements (FAIRE) (Giresi et al., 2007; Omidbakhshfard et al., 2014). Alternatively, Chromatin Immunoprecipitation sequencing (ChIP-Seq) experiments could be carried out using antibodies against enhancer-specific histone modifications such as monomethylation of lysine 4 on histone 4 (H4K4me1) (Mikkelsen et al., 2007). In maize, evidence for the influence of TE insertions on gene regulation by distant upstream control element has been found: the insertion of a MITE element into the conserved non-coding sequence locus *vegetative to generative transition 1* (*vgt1*) roughly 70 kb upstream of the AP2 transcription factor *ZmRap2.7* leads to

upregulation of *ZmRap2.7* expression, resulting in a late flowering phenotype (Salvi et al., 2007). Likewise, the integration of a *Hopscotch* retrotransposon 60 kb upstream of *teosinte branched 1* (*tb1*) enhances expression of *tb1*, causing a considerable reduction of branching in maize compared to its wild progenitor teosinte (Studer et al., 2011). In these studies, the causal relationship between the TE insertion and the transcriptional upregulation was elucidated by allele-specific expression differences of *ZmRap2.7* in hybrid lines (Salvi et al., 2007) and by changes in gene expression observed in protoplast transformations with reporter constructs containing different parts of the upstream regulatory region of *tb1* (Studer et al., 2011). Similar approaches could be undertaken in apple to prove that the *Gypsy*-44 insertion causes the differential expression of *dmr6-like*. As the *dmr6-like* alleles on the columnar and the non-columnar chromosome of Wijcik and P28 are identical, a second allele of *dmr6-like* suitable for the investigation of allele-specific expression would have to be created by targeted mutation or might be found in screenings of other heterozygous columnar trees. Alternatively, reporter constructs containing different parts of the *dmr6-like* upstream region could be used in transformation experiments.

Even the first model proposed here still has two major drawbacks. First, while chromatin looping is a well-established model to explain promoter-enhancer interactions (Carter et al., 2002; Kolovos et al., 2012; Ptashne and Gann, 1997; Tolhuis et al., 2002), it seems unlikely that a regulatory region can interact with a promoter when present at a high distance (measured in basepairs), but has no effect at a low distance. However, it has been shown that a few base pairs spacing between the promoter and the enhancer can be required for enhancer activity (Chen and Yoshimura, 1998), and that specific interjacent sequences and chromatin states facilitate DNA looping, thereby modulating the interaction of promoters and enhancers (Doyle et al., 2014; Li et al., 2006; Nolis et al., 2009; Ringrose et al., 1999). For example, the IFN-β enhancer is non-functional in synthetic enhancer-promoter constructs at a distance of > 560 bp, but DNA looping and thus transcriptional activity are restored when a heterologous transcription factor binding site is inserted at the promoter (Nolis et al., 2009). Second, the model does not explain why the expression level of *dmr6-like* in homozygous columnar primary roots is approximately equal to that observed in non-columnar primary roots. The silencing effect of a transcriptional repressor on both chromosomes in the homozygous columnar radicles would actually suggest that the expression levels are even more drastically reduced in homozygous columnar than in heterozygous columnar primary roots, as can be seen in the case of *MDP0000927098* (*ATL5K-like*). For *dmr6-like*, the influence of one (or several) other regulator(s) is conceivable. Since changes in the expression level of *dmr6-like* seem to have tremendous effects on the overall plant growth habit, it is likely that its expression is fine-tuned by a

complex interplay of several transcription factors. In addition, regulatory regions might be present within the *Gypsy-44* sequence itself, and these might be bound by insulators, activators or repressors, modulating the expression patterns of nearby genes (Gause et al., 2001; Hily et al., 2009; Xie et al., 2013).

Furthermore, epigenetic changes might occur due to the presence of *Gypsy*-44. Usually organisms try to limit the transcriptional activity of retrotransposons in order to avoid excessive jumping, which might have deleterious effects for the host. This can be achieved either by elimination of the retrotransposon via recombination, by generating antisense transcripts against its genes, or by methylation of the retrotransposon sequence. This methylation can stretch out over neighboring regions, leading to the silencing of nearby genes (Ahmed et al., 2011; Candaele et al., 2014; Eichten et al., 2012). On the other hand, it has recently been shown that tissue-specific hypomethylation of TE sequences can lead to the formation of enhancer marks such as H3K4me1 and enable the TEs to activate transcription of nearby genes (Xie et al., 2013). If this is the case for *Gypsy*-44, it would explain its high transcriptional activity and the upregulation of *dmr6-like* in SAMs. Even if differential methylation was not the causal link between the retrotransposon insertion and upregulation of *dmr6-like*, its modulatory effect might still contribute to the differences observed in the penetrance of columnar growth evident by the formation of intermediate growth types (Baldi et al., 2012; Hemmat et al., 1997; Ikase and Dumbravs, 2004). Therefore, the DNA methylation pattern within the *Co* target region in A14 and P28 was compared, within the scope of a master thesis (Langhanki, 2014), by MspI/HpaII digestion followed by Southern Blotting. The Southern Blots did show differences between A14 and P28 that provide strong evidence for an altered methylation pattern in P28 compared with A14. However, this might reflect cultivar-specific differences rather than effects of the retrotransposon insertion. Therefore, these results need to be validated in a comparison of McIntosh and Wijcik instead of A14 and P28. Furthermore, changes in the histone modification pattern are an important mechanism for the control of gene expression and have not been investigated at all. Thus, it would be worthwhile to conduct in-depth analyses of epigenetic changes within the *Co* target region in future.

Besides the altered expression of genes within the *Gypsy*-44 vicinity, *Gypsy-44* itself and/or other similar or identical transposons show upregulation in columnar shoot apical meristems, which means that an increased number of retrotransposon transcripts is present in columnar varieties. The transcription of LTR retrotransposons, at least of the *Copia* family, can be induced by various biotic and abiotic stresses because some of them carry promoter sequences related to those of defense genes (Wessler, 1996). Therefore, the high transcriptional activity of *Gypsy*-44 might be the consequence rather than the cause of the

upregulation of stress-associated genes in columnar apple trees. On the other hand, *Gypsy*-44 does carry several short ORFs of obscure function of which at least one is actively transcribed. Thus, understanding the role of these transcripts might be one of the keys to elucidate how the genotype determines the phenotype.

With regard to the question of the downstream effects observed in the whole-transcriptome analyses, the differentially regulated genes within the *Co* target region most likely cause the observed changes in phytohormone levels, which then converge on the formation of the unique growth habit (Fig. 4A). In apple, the function of *dmr6-like*, *ATL5K-like* and *ACC1-like* has not been investigated yet. *Arabidopsis* plants carrying a null mutation of *dmr6* show enhanced expression of a subset of defense-associated genes (van Damme et al., 2008). Therefore, the upregulation of *dmr6-like* in columnar SAMs most likely leads to differential regulation of defense- and stress-associated genes such as those for jasmonate biosynthesis. 2OG-Fe(II) oxygenases have also been shown to play a role in flavonoid, ethylene and gibberellin biosynthesis (Prescott and John, 1996), and flavonoids modulate polar auxin transport (Brown et al., 2001; Peer and Murphy, 2007), so that upregulation of *dmr6-like* might directly influence phytohormone concentrations. Furthermore, triggered by the defense response or by an increase in flavonoid biosynthesis, more IAA transporters are generated, which increases polar auxin transport and in turn causes higher apical dominance in columnar trees. In addition, gibberellins are downregulated and cytokinins are upregulated, possibly via cross-regulatory effects of altered phytohormone levels (El-Showk et al., 2013; Weiss and Ori, 2007). Alterations in genes encoding members of the cell wall and cell membrane metabolism observed in SAMs and primary roots also seem to be characteristic for columnar tissues and might lead to a decrease in cell size. These hypotheses could be tested by further in-depth transcriptome analyses, proteomic studies and detailed phytohormone and flavonoid measurements.

Besides phytohormones, microRNAs (miRNAs) play a major role in plant growth regulation (reviewed in Jones-Rhoades et al. (2006), Jin et al. (2013) and Wu (2013)). None of the miRNAs identified in apple so far (Gleave et al., 2007; Ma et al., 2014; Varkonyi-Gasic et al., 2010; Xia et al., 2012; H. Yu et al., 2011) are located within the *Co* target region. However, a miRNA annotated in *Arabidopsis lyrata* can be mapped to the apple BAC metacontig 400 bp downstream of *ATL5K-like* with two mismatches. Hence, this locus might also produce a miRNA precursor in apple, which might suffer changes in its expression due to the *Gypsy*-44 insertion. The function of this miRNA is unknown, and no transcripts of this locus were detected within our transcriptomic Illumina data. However, for the detection of miRNA transcripts the protocol for Illumina library preparation needs to be adjusted in

order to include small transcripts. In future, comparative small RNA sequencing experiments of McIntosh and Wijcik might deliver interesting results.

Choice of Plant Material and RNA-Seq Analysis Methods

The choice of plant material and analysis methods is crucial for the correct interpretation of results, especially in the case of RNA-Seq studies. For large parts of this thesis, the material used was not ideal. In general, we used *in vivo* material and varieties different from McIntosh and Wijcik because we wanted to conduct our investigations under natural conditions and in varieties that are used for breeding and crossing. While this makes sense from a breeder's point of view, it might not always be the best choice for a geneticist. In the initial genomic studies, many variations between A14 and P28 were detected, which all turned out to be uncorrelated with columnar growth when subjected to validation in McIntosh and Wijcik. For the SAM RNA-Seq analyses, A14 and P28 samples were used, which have a different genetic background, so that it is always doubtful whether the changes in gene expression detected are due to the presence of the *Co* mutation or to other genotypic differences. In case of the primary roots, the seeds they emerged from were obtained from apples that had been subjected to open pollination so that the male parent is unknown, raising the same problem as for the SAM data. Only the leaves were harvested from McIntosh and McIntosh Wijcik. However, the leaves subjected to Illumina sequencing were obtained on different days so that gene expression changes might be due to their different developmental stage or altered environmental conditions.

Investigations of differential gene expression of *in vivo* plant material generally provide great challenges because growth conditions can never be kept completely constant in a natural environment. This means that two trees whose transcriptome profiles are examined will probably have experienced altered availability of nutrients, water or light and have been subjected to different interactions with other organisms such as pathogens, pests or symbionts. Additionally, pesticides are applied to all trees whose fruits are intended for consumption, and the day of pesticide sprayings are not always the same for each cultivar. All these changing parameters might have an impact on gene expression, thereby limiting the explanatory power of RNA-Seq results with regard to one specific factor such as columnar growth. However, only in vivo material allows an insight into the processes that actually occur in trees on the field. There are two solutions to this problem: either the growth conditions must be kept as similar as possible or one must investigate a number of

samples large enough so that all effects except the differences in growth habits eventually lose their statistical significance.

We used the first option for qRT-PCR experiments in Paper 6, resorting to *in vitro* cultures of A14 and P28, which can be grown under very steady conditions. However, *in vitro* cultures do not reflect natural conditions so that the observed gene expression changes do not necessarily occur *in vivo*. For instance, the methylation pattern of specific genes of *in vitro* cultures differs from that of field-grown plants (Li et al., 2002). Besides, *in vitro* cultures of A14 and P28 still possess different genetic backgrounds. Therefore, the ideal plant material for RNA-Seq studies would be potted plants of McIntosh and McIntosh Wijcik grown in a greenhouse under equal conditions and then used for time course experiments. However, since seeds obtained from McIntosh and Wijcik apples differ in their genotype, grafting onto the same rootstock would be necessary. This would then allow the examination of gene expression levels in leaves and shoots, but not in own roots or seedlings. Furthermore, the branches used for grafting again might have been subjected to different environmental influences, some of which are unavoidable owing to the differences in growth habits, e.g. non-columnar trees being more prone to self-shading than columnar trees. In conclusion, completely equal conditions can never be reached.

Where material grown under the same conditions is unavailable, it is possible to "cancel out" differences observed due to influences other than the growth habit by sampling a large number of replicates and only taking into consideration those changes that are observed across all of them. However, such a high number of data can be difficult to analyze and interpret and necessitates advanced statistics. We pooled about 10 – 15 shoot tips, three primary roots and several *in vitro* cultures prior to RNA isolation for RNA-Seq and qRT-PCR experiments. In addition, we used two to three biological replicates for our RNA-Seq data analysis and another two replicates for qRT-PCR analyses. We found many genes to be consistently up- or downregulated across different materials and assay techniques, indicating high specificity. However, it would be interesting to expand the number of replicates and observe how gene expression differences behave.

With regard to the computational aspects of the project, varying analysis pipelines for RNA-Seq studies were employed. We preferred the mapping approach over assemblies because we were specifically interested in read counting rather than transcript detection and annotation, and the quantification was shown to be more consistent among replicates when raw reads rather than contigs were used (Paper 4). For assigning the raw reads to specific genes, either BLAST searches or mappings against the apple genome were used. Both delivered similar numbers of matches, but mappings confer the advantage that the result can be directly checked regarding the position assigned to a read without an extra

step of sequence extraction and assembling. For the evaluation of differential gene expression, two different methods were used: either calculation of R_j values and ranking according to them (Stekel et al., 2000), or DESeq analysis (Anders and Huber, 2010). The R_j calculation is based on a LOG likelihood statistic, asymptotically tending towards a χ^2 distribution (Stekel et al., 2000). By contrast, DESeq conducts a negative binomial test (Anders and Huber, 2010). DESeq has been shown to perform well for the analysis of differential gene expression (Kvam et al., 2012) and has been widely applied to transcriptome analyses in humans, animals, yeast, bacteria and plants (see, for example, Busby et al. (2011), Visconte et al. (2012), Champigny et al. (2013), Samborski et al. (2013), Shrestha et al. (2013), Shearman et al. (2013), Straub et al. (2013)). In addition, it can take biological replicates into consideration, which is why we preferred this tool for our later analyses. Our results were compared on two different levels of gene annotation: either the MDP level so that we were able to distinguish the expression of each individual gene, or collapsed SwissProt descriptions, a method that looks at effects rather than individual genes. For instance, the apple has 13 copies of phosphofructokinase, and each paralog can be individually regulated, one of them being upregulated and one downregulated. This can be biologically meaningful because the two paralogs might have slightly different functions or might be regulated by different factors. However, if the main question is whether the process or the outcome of its reaction in glycolysis is up-or downregulated, it makes more sense to compare the expression of all phosphofructokinase-encoding genes taken together. Therefore, we consider collapsing a suitable tool to get an overall idea of the transcriptome profile, but detailed analyses of individual genes such as those within the *Co* target region have a higher informative value. For the visualization of differential gene expression in MapMan, first the poplar mapping was used, whereas later a mapping file for apple became available and was thus favored. All in all, we consider our data to be comprehensive and statistically sound enough to enable reliable interpretation with regard to the molecular causes of columnar growth.

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., Venter, J.C., 1991. Complementary DNA Sequencing: Expressed Sequence Tags and Human Genome Project. Science 252, 1651–1656.

Ahmed, I., Sarazin, A., Bowler, C., Colot, V., Quesneville, H., 2011. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in *Arabidopsis*. Nucleic Acids Res. 39, 6919–6931.

Allan, A.C., Crowhurst, R., Gleave, A., Newcomb, R., Schaffer, R., 2009. Apple Functional Genomics, in: Folta, K.M., Gardiner, S.E. (Eds.), Genetics and Genomics of Rosaceae, pp. 121–142.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic Local Alignment Search Tool. J. Mol. Biol. 215, 403–410.

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

Arumuganathan, K., Earle, E.D., 1991. Nuclear DNA Content of Some Important Plant Species. Plant Mol. Biol. Report. 9, 208–218.

Bai, T., Zhu, Y., Fernández-Fernández, F., Keulemans, J., Brown, S., Xu, K., 2012. Fine Genetic Mapping of the *Co* Locus Controlling Columnar Growth Habit in Apple. Mol. Genet. Genomics 287, 437–450.

Baldi, P., Wolters, P.J., Komjanc, M., Viola, R., Velasco, R., Salvi, S., 2012. Genetic and physical characterisation of the locus controlling columnar habit in apple (*Malus* × *domestica* Borkh.). Mol. Breed. 31, 429–440.

Baldo, A., Norelli, J.L., Farrell Jr, R.E., Bassett, C.L., Aldwinckle, H.S., Malnoy, M., 2010. Identification of genes differentially expressed during interaction of resistant and susceptible apple cultivars (*Malus* × *domestica*) with *Erwinia amylovora*. BMC Plant Biol. 10, 1.

Bennetzen, J.L., 1996. The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol. 4, 347–353.

Bennetzen, J.L., 2005. Transposable elements, gene creation and genome rearrangement in flowering plants. Curr. Opin. Genet. Dev. 15, 621–627.

Bennetzen, J.L., Kellogg, E.A., 1997. Do Plants Have a One-Way Ticket to Genomic Obesity? Plant Cell 9, 1509–1514.

Boyer, J., Liu, R.H., 2004. Apple phytochemicals and their health benefits. Nutr. J. 3, 5.

Breyne, P., Dreesen, R., Cannoot, B., Rombaut, D., Vandepoele, K., Rombauts, S., Vanderhaeghen, R., Inzé, D., Zabeau, M., 2003. Quantitative cDNA-AFLP analysis for genome-wide expression studies. Mol. Genet. Genomics 269, 173–179.

Brown, D.E., Rashotte, A.M., Murphy, A.S., Normanly, J., Tague, B.W., Peer, W.A., Taiz, L., Muday, G.K. 2001. Flavonoids Act as Negative Regulators of Auxin Transport *in Vivo* in *Arabidopsis*. Plant Physiol. 126, 524-535.

Brunner, S., Fengler, K., Morgante, M., Tingey, S., Rafalski, A., 2005. Evolution of DNA Sequence Nonhomologies among Maize Inbreds. Plant Cell 17, 343–360.

Bureau, T.E., Wessler, S.R., 1992. Tourist: a Large Family of Small Inverted Repeat Elements Frequently Associated with Maize Genes. Plant Cell 4, 1283–1294.

Bureau, T.E., Wessler, S.R., 1994. Stowaway: a New Family of Inverted Repeat Elements Associated with the Genes of Both Monocotyledonous and Dicotyledonous Plants. Plant Cell 6, 907–916.

Busby, M.A., Gray, J.M., Costa, A.M., Stewart, C., Stromberg, M.P., Barnett, D., Chuang, J.H., Springer, M., Marth, G.T., 2011. Expression divergence measured by transcriptome sequencing of four yeast species. BMC Genomics 12, 635.

Campbell, C.S., Donoghue, M.J., Baldwin, B.G., Wojciechowski, M.F., 1995. Phylogenetic Relationships in Maloideae (Rosaceae): Evidence from Sequences of the Internal Transcribed Spacers of Nuclear Ribosomal DNA and its Congruence with Morphology. Am. J. Bot. 82, 903–918.

Canales, R.D., Luo, Y., Willey, J.C., Austermiller, B., Barbacioru, C.C., Boysen, C., Hunkapiller, K., Jensen, R. V, Knight, C.R., Lee, K.Y., Ma, Y., Maqsodi, B., Papallo, A., Peters, E.H., Poulter, K., Ruppel, P.L., Samaha, R.R., Shi, L., Yang, W., Zhang, L., Goodsaid, F.M., 2006. Evaluation of DNA microarray results with quantitative gene expression platforms. Nat. Biotechnol. 24, 1115–1122.

Candaele, J., Demuynck, K., Mosoti, D., Beemster, G.T.S., Inzé, D., Nelissen, H., 2014. Differential Methylation during Maize Leaf Growth Targets Developmentally Regulated Genes. Plant Physiol. 164, 1350–1364.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y., Fraser, P., 2002. Long-range chromatin regulatory interactions *in vivo*. Nat. Genet. 32, 623–626.

Casacuberta, J.M., Santiago, N., 2003. Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311, 1–11.

Chagné, D., Crowhurst, R.N., Troggio, M., Davey, M.W., Gilmore, B., Lawley, C., Vanderzande, S., Hellens, R.P., Kumar, S., Cestaro, A., Velasco, R., Main, D., Rees, J.D., Iezzoni, A., Mockler, T., Wilhelm, L., Van de Weg, E., Gardiner, S.E., Bassil, N., Peace, C., 2012. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One 7, e31745.

Champigny, M.J., Sung, W.W., Catana, V., Salwan, R., Summers, P.S., Dudley, S.A., Provart, N.J., Cameron, R.K., Golding, G.B., Weretilnyk, E.A., 2013. RNA-Seq effectively monitors gene expression in *Eutrema salsugineum* plants growing in an extreme natural habitat and in controlled growth cabinet conditions. BMC Genomics 14, 578.

Chen, H., Yoshimura, F.K., 1998. Spacing between the enhancer and promoter of the long terminal repeat of a murine leukaemia retrovirus is required for transcriptional activation in T cells. J. Gen. Virol. 79, 1101–1104.

Chevreau, E., Lespinasse, Y., Gallet, M., 1985. Inheritance of pollen enzymes and polyploid origin of apple (*Malus* × *domestica* Borkh.). Theor. Appl. Genet. 71, 268–277.

The ENCODE Project Consortium 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816.

Cornille, A., Gladieux, P., Smulders, M.J.M., Roldán-Ruiz, I., Laurens, F., Le Cam, B., Nersesyan, A., Clavel, J., Olonova, M., Feugey, L., Gabrielyan, I., Zhang, X.-G., Tenaillon, M.I., Giraud, T., 2012. New Insight into the History of Domesticated Apple: Secondary Contribution of the European Wild Apple to the Genome of Cultivated Varieties. PLOS Genet. 8, e1002703.

Costa, F., Alba, R., Schouten, H., Soglio, V., Gianfranceschi, L., Serra, S., Musacchi, S., Sansavini, S., Costa, G., Fei, Z., Giovannoni, J., 2010. Use of homologous and heterologous gene expression profiling tools to characterize transcription dynamics during apple fruit maturation and ripening. BMC Plant Biol. 10, 229.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., Zhou, D., Luo, S., Vasicek, T.J., Daly, M.J., Wolfsberg, T.G., Collins, F.S., 2006. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res. 16, 123–131.

Dal Cin, V., Barbaro, E., Danesin, M., Murayama, H., Velasco, R., Ramina, A., 2009. Fruitlet abscission: A cDNA-AFLP approach to study genes differentially expressed during shedding of immature fruits reveals the involvement of a putative auxin hydrogen symporter in apple (*Malus domestica* L. Borkh). Gene 442, 26–36.

Degenhardt, J., Al-Masri, A.N., Kürkcüoglu, S., Szankowski, I., Gau, A.E., 2005. Characterization by suppression subtractive hybridization of transcripts that are differentially expressed in leaves of apple scab-resistant and susceptible cultivars of *Malus domestica*. Mol. Genet. Genomics 273, 326–335.

Devos, K.M., Brown, J.K.M., Bennetzen, J.L., 2002. Genome Size Reduction through Illegitimate Recombination Counteracts Genome Expansion in *Arabidopsis*. Genome Res. 12, 1075–1079.

Dirlewanger, E., Cosson, P., Tavaud, M., Aranzana, M.J., Poizat, C., Zanetto, A., Arús, P., Laigret, F., 2002. Development of microsatellite markers in peach (*Prunus persica* (L.) Batsch) and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theor. Appl. Genet. 105, 127–138.

Dong, Y., Lu, X., Song, W., Shi, L., Zhang, M., Zhao, H., Jiao, Y., Lai, J., 2011. Structural characterization of helitrons and their stepwise capturing of gene fragments in the maize genome. BMC Genomics 12, 609.

Doyle, B., Fudenberg, G., Imakaev, M., Mirny, L.A., 2014. Chromatin loops as modulators of enhancer-promoter interactions in their vicinity. Cornell Univ. Libr.

Eichten, S.R., Ellis, N.A., Makarevitch, I., Yeh, C.-T., Gent, J.I., Guo, L., McGinnis, K.M., Zhang, X., Schnable, P.S., Vaughn, M.W., Dawe, R.K., Springer, N.M., 2012. Spreading of heterochromatin is limited to specific families of maize retrotransposons. PLoS Genet. 8, e1003127.

El-Showk, S., Ruonala, R., Helariutta, Y., 2013. Crossing paths: cytokinin signalling and crosstalk. Development 140, 1373–1383.

Engström, P.G., Steijger, T., Sipos, B., Grant, G.R., Kahles, A., Alioto, T., Behr, J., Bertone, P., Bohnert, R., Campagna, D., Davis, C. a, Dobin, A., Gingeras, T.R., Goldman, N., Guigó, R., Harrow, J., Hubbard, T.J., Jean, G., Kosarev, P., Li, S., Liu, J., Mason, C.E., Molodtsov, V., Ning, Z., Ponstingl, H., Prins, J.F., Rätsch, G., Ribeca, P., Seledtsov, I., Solovyev, V., Valle, G., Vitulo, N., Wang, K., Wu, T.D., Zeller, G., 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat. Methods 10, 1185–1191.

Evans, R.C., Campbell, C.S., 2002. The origin of the apple subfamily (Maloideae; Rosaceae) is clarified by DNA sequence data from duplicated GBSSI genes. Am. J. Bot. 89, 1478–1484.

Feschotte, C., Jiang, N., Wessler, S.R., 2002. Plant transposable elements: where genetics meets genomics. Nat. Rev. 3, 329–341.

Fisher, D. V., 1969. Spur-type Strains of McIntosh for High Density Plantings. Bristish Columbia Fruit Grow. Assoc. Quaterly Rep. 14, 3–10.

Fisher, D. V., 1995. The "Wijcik Spur Mclntosh". Fruit Var. J. 49, 212–213.

Flachowsky, H., Le Roux, P.-M., Peil, A., Patocchi, A., Richter, K., Hanke, M.-V., 2011. Application of a high-speed breeding technology to apple (*Malus* × *domestica*) based on transgenic early flowering plants and marker-assisted selection. New Phytol. 192, 364–377.

Flachowsky, H., Peil, A., Sopanen, T., Elo, A., Hanke, V., 2007. Overexpression of *BpMADS4* from silver birch (*Betula pendula* Roth.) induces early-flowering in apple (*Malus* × *domestica* Borkh.). Plant Breed. 126, 137–145.

Forsline, P.L., Aldwinckle, H.S., Dickson, E.E., Luby, J.J., Hokanson, S.C., 2003. Collection, Maintenance, Characterization, and Utilization of Wild Apples of Central Asia, in: Janick, J. (Ed.), Horticultural Reviews. John Wiley & Sons, Inc., pp. 1–62.

Fujimoto, R., Kinoshita, Y., Kawabe, A., Kinoshita, T., Takashima, K., Nordborg, M., Nasrallah, M.E., Shimizu, K.K., Kudoh, H., Kakutani, T., 2008. Evolution and Control of Imprinted FWA Genes in the Genus *Arabidopsis*. PLOS Genet. 4, e1000048.

Gause, M., Morcillo, P., Dorsett, D., 2001. Insulation of Enhancer-Promoter Communication by a Gypsy Transposon Insert in the *Drosophila cut* Gene : Cooperation between Suppressor of Hairy-wing and Modifier of mdg4 Proteins. Mol. Cell Biol. 21(14), 4807-4817.

Gelvonauskienë, D., Gelvonauskis, B., Sasnauskas, A., 2006. Impact of Rootstock on Columnar Apple Tree Growth in a Nursery. Sci. Work. Lith. Inst. Hortic. Lith. Univ. Agric. 25, 51–56.

Giorno, F., Guerriero, G., Baric, S., Mariani, C., 2012. Heat shock transcriptional factors in *Malus domestica*: identification, classification and expression analysis. BMC Genomics 13, 639.

Giovannoni, J., 2010. Harvesting the Apple Genome. Nat. Genet. 42, 822–823.

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., Lieb, J.D., 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res. 17, 877–885.

Gleave, A.P., Ampomah-Dwamena, C., Berthold, S., Dejnoprat, S., Karunairetnam, S., Nain, B., Wang, Y.-Y., Crowhurst, R.N., MacDiarmid, R.M., 2007. Identification and characterisation of primary microRNAs from apple (*Malus domestica* cv. Royal Gala) expressed sequence tags. Tree Genet. Genomes 4, 343–358.

Grandbastien, M.-A., Spielmann, A., Caboche, M., 1989. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337, 376–380.

Gusberti, M., Gessler, C., Broggini, G.A.L., 2013. RNA-Seq Analysis Reveals Candidate Genes for Ontogenic Resistance in *Malus* - *Venturia* Pathosystem. PLoS One 8, e78457.

Hackett, W.P., 1985. Juvenility, Maturation and Rejuvenility in Woody Plants. Hortic. Rev. (Am. Soc. Hortic. Sci). 7, 109–155.

Halász, J., Hegedüs, A., Pedryc, A., 2006. Review of the Molecular Background of Self-Incompatibility in Rosaceous Fruit Trees. Int. J. Hortic. Sci. 12, 7–18.

Harris, S.A., Robinson, J.P., Juniper, B.E., 2002. Genetic clues to the origin of the apple. Trends Genet. 18, 426–430.

Hegedüs, A., 2006. Review of the Self-Incompatibility in Apple (*Malus* × *domestica* Borkh., syn.: *Malus pumila* Mill.). Int. J. Hortic. Sci. 12, 31–36.

Hemmat, M., Weeden, N.F., Conner, P.J., Brown, S.K., 1997. A DNA Marker for Columnar Growth Habit in Apple Contains a Simple Sequence Repeat. J. Am. Soc. Hortic. Sci. 122, 347–349.

Hily, J.-M., Singer, S.D., Yang, Y., Liu, Z., 2009. A transformation booster sequence (TBS) from *Petunia hybrida* functions as an enhancer-blocking insulator in *Arabidopsis thaliana*. Plant Cell Rep. 28, 1095–1104.

Hyson, D.A., 2011. A Comprehensive Review of Apples and Apple Components and Their Relationship to Human Health. Adv. Nutr. 2, 408–420.

Ibarra-Laclette, E., Lyons, E., Hernández-Guzmán, G., Pérez-Torres, C.A., Carretero-Paulet, L., Chang, T.-H., Lan, T., Welch, A.J., Juárez, M.J.A., Simpson, J., Fernández-Cortés, A., Arteaga-Vázquez, M., Góngora-Castillo, E., Acevedo-Hernández, G., Schuster, S.C., Himmelbauer, H., Minoche, A.E., Xu, S., Lynch, M., Oropeza-Aburto, A., Cervantes-Pérez, S.A., de Jesús Ortega-Estrada, M., Cervantes-Luevano, J.I., Michael, T.P., Mockler, T., Bryant, D., Herrera-Estrella, A., Albert, V.A., Herrera-Estrella, L., 2013. Architecture and evolution of a minute plant genome. Nature 498, 94–98.

Ikase, L., Dumbravs, R., 2004. Breeding of Columnar Apple-Trees in Latvia. Biologija 2, 8–10.
Imelfort, M., Edwards, D., 2009. *De novo* sequencing of plant genomes using second-generation technologies. Brief. Bioinform. 10, 609–18.
The International Peach Genome Initiative, 2013. The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat. Genet. 45, 487–494.
International Rice Genome Sequencing Project, 2005. The map-based sequence of the rice genome. Nature 436, 793–800.
Janick, J., Cummins, J.N., Brown, S.K., Hemmat, M., 1996. Apples, in: Janick, J., Moore, J.N. (Eds.), Fruit Breeding, Tree and Tropical Fruits, pp. 1–77.
Janssen, B.J., Thodey, K., Schaffer, R.J., Alba, R., Balakrishnan, L., Bishop, R., Bowen, J.H., Crowhurst, R.N., Gleave, A.P., Ledger, S., McArtney, S., Pichler, F.B., Snowden, K.C., Ward, S., 2008. Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. BMC Plant Biol. 8, 16.
Jensen, P.J., Rytter, J., Detwiler, E.A., Travis, J.W., McNellis, T.W., 2003. Rootstock effects on gene expression patterns in apple tree scions. Plant Mol. Biol. 53, 493–511.
Jiang, N., Bao, Z., Zhang, X., Eddy, S.R., 2004. Pack-MULE transposable elements mediate gene evolution in plants. Nature 431, 569–573.
Jin, D., Wang, Y., Zhao, Y., Chen, M., 2013. MicroRNAs and Their Cross-Talks in Plant Development. J. Genet. Genomics 40, 161–170.
Jones, E., Chu, W.-C., Ayele, M., Ho, J., Bruggeman, E., Yourstone, K., Rafalski, A., Smith, O.S., McMullen, M.D., Bezawada, C., Warren, J., Babayev, J., Basu, S., Smith, S., 2009. Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays* L.) germplasm. Mol. Breed. 24, 165–176.
Jones-Rhoades, M.W., Bartel, D.P., Bartel, B., 2006. MicroRNAs and Their Regulatory Roles in Plants. Annu. Rev. Plant Biol. 57, 19–53.
Jung, S., Main, D., 2013. Genomics and bioinformatics resources for translational science in Rosaceae. Plant Biotechnol. Rep. 8, 49-64.
Juniper, B.E., Mabberley, D.J., 2006. The Story of the Apple. Timber Press.
Kajikawa, M., Okada, N., 2002. LINEs Mobilize SINEs in the Eel through a Shared 3' Sequence. Cell 111, 433–444.
Kalendar, R., Grob, T., Regina, M., Suoniemi, A., Schulman, A., 1999. IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor. Appl. Genet. 98, 704–711.
Kapitonov, V. V, Jurka, J., 2001. Rolling-circle transposons in eukaryotes. PNAS 98, 8714–8719.
Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P.A., Gingeras, T.R., 2002. Large-Scale Transcriptional Activity in Chromosomes 21 and 22. Science 296, 916–919.
Kellerhals, M., 2009. Introduction to Apple (*Malus* × *domestica*), in: Folta, K.M., Gardiner, S.E. (Eds.), Genetics and Genomics of Rosaceae, pp. 73–84.
Khan, M.A., Han, Y., Zhao, Y.F., Troggio, M., Korban, S.S., 2012. A Multi-Population Consensus Genetic Map Reveals Inconsistent Marker Order among Maps Likely Attributed to Structural Variations in the Apple Genome. PLoS One 7, e47864.
Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., Kakutani, T., 2007. Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. Plant J. 49, 38–45.
Kobayashi, S., Goto-Yamamoto, N., Hirochika, H., 2004. Retrotransposon-Induced Mutations in Grape Skin Color. Science. 304, 982.
Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7, 474.
Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R., Papantonis, A., 2012. Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin 5, 1.
Korban, S.S., Skirvin, R.M., 1994. Nomenclature of the cultivated apple. Hortic. Sci. 19, 177–180.
Korf, I., 2013. Genomics: the state of the art in RNA-seq analysis. Nat. Methods 10, 1165–1166.
Krost, C., 2012. Transkriptom des Meristemgewebes des Kolumnarapfels. Johannes Gutenber-Universität Mainz, Dissertation.
Kumar, A., Bennetzen, J.L., 1999. Plant Retrotransposons. Annu. Rev. Genet. 33, 479–532.
Kvam, V.M., Liu, P., Si, Y., 2012. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am. J. Bot. 99, 248–56.
Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S.,

Wensel, A., Huala, E., 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40, D1202–1210.
Langhanki, L., 2014. Epigenetische Marken in der Kolumnarmutante des Apfels. Johannes Gutenberg-Universität Mainz, Masterarbeit.
Lapins, K.O., 1969. Segregation of Compact Growth Types in Certain Apple Seedling Progenies. Can. J. Plant Sci. 49, 765–768.
Lapins, K.O., 1976. Inheritance of Compact Growth Type in Apple. J. Am. Soc. Hortic. Sci. 101, 133–135.
Lapins, K.O., Watkins, R., 1973. Genetics of Compact Growth Habit, Report of East Malling Research Station for 1972, 136.
Laten, H.M., Gaston, G.D., 2012. Plant Endogenous Retroviruses? A Case of Mysterious ORFs, in: Grandbastien, M.-A., Casacuberta, J.M. (Eds.), Plant Transposable Elements. Topics in Current Genetics 24, pp. 89–112.
Laten, H.M., Havecker, E.R., Farmer, L.M., Voytas, D.F., 2003. SIRE1, an Endogenous Retrovirus Family from *Glycine max*, is Highly Homogeneous and Evolutionarily Young. Mol. Biol. Evol. 20, 1222–1230.
Lauri, P.E., Lespinasse, J.M., 1993. The Relationship between Cultivar Fruiting-Type and Fruiting Branch Characteristics in Apple Trees. Acta Hortic. 349, 259–263.
Lee, M.H., Nicholson, P., 1997. Isolation of genomic DNA from plant tissues. Nat. Biotechnol. 15, 805–806.
Leprince, A.-S., Grandbastien, M.-A., Meyer, C., 2001. Retrotransposons of the Tnt1B family are mobile in *Nicotiana plumbaginifolia* and can induce alternative splicing of the host gene upon insertion. Plant Mol. Biol. 47, 533–541.
Levin, D. (Ed.), 2002. The role of chromosomal change in plant evolution. Oxford University Press
Li, Q., Barkess, G., Qian, H., 2006. Chromatin looping and the probability of transcription. Trends Genet. 22, 197–202.
Li, X., Xu, M., Korban, S.S., 2002. DNA methylation profiles differ between field- and in vitro-grown leaves of apple. J. Plant Physiol. 159, 1229–1234.
Lisch, D., 2009. Epigenetic Regulation of Transposable Elements in Plants. Annu. Rev. Plant Biol. 60, 43–66.
Lisch, D., 2013. How important are transposons for plant evolution? Nat. Rev. 14, 49–61.
Liu, C., Li, C., Liang, D., Wei, Z., Zhou, S., Wang, R., Ma, F., 2012. Differential expression of ion transporters and aquaporins in leaves may contribute to different salt tolerance in *Malus* species. Plant Physiol. Biochem. 58, 159–165.
Looney, N.E., Lane, W.D., 1984. Spur-Type Growth Mutants of McIntosh Apple: a Review of Their Genetics, Physiology and Field Performance. Acta Hort. 146, 31–46.
Luby, J.J., 2003. Taxonomic Classification and Brief History, in: Ferree, D.C., Warrington, I.J. (Eds.), Apples: Botany, Production and Uses. pp. 1–14.
Ma, C., Lu, Y., Bai, S., Zhang, W., Duan, X., Meng, D., Wang, Z., Wang, A., Zhou, Z., Li, T., 2014. Cloning and Characterization of miRNAs and Their Targets, Including a Novel miRNA-Targeted NBS-LRR Protein Class Gene in Apple (Golden Delicious). Mol. Plant 7, 218–230.
Ma, J., Devos, K.M., Bennetzen, J.L., 2004. Analyses of LTR-Retrotransposon Structures Reveal Recent and Rapid Genomic DNA Loss in Rice. Genome Res. 14, 860–869.
Mabberley, D.J., Jarvis, C.E., Juniper, B.E., 2001. The name of the apple. Telopea 9, 421–430.
Malladi, A., Johnson, L.K., 2011. Expression profiling of cell cycle genes reveals key facilitators of cell production during carpel development, fruit set, and fruit growth in apple (*Malus* × *domestica* Borkh.). J. Exp. Bot. 62, 205–219.
McClintock, B., 1951. Chromosome Organization and Genic Expression. Cold Spring Harb. Symp. Quant. Biol. 16, 13–47.
Meulenbroek, J., Verhaegh, J., Janse, J., 1998. Inheritance Studies with Columnar Type Trees. Acta Hort. 484, 255–258.
Micheletti, D., Troggio, M., Zharkikh, A., Costa, F., Malnoy, M., Velasco, R., Salvi, S., 2011. Genetic diversity of the genus *Malus* and implications for linkage mapping with SNPs. Tree Genet. Genomes 7, 857–868.
Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E.S., Bernstein, B.E., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553–560.

Morgan, D.R., Soltis, D.E., Robertson, K.R., 1994. Systematic and Evolutionary Implications of *rbcL* Sequence Variation in Rosaceae. Am. J. Bot. 81, 890–903.

Morgante, M., Brunner, S., Pea, G., Fengler, K., Zuccolo, A., Rafalski, A., 2005. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. Nat. Genet. 37, 997–1002.

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

Moriya, S., Okada, K., Haji, T., Yamamoto, T., Abe, K., 2012. Fine Mapping of *Co*, a Gene Controlling Columnar Growth Habit Located on Apple (*Malus* × *domestica* Borkh.) Linkage Group 10. Plant Breed. 131, 641-647.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The Transcriptional Landscape of the Yeast Genome Defined by RNA Sequencing. Science 320, 1344–1349.

Nakayashiki, H., 2011. The Trickster in the genome: contribution and control of transposable elements. Genes to Cells 16, 827–841.

Newcomb, R.D., Crowhurst, R.N., Gleave, A.P., Rikkerink, E.H.A., Allan, A.C., Beuning, L.L., Bowen, J.H., Gera, E., Jamieson, K.R., Janssen, B.J., Laing, W.A., Mcartney, S., Nain, B., Ross, G.S., Snowden, K.C., Souleyre, E.J.F., Walton, E.F., Yauk, Y.-K., 2006. Analyses of Expressed Sequence Tags from Apple. Plant Physiol. 141, 147–166.

Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., Thanos, D., 2009. Transcription factors mediate long-range enhancer-promoter interactions. PNAS 106, 20222–20227.

Okamoto, H., Hirochika, H., 2001. Silencing of transposable elements in plants. Trends Plant Sci. 6, 527–534.

Omidbakhshfard, M.A., Winck, F.V., Arvidsson, S., Riaño-Pachón, D.M., Mueller-Roeber, B., 2014. A step-by-step protocol for formaldehyde-assisted isolation of regulatory elements from *Arabidopsis thaliana*. J. Integr. Plant Biol. doi:10.1111/jipb.12151

Otto, D., 2013. Molekulare Analyse des Kolumnar-Gens beim Apfel (*Malus* × *domestica*). Johannes Gutenberg-Universität Mainz, Dissertation.

Pardue, M.-L., DeBaryshe, P.G., 2003. Retrotransposons Provide an Evolutionarily Robust Non-Telomerase Mechanism to Maintain Telomeres. Annu. Rev. Genet. 37, 485–511.

Peace, C., Norelli, J.L., 2009. Genetics and Genomics of Rosaceae. Springer New York, New York, NY.

Peer, W.A., Murphy, A.S., 2007. Flavonoids and auxin transport: modulators or regulators? Trens Plant Sci. 12, 556-563.

Pereira-Lorenzo, S., Ramos-Cabrer, A.M., Fischer, M., 2009. Breeding Apple (*Malus* × *domestica* Borkh), in: Jain, S.M., Priyadarshan, P.M. (Eds.), Breeding Plantation Tree Crops: Temperate Species, pp. 33–81.

Phipps, J.B., Robertson, K.R., Rohrer, J.R., Smith, P.G., 1991. Origins and Evolution of Subfam. Maloideae (Rosaceae). Syst. Bot. 16, 303–332.

Pichler, F.B., Walton, E.F., Davy, M., Triggs, C., Janssen, B., Wünsche, J.N., Putterill, J., Schaffer, R.J., 2007. Relative developmental, environmental, and tree-to-tree variability in buds from field-grown apple trees. Tree Genet. Genomes 3, 329–339.

Prescott, A.G., John, P., 1996. DIOXYGENASES: Molecular Structure and Role in Plant Metabolism. Annu. Rev. Plant. Physiol. Plant. Mol. Biol. 47, 245-271.

Ptashne, M., Gann, A., 1997. Transcriptional activation by recruitment. Nature 386, 569–577.

Rabinowicz, P.D., Palmer, L.E., May, B.P., Hemann, M.T., Lowe, S.W., McCombie, W.R., Martienssen, R.A., 2003. Genes and Transposons Are Differentially Methylated in Plants, but Not in Mammals. Genome Res. 13, 2658–2664.

Ramsay, L., Macaulay, M., Cardle, L., Morgante, M., degli Ivanissevich, S., Maestri, E., Powell, W., Waugh, R., 1999. Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J. 17, 415–425.

Ringrose, L., Chabanis, S., Angrand, P.-O., Woodroofe, C., Stewart, A.F., 1999. Quantitative comparison of DNA looping *in vitro* and *in vivo*: chromatin increases effective DNA flexibility at short distances. EMBO J. 18, 6630–6641.

Robinson, J.P., 2001. Taxonomy of the Genus Malus Mill. (Rosaceae) with Emphasis on the Cultivated Apple, *Malus domestica* Borkh. Plant Syst. Evol. 226, 35–58.

Sabot, F., Schulman, A.H., 2006. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. Heredity 97, 381–388.

Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K. a, Meeley, R., Ananiev, E. V, Svitashev, S., Bruggemann, E., Li, B., Hainey, C.F., Radovic, S., Zaina, G., Rafalski, J.-A., Tingey, S. V, Miao, G.-H., Phillips, R.L., Tuberosa, R., 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. PNAS 104, 11376–81.

Samborski, A., Graf, A., Krebs, S., Kessler, B., Bauersachs, S., 2013. Deep Sequencing of the Porcine Endometrial Transcriptome on Day 14 of Pregnancy. Biol. Reprod. 88, 1–13.

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y., Bennetzen, J.L., 1998. The paleontology of intergene retrotransposons of maize. Nat. Genet. 20, 43–45.

SanMiguel, P., Tikhonov, A., Jin, Y.-K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., Bennetzen, J.L., 1996. Nested Retrotransposons in the Intergenic Regions of the Maize Genome. Science 274, 765–768.

Sanzol, J., 2010. Dating and functional characterization of duplicated genes in the apple (*Malus domestica* Borkh.) by analyzing EST data. BMC Plant Biol. 10, 87.

Sax, K., 1933. The origin of the Pomoideae. Proc. Am. Soc. Hortic. Sci. 30, 147–150.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. Science 270, 467–470.

Schouten, H.J., Krens, F.A., Jacobsen, E., 2006. Cisgenic plants are similar to traditionally bred plants. EMBO Rep. 7, 750–753.

Shangguan, L., Han, J., Kayesh, E., Sun, X., Zhang, C., Pervaiz, T., Wen, X., Fang, J., 2013. Evaluation of Genome Sequencing Quality in Selected Plant Species Using Expressed Sequence Tags. PLoS One 8, e69890.

Shearman, J.R., Jantasuriyarat, C., Sangsrakru, D., Yoocha, T., Vannavichit, A., Tragoonrung, S., Tangphatsornruang, S., 2013. Transcriptome analysis of normal and mantled developing oil palm flower and fruit. Genomics 101, 306–312.

Shrestha, P.M., Rotaru, A.-E., Summers, Z.M., Shrestha, M., Liu, F., Lovley, D.R., 2013. Transcriptomic and Genetic Analysis of Direct Interspecies Electron Transfer. Appl. Environ. Microbiol. 79, 2397–2404.

Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., Burns, P., Davis, T.M., Slovin, J.P., Bassil, N., Hellens, R.P., Evans, C., Harkins, T., Kodira, C., Desany, B., Crasta, O.R., Jensen, R. V, Allan, A.C., Michael, T.P., Setubal, J.C., Celton, J.-M., Rees, D.J.G., Williams, K.P., Holt, S.H., Ruiz Rojas, J.J., Chatterjee, M., Liu, B., Silva, H., Meisel, L., Adato, A., Filichkin, S. a, Troggio, M., Viola, R., Ashman, T.-L., Wang, H., Dharmawardhana, P., Elser, J., Raja, R., Priest, H.D., Bryant, D.W., Fox, S.E., Givan, S. a, Wilhelm, L.J., Naithani, S., Christoffels, A., Salama, D.Y., Carter, J., Lopez Girona, E., Zdepski, A., Wang, W., Kerstetter, R. a, Schwab, W., Korban, S.S., Davik, J., Monfort, A., Denoyes-Rothan, B., Arus, P., Mittler, R., Flinn, B., Aharoni, A., Bennetzen, J.L., Salzberg, S.L., Dickerman, A.W., Velasco, R., Borodovsky, M., Veilleux, R.E., Folta, K.M., 2011. The genome of woodland strawberry (*Fragaria vesca*). Nat. Genet. 43, 109–116

Steijger, T., Abril, J.F., Engström, P.G., Kokocinski, F., Akerman, M., Alioto, T., Ambrosini, G., Antonarakis, S.E., Behr, J., Bertone, P., Bohnert, R., Bucher, P., Cloonan, N., Derrien, T., Djebali, S., Du, J., Dudoit, S., Gerstein, M., Gingeras, T.R., Gonzalez, D., Grimmond, S.M., Guigó, R., Habegger, L., Harrow, J., Hubbard, T.J., Iseli, C., Jean, G., Kahles, A., Lagarde, J., Leng, J., Lefebvre, G., Lewis, S., Mortazavi, A., Niermann, P., Rätsch, G., Reymond, A., Ribeca, P., Richard, H., Rougemont, J., Rozowsky, J., Sammeth, M., Sboner, A., Schulz, M.H., Searle, S.M.J., Solorzano, N.D., Solovyev, V., Stanke, M., Stevenson, B.J., Stockinger, H., Valsesia, A., Weese, D., White, S., Wold, B.J., Wu, J., Wu, T.D., Zeller, G., Zerbino, D., Zhang, M.Q., 2013. Assessment of transcript reconstruction methods for RNA-seq. Nat. Methods 10, 1177–1184.

Stekel, D.J., Git, Y., Falciani, F., 2000. The Comparison of Gene Expression from Multiple cDNA Libraries. Genome Res. 2055–2061.

Straub, D., Yang, H., Liu, Y., Tsap, T., Ludewig, U., 2013. Root ethylene signalling is involved in *Miscanthus sinensis* growth promotion by the bacterial endophyte *Herbaspirillum frisingense* GSF30(T). J. Exp. Bot. 64, 4603–4615.

Studer, A., Zhao, Q., Ross-Ibarra, J., Doebley, J., 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. Nat. Genet. 43, 1160–1163.

Sun, H.-Y., Dai, H.-Y., Zhao, G.-L., Ma, Y., Ou, C.-Q., Li, H., Li, L.-G., Zhang, Z.-H., 2008. Genome-wide Characterization of Long Terminal Repeat-Retrotransposons in Apple Reveals the Differences in Heterogeneity and Copy Number between Ty1-copia and Ty3-gypsy Retrotransposons. J. Integr. Plant Biol. 50, 1130–1139.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., Paterson, A.H., 2008. Synteny and collinearity in plant genomes. Science 320, 486–488.

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y., Stitt, M., 2004. Mapman: a User-Driven Tool to Display Genomics Data Sets onto Diagrams of Metabolic Pathways and Other Biological Processes. Plant J. 37, 914–939.

Tian, Y.-K., Wang, C.-H., Zhang, J.-S., James, C., Dai, H.-Y., 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 Hort. 159, 63–68.

Tobutt, K.R., 1994. Combining Apetalous Parthenocarpy with Columnar Growth Habit in Apple. Euphytica 77, 51–54.

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., de Laat, W., 2002. Looping and Interaction between Hypersensitive Sites in the Active β-globin Locus. Mol. Cell 10, 1453–1465.

Troggio, M., Gleave, A., Salvi, S., Chagné, D., Cestaro, A., Kumar, S., Crowhurst, R.N., Gardiner, S.E., 2012. Apple, from genome to breeding. Tree Genet. Genomes 8, 509–529.

Van Damme, M., Huibers, R.P., Elberse, J., Van den Ackerveken, G., 2008. *Arabidopsis DMR6* encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. Plant J. 54, 785–93.

Van de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L., Vandepoele, K., 2009. The flowering world: a tale of duplications. Trends Plant Sci. 14, 680–688.

Varagona, M.J., Purugganan, M., Wessler, S.R., 1992. Alternative Splicing Induced by Insertion of Retrotransposons into the Maize *waxy* Gene. Plant Cell 811–820.

Varkonyi-Gasic, E., Gould, N., Sandanayaka, M., Sutherland, P., MacDiarmid, R.M., 2010. Characterisation of microRNAs from apple (*Malus domestica* "Royal Gala") vascular tissue and phloem sap. BMC Plant Biol. 10, 159.

Varma, A., Padh, H., Shrivastava, N., 2007. Plant genomic DNA isolation: An art or a science. Biotechnol. J. 2, 386–392.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., Goremykin, V., Komjanc, M., Longhi, S., Magnago, P., Malacarne, G., Malnoy, M., Micheletti, D., Moretto, M., Perazzolli, M., Si-Ammour, A., Vezzulli, S., Zini, E., Eldredge, G., Fitzgerald, L.M., Gutin, N., Lanchbury, J., Macalma, T., Mitchell, J.T., Reid, J., Wardell, B., Kodira, C., Chen, Z., Desany, B., Niazi, F., Palmer, M., Koepke, T., Jiwan, D., Schaeffer, S., Krishnan, V., Wu, C., Chu, V.T., King, S.T., Vick, J., Tao, Q., Mraz, A., Stormo, A., Stormo, K., Bogden, R., Ederle, D., Stella, A., Vecchietti, A., Kater, M.M., Masiero, S., Lasserre, P., Lespinasse, Y., Allan, A.C., Bus, V., Chagné, D., Crowhurst, R.N., Gleave, A.P., Lavezzo, E., Fawcett, J.A., Proost, S., Rouzé, P., Sterck, L., Toppo, S., Lazzari, B., Hellens, R.P., Durel, C.-E., Gutin, A., Bumgarner, R.E., Gardiner, S.E., Skolnick, M., Egholm, M., Van De Peer, Y., Salamini, F., Viola, R., 2010. The Genome of the Domesticated Apple (*Malus* × *domestica* Borkh.). Nat. Genet. 42, 833–839.

Velculescu, V.E., Zhang, L., Vogelstein, B., Kinzler, K.W., 1995. Serial Analysis of Gene Expression. Science 270, 484–487.

Vignols, F., Rigau, J., Miguel, A.T., Capellades, M., Puigdomènech, P., 1995. The *brown midrib3* (*bm3*) Mutation in Maize Occurs in the Gene Encoding Caffeic Acid O-Methyltransferase. Plant Cell 7, 407–416.

Vimolmangkang, S., Zheng, D., Han, Y., Khan, M.A., Soria-Guerra, R.E., Korban, S.S., 2014. Transcriptome analysis of the exocarp of apple fruit identifies light-induced genes involved in red color pigmentation. Gene 534, 78–87.

Visconte, V., Rogers, H.J., Singh, J., Barnard, J., Bupathi, M., Traina, F., McMahon, J., Makishima, H., Szpurka, H., Jankowska, A., Jerez, A., Sekeres, M.A., Sauntharajah, Y., Advani, A.S., Copelan, E., Koseki, H., Isono, K., Padgett, R.A., Osman, S., Koide, K., O'Keefe, C., Maciejewski, J.P., Tiu, R. V, 2012. SF3B1 haploinsufficiency leads to formation of ring sideroblasts in myelodysplastic syndromes. Blood 120, 3173–3186.

Vitte, C., Panaud, O., 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet. Genome Res. 110, 91–107.

Vitte, C., Panaud, O., Quesneville, H., 2007. LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss. BMC Genomics 8, 218.

Volff, J.-N., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. BioEssays 28, 913–922.

Vuylsteke, M., Peleman, J.D., van Eijk, M.J.T., 2007. AFLP-based transcript profiling (cDNA-AFLP) for genome-wide expression analysis. Nat. Protoc. 2, 1399–1413.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. 10, 57–63.

Watanabe, M., Suzuki, A., Komori, S., Bessho, H., 2004. Comparison of Endogenous IAA and Cytokinins in Shoots of Columnar and Normal Type Apple Trees. J. Japan. Soc. Hort. Sci. 73, 19–24.

Weiss, D., Ori, N., 2007. Mechanisms of Cross Talk between Gibberellin and Other Hormones. Plant Physiol. 144, 1240–1246.

Wessler, S.R., 1996. Plant retrotransposons: Turned on by stress. Curr. Biol. 6, 959–961.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973–82.

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

Wright, D.A., Voytas, D.F., 2002. *Athila4* of *Arabidopsis* and *Calypso* of Soybean Define a Lineage of Endogenous Plant Retroviruses. Genome Res. 12, 122–131.

Wu, C., 1980. The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNaseI. Nature 286, 854–860.

Wu, C., Wong, Y.-C., Elgin, S.C., 1979. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. Cell 16, 807–814.

Wu, G., 2013. Plant MicroRNAs and Development. J. Genet. Genomics 40, 217–230.

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Khan, M.A., Tao, S., Korban, S.S., Wang, H., Chen, N.J., Nishio, T., Xu, X., Cong, L., Qi, K., Huang, X., Wang, Y., Zhao, X., Wu, J., Deng, C., Gou, C., Zhou, W., Yin, H., Qin, G., Sha, Y., Tao, Y., Chen, H., Yang, Y., Song, Y., Zhan, D., Wang, J., Li, L., Dai, M., Gu, C., Wang, Y., Shi, D., Wang, X., Zhang, H., Zeng, L., Zheng, D., Wang, C., Chen, M., Wang, G., Xie, L., Sovero, V., Sha, S., Huang, W., Zhang, S., Zhang, M., Sun, J., Xu, L., Li, Y., Liu, X., Li, Q., Shen, J., Wang, J., Paull, R.E., Bennetzen, J.L., Wang, J., Zhang, S., 2013. The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome Res. 23, 396–408.

Xia, R., Zhu, H., An, Y., Beers, E.P., Liu, Z., 2012. Apple miRNAs and tasiRNAs with novel regulatory networks. Genome Biol. 13, R47.

Xie, M., Hong, C., Zhang, B., Lowdon, R.F., Xing, X., Li, D., Zhou, X., Lee, H.J., Maire, C.L., Ligon, K.L., Gascard, P., Sigaroudinia, M., Tlsty, T.D., Kadlecek, T., Weiss, A., O'Geen, H., Farnham, P.J., Madden, P. a F., Mungall, A.J., Tam, A., Kamoh, B., Cho, S., Moore, R., Hirst, M., Marra, M.A., Costello, J.F., Wang, T., 2013. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nat. Genet. 45, 836–841.

Xiong, Y., Eickbush, T.H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9, 3353–3362.

Yamada, K., Lim, J., Dale, J.M., Chen, H., Shinn, P., Palm, C.J., Southwick, A.M., Wu, H.C., Kim, C., Nguyen, M., Pham, P., Cheuk, R., Karlin-Newmann, G., Liu, S.X., Lam, B., Sakano, H., Wu, T., Yu, G., Miranda, M., Quach, H.L., Tripp, M., Chang, C.H., Lee, J.M., Toriumi, M., Chan, M.M.H., Tang, C.C., Onodera, C.S., Deng, J.M., Akiyama, K., Ansari, Y., Arakawa, T., Banh, J., Banno, F., Bowser, L., Brooks, S., Carninci, P., Chao, Q., Choy, N., Enju, A., Goldsmith, A.D., Gurjal, M., Hansen, N.F., Hayashizaki, Y., Johnson-Hopson, C., Hsuan, V.W., Iida, K., Karnes, M., Khan, S., Koesema, E., Ishida, J., Jiang, P.X., Jones, T., Kawai, J., Kamiya, A., Meyers, C., Nakajima, M., Narusaka, M., Seki, M., Sakurai, T., Satou, M., Tamse, R., Vaysberg, M., Wallender, E.K., Wong, C., Yamamura, Y., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A., Ecker, J.R., 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302, 842–6.

Yao, J.-L., Dong, Y.-H., Morris, B.A.M., 2001. Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. PNAS 98, 1306–1311.

Yu, C., Zhang, J., Peterson, T., 2011. Genome rearrangements in maize induced by alternative transposition of reversed ac/ds termini. Genetics 188, 59–67.

Yu, H., Song, C., Jia, Q., Wang, C., Li, F., Nicholas, K.K., Zhang, X., Fang, J., 2011. Computational identification of microRNAs in apple expressed sequence tags and validation of their precise sequences by miR-RACE. Physiol. Plant. 141, 56–70.

Zhang, W., Zhang, T., Wu, Y., Jiang, J., 2012. Genome-wide identification of regulatory DNA elements and protein-binding footprints using signatures of open chromatin in *Arabidopsis*. Plant Cell 24, 2719–2731.

Zhu, H., Dardick, C.D., Beers, E.P., Callanhan, A.M., Xia, R., Yuan, R., 2011. Transcriptomics of shading-induced and NAA-induced abscission in apple (*Malus domestica*) reveals a shared pathway involving reduced photosynthesis, alterations in carbohydrate transport and signaling and hormone crosstalk. BMC Plant Biol. 11, 138.

Attachments

Paper 1

Tracing a Key Player in the Regulation of Plant Architecture – the Columnar Growth Habit of Apple Trees (*Malus* x *domestica*)

Romina Petersen[1], Clemens Krost[1]

[1] Department of Molecular Genetics, Johannes Gutenberg-University of Mainz

This paper was published as a review article in Planta in 2013 (*Planta* **238**(1): 1-22, with three figures and one table)

Abstract

Plant architecture is regulated by a complex interplay of some key players (often transcription factors), phytohormones and other signaling molecules such as microRNAs. The columnar growth habit of apple trees is a unique form of plant architecture characterized by thick and upright stems showing a compaction of internodes and carrying short fruit spurs instead of lateral branches. The molecular basis for columnar growth is a single dominant allele of the gene *Columnar*, whose identity, function and gene product is unknown. As a result of marker analyses this gene has recently been fine-mapped to chromosome 10 at 18.51 – 19.09 megabases (according to the annotation of the apple genome (Velasco et al. 2010)), a region containing a cluster of quantitative trait loci associated with plant architecture, but no homologs to the well-known key regulators of plant architecture. Columnar apple trees have a higher auxin/cytokinin ratio and lower levels of gibberellins and abscisic acid than normal apple trees. Transcriptome analyses corroborate these results and additionally show differences in cell membrane and cell wall function. It can be expected that within the next year or two, an integration of these different research methodologies will reveal the identity of the *Columnar* gene. Besides enabling breeders to efficiently create new apple (and maybe related pear, peach, cherry etc.) cultivars which combine desirable characteristics of commercial cultivars with the advantageous columnar growth habit using gene technology, this will also provide new insights into an elevated level of plant growth regulation.

Keywords

Apple, columnar, mapping, quantitative trait loci, phytohormones, transcriptome analysis

Abbreviations

A14, A14-190-93; ABA, Abscisic Acid; AFLP, Amplified Fragment Length Polymorphism; BA, 6-Benzyladenine; BAC, Bacterial Artificial Chromosome; BLAST, Basic Local Alignment Search Tool; BR, Brassinosteroid; CEN, Centroradialis; CK, Cytokinin; CO, Contstans; Co, Columnar; FT, Flowering Locus T; GA, Gibberellin; GC-MS-SIM, Gas Chromatography - Mass Spectrometry - Selected Ion Monitoring; GRAS, Gibberellic-Acid Insensitive – Repressor of Gibberellic-Acid Insensitive and Scarecrow; IAA, Indole-3-Acetic Acid; KNOX, Knotted-1-likehomeobox; LG, Linkage Group; Mb, Megabases; MDP, *Malus domestica* Protein; miRNA, microRNA; P28, Procats 28; QTL, Quantitative Trait Loci; RAPD, Random Amplification of Polymorphic DNA; SCAR, Sequence Characterized Amplified Region; SL, strigolactone; SNP, Single Nucleotide Polymorphism; SSR, Simple Sequence Repeat; TFL, Terminal Flower

Introduction

Genetic Control of Plant Architecture in Herbaceous Plants

The architecture of a plant defines the spatial arrangement of its individual (aerial) organs. With reference to the biological dogma of form and function being intrinsically tied to each other, studying plant architecture is crucial for the comprehension of plant function. Furthermore, controlling plant architecture has always been an important aim in plant breeding, as manipulating plants to grow to a smaller height and produce shorter branches, while at the same time maximizing yield, bears great economic advantages. This has been proven by the introduction of lodging-resistant semi-dwarf wheat and rice mutants that led to the Green Revolution in the 1960s (Peng et al. 1999) and by the adoption of dwarfing rootstocks in fruit tree breeding in the 1920s (Fideghelli et al. 2003).

Even though plant architecture is influenced by environmental factors such as light penetration, temperature, humidity and soil conditions including nutrient availability, the intrinsic body plan of the plant is genetically determined. During the past few years, significant progress has been made in the disclosure of genes that play a pivotal role in establishing the shape of the plant, mostly by examining herbaceous plants showing an aberrant architecture due to a mutation. This has been extensively reviewed elsewhere (see, for example, Reinhardt and Kuhlemeier 2001; Wang and Li 2008), so only a few key players and basic concepts which can be applied to most plant species are presented here (Fig. 1). For clarity, *Arabidopsis* nomenclature is used for genes and proteins unless indicated otherwise.

In the shoot apical meristem, a feedback loop between *wuschel* and *clavata* controls the balance between stem cell maintenance and organogenesis: *wuschel* (Laux et al. 1996; Mayer et al. 1998) maintains undifferentiated cells within the central zone and activates *clavata 3* (Clark et al. 1995), whose gene product in turn together with Clavata 1 and Clavata 2 (Clark et al. 1993) restricts *wuschel* expression to a defined region, thus enabling cells in the peripheral zone to undergo differentiation (Brand et al. 2000; Schoof et al. 2000). The products of the *knotted1-like homeobox* (*KNOX*) genes, with *shootmeristemless* as their best known member, are also responsible for stem cell maintenance (Long et al. 1996; Hay

and Tsiantis 2010). Furthermore, KNOX proteins define boundaries of newly formed organs via interactions with cup shaped cotyledon proteins (Aida et al. 1999).

When lateral organ primordia are initiated, they are arranged in distinct phyllotactic patterns. These patterns are mainly controlled by auxin levels, as the primordia are induced by a local auxin maximum (Reinhardt et al. 2000; Benková et al. 2003). Therefore, genes regulating phyllotaxis are either genes establishing the organization of the shoot apical meristem or genes involved in auxin transport and signaling like *pinformed 1* (Reinhardt and Kuhlemeier 2001). Only mutants of *terminal ear 1* (Veit et al. 1998) in maize and of *perianthia* (Running and Meyerowitz 1996) in *Arabidopsis* show alterations in phyllotactic patterns without also showing changes in the apical meristem. In the leaf primordia, *phan*, *phabulosa* and *phavoluta* establish the adaxial cell fate (McConnell et al. 2001), while *yabby* and *kanadi* promote abaxial cell fates (Siegfried et al. 1999; Kerstetter et al. 2001). Once the leaf has been formed, its shape is controlled by *angustifolia* in the lateral direction and *rotundifolia* in the longitudinal direction (Tsuge et al. 1996; Tsukaya 2005). *lanceolate* as well as *KNOX* induce the formation of compound leaves (Mathan and Jenkins 1962; Hareven et al. 1996).

In the axils of the leaves, axillary meristems either develop from cells of the shoot apical meristem that have maintained their stem cell identity (Garrison 1955; Sussex 1955) or from differentiated cells that undergo dedifferentiation (Snow and Snow 1942). The initiation of axillary meristems is regulated by *revoluta* (Otsuga et al. 2001), *lateral suppressor* (Greb et al. 2003) and *branched 1* (Aguilar-Martínez et al. 2007). For the maintenance of stem cell identity, *shootmeristemless* is induced by Cup Shaped Cotyledon 1 (Hibara et al. 2003), indicating similar regulatory loops in axillary meristems as in the apical meristem. Axillary meristems generate axillary buds, which can remain dormant or grow out to form lateral shoots (Shimizu-sato and Mori 2001). This is regulated by apical dominance, a concept described below. In addition, the angle at which lateral shoots grow out largely contributes to overall plant architecture. The control of the crotch angle has not yet been thoroughly researched, but recently *lazy 1* and *tac 1* have been shown to play important roles as a negative and a positive regulator, respectively, of tiller angle in maize (Li et al. 2007; Yu et al. 2007).

The final important decision in a plant's life is the phase change from vegetative to reproductive growth. For many herbaceous plants, producing a flower and subsequently fruit represents the end of their life cycle. This developmental switch therefore has to be precisely controlled. *revoluta*, *knotted1-like 1* and *erecta* are essential players in the regulation of inflorescence architecture (Douglas et al. 2002). In *Arabidopsis*, five genetically defined but crosstalking pathways have been identified in the control of flowering (recently reviewed in Srikanth and Schmid 2011): the vernalization pathway, the photoperiod pathway, the gibberellin pathway, the autonomous pathway and the aging pathway. Focusing on a simplified view of the photoperiod and autonomous pathways, under long days *constans* (*CO*) mRNA is stabilized at the end of the day (with the help of phytochrome A and due to the influence of *gigantea*) and activates the transcription of *flowering locus T* (*FT*) and *twin sister of FT* (Kardailsky et al. 1999; Kobayashi et al. 1999; Samach et al. 2000; Suarez-Lopez et al. 2001; Valverde et al. 2004). FT moves through the phloem into the meristem and, together with Flowering Locus D, activates *apetala 1*, *fruitful* and *suppressor of overexpression of constans* (Abe et al. 2005; Wigge et al. 2005; Yoo et al. 2005). Floral meristem identity genes like *apetala* (Irish and Sussex 1990) and *leafy* (Weigel et al. 1992)

transform indeterminate axillary meristems into determinate floral meristems. Leafy directly activates *Apetala 1* redundantly with FT and also induces the transcription of the homeotic genes *agamous* and *apetala 3* (Busch et al. 1999; Lamb et al. 2002). The antagonist of *leafy* is *terminal flower 1* (*TFL1*), which favors indeterminate vegetative growth (Shannon and Meeks-Wagner 1991).

While the individual parts of the plant are being built, intercalary meristems mediate elongation growth as well as secondary growth of the main axis and lateral organs. Elongation growth of the main axis occurs via cell divisions and subsequent internode elongation. Internode elongation is mainly hormonally controlled (see below), but regulators of cell division or cell wall remodeling and polyamine signals also influence cell elongation in the stem (Hanzawa et al. 2000; Zhou et al. 2006; Asano et al. 2010; Todaka et al. 2012). The increase in stem diameter is caused by the formation of secondary xylem and secondary phloem by the vascular cambium. This process seems to be regulated in essentially the same way as primary growth because it involves *wuschel-related homeobox* and *clavata3/ESR-related* factors (Schrader et al. 2004; Hirakawa et al. 2010) and requires *shootmeristemless* and *KNOX* genes for the maintenance of stem cell identity (Groover et al. 2006; Du et al. 2009; Zhang et al. 2011).

Phytohormones Regulating Plant Architecture in Herbaceous Plants

For a plant as a sessile organism it is crucial to establish a stable architecture while at the same time maintaining the possibility to modify the spatial arrangement of its individual parts in order to adapt to environmental changes. Thus, plant architecture is dynamic and extensive signal integration and crosstalk constantly take place for its precise control. Most long-distance signaling within the plant is accomplished via the highly interconnected network of the different phytohormones, and the regulation of architecture is no exception to this rule (Fig. 1).

Auxins, the most important representative being indole-3-acetic acid (IAA), are involved in nearly all aspects of plant growth and development (reviewed in Overvoorde et al. 2010; Müller and Leyser 2011; Durbak et al. 2012). As mentioned above, auxin levels determine the sites of organ primordia formation (Reinhardt et al. 2000). They also are the predominant mediator of apical dominance, which is defined as the control exerted by the apical portions of the main shoot over the outgrowth of lateral buds (Cline, 1991). The high auxin content in the shoot apex regulates ramification because it usually suppresses the outgrowth of lateral shoots so that the latter usually only take over in case the leader is damaged and its influence ceases (Thimann and Skoog 1933; Cline 1991), although outgrowth can also occur under “vigorous” growth conditions or on vigorous rootstocks. The regulation of apical dominance has been extensively discussed elsewhere (Leyser 2005; Rameau 2010; Domagalska and Leyser 2011) and will thus not be discussed in detail. Clearly, the polar basipetal IAA transport within the dominant stem, mediated via the asymmetric distribution of auxin influx carriers of the Auxin Resistant 1/Like Auxin Resistant family (Swarup et al. 2008; Péret et al. 2012; Swarup and Péret 2012) and auxin efflux carriers of the Pinformed family (Gälweiler et al. 1998; Palme and Gälweiler 1999; Křeček et al. 2009), plays a central role in the establishment of apical dominance (Friml and Palme 2002). The polar auxin transport in the main shoot either hinders the lateral buds from establishing their own auxin flux which then prevents their outgrowth (Li and Bangerth 1999), or it regulates the levels of a second, upwardly moving messenger of bud

sprouting (Snow 1929; Sachs and Thimann 1967). Cytokinins (CKs) and strigolactones (SLs) are the most likely candidates for this second messenger: direct CK application to the lateral bud promotes its outgrowth (Cline 1991), whereas SLs act as a branching inhibitor (Gomez-Roldan et al. 2008; Umehara et al. 2008). Furthermore, CK and SL levels have been shown to be regulated by IAA (Nordström et al. 2004; Johnson et al. 2006; Brewer et al. 2009). In turn, CK induces IAA biosynthesis (Jones et al. 2010), and genes involved in the biosynthesis and/or signaling of SLs act as negative regulators of polar auxin transport (Bennett et al. 2006). As a central component of plant development, auxin cross talks with nearly all other phytohormones (reviewed in Chandler 2009). The interplay between auxin and SLs not only determines sprouting but also stimulates secondary growth (Agusti et al. 2011).

Besides auxin, CKs are essential for plant growth, since they are generally stimulators of cytokinesis (Hartig and Beck 2006). As such, they maintain the activity of the vegetative and floral shoot apical meristem (Riou-Khamlichi et al. 1999), regulate cambial activity (Matsumoto-Kitano et al. 2008; Nieminen et al. 2008) and antagonize the inhibitory effect of auxin on the outgrowth of lateral buds. Contrary to their promotive role for growth of the aerial parts of the plant, they negatively regulate root apical meristem activity and lateral root formation (Werner et al. 2003).

In addition to cytokinesis, cell elongation is essential for growth to take place. Elongation growth is mainly mediated via gibberellins (GAs) and brassinosteroids (BRs) (reviewed in Phinney 1985; Müssig 2005), which stimulate internode elongation (Yamamuro et al. 2000; Dayan et al. 2012; Li et al. 2012). Ethylene signaling has been shown to influence internode elongation in rice adapted to deep water conditions (Hattori et al. 2009; Qi et al. 2011), and IAA contributes to elongation growth in pea (McKay et al. 1994). Low GA levels in the shoot apical meristem together with the high CK/IAA ratio favor the maintenance of stem cells, whereas high GA and low CK/IAA ratio induce the formation of lateral organs (Shani et al. 2006). GAs and BRs also promote flowering (Langridge 1957; Wilson et al. 1992; Domagalska et al. 2010). This effect is antagonized by abscisic acid (ABA), which generally has an inhibitory role on plant growth (Milborrow 1967).

The remaining three groups of phytohormones are ethylene, jasmonates and salicylic acid; however, they only play a minor role in the regulation of plant shape, except under stress conditions (Xu et al. 1994; Zhang and Turner 2008; Sehr et al. 2010). In contrast, the past few years have shown that other molecules like microRNAs (miRNAs) highly influence plant architecture (Chen et al. 2010; Jiao et al. 2010). It is often not yet clear at which points and in which way phytohormone or miRNA pathways and the key genes for the regulation of plant architecture converge. The input from several signaling pathways is probably integrated to fine-tune a few final developmental switches.

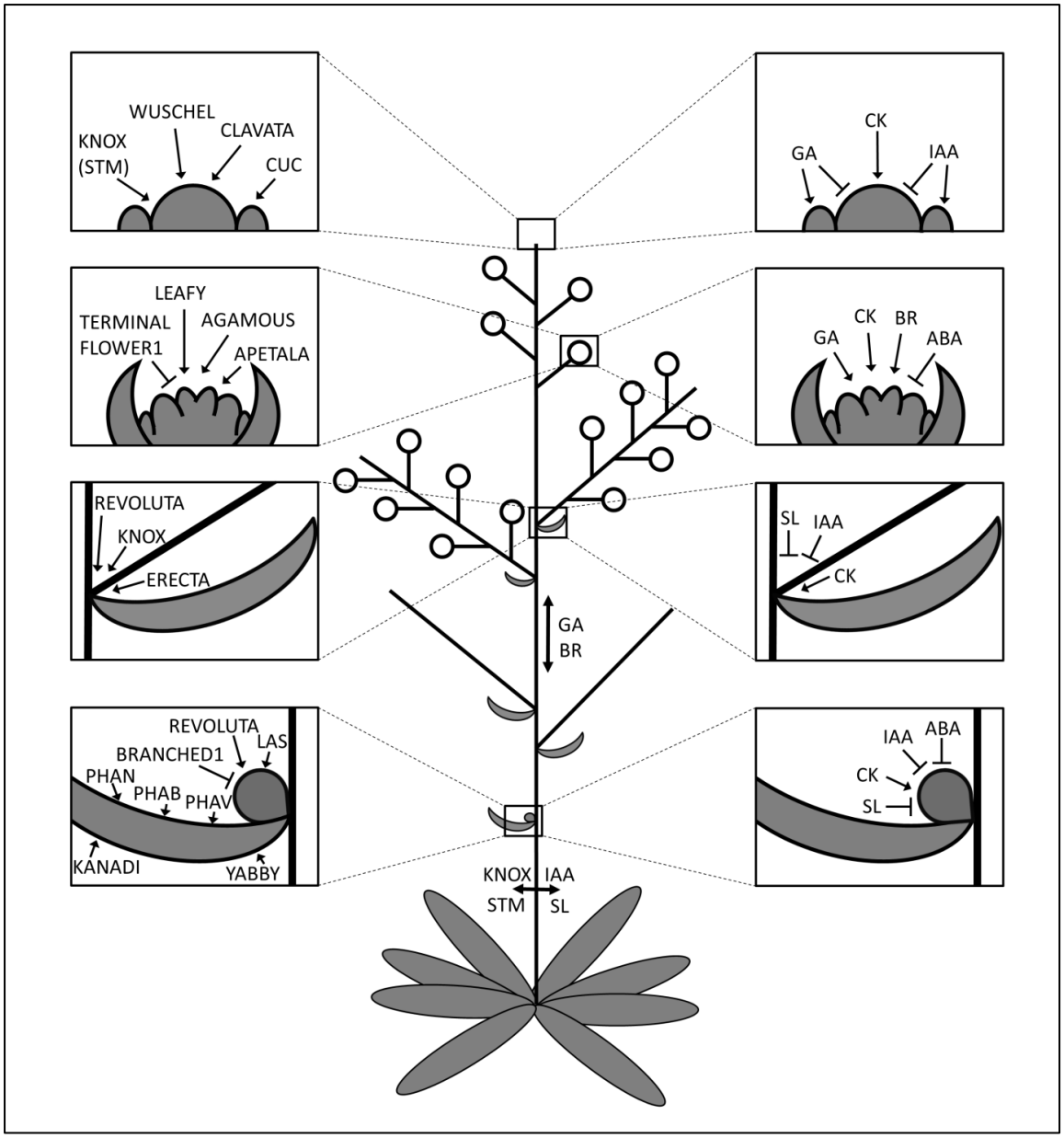


Fig. 1 Major aspects of plant architecture regulation. The shape of a dicotyledonous plant is regulated by some key proteins (left hand side) and phytohormones (right hand side) that show promotive (arrows) or inhibitory (bar-headed arrow) effects on shoot apical meristem activity (top), floral meristem activity, inflorescence branching and vegetative branching (bottom) as well as on elongation growth (double-headed vertical arrow) and secondary growth (double-headed horizontal arrow). For detailed explanations see text. ABA – abscisic acid, BR – brassinosteroid, CK – cytokinin, CUC – Cup Shaped Cotyledon, GA – giberellic acid, IAA – indole-3-acetic acid, KNOX – Knotted1-like Homeobox, LAS – Lateral Suppressor, Phab – Phabulosa, Phav – Phavoluta, SL – strigolactone, STM – Shootmeristemless.

Approaching Tree Architecture

To date, our understanding of plant architecture establishment as described above is mostly based on results obtained from the herbaceous model organism *Arabidopsis thaliana* and the economically important cereals such as rice or maize. In contrast, research on tree

architecture has long been mainly descriptive (e.g. Ceulemans et al. 1990; Costes et al. 2006). However, during the past few years, the genus *Populus* (including poplars and aspen) has emerged as a model for the study of tree architecture and physiology, supported by the sequencing of its genome, which was the first tree genome to be completed (Tuskan et al. 2006; Wullschleger et al. 2013). As the number of available genome sequences of perennial plants continues to increase (Jaillon et al. 2007; Velasco et al. 2010; Shulaev et al. 2011; Verde et al. 2013), the unraveling of the regulation of tree architecture has been greatly accelerated. Most of the genes and mechanisms described in the previous sections have homologs and equivalents in woody plants. Sometimes two or more orthologs to each *Arabidopsis* gene can be found due to gene duplication, and these paralogous genes might show different expression patterns and fulfill slightly different roles. For instance, *Populus* has several *wuschel* and *shootmeristemless* orthologs, which in addition to regulating stem cell maintenance in the SAM also play a similar role in the vascular cambium (Schrader et al. 2004; Groover 2005; Groover et al. 2006; Bao et al. 2009). With regard to the regulation of flowering, *Populus* has two orthologs of *FT* (see below), *TFL 1* (Mohamed et al. 2010) and *agamous* (Brunner et al. 2000), respectively, and in apple, two *leafy* orthologs with different expression patterns and thus possibly different functions can be detected (Wada et al. 2002). Functional differences of certain genes in *Arabidopsis* and trees have been reported, especially in the case of heterologous transformation (Gocal et al. 2001; Flachowsky et al. 2010).

Trees and herbaceous plants also share all key concepts of phytohormone regulation. In trees, IAA is involved in the control of apical dominance and apical control, acting in combination with cytokinins (Wilson 2000; Cline and Dong-Il 2002). For perennial plants, “apical dominance” only refers to the decision between outgrowth or bud formation of the current year’s axillary meristems (yielding sylleptic branches), whereas the term “apical control” is used to describe the influence of apical parts of the tree on the growth of lateral shoots and of previously dormant buds in subsequent years (yielding sylleptic and proleptic branches) (Brown et al. 1967; Cline 1997). Apical dominance and apical control are dynamic in time and can be modified in response to environmental effects. IAA also regulates secondary growth, which is more pronounced in trees because their longevity combined with the indeterminate growth of plants leads to a higher average plant size and biomass compared with annual plants, necessitating the reinforcement of the plant body. In this context, the formation of a radial auxin gradient with a peak in the cambium and the adjacent first few layers of xylem cells and its synergistic action with GA in wood formation have been intensively researched (Uggla et al. 1996; Uggla et al. 1998; Eriksson et al. 2000; Isreaelsson et al. 2005; Björklund et al. 2007; Nilsson et al. 2008; Mauriat and Moritz 2009; Han et al. 2011; Chen et al. 2013). Furthermore, GAs affect elongation growth (Han et al. 2011; Elias et al. 2012) as well as flowering of perennial plants (Zawaski et al. 2011; Randoux et al. 2012). They also control seed dormancy together with ABA (reviewed in Graeber et al. 2012). ABA regulates responses to abiotic stresses, especially drought (Li et al. 2004; Popko et al. 2010; Ji et al. 2013).

To gain an advantage in the competition for light and nutrients during their first few years of life and to build up constructional and photosynthetically active organs before the formation of reproductive structures, most trees undergo a juvenile phase in which they cannot be induced to flower (Hackett 1985). This phase can last several years; for instance, poplar and apple flower for the first time after 7 – 10 and 4 – 8 years, respectively (Hackett

et al. 1985; Hsu et al. 2006). The transition from the juvenile to the adult phase is regulated by the CO/FT regulatory module, similar to the photoperiod pathway of the transition from the vegetative to the reproductive phase in *Arabidopsis* described above. *FT* transcription gradually increases during the juvenile phase (Böhlenius et al. 2006), and *Populus* plants overexpressing the *FT* homologs *FT1* or *FT2* as well as plum plants transformed with poplar *FT1* show an early flowering phenotype (Böhlenius et al. 2006; Hsu et al. 2006; Hsu et al. 2011; Srinivasan et al. 2012). Downregulation of the *Populus TFL* homologs *Populus Centroradialis 1* (*PopCEN1*) and *Populus Centroradialis 2* (*PopCEN2*) leads to precocious maturity (Mohamed et al. 2010). In addition, miR156 and miR172 contribute to the control of vegetative phase change (reviewed in Huijser and Schmid 2011).

Perennial plants of temperate and boreal regions need to adapt to seasonality and develop a strategy to survive the winter period with its unfavorable growth conditions. For this purpose, most trees initiate bud set in late summer and then undergo a period of bud dormancy (for a recent comprehensive review, see Cooke et al. 2012). From a physiological point of view, dormancy has been divided into three phases: paradormancy, caused by inhibitors in leaves and terminal buds, ecodormancy, due to unfavorable environmental conditions, and endodormancy, caused by inhibitors within the bud itself (Lang 1987; Lang et al. 1987). Later it has been redefined independently of external and internal stimuli as the inability of a meristem to resume growth under favorable conditions (Rohde and Bhalerao 2007). This can be applied to the apical meristem as well as to the axillary meristems and the cambium. While the cambium is sheltered by the bark, the SAM and the axillary meristems are enclosed in buds, providing protection for the winter period. Buds are highly important for both vegetative and reproductive growth of trees since they are in fact undeveloped shoots. While most trees produce vegetative buds that develop into vegetative shoots and flower buds that develop into flowers, some species such as apple and pear produce vegetative and mixed buds, the latter of which can develop into leafy shoots as well as flowers (Mimida et al. 2009). In these species, a mixed unit ("bourse") containing vegetative and floral organs can occur. A bourse can subsequently form a sylleptic axillary shoot ("bourse shoot") that can finally develop into a short or long shoot (Costes and Guédon 2002). The decision whether a bud turns into a flower/fruit or shoot is controlled by various factors such as cultivar, rootstock, shoot growth and phytohormones (Hoad 1984; Buban 1996; Koutinas et al. 2010).

Cessation of apical elongation growth and bud set are the first steps to winter dormancy. Most plants induce these processes in response to shortening of day length (Heide 1974; Junttila 2007), whereas species of the Rosaceae family such as apple react to lower temperatures (Heide and Prestrud 2005; Heide 2008). Species with a strictly determinate growth pattern, in which the terminal bud contains all preformed primordia and internodes for the subsequent growth period, show autonomous control of bud set and growth cessation with almost no influence of environmental changes (Junttila 1976). Dormancy release occurs in answer to the fulfillment of a chilling requirement and/or a longer photoperiod (Murray et al. 1989; Heide et al. 1993; Junttila and Hänninen 2012). The longer and the colder the chilling period is, the lower are the time and temperature sum to bud burst (Junttila and Hänninen 2012). The master switch in the regulation of onset as well as release of seasonal arrest is again the CO/FT module sensing day length in relation to the circadian clock (for a recent review on the circadian clock, see Farrè et al. 2011). Additionally, the circadian clock might be able to sense temperature in *Arabidopsis*

(Edwards et al. 2006; Gould et al 2006). In *Populus*, *FT2* controls growth cessation, bud set and dormancy induction, and *FT1* is expressed during the chilling period to induce the transition from the vegetative to the reproductive phase (Böhlenius et al. 2006, Hsu et al. 2006; Hsu et al. 2011; Rinne et al. 2011). Furthermore, overexpression of *CEN/TFL1* in *Populus* causes delayed bud break and altered chilling requirements (Mohamed et al 2010). *Aintegumenta*-like genes, which are regulators of cell division, and dormancy-induced *MADS-box* genes are downstream targets of CO/FT (Karlberg et al. 2011; Yamane et al. 2011). Additional downstream effects related to dormancy induction are the upregulation of genes associated with cold hardiness and drought, defense, carbohydrate synthesis and transport, cell wall biosynthesis or modification as well as RNA metabolism and chromatin modification/remodeling (Ruttink et al. 2007; Park et al. 2008; Ko et al. 2011). In contrast, the transition from dormancy to active growth is characterized by the induction of flowering pathways, RNA metabolism and protein biosynthesis and transport (Larisch et al. 2012). It has been found that SAM cells are symplastically isolated during the dormant period due to the formation of a callose block at the plasmodesmata (Rinne and van der Schoot 1998; Rinne et al. 2001, Rinne et al. 2011). However, it is not yet clear whether there is a causal relationship between the cellular isolation and the activity - dormancy cycle. Phytohormone levels are also altered by the circadian clock in response to dormancy. A short photoperiod causes downregulation of GA20 oxidase and correspondingly decreases GA levels, which induces growth cessation (Eriksson and Moritz 2002). ABA has a central role in seed dormancy and has long been thought to mediate bud dormancy as well (Knox and Wareing 1984). Its precise role in bud dormancy is still unclear, but it seems to be involved in the control of bud development and maturation, as *abscisic acid insensitive 3* overexpressing plants develop defective buds, but normal dormancy (Ruttink et al. 2007). Ethylene might crosstalk with ABA to regulate bud dormancy induction and bud morphology (Ruttink et al. 2007). The decreased cell division capcity of the cambium during winter dormancy has been shown to be accompanied by decreased auxin sensitivity (Schrader et al. 2004; Baba et al. 2011). Furthermore, there might be epigenetical (Santamaria et al. 2009; 2011), miRNA (Wang et al. 2011) and metabolism (sugars, energy status, redox state and reactive oxygen species) apects of the regulation of winter dormancy (Halaly et al. 2008; Ophir et al. 2009), but research on these topics is still underway.

It would be of great value to gain an even deeper insight into the control of plant architecture in trees, with the possibility to transfer the new knowledge to other plant species. The columnar growth phenotype of apple is a natural mutation with the potential to reveal a key player in the regulation of tree architecture since it is dominantly inherited. As columnar apple trees show a thicker main stem with shorter internodes than apple trees with a standard growth habit and since many short fruit spurs emanate from the main stem of columnar apple trees at a narrower crotch angle than the long lateral shoots of standard apple trees (Fig. 2), its causative gene, *Columnar* (*Co*), seems to influence nearly all aspects of plant architecture. At the same time it could equip apple breeders with an important tool for the generation of new cultivars with high economic importance. Therefore this review will briefly address the history and development of the columnar growth habit and then focus on the different approaches that are currently being taken to identify *Co* and its function.

History and Development of the Columnar Growth Habit

In the 1960s, researchers at East Malling in Kent were trying to revolutionize fruit tree breeding by the induction of mutations in apple (*Malus* x *domestica*) and the subsequent examination of the resulting phenotypes, focusing on spur type trees. The spur type growth habit was first recorded in the 1920s and is characterized by the formation of numerous short fruit spurs instead of large side branches (Quinlan and Tobutt 1990). However, the spur type is a fruiting type rather than a growth habit (Fideghelli et al. 2003), and an extreme of a continuous variation rather than a distinct heritable trait (Looney and Lane 1984), so in targeted breeding experiments almost none of the spur type varieties transferred the desired phenotype to their progeny to a significant extent (Lapins 1969).

Being trained to watch out specifically for spur type mutants, in 1961 grower Anthony Wijcik spotted a very compact and spurry limb sport atop a 50-year-old McIntosh tree at the Summerland Research Station, British Columbia, which arose as the result of a spontaneous somatoclonal mutation (Fisher 1969; Fisher 1995). Fruits from this sport reached maturity slightly later than that of the rest of the tree and had a less intensive color (Fisher 1995). Vegetative propagation of this sport, later called "McIntosh Wijcik" (commercially also known as "Starkspur Compact Mac"), and subsequent crosses with plants showing a normal growth habit, demonstrated that almost 50 % of its progeny showed the compact phenotype (Lapins 1969; Lapins and Watkins 1973; Lapins 1974), which was later referred to as columnar growth habit. A number of similar sports were found on top of other aging McIntosh trees, e.g. in the Bendig orchard at Summerland, but were not propagated (Looney and Lane 1984; Fisher 1995).

Using McIntosh Wijcik as a parent, thousands of crossings performed at East Malling yielded six more columnar apple varieties until 1991: Telamon (Waltz), Trajan (Polka), Tuscan (Bolero), Obelisk (Flamenco), Charlotte (Hercules) and Maypole (Tobutt 1994). These columnar apple trees of the first generation, known as the "Ballerina" trees due to their commercial names, were susceptible to scab, showed biennial bearing and their fruits were not competitive with those of the most popular commercial apple tree varieties such as Golden Delicious, Jonagold or Gala. Breeding approaches in Canada, the USA, China, Korea, Belgium, Lithuania, Russia, Great Britain, Germany and France have since produced columnar varieties like Arbat, Moonlight and Goldlane that bear fruits of higher quality than the original columnar varieties and are resistant to scab and other common diseases (Gelvonauskienë et al. 2006). However, breeding apples by conventional crossing is time-consuming and cost-intensive. Apples are highly heterozygous and self-incompatible (Hegedűs 2006; Newcomb et al. 2006), and many agronomically important traits are under polygenic control, which constrains the production of varieties with a specific combination of advantageous traits and increases the need for large progenies of crossings (Velasco et al. 2010). Furthermore, apples have a long juvenile period so that fruit quality can only be assessed after several years (Hackett 1985). Thus, the demand for methods of early detection of the growth phenotype and more efficient creation of new columnar apple cultivars has accelerated research of this interesting phenotype and its molecular cause.

Phenotype Characteristics of Columnar Type Apple Trees

The compact growth of columnar type apple trees is based on their very thick and upright stems with almost no difference in diameter between the top and the bottom and short internodes, overall looking like a sturdy cordon (Fig. 2b) (Tobutt 1985; Tobutt 1994). They produce short fruit spurs rather than long lateral branches (Fig. 2c). Rarely, the axillary buds do develop into long lateral shoots which then grow almost parallel to the stem at a very narrow crotch angle (Hemmat et al. 1997; Bai et al. 2012). The development of long side shoots is favored if the central leader is damaged, in which case two to three spurs near the top grow to about 50 cm of length (Tobutt 1985; Watanabe et al. 2006), which implicates that the lateral buds are under tight apical control. The new lateral shoots also show the columnar habit (Kenis and Keulemans 2007). If a number of lateral buds of one-year old branch sections are removed, then the reaction is that the more buds are removed the more likely the remaining ones are to grow out (Looney and Lane 1984), indicating a competition between individual spurs. Columnar trees have less sylleptic shoots and thus show higher apical dominance. They exhibit a lower level of acrotony and develop more proleptic shoots, which is in line with the hypothesis of the buds being under higher apical control than those of trees with standard growth habit (De Wit et al. 2000). Even though shoots of columnar apple trees grow longer during one vegetative season than shoots of normal trees (Watanabe et al. 2004), the central leader of McIntosh Wijcik grows to only about 55 % the size of a standard McIntosh tree, and many of the other columnar and spur type trees are also smaller than normal trees (Lane and Looney 1982; Kelsey and Brown 1992). However, since this does not apply to all varieties, it is likely that columnar growth habit and dwarfing are two distinct traits that segregate independently (Eaton and Lapins 1970).

While the number of leaves per shoot is similar between normal and columnar apple trees (Lee and Looney 1977), the total leaf area is greater in columnar seedlings at three years of age (Zhang and Dai 2011), and the leaves themselves also show some differences. Leaves of columnar apple trees are dark green and very thick with long petioles and usually have a serrate or crenate margin (Lapins 1969; Tobutt 1988a; Tobutt 1988b; Tobutt 1988c; Tobutt 1988d; Sarwar et al. 1998). This is a characteristic of most other spur type growth habits as well (Liu and Eaton 1970). Microscopic examinations and detailed measurements have demonstrated that the leaves of columnar apple trees have a thicker palisade parenchyma as well as a greater dry weight and chlorophyll content (Gelvonauskis et al. 2006; Zhang and Dai 2011). This results in a higher net photosynthetic rate and transpiration rate of columnar compared with normal type apple trees (Zhang and Dai 2011).

The diameter of xylem vessels is bigger in shoots and roots of columnar than standard apples and the number of xylem vessels is also higher in roots of columnar trees (Zhang and Dai 2011). This together with the higher photosynthetic rate explains why these trees can produce high yields of fruit despite their compact growth, at least when they are grown on typical commercial rootstocks like M9: they are able to efficiently transport water and minerals from the soil up through the stem and to produce a higher amount of sugar compounds. Regarding the number or width of phloem vessels, no differences were found (Zhang and Dai 2011).

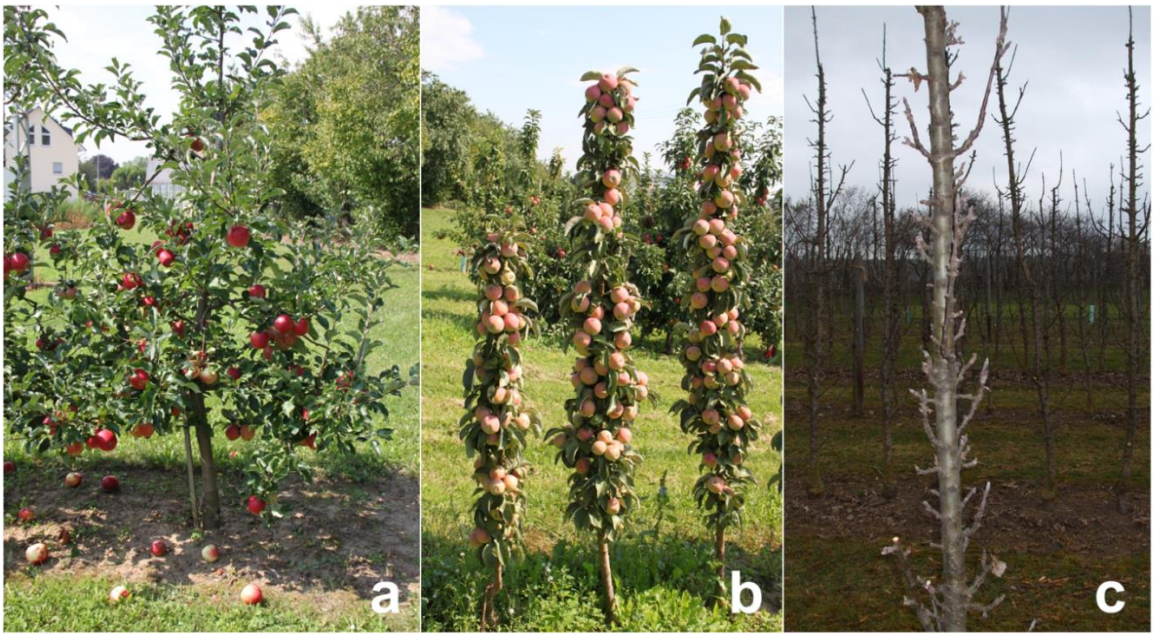


Fig. 2 Comparison of the plant architecture of standard and columnar type apple trees. (a) Apple trees with standard growth habit (variety A14-190-93K) have long lateral branches with a wide crotch angle and usually require staking. (b) By contrast, columnar apple trees (variety Goldcats) do not require wood stakes and show compact growth. (c) The pillar-like growth of columnar trees (here variety A73-75-97K) is due to short fruit spurs and few longer lateral branches emanating at a narrow crotch angle and growing almost parallel to the stem. Pictures were taken at the Geisenheim University in early fall 2011 (a, b) and early spring 2013 (c).

Another characteristic of columnar apple varieties is that they often show frost and drought resistance, which might be due to their Canadian origin (Jacob 2010).

The columnar growth habit can be detected as soon as two weeks to two months after germination (Lee and Looney 1977; Meulenbroek et al. 1998). However, this early examination is often erroneous. A reliable verification of the phenotype is possible after about two to three years (Blazek 1992; Baldi et al. 2012). Even then, classification can be difficult because there is not always a clear distinction between different growth phenotypes and many intermediate types exist (Hemmat et al. 1997; Kim et al. 2003; Ikase and Dumbravs 2004; Moriya et al. 2009; Baldi et al. 2012). In addition to the age of the plants, the proportion of progeny with an intermediate growth habit resulting from a cross with a columnar parent seems to be dependent on the columnar variety used (Table 1) as well as on the growth conditions (Tobutt 1985; Brown et al. 2004). The columnar growth habit is still evident when columnar plants are grafted on rootstocks obtained from apple varieties with standard growth. However, the choice of rootstock influences tree height, stem diameter and shoot number (Gelvonauskienë et al. 2006). This can either be caused by the lack of expression of columnar-specific genes in the roots of normal type rootstocks, or by the general influence of the rootstock on plant growth characteristics like tree height, trunk diameter, number of shoots and flowering onset, which can be observed to act upon standard type apple trees after grafting (Seleznyova et al. 2008). As for normal apple trees, dwarfing rootstocks such as M9 are efficient tools for controlling the height and shoot number of columnar apple trees (Gelvonauskienë et al. 2006).

Table 1 Results of crosses with columnar cultivars. The columnar parent is underlined. Plant age is indicated at the last date of phenotypic evaluation where mentioned. Plants were grown on own roots unless indicated otherwise. Only crosses with a total plant number higher than 50 and precise numbers of progeny which was not pre-selected were included. Other research groups reported an approximation of a 1:1 ratio of columnar versus non-columnar individuals without giving precise numbers (Lee and Looney 1978, Hemmat et al. 1997, Tian 2005, Zhu et al. 2007).

Cross (plant age)	Total plant no.	co plants	non-co plants	Inter-mediate/ non-classified plants	Per-centage of co plants	Literature
Golden Delicious x McIntosh Wijcik (2)	107	47	60	0	44	Lapins (1969)
Wellington Bloomless x McIntosh Wijcik (2)	140	46	94	0	33	Tobutt (1994)
Spencer Seedless x SAxy[1] (2)	604	297	307	0	49	Tobutt (1994)
McIntosh Wijcik x NY75441-67 (4)	126	44	40	42	35	Hemmat et al. (1997)
Telamon x Braeburn (2)	59	21	38	0	36	De Wit et al. (2000)
Telamon x Sunrise (2)	69	23	46	0	33	De Wit et al. (2000)
Telamon x 110 (2)	82	37	39	0	42	De Wit et al. (2000)
Fuji x Tuscan (3)	227	69	41	117	30	Kim et al. (2003)
Arbat x Forele (7)	66	42	21	3	64	Ikase and Dumbravs (2004)
KV-11 x Melba (7)	52	27	8	17	52	Ikase and Dumbravs (2004)
Telamon x Braeburn[2] (2)	247	108	134	5	44	Kenis and Keulemans (2007)
Fiesta x Totem	85	44	37	4	52	Férnandez-Férnandez et al. (2008)
Fuji x 5-12786 (2)	68	30	38	0	44	Moriya et al. (2009)
Fuji x NYCO7-G (9)	271	156	104	11	58	Bai et al. (2012)
Telamon x Braeburn[2,3] (5)	222	105	113	4	47	Bai et al. (2012)
6-837 x 5-8246 (2)	100	46	54	0	46	Moriya et al. (2012)
Golden Delicious x Wijcik[2] (6)	101	40	61	0	40	Baldi et al. (2012)
Goldrush x Wijcik[2] (4)	141	54	84	3	38	Baldi et al. (2012)
Galaxy x Wijcik[2] (4)	63	28	34	1	44	Baldi et al. (2012)
Golden Delicious x Wijcik (3)	399	199	175	25	50	Baldi et al. (2012)
Golden Delicious x Wijcik (3)	898	442	434	22	49	Baldi et al. (2012)

[1] pooled data from four crosses
[2] plants grafted on M9 rootstocks at two years age
[3] The Telamon x Braeburn population used by Bai et al. (2012) is the same as the one used by Kenis and Keulemans (2007)

Most of the phenotype characteristics of columnar apple tree varieties provide economic benefits, which is why it has always attracted the attention of breeders and subsequently researchers. Columnar trees can be planted only 0.5 – 1 m apart in orchards of about 10,000 trees per hectare (Tobutt 1994). They require no staking due to their thick and upright stems and only need to be pruned to control their height. Flowering for the first time about four years after germination, they have a lifespan of about 20 years and could pay back for

the expenses of planting after about four years. Furthermore, mechanical harvesters could be used in orchards of this kind, which would save additional cost and labor. It has also been proposed that columnar apple trees be used as space-saving pollinators for conventional orchards or as ornamentation in gardens or streets (Tobutt 1985).

Mapping and Analyzing the *Columnar* Gene Region

The columnar growth habit represents a class of fruit tree architecture on its own (Fideghelli et al. 2003), so it would be highly interesting to determine its molecular basis. Lapins (1969) first proposed that the columnar growth habit could be attributed to the dominant allele of a single gene, *Columnar* (*Co*). Since crosses between a columnar and a non-columnar parent usually yield less than 50 % columnar progeny (Table 1), Lapins (1976) suggested that one or two modifier genes might be involved. In contrast, Blazek (1992) deduced that the columnar growth habit might be a double recessive trait. However, the latter hypothesis can be rejected because all commercially available columnar cultivars have been found to be heterozygous for *Co* (Tian et al. 2005) and crosses between two columnar cultivars have yielded only up to 75 % columnar progeny (Lapins 1976; Meulenbroek et al. 1998). It has been proposed that the deficiency of columnar type trees in the progeny might be caused by a negative influence of *Co* or a linked gene on the viability of pollen, seeds or emerging seedlings (Meulenbroek et al. 1998). Baldi et al. (2012) found the lack of columnar F_1 plants to be more pronounced for grafted trees than for trees grown on their own roots, so they concluded that columnar plants might be lost during and/or shortly after the grafting process, which might be amplified when dwarfing rootstocks (in their case M9) are used.

At present, the identity and function of *Co* as well as its gene product or the type of mutation are still unknown. Similar growth phenotypes have been found for other Rosaceae species such as peach and sour cherry, but their genetic background is either unclear or is distinct from apple (Scorza et al. 2002; Schuster 2009). Since *Co* is dominant, it is rather unlikely that it is a loss-of-function allele, unless it has a dose-dependent effect. It might carry an amino acid-changing single nucleotide polymorphism (SNP) that causes a gain-of-function in a protein. It might also be a small RNA or a transposon insertion/deletion. Since columnar trees still show their compact growth even when grafted on normal type rootstocks (Fisher 1995), the *Co* gene product probably exerts its effect in the shoot rather than working up from the roots. Interestingly, overexpression of the *Arabidopsis leafy* gene in apple trees causes the plants to develop a columnar growth habit (Flachowsky et al 2010), so *Co* might somehow be involved in the developmental switch from indeterminate vegetative to determinate reproductive growth. However, these are mere speculations and different scientific approaches are needed to form well-founded hypotheses.

During the past 15 years, several research groups have tried to determine the chromosomal location of *Co* using marker analyses so that we can now conclude that it is located on chromosome 10 within the region of 18.51 – 19.09 megabases (Mb) (Fig. 3a). At the same time the genetic linkage map of apple was designed and constantly refined, providing more markers for coupling analyses (Maliepaard et al. 1998; Liebhard et al. 2002; Liebhard et al. 2003; Silfverberg-Dilworth et al. 2006; Wang et al. 2012). The different research groups used PCR-based markers, mainly RAPD (Random Amplification of

Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism) or SSR (Simple Sequence Repeats), the former two sometimes being converted to SCAR (Sequence Characterized Amplified Region) markers.

Before the publication of the apple genome sequence (Velasco et al. 2010), *Co* was mapped to linkage group (LG) 10 of the apple linkage map and its location was gradually narrowed down to 17.0 – 19.5 Mb based on seven pivotal markers (Fig. 3b). The very first *Co*-linked marker, SSR^{Co} (Hemmat et al. 1997), whose locus is also amplified by the COL primers designed by Gianfranceschi et al. 1998 and used for linkage analysis by Kenis and Keulemans (2007), seems to be at a distance of at least 20 cM from *Co*. Markers in closer proximity to *Co* are $UBC818_{1000}$ (Zhu et al. 2007) and $SCAR_{216}$ (Tian et al. 2005), followed by Hi01a03 (Moriya et al. 2009), $SCAR_{682}$ (Tian et al. 2005) and CH03d11 (Fernández-Fernández et al. 2008; Liebhard et al. 2002; Tian et al. 2005). Different values for recombination frequencies and genetic distances of these markers to *Co* were found by different research groups depending on the apple varieties and the number of individuals used in their linkage studies, but their sequential arrangement remained roughly the same. In a comprehensive marker analysis involving segregating progenies of three crosses and a high number of columnar plants of different varieties, SSR markers CH03d11 and Hi01a03 were found to be the most tightly linked markers flanking *Co* on either side with maximum genetic distances of 7.4 cM and 1.4 cM, respectively (Moriya et al. 2009). This and two other studies also detected linkage of marker CH02a10, identical to SSRCO F/R (Tian et al. 2005; Bendokas et al. 2007; Moriya et al. 2009), and a fourth study identified SCAR marker $SCB82_{670}$ as linked (Kim et al. 2003). However, Basic Local Alignment Search Tool (BLAST) searches (Altschul et al. 1990) against the annotated apple genome (Velasco et al. 2010) show that the sequences of CH02a10 and $SCB82_{670}$ map to chromosome 3 at about 30 Mb. $SCB82_{670}$ was later shown to amplify a fragment from the paternal non-columnar parent of the columnar variety used by Kim et al. (2003), so that it cannot be linked to *Co* (Tian et al. 2005; Fernández-Fernández et al. 2008; Moriya et al. 2009). In contrast, the reason for co-segregation of CH02a10 despite the lack of physical coupling remains unclear. It might indicate a genomic region of major importance for maintaining the viability of columnar plants and is thus covered by a selective sweep as proposed by Krost et al. (2013). Alternatively, the genomic contig might have been incorrectly assembled to chromosome 3.

Since the Golden Delicious genome sequence was released in 2010 (Velasco et al. 2010), the possibility of generating sequence-based markers has significantly simplified and accelerated the fine-mapping of the *Co* region (Fig. 3a). Bai et al. (2012) developed 88 SSR markers based on genomic apple contigs originating from chromosome 10 around the target region as well as from two unanchored contigs which they identified as being located within this region via a synteny approach using the peach genome. They used four segregating progenies as well as 290 columnar selections to evaluate the quality of 18 already published markers and the new SSR markers. 47 plants showing recombination between $SCAR_{682}$ and Hi01a03 as well as one double-recombinant were used for screening with the new markers, and six key recombinants finally served to delimit the *Co* gene region to 193 kb between markers C1753-3520 at 18.90 Mb and C7629-22009 at 19.09 Mb. Marker C18470-25831 (19.0 Mb) was found to co-segregate with *Co*. The newly delimited region contains 20 annotated genes and seven predicted genes, three of which code for homologs of Lateral Organ Boundaries Domain transcription factors in *Arabidopsis* (Majer and

Hochholdinger 2011) and were thus considered the most likely candidates for *Co* (Bai et al. 2012).

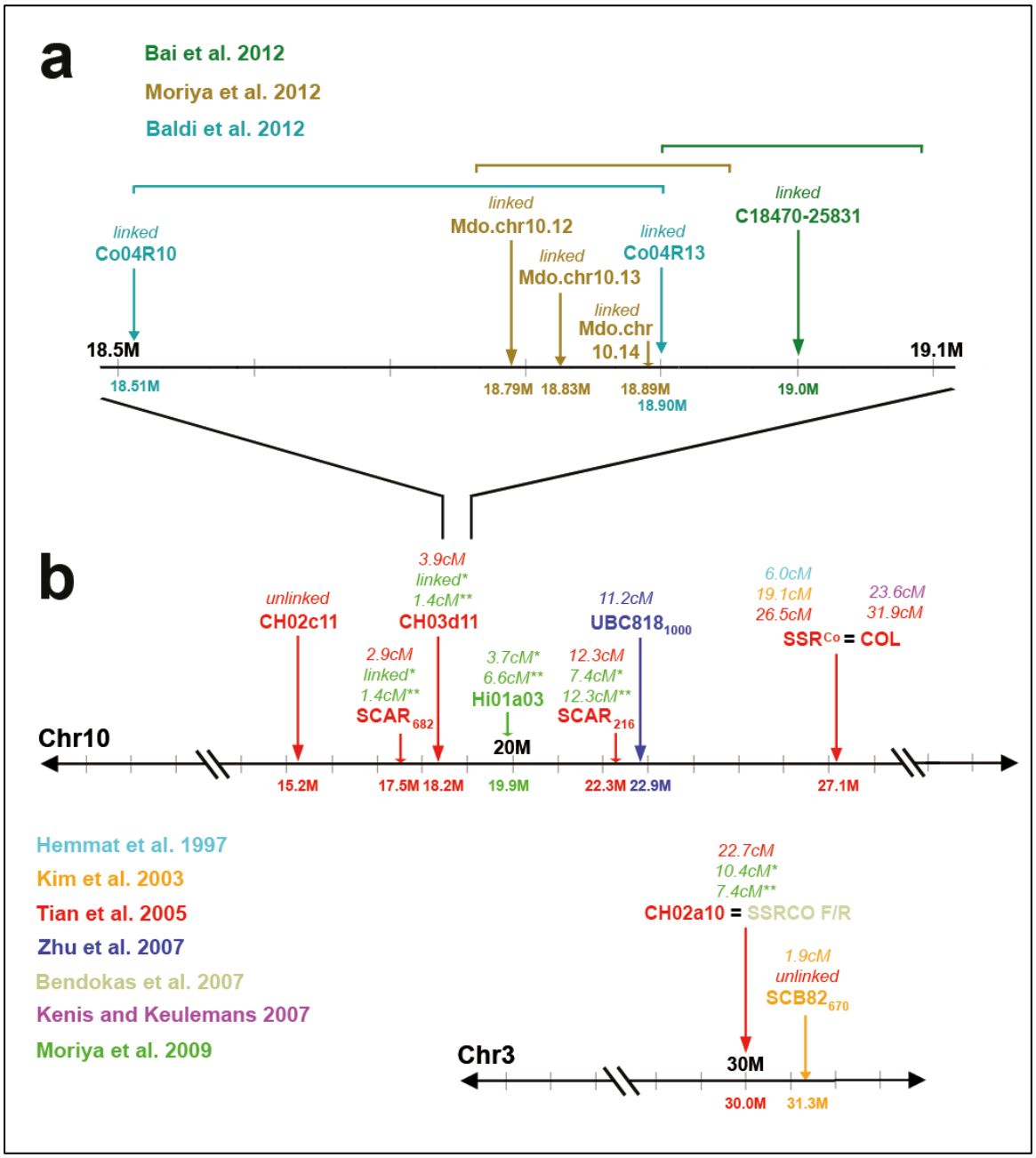


Fig. 3 Molecular markers in the *Co* gene region. (a) *Co* has recently been mapped on chromosome 10 at 18.52-19.09 Mb based on the apple genome sequence (Velasco et al. 2010). (b) Before this publication, *Co* had already roughly been mapped to 17.0 – 19.5 Mb. Markers used are shown at the location they have been mapped to via BLAST searches against the apple genome (Velasco et al 2010). Marker names are in bold, recombination frequencies are in italics, corresponding publications are designated by colors. */** indicates recombination frequencies found for different varieties. Horizontal lines above the markers in (a) designate the newly delimited *Co* regions by Bai et al. (2012), Moriya et al. (2012) and Baldi et al. (2012)

Moriya et al. (2012) also developed SSR markers based on the Golden Delicious genome sequence and used 1000 F_1 individuals from 31 populations for linkage analysis. They delimited the *Co* gene region to 196 kb, a similar size as Bai et al. (2012). However, the

markers developed by Moriya et al. (2012) flanking this region are located at 18.76 Mb (Mdo.chr10.11) and 18.96 Mb (Mdo.chr10.15). Additionally, markers Mdo.chr10.12 (18.79 Mb), Mdo.chr10.13 (18.83 Mb) and Mdo.chr10.14 (18.89 Mb) co-segregated with *Co*.

In a third study, Baldi et al. (2012) first used three adult segregating progenies (301 F_1 plants in total) and the early published markers to roughly define the *Co* gene region and then refined it with two large populations treated as a single segregating progeny of 1,250 individuals which they subjected to genotypic mapping with seven newly designed SSR markers. *Co* was localized between co-mapping markers Co04R10 and Co04R11 on one side and Co04R13 on the other side, which yielded a genomic region of 0.56 cM, corresponding to 393 kb at 18.52 – 18.90 Mb. A thorough open reading frame analysis revealed 36 potential genes within this region, several of which code for transcription factors of the MYB, basic helix-loop-helix and AP2/ERF classes, members of which play roles in the regulation of plant architecture. The target region overlaps with the region identified by Moriya et al. (2012), but does not span the chromosomal location predicted for *Co* by Bai et al. (2012). Furthermore, marker C18470-25831, which Bai et al. (2012) found to be linked to *Co*, showed three recombinants in the mapping population of Baldi et al. (2012). Possible reasons for the different locations of the *Co* region boundaries are that the three research groups used different apple genotypes which possibly have distinct recombination frequencies and that the marker defining the right border of Bai et al. (2012) originates from one of the contigs which was unanchored in the genome project, so its precise location is not well-founded. There might also have been some difficulties in the phenotypic classification of individual plants. Taken together, *Co* is most likely located between 18.76 and 18.9 Mb, a region comprising approximately 30 annotated genes, none of which have so far been identified as having a profound influence on plant architecture.

Detailed analyses of the *Co* target region are already underway: Baldi et al. (2012) constructed a Bacterial Artificial Chromosome (BAC) library based on genomic DNA extracted from leaves of McIntosh Wijcik with an average insert size of 145 kb and found ten BAC clones to originate from the *Co* target region, of which a minimum of four clones is needed to span the entire chromosomal section of interest. These clones will enable a comparative analysis of the columnar and the non-columnar allele within the next months.

Otto et al. (2013) also constructed BAC libraries using genomic DNA of the heterozygous columnar cultivar Procats 28. They obtained more than 100,000 clones with an average insert size of 27 kb, 37 of which could be assigned to chromosome 10 at 18.0 – 19.0 Mb. Assembly of their sequences yielded two metacontigs of 590 kb and 190 kb in size. Comparison of these sequences to the Golden Delicious reference has already been used to create four Indel-based markers linked to *Co* and is ongoing in order to detect more differences between the genomic organization of columnar and non-columnar apple trees.

Alleles showing consistent differences between columnar and non-columnar cultivars can be expected to be further analyzed and prepared for transformation in the near future.

Analysis of Quantitative Trait Loci in the *Columnar* Gene Region

To gain a better understanding of the many distinct but interconnected factors that influence plant architecture, mapping of QTLs associated with plant growth on progenies of columnar trees have been carried out alongside marker analyses. Several research groups

have found clusters of QTLs associated with plant architecture on LG 10 in the vicinity of *Co*. For 172 juvenile apple trees of the cross McIntosh Wijcik x NY 75441-58, internode number in year 1 and year 2 of their life, the stem base diameter increment in year 1 and year 3 as well as branch number and internode length in year 1 were associated with regions located on LG 10 at roughly the locus of the marker P459z (Conner et al. 1998). The *terminal bearing* (*Tb*) locus correlating with branching habit and influencing vegetative bud break (Lawson et al. 1995) also maps close to this region (Conner et al. 1998); however, it is distinct from *Co*. Based on the investigations of a Telamon x Braeburn progeny comprising 257 individuals, a QTL cluster for a wide range of phenotype characteristics such as total growth increment, total branch number and branch length, internode length, main axis growth rate and main axis height increment was detected on LG 10 (Kenis and Keulemans 2007). The authors deduced clustering of different genes or a pleiotropic effect of a single gene, preferring the latter. They also found the *Co* gene influence on branch length (apical control) to be more pronounced than its influence on branch number (apical dominance).

Using the same mapping population, loci for plant architectural traits and QTLs associated with fruit quality were both found on the same linkage group as *Co* (Davey et al. 2006; Kenis et al. 2008). This includes QTLs for fruit flesh weight, flesh L-ascorbic acid content, soluble sugar, firmness and acidity. Some of these QTLs account for up to 60 % of the phenotypic differences observed. A possible theory is that this specific area on LG 10 controls aspects of plant growth and development and these then have pleiotropic effects which affect fruit quality traits (Kenis et al. 2008). This led Moriya et al. (2009) to the conclusion that *Co* might be tightly linked to genes conferring low fruit quality, and thus it would be desirable to use gene technology for the production of new columnar cultivars because classical breeding approaches would in most cases transfer *Co* in combination with these unpopular traits.

In summary, the *Co* gene region on LG 10 seems to be pivotal for plant growth and development, either due to the influence of one gene (possibly *Co*) or due to the combined action of several genes in a cluster. In any case, changes in growth habit would most likely be accompanied by alterations in the phytohormone levels of the plant.

Phytohormone Levels in Columnar Apple Trees

Several attempts have been made to correlate the columnar growth habit with changes in phytohormone levels. Unfortunately, phytohormone concentrations are difficult to measure and they show extensive variations depending on the age of the plant, the season and the environmental factors. Additionally, it is challenging to decide whether changes in hormone levels are a cause or a consequence of the columnar growth habit. Only four hormone groups have been investigated so far in the columnar apple, and direct connections between the results and the phenotypic manifestation are scarce.

Since columnar apple trees form short fruit spurs rather than long lateral branches, it was suggested that they might have stronger apical dominance and apical control than normal type apple trees and thus a focus has been put on the IAA levels in columnar plants. Measurements of IAA content with the indole-pyrone fluorescence method indicated that in shoot tips, free IAA levels correlated positively with the degree of compactness. Additionally, axillary buds of columnar apple trees have a higher ratio of free IAA to total

IAA than normal type apple trees due to a lower level of conjugated IAA (Looney and Lane 1984). However, when the polar auxin transport is blocked by application of the inhibitor triiodobenzoic acid, branch attachment angles and the numbers of spurs on both genotypes were increased, which contradicts the hypothesis of stronger apical dominance in columnar apple trees (Looney and Lane 1984). The authors concluded that in McIntosh Wijcik, an IAA-hydrolyzing enzyme might be active at the time when lateral buds are formed, which results in very strong buds easily breaking dormancy and growing into spurs during the following season. Due to the competition for nutrients among the axillary organs, fruit spurs rather than long branches are formed (Looney and Lane 1984). However, it seems more likely that the preferential formation of spurs compared with lateral shoots is caused by a lack of growth-promoting factors such as GA. Watanabe et al. (2004; 2006; 2008) measured the IAA concentration of the central and axillary shoots (arising from just below the previous year's pruning cut) of 3- to 5-year-old columnar trees grafted on seedling rootstocks using gas chromatography-mass spectrometry-selected ion monitoring (GC-MS-SIM) and found the central leader to have more IAA than the lateral branches. Furthermore, in July, total IAA was higher in axillary shoots of columnar type apple trees than in axillary shoots of one- or two-year-old branches of 38-year-old normal type McIntosh trees on seedling rootstocks, which might be correlated with the observation that the former still grow vigorously until October, whereas the latter already cease growing in July. No significant differences were found with respect to the total IAA content in axillary shoots of columnar and normal type apple trees during the entire growth season. Unfortunately, since Watanabe et al. (2004; 2006; 2008) compared 3- to 5-year-old columnar trees of the varieties Maypole and Tuscan (grown on seedling rootstocks) to normal McIntosh trees at 38 to 40 years of age (also grown on seedling rootstocks), the differences might be a result of age rather than of growth habit. Furthermore, the number of plants analyzed was low (maximum of n=15). Therefore, it would be interesting to compare IAA levels within a high number of columnar and non-columnar trees of the same age. Since it is usually the IAA transport within the shoot and not the absolute IAA level that regulates growth processes, measuring the IAA movement within the stem of columnar apple trees would also be of great interest. In summary, the results indicate a higher IAA content in columnar apple trees, which would be in agreement with higher apical control.

Focusing on CKs, in the soybean hypocotyl section bioassay, McIntosh Wijcik was found to have significantly higher levels of zeatin-like growth substances than other McIntosh strains (Looney and Lane 1984). In GC-MS-SIM experiments of the same plants as mentioned above, the endogenous concentration of zeatin riboside, the predominant CK associated with budburst in apple, was found to be higher in both apical and lateral shoots of columnar trees during the course of a year (Watanabe et al. 2004; Watanabe et al. 2006). This could explain the high number of spurs. However, like in the IAA measurements, the authors again only compared a low number 3- to 5-year old columnar trees, which grow vigorously, with 38- to 40-year-old normal trees, which are at the end of their growth period, so further studies comparing a statistically significant number of trees of the same age would be needed. The ratio of isopentenyl adenosine to total cytokinin was highest in lateral shoots and buds of normal trees in July and of columnar ones in November, so it is probably associated with the onset of winter dormancy (Watanabe et al. 2008).

Growth of normal apple trees after exogenous application of CKs follows an optimum curve: the higher the amount of CKs given, the more vigorously does the plant grow when

growth is defined as the number of shoots initiated which develop to shoots longer than 1 cm or as culture weight gain for in vitro cultures. If the concentration exceeds the optimum level, plant growth is more and more inhibited. In vitro cultures of columnar apple trees show a similar reaction of culture weight gain to exogenously applied CKs, and the optima for 6-benzyladenine (BA) (5 µM) and Thidiazuron (3 µM) are similar to the ones of normal type apple trees (Lane and Looney 1982). However, at very low BA concentrations, normal type apple cultures on modified Murashige and Skoog medium grow better, whereas columnar apple cultures on modified Murashige and Skoog medium show a significantly higher tolerance to supra-optimal concentrations of CKs: the maximum concentrations at which columnar and normal type plants still grow are 25 µM (Sarwar et al. (1998): 50 µM) versus 5 µM for BA, 60 µM versus 25 µM for kinetin, 25 µM versus 20 µM for 2-isopentyladenine, and 40 µM versus 25 µM for thidiazuron, respectively (Lane and Looney 1982; Sarwar et al. 1998). In order to perform shoot regeneration from leaf explants, higher concentrations of CKs are needed for columnar apple plants than for normal ones (Sarwar and Skirvin 1997). The authors concluded that either columnar plants can metabolize excess levels of CKs or they compensate for it by adjusting the level of other growth regulators (Lane and Looney 1982). Another explanation would be that columnar apples do not take up excessive amounts of CKs, e. g. because they have thicker cell layers than standard type apple trees (Sarwar et al. 1998).

Taken together, the levels of CKs seem to be higher in columnar than in normal apples and columnar apple trees seem to have an altered CK metabolism. Watanabe et al. (2008) hypothesized that the large leaf area of columnar apple trees might contribute in part to the increased production of CKs. However, as CKs are predominantly produced in the roots and to a smaller extent in young leaves only, this explanation seems rather unlikely.

Some of the phenotype characteristics of columnar trees such as stunted growth and dark green leaves resemble characteristics of GA-deficient mutants (Koornneef and Veen 1980; Talon et al. 1990; Sun and Kamiya 1994; Peng and Harberd 1997). Thus, GAs constitute another class of phytohormones that has attracted the attention of researchers focusing on columnar growth. Looney and Lane (1984) summarized their findings on GA levels in columnar apple trees as follows: in the dwarf pea bioassay, shoot tip extracts of columnar apple seedlings did not promote growth as much as those of normal type seedling, indicating lower GA-like activity. Low polar GAs were also found in actively growing shoot tips of McIntosh Wijcik when examined using silica gel partition column chromatography and the dwarf rice bioassay. However, polar GAs are not necessarily bioactive (Atzorn and Weiler 1983). After exogenously applied GA_3, columnar seedlings showed a greater percentage of growth increase, but still did not reach the height of their normal type counterparts. The conclusion was that low GA levels probably correlate with dwarfing of McIntosh Wijcik rather than its spurriness, and dwarfing is a phenotype characteristic independent of compact growth (Looney and Lane 1984).

With regard to ABA, lower levels of free *cis*-ABA were found in actively growing shoot tips including five expanded terminal leaves of young columnar progeny of McIntosh Wijcik crosses with three different non-columnar varieties grafted on M7 than in their normal type siblings (Lee and Looney 1977). These data were statistically significant on a per shoot tip basis, whereas data on a fresh weight basis had a tendency towards lower levels, but no statistical significance was achieved. The bourse buds of McIntosh Wijcik trees also had less free and conjugated ABA than those of standard McIntosh on a fresh weight basis, even

though, due to their larger size, the total ABA amount per bud was higher (Looney and Lane 1984). ABA in general is higher per fresh weight in rapidly elongating shoots (Feucht et al. 1974), so the lower ABA levels are probably a consequence – not the cause – of the slower growth of fruit spurs compared with lateral shoots (Looney and Lane 1984).

In the seeds and early seedlings of a progeny of controlled crosses of McIntosh Wijcik with a non-columnar variety, levels of ABA and GA were similar for both varieties, indicating that the hormonal differences that characterize the compact seedlings are probably established at a later stage of development (Lee and Looney 1978).

Taken together, these results suggest that columnar apple varieties do show differences in phytohormone levels, but most of them are fairly subtle, and only hypotheses of their correlation with the phenotype can be made. Due to the high IAA levels and lower IAA/CK ratio of shoots, the apical dominance of columnar trees is higher than that of normal type trees and thus they do not produce long axillary shoots. Fruit spurs are produced because of high levels of CKs in combination with low levels of ABA, which favors bud break, but slows extension growth. Only in some occasions (e. g. when the central leader is damaged) are a few spurs able to overcome the apical dominance and grow out. The lower levels of GA might inhibit long extension growth and are related to the dwarfing of most columnar trees.

Transcriptome Analyses of Columnar Type Apple Trees

Another approach to unravel the function of *Co* is the analysis of transcriptional changes in columnar compared to normal type apple trees. Taking into consideration the fundamental phenotypic changes, it can be expected that the expression levels of several genes are altered, but it is difficult to decide on one specific pathway which is definitely influenced, so the method of choice is a whole transcriptome study.

Three recent studies have analyzed the transcriptomes of columnar and normal type apple trees via RNA-seq. Zhang et al. (2012) collected new lateral shoots of 4-year-old columnar and standard seedlings of the progeny of Fuji x Telamon. They used three time points between May and July 2010 and pooled the material of the three time points in order to obtain two samples, one from columnar and one from non-columnar apple. They performed a total RNA extraction and mRNA purification, and generated about 4 million reads for each sample in an Illumina HiSeq 2000 next generation sequencing run. 80 % of the reads were mapped to the apple genome (Velasco et al. 2010). Assembling yielded about 57,000 non-redundant unigenes with contigs having an N_{50} of about 420. Comparing differential gene expression based on reads per kilobase per million mapped reads values, 5,237 genes were found to be differentially expressed by more than twofold, 2,704 being upregulated and 2,533 being downregulated in the columnar versus the normal type apple. Of the differentially expressed genes, 15 % were involved in the biosynthesis of secondary metabolites, and 24 % had a role in metabolic pathways, some of them being key players in gibberellin, IAA and brassinosteroid biosynthesis. Unfortunately, no specific genes or their precise regulation were mentioned. In addition, 287 genes involved in pathways crucial for the regulation of plant architecture were identified, most likely based on the function of their homologs identified by BLAST searches (although the authors did not mention how they identified their function), but again no values for their expression were given. Among these 287 genes, 31 were mapped to chromosome 10, and 25 were Gibberellic-Acid

Insentitive, Repressor of Gibberellic-Acid Insensitive and Scarecrow (GRAS) transcription factors like DELLAs, which play a role in the gibberellin signal transduction. Some of these genes of interest are intended to be transferred into Gala apple trees via *Agrobacterium*-mediated transformation.

Krost et al. (2012) also compared gene expression levels of columnar versus normal type apple trees using RNA-seq. They collected shoot apical meristems of spurs of columnar Procats 28 (P28), which has a Telamon ancestor, and of branches of normal A14-190-93 (A14), isolated the total RNA and purified the mRNA followed by mRNA amplification. May 2009 and September 2009 were used as time points for A14 and P28, respectively. 454 as well as Illumina sequencings were performed with these samples, and about 250,000 reads were obtained for each 454 library as well as about 80 million reads for each Illumina library. They conducted BLAST searches of the raw sequences against all annotated *Malus domestica* proteins (MDPs) and compared those to UniProtKB. Sunsequently, differential expression was determined and differentially expressed genes were grouped into distinct categories. Genes of categories representing light reactions, mitochondrial electron transport, lipid metabolism and cell wall modification (expansins and xyloglucan endotransglucosylases/hydrolases) were significantly downregulated in the columnar variety, whereas another group of genes involved in cell wall modification, terpenoid and tryptophan synthesis (the precursors of IAA biosynthesis) were upregulated. Genes involved in DNA synthesis, RNA processing and protein synthesis were downregulated, which correlates with reduced growth of the columnar plants, whereas those involved in transport and protein modification were upregulated. Considering phytohormone metabolism, genes of biosynthesis and signal transduction of IAA and jasmonates were induced. The opposite was reported for genes associated with GA biosynthesis and signal transduction. These results agree with the findings on phytohormone levels described in the previous section. Furthermore, there were hints to a cell cycle arrest in G2 in columnar apples, which would result in a lower number of cells and would thus explain the stunted growth. In summary, results suggested an alteration in cell wall and cell membrane formation resulting in smaller cells which in combination with the cell cycle arrest would lead to short spurs instead of long branches. Functions associated with membrane integrity like transport, photosynthesis and mitochondrial electron transport were also changed, which was considered to be a consequence of columnar growth.

In a second gene expression study, Krost et al. (2013) narrowed down the genes of interest to those involved in phytohormone biosynthesis, signaling and transport. At the same time they expanded the amount of data to six Illumina data sets, totaling almost 500 million reads, from three different time points (May 2009, September 2009 and July 2010) and validated their data with results of a microarray chip hybridized with RNAs of four collection dates between April and May as well as of quantitative real-time PCR carried out on RNA isolated from in vitro cultures of P28 and A14. It is doubtful whether in vitro cultures represent a suitable model for the study of transcriptional changes associated with altered tree architecture since they do not reach the same age and architectural complexity as trees, and in vitro-grown leaves have been shown to have altered methylation patterns compared with field-grown leaves, which might influence gene expression (Li et al. 2002). However, since the in vitro cultures were only used to confirm results that had already been obtained by RNA-seq and microarray analyses of in vivo material in this study, they provide the possibility to test the transferability of the results to plants grown under tightly

controlled conditions. Out of 619 genes found to be significantly differentially regulated by Krost et al. (2013), 16 were detected to be involved in the regulation of all major phytohormone groups, as can be expected for a phenotype affecting many different aspects of plant growth. An integration of the results suggested that an increase of IAA levels together with a higher basipetal IAA transport (due to upregulation of *Auxin Resistant* and *Pinformed 1*) is responsible for the high level of apical dominance and apical control of columnar apple trees. This is overcome by the elevated level of CKs when spurs are formed; however, they cease their elongation growth early. Additionally, the complex interplay between growth-promoting factors (IAA, CKs and apolar GAs) and growth-inhibitory factors (JAs and XTHs), which all showed upregulation in columnar trees in this study, is responsible for the achievement of normal plant height. Another 16 genes showing constant differential regulation in at least two of the three gene expression studies conducted were analyzed with regard to their chromosomal location. Five of them were found to reside from chromosome 10, indicating a statistically approved enrichment of differentially regulated genes in the *Co* gene vicinity. Four of them are associated with phytohormonal regulations, so Krost et al. (2013) drew the conclusion that the region is most likely covered by a selective sweep due to the influence of the *Co* gene.

Further time course gene expression experiments will be necessary in order to validate and increase our knowledge about the function of the *Co* gene product. It would also be helpful to expand these analyses to other tissues and developmental stages, yielding a comprehensive picture of the physiological state of the plant.

Conclusion

While a number of individual genes and signaling molecules regulating plant architecture have already been identified in model plants, it has not yet been possible to create a concise picture of the whole process since some key players still remain elusive and knowledge about the interaction of the individual pathways is scarce. The columnar growth habit of apple is a natural mutant showing altered plant shape and thus might be the key to an important factor in the regulation of tree architecture. During the past few years, a lot of progress in the understanding of the columnar growth habit has been made. It is caused by the dominant allele of the *Columnar* gene on chromosome 10 which probably influences IAA, CK and GA metabolism and signal transduction leading to stunted growth and nearly absent branching. Molecular markers created with the help of the apple genome sequence (Velasco et al. 2010) have successfully been used to fine map the *Co* locus to 18.51 – 19.09 Mb on chromosome 10. Since in-depth analysis of this region is already underway, disclosure of the identity of *Co* can be anticipated for the near future. Additional comparisons of growth regulator concentrations in columnar and normal type apple trees as well as in-depth transcriptome analyses would facilitate the discovery of *Co* function and affected signal pathways.

Acknowledgements

We are grateful to Prof. E.R. Schmidt for the critical review of the manuscript, to Ramona Petersen for linguistic proofreading, and to Bastienne Brauksiepe for allowing us to take pictures of the apple trees at the Geisenheim University. We apologize to colleagues whose work could not be cited due to space limitations. Our work was supported by grants of the Federal Ministry of Agriculture and Nutrition (no. 511-06.01-28-1-43.042-07 and no. 313-06.01-28-1-43.042-07). The authors declare that they have no conflict of interest.

References

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309:1052–1056

Aguilar-Martínez JA, Poza-Carrión C, Cubas P (2007) *Arabidopsis BRANCHED1* Acts as an Integrator of Branching Signals within Axillary Buds. Plant Cell 19:458–472

Agusti J, Herold S, Schwarz M, Sanchez P, Ljung K, Dun EA, Brewer PB, Beveridge CA, Sieberer T, Sehr EM, Greb T (2011) Strigolactone signaling is required for auxin-dependent stimulation of secondary growth in plants. PNAS 108:20242–20247

Aida M, Ishida T, Tasaka M (1999) Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the *CUP-SHAPED COTYLEDON* and *SHOOT MERISTEMLESS* genes. Development 126:1563–1570

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. J Mol Biol 215:403–410

Asano K, Miyao A, Hirochika H, Kitano H, Matsuoka M, Ashikari M (2010) SSD1, which encodes a plant-specific novel protein, controls plant elongation by regulating cell division in rice. P Jpn Acad B-Phys 86(3):265-273

Atzorn R, Weiler EW (1983) The Immunoassay of Gibberellins. Planta 159:7–11

Baba K, Karlber A, Schmidt J, Schrader J, Hvidsten TR, Bako L, Bhalerao RP (2011) PNAS 108:3418-3423

Bai T, Zhu Y, Fernández-Fernández F, Keulemans J, Brown S, Xu K (2012) Fine Genetic Mapping of the *Co* Locus Controlling Columnar Growth Habit in Apple. Mol Genet Genomics 287:437–450

Baldi P, Wolters PJ, Komjanc M, Viola R, Velasco R, Salvi S (2012) Genetic and physical characterisation of the locus controlling columnar habit in apple (*Malus × domestica* Borkh.). Mol Breeding. doi: 10.1007/s11032-012-9800-1

BaoY, Dharmwardhana P, Arias R, Allen MB, Ma C, Strauss SH (2009) *WUS* and *STM*-based reporter genes for studying mersistem development in poplar. Plant Cell Rep 28:947-962

Bendokas V, Gelvonauskienė D, Gelvonauskis B, Vinskienė J, Stanys V (2007) Identification of Apple Columnar Hybrids in Juvenile Phase Using Molecular Markers. Scientific Works of the Lithuanian Institute of Horticulture and Lithuanian University of Agriculture 26(3):289–295

Benková E, Michniewicz M, Sauer M, Teichmann T, Seifertová D, Jürgens G, Friml J (2003) Local, efflux-dependent auxin gradients as a common module for plant organ formation. Cell 115:591–602

Bennett T, Sieberer T, Willett B, Booker J, Luschnig C, Leyser O (2006) The *Arabidopsis MAX* pathway controls shoot branching by regulating auxin transport. Curr Biol 16:553–56

Björklund S, Antti H, Uddestrand I, Moritz T, Sundberg B (2007) Cross-talk between gibberellin and auxin in development of *Populus* wood: gibberellin stimulates polar auxin transport and has a common transcriptome with auxin. Plant J 52:499-511

Blazek J (1992) Segregation and General Evaluation of Spur Type or Compact Growth Habits in Apples. Acta Hort 317:71–79

Böhlenius H, Huang T, Charbonnel-Campaa L, Brunner AM, Jansson S, Strauss SH, Nilsson O (2006) *CO/FT* Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. Science 312:1040-1043

Brand U, Fletcher JC, Hobe M, Meyerowitz EM, Simon R (2000) Dependence of Stem Cell Fate in *Arabidopsis* on a Feedback Loop Regulated by *CLV3* Activity. Science 289:617–619

Brewer PB, Dun EA, Ferguson BJ, Rameau C, Beveridge CA (2009) Strigolactone acts downstream of auxin to regulate bud outgrowth in pea and *Arabidopsis*. Plant Physiol 150:482–93

Brown CL, McAlpine RG, Kormanik PP (1967) Apical Dominance and Form in Woody Plants: A Reappraisal. Am J Bot 54:153–162

Brown SK, Maloney KE, Hemmat M, Aldwinckle HS (2004) Apple Breeding at Cornell: Genetic Studies of Fruit Quality, Scab Resistance and Plant Architecture. Acta Hort 663:693–698

Brunner AM, Rottmann WH, Sheppard LA, Krutovskii K, DiFazio SP, Leonardi S, Strauss SH (2000) Structure and expression of duplicate AGAMOUS orthologues in poplar. Plant Mol Biol 44:619-634

Buban T (1996) In: Nieki J, Soltesz M (eds) Floral biology of temperate zone fruit trees and small fruits, Akademiai Kiado, Budapest, pp 3-54

Busch MA, Bomblies K, Weigel D (1999) Activation of a Floral Homeotic Gene in *Arabidopsis*. Science 285:585–587

Ceulemans R, Stettler RF, Hinckley TM, Isebrands JG, Heilmann PE (1990) Crown architecture of *Populus* clones as determined by branch orientation and branch characteristics. Tree Physiol 7:157-167

Chandler JW (2009) Auxin as compère in plant hormone crosstalk. Planta 231:1–12

Chen X, Zhang Z, Liu D, Zhang K, Li A, Mao L (2010) *SQUAMOSA* Promoter-Binding Protein-Like Transcription Factors: Star Players for Plant Growth and Development. J Integr Plant Bio 52:946–951

Chen Y, Yordanov YS, Ma C, Strauss S, Busov VB (2013) DR5 as a reporter system to study auxin response in *Populus*. Plant Cell Rep 32:453-463

Clark SE, Running MP, Meyerowitz EM (1993) *CLAVATA1*, a regulator of meristem and flower development in *Arabidopsis.* Development 119:397–418

Clark SE, Running MP, Meyerowitz EM (1995) *CLAVATA3* is a specific regulator of shoot and floral meristem development affecting the same processes as *CLAVATA1*. Development 2057–2067

Cline MG (1991) Apical Dominance. Bot Rev 57:318–358

Cline MG (1997) Concepts and Terminology of Apical Dominance. Am J Bot 84:1064–1069

Cline MG, Dong-Il K (2002) A Preliminary Investigation of the Role of Auxin and Cytokinin in Sylleptic Branching of Three Hybrid Poplar Clones Exhibiting Contrasting Degrees of Sylleptic Branching. Ann Bot-London 90:417-421

Conner PJ, Brown SK, Weeden NF (1998) Molecular Marker Analysis of Quantitative Traits for Growth and Development in Juvenile Apple Trees. Theor Appl Genet 96:1027–1035

Cooke JEK, Eriksson ME, Junttila O (2012) The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. Plant Cell Environ 35:1707-1728

Costes E, Guédon Y (2002) Modelling Branching Patterns on 1-year-old Trunks of Six Apple Cultivars. Ann Bot-London 89:513-524

Costes E, Lauri PE, Regnard JL (2006) Analyzing Fruit Tree Architecture: Implications for Tree Management and Fruit Production. Horticultural Reviews 32:1–61

Davey MW, Kenis K, Keulemans J (2006) Genetic Control of Fruit Vitamin C Contents. Plant Physiol 142:343–351

Dayan J, Voronin N, Gong F, Sun T-P, Hedden P, Fromm H, Aloni R (2012) Leaf-Induced Gibberellin Signaling Is Essential for Internode Elongation, Cambial Activity, and Fiber Differentiation in Tobacco Stems. Plant Cell 24(1):66-79

Domagalska MA, Leyser O (2011) Signal Integration in the Control of Shoot Branching. Nat Rev 12:211–221

Domagalska MA, Sarnowska E, Nagy F, Davis SJ (2010) Genetic Analyses of Interactions among Gibberellin, Abscisic Acid, and Brassinosteroids in the Control of Flowering Time in *Arabidopsis thaliana*. PLoS One 5:e14012

Douglas SJ, Chuck G, Dengler RE, Pelecanda L, Riggs CD (2002) *KNAT1* and *ERECTA* Regulate Inflorescence Architecture in *Arabidopsis*. Plant Cell 14:547–558

Du J, Groover A (2010) Transcriptional regulation of secondary growth and wood formation. J Integr Plant Biol 52(1):17-27

Du J, Mansfield SD, Groover AT (2009) The *Populus* homeobox gene *ARBORKNOX2* regulates cell differentiation during secondary growth. Plant J 60:1000–1014

Durbak A, Yao H, McSteen P (2012) Hormone signaling in plant development. Curr Opin Plant Biol 15:92–96

Eaton GW, Lapins KO (1970) Identification of Standard and Compact Apple Trees by Discriminant Function Analysis. J Appl Ecol 7:267–272

Edwards KD, Anderson PE, Hall A, Salathia NS, Locke JCW, Lynn JR, Straume M, Smith JQ, Millar AJ (2006) *FLOWERING LOCUS C* mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock. Plant Cell 18:639–650

Elias AA, Busov VB, Kosola KR, Ma C, Etherington E, Shevchenko O, Gandhi H, Pearce DW, Rood SB, Strauss SH (2012) Green Revolution Trees: Semidwarfism Transgenes Modify Gibberellins, Promote Root Growth, Enhance Morphological Diversity, and Reduce Competitiveness in Hybrid Poplar. Plant Physiol 160:1130-1144

Eriksson ME, Israelsson M, Olsson O, Moritz T (2000) Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. Nat Biotech 18:784-788

Eriksson ME, Moritz T (2002) Daylength and spatial expression of a gibberellin 20-oxidase isolated from hybrid aspen (*Populus tremula* L. x *P. tremuloides* Michx.). Planta 214:920–930

Farré EM (2012) The regulation of plant growth by the circadian clock. Plant Biol 14:401-410

Fernández-Fernández F, Evans KM, Clarke JB, Govan CL, James CM, Marič S, Tobutt KR (2008) Development of an STS map of an interspecific progeny of *Malus*. Tree Genet Genomes 4:469–479

Feucht W, Khan MZ, Daniel P (1974) Abscisic Acid in *Prunus* Trees : Isolation and the Effect on Growth of Excised Shoot Tissue. Physiol Plantarum 32:247–252

Fideghelli C, Sartori A, Grassi F (2003) Fruit Tree Size and Architecture. Acta Hort 622:279–293

Fisher DV (1969) Spur-type Strains of McIntosh for High Density Plantings. Bristish Columbia Fruit Growers' Association Quaterly Report 14:3–10

Fisher DV (1995) The "Wijcik Spur Mclntosh". Fruit Varieties J 49:212–213

Flachowsky H, Hättasch C, Höfer M, Peil A, Hanke MV (2010) Overexpression of *LEAFY* in Apple Leads to a Columnar Phenotype with Shorter Internodes. Planta 231:251–263

Friml J, Palme K (2002) Polar Auxin Transport – Old Questions and New Concepts? Plant Mol Biol 49:273–284

Gälweiler L, Guan C, Müller A, Wisman E, Mendgen K, Yephremov A, Palme K (1998) Regulation of Polar Auxin Transport by AtPIN1 in *Arabidopsis* Vascular Tissue. Science 282:2226-2230

Garrison R (1955) Studies in the Development of Axillary Buds. Am J Bot 42:257–266

Gelvonauskienë D, Gelvonauskis B, Sasnauskas A (2006) Impact of Rootstock on Columnar Apple Tree Growth in a Nursery. Scientific Works of the Lithuanian Institute of Horticulture and Lithuanian University of Agriculture 25:51–56

Gelvonauskis B, Brazaitytë A, Sasnauskas A, Duchovskis P, Gelvonauskienë D. (2006) Morphological and Physiological Characteristics of Columnar Apple Trees. Scientific Works of the Lithuanian Institute of Horticulture and Lithuanian University of Agriculture 25:350–356

Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C (1998) Simple Sequence Repeats for the Genetic Analysis of Apple. Theor Appl Genet 96:1069–1076

Gocal GF, King RW, Blundell CA, Schwartz OM, Andersen CH, Weigel D (2001) Evolution of Floral Meristem Identity Genes. Analysis of *Lolium temulentum* Genes Related to *APETALA1* and *LEAFY* of *Arabidopsis*. Plant Physiol 125:1788–1801

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

Gould PD, Locke JCW., Larue C, et al. (2006) The molecular basis of temperature compensation in the *Arabidopsis* circadian clock. Plant Cell 18:1177–1187

Graeber K, Nakabayashi K, Miatton E, Leubner-Metzger G, Soppe WJJ (2012) Molecular mechanisms of seed dormancy. Plant Cell Environ 35:1769-1786

Greb T, Clarenz O, Schäfer E, Muller D, Herrero R, Schmitz G, Theres K (2003) Molecular analysis of the *LATERAL SUPPRESSOR* gene in *Arabidopsis* reveals a conserved control mechanism for axillary meristem formation. Gene Dev 17:1175–1187

Groover AT (2005) What genes make a tree a tree? Trends Plant Sci 10:210–214

Groover AT, Mansfield SD, DiFazio SP, Dupper G, Fontana JR, Millar R, Wang Y (2006) The *Populus* homeobox gene *ARBORKNOX1* reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. Plant Mol Biol 61:917–932

Hackett WP (1985) Juvenility, Maturation and Rejuvenility in Woody Plants. Horticultural Reviews 7:109–155

Halaly T, Pang X, Batikoff T, Crane O, Keren A, Venkateswari J, Ogrodovitch A, Sadka A, Lavee S, Or E (2008) Similar mechanisms might be triggered by alternative external stimuli that induce dormancy release in grape buds. Planta 228:79–88

Han KM, Dharmawardhana P, Arias RS, Ma C, Busov V, Strauss SH (2011) Gibberellin-associated cisgenes modify growth, stature and wood properties in *Populus*. Plant Biotech J 9:162-178

Hanzawa Y, Takahashi T, Michael AJ, Burtin D, Long D, Pineiro M, Coupland G, Komeda Y (2000) *ACAULIS5*, an Arabidopsis gene required for stem elongation, encodes a spermine synthase. EMBO J 19(16):4248-4256

Hareven D, Gutfinger T, Parnis A, Eshed Y, Lifschitz E (1996) The making of a compound leaf: genetic manipulation of leaf architecture in tomato. Cell 84:735–44

Hartig K, Beck E (2006) Crosstalk between Auxin, Cytokinins, and Sugars in the Plant Cell Cycle. Plant Biol 8:389–396

Hattori Y, Nagai K, Furukawa S, et al. (2009) The ethylene response factors *SNORKEL1* and *SNORKEL2* allow rice to adapt to deep water. Nature 460:1026-1030

Hay A, Tsiantis M (2010) *KNOX* genes: versatile regulators of plant development and diversity. Development 137:3153–3165

Hegedűs A (2006) Review of the Self-Incompatibility in Apple (*Malus × domestica* Borkh., syn.: *Malus pumila* Mill.). J Hortic Sci 12:31–36

Heide OM (1974) Growth and dormancy in Norway spruce ecotypes (*Picea abies*) I. Interaction of photoperiod and temperature. Physiol Plantarum 30:1–12

Heide OM (1993) Daylength and thermal time responses of bud burst during dormancy release in some northern deciduous trees. Physiol Plantarum 88:531–540

Heide OM (2008) Interaction of photoperiod and temperature in the control of growth and dormancy of *Prunus* species. Sci Hortic 115:309–314

Heide OM, Prestrud AK (2005) Low temperature, but not photoperiod, controls growth cessation and dormancy induction and release in apple and pear. Tree Physiol 25:109–114

Hemmat M, Weeden NF, Conner PJ, Brown SK (1997) A DNA Marker for Columnar Growth Habit in Apple Contains a Simple Sequence Repeat. J Am Society Hortic Sci 122:347–349

Hibara K, Takada S, Tasaka M (2003) *CUC1* gene activates the expression of SAM-related genes to induce adventitious shoot formation. Plant J 36:687–696

Hirakawa Y, Kondo Y, Fukuda H (2010) *TDIF* Peptide Signaling Regulates Vascular Stem Cell Proliferation via the *WOX4* Homeobox Gene in *Arabidopsis*. Plant Cell 22:2618–2629

Hoad GV (1984) Hormonal regulation of fruit-bud formation in fruit trees. Acta Hort 149:13-24

Hsu C-Y, Adams JP, Kim H, et al. (2011) *FLOWERING LOCUS T* duplication coordinates reproductive and vegetative growth in perennial poplar. PNAS 108(26):10756-10761

Hsu C-Y, Liu Y, Luthe DS, Yuceer C (2006) Poplar *FT2* Shortens the Juvenile Phase and Promotes Seasonal Flowering. Plant Cell 18:1846-1861

Huijser P, Schmid M (2011) The control of developmental phase transitions in plants. Development 138:4117-4129

Ikase L, Dumbravs R (2004) Breeding of Columnar Apple Trees in Latvia. Biologija 2:8–10

Irish VF, Sussex IM (1990) Function of the *apetala-1* Gene during *Arabidopsis* Floral Development. Plant Cell 2:741–753

Israelsson M, Sundberg B, Moritz T (2005) Tissue-specific localization of gibberellins and expression of gibberellin-biosynthetic and signaling genes in wood-forming tissue in aspen. Plant J 44:494-504

Jacob HB (2010) Breeding Experiments of Apple Varieties with Columnar Growth and Low Chilling Requirements. Acta Hort 159–164

Jaillon O, Aury J-M, Noel B, et al. (The French-Italian Public Consortium for Grapevine Genome Characterization) (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463-467

Ji L, Wang J, Ye M, Li Y, Guo B, Chen Z, Li H, An X (2013) Identification and Characterization of the *Populus* AREB/ABF Subfamily. J Integr Plant Biol 55(2):177-186

Jiao Y, Wang Y, Xue D, et al. (2010) Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. Nat Genet 42:541–544

Johnson X, Brcich T, Dun EA, Goussot M, Haurogné K, Beveridge CA, Rameau C (2006) Branching Genes are Conserved across Species. Genes Controlling a Novel Signal in Pea are Coregulated by Other Long-Distance Signals. Plant Physiol 142:1014–1026

Jones B, Gunnerås SA, Petersson S V, Tarkowski P, Graham N, May S, Dolozal K, Sandberg G, Ljung K (2010) Cytokinin Regulation of Auxin Synthesis in *Arabidopsis* Involves a Homeostatic Feedback Loop Regulated via Auxin and Cytokinin Signal Transduction. Plant Cell 22:2956–2969

Junttila O (1976) Growth cessation and shoot tip abscission in *Salix*. Physiol Plantarum 38:278–286

Junttila O (2007) Regulation of annual shoot growth cycle in northern tree species. In: Taulavuori E, Taulavuori K (eds) Physiology of Northern Plants under Changing Environment, Research Signpost, Kerala, India pp. 177-210

Junttila O, Hänninen H (2012) The minimum temperature for budburst in *Betula* depends on the state of dormancy. Tree Physiol 32:337–345

Kardailsky I, Shukla VK, Ahn JH, Dagenais N, Christensen SK, Nguyen JT, Chory J, Harrison MJ, Weigel D (1999) Activation tagging of the floral inducer *FT*. Science 286:1962–1965

Karlberg A, Bakó L, Bhalerao RP (2011) Short day-mediated cessation of growth requires the downregulation of AINTEGUMENTALIKE1 transcription factor in hybrid Aspen. PLoS Genetics 7:e1002361

Kelsey DF, Brown SK (1992) ‚McIntosh Wijcik': A Columnar Mutation of ‚McIntosh' Apple Proving Useful in Physiology and Research. Fruit Varieties J 46:83–87

Kenis K, Keulemans J (2007) Study of Tree Architecture of Apple (*Malus × domestica* Borkh.) by QTL Analysis of Growth Traits. Mol Breeding 19:193–208

Kenis K, Keulemans J, Davey MW (2008) Identification and Stability of QTLs for Fruit Quality Traits in Apple. Tree Genet Genomes 4:647–661

Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS (2001) *KANADI* regulates organ polarity in *Arabidopsis*. Nature 411:706–709

Kim MY, Song KJ, Hwang J-H, Shin YU, Lee HJ (2003) Development of RAPD and SCAR markers Linked to the *Co* Gene Conferring Columnar Growth Habit in Apple (*Malus pumila* Mill.). J Hort Sci Biotech 78:512–517

Knox JP, Wareing PF (1984) Apical Dominance in *Phaseolus vulgaris* L.: The Possible Roles of Abscisic and Indole-3-Acetic Acid. J Exp Bot 35:239–244

Ko J-H, Prassinos C, Keathley D, Han K-H (2011) Novel aspects of transcriptional regulation in the winter survival and maintenance mechanism of poplar. Tree Physiol 31:208-225

Kobayashi Y, Kaya H, Goto K, Iwabuchi M, Araki T (1999) A pair of related genes with antagonistic roles in mediating flowering signals. Science 286:1960–1962

Koornneef M, Veen JH (1980) Induction and analysis of gibberellin sensitive mutants in *Arabidopsis thaliana* (L.) heynh. Theor Appl Genet 58:257–263

Koutinas N, Pepelyankov G, Lichev V (2010) Flower Induction and Flower Bud Development in apple and sweet cherry. Biotechnol & Biotechnol Eq. 24(1):1549-1558

Křeček P, Skůpa P, Libus J, Naramoto S, Tejos R, Friml J, Zažímalová E (2009) The PIN-FORMED (PIN) protein family of auxin transporters. Genome Biol 10:249

Krost C, Petersen R, Schmidt ER (2012) The Transcriptomes of Columnar and Standard Type Apple Trees (*Malus x domestica*) – a Comparative Study. Gene 498(2):223-230

Krost C, Petersen R, Lokan S, Brauksiepe B, Braun P, Schmidt ER (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

Lamb RS, Hill TA, Tan QK-G, Irish VF (2002) Regulation of *APETALA3* floral homeotic gene expression by meristem identity genes. Development 129:2079–2086

Lang GA (1987) Dormancy: a new terminology. J Hortic Sci 22:817-820

Lang GA, Early JD, Martin GC, Darnell RL (1987) Endo-, para- and ecodormancy: physiological terminology and classification for dormancy research. J Hortic Sci 22:371-377

Lane WD, Looney NE (1982) A Selective Tissue Culture Medium for Growth of Compact (Dwarf) Mutants of Apple. Theor Appl Genet 61:219–223

Langridge J (1957) Effect of Day-length and Gibberellic Acid on the Flowering of *Arabidopsis*. Nature 180:36–37

Lapins KO (1969) Segregation of Compact Growth Types in Certain Apple Seedling Progenies. Can J Plant Sci 49:765–768

Lapins KO (1974) Spur-type Growth Habit in 60 Apple Progenies. J Am Soc Hortic Sci 99:568–572

Lapins KO (1976) Inheritance of Compact Growth Type in Apple. J Am Soc Hortic Sci 101:133–135

Lapins KO, Watkins R (1973) Genetics of Compact Growth Habit. Report of East Malling Research Station for 1972 136

Larisch C, Dittrich M, Widhagen H, Lautner S, Fromm J, Polle A, Hedrich R, Rennenberg H, Müller T, Ache P (2012) Poplar Wood Rays Are Involved in Seasonal Remodeling of Tree Physiology. Plant Physiol 160:1515-1529

Laux T, Mayer KF, Berger J, Jürgens G (1996) The *WUSCHEL* gene is required for shoot and floral meristem integrity in *Arabidopsis*. Development 122:87–96

Lawson DM, Hemmat M, Weeden NF (1995) The Use of Molecular Markers to Analyze the Inheritance of Morphological and Developmental Traits in Apple. J Am Soc Hortic Sci 120:532–537

Lee JM, Looney NE (1977) Abscisic Acid Levels and Genetic Compaction in Apple Seedlings. Can J Plant Sci 5:81–85

Lee JM, Looney NE (1978) Changes in Abscisic Acid and Gibberellin Levels in Apple Seeds During Stratification and their Relationship to Genetic Compactation. Can J Plant Sci 58:761–767

Leyser O (2005) The Fall and Rise of Apical Dominance. Curr Opin Genet Dev 15:468–471

Li C, Bangerth F (1999) Autoinhibition of indoleacetic acid transport in the shoots of two-branched pea (*Pisum sativum*) plants and its relationship to correlative dominance. Physiol Plantarum 106:415–420

Li C, Yin C, Liu S (2004) Different responses of two contrasting *Populus davidiana* populations to exogenous abscisic acid application. Environ Exp Bot 51:237-246

Li J, Sima W, Ouyang B, et al. (2012) Tomato *SIDREB* gene restricts leaf expansion and internode elongation by downregulating key genes for gibberellin biosynthesis. J Exp Bot 63(18):6407-6420

Li P, Wang Y, Qian Q, Fu Z, Wang M, Zeng D, Li B, Wang X, Li J (2007) *LAZY1* controls rice shoot gravitropism through regulating polar auxin transport. Cell Res 17:402–410

Li X, Xu M, Korban SS (2002) DNA methylation profiles differ between field- and *in vitro*-grown leaves of apple. J Plant Physiol 159:1229-1234

Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van De Weg E, Gessler C (2002) Development and Characterisation of 140 New Microsatellites in Apple (*Malus x domestica* Borkh.). Mol Breeding 10:217–241

Liebhard R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C (2003) Mapping Quantitative Physiological Traits in Apple (*Malus x domestica* Borkh.). Plant Mol Biol 52:511–526

Liu A, Eaton GW (1970) Comparative Leaf Anatomy of Two Standard and Two Compact Apple Mutants. Can J Plant Sci 50:733–735

Long JA, Moan EI, Medford JI, Barton MK (1996) A member of the *KNOTTED* class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. Nature 379:66-69

Looney NE, Lane WD (1984) Spur-Type Growth Mutants of McIntosh Apple: A Review of their Genetics, Physiology and Field Performance. Acta Hort 146:31–46

Majer C, Hochholdinger F (2011) Defining the Boundaries: Structure and Function of LOB Domain Proteins. Trends Plant Sci 16:47–52

Maliepaard C, Alston FH, Van Arkel G, et al. (1998) Aligning Male and Female Linkage Maps of Apple (*Malus pumila* Mill.) Using Multi-allelic Markers. Theor Appl Genet 97:60–73

Mathan DS, Jenkins JA (1962) A Morphogenetic Study of *lanceolate*, a Leaf-Shape Mutant in the Tomato. Am J Bot 49:504–514

Matsumoto-Kitano M, Kusumoto T, Tarkowski P, Kinoshita-Tsujimura K, Václavíková K, Kakimoto T (2008) Cytokinins are central regulators of cambial activity. PNAS 105:20027–20031

Mauriat M, Moritz T (2009) Analyses of *GA20ox*- and *GID1*-over-expressing aspen suggest that gibberellins play two distinct roles in wood formation. Plant J 58:989-1003

Mayer KF, Schoof H, Haecker A,Lenhard M, Jürgens G, Laux T (1998) Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. Cell 95:805–815

McConnell JR, Emery J, Eshed Y, Bao N, Bowman J, Barton MK (2001) Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. Nature 411:709–713

McKay M, Ross JJ, Lawrence NL, Cramp RE, Beveridge CA, Reid JB (1994) Control of Internode Length in *Pisum sativum*. Plant Physiol 106:1521-1526

Meulenbroek J, Verhaegh J, Janse J (1998) Inheritance Studies with Columnar Type Trees. Acta Hort 484:255–258

Milborrow BV (1967) The Identification of (+)-Abscisin II [(+)-Dormin] in Plants and Measurement of its Concentrations. Planta 76:93–113

Mimida N, Kotoda N, Ueda T, Igarashi M, Hatsuyama Y, Iwanami H, Moriya S, Abe K (2009) Four *TFL1/CEN*-Like Genes on Distinct Linkage Groups Show Different Expression Patterns to Regulate Vegetative and Reproductive Development in Apple (*Malus* x *domestica* Borkh.) Plant Cell Physiolog 50(2):394-412

Mohamed R, Wang C-T, Ma C, et al. (2010) *Populus CEN/TFL1* regulates first onset of flowering, axillary meristem identity and dormancy release in *Populus*. Plant J 62:674-688

Moriya S, Iwanami H, Kotoda N, Takahashi S, Yamamoto T, Abe K (2009) Development of a Marker-Assisted Selection System for Columnar Growth Habit in Apple Breeding. J Japan Soc Hort Sci 78:279–287

Moriya S, Okada K, Haji T, Yamamoto T, Abe K (2012) Fine Mapping of *Co*, a Gene Controlling Columnar Growth Habit Located on Apple (*Malus × domestica* Borkh.) Linkage Group 10. Plant Breeding. doi: 10.1111/j.1439-0523.2012.01985.x

Müller D, Leyser O (2011) Auxin, cytokinin and the control of shoot branching. Ann Bot 107:1203–1212

Murray MB, Cannell MGR, Smith RI (1989) Date of bud burst of fifteen tree species in Britain following climatic warming. J Appl Ecol 26:693–700

Müssig C (2005) Brassinosteroid-Promoted Growth. Plant Biol 7:110–117

Newcomb RD, Crowhurst RN, Gleave AP, et al. (2006) Analyses of Expressed Sequence Tags from Apple. Plant Physiol 141:147–166

Nieminen K, Immanen J, Laxell M, et al. (2008) Cytokinin signaling regulates cambial development in poplar. PNAS 105:20032–20037

Nilsson J, Karlberg A, Antti H, Lopez-Vernaza M, Mellerowicz E, Perrot-Rechenmann C, Sandberg G, Bhalerao RP (2008) Dissecting the Molecular Basis of the Regulation of Wood formation by Auxin in Hybrid Aspen. Plant Cell 20:843-855

Nordström A, Tarkowski P, Tarkowska D, Norbaek R, Âstot C, Dolezal K, Sandberg G (2004) Auxin regulation of cytokinin biosynthesis in *Arabidopsis thaliana*: a factor of potential importance for auxin-cytokinin-regulated development. PNAS 101:8039–8044

Ophir R, Pang X, Halaly T, Venkateswari J, Lavee S, Galbraith D, Or E (2009) Gene-expression profiling of grape bud response to two alternative dormancy-release stimuli expose possible links between impaired mitochondrial activity, hypoxia, ethylene-ABA interplay and cell enlargement. Plant Mol Biol 71:402–423

Otsuga D, DeGuzman B, Prigge MJ, Drews GN, Clark SE (2001) *REVOLUTA* regulates meristem initiation at lateral positions. Plant J 25:223–236

Otto D, Petersen R, Krost C, Brandl R, Brauksiepe B, Braun P, Schmidt ER (2013) Molecular Characterization of the *Co* Gene Region in *Malus* x *domestica*. Acta Hort, in press

Overvoorde P, Fukaki H, Beeckman T (2010) Auxin Control of Root Development. Cold Spring Harb Perspect Biol 2(6):a001537

Palme K, Gälweiler L (1999) PIN-pointing the molecular basis of auxin transport. Curr Opin Plant Biol 2:375-381

Park S, Keathley DE, Han KH (2008) Transcriptional profiles of the annual growth cycle in *Populus deltoides*. Tree Physiol 28:321–329

Peng J, Harberd NP (1997) Gibberellin deficiency and response mutations suppress the stem elongation phenotype of phytochrome-deficient mutants of *Arabidopsis*. Plant Physiol 113:1051–1058

Peng J, Richards DE, Hartley NM, et al. (1999) "Green revolution" genes encode mutant gibberellin response modulators. Nature 400:256–261

Péret B, Swarup K, Ferguson A, et al. (2012) *AUX/LAX* Genes Encode a Family of Auxin Influx Transporters That Perform Distinct Functions during *Arabidopsis* Development. Plant Cell 24:2874-2885

Phinney BO (1985) Gibberellin A1 Dwarfism and Shoot Elongation in Higher Plants. Biologia Plantarum 27:172–179

Popko J, Hänsch R, Mendel R-R, Polle A, Teichmann T (2010) The role of abscisic acid and auxin in the response of poplar to abiotic stress. Plant Biol 12:242-258

Qi W, Sun F, Wang Q, Chen M, Huang Y, Feng Y-Q, Luo X, Yang J (2011) Rice Ethylene-Response AP2/ERF Factor *OsEATB* Restricts Internode Elongation by Down-Regulating a Gibberellin Biosynthetic Gene. Plant Physiol 157:216-228

Quinlan JD, Tobutt KR (1990) Manipulating Fruit Tree Structure Chemically and Genetically for Improved Performance. Hort Sci 25:60–64

Rameau C (2010) Strigolactones, a novel class of plant hormone controlling shoot branching. Comptes rendus biologies 333:344–9

Randoux M, Jeauffre J, Thouroude T, Vasseur F, Hamama L, Juchaux M, Sakr S, Foucher F (2012) Gibberellins regulate the transcription of the continuous flowering regulator, *RoKSN*, a rose *TFL1* homologue. J Exp Bot 63(18):6543-6554

Reinhardt D, Kuhlemeier C (2001) Phyllotaxis in higher plants. In: McManus MT, Veit B (eds.) Meristematic Tissues in Plant Growth and Development, 1st edn, Wiley, New York, pp 172 – 212

Reinhardt D, Mandel T, Kuhlemeier C (2000) Auxin regulates the initiation and radial position of plant lateral organs. Plant Cell 12:507–518

Rinne PLH, Welling A, Vahala J, Ripel L, Ruonala R, Kangasjärvi J, van der Schoot C. (2011) Chilling of dormant buds hyperinduces FLOWERING LOCUS T and recruits GA-inducible 1,3-β-glucanases to reopen signal conduits and release dormancy in *Populus*. Plant Cell 23:130–146

Riou-Khamlichi C, Huntley R, Jacqmard A, Murray JAH (1999) Cytokinin Activation of *Arabidopsis* Cell Division Through a D-Type Cyclin. Science 283:1541–1544

Rohde A, Bhalerao RP (2007) Plant dormancy in the perennial context. Trends Plant Sci 12(5):217-222

Running MP, Meyerowitz EM (1996) Mutations in the *PERIANTHIA* gene of *Arabidopsis* specifically alter floral organ number and initiation pattern. Development 122:1261–1269

Sachs T, Thimann KV (1967) The Role of Auxins and Cytokinins in the Release of Buds From Dominance. Am J Bot 54:136–144

Ruttink T, Arend M, Morreel K, Storme V, Rombauts S, Fromm J, Bhalerao RP, Boerjan W, Rohde A (2007) A molecular timetable for apical bud formation and dormancy induction in poplar. Plant Cell 19:2370–2390

Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G (2000) Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. Science 288:1613–1616

Santamaria ME, Hasbun R, Valera MJ, Meijon M, Valledor L, Rodriguez JL, Toorop PE, Canal MJ, Rodriguez R (2009) Acetylated H4 histone and genomic DNA methylation patterns during bud set and bud burst in *Castanea sativa*. J Plant Physiol 166:1360–1369.

Santamaria ME, Rodriguez R, Canal MJ, Toorop PE (2011) Transcriptome analysis of chestnut (*Castanea sativa*) tree buds suggests a putative role for epigenetic control of bud dormancy. Ann Bot-London 108:485–498

Sarwar M, Skirvin RM (1997) Effect of Thidiazuron and 6-Benzylaminopurine on Adventitious Shoot Regeneration from Leaves of Three Strains of "McIntosh" Apple (*Malus* x *domestica* Borkh.) in vitro. Sci Hortic 68:95–100

Sarwar M, Skirvin RM, Kushad M, Norton MA (1998) Selecting Dwarf Apple (*Malus × domestica* Borkh.) Trees in vitro: Multiple Cytokinin Tolerance Expressed among Three Strains of "McIntosh" that Differ in their Growth Habit under Field Conditions. Plant Cell Tiss Org 54:71–76

Schoof H, Lenhard M, Haecker A, Mayer KF, Jürgens G, Laux T (2000) The Stem Cell Population of *Arabidopsis* Shoot Meristems Is Maintained by a Regulatory Loop between the *CLAVATA* and *WUSCHEL* Genes. Cell 100:635–644

Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M, Sandberg G (2004) A High-Resolution Transcript Profile across the Wood-Forming Meristem of Poplar Identifies Potential Regulators of Cambial Stem Cell Identity. Plant Cell 16:2278–2292

Schuster M (2009) Sour Cherries *Prunus cerasus* L. with Columnar Tree Habit. Acta Hort 814:325–328

Scorza R, Bassi D, Liverani A (2002) Genetic Interactions of Pillar (Columnar), Compact, and Dwarf Peach Tree Genotypes. J Am Soc Hortic Sci 127:254–261

Sehr EM, Agusti J, Lehner R, Farmer EE, Schwarz M, Greb T (2010) Analysis of secondary growth in the *Arabidopsis* shoot reveals a positive role of jasmonate signaling in cambium formation. Plant J 63(5):811-822

Seleznyova AN, Tustin DS, Thorp TG (2008) Apple Dwarfing Rootstocks and Interstocks Affect the Type of Growth Units Produced during the Annual Growth Cycle: Precocious Transistion to Flowering Affects the Composition and Vigour of Annual Shoots. Ann Bot-London 101:679-687

Shani E, Yanai O, Ori N (2006) The role of hormones in shoot apical meristem function. Curr Opin Plant Biol 9:484–489

Shannon S, Meeks-Wagner DR (1991) A Mutation in the *Arabidopsis TFL1* Gene Affects Inflorescence Meristem Development. Plant Cell 3:877–892

Shimizu-sato S, Mori H (2001) Control of Outgrowth and Dormancy in Axillary Buds. Plant Physiol 127:1405–1413

Shulaev V, Sargent DJ, Crowhurst RN, et al. (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43:109-116

Siegfried KR, Eshed Y, Baum SF, Otsuga D, Drews GN, Bowman JL (1999) Members of the *YABBY* gene family specify abaxial cell fate in *Arabidopsis*. Development 126:4117–4128

Silfverberg-Dilworth E, Matasci CL, Van de Weg WE, et al. (2006) Microsatellite Markers Spanning the Apple (*Malus* x *domestica* Borkh.) Genome. Tree Genet Genomes 2:202–224

Snow M, Snow R (1942) The Determination of Axillary Buds. New Phytol 41:13–22

Snow R (1929) The Transmission of Inhibition through Dead Stretches of Stem. Ann Bot 3:261–267

Srikanth A, Schmid M (2011) Regulation of flowering time: all roads lead to Rome. Cell Mol Life Sci 68:2013-2037

Srinivasan C, Dardick C, Callahan A, Scorza R (2012) Plum (*Prunus domestica*) Trees Transformed with Poplar *FT1* Result in Altered Architecture, Dormancy Requirement, and Continuous Flowering. PLoS ONE 7:e40715

Suarez-Lopez P,Wheatley K, Robson F, Onouchi H, Valverde F, Coupland G (2001) *CONSTANS* mediates between the circadian clock and the control of flowering in *Arabidopsis*. Nature 410:1116–1120

Sun T-P, Kamiya Y (1994) The *Arabidopsis GA1* Locus Encodes the Cyclase ent-Kaurene Synthetase A of Gibberellin Biosynthesis. Plant Cell 6:1509–1518

Sussex IM (1955) Morphogenesis in *Solanum tuberosum* L.: experimental investigation of leaf dorsoventrality and orientation in the juvenile shoot. Phytomorphology 5:286–300

Swarup K, Benkovà E, Swarup R, et al. (2008) The auxin influx carrier LAX3 promotes lateral root emergence. Nat Cell Biol 10(8):946-954

Swarup R, Péret B (2012) AUX/LAX family of auxin influx carriers – an overview. Front Plant Sci 3:225

Talon M, Koornneef M, Zeevaart JAD (1990) Endogenous gibberellins in *Arabidopsis thaliana* and possible steps blocked in the biosynthetic pathways of the semidwarf *ga4* and *ga5* mutants. PNAS 87:7983–7987

Thimann KV, Skoog F (1933) Studies on the Growth Hormone of Plants. III. The Inhibiting Action of the Growth Substance on Bud Development. PNAS 19:714–716

Tian Y-K, Wang C-H, Zhang J-S, James C, Dai H-Y (2005) Mapping *Co*, a Gene Controlling the Columnar Phenotype of Apple, with Molecular Markers. Euphytica 145:181–188

Tobutt KR (1994) Combining Apetalous Parthenocarpy with Columnar Growth Habit in Apple. Euphytica 77:51–54

Tobutt KR (1985) Breeding Columnar Apple Trees at East Malling. Acta Hort 159:63–68

Tobutt KR (1988a) Columnar Apple Tree - Maypole Variety. US Plant Patent No 6,184

Tobutt KR (1988b) Columnar Apple Tree - Telamon Variety. US Plant Patent No 6,224

Tobutt KR (1988c) Columnar Apple Tree - Tuscan Variety. US Plant patent No 6,225

Tobutt KR (1988d) Columnar Apple Tree - Trajan Variety. US Plant patent No 6,226

Todaka D, Nakashima K, Maruyama K, et al. (2012) Rice phytochrome-interacting factor-like protein OsPIL1 functions as a key regulator of internode elongation and induces a morphological response to drought stress. PNAS 109(39):15947-15952

Tsuge T, Tsukaya H, Uchimiya H (1996) Two independent and polarized processes of cell elongation regulate leaf blade expansion in *Arabidopsis thaliana* (L.) Heynh. Development 122:1589–1600

Tsukaya H (2005) Leaf shape: genetic controls and environmental factors. Int J Dev Biol 49:547–555

Tuskan GA, DiFazio S, Jansson S, et al (2006) The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313(5793):1596-160

Uggla C, Mellerowicz EJ, Sundberg B (1998) Indole-3-Acetic Acid Controls Cambial Growth in Scots Pine by Positional Signaling. Plant Physiol 117:113-121

Uggla C, Moritz T, Sandberg G, Sundberg B (1996) Auxin as a positional signal in pattern formation in plants. PNAS 93:9282-9286

Umehara M, Hanada A, Yoshida S, et al. (2008) Inhibition of shoot branching by new terpenoid plant hormones. Nature 455:195–200

Valverde F, Mouradov A, Soppe W, Ravenscroft D, Samach A, Coupland G (2004) Photoreceptor Regulation of CONSTANS Protein in Photoperiodic Flowering. Science 303(5660):1003-1006

Veit B, Briggs SP, Schmidt RJ, Yanofsky MF, Hake S (1998) Regulation of leaf initiation by the *terminal ear 1* gene of maize. Nature 393:166–168

Velasco R, Zharkikh A, Affourtit J, et al. (2010) The Genome of the Domesticated Apple (*Malus × domestica* Borkh.). Nat Genet 42:833–839

Verde I, Abbott AG, Scalabrin S, et al. (The International Peach Genome Initiative) (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45(5):487-494

Wada M, Cao Q, Kotoda N, Soejima J, Masuda T (2002) Apple has two orthologues of *FLORICAULA/LEAFY* involved in flowering. Plant Mol Biol 49:567-577

Wang A, Aldwinckle H, Forsline P, Main D, Fazio G, Brown S, Xu K (2012) EST Contig-Based SSR Linkage Maps for *Malus × domestica* cv Royal Gala and an Apple Scab Resistant Accession of *M. sieversii*, the Progenitor Species of Domestic Apple. Mol Breeding 29:379–397

Wang J-W, Park MY,Wang L-J, Koo Y, Chen X-Y, Weigel D, Poethig RS (2011) MiRNA control of vegetative phase change in trees. PLoS Genetics 7, e1002012

Wang Y, Li J (2008) Molecular basis of plant architecture. Annu Rev Plant Biol 59:253–79

Watanabe M, Suzuki A, Komori S, Bessho H (2004) Comparison of Endogenous IAA and Cytokinins in Shoots of Columnar and Normal Type Apple Trees. J Japan Soc Hort Sci 73:19–24.

Watanabe M, Suzuki A, Komori S, Bessho H (2006) Effects of Heading-back Pruning on Shoot Growth and IAA and Cytokinin Concentrations at Bud Burst of Columnar-type Apple Trees. J Japan Soc Hortic Sci 75:224–230

Watanabe M, Bessho H, Suzuki A, Komori S (2008) Seasonal Changes of IAA and Cytokinin in Shoots of Columnar Type Apple Trees. Acta Hort 774:75–80

Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. Cell 69:843–859.

Werner T, Motyka V, Laucou V, Smets R, Van Onckelen H, Schmülling T (2003) Cytokinin-Deficient Transgenic Arabidopsis Plants Show Multiple Developmental Alterations Indicating Opposite Functions of Cytokinins in the Regulation of Shoot and Root Meristem Activity. Plant Cell 15:2532–2550

Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. Science 309:1056–1059

Wilson BF (2000) Apical control of branch growth and angle in woody plants. Am J Bot 87(5):601-607

Wilson RN, Heckman JW, Somerville CR (1992) Gibberellin Is Required for Flowering in *Arabidopsis thaliana* under Short Days. Plant Physiol 100:403–408

De Wit I, Pauwels E, Keulemans J (2000) Different Growth Habits in Apple and Correlation Between Growth Characteristics in Progenies with a Common *Co*-Gene Parent. Acta Hortic 325–330

Wullschleger SD, Weston DJ, DiFazio SP, Tuskan GA (2013) Revisiting the sequencing of the first tree genome: *Populus trichocarpa*. Tree Physiol 33(4):357-364

Xu Y, Chang P-FL, Liu D, Narasimhan ML, Raghothama KG, Hasegawa PM, Bressan RA (1994) Plant Defense Genes Are Synergistically Induced by Ethylene and Methyl Jasmonate. Plant Cell 6:1077-1085

Yamane H, Ooka T, Jotatsu H, Hosaka Y, Sasaki R, Tao R (2011) Expressional regulation of *PpDAM5* and *PpSAM6*, peach (*Prunus persica*) dormancy associated MADS-box genes, by low temperature and dormancy-breaking reagent treatment. J Exp Bot 62:3481–3488

Yamamuro C, Ihara Y, Wu X, Noguchi T, Fujioka S, Takatsuto S, Ashikari M, Kitano H, Matsuoka M (2000) Loss of Function of a Rice *brassinosteroid insensitive1* Homolog Prevents Internode Elongation and Bending of the Lamina Joint. Plant Cell 12:1591-1605

Yoo SK, Chung KS, Kim J, Lee JH, Hong SM, Yoo SJ, Yoo SY, Lee JS, Ahn JH (2005) *Constans* Activates *Suppressor of Overexpression of Constans 1* through *Flowering Locus T* to promote Flowering in *Arabidopsis*. Plant Physiol 139:770-778

Yu B, Lin Z, Li H, et al. (2007) *TAC1*, a major quantitative trait locus controlling tiller angle in rice. Plant J 52:891–898

Zawaski C, Kadmiel M, Pickens J, Ma C, Strauss S, Busov V (2011) Repression of gibberellin biosynthesis or signaling produces striking alterations in poplar growth, morphology and flowering. Planta 234:1285-1298

Zhang J, Elo A, Helariutta Y (2011) *Arabidopsis* as a model for wood formation. Curr Opin Biotech 22:293-299

Zhang Y, Dai HY (2011) Comparison of Photosynthetic and Morphological Characteristics, and Micro-Structure of Roots and Shoots, between Columnar Apple and Standard Apple Trees of Hybrid Seedlings. Phyton 80:119–125

Zhang Y, Turner JG (2008) Wound-Induced Endogenous Jasmonates Stunt Plant Growth by Inhibiting Mitosis. PLoS One 3(11):e3699

Zhang Y, Zhu J, Dai H (2012) Characterization of Transcriptional Differences Between Columnar and Standard Apple Trees Using RNA-Seq. Plant Mol Biol Reporter. doi: 10.1007/s11105-011-0396-0

Zhou H-L, He S-J, Cao Y-R, Chen T, Du B-X, Chu C-C, Zhang J-S, Chen S-Y (2006) OsGLU1, a putative membrane-bound endo-1,4-β-D-glucanase from rice, affects plant internode elongation. Plant Mol Biol 60:137-151

Zhu YD, Zhang W, Li GC, Wang T (2007) Evaluation of Inter-Simple Repeat Analysis for Mapping the *Co* Gene in Apple (*Malus pumila* Mill.). J Hortic Sci Biotech 82:371–376

Paper 2

Molecular Characterization of the *Co* Gene Region in *Malus* x *domestica*

Dominik Otto[1], Romina Petersen[1], Clemens Krost[1], Rachel Brandl[2], Bastienne Brauksiepe[2], Peter Braun[2], Erwin R. Schmidt[1]

[1] Department of Molecular Genetics, Johannes Gutenberg-University of Mainz
[2] Departement of Pomology, Geisenheim Research Centre

This article will be published in the BiotechFruit2012 special issue of Acta Horticulturae in 2014 and has been included in this thesis with permission.

Keywords

apple, BAC library, *Co* gene, columnar growth, molecular markers, NGS

Abstract

The columnar phenotype in apple is characterized by a stunted growth with short internodes and lateral branches being converted into fruit spurs. Columnar growth is due to a spontaneous mutation in the cultivar 'McIntosh' discovered in the 1960s. It seems to be controlled by a single dominant allele of the *Co* gene, which displays a typical Mendelian way of inheritance. So far all tested columnar apple trees had the *Co* gene in an heterozygous state. We developed PCR assays on the basis of molecular markers closely linked to the *Co* gene by screening the F1 progeny of a cross between a columnar and a non-columnar cultivar. These markers, combined with markers published earlier by other groups, allowed us to define a genomic region of approximately 1 Mbp on chromosome 10 probably containing the *Co* locus. In order to analyze this region in detail we constructed several genomic BAC libraries of a columnar cultivar. The BAC libraries consisted of more than 100,000 clones representing about 4x haploid apple tree genome equivalents. PCR-based screening and colony filter hybridization of the BAC libraries delivered about 37 BAC clones originating from the *Co* region. These clones were sequenced using a combination of Sanger and Illumina next generation sequencing. These sequences were assembled into a contig spanning most of the *Co* region. Selected genes from this region were analyzed in detail with regard to gene structure and single nucleotide polymorphisms to identify potential candidate genes for the *Co* gene. 24 genes displayed mutations leading to amino acid substitutions in the columnar growing cultivar and hence are considered as candidates for the *Co* gene.

Introduction

In the early 1960s the columnar growth phenotype in apple was discovered as a spontaneous mutation of the cultivar 'McIntosh' in British Columbia (Fisher, 1969). The columnar phenotype is characterized by short internodes, lateral branches being converted into fruit spurs and a compact, dense growth (Lapins, 1969). It is an extreme form of spur type growth (Eaton and Lapins, 1970). The very compact growth habit gives rise to economic advantages like high density planting, minimal pruning and mechanical harvesting and therefore provides valuable resources for apple breeding (Tobutt, 1985; Jacob, 2010). The columnar growth is determined by a single dominant allele of the *Co* gene (Lapins, 1969; Looney and Lane, 1984). Using molecular markers, it was shown that the *Co* gene is located on chromosome 10 and that it occurs in a heterozygous state (Tian et al., 2005).

Molecular markers are important tools for localizing and fine mapping of a gene region (Tian et al., 2005). When we started our project, several molecular markers linked to the *Co* gene were already available such as Ch03d11 (Liebhard et al., 2002), Hi01a03 (Silfverberg-Dilworth et al., 2006), $SCAR_{216}$ and $SCAR_{682}$ (Tian et al., 2005). The marker analysis showed that recombination frequencies of different cultivars vary significantly concerning the linkage of a columnar marker to the *Co* gene (Moriya et al., 2009).

The motivation for the current study was the detailed analysis of the structure of the *Co* gene region, the identification of potential candidate genes for *Co* and narrow down the region as much as possible. For that purpose we had to develop new molecular markers tightly linked to the *Co* gene and we tried to clone the entire region from a cultivar carrying the *Co* allele heterozygously. Our cloning strategy was to construct a genomic BAC library, using genomic DNA from *in vitro* cultured columnar specimens. We analyzed 61 genes in the target region and finally ended up with 24 candidate genes for *Co*.

Material and Methods

Plant Material

In vitro cultures of the columnar cultivar 'Procats28' (Geisenheim Research Center) were used for the construction of BAC libraries. *In vitro* cultures were alternately grown at 22°C on MS medium (Murashige and Skoog, 1962) containing sorbitol (30 g/L), gluconic acid (0,625 g/L), 6-benzylaminopurine (1 mg/L), indole-3-butyric acid (0,1 mg/L) and agar (8 g/L) and on one-half strength MS medium containing sorbitol (30 g/L), gluconic acid (0,625 g/L), 6-benzylaminopurine (0,5 mg/L), indole-3-butyric acid (0,05 mg/L) and agar (8 g/L). For the linkage studies, 95 individual offspring from crosses performed in 2007 between the columnar cultivar 'Procats 28' and 'A14-190-93' were used. 'Procats 28' originated from a cross between columnar 'Flamenco' and non-columnar 'Topaz', 'A14-190-93' from 'Golden Delicious Weinsberg' and 'Waltz'. All progeny plants were evaluated regarding columnar growth in September 2009, 2010 and 2011.

DNA Extraction and BAC Library Construction

The *in vitro* cultures were homogenized by a hand-held blender. Isolation of nuclei was performed according to the protocols of Burgoyne et al. (1970) and Burgoyne (1977) with some modifications. The suspension of nuclei was incubated in lysis buffer containing 2 % SDS and 1 mg/ml proteinase K for 10 min at 60°C, 60 min at 37°C and then dialyzed overnight against ddH_2O. After two phenol chloroform isoamyl alcohol and two chloroform isamyl alcohol extractions, the DNA was digested for 1.5 min using *Sau*3A1 (Roche, Mannheim, Germany) immediately followed by another phenol chloroform isoamyl alcohol extraction. Size fractionation was performed twice using 0.6 % agarose gels ("Horizontal System for Submerged Gel Electrophoresis", Life Technologies™, Gaithersburg, USA). High molecular weight DNA was ligated into *Bam*H1 digested pBeloBAC11 (Shizuya et al., 1992) for 72 h at RT and transformed into *Escherichia coli* strain DH10B by electroporation (Sheng et al., 1995). Probes to be used for library screening were designed on the basis of the published genome sequence (Velasco et al., 2010). After *in vitro* amplification the probes were tested for repetitive sequences using Southern and Pirotta hybridization (Southern, 1975; Buhariwalla et al., 1995). The colony hybridization was performed according to Grunstein and Hogness (1975). For the linkage studies, DNA from leaves of trees was extracted according to Eimert et al. (2003).

Sequencing and Assembly of BAC Clones

Using the PureYield™ Plasmid Maxiprep System (Promega, Madison, USA) plasmids from positively selected colonies were isolated. Based on BAC end sequencing (Sanger) and BLASTn searches against the 'Golden Delicious' genome (Velasco et al., 2010) the correct localization of the BAC clones could be validated. BACs were sequenced using a combination of Illumina® next generation sequencing (Illumina® HiSeq 2000) and Sanger sequencing (GENterprise Genomics, Mainz, Germany). Assembly of BAC sequences was conducted using the *de novo* assembly tool of CLC Genomics Workbench (CLC bio, Aarhus, Denmark) and SeqMan NGEN2® (DNASTAR®, Madison, USA). Gaps, repeats and critical positions were verified by PCR and Sanger sequencing. To create metacontigs, the BAC sequences were assembled using SeqMan™ and MegAlign™ (DNASTAR®, Madison, USA). Gaps in the metacontig were closed by long range PCR. The long range PCRs were performed using TaKaRa LA *Taq*® DNA polymerase (TaKaRa/ Clontech, Mountain View, USA).

Generation of Molecular Markers

The comparison between the BAC sequences of the columnar cultivar 'Procats28' und the published 'Golden Delicious' genome (Velasco et al., 2010) enabled us to identify insertions and deletion (Indels, Vali et al. 2008). Primers flanking these Indels were designed, purchased by Invitrogen™ (Darmstadt, Germany) and tested on genomic DNA of 'Procats28' (columnar) and 'A14-190-93' (non-columnar). After a first round of PCR only those primer pairs were further considered which showed polymorphic bands between 'Procats28' and 'A14-190-93'. These primers were tested on an F1 population of a cross between 'Procats28' and 'A14-190-93'. The amplified fragments were cloned into the pGEM®-T Easy Vector System (Promega, Madison, USA) and sequenced to validate the correct genomic localization. All PCR reactions were performed with an initial denaturation of 95°C for 5 min followed by 40 cycles of 94°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec. A final

extension of 72°C for 10 min was included. PCR products were analysed by horizontal agarose gel electrophoresis.

Comparative SNP Detection in Individual Genes

For SNP detection, genomic DNA was isolated from *in vitro* cultures of 'A14-190-93' and 'Procats28' using the innuPrep Plant DNA kit (Analytik Jena, Jena, Germany). Primers were designed in order to obtain PCR products covering all annotated exons as well as upstream and downstream regions of 500 bp of genes in the region. PCR reactions were performed as described above. PCR products were digested with 5 U Exonuclease I and 1 U Fast Alkaline Phosphatase (both Fermentas, St. Leon-Roth, Germany) at 37°C for 25 min and then heated to 75°C for 15 min.

Gene Structure Analysis

For the analysis of intron-exon structure, total RNA was isolated from in vitro cultures of 'A14-190-93' and 'Procats28' using the innuPREP Plant RNA kit (Analytik Jena, Jena, Germany). 800 ng of good quality RNA as tested on a Agilent 2100 Bioanalyzer (Agilent, Böblingen, Germany), was converted to cDNA using SuperScript® III Reverse Transcriptase (Invitrogen™, Darmstadt, Germany). Primers were designed to span introns and were purchased by Invitrogen™ (Darmstadt, Germany). PCR amplification, digestion with Exonuclease I and Fast Alkaline Phosphatase as well as sequencing and sequence analysis were conducted as described above.

Results and Discussion

Localization of the *Co* Gene Region

As a starting point for the localization of the *Co* gene region, we used the two markers $SCAR_{682}$ and $SCAR_{216}$ (Tian et al., 2005). The genetic distances are 2.9 and 12.3 cM to *Co*, respectively. With BLASTn searches against the 'Golden Delicious' genome (Velasco et al., 2010) we calculated the physical distance between the two SCAR markers. $SCAR_{682}$ was mapped to 17.542, while $SCAR_{216}$ was mapped to 22.271 Mbp of chromosome 10. The resulting physical distance between the two SCAR markers is about 4.729 Mbp. Using the equation $P_{xz} = P_{xy} + P_{yz} - 2P_{xy}P_{yz}$ (Lynch and Walsh, 1998), x being $SCAR_{682}$, y being *Co* and z being $SCAR_{216}$, we were able to locate *Co* roughly between 18 and 19.9 Mbp.

In order to identify indel based markers, the consensus sequence of all sequenced BACs was aligned to the 'Golden Delicious' genome (Velasco et al., 2010) and screened for appropriate (> 15 bp) insertions or deletions. Flanking primers were generated and tested on non-columnar 'A14-190-93' and columnar 'Procats28'. Following this approach, we obtained four markers which showed polymorphic bands for the columnar cultivar 'Procats28' (heterozygous) and a single band for non-columnar 'A14-190-93' (homozygous), therefore highly suitable for bulked segregant analysis (BSA). Subsequently, all four markers (distributed between 18.3 and 18.9 Mbp) were chosen for fine mapping and named after the BAC clone, in which they were found (Fig. 1). For the calculation of recombination frequencies (RFs), K25_M1, H1_M1, I2_3_M1 and 1C3_M2 (Table 1) were screened over a mapping population consisting of 95 progeny plants, 48 of which were columnar while 47

were non-columnar. This resulted in RFs of 0 % for all four markers (Fig. 2), indicating 100 % linkage to the *Co* gene. Additionally, three previous markers $SCAR_{216}$, (Tian et al., 2005), Hi01a03 (Silfverberg-Dilworth et al., 2006) and Ch03d11 (Liebhard et al., 2002) were included. As expected, these showed RFs of 6.32 % ($SCAR_{216}$), 3.16 % (Hi01a03) and 1.05 % (Ch03d11) proportional to their expected longer distance from *Co* (Fig. 2). The *Co* gene region could be delimited to 18 Mbp with Ch03d11 and K25_M1 in the proximal direction. Unfortunately, our linkage analysis does not provide us with a clear answer as to where the *Co* region ends distally beyond the 19 Mbp position. The next marker, Hi01a03, which shows recombination is located at 19.9 Mbp. We were not able to further fine-map this region because of the lack or recombination among the new markers located between 18.3 and 18.9 Mbp. It should be possible to further delimit the target region using larger numbers of progeny plants.

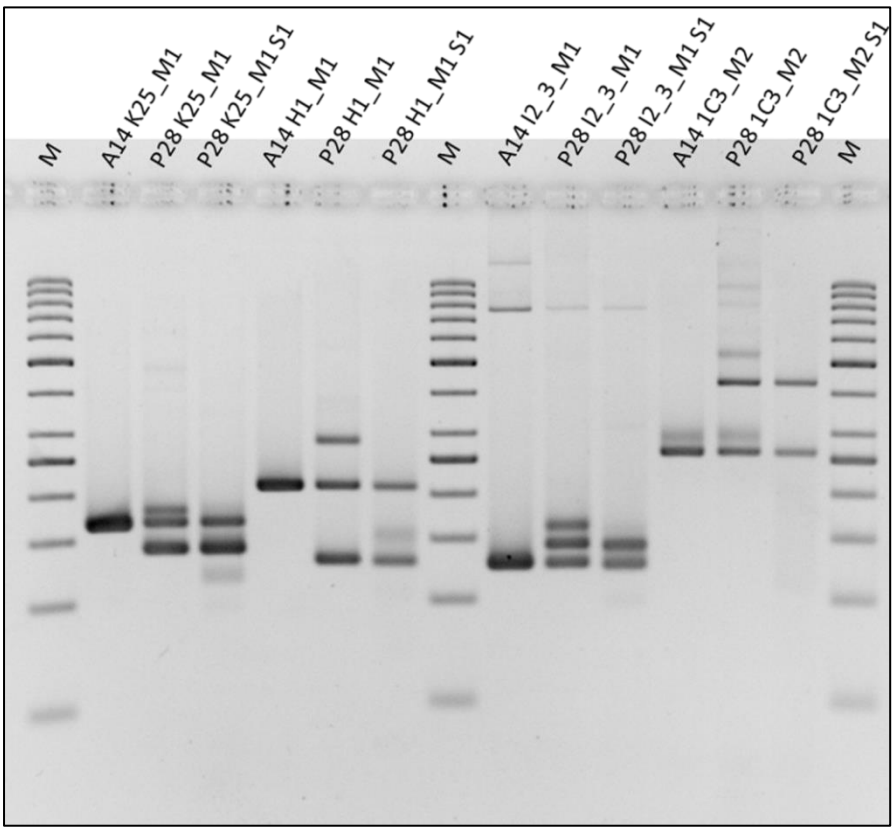


Fig. 1 PCR products of markers K25_M1, H1_M1, I2_3_M1 and 1C3_M2
A14: 'A14-190-93' (non-columnar); P28: 'Procats 28' (columnar); M: GeneRuler™ 50 bp DNA Ladder (Fermentas, St. Leon-Roth, Germany) electrophoresed on an agarose gel; S1: samples treated with S1 nuclease (Fermentas, St. Leon-Roth, Germany) to remove slow migrating bands due to open loops in heteroduplex DNA.

Table 1: Primer sequences, chromosomal localization and size of indel markers used for distinguishing columnar (C) and non-columnar (NC) apple trees.

Indel marker	Primer (5' – 3')	Localization (Chr10/Mbp)	Size (bp) C/NC
1C3_M2	Forward: AAGGTGAAGGTCCAGATCTG Reverse: AAGAGACCGATTGAGCTTGC	18.909	414/248
I2_3_M1	Forward: GCTACAGTGCTAACTAACTCTC Reverse: TATGTGCATCTGTATGCCGC	18.767	143/126
H1_M1	Forward: TTTGCGGTTTATATTCGAGCTG Reverse: CCGTAATTTCATGCGCTGAC	18.514	128/206
K25_M1	Forward: CCAAAGGAATAGAGACCCTG Reverse: GCAAGTGCTAAACCCTGTTC	18.304	136/157

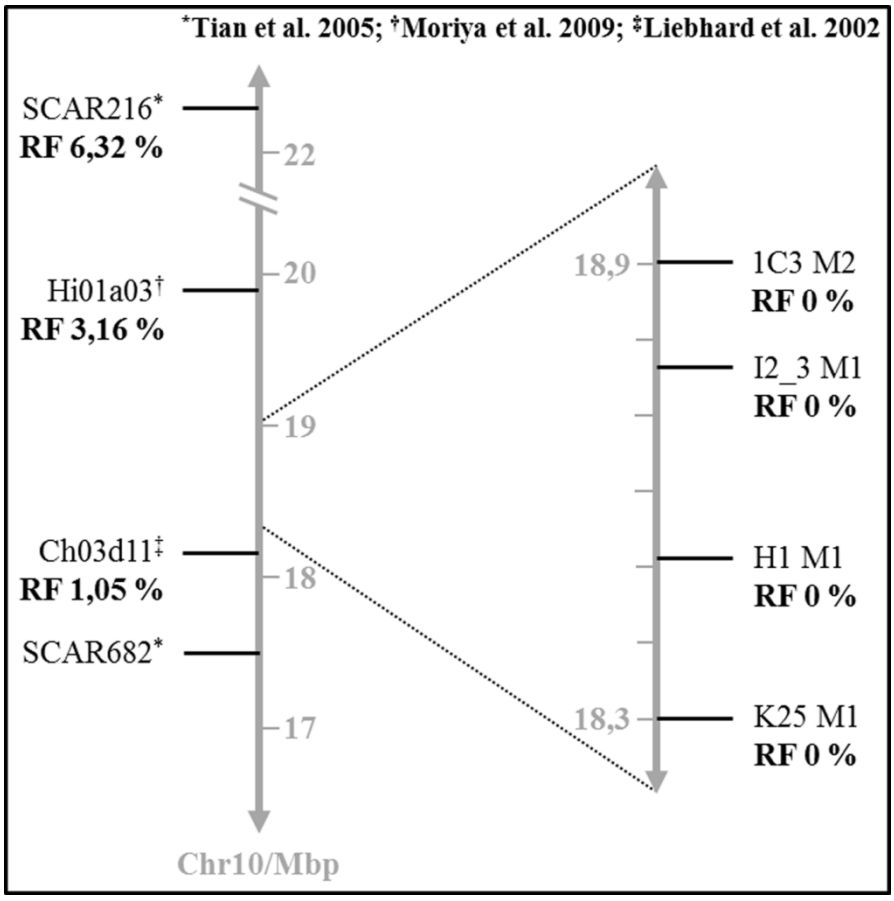


Fig. 2 Recombination frequencies (RFs) in the fine-mapped *Co* gene region
The figure shows the localization and RF of four previously reported ($SCAR_{216}$, Hi01a03, Ch03d11, $SCAR_{682}$) and four newly established markers (1C3_M2, I2_3_M1, H1_M1, K25_M1) between 17.5 and 22.3 Mbp on apple chromosome 10. RFs were calculated on the basis of 95 individuals. $SCAR_{682}$ was not suitable for analysis and therefore excluded.

BAC Library Construction, Screening and Contig Construction

The BAC libraries representing the genomic DNA of the columnar cultivar 'Procats28' consisted of more than 100,000 clones with an average insert size of 27 kb. With an estimated genome size of 742.3 Mb/haploid (Velasco et al., 2010) the libraries represent

about 4x haploid apple tree genome equivalents. Out of these 100,000 clones 37 were shown to originate from the target region. These 37 BACs were sequenced and assembled into individual contigs. Gaps were closed by long range PCR. Finally, all sequences were assembled into two metacontigs, leaving only one gap around position 18.7 Mb (Fig. 3). These two metacontigs are 590 and 190 kb in length, with a gap of about 35 kb between them. Unfortunately, it was not possible to span the gap with long range PCR, most likely due to its length. In future we will try to perform this using high throughput sequencing of Illumina mate pair libraries. Using *Co* gene linked markers the BAC clones could be assigned to the columnar and the non-columnar chromosome (Fig. 3). About 550 kbp originated from the columnar, about 400 kbp from the non-columnar chromosome. A detailed comparison of the metacontig sequences with the 'Golden Delicious' genome (Velasco et al., 2010) is still in progress.

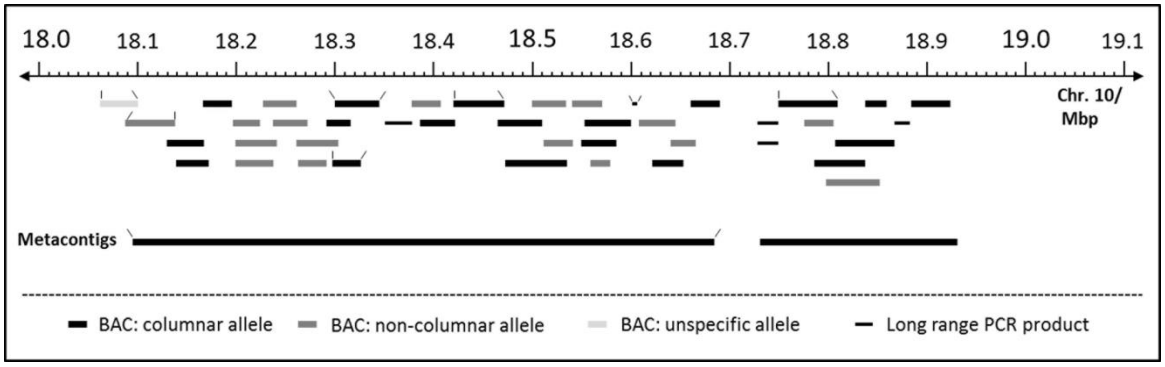


Fig. 3 Mapping of BAC clones and long range PCR products localized in the *Co* gene region on chromosome 10.

Comparative Analysis of Putative Candidate Genes

According to the 'Golden Delicious' genome sequence (Velasco et al., 2010), the newly defined target region of the Co gene contains 98 genes. We performed BLASTx searches of these genes against SwissProt and TAIR (The Arabidopsis Information Resource, Lamesch et al. 2011) and identified 61 genes which potentially could play a role in plant growth regulation. The coding regions of these 61 genes were sequenced as well from 'Procats28' as from 'A14-190-93'. 55 genes show differences in nucleotide sequence between these cultivars and among these, 24 genes carry up to seven missense mutations occurring heterozygously in 'Procats28' while being absent in 'A14-190-93'. Among them are transcription factors, kinases and phosphatases and proteins involved in phytohormone biosynthesis (Table 2). We are currently analyzing the intron-exon structure of these genes in order to verify the open reading frames. Furthermore, we will investigate whether the missense mutations are also found in other columnar and non-columnar cultivars. Eventually, the potential candidate genes will be transformed into non-columnar genotypes to validate their ability to generate a columnar growth phenotype.

Table 2: Genes located within the *Co* gene target region and carrying heterozygous missense mutations in columnar 'Procats28' compared with non-columnar 'A14-190-93'. Genes are categorized by their probable functional group and are indicated with their *Malus domestica* Protein (MDP) number, chromosomal region (in Mbp) and best BLASTx hit in the TAIR database.

Functional Group	MDP	TAIR Hit
Transcription factors	MDP0000286915	ERF/AP2 transcription factor family
	MDP0000307619	Scarecrow
Chromatin modification	MDP0000214258	Methyltransferases superfamily protein
	MDP0000191925	Lysine-specific demethylase PKDM7D
Kinases	MDP0000320862	Protein kinase family protein with LRR
	MDP0000308878	AGC kinase family protein
	MDP0000214257	Serine/threonine-protein kinase 19
RNA or protein degradation	MDP0000157859	ARM repeat superfamily protein
	MDP0000766466	RING/U-box superfamily protein
	MDP0000927098	RING/U-box superfamily protein
Protein-protein interactions	MDP0000320861	Protein containing two tandem BRCA1 domains
Transporter	MDP0000423958	Cation efflux family protein
	MDP0000254096	Mitochondrial import inner membrane translocase subunit AtTIM44-2
	MDP0000154627	ABC transporter C family member 14
	MDP0000927091	Autophagy 9
	MDP0000169714	Transducin family protein
Membrane modification	MDP0000301011	Glycolipid transfer protein 1
	MDP0000269126	Cellulase 1
Phytohormone biosynthesis and transport	MDP0000163720	Similar to a 2-oxoglutarate dioxygenase
	MDP0000284965	Root UB-B sensitive 2
Unknown function	MDP0000582151	Protein of unknown function
	MDP0000350168	No significant hit
	MDP0000513356	No significant hit
	MDP0000186457	Unknown protein

Conclusion

One hundred thousand BAC clones containing genomic DNA derived from a columnar apple cultivar with an average insert size of 27 kb of insert size were generated. 37 BACs were identified originating from the *Co* gene region. After sequencing, we were able to assemble these BAC sequences into two metacontigs of 590 and 190 kb in size leaving only one gap of roughly 35 kbp around position 18.7 Mb. We established four indel-based molecular markers linked to the *Co* gene with no recombination in a population of 95 individuals. DNA from other cultivars, columnar as well as non-columnar, was screened with these markers and showed the expected results. The only exception is 'Topaz', a non-columnar cultivar which showed the same banding profile as columnar cultivars. The reason for this is yet unclear. We identified 24 genes carrying heterozygous missense mutations in

the columnar cultivar while the non-columnar cultivar only exhibited one allele only. These genes are considered as potential candidate genes. Gene structure and the occurrence of SNPs in other cultivars are currently under investigation. The still existing gap of approximately 35 kb will be closed shortly.

Acknowledgements

We are grateful to Steffen Rapp and Benjamin Rieger (Department of Molecular Genetics) for their support in bioinformatic analysis. The work was supported by grant of the Federal Ministry of Agriculture and Nutrition (Nr. 511-06.01-28-1-43.042-07).

Literature Cited

Buhariwalla C, Greaves S, Magrath R, Mithen R (1995) Development of specific PCR primers for the amplification of polymorphic DNA from the obligate root pathogen Plasmodiophora brassicae. Physiological and Molecular Plant Pathology 47: 83-94

Burgoyne LA (1977) The standard preparation of polyamine-stabilized nuclei. School of Biological Sciences, Flinders University

Burgoyne LA, Wagar MA, Atkinson MR (1970) Calcium-dependent priming of DNA synthesis in isolated rat liver nuclei. Biochem Biophys Res Commun 39: 254-259

Eaton GW, Lapins KO (1970) Identification of standard and compact apple trees by discriminate function analysis. J Appl Ecol 7: 267-272

Eimert K, Reutter G, Strolka B (2003) Fast and reliable detection of doubled-haploids in Asparagus officinalis by stringent RAPD-PCR. The Journal of Agricultural Science 141: 73-78

Fisher DV (1969) Spur-type strains of McIntosh for high density plantings. British Columbia Fruit Growers' Association Quarterly Report 14: 3-10

Grunstein M, Hogness DS (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proceedings of the National Academy of Sciences 72: 3961-3965

Jacob HB (2010) Breeding experiments of apple varieties with columnar growth and low chilling requirements. Acta Hortic. 872

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research 40: D1202-D1210

Lapins KO (1969) Segregation of compact growth types in certain apple seedling progenies. Canad J Plant Sci 49: 765-768

Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C (2002) Development and characterisation of 140 new microsatellites in apple (*Malus x domestica* Borkh.). Molecular Breeding 10: 217-241

Looney NE, Lane WD (1984) Spur-type growth mutants of McIntosh apple: A review of their genetics, physiology and field performance. Acta Hortic. 146: 31-46

Lynch M, Walsh JB (1998) Genetics and Analysis of Quantitative Traits. Sinauer Assocs., Inc., Sunderland, MA.

Moriya S, Iwanami H, Kotoda N, Takahashi S, Yamamoto T, Kazuyuki A (2009) Development of a Marker-assisted Selection System for Columnar Growth Habit in Apple Breeding. J. Japan. Soc. Hort. Sci. 78: 279-287

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. PHYSIOLOGIA PLANTARUM 15: 473-497

Sheng Y, Mancino V, Birren B (1995) Transformation of Escherichia coli with large DNA molecules by electroporation. Nucleic Acids Res 23: 1990-1996

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci U S A 89: 8794-8797

Silfverberg-Dilworth E, Matasci CL, Van de Weg WE, Van Kaauwen MPW, Walser M, Kodde LP, Soglio V, Gianfranceschi L, Durel CE, Costa F, Yamamoto T, Koller B, Gessler C, Patocchi A (2006) Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome. Tree Genetics & Genomes 2: 202-224

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503-517

Tian YK, Wang CH, Zhang JS, James C, Dai HY (2005) Mapping Co, a gene controlling the columnar phenotype of apple, with molecular markers. Euphytica 145: 181–188

Tobutt KR (1985) Breeding columnar apples at east malling. Acta Hortic. 159: 63-68

Vali U, Brandstrom M, Johansson M, Ellegren H (2008) Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. BMC Genetics 9: 8

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagne D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouze P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2010) The genome of the domesticated apple (Malus x domestica Borkh.). Nat Genet 42: 833-839

Paper 3

The Columnar Mutation ("*Co* Gene") of Apple (*Malus* x *domestica*) Is Associated with an Integration of a Gypsy-like Retrotransposon

Dominik Otto[1*], Romina Petersen[1*], Bastienne Brauksiepe[2], Peter Braun[3], Erwin R. Schmidt[1]

[1] Department of Molecular Genetics, Johannes Gutenberg-University of Mainz
[2] Department of Botany, Geisenheim University
[3] Department of Pomology, Geisenheim University
[*] alphabetical order, authors contributed equally

This paper was published as a research article in Molecular Breeding in 2014 (*Mol Breeding* **33**: 863-880, with three figures and two supplementary files)

Abstract

The columnar growth habit of apple trees (*Malus* x *domestica* Borkh.) is a unique plant architecture phenotype that arose as a bud sport mutation of a McIntosh tree in the 1960s. The mutation ("*Co* gene") lead to trees (McIntosh Wijcik) with thick, upright main stems and short internodes that generate short fruit spurs instead of long lateral branches. Although *Co* has been localized to chromosome 10, in a region approximately between 18.5–19 megabases, its molecular nature is unknown. In a classical positional cloning approach in combination with the analysis of NGS data we cloned and analyzed the *Co* region. Our results show that the insertion of a Ty3/Gypsy retrotransposon into a non-coding region at position 18.8 Mb is the only detectable genomic difference between McIntosh and McIntosh Wijcik and is found in all columnar cultivars. The genetic effect of the insertion is unclear; however, Illumina® RNA-seq data of McIntosh and McIntosh Wijcik suggest that the columnar growth habit is associated with differential expression of the retrotransposon transcript, causing changes of the expression levels of many protein coding genes. The mechanism how the gypsy retrotransposon is involved generating the columnar is not yet clear; our findings are the basis to tackle this question.

Keywords

map-based cloning, apple columnar gene, Ty3/Gypsy retrotransposon, Illumina, RNA-seq, leaf transcriptome

Introduction

Apple (*Malus* x *domestica* Borkh.) trees have been a subject of domestication for at least 4000 years (Zohary and Hopf 2000). Today, apple is one of the most widespread fruits in the world. In 2011, more than 75.6 million tonnes of apples were produced worldwide (FAOSTAT, Food and Agriculture Organisation, faostat.fao.org). This corresponds to a market value of at least $7.5 -$9.5 billion. Therefore breeding of apple for the improvement of crop quality and other agronomically important traits is a major issue of plant breeders.

Plant architecture is an agronomically important trait. It is crucial for planting density in orchards as well as efficient harvesting, and influences pruning efforts. One goal of many breeding attempts was therefore to generate tree varieties that show a more compact growth habit. In 1961 a very compact limb sport mutation was detected by the apple grower Anthony Wijcik on a McIntosh apple tree, which turned out to be the result of a spontaneous somatic mutation (Fisher 1969, 1995). In honor of the discoverer, this variety has later been named McIntosh Wijcik. Due to the extreme compact growth type of the McIntosh Wijcik variety with thick, upright main stems and short internodes that generate short fruit spurs instead of long lateral branches, this growth habit has been named "columnar". The interpretation of the results of numerous crossings with the McIntosh Wijcik mutant yielding approximately 50 % of the offspring displaying the typical extreme compact growth phenotype of the McIntosh Wijcik mutation) (Lapins 1969; Lapins and Watkins 1973; Lapins and Fisher 1974; Zhang and Dai 2011) has led to the assumption that a dominant gene (or allele) called "*Columnar*" ("*Co* gene") (Lapins 1976) must be responsible for the compact growth habit of columnar trees (Lapins and Fisher 1974; Kesley and Brown 1992). The *Co* gene is acting alone in a heterozygous (possibly hemizygous) state or in combination with several modifiers leading to a variety of intermediate phenotypes (Hemmat et al. 1997; Kim et al. 2003). In almost all cases, in which the genotype has been analyzed in detail, for example by molecular markers closely linked to the *Co* gene (Baldi et al. 2013; Moriya et al. 2012; Bai et al. 2012), the columnar trees proved to be in a heterozygous condition. Why this is the case is not clear. Crosses between two columnar cultivars can yield up to 75 % columnar progeny (Lapins 1976; Meulenbroek et al. 1998), which implicates the existence of homozygous individuals. The heterozygous individuals might simply have been favored by breeders since they have a better chance of combining columnar growth with advantageous traits of non-columnar trees such as high fruit quality. Alternatively, a negative influence of *Co* or a linked gene on the fitness of pollen, seeds or early seedlings has been proposed (Meulenbroek et al. 1998). However, recently we were able to identify one specimen among our own collection of columnar trees that is homozygous for the chromosomal region carrying the *Co* locus, A73-19-97K (A73). So the *Co* gene is not a homozygous lethal mutation. Because the compact growth habit could be advantageous for apple growers, many breeding programs worldwide have been launched to transfer this beneficial trait into new varieties of apples (Tobutt 1994; Gelvonauskienė et al. 2006). However, especially the first columnar apple trees showed a number of disadvantageous traits like poor fruit quality, biennial bearing and susceptibility to scab (Meulenbroek et al. 1998; Moriya et al. 2009; Tobutt 1985; Gelvonauskienė et al. 2006; Lauri and Lespinnasse 1993; Petersen and Krost 2013). Combining the columnar growth habit with high fruit quality by conventional crossing has proven to be very difficult due to the self-incompatibility of *Malus* x *domestica*, its high degree of heterozygosity and a long juvenile

period (Hemmat et al. 1994). Therefore, the identification and characterization of the *Co* gene could be of great value for apple breeding.

A large number of scientific efforts have been made to localize the *Co* gene and to find molecular markers for fine mapping. These have led to mappings of the *Co* locus on linkage group 10 (Conner et al. 1997; Hemmat et al. 1997), now referred to as chromosome 10 according to the apple genome assembly of the Golden Delicious (GD) genome as described by Velasco et al. (Velasco et al. 2010). Due to increasing numbers of available markers (Liebhard et al. 2002; Moriya et al. 2009; Tian et al. 2005), more accurate mappings were possible (reviewed in Petersen and Krost (2013)). Most recently, Moriya et al. (2012) specified the most likely *Co* location as 18.76 – 18.96 Mb, Baldi et al. (2013) detected *Co* to be located between 18.52 and 18.92 Mb and Bai et al. (2012) proposed 18.90 and 19.09 Mb to be the borders for the possible *Co* locus. We consider the overlapping region of Baldi et al. (2013) and Moriya et al. (2012), i.e. 18.76 – 18.92 Mb, as the most likely *Co* region because the results of these two marker analyses are in good agreement with each other: the left border of Moriya et al. (2012) represents a further delimitation of the left border detected by Baldi et al. (2013) and the right borders are almost identical. In contrast, the marker of Bai et al. (2012) showing recombination at 19.09 Mb was located on a GD contig that was unanchored in the apple genome and whose position is therefore possibly not as clear as the position of the other markers, even though we are familiar with problems of the marker analyses having been carried out on different apple varieties etc. (discussed in Petersen and Krost (2013)). Furthermore, one marker that was found to be linked to *Co* by Bai et al. (2012) showed three recombinants when tested by Baldi et al. (2013). Therefore we think that the results of Baldi et al. (2013) and Moriya et al. (2012) are more reliable. Marker analyses have already led to the nomination of some genes as potential candidates for *Co*: Bai et al. (2012) and Baldi et al. (2013) proposed genes for Lateral Organ Boundary Domain proteins and different transcription factors, respectively, as possible growth habit regulators, due to their potential function according to BLAST searches. In spite of this progress, the nature of the so-called *Co* gene has remained obscure.

In order to identify and analyze the *Co* gene we applied a classical positional cloning approach using BAC libraries of the variety Procats 28 (P28) carrying the *Co* gene heterozygously and displaying the full columnar phenotype. The predominant reason to use Procats 28 (P28) instead of McIntosh Wijcik for the construction of BAC libraries and sequencing was mainly because all our genetical and mapping analysis was performed with the cultivars P28 and A14-190-93 (A14). In particular our newly developed molecular markers were tested and validated on a crossing between P28 and A14. In a second step we performed a deep sequencing of the genomes of McIntosh and McIntosh Wijcik. The sequences of the columnar and non-columnar BAC metacontigs were used as a reference for sequence mappings of genomic Illumina data from an original McIntosh tree and from a McIntosh Wijcik tree. With this approach, we were able to identify a new integration of a retrotransposon belonging to the Ty3/Gypsy-type of retrotransposons (for comprehensive database of Gypsy retrotransposons see gydb.org/index.php/Main_Page; (Llorens et al. 2011)), which we consider to be associated with the *Co* mutation. There is no obvious gene in the target site region that might have been inactivated by the Gypsy integration, so that there is no straight forward explanation of whether and how the columnar phenotype might be generated by the retrotransposon integration. However, thanks to our extensive analysis of all annotated genes in our region concerned, we know that there are no other mutations

between McIntosh and McIntosh Wijcik within coding regions that could more easily be made responsible for the columnar phenotype. Using next generation sequencing we have also analyzed the transcriptomes of columnar versus non-columnar trees (apical meristem and leaves) with the goal to identify the difference in gene expression between the two varieties (Krost et al. 2012; Krost et al. 2013).

Materials and Methods

Plant materials

Since McIntosh (-/-) and its columnar bud sport McIntosh Wijcik (*Co*/-) should have identical genomes except from the mutation that led to the columnar growth habit, leaves of McIntosh and McIntosh Wijcik were collected at the Geisenheim University and used for Illumina® whole genome sequencing, genomic analysis of candidate genes and transcriptomic studies. Furthermore, we chose the heterozygous columnar variety P28 (*Co*/-) for the construction of genomic BAC libraries and intron-exon structure analysis because we performed all our genetical mappings and the development of new molecular markers on the offspring of a crossing of P28 x A14-190-93 (A14) and because in vitro cultures of A14 and P28 were available in high amounts all year round. In vitro cultures of P28 (*Co*/-) and A14 (-/-) grown on Murashige and Skoog medium (Murashige and Skoog 1962) were provided by the Departments of Pomology and Botany, Geisenheim University. P28 originates from a crossing of the columnar cultivar Flamenco (*Co*/-) and the non-columnar cultivar Topaz (-/-), with Flamenco being a direct descendant of McIntosh Wijcik (Supplementary table 1). A14 is a non-columnar descendant of the columnar cultivar Telamon (*Co*/-), whose mother is McIntosh Wijcik, and the non-columnar cultivar GD (-/-) (Supplementary table 1). The progeny of the P28 x A14 crossing comprises 95 progeny plants (48 columnar (*Co*/-), 47 non-columnar (-/-)) and was originally created for breeding purposes, attempting to combine the high fruit quality of A14 with the desirable growth habit of P28.

In order to assess whether the molecular markers can reliably detect the correct genotype with regard to the presence of *Co* independent of the apple cultivar tested, leaves of a collection of columnar and non-columnar varieties available from the fields of the Geisenheim University were used for molecular marker analysis. The collection includes the heterozygous columnar cultivars Procats 1, Procats 4, Procats 11, Procats 13, Greencats, Pomforyou, Pomfital, Kordonia, A10-28, McIntosh Wijcik, Pomgold and HL 4 K, the non-columnar varieties McIntosh, Elswout, Gala, Jonagold, Pinova and Topaz and the homozygous columnar cultivar A73. Details about their pedigrees can be found in Supplementary table 1.

BAC library construction and screening

For construction of genomic BAC libraries, isolation of high molecular weight (HMW) DNA is required. In vitro cultures of P28 were homogenized by a hand-held blender. Isolation of nuclei was performed according to the protocol of Burgoyne et al. (1970) with some modifications. Nuclei were incubated in lysis buffer containing 2 % SDS and 1 mg/ml proteinase K for 10 min at 60 °C, 60 min at 37 °C and then dialyzed overnight against water.

Two phenol chloroform isoamyl alcohol and two chloroform isoamyl alcohol extractions were performed to purify the DNA.

Partial digestion of the HMW DNA was carried out using 0.5 U Sau3A1 (Roche, www.roche.de) at 37 °C for 1.5 min immediately followed by a phenol chloroform isoamyl alcohol extraction for enzyme inactivation. Size fractionation was performed twice using 0.6 % agarose gels ("Horizontal System for Submerged Gel Electrophoresis", Life Technologies™, www.lifetechnologies.com). HMW DNA fragments were ligated into BamH1 digested pBeloBAC11 (Shizuya et al. 1992) for 72 h at RT and transformed into *Escherichia coli* strain DH10B by electroporation (Sheng et al. 1995).

Supplementary Table 1. Phenotypes and pedigrees of apple cultivars used for marker analyses. The upper part lists the phenotype and parents of all apple varieties used in this study (where known) and the lower part lists the phenotype and parentage of the parent cultivars with McIntosh Wijcik origin.

	Phenotype	**Mother**	**Father**
Procats 1	columnar	Charlotte	Topaz
Procats 4	columnar	Waltz	Topaz
Procats 11	columnar	Bolero	
Procats 13	columnar	Waltz	AK18-182-93
Procats 27	columnar	Polka	Topaz
Procats 28	columnar	Flamenco	Topaz
Greencats	columnar	Bolero	Golden Delicious
Kordonia	columnar		
Pomforyou	columnar	Maypole	R. Elstar
Pomgold	columnar	Waltz	Calagolden
Pomfital	columnar	Maypole	R. Elstar
A10-28	columnar		
HL-4_K	columnar		
A73-19-97K	columnar	Polka	open pollination
A14-190-93	non-columnar	Golden Delicious T. Weinsberg	Waltz
Jonagold	non-columnar	Golden Delicious	Jonathan
Gala	non-columnar	Kidds Orange	Golden Delicious
Elswout	non-columnar	Golden Delicious	Ingrid Marie
Pinova	non-columnar	Clivia	Golden Delicious
Waltz (Telamon)	columnar	McIntosh Wijcik	Golden Delicious
Flamenco (Obelisk)	columnar	(Cox's Orange Pippin x Court Pendu)	McIntosh Wijcik
Polka (Trajan)	columnar	Golden Delicious	McIntosh Wijcik
Bolero (Tuscan)	columnar	McIntosh Wijcik	Greensleeves
Charlotte (Hercules)	columnar	McIntosh Wijcik	Greensleeves
AK18-182-93 (syn. Geisenheimer Goldrenette)	columnar	R. Alkmene	Waltz

For library screening, PCR-based probes were designed based on the published genome sequence (Velasco et al. 2010) and tested for repetitive sequences using Southern and Pirotta hybridization (Buhariwalla et al. 1995; Southern 1975). The colony hybridization

was performed according to Grunstein and Hogness (1975). Positively selected colonies, i.e. colonies harboring sequences putatively originating from chromosome 10, 18.76 – 18.92 Mb, were isolated using PureYield™ Plasmid Maxiprep System (Promega, www.promega.com).

BAC Sequencing and Assembly

BACs were sequenced using a combination of Illumina® next generation sequencing (see below) and Sanger sequencing. Sanger sequencing was performed by StarSEQ (www.starseq.com) with the corresponding primers. Assemblies of BAC sequences were performed using the de novo assembly tool of CLC Genomics Workbench (CLC bio, www.clcbio.com) and NGEN2® (DNASTAR®, www.dnastar.com) with default parameters. Gaps, repeats and critical positions were verified by PCR and Sanger sequencing. To create contigs, the BAC sequences were assembled using SeqMan™ and MegAlign™ (DNASTAR®, www.dnastar.com). Gaps between contigs were closed by long range PCRs.

Generation of PCR-based Molecular Markers

The comparison between the BAC sequences of the columnar cultivar P28 and the published GD genome (Velasco et al. 2010) enabled us to identify insertions and deletions (indels (Vali et al. 2008)). Primers flanking these indels were designed and purchased from Invitrogen™ (www.invitrogen.com). DNAs of leaves of an F1 population comprising 95 progeny plants of a cross between P28 and A14 was provided by the Geisenheim University, extracted according to Eimert et al. (2003). The DNAs of the F1 progeny were used as templates for marker PCRs. All PCR reactions were performed as described below. The amplified fragments were cloned into the pGEM®-T Easy Vector System (Promega, www.promega.com) and sequenced to validate the correct genomic localization.

Two indel-based novel markers, I2_3_M1 and 1C3_M2 (Fig. 1), showed 100 % linkage to the *Co* gene (Otto et al. 2013). BAC sequences were assigned to the columnar or to the non-columnar chromosome using either markers I2_3_M1 and 1C3_M2 or alignments of large identical sequence overlaps (> 5,000 bp) with already assigned BAC clones.

DNAs of leaves of other apple varieties were extracted according to Eimert et al. (2003) in order to test markers for species-specificity.

PCRs

For genomic PCRs, about 10 ng of DNA was used as a template. All primers were purchased from Invitrogen™ (www.invitrogen.com). PCRs were carried out in 50 µL reaction volume as follows: initial denaturation at 95 °C for 5 min, followed by 40 amplification cycles at 95 °C for 30 sec, the appropriate annealing temperature (58 – 60 °C) for 30 sec and elongation at 72 °C for 1 min. A final elongation step at 72 °C for 10 min was included. Touchdown PCRs of 64 – 54 °C were performed where PCR products could not be obtained with the normal PCR protocol. Long range PCRs were performed where necessary using TaKaRa LA Taq® DNA polymerase (TaKaRa/ Clontech, www.clontech.com) according to the manufacturer's instructions. PCR products were analyzed on agarose gels and were digested using 5 U Exonuclease I and 1 U Fast Alkaline Phosphatase (both Fermentas, www.thermoscientificbio.com/fermentas) at 37 °C for 25 min followed by enzyme inactivation at 75 °C for 15 min. In case unspecific bands occurred on the gel, the band of the right size was excised from the agarose gel and purified using the NucleoSpin® Gel and PCR

Clean-up kit (Macherey-Nagel, www.mn-net.com). Purified PCR products were sequenced by Sanger at StarSEQ (www.starseq.com) and sequences were analyzed with the (DNASTAR® Lasergene package (www.dnastar.com)).

Analysis of candidate genes

For genomic analysis, PCR primers were designed to obtain PCR products covering all annotated exons as well as the regions 500 bp upstream and downstream of genes. DNA was isolated from adult leaves of McIntosh and McIntosh Wijcik using the innuPREP Plant DNA kit (Analytik Jena, www.analytik-jena.de) and was used as PCR template. All PCR reactions were performed as described above.

For intron exon structure analysis, in vitro cultures of P28 and A14 were harvested just prior to RNA isolation. RNA was isolated using the innuPREP Plant RNA kit (Analytik Jena, www.analytik-jena.de) following the manufacturer's instructions. RNA concentration and integrity was analyzed on a Bioanalyzer RNA Nano chip (Agilent, www.agilent.com). Subsequently, cDNA was reverse transcribed from 1 µg of RNA using the SuperScriptIII Reverse Transcriptase (Invitrogen™, www.invitrogen.com) and 50 µM oligo$(dT)_{20}$ Primer (Fermentas, www.thermoscientificbio.com/fermentas) as described in the protocol provided by the manufacturer. 1 µl of a 1:10 dilution of the cDNA was used as PCR template. PCRs were conducted as described above.

Illumina Next Generation Sequencing

For Illumina® whole genome sequencing, adult leaves of McIntosh and McIntosh Wijcik were used for DNA isolation with the innuPREP Plant DNA kit (Analytik Jena, www.analytik-jena.de) following the manufacturer's instructions.

For RNA isolation with the purpose of Illumina sequencing, young but fully developed leaves of about 5 cm size were collected on May 6th 2012 (around 11 am) from juvenile trees of McIntosh Wijcik and on May 23rd 2012 (around 2.30 pm) from juvenile trees of McIntosh. They were frozen on dry ice and stored at -80 °C until usage. About one quarter of one leaf each of McIntosh and McIntosh Wijcik was ground under liquid nitrogen and RNA was isolated using the innuPREP Plant RNA kit (Analytik Jena, www.analytik-jena.de) following the manufacturer's instructions. RNA concentration and integrity was analyzed on a Bioanalyzer RNA Nano chip (Agilent, www.agilent.com).

Library preparation for Illumina® DNA and RNA sequencing was performed from 5 µg of sample by GENterprise Genomics (www.genterprise.de). Subsequently, a 100 bp paired end run was conducted on the Illumina® HiSeq2000 (www.illumina.com) of the IMSB Mainz.

Trimming of Illumina Data

Illumina® sequencing data were imported into the CLC genomics workbench (CLC bio, www.clcbio.com) as paired end reads and were quality trimmed using an error probability limit of 0.05 and allowing a maximum of 3 ambiguous bases. Subsequently, 5 nucleotides at the 5' end and 10 nucleotides at the 3' end were removed. After this step, all sequences with a length of < 15 bp were discarded.

Mappings and Mapping Analyses of Genomic Data

All mappings were conducted in the CLC genomics workbench (CLC bio, www.clcbio.com). For genomic data, "Map Reads to Reference" was applied. For a normal

mapping, a similarity of at least 98 % was demanded for at least 90 % of the read length. Parameters “mismatch cost”, “insertion cost” and “deletion cost” were set to 3, and parameter “non-specific match handling” was set to “map randomly”. For stringent mappings, parameters “similarity” and “length fraction” were both set to 1.0 and all other parameters were chosen as for a normal mapping.

SNP and DIP detections (now called probabilistic variant detections) were performed demanding a minimum variant frequency of 0.25, a minimum coverage of 10 and a maximum of expected variants of 2. Parameters “ignore non-specific matches” and “require presence in both forward and reverse reads” were both set to “no” Structural variation detection was conducted with the default p value threshold of 1.0E-04 and including the calculation of interchromosomal variations.

Analysis of Transcriptomic Data

For mappings of RNA-seq reads against target region sequences, either “Map Reads to Reference” was applied with the same parameters as for genomic data, or “Large Gap Read Mappings” were conducted using the same parameters as for a stringent “Map Reads to Reference” and the default parameters of “maximum no. of hits for a read” set to 10 and “maximum distance from ‘seed’” set to 50,000.

In order to perform a whole transcriptome analysis, RNA-seq mappings of transcriptomic Illumina reads against the GD reference genome were conducted with similarity 98 % and length fraction 90 % and the same parameters as for “Map Reads to Reference”. Additionally, “maximum no. of hits for a read” was set to 10, annotated gene regions were extended by 50 bp of upstream and downstream flanking residues, and the “include broken pairs” counting scheme was applied, i.e. single reads of broken pairs were counted as 1 and a pair of paired reads was counted as 2. Total gene read counts were extracted and differential expression was analyzed as described in Krost et al. 2012. Briefly, MDPs were assigned to SwissProt hits and SwissProt hits were ranked according to their differential expression using R_j values (Stekel et al. 2000). Subsequently, the LOG2 fold change of read counts was determined, normalized to expression in McIntosh. Visualization was achieved using MapMan (Thimm et al. 2004) and mapping against the *Populus trichocarpa* genome since *Populus* is the closest relative to apple for which a MapMan mapping file exists.

Database searches

Basic Local Alignment Search Tool (BLAST) searches (Altschul et al. 1990) against the apple genome (Velasco et al. 2010) were carried out on the homepage of the Institute Agrario di San Michele all’Adige (genomics.research.iasma.it) or at the homepage of the Genome Database of Rosaceae (www.rosaceae.org). BLAST searches against the Arabidopsis genome were conducted using The Arabidopsis Information Resource (TAIR) (Lamesch et al. 2012). Sequences of interest were analyzed online with CENSOR (Kohany et al. 2006), which performs BLAST searches against transposon databases, and GENSCAN (Burge and Karlin 1997), which can detect open reading frames (ORFs).

Accession numbers

Sequence data from this article can be found in the EMBL/GenBank data libraries under accession numbers ERP002574 (genomic Illumina data), ERP002576 (transcriptomic

Illumina data), HF968765 (metacontig of the *Co* gene region, columnar allele) and HF968766 (metacontig of the *Co* gene region, non-columnar allele).

Results

The difference between the columnar and the non-columnar apple tree is an 8.2 kb insertion at 18.8 Mb on linkage group 10

In order to identify the somatic mutation that is responsible for the columnar growth phenotype we resequenced the region putatively containing the *Co* gene. We chose P28 for sequencing because we performed all our genetical mappings and the development of new molecular markers on the offspring of a crossing of P28 x A14. The BAC libraries contained approximately 150,000 clones with an average insert size of 26 kb, representing about 5.4 x haploid genome equivalents. Out of 150,000 clones, seven clones were identified by colony filter hybridization and could be anchored between positions 18.75 and 18.92 Mb on LG 10 of the GD genome sequence based on their sequence. Based on the novel molecular markers I2_3_M1 and 1C3_M2, BAC sequences were assembled into three contigs (Fig. 1). Two contigs of 125 kb (II) and 39 kb (V) in size represent the sequence of the columnar chromosome, whereas the third contig (76 kb, III) corresponds to the sequence of the non-columnar chromosome. The remaining gap (15 kb) between the two contigs of the columnar chromosome was closed by long range PCRs (IV). Furthermore long range PCR products (I) of altogether 21 kb extend the available sequence from 18.75 to 18.73 Mb. Thus, the metacontig containing the columnar sequence where available stretches from 18.73 to 18.92 Mb (I, II, IV and V).

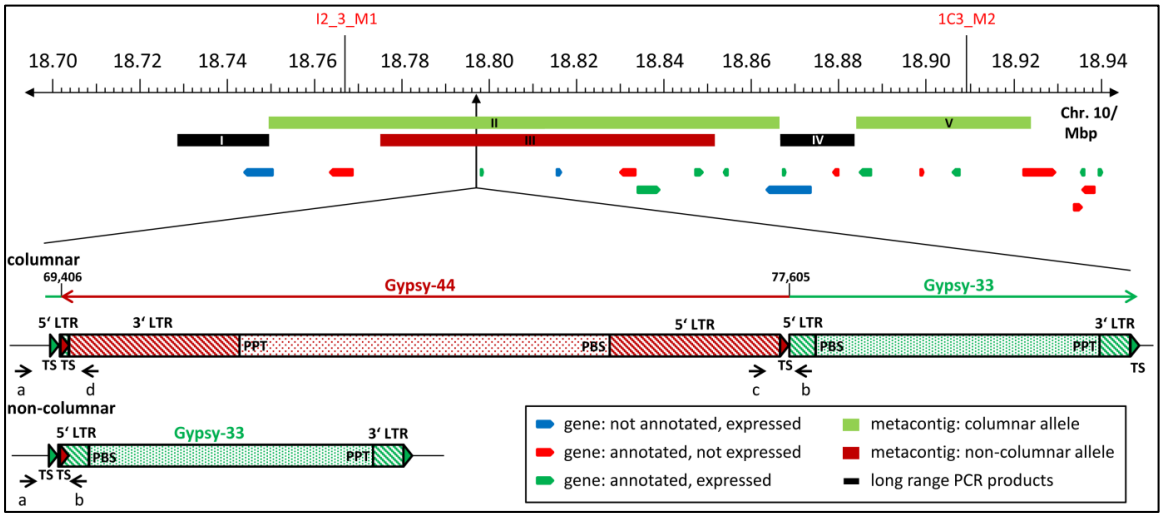


Fig. 1 Overview of the *Co* target region on apple chromosome 10
Upper Part The potential *Co* region of 18.73 – 18.92 Mb was cloned from the heterozygous columnar cultivar Procats 28 by constructing a genomic BAC library and screening with molecular probes from the *Co* region and eventually resequenced. BAC clones were assembled into contigs (II, III and V) and assigned to the columnar (II, V) or the non-columnar chromosome (II) using markers I2_3_M1 and 1C3_M2. Long range PCR products (I, IV) are indicated by black bars. Genes that had previously been annotated in the GD genome (Velasco et al. 2010) and newly detected genes are shown at their respective positions. **Lower Part** Structure and Location of Gypsy-44. On the columnar chromosome

(columnar), Gypsy-44 is found to be inserted into the 5'-LTR of Gypsy-33 in the 3' to 5' direction, position 69,406 – 77,605. Both transposons show long terminal repeats (LTRs), primer binding sites (PBS) and poly purine tracts (PPT) and are flanked by target site (TS) duplications. In the non-columnar chromosome (non-columnar), only Gypsy-33 is present. a – d indicate locations of primers used for PCR analysis.

For the identification of the putative *Co* mutation we resequenced the corresponding chromosomal region from McIntosh and McIntosh Wijcik apple trees. The genomes of these two varieties should be identical except for the somatic mutation that causes the columnar growth phenotype. An Illumina® HiSeq2000 100 bp paired end run of genomic DNA extracted from leaves of McIntosh and McIntosh Wijcik yielded 422,811,486 and 444,534,268 reads for McIntosh and McIntosh Wijcik, respectively. After quality and end trimming, about 414 million (McIntosh) and about 434 million (McIntosh Wijcik) reads with an average length of 81 bp remained, corresponding to a genome coverage of 45 x for McIntosh (414 million reads) and 46 x for McIntosh Wijcik (434 million reads). Approximately 2.6 million reads (McIntosh) and 2.8 million reads (McIntosh Wijcik) matched with the putative *Co* gene region, providing an average coverage of 891 x for McIntosh and 952 x for McIntosh Wijcik. In order to identify the expected genomic difference(s) within this region (corresponding to the *Co* mutation), structural variation detection as well as probabilistic variant detection was conducted on both mappings. At position 77,605 of the BAC metacontig (corresponding to chromosome 10, position 18,796,887 Mb on the GD genome sequence), a structural variation was detected only in the McIntosh sequence and not in McIntosh Wijcik with a very high probability (p value of $4.5x10^{-12}$), and was classified as "deletion" by CLC if compared to the metacontig sequence of the columnar chromosome.

The sequence reads obtained from McIntosh Wijcik yield perfect coverage of the metacontig representing the chromosome carrying the *Co* mutation, without any gaps or discontinuities. The reads obtained from the homozygous non-columnar McIntosh also match along the entire columnar metacontig sequence. However, there is not a single read spanning position 77,605 from 5' to 3' or vice versa. In addition, there is a complete lack of paired reads (i.e. a forward and a reverse read originating from the two ends of the same DNA molecule) with one end being located upstream and the other end downstream of position 77,605 (Supplementary Fig. 1). The same result is found for position 69,406 of the columnar metacontig. This shows that the 8.2 kb sequence between positions 69,406 and 77,605 of the columnar metacontig is not present in McIntosh at this chromosomal position. It is in fact not a deletion in McIntosh but an insertion in McIntosh Wijcik.

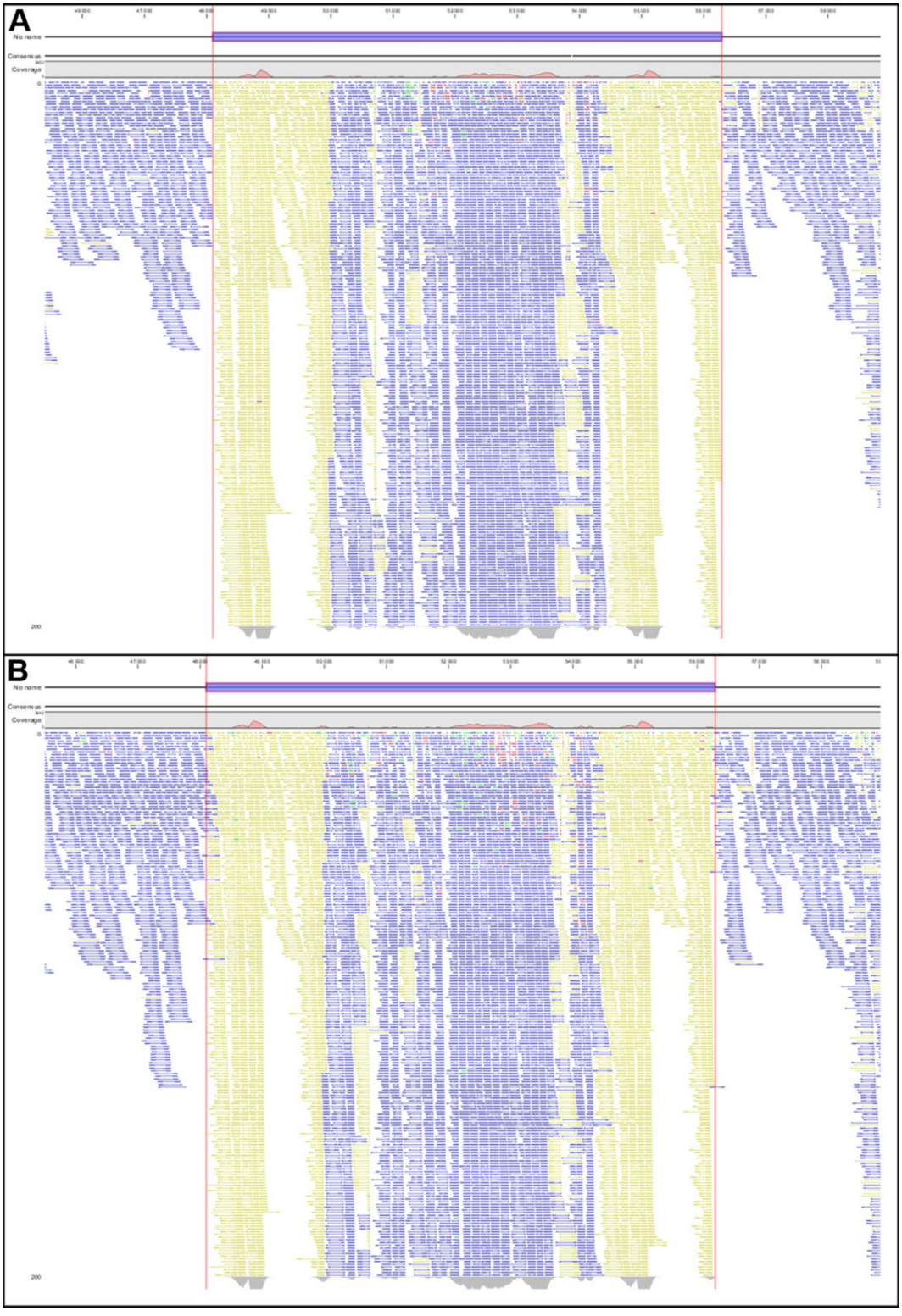
To test this hypothesis we designed primers spanning the left or the right border of the putative 8.2 kb insertion as well as primers located up- and downstream of the entire 8.2 kb insertion and then carried out PCRs using genomic DNAs as templates that were obtained from seven different non-columnar and 14 columnar apple varieties. For all specimens used, the phenotype (columnar vs. non-columnar) was clearly known, and we also determined the genotype (*Co*/*Co*; *Co*/-; -/-) using the 100 % linked molecular markers. The results were very clear: for primer pairs spanning the left or right border of the putative insertion, the expected fragments could be amplified only from template DNA of phenotypically columnar varieties, whereas DNAs from non-columnar varieties did not yield a PCR product of the expected size (Fig. 2 and Supplementary Fig. 2). However, regularly some additional PCR products were generated, but sequencing showed that these fragments originated from other chromosomal locations. The primer pair flanking the entire 8.2 kb insertion produced

a small fragment of the expected size amplified from the non-columnar chromosome (which does not contain the insertion) only in non-columnar and heterozygous columnar varieties, but not in the homozygous columnar cultivar (Fig. 2). The fragment of about 9 kb that is expected as the amplification product from the chromosome containing the 8.2 kb insertion could not be obtained, which might be explained by the very high AT-content (64 %) that might interfere with an efficient *in vitro* amplification.

The 8.2 kb insertion is a Ty3/Gypsy-like retrotransposon

In order to analyze the nature and origin of the insertion, its sequence was subjected to further mappings and database analyses. In a BLAST search against the CENSOR database (Kohany et al. 2006), the fragment was identified as a Long Terminal Repeat (LTR) retrotransposon containing an internal portion of Gypsy-44_Mad-I (Gy-44), a Ty3/Gypsy retrotransposon already known to be present in apple genomes. Relative to the annotated apple genome sequence its orientation is 3' to 5'. Gy-44 carries large LTRs of 1,951 bp, beginning with a TG and ending with a CA (Fig. 3). The LTRs show 100 % sequence identity to each other, which indicates a rather recent insertion event. There is a 5 bp long target site duplication (5'-AGG AC-3'). A typical primer binding site (PBS) 5'-TGG CGC CGT TGC CGG-3' is found 5 bp downstream of the end of the 5'-LTR, and a poly purine tract (PPT) 5'-CAA AAG TTA AGT GTG GGG GTA TT-3' is present in front of the start of the 3'-LTR. The internal portion does not show the typical open reading frames (ORFs) encoding the group specific antigen (Gag) and polymerase (Pol) proteins of retrotransposons. Instead, several short ORFs can be detected. The longest ORF has a length of 609 bp and is located 3' to 5' relative to the Gy-44 orientation at 600 bp distance from its 5'-LTR. Via strand-specific reverse transcription we were able to show that both strands are transcribed in the region of this ORF. The inferred amino acid sequences of all detected ORFs do not give any significant match when searched in various sequence data libraries. From the significantly increased coverage of the transposon by sequence reads (retrotransposon vs. genome = 2,871 x vs. 45 x in McIntosh and 2,982 x vs. 46 x in McIntosh Wijcik), a copy number of about 60 can be estimated for Gy-44.

Analyzing the immediate flanking regions of the retrotransposon with CENSOR (Kohany et al. 2006), it turns out that the McIntosh Wijcik-specific Gy-44 has integrated into the 5'-LTR of another Ty3/Gypsy retrotransposon, Gypsy-33_Mad-I (Gy-33) (Fig. 1). The only annotated gene within this region, MDP0000766466, seems to correspond to the *gag* gene of Gy-33 because it contains a typical Cys/His finger motif. No ORF for *pol* can be found in Gy-33. At a distance of 179 bp upstream of Gy-33 another transposon can be found, which was identified as the DNA transposon DNA9-5_Mad by CENSOR. In conclusion, the Ty3/Gypsy retrotransposon Gy-44 has recently "jumped" into an extremely transposon-rich region.



Supplementary Fig. 1 Screenshots of mappings of genomic paired end reads against the 8.2 kb insertion

Mappings were conducted in CLC Genomics Workbench (CLC bio) against the BAC metacontig sequence anchored around position 18.8 Mb on chromosome 10. (A) Mapping of genomic paired end reads of McIntosh. (B) Mapping of genomic paired end reads of McIntosh Wijcik. The left and right borders of the insertion are highlighted in red; the insertion sequence is marked in purple. Paired reads with matching partners are blue, forward reads are green, reverse reads are red and reads matching to more than one position are shown in yellow.

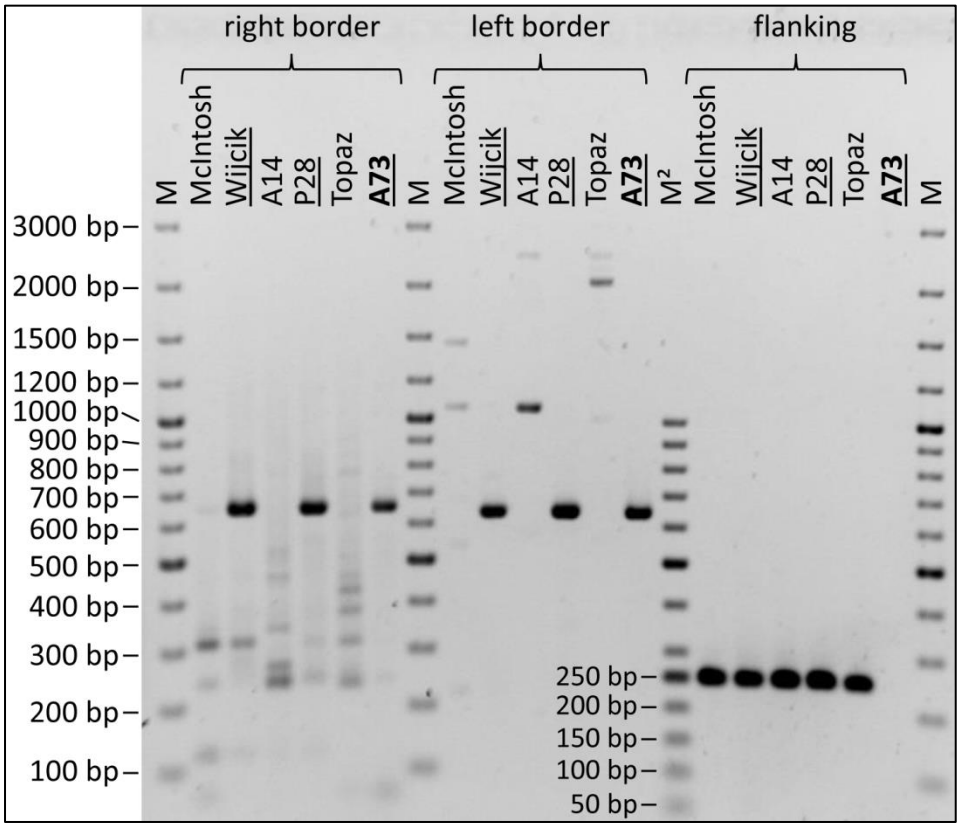
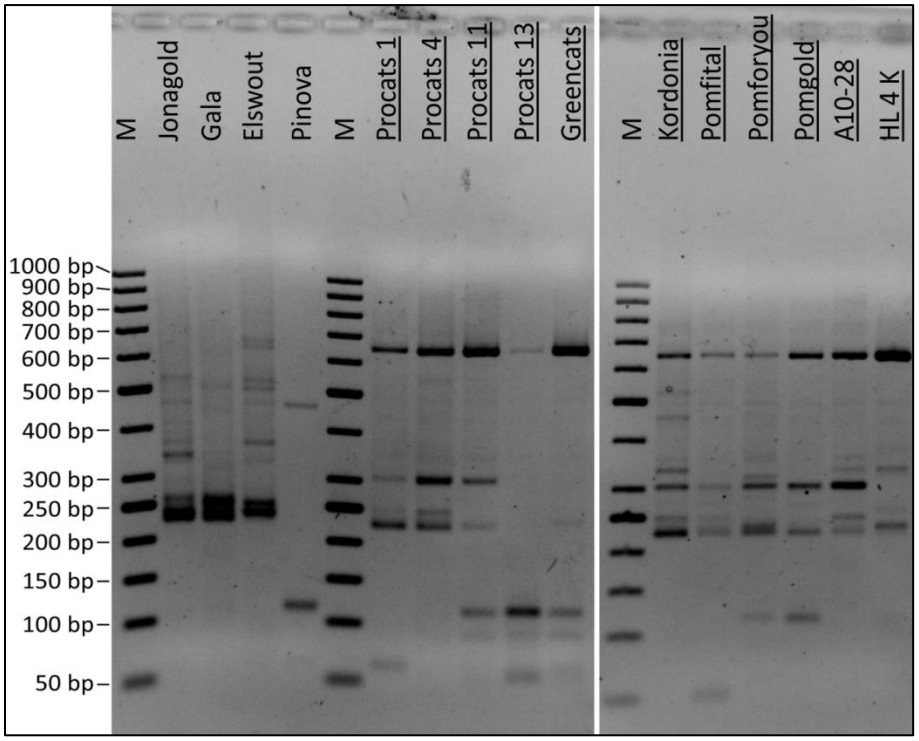


Fig. 2 Verification of the presence of the transposon
The presence or absence of the retrotransposon was verified by PCRs with pairs of primers spanning the left or right border of the retrotransposon or flanking the entire insertion using genomic DNA of the non-columnar cultivars McIntosh, A14-190-93K and Topaz, the heterozygous columnar cultivars McIntosh Wijcik and P28 (underlined) as well as the homozygous columnar cultivar A73 (underlined, bold). The expected fragments of 628 bp (primers spanning left border) and 633 bp (primers spanning right border) were obtained only from columnar cultivars. In case of primers spanning the whole insertion, the expected 242 bp product amplified from the non-columnar chromosome was obtained from all cultivars except the homozygous columnar cultivar A73. The fragment of about 9 kb that is expected as the amplification product from the columnar chromosome containing the 8.2 kb insertion could not be obtained, which might be explained by the very high AT-content (64 %) that might interfere with an efficient *in vitro* amplification. M: GeneRuler™ 100 bp Plus DNA Ladder (Fermentas), M²: GeneRuler™ 50 bp DNA Ladder (Fermentas).



Supplementary Fig. 2 Verification of the transposon presence on different columnar and non-columnar varieties
In addition to the results shown in Fig. 2, PCRs with the primer pair spanning the right border of Gy-44 were conducted on template DNA of an additional four non-columnar and 11 columnar (underlined) cultivars. The expected fragment of 633 bp was only obtained from columnar cultivars. M: GeneRuler™ 50 bp DNA Ladder (Fermentas)

The insertion of Gypsy-44 is the only genomic difference between McIntosh and McIntosh Wijcik within the potential *Co* gene region

The mappings of the genomic Illumina sequence data of McIntosh and McIntosh Wijcik against the BAC metacontig were subjected to further single nucleotide polymorphism (SNP), deletion insertion polymorphism (DIP) and structural variation analyses. In all mappings, no SNPs, DIPs or structural variations of statistical significance were detected between McIntosh and McIntosh Wijcik. Furthermore, pairs of primers were designed to span the coding region as well as 500 bp upstream and downstream of the coding regions of all annotated genes located in the BAC metacontig region. Additionally, all genes located within the *Co* target region of Bai et al. (2012) and candidate genes proposed by Baldi et al. (2013) that do not reside in the metacontig region were analyzed. Genomic DNA from McIntosh and McIntosh Wijcik was used as a template for PCR amplification. PCR products were sequenced by Sanger sequencing, and sequences of McIntosh and McIntosh Wijcik were compared. Based on these genomic PCRs, all annotated genes were shown to be 100 % identical between McIntosh and McIntosh Wijcik. In contrast, numerous nucleotide polymorphisms are detected when the sequences of McIntosh or McIntosh Wijcik are aligned to the annotated GD sequence.

Since the annotation of the GD apple genome was carried out solely automatically (Velasco et al. 2010), we mapped two RNA-seq data sets of A14 and P28 obtained from

shoot apical meristems in previous studies (Krost et al. 2012; Krost et al. 2013) to the BAC metacontig as well as to the corresponding annotated GD contigs with the aim to detect any genes or exons that might have been missed by the bioinformatical annotation. Results were verified by cDNA synthesis and subsequent PCRs and Sanger sequencing. We identified several differences with regard to the structure of annotated genes and additionally found three new genes that had not been annotated in the GD genome data base entry (Fig. 1). The newly detected genes and exons were included in our comparative genomic analysis of McIntosh and McIntosh Wijcik. As for the previously annotated genes, again no differences were found in McIntosh Wijcik compared with McIntosh in the newly detected coding regions. We also analyzed the RNA-seq data sets of A14 and P28 for splice variants. There were no indications for alternative splicing in P28 relative to A14 in all 19 analyzed genes located within the *Co* gene region. Detailed gene annotations can be found under NCBI accession numbers HF968765 and HF968766.

Gypsy-44 is transcriptionally active and shows differential regulation in McIntosh Wijcik

In order to determine the expression levels of Gy-44 and Gy-33, we conducted an Illumina RNA-seq analysis of McIntosh and McIntosh Wijcik. As our previous studies had already dealt with RNA-seq data of shoot apical meristems (Krost et al. 2012; Krost et al. 2013), in this study we used RNA extracted from leaves. We obtained 177,767,620 and 169,702,766 reads for McIntosh and McIntosh Wijcik, respectively. Reads were quality filtered and end trimmed in the same way as the genomic reads, so that 177,462,922 (McIntosh) and 169,449,200 (McIntosh Wijcik) reads with an average length of 85 bp remained. Reads were mapped to the columnar BAC metacontig as reference, demanding 100 % sequence identity in order to exclude mapping and counting of any reads originating from transcripts of similar but not identical copies of Gy-44, which are probably present elsewhere in the McIntosh genome. 4,447 and 5,995 reads were mapped to Gy-44 in McIntosh and McIntosh Wijcik, respectively, corresponding to 1.4 x higher read counts in McIntosh Wijcik (normalized to the number of total reads). This means that the one additional columnar-specific Gy-44 copy in McIntosh Wijcik results in 40 % higher amount of transcripts of all Gy-44 copies present in McIntosh Wijcik. In contrast, Gy-33 was only covered by 2,443 (McIntosh) and 2,617 (McIntosh Wijcik) reads, which were confined to a few defined regions of the retrotransposon and thus represent fragments of other transcripts.

McIntosh Wijcik shows gross changes in gene expression in leaves

With the aim of better understanding the potential influence of the columnar Gy-44 insertion we comparatively analyzed the pattern of gene expression in leaves of McIntosh Wijcik and McIntosh. For this purpose, RNA-seq mappings of the transcriptomic Illumina data against the GD reference genome were carried out and quantitatively analyzed. In total, 115.556.593 reads (McIntosh; 65 % of total reads) and 108.030.064 reads (McIntosh Wijcik; 64 % of total reads) were mapped against the annotated GD genome (Velasco et al. 2010). Read counts for MDPs were extracted and MDPs were “collapsed” to UniprotKB unigenes in order to avoid redundancy caused by the fact that several different MDPs can encode proteins with the same function (for details see Krost et al. (2012)). Thus, read counts were assigned to unigenes and unigenes were ranked according to their level of differential

expression (Supplementary table 2). In total, 9,961 unigenes were found to be expressed in leaves of McIntosh and/or McIntosh Wijcik. Of these, 5,751 genes (58 %) were found to be significantly differentially regulated. Up- or downregulation in McIntosh Wijcik compared with McIntosh was visualized in MapMan (Fig. 3). Genes involved in photosynthesis, protein biosynthesis and nucleotide metabolism are downregulated in McIntosh Wijcik compared with McIntosh, whereas genes associated with secondary metabolism, especially lignin and terpene biosynthesis, are strongly upregulated (Fig. 3a). With regard to phytohormones, genes involved in auxin, jasmonate and ethylene metabolism and/or signaling are highly upregulated. In contrast, genes encoding for enzymes managing the redox state of the plant like thioredoxin, dismutase/catalase and peroxidase are downregulated in McIntosh Wijcik; only genes for glutathione-S-transferases are up. The most striking differential regulation is observed within the category "pathogen/pest attack" (Fig 3b): almost all genes encoding proteins involved in defense or stress reactions such as pathogen recognition receptors (PR) and heat shock proteins are highly upregulated in McIntosh Wijcik leaves compared with McIntosh. Focusing on genes located in or near the *Co* target region, two genes, MDP0000934866 and MDP0000878773, which can be found upstream of the Gy-44 insertion at positions 18.88 Mb and 18.95 Mb, respectively, are noticeably upregulated in McIntosh Wijcik.

In order to further investigate the potential effect of the insertion of the Gy-44 on the transcription of nearby genes, we carried out the same RNA-seq data analyses with the data sets of A14 (non-columnar) and P28 (columnar) representing the transcriptomes of shoot apical meristems (Krost et al. 2012; Krost et al. 2013). One of the annotated genes, MDP0000934866, located 90 kb upstream of the Gy-44 insertion site, showed approximately 2-fold upregulation in P28 compared with A14. Furthermore, one of the newly detected genes, which is located upstream (at 18.81 Mb) in close vicinity to the Gy-44 insertion, was found to be expressed exclusively in the columnar P28 and not in non-columnar A14.

In conclusion, we found an integration of the gypsy retrotransposon Gy-44 at position 18,796,887 Mb on linkage group 10 (apple genome annotation) that is specific to columnar apple tree varieties and thus might be linked to the establishment of the columnar phenotype. It interrupts the LTR of a preexisting retrotransposon of the same class of retrotransposons (Gy-33). Gy-44 is highly transcribed, but possesses no genes typical for an autonomously retrotransposing element. No obvious protein coding gene is directly inactivated by the insertion. However, the expression patterns of a large number of genes all around the genome are changed in columnar McIntosh Wijcik leaves.

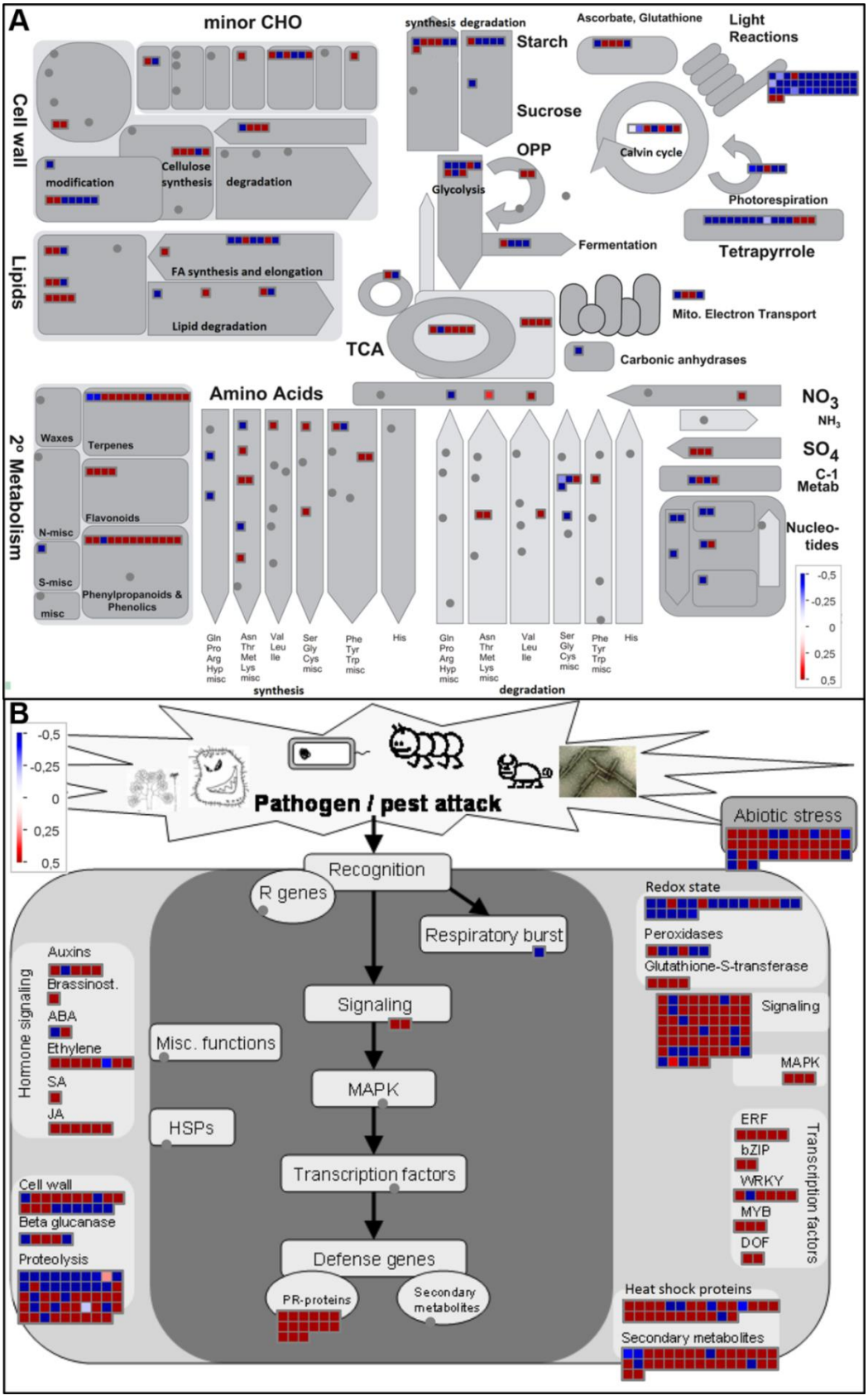


Fig. 3 Visualization of differential gene expression in leaves of Wijcik compared with McIntosh Blue squares indicate genes downregulated in McIntosh Wijcik, red squares indicate upregulated genes involved in metabolism (A) or defense (B) categories of MapMan.

Discussion

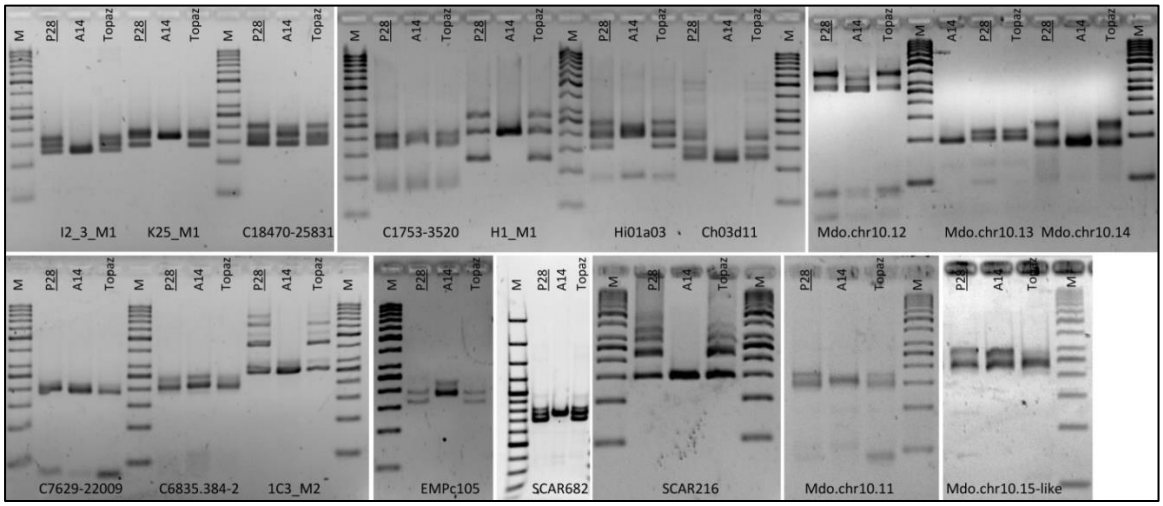
Does the Gypsy-44 insertion represent the original *Co* mutation?

In this study we have shown that the genomes of McIntosh and its columnar bud sport McIntosh Wijcik differ in the insertion of the LTR retrotransposon Gy-44 at chromosome 10, 18.8 Mb. This transposon insertion resides within the borders of the potential *Co* gene region determined by Moriya et al. (2012) and Baldi et al (2013). It is not located within the region proposed by Bai et al. (2012) as the most likely *Co* target region, which is located downstream of the regions of Baldi et al. (2013) and Moriya et al. (2012). However, as explained in the introduction, we consider the results of Baldi et al. (2013) and Moriya et al. (2012) more reliable, so that we think that the Gy-44 insertion could correspond to the original *Co* mutation or is at least tightly associated with it.

In their publications, Bai et al. (2012) and Baldi et al. (2013) considered genes for Lateral Organ Boundary Domain proteins and different transcription factors, respectively, to be the most likely candidates for *Co*, due to their location within the probable *Co* region. However, we could not detect any genomic differences between McIntosh and McIntosh Wijcik in any of these genes. Instead, our mapping and PCR analyses identified the Gy-44 insertion as the only genomic difference within the region identified as the potential *Co* locus. The McIntosh reads observed to match the 8.2 kb Gy-44 insertion in the columnar BAC sequence most probably originate from different locations of Gy-44 in the McIntosh genome, which is a typical result for mappings against repetitive sequences. Even though our BAC metacontigs originating from the columnar chromosome do not cover the whole region so that we used long-range PCR products as a reference within the range of 18.73 – 18.75 Mb and 18.869 – 18.884 Mb, we are confident that we did not overlook any genomic differences. Mappings of genomic McIntosh Wijcik data against the non-columnar contig in comparison with mappings of genomic McIntosh data against the non-columnar contig would have indicated any structural variations or other mutations that only occur in McIntosh Wijcik while being absent from McIntosh. The only possible genomic differences we would have been likely to overlook would thus be differences located outside the region we focused on.

Important indications that the integration of the Gypsy-like retrotransposon is likely identical with the columnar mutation are the results of the PCRs with primers spanning the borders of the 8.2 kb insertion. All apple varieties tested showed the expected DNA bands in the PCR after gel electrophoresis (see Fig. 2 and Supplementary Fig. 2). Some unspecific bands were detected in PCRs with the two primer pairs spanning the insertion borders, which is not surprising as one of the primers of each pair binds within the Gy-44 sequence and can therefore hybridize at about 60 locations within the apple genome. Of particular interest are the results obtained with McIntosh, Topaz and A73 (Fig. 2). We previously tested several important published molecular markers that have been shown to be linked to the *Co* gene using genomic DNA of the non-columnar cultivars McIntosh and Topaz as template. For each of them, the phenotypically non-columnar cultivar McIntosh showed the same pattern of bands as the heterozygous columnar cultivar McIntosh Wijcik (data not shown). This is due to the fact that these markers only detect sequence characteristics (such as simple sequence repeats) of the McIntosh chromosome that later acquired the *Co* mutation (being transformed into the columnar McIntosh Wijcik chromosome) and not the mutation itself. In addition, the phenotypically non-columnar cultivar Topaz showed the molecular marker genotype that is typical for the heterozygous columnar genotype *Co*/-

(see Supplementary Fig. 3). This means that the non-columnar Topaz has a genotype which is in the region of interest very similar if not identical to the non-columnar McIntosh. Only with the primers spanning the left and right borders of the Gy-44 insertion we were able to obtain the non-columnar pattern of bands for McIntosh and Topaz (Fig. 2), which is in agreement with their non-columnar phenotype. Thus, the primers for Gy-44 in this position are so far the only reliable molecular discriminator of the columnar and the non-columnar genotype.



Supplementary Fig. 3 Marker analysis of the non-columnar cultivar Topaz
PCRs were carried out with the particular published pair of primers. Genomic DNAs of the non-columnar cultivar A14, the non-columnar cultivar Topaz and the columnar cultivar P28 (underlined) were used as templates for the PCRs. C18470-25831, C1753-3520, C7629-2209 and C6835.384-2 were designed by Bai et al. (2012). Mdo.chr10.11 - Mdo.chr10.14 and Mdo.chr10.15-like are derived from Moriya et al. (2012) and Hi01a03 from Moriya et al. (2009). Primers Mdo.chr10.15-like are more specific than the published primers for Mdo.chr10.15. SCAR682 and SCAR216 originate from Tian et al. (2005), EMPc105 and Ch03d11 from Fernández-Fernández et al. (2008) and I2_3_M1, K25_M1, H1_M1 and 1C3_M2 are our own indel-based primers (Otto et al. (2013)). M: GeneRuler™ 50 bp DNA Ladder (Fermentas)

Furthermore, we discovered in a cross between a columnar and a non-columnar selection (A73-19-97K x Topaz) that the offspring was 100% columnar showing that the columnar parent was homozygous for *Co* (*Co/Co*) (Braun, unpublished data). The offspring was clearly columnar and phenotypically intermediate types often found in crosses of a heterozygous columnar parent with non-columnar varieties were not present. This obviously *Co/Co* homozygous individual is also homozygous for the Gy-44 insertion as no PCR products are detected with the primers amplifying only the sequence with the empty target site (Fig. 2). We therefore assume that the Gy-44 retrotransposon insertion represents the original *Co* mutation. As the Gy-44 LTRs show 100 % sequence identity, its insertion must have been a recent event, which would be in line with the *Co* mutation having occurred only 50 years ago. If the Gy-44 insertion was not the *Co* mutation, then one had to assume two simultaneous events having occurred independently or linked to each other. To us it seems highly unlikely that two independent events (one leading to the *Co* mutation and one leading to the Gy-44 integration) happened within such a short chromosomal distance at around the same time.

Bud sport mutations are known to be frequently caused by transposon insertions (Asíns et al. 1999; Venturi et al. 2006). For example, in apple, the apetalous varieties Rae Ime, Spencer Seedless and Wellington Bloomless originate from the insertion of an LTR retrotransposon into an intron of the apple *pistillata* homolog *MdPI*, resulting in reduced *MdPI* expression (Yao et al. 2001). Since the columnar growth phenotype of apple also arose as a bud sport mutation, it is not too surprising that the *Co* mutation could also be a retrotransposon insertion event. In apple, Ty3/Gypsy retrotransposons are the biggest family of transposable elements, representing about 25 % of the whole genome (Sun et al. 2008; Velasco et al. 2010). Even though Gy-44 possesses most of the cis elements required for transposition (LTRs, PPT and PBS) and is actively transcribed in the columnar apple varieties investigated here, it does not harbor the typical ORFs encoding GAG and a functional reverse transcriptase. However, we assume that Gy-44 is a non-autonomous transposon that has recently been mobilized by an autonomous "helper" transposon. A similar situation occurred in rice, in which the non-autonomous Ty3/Gypsy retrotransposon *Dasheng* was recently mobilized by the autonomous transposon *RIRE2*, whose LTR, PBS and PPT sequences are related to those of *Dasheng* (Jiang et al. 2002a; Jiang et al. 2002b).

What is the link between the Gy-44 insertion and the columnar phenotype?

So far it is unclear whether and how precisely the columnar phenotype could be generated by the Gy-44 insertion. It is well known that retrotransposons preferentially insert into each other at transposon-rich regions (SanMiguel et al. 1996; Kalendar et al. 1999; Ramsay et al. 1999), similar to our observation that the columnar-specific Gy-44 copy is located within Gy-33 and downstream of DNA9-5_Mad. However, to the best of our knowledge such nested retrotransposons have never been found to be linked to a specific phenotype before. Transposon insertions can induce phenotypes in many different ways (reviewed in (Lisch 2013)). They can introduce null mutations (Grandbastien et al. 1989; Vignols et al. 1995), change gene expression by destroying or providing regulatory elements (Kobayashi et al. 2004; Yao et al. 2001; Studer et al. 2011) or affecting the methylation pattern of a chromosomal region (Kinoshita et al. 2007; Fujimoto et al. 2008), introduce new genetic information (Jiang et al. 2004; Du et al. 2009) or lead to large-scale chromosomal insertions/deletions and rearrangements due to recombination between two transposon copies (Yu et al. 2011).

For the Gy-44 insertion, the disruption of a protein coding gene is highly improbable because Gy-44 is located in the 5'-LTR of the Ty3/Gypsy retrotransposon Gy-33. We also did not observe any chromosomal insertions/deletions or rearrangements within the potential *Co* gene region. As we focused on a region spanning only roughly 200 kb, a gross chromosomal rearrangement might have escaped our notice, necessitating a deeper comparative analysis of the genomic McIntosh and McIntosh Wijcik data. However, a chromosomal rearrangement caused by the recombination of the newly integrated Gy-44 with another (older) Gy-44 copy present at some distance would cause a genomic sequence flanking the "old" Gy-44 in McIntosh to be placed beside the "new" Gy-44 in McIntosh Wijcik while residing at a different chromosomal location in McIntosh. Consequently, a mapping of genomic McIntosh reads against the columnar BAC metacontig sequence from which the columnar-specific Gy-44 sequence was removed should show a chromosomal breakpoint at the empty target site. This was not the case in our mappings: If McIntosh reads are mapped against the columnar BAC metacontig sequence from which the Gy-44 sequence has been

eliminated, continuous coverage can be observed. Therefore, the only chromosomal rearrangement we would not have noticed in our mappings would be a recombination that occurred independently of the Gy-44 insertion outside the borders of the region we analyzed. This again seems to be statistically improbable as the two events (the rearrangement and the transposon insertion) would have occurred independently at the same time involving the same locus.

Gy-44 does carry one ORF that has been shown to be transcribed. However, its potential function is obscure. On the other hand, rather than providing new genetic information itself, Gy-44 could influence the level of transcription of the nearby gene MDP0000934866. This gene appears to be significantly upregulated in McIntosh Wijcik compared with McIntosh as well as in P28 compared with A14. MDP0000934866 encodes a basic helix-loop-helix protein exhibiting sequence similarity to the *Arabidopsis* transcription factor gene *embryo defective 1444* (TAIR BLAST e-Value 1e-98), which has not been characterized in detail. Since transcription factors can regulate many distinct downstream genes, this could explain the differential regulation of a large number of genes belonging to many different pathways in columnar compared with non-columnar apples. However, MDP0000934866 is located approximately 90 kb downstream of the Gy-44 insertion, and the expression levels of several genes located between Gy-44 and this gene are not affected by the presence of the additional transposon copy. It has been shown that transposons integrated into non-coding regions can affect genes that are localized several kilobases downstream of the insertion sites. For instance, the insertion of a MITE element into the conserved non-coding sequence locus *Vegetative to generative transition 1* located roughly 70 kb upstream of the AP2 transcription factor *ZmRap2.7* in maize leads to upregulation of *ZmRAp2.7* expression resulting in a late flowering phenotype (Salvi et al. 2007). Similarly, the integration of a *Hopscotch* retrotransposon 60 kb upstream of the *teosinte branched 1* (*tb1*) locus enhances expression of *tb1* in maize, which results in a severe reduction of branching in maize compared to its wild progenitor teosinte (Studer et al. 2011).

Very close to the insertion site we detected a new gene (not annotated in the apple genome sequence (Velasco et al. 2010)). This new gene is also clearly upregulated in columnar P28 shoot apical meristem compared with A14, but it showed no expression in leaves, neither of McIntosh nor of McIntosh Wijcik. According to a BLASTx search against the *Arabidopsis* genome, it encodes Downy Mildew Resistant 6 (DMR6, e-value 1e-60), a regulator of the defense response (Van Damme et al. 2005; Van Damme et al. 2008). This could explain the differential regulation of defense-related genes in columnar compared with non-columnar apple trees. Alternatively, the transcriptionally highly active Gy-44 copy might be "mistaken" by the plant as an endogenous retrovirus leading to an internal state similar to a virus infection. This speculative hypothesis is supported by the transcriptome profiles of columnar apple trees, which clearly resemble those of plants infected by RNA viruses (Smith et al. 2004; Góngora-Castillo et al. 2012). Despite the lack of an *env-like* ORF, Gy-44 might appear as virus-like particles (Wright and Voytas 1998; Laten et al. 1998; Peterson-Burch et al. 2000; Marco and Marín 2005) wrapped in protein acquired from another retroelement in trans, and this might trigger the defense response within the plant, leading to a growth habit reminiscent of an infected plant.

An obvious drastic change is the strong transcription of Gy-44 and/or other similar or identical transposons in the columnar apple varieties investigated. In maize, at least 1.5 % of expressed sequence tags originate from transposons, and *gypsy* transposons show the

highest transcriptional activity of all transposon classes (Vicient 2010). The transcription of LTR retrotransposons, at least of the *copia* family, can be induced by various biotic and abiotic stresses because some of them carry promoter sequences related to those of defense genes (Wessler 1996). Therefore, the high transcriptional activity of Gy-44 might be the consequence rather than the cause of the upregulation of stress associated genes in columnar apple trees. However, it would not explain why the upregulation of defense genes is persistent in different varieties and tissues tested, unless there was a positive feedback loop between the transposon transcripts and the defense genes. Thus, understanding the role of these transcripts might be one of the keys to elucidate the way how the genotype determines the phenotype.

It is possible that the transcriptomic changes detected in leaves result from the fact that in vivo material was used, which was not grown under tightly controlled and therefore equal conditions for McIntosh and McIntosh Wijcik and not harvested at exactly the same date and time of the day. However, upregulation of defense-associated genes, Gy-44 transcripts and of the two candidate genes mentioned above was detected not only in leaves of McIntosh Wijcik compared with McIntosh, but also in the shoot apical meristem of P28 compared with A14 based on several data sets (including biological replicates) (Krost et al. 2013). As the shoot apical meristem is responsible for the regulation of the growth habit of aerial plant organs, we consider this a crucial hint that these differences are associated with columnar growth.

In conclusion, our results suggest that the columnar phenotype is linked to a *de novo* insertion of the Gy-44 retrotransposon, which in turn leads to differential expression of nearby genes and/or to the triggering of a defense-like response within the plant. A positive proof of our hypothesis is difficult to deliver. Transformation experiments introducing the transposon sequence into non-columnar apple varieties or knocking out/down Gy-44 expression in columnar apple varieties would provide further details about the role of the Gy-44 insertion and are underway. The transformation of a transposon is a daunting task, so an alternative might be the overexpression or downregulation of the differentially regulated genes in the vicinity of Gy-44 such as *dmr6* and *bHLH*. However, transformation experiments in apple trees are difficult and time intensive and a reliable detection of an altered growth habit can only be achieved after several years. Meanwhile, in-depth transcriptome studies dealing with gene expression in different tissues and at different developmental stages in a range of genotypes could provide further evidence for a causal relationship between the Gy-44 insertion and the upregulation of its nearby genes. Transcriptomic studies might also shed light on the downstream factors influenced by the Gy-44 presence. Furthermore, the investigation of the DNA methylation state and chromatin structure of the *Co* region could be worthwhile as epigenetic changes as a consequence of the Gy-44 integration are conceivable since plants have the tendency to hypermethylate transposon sequences in order to hinder their mobility (Bennetzen et al. 1994; Lisch 2009).

Acknowledgements

This work was supported by grants of the Federal Ministry of Agriculture and Nutrition (no. 511-06.01-28-1-43.042-07 and no. 313-06.01-28-1-43.042-07). We thank Benjamin Rieger and Dr. Steffen Rapp for the bioinformatics support.

Conflict of Interest

The authors declare no conflict of interest.

Authors' Contributions

D.O. performed all work associated with BAC libraries including marker analyses and BAC sequence analysis. R.P. analyzed genomic and transcriptomic Illumina data and performed PCRs on target region genes. P.B. and B.B. provided all plant material. E.R.S. and P.B. initiated the project and E.R.S supervised the project throughout. D.O., R.P. and E.R.S wrote the manuscript.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology 215 (3):403-410. doi:10.1006/jmbi.1990.9999S0022-2836(05)80360-2 [pii]

Asíns MJ, Monforte AJ, Mestre PF, Carbonell EA (1999) Citrus and Prunuscopia-like retrotransposons. Theor Appl Genet 99 (3-4):503-510. doi:10.1007/s001220051263

Bai T, Zhu Y, Fernández-Fernández F, Keulemans J, Brown S, Xu K (2012) Fine genetic mapping of the *Co* locus controlling columnar growth habit in apple. Molecular Genetics and Genomics 287 (5):437-450. doi:10.1007/s00438-012-0689-5

Baldi P, Wolters P, Komjanc M, Viola R, Velasco R, Salvi S (2013) Genetic and physical characterisation of the locus controlling columnar habit in apple (*Malus* × *domestica* Borkh.). Mol Breeding 31(2):429-440. doi:10.1007/s11032-012-9800-1

Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome 37 (4):565-576. doi:10.1139/g94-081

Buhariwalla C, Greaves S, Magrath R, Mithen R (1995) Development of specific PCR primers for the amplification of polymorphic DNA from the obligate root pathogen *Plasmodiophora brassicae*. Physiological and Molecular Plant Pathology 47 (2):83-94

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology 268 (1):78-94. doi:http://dx.doi.org/10.1006/jmbi.1997.0951

Burgoyne LA, Wagar MA, Atkinson MR (1970) Calcium-dependent priming of DNA synthesis in isolated rat liver nuclei. Biochem Biophys Res Commun 39 (2):254-259

Conner PJ, Brown SK, Weeden NF (1997) Randomly Amplified Polymorphic DNA-based Genetic Linkage Maps of Three Apple Cultivars. J Amer Soc Hort Sci 122 (3):350-359

Du C, Fefelova N, Caronna J, He L, Dooner HK (2009) The polychromatic Helitron landscape of the maize genome. Proceedings of the National Academy of Sciences 106 (47):19916-19921

Eimert K, Reutter G, Strolka B (2003) Fast and reliable detection of doubled-haploids in *Asparagus officinalis* by stringent RAPD-PCR. The Journal of Agricultural Science 141 (01):73-78

Fernández-Fernández F, Evans K, Clarke J, Govan C, James C, Marič S, Tobutt K (2008) Development of an STS map of an interspecific progeny of *Malus*. Tree Genetics & Genomes 4 (3):469-479. doi:10.1007/s11295-007-0124-y

Fisher DV (1969) Spur-type strains of McIntosh for high density plantings. British Columbia Fruit Growers' Association Quarterly Report 14 (2):3-10

Fisher DV (1995) The 'Wijick Spur McIntosh'. Fruit Var J 49:212-213

Fujimoto R, Kinoshita Y, Kawabe A, Kinoshita T, Takashima K, Nordborg M, Nasrallah ME, Shimizu KK, Kudoh H, Kakutani T (2008) Evolution and Control of Imprinted *FWA* Genes in the Genus *Arabidopsis*. PLoS Genet 4 (4):e1000048. doi:10.1371/journal.pgen.1000048

Gelvonauskienė D, Gelvonauskis B, Sasnauskas A (2006) Impact of rootstocks on columnar apple tree growth in a nursery. Sodininkystė ir daržininkystė: 25(3):51-56

Góngora-Castillo E, Ibarra-Laclette E, Trejo-Saavedra DL, Rivera-Bustamante RF (2012) Transcriptome analysis of symptomatic and recovered leaves of geminivirus-infected pepper (*Capsicum annuum*). Virology journal 9 (1):1-16

Grandbastien M-A, Spielmann A, Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337 (6205):376-380

Grunstein M, Hogness DS (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proceedings of the National Academy of Sciences 72 (10):3961-3965

Hemmat M, Weeden NF, Conner PJ, Brown SK (1997) A DNA Marker for Columnar Growth Habit in Apple Contains a Simple Sequence repeat. J Amer Soc Hort Sci 122 (3):347-349

Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. The Journal of Heredity 85 (1):4-11

Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler SR (2002a) Dasheng: A Recently Amplified Nonautonomous Long Terminal Repeat Element That Is a Major Component of Pericentromeric Regions in Rice. Genetics 161 (3):1293-1305

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431 (7008):569-573. doi:http://www.nature.com/nature/journal/v431/n7008/suppinfo/nature02953_S1.html

Jiang N, Jordan IK, Wessler SR (2002b) Dasheng and RIRE2. A Nonautonomous Long Terminal Repeat Element and Its Putative Autonomous Partner in the Rice Genome. Plant Physiology 130 (4):1697-1705

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman A (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98 (5):704-711. doi:10.1007/s001220051124

Kesley DF, Brown SK (1992) "McIntosh Wijcik": A columnar mutation of "McIntosh" apple proving useful in physiology and breeding research. Fruit Var J 46:83-87

Kim MY, Song KJ, Hwang J-H, Shin Y-U, Lee HJ (2003) Development of RAPD and SCAR markers linked to the *Co* gene conferring columnar growth habit in apple (*Malus pumila* Mill.). Journal of Horticultural Science & Biotechnology 78 (4):512-517

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M, Kakutani T (2007) Control of *FWA* gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. The Plant Journal 49 (1):38-45. doi:10.1111/j.1365-313X.2006.02936.x

Kobayashi S, Goto-Yamamoto N, Hirochika H (2004) Retrotransposon-Induced Mutations in Grape Skin Color. Science 304 (5673):982-982

Kohany O, Gentles A, Hankus L, Jurka J (2006) Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics 7 (1):474

Krost C, Petersen R, Lokan S, Brauksiepe B, Braun P, Schmidt E (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. doi:10.1007/s11103-012-9992-0

Krost C, Petersen R, Schmidt ER (2012) The transcriptomes of columnar and standard type apple trees (*Malus* x *domestica*) - A comparative study. Gene 498:223-230. doi:10.1016/j.gene.2012.01.078

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research 40 (D1):D1202-D1210

Lapins KO (1969) Segregation of compact growth types in certain apple seedling progenies. Canadian Journal of Plant Science 49:765-768

Lapins KO (1976) Inheritance of compact growth type in apple. J Am Soc Hort Sci 101:133-135

Lapins KO, Fisher DV (1974) Four natural spur-type mutants of McIntosh apple. Canad J Plant Sci 54 (2):359-362

Lapins KO, Watkins R (1973) Genetics of compact growth habit. Report of East Malling Research Station for 1972 136

Laten HM, Majumdar A, Gaucher EA (1998) SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. Proceedings of the National Academy of Sciences 95 (12):6897-6902

Lauri PE, Lespinnasse JM (1993) The Relationship between Cultivar Fruiting-Type and Fruiting Branch Characteristics in Apple Trees. Acta Hort 349:259-263

Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C (2002) Development and characterisation of 140 new microsatellites in apple (*Malus* x *domestica* Borkh.). Mol Breeding 10:217-241

Lisch D (2009) Epigenetic regulation of transposable elements in plants. Annual review of plant biology 60:43-66

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

Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A, Moya A (2011) The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Research 39 (suppl 1):D70-D74

Marco A, Marín I (2005) Retrovirus-like elements in plants. Recent Res Devel Plant Sci 3:15-24

Meulenbroek B, Verhaegh J, Janse J (1998) Inheritance Studies with Columnar Type Trees. Acta Hort 484:255-260

Moriya S, Iwanami H, Kotoda N, Takahashi S, Yamamoto T, Kazuyuki A (2009) Development of a Marker-assisted Selection System for Columnar Growth Habit in Apple Breeding. J Japan Soc Hort Sci 78 (3):279-287

Moriya S, Okada K, Haji T, Yamamoto T, Abe K (2012) Fine mapping of *Co*, a gene controlling columnar growth habit located on apple (*Malus* × *domestica* Borkh.) linkage group 10. Plant Breeding:641-647

Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiologia Plantarum 15 (3):473-497. doi:10.1111/j.1399-3054.1962.tb08052.x

Otto D, Petersen R, Krost C, Brandl R, Brauksiepe B, Braun P, Schmidt ER (2013) Molecular Characterization of the *Co* Gene Region in *Malus* x *domestica*. Acta Horticulturae:in press

Petersen R, Krost C (2013) Tracing a key player in the regulation of plant architecture: the columnar growth habit of apple trees (*Malus* × *domestica*). Planta 238:1-22. doi:10.1007/s00425-013-1898-9

Peterson-Burch BD, Wright DA, Laten HM, Voytas DF (2000) Retroviruses in plants? Trends in Genetics 16 (4):151-152

Ramsay L, Macaulay M, Cardle L, Morgante M, Ivanissevich Sd, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. The Plant Journal 17 (4):415-425. doi:10.1046/j.1365-313X.1999.00392.x

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashev S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski JA, Tingey SV, Miao G-H, Phillips RL, Tuberosa R (2007) Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proceedings of the National Academy of Sciences 104 (27):11376-11381

SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested Retrotransposons in the Intergenic Regions of the Maize Genome. Science 274 (5288):765-768

Sheng Y, Mancino V, Birren B (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. Nucleic Acids Res 23 (11):1990-1996

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proceedings of the National Academy of Sciences of the United States of America 89 (18):8794-8797

Smith CM, Rodriguez-Buey M, Karlsson J, Campbell MM (2004) The response of the poplar transcriptome to wounding and subsequent infection by a viral pathogen. New Phytologist 164 (1):123-136. doi:10.1111/j.1469-8137.2004.01151.x

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98 (3):503-517

Stekel DJ, Git Y, Falciani F (2000) The Comparison of Gene Expression from Multiple cDNA Libraries. Genome Research 10 (12):2055-2061

Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene tb1. Nat Genet 43 (11):1160-1163. doi:http://www.nature.com/ng/journal/v43/n11/abs/ng.942.html#supplementary-information

Sun H-Y, Dai H-Y, Zhao G-L, Ma Y, Ou C-Q, Li H, Li L-G, Zhang Z-H (2008) Genome-wide Characterization of Long Terminal Repeat -retrotransposons in Apple Reveals the Differences in

Heterogeneity and Copy Number between Ty1-copia and Ty3-gypsy Retrotransposons. Journal of Integrative Plant Biology 50 (9):1130-1139. doi:10.1111/j.1744-7909.2008.00717.x

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal 37 (6):914-939. doi:10.1111/j.1365-313X.2004.02016.x

Tian Y-K, Wang C-H, Zhang J-S, James C, Dai H-Y (2005) Mapping *Co*, a gene controlling the columnar phenotype of apple, with molecular markers. Euphytica 145 (1):181-188

Tobutt KR (1985) Breeding columnar apples at East Malling. Acta Hortic 159:63-68

Tobutt KR (1994) Combining apetalous parthenocarpy with columnar growth habit in apple. Euphytica 77 (1-2):51-54. doi:10.1007/bf02551460

Vali U, Brandstrom M, Johansson M, Ellegren H (2008) Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. BMC Genetics 9 (1):8

Van Damme M, Andel A, Huibers RP, Panstruga R, Weisbeek PJ, Van den Ackerveken G (2005) Identification of Arabidopsis Loci Required for Susceptibility to the Downy Mildew Pathogen *Hyaloperonospora parasitica*. Molecular Plant-Microbe Interactions 18 (6):583-592. doi:10.1094/mpmi-18-0583

Van Damme M, Huibers RP, Elberse J, Van den Ackerveken G (2008) Arabidopsis *DMR6* encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. The Plant Journal 54 (5):785-793. doi:10.1111/j.1365-313X.2008.03427.x

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagne D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouze P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2010) The genome of the domesticated apple (*Malus* x *domestica* Borkh.). Nat Genet 42 (10):833-839. doi:ng.654 [pii]10.1038/ng.654

Venturi S, Dondini L, Donini P, Sansavini S (2006) Retrotransposon characterisation and fingerprinting of apple clones by S-SAP markers. Theor Appl Genet 112 (3):440-444. doi:10.1007/s00122-005-0143-8

Vicient C (2010) Transcriptional activity of transposable elements in maize. BMC Genomics 11 (1):601

Vignols F, Rigau J, Torres MA, Capellades M, Puigdomènech P (1995) The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. The Plant Cell Online 7 (4):407-416

Wessler SR (1996) Plant retrotransposons: Turned on by stress. Current Biology 6 (8):959-961. doi:http://dx.doi.org/10.1016/S0960-9822(02)00638-3

Wright DA, Voytas DF (1998) Potential Retroviruses in Plants: Tat1 Is Related to a Group of *Arabidopsis thaliana* Ty3/gypsy Retrotransposons That Encode Envelope-Like Proteins. Genetics 149 (2):703-715

Yao J-L, Dong Y-H, Morris BAM (2001) Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. Proceedings of the National Academy of Sciences 98 (3):1306-1311

Yu C, Zhang J, Peterson T (2011) Genome Rearrangements in Maize Induced by Alternative Transposition of Reversed Ac/Ds Termini. Genetics 188 (1):59-67

Zhang YG, Dai HY (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

Zohary D, Hopf M (2000) Domestication of plants in the Old World. 3rd edition edn. Oxford University Press, Oxford

Electronic Supplementary Material

Supplementary Table 2. Analysis of differential gene expression. Reads of the leaf transcriptomes of McIntosh and McIntosh Wijcik were mapped to the apple genome (Velasco et al. 2010) and the number of total gene reads were extracted for each MDP. All annotated genes (MDPs) were assigned to UniProtKB/SwissProt hits and differential expression of unigenes was evaluated based on R_j values (Stekel et al. 2000). Unigenes are ranked according to their significance of differential expression.
This table is included in the electronic supplementary material accompanying this thesis

Paper 4

The Transcriptomes of Columnar and Standard Type Apple Trees (*Malus x domestica*) – a Comparative Study

Clemens Krost[1], Romina Petersen[1], Erwin R. Schmidt[1]

[1] Department of Molecular Genetics, Johannes Gutenberg-University of Mainz

This paper was published as a research article in Gene in 2012 (*Gene* **498**(2): 223-230) and has been included in this thesis with permission.

Abstract

Columnar apple trees (*Malus x domestica*) provide several economic advantages due to their specific growth habit. The columnar phenotype is the result of the dominant allele of the gene *Co* and is characterized by thick stems with short internodes and reduced lateral branching. *Co* is located on chromosome 10 and often appears in a heterozygous state (*Co*/*co*). The molecular explanation of columnar growth is not well established. Therefore, we studied the transcriptomes of columnar and standard type apple trees using 454 and Illumina next generation sequencing (NGS) technologies. We analyzed the transcriptomes of shoot apical meristems (SAMs) because we expect that these organs are involved in forming the columnar growth phenotype. The results of the comparative transcriptome analysis show significant differences in expression levels of hundreds of genes. Many of the differentially expressed genes are associated with membrane and cell wall growth or modification and can be brought in line with the columnar phenotype. Additionally, earlier findings on the hormonal state of shoots of columnar apples could be affirmed. Our study resulted in a large number of genes differentially expressed in columnar vs. standard type apple tree SAMs. Although we have not unraveled the nature of the *Co* gene, we could show that the modified expression of these genes, most likely due to the presence of *Co,* can determine the columnar phenotype. Furthermore, the usefulness of NGS for the analysis of the molecular basis of complex phenotypes is discussed.

Keywords

Co gene, NGS, Illumina, 454, *de novo* assembly, MAPMAN

Introduction

Columnar growth in apple (*Malus x domestica*) was initially discovered in Canada in the early 1960s as a natural mutation of the cultivar 'McIntosh' (Fisher, 1969), described as 'McIntosh Wijcik'. The columnar phenotype is similar to spur type trees (Blažek, 1992) and is mainly characterized by short internodes and axillary buds growing into spurs (Tobutt, 1985). Due to short lateral branches and a limited canopy, columnar apple trees are well adapted to high density plantings. Economic advantages of such orchards would be higher yields per hectare, minimal pruning and possible use of mechanical harvesting (Tobutt, 1985). Several genetic studies performed by crossing columnar parents showed that columnar growth phenotype is controlled by a dominant allele of the gene *Columnar* (*Co*) (Lapins, 1969; Lapins and Watkins, 1973; Blažek, 1992). So far, all available columnar apple varieties seem to be heterozygous for *Co* (Tian et al., 2005). Studies using molecular markers revealed that *Co* is located on linkage group 10 (Tian et al., 2004; Tian et al., 2005). The involvement of growth regulating hormones on the columnar phenotype seems to be obvious. Several studies showed differences in endogenous levels of abscisic acid (ABA), indoleacetic acid (IAA), gibberellic acid (GA) and cytokinins (Looney and Lane, 1984; Looney et al., 1988; Watanabe et al., 2004; Watanabe et al., 2008). If this is directly affected by Co or just a cross-regulatory effect (Kuppusamy et al., 2009) could not be determined. Recent studies revealed that overexpression of the flower meristem identity gene *LEAFY* (*LFY*) leads to a columnar phenotype (Flachowsky et al., 2009). However, in spite of the knowledge about the *Co* gene, the molecular mechanisms leading to the columnar growth phenotype are not yet known.

NGS technologies have been successfully used to study transcriptomes of both model and non-model organisms (Morozova et al., 2009; Wall et al., 2009). The use of high-throughput technology is especially expedient when no expression studies have been performed before. There is a variety of applications for the global characterization of special tissues (Wang et al., 2009; Wang et al., 2010), discovery of putative and novel genes (Sun et al., 2010) or comparison of genotypes (Alagna et al., 2009). Therefore, we decided to use next generation sequencing to analyze the transcriptomes of columnar and standard type apple trees.

In this study, we used deep sequencing with 454 and Illumina NGS technologies for the comprehensive transcriptome characterization of SAMs in columnar and standard type apple trees. The comparison of the transcriptomes should display differences in the expression pattern of genes somehow involved in the columnar phenotype. We do not expect to identify the *Co* gene by the comprehensive analysis of the gene network involved in apical growth. However, it should give some indications how the columnar phenotype might arise. Furthermore, we describe a way of data processing which can be universally used for rapid comparison of plant transcriptomes.

Material and methods

Plant material

Two apple cultivars, standard type 'A14-190-93' (A14) and columnar type 'Procats28' (P28), grown at the Department of Pomology, Geisenheim Research Centre, were used in

this study. For 454 sequencing, SAM spur samples of columnar P28 and of branches of non-columnar A14 were collected in late May 2009, kept on ice during transport and stored at -80°C. A second collection for SAMs of P28 was performed in late September 2009. For Illumina sequencing, only A14 material collected in May and P28 material collected in September was used.

Sample preparation

SAMs were roughly freed from leaf, petiole and lower shoot tissue. Total RNA was isolated from SAMs using liquid nitrogen for disruption of plant material and the innuPREP Plant RNA Kit (Analytik Jena, Jena, Germany). Up to 120 µg of total RNA was further purified using the NucleoTrap mRNA Kit (Macherey-Nagel, Düren, Germany). Purified mRNA was amplified using the MessageAmp II aRNA Amplification Kit (Ambion, Austin, USA). For 454 sequencing, cDNA synthesis was carried out with SuperScript III Reverse Transcriptase (Invitrogen, Darmstadt, Germany) and 5 µM Random Hexamer Primer (Fermentas, St. Leon-Roth, Germany). Second strand cDNA synthesis and blunting was conducted following the Fermentas Second Strand cDNA synthesis protocol available at http://fermentas.com/en/support/Application-Protocols/.

Sequencing

About 5 µg of double stranded (ds) cDNA was used for 454 pyrosequencing using GS FLX Titanium chemistry (Seq-IT, Kaiserslautern, Germany). Amplified antisense RNA (aRNA) was sequenced on the Genome Analyzer GA_{IIx} using 2 x 100 bp paired end sequencing mode (Cologne Center for Genomics, Cologne, Germany). Quality control of total RNA, aRNA and ds cDNA was accomplished on a Bioanalyzer (Agilent, Böblingen, Germany) RNA Nano, RNA Pico and DNA High Sensitivity Chips, respectively.

Assembly

De novo assembly of 454 raw sequencing data was performed using the assembler software SeqMan NGen (DNASTAR, Madison, USA) with default parameters. Adapter and quality trimming was included to obtain high-quality reads. De novo assembly of Illumina raw sequencing data was conducted using ABySS (Assembly BY Short Sequences) 1.2.5 (Birol et al., 2009; Simpson et al., 2009) with the following parameters: k-mer length k=83, k-mer coverage c=10, endtrimming e=5, paired-end data n=2, seed length s=166 (=2k).

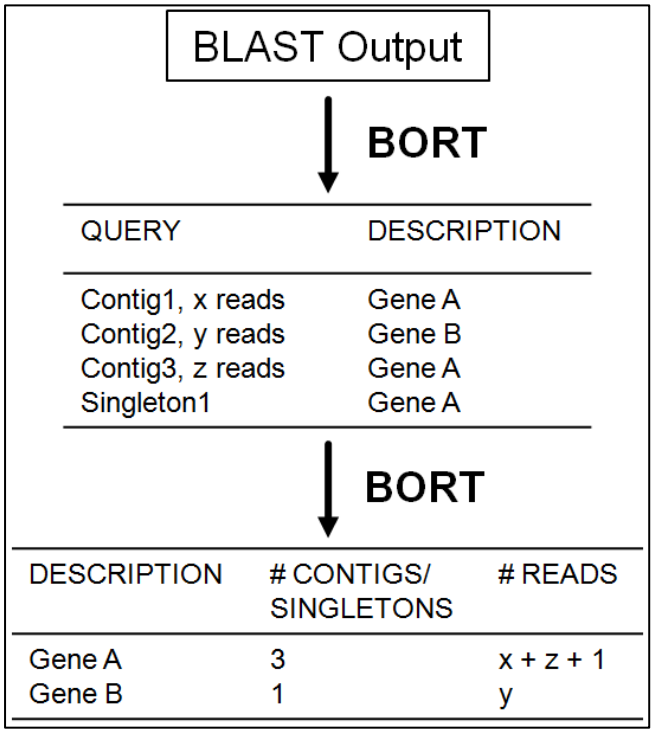
Data processing

Contigs only (Illumina) or contigs and singletons (454) were analyzed by highly parallel BLASTx similarity searches (Altschul et al., 1990) against the UniProtKB database (UniProt-Consortium, 2010). The BLASTx output was trimmed and quantified using an in-house Perl script named BORT (Blast Output Refinement Tool) version 1.3 (Additional file 1) or 2.1 (Additional file 2). BORT is programmed to only consider the best BLAST hit for each query and count 454 or Illumina reads (or k-mers) in order to quantify unique genes for every data set (BORT1.3 subitem “quantification” for 454 data and BORT2.1 subitem “quantification by standard blasthits” for Illumina data). BORT2.1 was introduced because it can handle large BLAST outputs due to independence of local RAM. The function of BORT in NGS data processing is illustrated (Additional file 3). Unigenes can be further collapsed (BORT1.3 subitem “quantification swissprot”) to unique UniProtKB entries if a single gene

appears in more than one species. Differential regulation of the unigenes obtained was assessed by calculating R_j values (Stekel et al., 2000) and ranking according to them. As an assembly control we also conducted BLASTx searches of 454 and Illumina raw sequencing data against annotated *Malus x domestica* proteins (Velasco et al., 2010) and processed the output as described above.

MAPMAN Analysis

For appropriate mapping, a BLASTp search of differentially regulated genes against *Populus trichocarpa* proteins (Tuskan et al., 2006) was performed and used for functional characterization and visualization with MAPMAN software 3.5.0 (Thimm et al., 2004). Illumina read counts for all genes were normalized (A14 set to 1) and the natural logarithm was chosen for conformable scaling according to P28 fold change.



Additional File 3 - Data processing using BORT for comparison of large transcriptome data sets.
BORT can be used to perform a user-defined trimming of BLAST outputs from highly parallelized similarity searches (upper arrow) and quantification of unique descriptions (lower arrow). Example is given for four queries coding for two genes A and B (Contig1,2,3 and Singleton1 as a typical result of 454 data assembly using SeqMan NGen; x,y,z representing number of reads assembled in each contig).

Results

Assembly and data processing

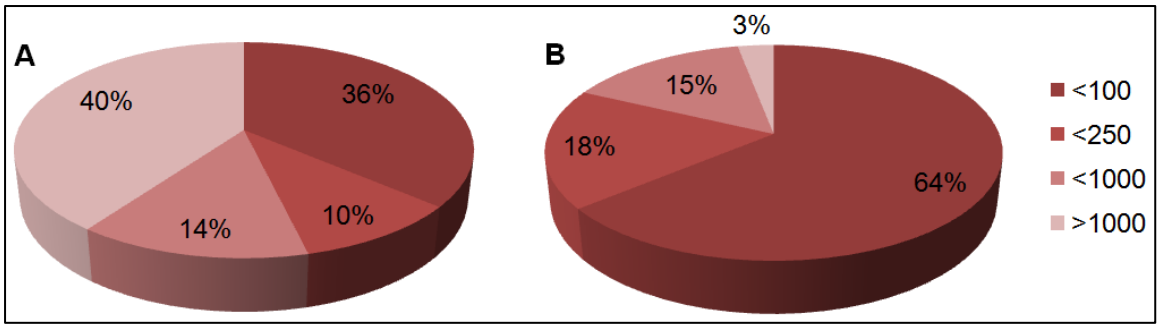
To identify genes differentially expressed in columnar habit type apple (*Malus x domestica*), we sequenced cDNAs generated from SAMs of the two accessions Procats28 (P28 - columnar) and A14-190-93 (A14 - standard type) using 454 and Illumina next generation sequencing (NGS) technologies (EBI Sequence Read Archive accession ERP000629). The statistics of the sequence data obtained is summarized in Table 1. The first 454 sequencing run of P28 SAMs (collected in May) only yielded 32,609 reads due to multiple technical problems and was therefore repeated with SAMs collected in September of the same year. All three 454 data sets were used for expression studies, but only the larger one is displayed in Table 1 for increased comparability with the two Illumina data sets of the same time points. De novo assembly of 454 reads was performed using SeqMan NGen (DNASTAR) and yielded 24,146 contigs for P28 and 24,620 contigs for A14. 39,194 contigs (P28) and 30,643 contigs (A14) were obtained from de novo assembly of Illumina paired-end reads using ABySS (Birol et al., 2009; Simpson et al., 2009) with optimized k-mer values (Surget-Groba and Montoya-Burgos, 2010). Both methods had a total of about 18 million bases assembled into contigs with mean coverage ratios of 5 (454) and 400 (Illumina).

Table 1 - 454 and Illumina raw sequence data and assembly.
Minimum contig and singleton length is dependent on trimming parameters (SeqMan NGen) and choice of k-mer length (ABySS).

	454 (SeqMan NGen)		**Illumina (ABySS)**	
	Procats28	*A14-190-93*	*Procats28*	*A14-190-93*
Sequencing				
Number of reads	245,297	276,038	76,177,216	80,289,084
# total bases	98,471,296	109,909,375	7,617,721,563	8,028,908,497
Average read length	401	398	100	100
Contigs				
Number of contigs	24,146	24,620	39,194	30,643
# bases assembled as contigs	18,087,985	18,038,940	18,416,635	18,192,061
Average contig length	749	733	470	594
Range of contig length	50 – 3,506	50 – 3,302	83-2,864	83-2,227
Singletons				
Number of singletons (# bases)	27,251 (12,782,403)	23,024 (8,294,039)	-	-
Average length of singletons	469	360	-	-
Range of singleton length	53 – 1,200	50 - 622	-	-
Unique sequences (bases)	**51,397 (30,870,388)**	**47,644 (26,332,979)**	**39,194 (18,416,635)**	**30,643 (18,192,061)**

Contigs and singletons (454) as well as contigs only (Illumina) were further used for highly parallel BLASTx similarity searches against UniProtKB database (E-value cutoff 10^{-3}). Trimming of BLAST output and re-formatting into unique genes was done using Blast Output Refinement Tool (BORT, see Additional file 3). Data processing as described above resulted in 9,030 (454) (Additional file 4) and 7,252 (Illumina) unigenes (Additional file 5). Statistical analysis (Stekel et al., 2000) was used to rank unigenes according to the level of statistical significance of the differential expression (R_j). The order (not the presence) of genes varied clearly between the two methods, but the ratio of reads (NGen) and k-mers (ABySS) between P28 and A14 was very similar. However, 40% of genes taken into account were found at highly different positions (more than 1000 positions apart) in each Unigene list (Additional files 4 and 5) showing divergent read or k-mer ratios (Additional file 6A).

To test the quality and applicability of both assemblies, 454 and Illumina raw sequencing data were simultaneously BLASTx searched (E-value cutoff 10^{-10}) against all *Malus x domestica* predicted proteins (MDPs; Velasco et al., 2010) and processed as described above. 8,474 unigenes (Additional file 7) were obtained from 454 data while processing of Illumina data yielded 9,819 unigenes (Additional file 8). MDPs were BLASTp searched (E-value 10^{-3}) against UniProtKB database to obtain unigene descriptions. The similarity of generated lists significantly increased when raw sequencing data and MDPs were used instead of assembled contigs as a reference (Additional file 6B), meaning that most of the genes taken into account were located in similar positions in each unigene list (Additional files 7 and 8). Therefore, the unigene lists created by processing of Illumina raw sequencing data without assembling were chosen for further analysis due to highest coverage and reliability.



Additional File 6 - Comparison of unigene lists created by two different ways of data processing.
Top 100 differentially expressed genes from generated lists. The figure shows distributions resulting from (A) processed contigs and (B) processed raw sequencing data. Numbers adjacent to color code display the distance of genes (position in each unigene list) between the lists generated from 454 and Illumina data.

MAPMAN analysis

MAPMAN software (Thimm et al., 2004) was used for the visualization of differential gene expression in columnar P28 SAM. MAPMAN allows the visualization of individual gene expression changes in a diagram, with genes grouped in "BINs" by function or class (Pauwels et al., 2008). The MAPMAN diagram of *Malus* metabolism (using *Populus* mapping) of top 500 unigenes from the comparison of A14 and P28 sampled in May (Figure 1A) and A14 sampled in May and P28 sampled in September (Figure 1B) showed significant differences in SAM gene expression in columnar versus standard growth type, mostly

independent of sampling time. Genes of BINs representing light reactions, mitochondrial electron transport, lipid metabolism (in particular fatty acid (FA) synthesis and FA elongation) and cell wall (subBIN "modification"), the latter mainly consisting of expansins and xyloglucan endotransglucosylases/hydrolases (Liu et al., 2007a), were significantly downregulated. Another group of genes involved in cell wall modification, summarized in subBINs "pectate lyases and polygalacturonases" and "cellulose synthase" of BIN cell wall were upregulated in both P28 samples, together with BINs corresponding to terpenoid and tryptophan synthesis (Figure 1). Analyses were repeated using all unigene lists described to eliminate the possibility of a bias caused by the way of data processing. As expected, all BINs described to be over- or underrepresented in P28 could hereby be confirmed.

To improve our insights into the global state of P28 SAM cells and to gather possible effects of *Co* gene product, MAPMAN analysis was extended to cell functions using the top 200 unigenes found to be differentially regulated. Genes associated with DNA synthesis (especially subBIN "chromatin structure"), RNA processing and protein synthesis were downregulated as would be expected for a mutant with reduced growth phenotype (Figure 2A and B). Genes upregulated in both P28 samples were mainly grouped into BINs corresponding to transport and protein modification. All genes of BINs corresponding to cell wall and membrane modification that showed the same regulation in P28 samples of both sampling times including the rate of regulation and RPKM values are summarized in Table 2.

Since the columnar phenotype is the result of abnormal growth, we focused on genes involved in hormone metabolism. Phytohormones have major effects on plant growth and development, especially in meristematic tissues (Galinha et al., 2009). Therefore, genes in the BIN hormone metabolism were analyzed in detail. This BIN was subdivided into subBINs "hormone metabolism" of indole-3-acetic acid (IAA), jasmonate (JA) and gibberellic acid (GA) (Figure 2, red frame). Genes associated with the synthesis of precursors and signal transduction of IAA and JA were upregulated whereas genes associated with GA signal transduction were downregulated. These results match the finding that the endogenous level of free IAA is higher in shoots and spur buds of columnar apple than in apple trees with standard habit. GA level, on the contrary, is lower in columnar compared to normal apple trees (Looney and Lane, 1984; Looney et al., 1988; Watanabe et al., 2004). Jasmonates are known to act as growth inhibitors (Wasternack, 2007), presumably causing a cell cycle arrest in G2 (Pauwels et al., 2008). We found cell cycle regulators expressed in G2 phase and M phase such as cyclin B and Cdk B (Hartig and Beck, 2006) to be upregulated in columnar SAM, whereas D type cyclins occurring in G1 as well as cyclin A3-1 which promotes the G1/S transition were downregulated. This suggests that in contrast to normal apple, cells of the columnar apple are more likely to be in G2 than in G1, which would be in line with a cell cycle arrest in this phase and would thus provide a possible reason for the stunted growth.

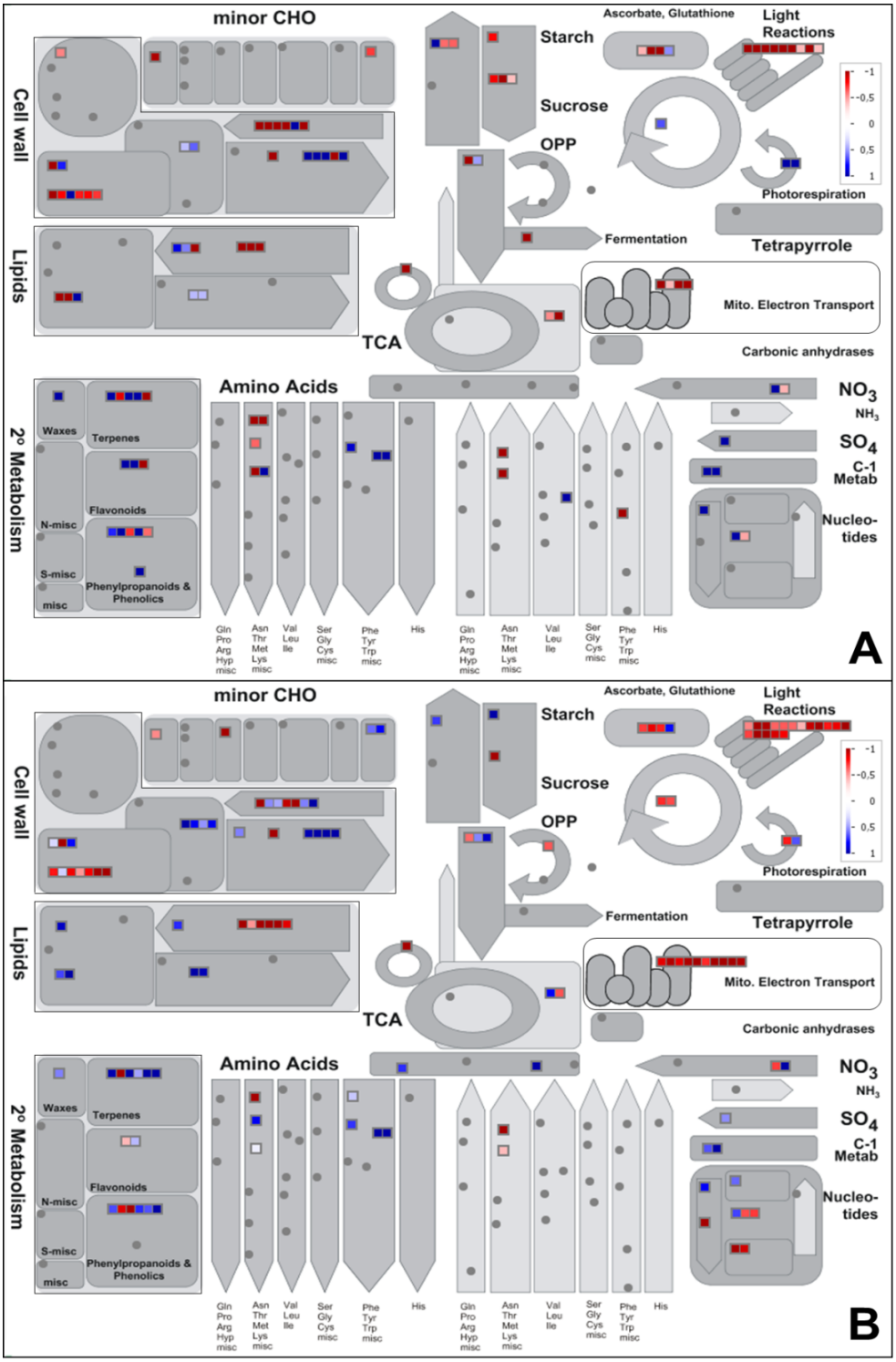


Figure 1 - MAPMAN visualization of differences in gene expression of metabolic processes.
The figure shows differences in gene expression found in SAMs of columnar P28 sampled in May (A) and September (B) compared to standard type A14. Red and blue boxes represent repressed and induced genes, respectively. Values in the key imply ln of normalized Illumina read counts. Top 500 unigenes from raw sequencing data processing ranked by R_j values and normalized to A14 were used for analyses. Genes of BINs associated with light reactions, mitochondrial electron transport, cell wall modification and fatty acid synthesis and elongation are repressed while genes of BINs associated with terpenoid metabolism, cellulose synthesis, cell wall degradation and tryptophan synthesis are induced in P28 transcriptome.

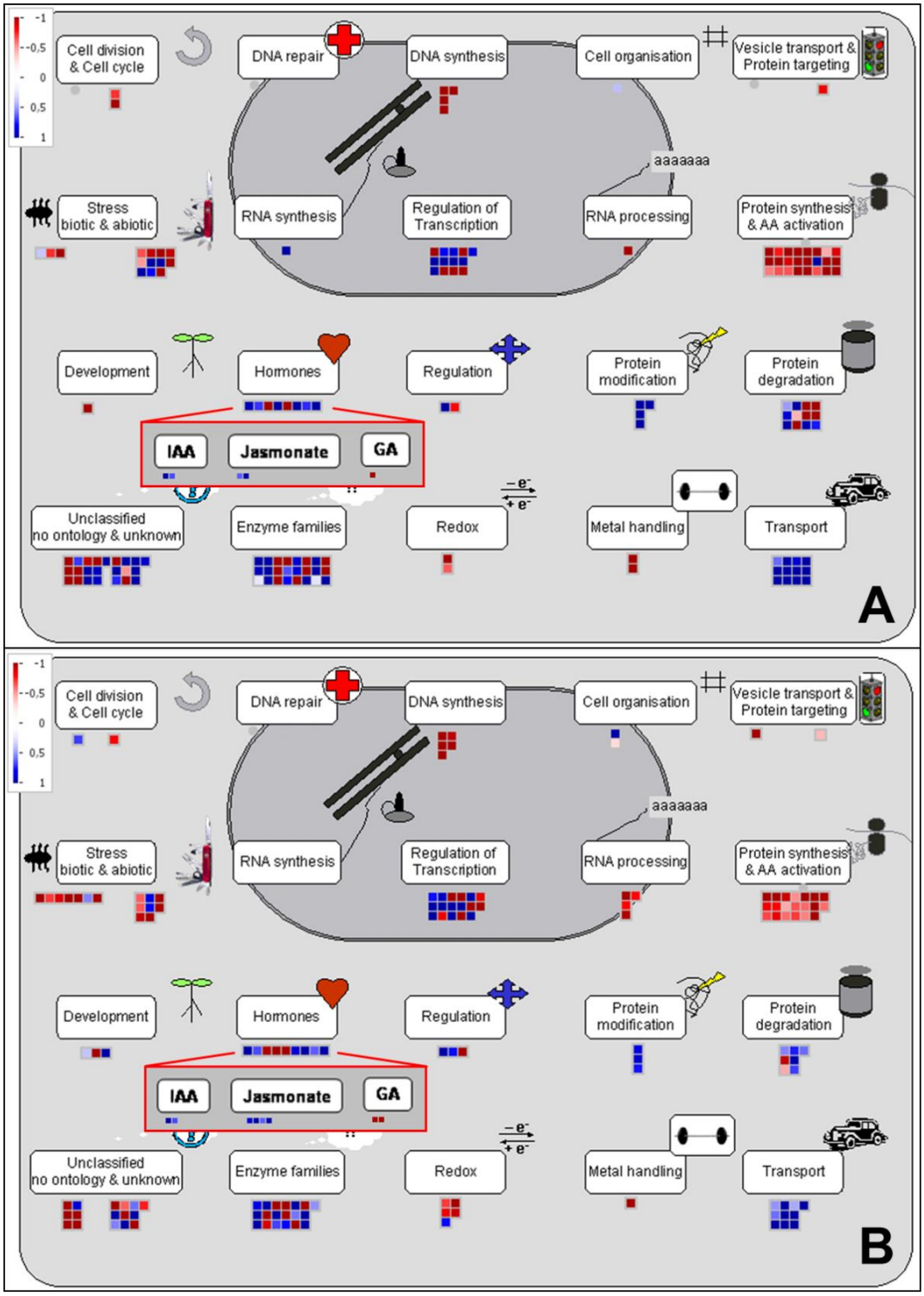


Figure 2 - MAPMAN visualization of differences in gene expression levels for various categories of genes.

The figure shows differences in gene expression found in SAMs of columnar P28 sampled in May (A) and September (B) compared to standard type A14. Red and blue boxes represent repressed and induced genes, respectively. Values in the key imply ln of normalized Illumina read counts. Top 200 unigenes from raw sequencing data processing sorted by R_j values and normalized to A14 were used for analyses. Genes of BINs associated with DNA synthesis, RNA processing and protein synthesis are repressed while genes of BINs associated with transport and protein modification are induced. BIN hormone metabolism is shown in detail for IAA, JA and GA (red frame).

BIN	subBIN	UniProtKB description	RPKM A14 (May)	RPKM P28 (May)	RPKM P28 (September)	Mean P28 expression
Cell wall	cellulose synthesis	protein COBRA	35.2	71.0	115.7	**2.3** ↑
	degradation.pectate lyases and polygalacturonases	probable pectate lyase 5	26.9	108.3	105.6	**5.1** ↑
		dehydration-responsive protein rd22	122.2	656.6	1433.0	**7.7** ↑
	modification	xyloglucan endotransglucosylase/hydrolase protein 9	1821.0	525.0	1048.2	**2.0** ↓
		brassinosteroid-regulated protein bru1	310.5	78.1	140.7	**2.8** ↓
		probable xyloglucan endotransglucosylase/hydrolase protein 23	80.6	37.2	8.7	**5.9** ↓
Lipid metabolism	FA synthesis and elongation.beta ketoacyl CoA synthase	3-ketoacyl-CoA synthase 1	58.6	1.2	9.2	**5.7** ↓
		3-ketoacyl-CoA synthase 10	505.1	154.6	58.5	**3.8** ↓
		3-ketoacyl-CoA synthase 6	166.1	21.7	47.7	**5.0** ↓
Transport	sugars	polyol transporter 5	101.8	378.0	308.9	**3.2** ↑
	ABC transporters and multidrug resistance systems	abc transporter g family member 11	19.6	97.8	249.6	**6.3** ↑
	calcium	vacuolar calcium ion transporter	170.1	606.5	570.7	**3.2** ↑
		cation/calcium exchanger 4	17.5	528.3	92.4	**6.9** ↑
		vacuolar cation/proton exchanger 3	69.4	517.5	374.7	**5.3** ↑
Hormone metabolism	auxin.induced-regulated-responsive-activated	DUF246 domain-containing protein	83.2	521.7	314.3	**4.2** ↑
	jasmonate.synthesis-degradation.lipoxygenase	lipoxygenase 2, chloroplastic	244.2	629.4	1871.1	**2.7** ↑
	jasmonate.synthesis-degradation.allene oxide cyclase	allene oxide cyclase 4, chloroplastic	440.6	1728.9	1160.9	**3.0** ↑
	gibberellins.signal transduction	Scarecrow-like protein 23	216.8	53.5	72.1	**3.4** ↓

Table 2 - Expression levels of genes involved in membrane and cell wall modification, transport and hormone metabolism.
The table shows BINs and subBINs as provided by MAPMAN. Expression levels are given in RPKM values and mean P28 expression in fold change (up ↑ or down ↓) relative to levels in A14 as derived from NGS read counts.

Discussion

cDNA library comparability depends on NGS data processing

The pivotal problem of NGS data mining is that almost every application needs individual data processing because no standardized pipelines exist. Furthermore, for non-model organisms, de novo assembly of transcriptome data produced by NGS seems to be obligatory for detailed analysis and gene quantification. De novo assembly of 454 and Illumina sequence data using appropriate assembly tools resulted in highly similar numbers of bases assembled into contigs. The results of 454 data assembly for both libraries were highly comparable in number of total contigs, mean contig length, range of contig length and number of singletons, whereas these values were divergent for Illumina data assembly output. This could be attributed to the range of overlapping paired-end reads which was higher for the P28 cDNA library, whose assembly resulted in a higher number of total contigs with reduced average lengths. Despite this high comparability, unigene lists created by data processing as described above showed obvious differences and therefore do not provide a suitable basis for comparative analysis. However, we were able to show that all unigene lists yielded the same results when subjected to MAPMAN analyses. The comparability of unigene lists obtained from contig processing was significantly improved when data were processed without any kind of assembly and with an additional step of comparing raw reads with annotated MDPs prior to comparison with UniProtKB. Unfortunately, these methods can only be applied to organisms where appropriate gene sets or sequenced genomes exist, such as for apple. However, for non-model organisms where full genome sequences are not available, straight BLAST searches of raw sequencing reads against UniProtKB and subsequent statistical comparison of reads counts as described would be the best possible approach for differential expression analysis.

Evidence on the involvement of membrane modification and cell wall related genes in the development of the columnar phenotype

MAPMAN analysis revealed a total of five groups of genes associated with modifications of membranes and synthesis of the cell wall that are significantly up- or downregulated in the columnar SAM. This includes genes for proteins involved in transport, pectin degradation, cellulose synthesis and deposition, cell wall modification and FA synthesis and elongation (see Table 2).

One subgroup of genes, which is downregulated, codes for expansins and xyloglucan endotransglucosylases/hydrolases (XTHs). These proteins have been shown to play an important role in cell wall growth (Liu et al., 2007a) where the maintenance of the cellulose/xyloglucan framework is a critical step in plant cell expansion (Liu et al., 2007b). Downregulation or knock-out of XTH family members like XTH9 is characterized by decreased length of intermodal cells resulting in short internodes, which are typical of columnar apple trees (Hyodo et al., 2003).

A second subgroup containing genes involved in cellulose synthesis and pectin degradation was found to be upregulated in P28, which should result in a higher content of cellulose and a reduced content of pectin in the cell wall. This may lead to an impairment of cell wall growth which would result in reduced cell size in columnar type apple trees. Since the trunk of columnar apple trees reaches approximately normal height and spurs develop

into longer shoots if the leader is damaged (Tobutt, 1985), apical dominance seems to overcome the growth inhibition in main shoots, but not in spurs.

We also found down- or upregulation of genes involved in membrane formation and modification. A group of 3-ketoacyl-CoA synthases and acyl carrier proteins (ACPs) was found to be downregulated, which should result in reduced FA synthesis and elongation. This could lead to reduced growth due to modified lipid composition as FAs are involved in various aspects of plant growth (Ohlrogge and Browse, 1995). The group of membrane modifying transport proteins (see Table 2) is significantly upregulated in P28, but an obvious relation to columnar growth is hard to spot. A major subgroup of those upregulated genes is involved in calcium transport. Family members of calcium exchangers (CAXs) and cation calcium exchangers (CCXs) are known to play important roles in growth and ion homeostasis (Cheng et al., 2005; Morris et al., 2008) and overexpression of AtCCX3 in tobacco (*Nicotiana tabacum*) was shown to result in stunted growth (Morris et al., 2008). Hence it is not too surprising to find an upregulation of such genes in the columnar growth mutant.

In summary, our results show that the columnar growth phenotype is linked to relevant changes in the expression patterns of many genes. Particularly, genes involved in membrane and cell wall function obviously participate in the formation of the columnar phenotype. In addition, there is a large number of genes which are differentially regulated but were not categorized into one of the above mentioned BINs (data not published). Genes involved in RNA processing and protein synthesis were found to be significantly downregulated which is in agreement with reduced cell growth activity in developing SAMs of P28. Since chromatin synthesis (BIN DNA synthesis, Figure 2) is usually upregulated in meristematic tissues, our findings would improve the hypothesis that cell cycle regulators contribute to a potential cell cycle arrest in G2 in columnar SAMs.

Membrane modifications lead to impairment of mitochondria and chloroplast functions

The reduced expression of members of the 3-ketoacyl-CoA synthase family mentioned above is also responsible for the impairment of mitochondria and chloroplast functions (Wu and Xue, 2010). This can be observed in the regulation of genes involved in membrane-associated processes like light reactions (chloroplast) and mitochondrial electron transport which are strongly downregulated in P28. It could not be assessed whether this is limited to SAMs only, but size reduction or significant phenotypes resulting from impaired chloroplast function could not be observed in other tissues. Additionally, redox reactions seem to be globally impaired, which can be inferred from BIN redox (Figure 2) and BIN ascorbate and glutathione (Figure 1). This could be due to the possible disturbance of ion homeostasis which is necessary for cellular redox reactions.

Jasmonates and Auxin are induced while Gibberellins are repressed

The data show that genes involved in biosynthesis of jasmonates and auxins (including tryptophan and anthranilate synthesis as precursors of IAA (Tromas and Perrot-Rechenmann, 2010)) are strongly upregulated while genes involved in GA signaling are mainly downregulated. Furthermore, none of the genes corresponding to terpenoid biosynthesis (as a potential source of GA precursors) could be assigned to GA, but instead to linalool, myrcene (up) and tocopherol (down) synthesis which can be ascribed to increased

JA biosynthesis and signaling (Munne-Bosch and Falk, 2004; Huber et al., 2005; Munne-Bosch et al., 2007; van Schie et al., 2007). These findings perfectly agree with studies on the concentrations of GA (Looney et al., 1988) and IAA (Watanabe et al., 2004) in shoots of columnar apple. Detailed analysis of hormone biosynthesis, signaling and interaction as well as its effects on SAM growth is in progress.

Conclusions

Analysis of the transcriptome of columnar apple tree SAMs revealed a vast number of differentially regulated genes grouped into eight functional categories (BINs). Three groups of genes are known to be important for cell growth and elongation, so it is likely that the differential expression of these genes is responsible for the columnar growth phenotype. Genes of the five remaining categories (photosynthesis, mitochondrial electron transport, DNA synthesis, RNA processing and protein synthesis) were assumed to be regulated as a consequence of the impairment of normal growth. As this is the first study approaching the transcriptome of columnar apple and our results are extensive, further approaches are necessary to discover the molecular function(s) of Co.

Authors' contributions

CK performed the sample preparation for 454 and Illumina sequencing, created the pipeline for NGS data analysis, analyzed the data, supervised the work of RP and wrote the paper. RP contributed to the sample preparation, the bioinformatics analysis, literature inquiries and the first draft of the manuscript. ERS initiated the project and was responsible for supervision and budget. He also contributed to the preparation of the manuscript.

Acknowledgments

We are grateful to Prof. Peter Braun (Department of Pomology, Geisenheim Research Center) and Bastienne Brauksiepe (Department of Botany, Geisenheim Research Center) for providing the apple plant material grown by the Department of Pomology. The help of Benjamin Rieger in the development and programming of Perl scripts is highly acknowledged. The work was supported by grant of the Federal Ministry of Agriculture and Nutrition (Nr. 511-06.01-28-1-43.042-07).

References

Alagna, F., D'Agostino, N., Torchia, L., Servili, M., Rao, R., Pietrella, M., Giuliano, G., Chiusano, M.L., Baldoni, L. and Perrotta, G. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* **10** (2009), p. 399.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **215** (1990), pp. 403-10.

Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., Horsman, D.E., Connors, J.M., Gascoyne, R.D., Marra, M.A. and Jones, S.J. De novo transcriptome assembly with ABySS. *Bioinformatics* **25** (2009), pp. 2872-7.

Blažek, J. Segregation and general evaluation of spur type or compact growth habits in apples. *Acta Hortic.* **317** (1992), pp. 71-79.

Cheng, N.H., Pittman, J.K., Shigaki, T., Lachmansingh, J., LeClere, S., Lahner, B., Salt, D.E. and Hirschi, K.D. Functional association of Arabidopsis CAX1 and CAX3 is required for normal growth and ion homeostasis. *Plant Physiol* **138** (2005), pp. 2048-60.

Fisher, D.V. Spur-type strains of McIntosh for high density plantings. *British Columbia Fruit Growers' Association Quarterly Report* **14** (1969), pp. 3-10.

Flachowsky, H., Hattasch, C., Hofer, M., Peil, A. and Hanke, M.V. Overexpression of LEAFY in apple leads to a columnar phenotype with shorter internodes. *Planta* (2009).

Galinha, C., Bilsborough, G. and Tsiantis, M. Hormonal input in plant meristems: A balancing act. *Semin Cell Dev Biol* **20** (2009), pp. 1149-56.

Hartig, K. and Beck, E. Crosstalk between auxin, cytokinins, and sugars in the plant cell cycle. *Plant Biol (Stuttg)* **8** (2006), pp. 389-96.

Huber, D.P., Philippe, R.N., Madilao, L.L., Sturrock, R.N. and Bohlmann, J. Changes in anatomy and terpene chemistry in roots of Douglas-fir seedlings following treatment with methyl jasmonate. *Tree Physiol* **25** (2005), pp. 1075-83.

Hyodo, H., Yamakawa, S., Takeda, Y., Tsuduki, M., Yokota, A., Nishitani, K. and Kohchi, T. Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, XTH9, in inflorescence apices is related to cell elongation in *Arabidopsis thaliana*. *Plant Mol Biol* **52** (2003), pp. 473-82.

Kuppusamy, K.T., Walcher, C.L. and Nemhauser, J.L. Cross-regulatory mechanisms in hormone signaling. *Plant Mol Biol* **69** (2009), pp. 375-81.

Lapins, K.O. Segregation of compact growth types in certain apple seedling progenies. *Canad J Plant Sci* **49** (1969), pp. 765-768.

Lapins, K.O. and Watkins, R. Genetics of compact growth habit. *Report of East Malling Research Station for 1972* **136** (1973).

Liu, Y., Liu, D., Zhang, H., Gao, H., Guo, X., Wang, D., Zhang, X. and Zhang, A. The alpha- and beta-expansin and xyloglucan endotransglucosylase/hydrolase gene families of wheat: molecular cloning, gene expression, and EST data mining. *Genomics* **90** (2007a), pp. 516-29.

Liu, Y.B., Lu, S.M., Zhang, J.F., Liu, S. and Lu, Y.T. A xyloglucan endotransglucosylase/hydrolase involves in growth of primary root and alters the deposition of cellulose in Arabidopsis. *Planta* **226** (2007b), pp. 1547-60.

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

Looney, N.E., Taylor, J.S. and Pharis, R.P. Relationship of endogenous gibberellin and cytokinin levels in shoot tips to apical form in four strains of 'McIntosh' apple. *J Amer Soc Hort Sci* **113** (1988), pp. 395-398.

Morozova, O., Hirst, M. and Marra, M.A. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet* **10** (2009), pp. 135-51.

Morris, J., Tian, H., Park, S., Sreevidya, C.S., Ward, J.M. and Hirschi, K.D. AtCCX3 is an Arabidopsis endomembrane H+ -dependent K+ transporter. *Plant Physiol* **148** (2008), pp. 1474-86.

Munne-Bosch, S. and Falk, J. New insights into the function of tocopherols in plants. *Planta* **218** (2004), pp. 323-6.

Munne-Bosch, S., Weiler, E.W., Alegre, L., Muller, M., Duchting, P. and Falk, J. Alpha-tocopherol may influence cellular signaling by modulating jasmonic acid levels in plants. *Planta* **225** (2007), pp. 681-91.

Ohlrogge, J. and Browse, J. Lipid biosynthesis. *Plant Cell* **7** (1995), pp. 957-70.

Pauwels, L., Morreel, K., De Witte, E., Lammertyn, F., Van Montagu, M., Boerjan, W., Inze, D. and Goossens, A. Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured Arabidopsis cells. *Proc Natl Acad Sci U S A* **105** (2008), pp. 1380-5.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J. and Birol, I. ABySS: a parallel assembler for short read sequence data. *Genome Res* **19** (2009), pp. 1117-23.

Stekel, D.J., Git, Y. and Falciani, F. The comparison of gene expression from multiple cDNA libraries. *Genome Res* **10** (2000), pp. 2055-61.

Sun, C., Li, Y., Wu, Q., Luo, H., Sun, Y., Song, J., Lui, E.M. and Chen, S. De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics* **11** (2010), p. 262.

Surget-Groba, Y. and Montoya-Burgos, J.I. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res* **20** (2010), pp. 1432-40.

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., Selbig, J., Muller, L.A., Rhee, S.Y. and Stitt, M. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* **37** (2004), pp. 914-39.

Tian, Y.K., Wang, C.H., Dai, H.Y. and Zhang, J.S. Screening of a RAPD marker tightly linked to *Co* gene in apple and the SCAR marker conversion. *Yi Chuan Xue Bao* **31** (2004), pp. 919-25.

Tian, Y.K., Wang, C.H., Zhang, J.S., James, C. and Dai, H.Y. Mapping *Co*, a gene controlling the columnar phenotype of apple, with molecular markers. *Euphytica* **145** (2005), pp. 181–188.

Tobutt, K.R. Breeding columnar apples at east malling. *Acta Hortic.* **159** (1985), pp. 63-68.

Tromas, A. and Perrot-Rechenmann, C. Recent progress in auxin biology. *C R Biol* **333** (2010), pp. 297-306.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313** (2006), pp. 1596-604.

UniProt-Consortium The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res* **38** (2010), pp. D142-8.

van Schie, C.C., Haring, M.A. and Schuurink, R.C. Tomato linalool synthase is induced in trichomes by jasmonic acid. *Plant Mol Biol* **64** (2007), pp. 251-63.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M., Pruss, D., Salvi, S., Pindo, M., Baldi, P., Castelletti, S., Cavaiuolo, M., Coppola, G., Costa, F., Cova, V., Dal Ri, A., Goremykin, V., Komjanc, M., Longhi, S., Magnago, P., Malacarne, G., Malnoy, M., Micheletti, D., Moretto, M., Perazzolli, M., Si-Ammour, A., Vezzulli, S., Zini, E., Eldredge, G., Fitzgerald, L.M., Gutin, N., Lanchbury, J., Macalma, T., Mitchell, J.T., Reid, J., Wardell, B., Kodira, C., Chen, Z., Desany, B., Niazi, F., Palmer, M., Koepke, T., Jiwan, D., Schaeffer, S., Krishnan, V., Wu, C., Chu, V.T., King, S.T., Vick, J., Tao, Q., Mraz, A., Stormo, A., Stormo, K., Bogden, R., Ederle, D., Stella, A., Vecchietti, A., Kater, M.M., Masiero, S., Lasserre, P., Lespinasse, Y., Allan, A.C., Bus, V., Chagne, D., Crowhurst, R.N., Gleave, A.P., Lavezzo, E., Fawcett, J.A., Proost, S., Rouze, P., Sterck, L., Toppo, S., Lazzari, B., Hellens, R.P., Durel, C.E., Gutin, A., Bumgarner, R.E., Gardiner, S.E., Skolnick, M., Egholm, M., Van de Peer, Y., Salamini, F. and Viola, R. The genome of the domesticated apple (*Malus* x *domestica* Borkh.). *Nat Genet* **42** (2010), pp. 833-9.

Wall, P.K., Leebens-Mack, J., Chanderbali, A.S., Barakat, A., Wolcott, E., Liang, H., Landherr, L., Tomsho, L.P., Hu, Y., Carlson, J.E., Ma, H., Schuster, S.C., Soltis, D.E., Soltis, P.S., Altman, N. and dePamphilis, C.W. Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics* **10** (2009), p. 347.

Wang, W., Wang, Y., Zhang, Q., Qi, Y. and Guo, D. Global characterization of *Artemisia annua* glandular trichome transcriptome using 454 pyrosequencing. *BMC Genomics* **10** (2009), p. 465.

Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., Chen, X. and Li, Y. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (Ipomoea batatas). *BMC Genomics* **11** (2010), p. 726.

Wasternack, C. Jasmonates: an update on biosynthesis, signal transduction and action in plant stress response, growth and development. *Ann Bot* **100** (2007), pp. 681-97.

Watanabe, M., Bessho, H., Suzuki, A. and Komori, S. Seasonal Changes of IAA and Cytokinin in Shoots of Columnar Type Apple Trees. *Acta Hortic.* **774** (2008), pp. 75-80.

Watanabe, M., Suzuki, A., Komori, S. and Bessho, H. Comparison of Endogenous IAA and Cytokinins in Shoots of Columnar and Normal Type Apple Trees. *J. Japan. Soc. Hort. Sci.* **73** (2004), pp. 19-24.
Wu, G.Z. and Xue, H.W. Arabidopsis {beta}-Ketoacyl-[Acyl Carrier Protein] Synthase I Is Crucial for Fatty Acid Synthesis and Plays a Role in Chloroplast Division and Embryo Development. *Plant Cell* **22** (2010), pp. 3726-44.

Additional Files

These files are included in the electronic supplementary material accompanying this thesis.

Additional File 1 - Perl script BORT1.3
Source code of in-house Perl script BORT1.3 that is able to process large BLAST output files in a user-defined manner.

Additional File 2 - Perl script BORT2.1
Source code of in-house Perl script BORT2.1 that is able to process large BLAST output files in a user-defined manner and independent of local RAM.

Additional File 4 - Complete unigene list obtained from NGen assembled 454 sequencing data
The table shows all UniProtKB entries obtained from BLAST similarity searches of NGen assembled contigs and singletons after processing using BORT1.3. Absolute numbers of reads for each gene and both P28 and A14 are included. Genes are sorted by R_j values corresponding to the level of significant differential regulation.

Additional File 5 - Complete unigene list obtained from ABySS assembled Illumina sequencing data.
The table shows all UniProtKB entries obtained from BLAST similarity searches of ABySS assembled contigs after processing using BORT2.1. Absolute numbers of k-mers for each gene and both P28 and A14 are included. Genes are sorted by R_j values corresponding to the level of significant differential regulation.

Additional File 7 - Complete unigene list obtained from 454 raw sequencing data processing.
The table shows all entries (corresponding UniProtKB description) obtained from BLAST similarity searches of 454 raw sequencing data against *Malus x domestica* proteins (MDPs) after processing using BORT2.1. Absolute numbers of reads for each gene and both P28 and A14 are included. Genes are sorted by R_j values corresponding to the level of significant differential regulation.

Additional File 8 - Complete unigene list obtained from Illumina raw sequencing data processing.
The table shows all entries (corresponding UniProtKB description) obtained from BLAST similarity searches of Illumina raw sequencing data against *Malus x domestica* proteins (MDPs) after processing using BORT2.1. Absolute numbers of reads for each gene and both P28 and A14 are included. Genes are sorted by R_j values corresponding to the level of significant differential regulation.

Paper 5

Deep Sequencing of the Shoot Apical Meristem Transcriptome of Columnar Apple Trees (*Malus* x *domestica*)

Clemens Krost[1], Romina Petersen[1], Peter Braun[2], Erwin R. Schmidt[1]

[1] Department of Molecular Genetics, Johannes Gutenberg-University of Mainz
[2] Department of Pomology, Geisenheim University

This article will be published in the BiotechFruit2012 special issue of Acta Horticulturae in 2014 and has been included in this thesis with permission.

Keywords

apple, columnar, NGS, RNA-seq, growth factors

Abstract

Columnar apple trees (*Malus* x *domestica*) are an interesting research object for plant breeders as well as molecular geneticists because of their striking growth habit and the fact that it is caused by the presence of a single dominant allele of the *Co* gene. The phenotype is characterized by nearly absent lateral branching and an unusual thick stem. Since almost nothing is known about the molecular function(s) of this gene, we performed Illumina RNA-seq next generation sequencing to compare the transcriptomes of columnar and non-columnar shoot apical meristems (SAMs). With this approach, we found hundreds of genes to be differentially regulated, even at different sample collection dates. Functional assignment using gene ontology (GO) tools revealed that many of these genes can potentially be associated with the columnar phenotype. There is noticeable evidence that the expression levels of many genes involved in phytohormone synthesis and signaling are changed. Thus, we hypothesize that columnar growth is in part a consequence of a disregulated phytohormone network.

Introduction

Columnar growth in apple (*Malus* x *domestica*) was initially discovered in the early 1960s as a somaclonal variation of 'McIntosh' (Fisher, 1969). The columnar phenotype is mainly characterized by short internodes and transformation of branches into fruit spurs (Tobutt, 1985). Hence, columnar apple trees provide several economic advantages, including

suitability for high-density orchards. The phenotype is caused by the presence of a single dominant allele of the *Columnar* (*Co*) gene (Lapins, 1969; Lapins and Watkins, 1973; Blažek, 1992), located on chromosome 10, which appears predominantly in a heterozygous state (Tian et al., 2005). The molecular basis of this mutant phenotype is largely unknown, but it seems to be obvious that phytohormones play a significant role. It has been demonstrated that concentrations of gibberellic acid (GA), cytokinin (CK), indole-3-acetic acid (IAA) and abscisic acid (ABA) are different in shoots of columnar apple trees compared with trees showing standard growth habit (Looney and Lane, 1984; Looney et al., 1988; Watanabe et al., 2004; Watanabe et al., 2008).

The aim of our work was to identify genes which are differentially regulated in shoot apical meristems (SAMs) of columnar trees using comparative whole transcriptome analysis (Krost et al., 2012; Zhang et al., 2012). Genes associated with growth regulation, which might provide special insights how the phenotype is realized, are of particular interest.

Materials and Methods

Shoot tips of two apple cultivars, columnar 'Procats 28' (P28) and non-columnar 'A14-190-93' (A14) were sampled in late May 2009, late July 2010 and late September 2009. Shoot meristem tissue was collected from shoot tips of lateral shoots/spurs.

Shoot tips were processed to mainly obtain meristems by removing surrounding tissues. The remaining material ("SAM") was instantly frozen and disrupted in liquid nitrogen. RNA isolation was performed using the innuPREP Plant RNA Kit (Analytik Jena, Jena, Germany) and 100 µg of total RNA was poly-A^+ purified using the NucleoTrap mRNA Kit (Macherey-Nagel, Düren, Germany). Linear amplification of mRNAs was included using MessageAmp II aRNA Amplification Kit (Ambion, Austin, USA). The quality of total RNAs, mRNAs and amplified aRNAs (antisense RNA) was tested on a Bioanalyzer (Agilent, Santa Clara, USA).

5 µg of each aRNA was used for all sequencing library constructions. Two samples were sequenced on a Genome Analyzer IIx (Cologne Center for Genomics, Cologne, Germany) and four samples were run on an Illumina HiSeq 2000 (Institute for Molecular Genetics, Mainz, Germany) using 2 x 100 bp paired end sequencing mode. All sequences can be obtained from EBI Sequence Read Archive accession ERP000629.

Results and Discussion

Using Next Generation Sequencing (NGS), we compared shoot apical meristem (SAM) transcriptomes of columnar (P28) and non-columnar (A14) apple trees. RNAs from three different sample collection dates throughout the growth period were used for sequencing. Table 1 summarizes the number of reads obtained from six sequencing runs that were used for our analysis. After adequate filtering, processing and quantification the data were fed into the DESeq software (Anders and Huber, 2010). DESeq analysis resulted in the identification of more than 600 differentially regulated genes ($p < 0.005$, fold change > 3) in the columnar cultivar independent of collection date. This number is significantly smaller than the 5,237 genes found to be differentially expressed in columnar trees by Zhang et al.

(2012), which is mainly due to differences in the material investigated (as Zhang et al. (2012) examined new shoots instead of shoot tips of lateral spurs) and in the parameters applied to define differential expression (fold change > 2 and false discovery rate < 0.001 in the study of Zhang et al. (2012)). Nevertheless the differential regulation of more than 600 genes shows the massive impact of the *Co* gene on the general gene expression pattern in the SAM, which is not too surprising in such a severe growth mutant phenotype. All genes were subjected to gene ontology (GO) classification resulting in seven overrepresented functional categories ($p < 0.05$): (1) biotic stress, (2) protein modification and degradation, (3) cell wall degradation, (4) amino acid metabolism, (5) regulation of transcription, (6) mitochondrial electron transport and (7) transport proteins. Many of these differentially regulated genes can be directly linked with the columnar phenotype, mainly based on changes in membrane and cell wall composition (Krost et al., 2012). This could be a possible explanation for the thicker stem and shorter internodes, two conspicuous phenotypic characteristics of columnar compared to normal type apple trees.

Table 1. Next Generation Sequencing results. The table shows the number of reads obtained from each of the six cDNA libraries using Illumina GA IIx and Illumina HiSeq 2000 platforms.

Collection date	Sequencing platform	Reads 'A14-190-93'	Reads 'Procats28'
May 2009	Illumina GA IIx	80,289,084	-
	Illumina HiSeq 2000	44,134,596	67,515,702
July 2010	Illumina HiSeq 2000	58,616,798	162,540,416
September 2009	Illumina GA IIx	-	76,177,216
		183,040,478	**306,233,334**

Furthermore, we explicitly analyzed genes which are associated with phytohormone synthesis, transport and signal transduction, since these are known to be directly involved in growth regulation. Summarizing our findings, we detected a number of important genes that are affected in their regulation by the presence of *Co*, which probably affect hormone concentrations such as IAA, CK, GA, jasmonic acid and brassinosteroids (Fig. 1). Concerning IAA, CK and GA, these results are in line with earlier findings on the hormonal state of shoot tips of columnar apples (Lee and Looney, 1977; Looney and Lane, 1984; Watanabe et al., 2008). Unfortunately, we were not able to decide on the basis of our results whether the *Co* gene or the *Co* gene product directly affects the genes involved in phytohormone synthesis, or whether it is an indirect effect.

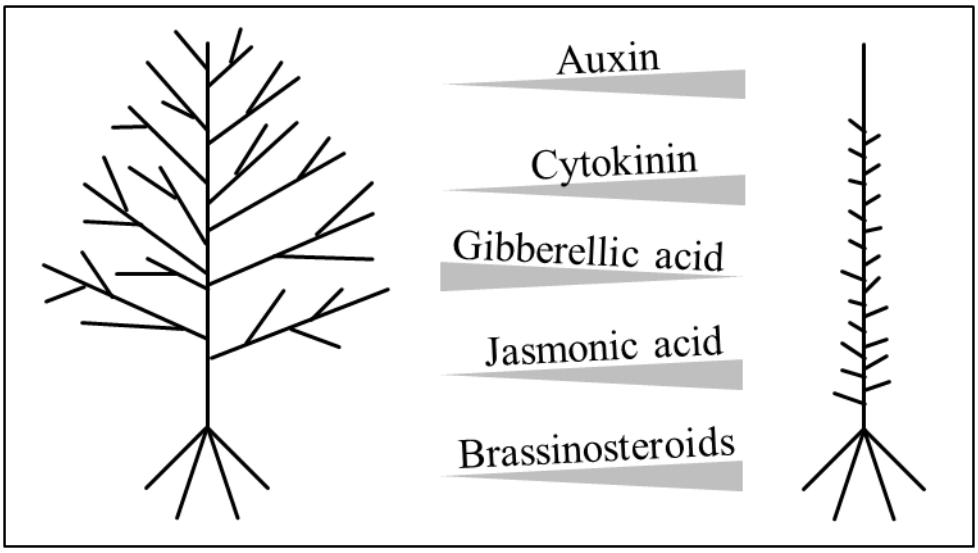


Fig. 1. Putative altered hormone concentrations due to the action of the *Co* gene in columnar apple trees (right).

Conclusions

Studying transcriptomes with NGS methods is an efficient way to get comprehensive insights into uncharacterized tissues of certain phenotypes without prior knowledge of their genetics. The presence of a single dominant gene *Co* in columnar apple trees was shown to lead to massive changes in the gene expression pattern in SAMs, a tissue that is crucial for the final shape of the plant. Even though the molecular nature of the molecular gene function remains unresolved, it is clear that some of the directly or indirectly affected genes may explain the emergence of this interesting phenotype.

Acknowledgements

We are grateful to Benjamin Rieger for the development and programming of Perl scripts. The work was supported by grant of the Federal Ministry of Agriculture and Nutrition (Nr. 511-06.01-28-1-43.042-07).

Literature Cited

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

Blažek, J. Segregation and general evaluation of spur type or compact growth habits in apples. 1992. Acta Hortic. 317: 71-79.

Fisher, D.V. 1969. Spur-type strains of McIntosh for high density plantings. British Columbia Fruit Growers' Association Quarterly Report 14: 3-10.

Krost, C., Petersen, R. and Schmidt, E.R. 2012. The transcriptomes of columnar and standard type apple trees (*Malus x domestica*) - a comparative study. Gene 498: 223-30.

Lapins, K.O. 1969. Segregation of compact growth types in certain apple seedling progenies. Canad. J. Plant Sci. 49: 765-768.

Lapins, K.O. and Watkins, R. 1973. Genetics of compact growth habit. Report of East Malling Research Station for 1972 136

Lee, J.M. and Looney, N.E. 1977. Abscisic acid levels and genetic compaction in apple seedlings. Canad. J. Plant Sci. 57: 81-85.

Looney, N.E. and Lane, W.D. 1984. Spur-type growth mutants of McIntosh apple: a review of their genetics, physiology and field performance. Acta Hort. 146: 31-46.

Looney, N.E., Taylor, J.S. and Pharis, R.P. 1988. Relationship of endogenous gibberellin and cytokinin levels in shoot tips to apical form in four strains of 'McIntosh' apple. J. Amer. Soc. Hort. Sci. 113: 395-398.

Tian, Y.K., Wang, C.H., Zhang, J.S., James, C. and Dai, H.Y. 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 Hort. 159: 63-68.

Watanabe, M., Bessho, H., Suzuki, A. and Komori, S. 2008. Seasonal changes of IAA and cytokinin in shoots of columnar type apple trees. Acta Hort. 774: 75-80.

Watanabe, M., Suzuki, A., Komori, S. and Bessho, H. 2004. Comparison of endogenous IAA and cytokinins in shoots of columnar and normal type apple trees. J. Japan. Soc. Hort. Sci. 73: 19-24.

Zhang, Y., Zhu, J. and Dai, H. 2012. Characterization of transcriptional differences between columnar and standard apple trees using RNA-Seq. Plant Mol. Biol. Rep. 30: 957-965.

Paper 6

Evaluation of the Hormonal State of Columnar Apple Trees (*Malus x domestica*) Based on High Throughput Gene Expression Studies

Clemens Krost[1], Romina Petersen[1], Stefanie Lokan[1], Bastienne Brauksiepe[2], Peter Braun[3], Erwin R. Schmidt[1]

[1] Department of Molecular Genetics, Johannes Gutenberg-University of Mainz
[2] Department of Botany, Geisenheim University
[3] Department of Pomology, Geisenheim University

This paper was published as a research article in Plant Molecular Biology in 2013 (*Plant Mol Biol* **81**(3): 211-220, with three figures, two tables and two supplementary files)

Abstract

The columnar phenotype of apple trees (*Malus x domestica*) is characterized by a compact growth habit with fruit spurs instead of lateral branches. These properties provide significant economic advantages by enabling high density plantings. The columnar growth results from the presence of a dominant allele of the gene *Columnar* (*Co*) located on chromosome 10 which can appear in a heterozygous (*Co*/*co*) or homozygous (*Co*/*Co*) state. Although two deep sequencing approaches could shed some light on the transcriptome of columnar shoot apical meristems (SAMs), the molecular mechanisms of columnar growth are not yet elaborated. Since the influence of phytohormones is believed to have a pivotal role in the establishment of the phenotype, we performed RNA-Seq experiments to study genes associated with hormone homeostasis and clearly affected by the presence of *Co*. Our results provide a molecular explanation for earlier findings on the hormonal state of columnar apple trees. Additionally, they allow hypotheses on how the columnar phenotype might develop. Furthermore, we show a statistically approved enrichment of differentially regulated genes on chromosome 10 in the course of validating RNA-Seq results using additional gene expression studies.

Keywords

Columnar, Apple, Next generation sequencing RNA-Seq, Phytohormones, Selective sweep

Abbreviations

Co – *Columnar*; NGS – next generation sequencing; SAM – shoot apical meristem; IAA – indole-3-acetic acid; CK – Cytokinin; ABA – abscisic acid; GA – gibberellic acid; JA – jasmonic acid; A14 – A14-190-93 (non-columnar type apple); P28 – Procats 28 (columnar type apple); BR – brassinosteroid

Introduction

Columnar growth in apple (*Malus x domestica*) was discovered as a somatoclonal variation of cv. McIntosh (Fisher 1969) in the early 1960s. The columnar habit represents an extreme form of spur type growth (Blažek 1992, Eaton and Lapins 1970) and is characterized by short internodes, thick stems and nearly absent branching. The phenotype is caused by the presence of a single dominant gene *Co* (Lapins 1969, Lapins and Watkins 1973, Looney and Lane 1984) located on chromosome 10 that can either occur in a heterozygous (Tian et al. 2005) or homozygous state (P. Braun, pers. communication). The main economic advantage of columnar varieties is the possibility of high density plantings resulting in increased yields per hectare (Jacob 2010). To date, little is known about the molecular characteristics of this growth mutation and the function of the *Co* gene product still remains a mystery. Two whole transcriptome approaches were performed to identify differentially regulated genes between columnar and non-columnar cultivars (Krost et al. 2012, Zhang et al. 2012). Even though these studies could identify a large number of growth associated genes to be differentially regulated, a potential candidate gene for *Co* was not identified. In Zhang et al. (2012), 287 out of 5,237 differently regulated genes were found to be related to plant architecture, 60 % being involved in branch formation, 20 % in plant height and 15 % in plant architecture formation. Out of these 287 genes, 31 were localized on chromosome 10; however, no specific candidate gene was identified. In contrast to this broad-scale attempt, Krost et al. (2012) focused on highly expressed genes with strong differential regulation and found eight functional categories to be overrepresented. Especially genes of groups associated with cell membrane and cell wall modification were brought in line with the columnar phenotype. Uniformly, both studies identified differently regulated genes associated with IAA (indole-3-acetic acid) and GA (gibberellic acid) phytohormone regulation, consistent with earlier studies dealing with direct measurements of phytohormone concentrations (Watanabe et al. 2008, Watanabe et al. 2006). Furthermore, BR (Brassinosteroid) associated genes were found to be affected in Zhang et al. (2012) while JA (Jasmonate) associated genes were up regulated in Krost et al. (2012) Therefore, the influence of modified phytohormone concentrations on the establishment of the phenotype seems to be obvious. Several earlier studies showed higher concentrations of free IAA and CK (cytokinin) in shoot tips of columnar compared with normal type trees (Looney and Lane 1984, Watanabe et al. 2008, Watanabe et al. 2004). Furthermore, columnar trees can tolerate higher levels of exogenous CKs before supraoptimal concentrations are reached (Lane and Looney 1982, Sarwar and Skirvin 1997, Sarwar et al. 1998). Endogenous concentrations of ABA (abscisic acid) and polar GAs were lower in columnar cultivars (Lee and Looney 1977, Looney and Lane 1984). Especially because

higher IAA and CK contents were found to be consistent throughout the whole growing period, both hormone levels are thought to be involved in the development of the columnar growth (Watanabe et al. 2008, Watanabe et al. 2006).

In this study, we used RNA-Seq (Garber et al. 2011, Wang et al. 2009) to investigate the hormonal state of SAM (shoot apical meristem) of columnar trees analyzing the regulation of genes mainly associated with phytohormone biosynthesis, signal transduction and transport. This expression profile should give some hints for the understanding of the molecular processes linked to columnar growth and give rise to new insights into the hormonal interplay leading to columnar specific growth habit. We used microarrays and qPCR (quantitative real-time PCR) to validate the results obtained from high throughput sequencing.

Methods

Plant material

For all Illumina sequencing approaches, shoot tips of columnar apple cv. 'Procats 28' (P28) and non-columnar cv. 'A14-190-93' (A14) were sampled in May 2009 (22nd), July 2010 (20th) and September 2009 (29th). For microarray analysis, shoot tips of both cultivars were collected in April 2010 (6th and 20th), May 2010 (4th) and May 2009 (22nd). For qPCR experiments, *in vitro* cultures of A14 and P28 were used. *In vitro* cultures were grown on MS Medium (Murashige and Skoog 1962) containing sorbitol (30 g/L), gluconic acid (1,3 g/L), 6-Benzylaminopurine (1 mg/L) and Indole-3-butyric acid (0,1 mg/L). The plant material was provided by the Departments of Pomology and Botany, Geisenheim Research Center.

Sample preparation

Shoot tips were roughly freed from all surrounding tissues to mainly obtain SAMs. Callus tissue was removed from *in vitro* cultures before preparation. The plant material was instantly frozen using liquid nitrogen and disrupted in a precooled mortar and pestle. RNA isolation was performed using innuPREP Plant RNA Kit (Analytik Jena, Jena, Germany). For Illumina sequencing, 100 µg of total RNA was poly-A+ purified using NucleoTrap mRNA Kit (Macherey-Nagel, Düren, Germany). After purification, mRNA was amplified using MessageAmp II aRNA Amplification Kit (Ambion, Austin, USA). The quality of total RNAs, mRNAs and amplified aRNAs (antisense RNA) was tested on a Bioanalyzer (Agilent, Santa Clara, USA). For microarrays and qPCR, only total RNAs with RIN (RNA integrity number) > 8.0 were used.

Sequencing

For each cultivar and collection date, about 5 µg of aRNA were subjected to Illumina library construction. Two libraries were sequenced on two separate lanes of a Genome Analyzer IIx (Cologne Center for Genomics, Cologne, Germany) while four libraries were multiplexed and sequenced on two lanes of an Illumina HiSeq2000™ (Institute for Molecular Genetics, Mainz, Germany). Both run modes were 2 x 100 bp paired end sequencing.

Microarrays

High quality RNA of four collection dates throughout the growth period was used for microarrays. Two 50 bp oligonucleotides each for 1500 differentially regulated genes obtained from provisional studies were used as probes. Biochip synthesis, sample labeling, hybridization and detection was performed by Febit Holding GmbH (Heidelberg, Germany).

qPCR

Reverse Transcription of high quality RNA from A14 and P28 *in vitro* cultures was carried out using SuperScript III Reverse Transcriptase (Invitrogen, Darmstadt, Germany) and 50 µM oligo(dT)$_{20}$ Primer (Fermentas, St. Leon-Roth, Germany). As an exogenous control, 100 ng of human MDA-MB468 RNA was added to every 800 ng of A14 or P28 RNA per reaction. All qPCR experiments were performed in a 7500 Fast Real-Time PCR System using Power SYBR Green PCR Master Mix (Applied Biosystems, Carlsbad, USA). For each sample, reactions were performed in triplicates using relative quantification. Target gene expression was normalized to human GAPDH (exogenous control). For each gene, at least three biological and two technical replicates were measured.

Bioinformatics

The obtained sequence information (QSEQ files) was converted into FASTA format using in-house Perl scripts and all reads with chastity = 0 were removed. For all libraries, highly parallel BLASTn searches (Altschul 1990) against NCBI *Malus x domestica* UniGene database were performed and read counts were scored using an E-value threshold of 10^{-10}. A14 and P28 libraries were marked as reference and sample, respectively, and differential gene expression was analyzed using DESeq (Anders and Huber 2010) with default parameters.

Results

Global comparison of the various transcriptomes

In order to find genes differentially regulated under the influence of Co throughout the whole vegetative growth period, we collected shoot tips of columnar apple cv. Procats 28 (P28) and non-columnar apple cv. A14-190-93 (A14) in May, July and September. The poly-A+ RNA/cDNA extracted from isolated SAMs were sequenced using Illumina GAIIx and Illumina HiSeq2000™ platforms. The raw datasets can be received from EBI Sequence Read Archive accession ERP000629. A total of 489,273,812 reads were obtained from six different libraries and further used for BLASTn similarity searches against the NCBI *Malus x domestica* UniGene database. Only hits with an E-value cutoff of 10^{-10} and lower were used for read counts resulting in 289,052,563 mapped reads (59.1 %). For the comparison of the sequenced transcriptomes, Pearson correlation coefficients were calculated for each pair of libraries (Figure 1A). These values show a strong correlation between all libraries (Pearson coefficients 0.71 - 0.87) except the non-columnar one based on SAM material sampled in July (**A14** *07-2010*, Pearson coefficients 0.33 - 0.71). Even though the correlation between most libraries is strong, there are many genes differentially regulated between columnar and non-columnar cultivar (all collection dates) and between samples of the same cultivar collected at different time points. An exemplary scatterplot analysis is shown in Figure 1B.

In order to extract genes which were differentially regulated independent of collection date, all read counts of non-columnar A14 and columnar P28 libraries were defined as reference and sample, respectively, and entered into DESeq software (Anders and Huber 2010). Using default parameters, a total of 1173 differentially regulated genes with $p < 0.005$ and fold change > 3.0 were obtained (Online Resource 1). To eliminate genes that show a basal expression, we introduced another cutoff (base mean < 150) to finally obtain a total of 619 differentially regulated genes with $p < 0.005$, change fold > 3.0 and base mean > 150. These genes were functionally analyzed using MAPMAN (Thimm et al. 2004) to revise earlier findings on gene regulation in SAMs of columnar trees (Krost et al. 2012). Most of the results on overrepresented categories (BINs) and gene regulation could thereby be affirmed using six instead of three datasets (Online resource 2).

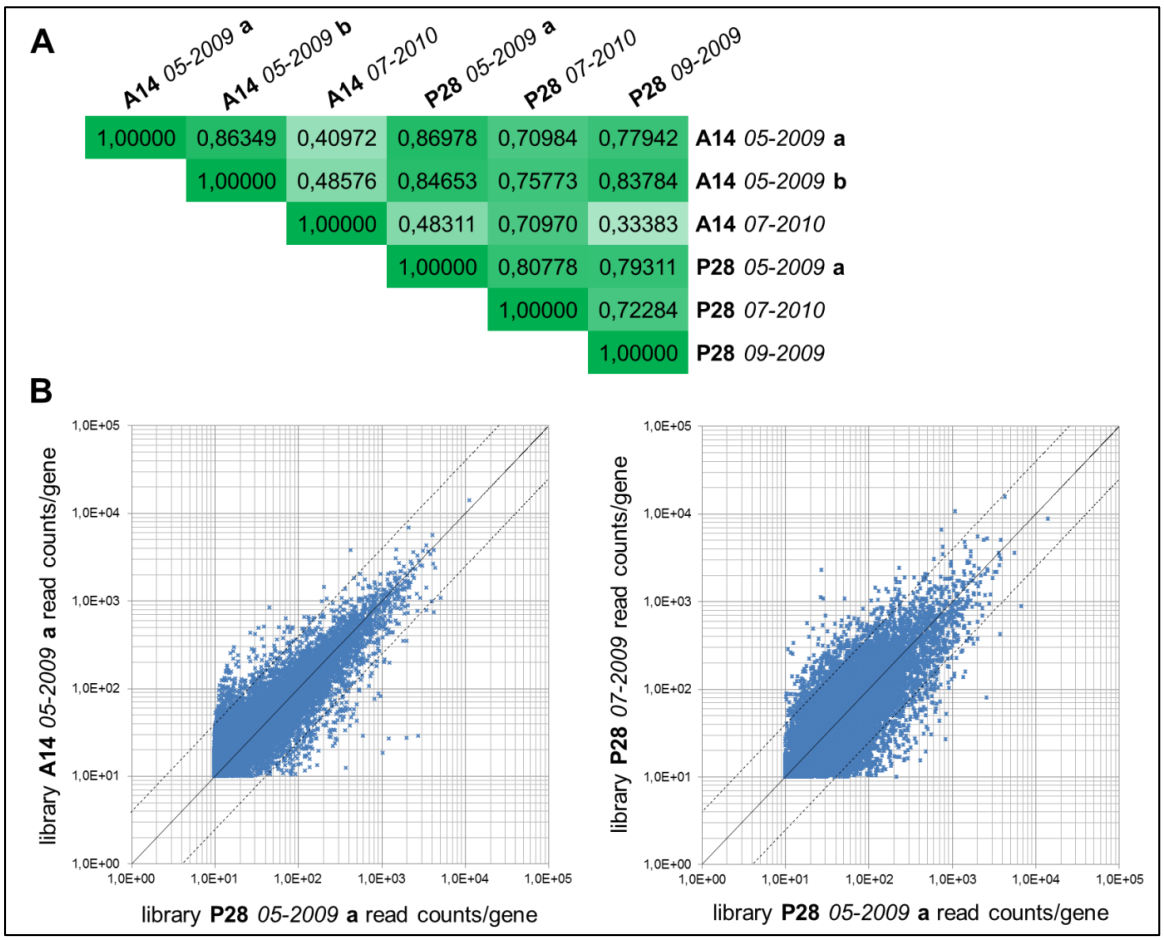
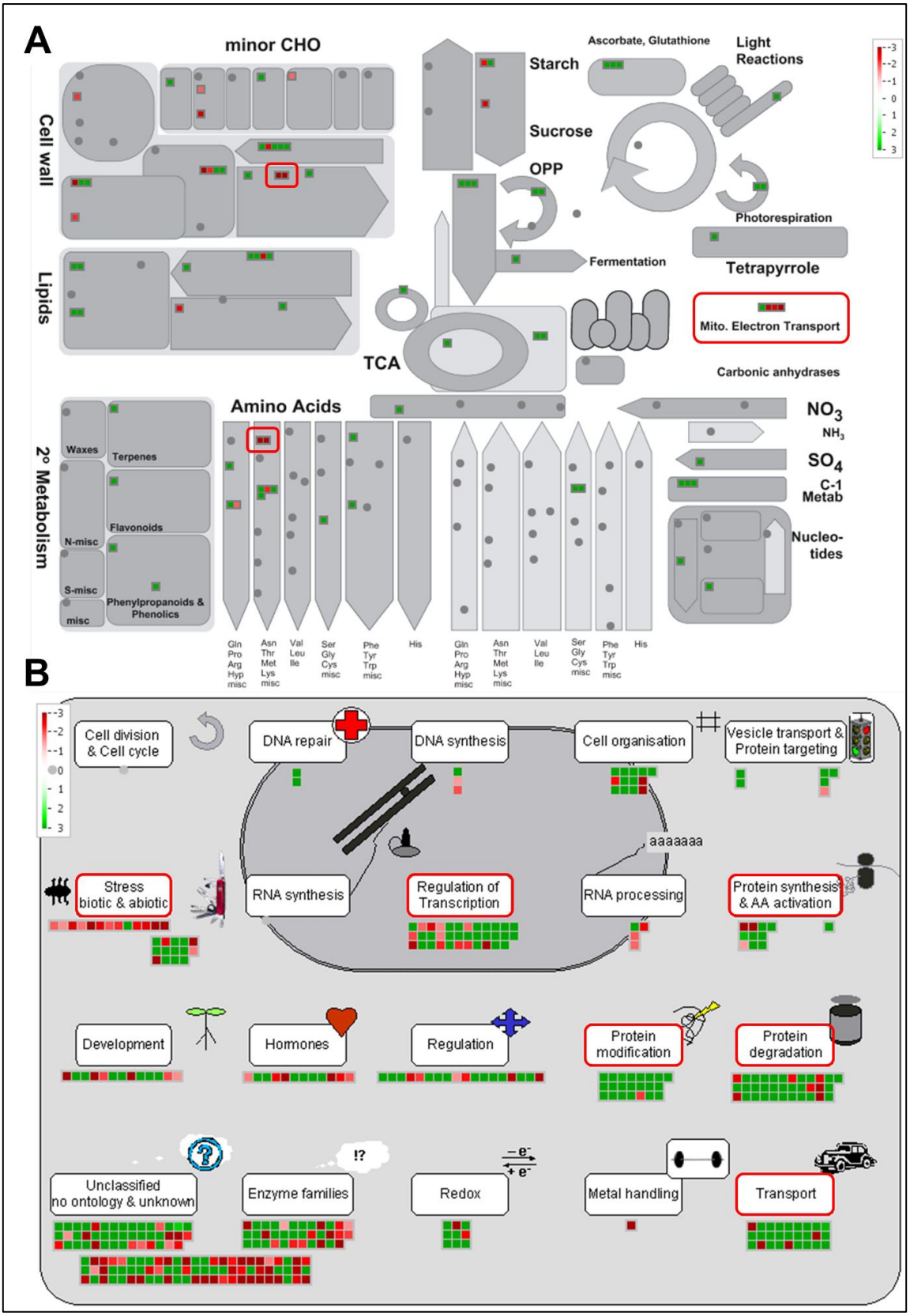


Figure 1 – Comparison of RNA-Seq data
(A) Pearson correlation coefficients for the global analysis of pairs of RNA-Seq libraries. The coefficients are highlighted as shades of dark green (strong correlation) to white (no correlation). Libraries A14 05-2009 a/b were generated using slightly different methods and therefore are no technical replicates.
(B) Scatter plots of RPKM-normalized read counts for the comparison of two RNA-Seq libraries obtained from SAMs of columnar P28 and non-columnar A14 of the same collection date (left) and of P28 libraries based on different collection dates (right). Read counts were adjusted to a minimum of 10 for better illustration. Dashed lines indicate a fold change of 4.



Online Resource 2 – MAPMAN analysis of differentially regulated genes
The figure visualizes all genes found to be differentially expressed by DESeq for metabolic processes (A) and various functional categories (B). Overrepresented categories that confirm earlier findings (Krost et al. 2012) are framed in red. Red and green boxes represent repressed and induced genes, respectively. Values in the key imply $\log_2$ of normalized Illumina read counts

Differentially regulated, phytohormone associated genes

Out of the 619 genes obtained from DESeq, we extracted genes associated with phytohormone biosynthesis, transport and signal transduction, with the goal to reassess earlier published data received from several bioassays on the hormonal status of columnar apple trees (Lee and Looney 1977, Looney and Lane 1984, Watanabe et al. 2008, Watanabe et al. 2004) and to learn more about the underlying molecular networks.

A total of 16 genes were unambiguously identified to be involved in the regulation of IAA (6), CK (3), ABA (3), BRs (2), GAs (1) and JAs (1) by comparison with KEGG database (Kanehisa et al. 2012). These findings are summarized in table 1. The potential functions of these genes are visualized in the context of their biosynthesis, transport or signaling pathways in figure 2. Regarding CK, GA, BR, JA and IAA biosynthesis in P28 SAM (figure 2A-E), a consistent change of expression of the genes mentioned here could have a tremendous effect on the levels of bioactive phytohormones when protein levels correlate with gene expression levels. While genes for cytokinin hydroxylase CYP735A (Takei et al. 2004), squalene monooxygenase (SQLE), delta(24)-sterol reductase DIM/DWF (Takahashi et al. 1995) and jasmonic acid-amido synthetase JAR1 (Suza and Staswick 2008) are up regulated, the genes for gibberellin 2-beta-dioxygenase GA2ox1 (Thomas et al. 1999) and indole-3-acetic acid-amido synthetases GH3.1 and GH3.3 (Staswick et al. 2005) are down regulated. Differential regulation of ABA associated genes was detected for PYL4, coding for an ABA receptor also involved in JA signaling (Lackman et al. 2011, Santiago et al. 2009), as well as for SnRK2.5 and AKIN10 encoding SNF1-related kinases (Kobayashi et al. 2005) (figure 2F).

Additionally, the genes for two main IAA carriers, AUX1 (auxin transporter protein 1) and PIN1 (auxin efflux carrier component 1) (Kleine-Vehn et al. 2006) were found to be up regulated in P28 probably resulting in an increased IAA transport in columnar SAM. Furthermore, the gene for an auxin-binding protein (ABP19a) and thus a potential receptor for IAA (Ohmiya et al. 1998) was found to be consistently down regulated at all collection dates analyzed here. Due to the fact that the genes for GH3.1, GH3.3 and xyloglucan endotransglucosylase/hydrolase protein 20 (XTH20) (Rose et al. 2002) are auxin-responsive and are also down regulated (Figure 2E), a direct regulation by ABP19a seems to be realistic. In spite of this, the down regulation of these genes will most probably lead to a reduced level of bioactive IAA and cell wall growth promoting factor XTH20.

Validation of NGS results using microarrays and qPCR

For the integration of several gene expression studies, we compared our results obtained by NGS to microarrays based on four different collection dates and qPCR based on *in vitro* grown cultures of A14 and P28. Since the microarrays used in our experiments were designed before our NGS datasets were available, they did not contain probes of all genes determined to be differentially regulated by DESeq. Only 105 out of the 619 differentially regulated genes could therefore be measured by the microarray hybridization experiment. RNA from *in vitro* cultures were chosen for qPCR because these are grown under constant conditions and are largely independent of environmental influences.

We identified 16 highly expressed genes whose differential regulation could be validated by at least two of three gene expression studies (table 2). As shown in table 2, five genes mapped to chromosome 10 (where *Co* is located) while the remaining eleven genes seemed to be randomly distributed over the other sixteen chromosomes. An appropriate χ^2 test

statistically validates an enrichment of differentially regulated genes on chromosome 10 with a significance level of 0.025.

Table 1 – Phytohormone associated genes
The table shows all DESeq derived, phytohormone associated genes with Gene ID according to Velasco et al. (2010), E-value and regulation (log_2 change fold). Reg. - Regulation.

Hormone	Gene name	Protein name	Gene ID	E-value	Reg.
IAA	AUX1	Auxin transporter protein 1	MDP0000749280	2×10^{-116}	↑ **4.9**
	PIN1	Auxin efflux carrier component 1	MDP0000138035	6×10^{-51}	↑ **4.8**
	ABP19a	Auxin-binding protein ABP19a	MDP0000397306	1×10^{-112}	↓ **-1.8**
	GH3.3	Indole-3-acetic acid-amido synthetase GH3.3	MDP0000873893	4×10^{-71}	↓ **-2.7**
	GH3.1	Probable indole-3-acetic acid-amido synthetase GH3.1	MDP0000209432	4×10^{-70}	↓ **-2.5**
	XTH20	Xyloglucan endotransglucosylase/ hydrolase protein 20	MDP0000320017	3×10^{-81}	↓ **-2.2**
CK	CYP735A1	Cytokinin hydroxylase	MDP0000909874	6×10^{-95}	↑ **4.9**
	ARR12	Two-component response regulator ARR12	MDP0000140568	3×10^{-45}	↑ **4.6**
	APRR2	Two-component response regulator-like APRR2	MDP0000237396	4×10^{-56}	↑ **5.3**
GA	GA2ox1	Gibberellin 2-beta-dioxygenase 1	MDP0000137705	3×10^{-171}	↓ **-2.5**
ABA	PYL4	Abscisic acid receptor PYL4	MDP0000228470	2×10^{-67}	↓ **-2.4**
	SnRK2.5	Serine/threonine-protein kinase	MDP0000247791	4×10^{-50}	↑ **4.3**
	AKIN10	SNF1-related protein kinase	MDP0000320932	5×10^{-136}	↑ **4.1**
BR	SQLE	Squalene monooxygenase	MDP0000161955	0	↑ **3.4**
	DIM/DWF1	Delta(24)-sterol reductase	MDP0000682675	2×10^{-129}	↑ **5.1**
JA	JAR1	Jasmonic acid-amido synthetase	MDP0000786650	2×10^{-89}	↑ **4.8**

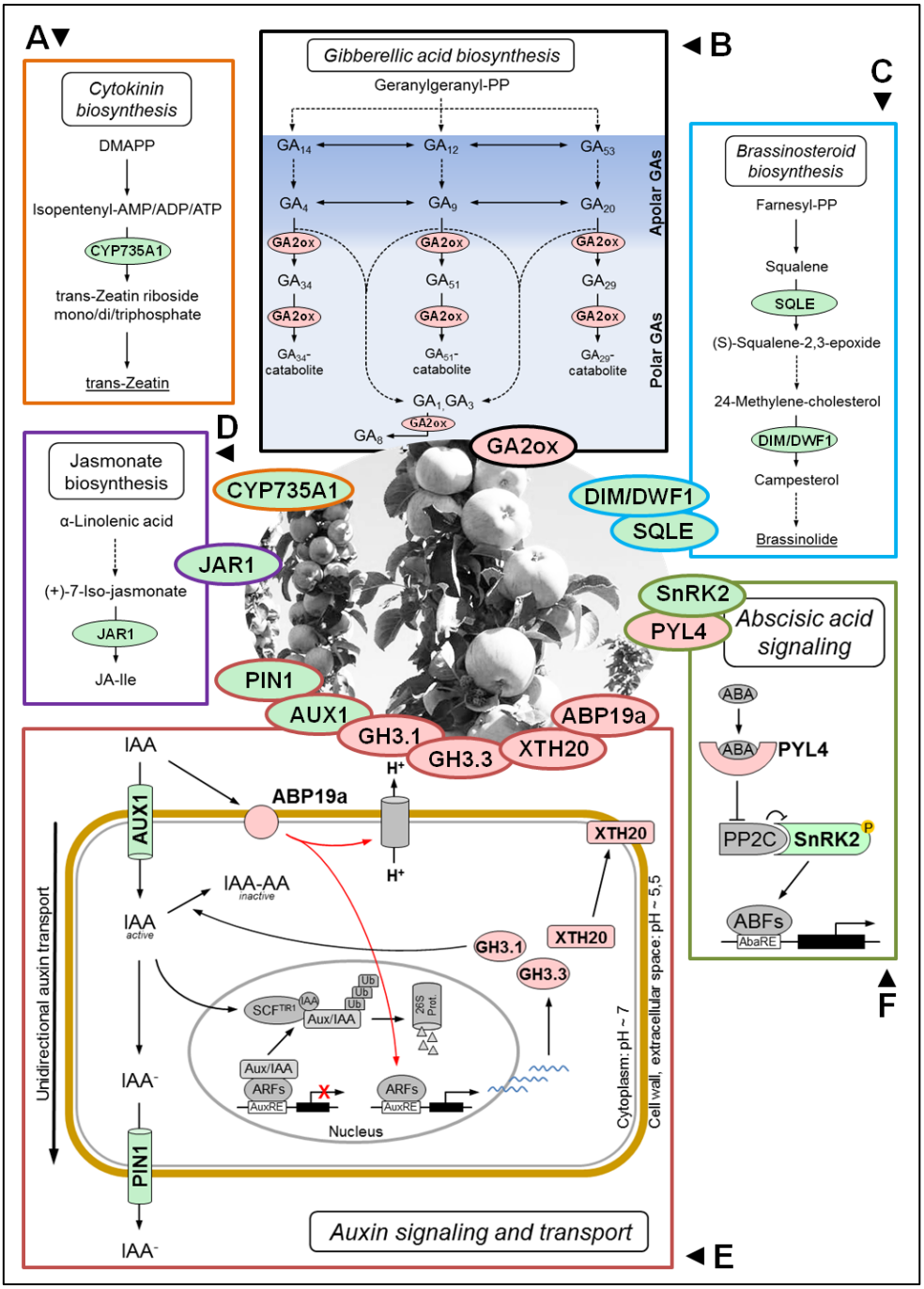


Figure 2 - Hormonal state of columnar SAM

The figure shows all differentially regulated, phytohormone associated genes in SAMs of columnar apple (central image) obtained by DESeq analysis. Identically colored frames indicate the affiliation of genes to the pathway. Up and down regulation in P28 is indicated by light green and light red background colors of genes and gene products, respectively. Black lines/arrows indicate single-step reactions while broken lines indicate multi-step reactions. Red arrows represent hypothetical interactions. (A) Cytokinin (CK) biosynthesis from DMAPP (dimethylallyl pyrophosphate) to trans-Zeatin. (B) Gibberellic acid (GA) biosynthesis from precursor geranylgeranyl pyrophosphate to bioactive (GA_4, GA_9, GA_1, GA_3) and inactive (GA_{34}, GA_{51}, GA_{29}, GA_8) forms. Polarity is indicated by blue shades. (C) Brassinosteroid (BR) biosynthesis from farnesyl pyrophosphate to bioactive Brassinolide (BL). (D) Jasmonate (JA) biosynthesis from precursor α-linolenic acid to bioactive JA-Isoleucin (JA-Ile). (E) Transport and signaling of auxins, represented by IAA which is transported in unidirectional direction based on the asymmetric distribution of influx and efflux carriers AUX1 and PIN1. Free IAA (active) is inactivated by conjugation to amino acids (AA) influenced by GH3 proteins. While free IAA generally activates IAA dependent genes by releasing repressing Aux/IAA protein that are subsequently degraded, down regulated ABP19a could be involved in the regulation of these downstream genes including GH3 and XTH20. (F) Abscisic acid (ABA) signaling cascade.

Discussion

RNA-Seq is an innovative high-throughput method for fast and global comparisons of transcriptomes with respect to gene expression, SNP detection, splice variants etc. Since technologies continually improve to produce hundreds of gigabases of sequence information per run, depth of data analysis is meanwhile limited to computational power and available software. Here, we used highly parallel BLASTn similarity searches for read counting. Out of many tools for the detection of differential expression (Garber et al. 2011, Kvam et al. 2012), we chose DESeq (Anders and Huber 2010) as an R/Bioconductor package because it provides exact test based p-values. With the help of this data pipeline, nearly 300 million Illumina reads out of six libraries were mapped to 23,777 unigenes resulting in 619 differentially regulated genes independent of collection date with statistically approved p-values, change fold and expression values.

Investigating five similar libraries (based on strong Pearson correlation), the only outgroup (non-columnar SAMs collected in July) can be attributed to a growth cessation in summer while columnar trees continue vegetative growth until late autumn (Watanabe et al. 2003, Watanabe et al. 2004), likely due to altered hormone concentrations. The only exception of this library is given by the strong correlation with the columnar SAM library of the same collection date, probably due to comparable environmental conditions.

Hormonal state is strongly affected by the presence of Co

We identified 16 phytohormone synthesis and transport relevant genes with striking differences in gene expression between columnar and non-columnar apple trees throughout the whole growth period. Seven genes code for products (GA2ox1, DIM/DWF1, SQLE, GH3.1, GH3.3, JAR1 and CYP735A1) that are involved in phytohormone anabolism and therefore the differential expression of these genes is assumed to directly affect (increase) the endogenous amounts of bioactive GAs, BRs, IAA, CKs (generally growth promoting) and JAs (growth inhibiting) in SAM tissue. Since CYP735A1, JAR1, SQLE and DIM/DWF1 are positive regulators whereas GA2ox1, GH3.1 and GH3.3 are negative regulators of active hormone concentrations, the regulation observed here probably leads to enhanced levels of active CKs, GAs, BRs, JAs and IAA. As GA2ox1 is necessary for the conversion of apolar GAs into polar GAs (figure 2B), these findings perfectly agree with results of concentration measurements on CKs, polar/apolar GAs (Atzorn 1983) and IAA in shoot tips of columnar apples. Endogenous levels of BRs and JAs have not been analyzed before. The situation is not as clear for ABA which had lower levels in columnar apple (Lee and Looney 1977). Cross-regulatory effects (Depuydt and Hardtke 2011, Gazzarrini and McCourt 2003) have to be taken into account here, such as the stimulatory effect of BRs on the biosynthesis of JAs (Zhang et al. 2009), the interaction of GAs with IAA and ABA (Gou et al. 2010) and the up regulation of GAs by IAA via a reduction of GA2ox levels (Reid et al. 2011). Furthermore, IAA transport and IAA, ABA and CK signaling are also affected. Taken together, these modifications in hormone homeostasis can be used to hypothesize how the columnar phenotype might arise: 1) Decreased levels of ABP19a lead to reduced transcription of early auxin-regulated genes XTH20, GH3.1 and GH3.3 as could similarly be shown in heterozygous ABP1 insertion mutants (Effendi et al. 2011). This, in turn, results in enhanced levels of free active IAA and impairment of cell wall growth due to a lack of XTH20 gene product that could explain the shortened internodes (Hyodo et al. 2003).

2) The expectably dwarfed phenotype because of XTH20 down regulation (Hyodo et al. 2003, Romo et al. 2005), which cannot be seen in columnar trees, is overcome by higher concentrations of growth promoting hormones like IAA, CKs and GAs. Solely, decreased GA2ox levels leads to a hyper-elongated phenotype called SLENDER (Lester et al. 1999, Martin et al. 1999). Since hyper-elongation does not occur in columnar trees either, a combination of growth inhibiting (XTH20, JAs) and growth promoting factors (GA, CK, IAA level) seems to define this special phenotype which allows plants to reach almost normal height despite their shorter internodes. Overexpression of DWF1 (BRs) leads to an inconspicuous growth phenotype with higher tolerance against biotic stress (Wang et al. 2011). This is in line with findings obtained from our DESeq analysis indicating an overrepresentation of differentially regulated genes associated with biotic stress (represented by the functional category "stress biotic" in Online Resource 2B).

3) The characteristics of the columnar phenotype indicate a high level of apical dominance (Cline 1994) and apical control (Wilson 2000) as attested by the low number of sylleptic and proleptic shoots and the low level of acrotony (De Wit et al. 2000). These properties can be attributed to the higher level and the enhanced basipetal transport of bioactive IAA which results in a shift of the hormone gradient. The high number of fruit spurs (Tobutt 1985) which is against the principle of apical dominance is not in line with these properties, but it can be explained by the higher level of CK acting as an antagonist of IAA and promoting the outgrowth of axillary buds (Domagalska and Leyser 2011, Shimizu-Sato et al. 2009). The early growth cessation of the side branches leading to the formation of short spurs was hypothesized to be caused by insufficient nutrient supply due to the high competition between numerous outgrowing axillary buds (Looney and Lane 1984). Single branches that overcome this growth arrest are themselves columnar (Kenis and Keulemans 2007).

When comparing these results to earlier transcriptome analyses (Krost et al. 2012, Zhang et al. 2012), we were able to confirm all findings with respect to IAA, GA, BR and JA regulation. However, the impact of further, until recently unidentified genes as well as the effect of other differentially regulated genes associated with growth cannot be ruled out (Online Resource 2 and Krost et al. (2012)).

Co affects nearby located genes

By analyzing the localization of differentially regulated genes whose regulation was consistent for all collection dates (table 2), a statistically approved enrichment of these genes on chromosome 10 was discovered (figure 3). Since four of them are associated with hormones whose concentration divergences to normal type apple trees most likely make a major contribution to the formation of the columnar phenotype, a direct effect on gene regulation by the presence of *Co* is probable. At the moment it can only be guessed how this is achieved, but a possible explanation would be a selective sweep. This term describes a decrease of variation near a mutation as a result of strong positive selection, e.g. more than 15 genes lost genetic diversity during maize domestication (Pokalyuk 2012, Tian et al. 2009). In our particular case a selective sweep would have led to a somehow fixed expression of genes near *Co*, in line with the fact that progenies of columnar x non-columnar crosses generally do not reach a 1:1 ratio as expected for inheritance studies with focus on a single dominant gene (Hemmat et al. 1997, Kenis and Keulemans 2007, Kim et al. 2003, Lapins 1969, Moriya et al. 2009, Tobutt 1985). Additionally, not all columnar progenies

show the same degree of phenotypic penetrance (De Wit et al. (2004) and P. Braun, pers. communication). Interestingly, QTL (quantitative trait loci) analysis in columnar apple trees revealed single regions on linkage group 10 (now chromosome 10) associated with internode number, internode length and base diameter increment (Conner et al. 1998) – traits which are seriously affected in the columnar phenotype. A third evidence could be the findings of molecular markers CH02a10 and $SCB82_{670}$ showing a linkage disequilibrium (linked to the *Co* gene but located on chromosome 3 (Kim et al. 2003, Moriya et al. 2009, Tian et al. 2005)) which also points to a selective sweep (McVean 2007).

Conclusions

The transcriptome of columnar apple tree SAM shows hundreds of genes to be differentially regulated throughout the whole growth period and therefore are likely to be affected by the presence of *Co*. Focusing on phytohormone associated genes, earlier results on the hormonal state of columnar trees were mostly validated by our transcriptome analysis. By combining RNA-Seq, microarrays and qPCR, a significant enrichment of differentially regulated genes on chromosome 10 likely due to a selective sweep was identified.

Table 2 - Localization of differentially regulated genes.
The table contains 16 differentially regulated genes from columnar SAM validated by at least two different gene expression studies (DESeq based on NGS data, microarray or qPCR) with their genomic localization (chromosome number). Numbers on the right indicate the regulation as $\log_2$ change fold, light green and light red colors show up and down regulation, respectively, according to the scale bar. Chr. - Chromosome; n.d. - not determined. Gene ID is according to the declaration by Velasco et al. (2010).

	Description (Gene ID)	Chr.	DESeq	Microarray	qPCR
1	Xyloglucan endotransglycosylase (MDP0000320017)	**10**	-2.20	-0.73	-1.20
2	Thaumatin-like protein 1b (MDP0000122958)	?	-2.59	-0.99	-1.33
3	Sieve element occlusion b (MDP0000220190)	13	-3.19	-1.53	n.d.
4	Gibberellin 2-beta-dioxygenase 1 (MDP0000137705)	**10**	-2.50	n.d.	-2.30
5	Thioredoxin O1, mitochondrial (MDP0000247363)	8	3.36	1.53	4.72
6	Cytokinin hydroxylase CYP735A (MDP0000909874)	**10**	4.10	2.50	2.78
7	Probable peptide/nitrate transporter (MDP0000565104)	13	5.30	1.07	1.20
8	Vacuolar cation/proton exchanger 3 (MDP0000141538)	12	5.63	1.19	0.92
9	Peroxidase 42 (MDP0000192235)	**10**	-	0.91	1.51
10	UDP-glycosyltransferase 87A1 (MDP0000276057)	11	-	2.40	0.77
11	Interferon-induced GTP-binding protein mx (MDP0000613462)	2	-	1.25	1.20
12	Unknown (MDP0000679810)	1	6.89	1.85	6.00
13	Glycerol-3-phosphate acyltransferase (MDP0000182098)	8	3.86	0.77	n.d.
14	Probable inactive purple acid phosphatase 27 (MDP0000245584)	14	4.17	0.84	n.d.
15	High mobility group B protein (MDP0000268093)	14	-	0.82	1.31
16	Squalene monooxygenase (MDP0000161955)	**10**	3.39	n.d.	0.58

-1 **0** **1**

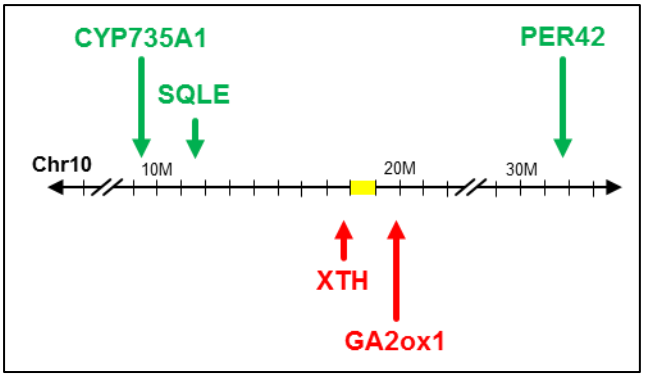


Figure 3 - Differentially regulated genes on chromosome 10
The diagram shows the localization of statistically enriched genes on chromosome 10. Green and red color indicates up and down regulation in columnar cultivar P28, respectively. The yellow square shows the most promising locus where the *Co* gene is located. M – megabases; CYP735A1 – Cytokinin hydroxylase; SQLE – squalene monoonxygenase; XTH – xyloglucan endotransglycosylase; GA2ox1 – Gibberellin 2-beta-dioxygenase 1; PER42 – Peroxidase 42.

Authors' contributions

CK performed the sample preparation for Illumina sequencing, analyzed the NGS and microarray data, supervised the work of SL and wrote the paper. RP was involved in data analysis, interpretation and literature inquiries. SL performed all qPCR experiments. BB was involved in sample preparation and bioinformatics of microarrays. PB and ERS had the idea and initiated the project, were responsible for supervision and contributed to the preparation of the manuscript.

Acknowledgments

We thank Benjamin Rieger and Dr. Steffen Rapp for their work in the development of Perl scripts and their bioinformatics support. We are grateful to Dr. Walker Jackson for his many helpful comments on the manuscript. The work was supported by grants of the Federal Ministry of Agriculture and Nutrition (Nr. 511-06.01-28-1-43.042-07 and Nr. 313-06.01-28-1-43.042-07).

Conflict of Interest

The authors declare that they have no conflict of interest.

Literature

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

Atzorn R (1983) The immunoassay of gibberellins. PLANTA 159: 1-6.

Blažek J (1992) Segregation and general evaluation of spur type or compact growth habits in apples. Acta Hortic. 317: 71-79.

Cline MG (1994) The role of hormones in apical dominance. New approaches to an old problem in plant development. PHYSIOLOGIA PLANTARUM 90: 230-237.

Conner PJ, Brown SK, Weeden NF (1998) Molecular-marker analysis of quantitative traits for growth and development in juvenile apple trees. Theor. Appl. Genet. 96: 1027-1035.

De Wit I, Cook NC, Keulemans J (2004) Characterization of Tree Architecture in two-year-old apple seedling populations of different progenies with a common columnar gene parent. Acta Hortic. 663: 363-368.

De Wit I, Pauwels E, Keulemans J (2000) Different growth habits in apple and correlation between growth characteristics in progenies with a common *Co*-gene parent. Acta Hortic. 538: 325-330.

Depuydt S, Hardtke CS (2011) Hormone signalling crosstalk in plant growth regulation. Curr Biol 21: R365-73.

Domagalska MA, Leyser O (2011) Signal integration in the control of shoot branching. Nat Rev Mol Cell Biol 12: 211-21.

Eaton GW, Lapins KO (1970) Identification of standard and compact apple trees by discriminate function analysis. J Appl Ecol 7: 267-272.

Effendi Y, Rietz S, Fischer U, Scherer GF (2011) The heterozygous *abp1/ABP1* insertional mutant has defects in functions requiring polar auxin transport and in regulation of early auxin-regulated genes. Plant J 65: 282-94.

Fisher DV (1969) Spur-type strains of McIntosh for high density plantings. British Columbia Fruit Growers' Association Quarterly Report 14: 3-10.

Garber M, Grabherr MG, Guttman M, Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods 8: 469-77.

Gazzarrini S, McCourt P (2003) Cross-talk in plant hormone signalling: what *Arabidopsis* mutants are telling us. Ann Bot 91: 605-12.

Gou J, Strauss SH, Tsai CJ, Fang K, Chen Y, Jiang X, Busov VB (2010) Gibberellins regulate lateral root formation in *Populus* through interactions with auxin and other hormones. Plant Cell 22: 623-39.

Hemmat M, Weeden NF, Conner PJ, Brown SK (1997) A DNA Marker for Columnar Growth Habit in Apple Contains a Simple Sequence repeat. J Amer Soc Hort Sci 122: 347-349.

Hyodo H, Yamakawa S, Takeda Y, Tsuduki M, Yokota A, Nishitani K, Kohchi T (2003) Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, XTH9, in inflorescence apices is related to cell elongation in Arabidopsis thaliana. Plant Mol Biol 52: 473-82.

Jacob HB (2010) Breeding experiments of apple varieties with columnar growth and low chilling requirements. Acta Hortic. 872.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40: D109-14.

Kenis K, Keulemans J (2007) Study of Tree Architecture of Apple (*Malus × domestica* Borkh.) by QTL Analysis of Growth Traits. Mol Breeding 19: 193-208.

Kim MY, Song KJ, Hwang J-H, Shin Y-U, Lee HJ (2003) Development of RAPD and SCAR markers linked to the *Co* gene conferring columnar growth habit in apple (*Malus pumila* Mill.). Journal of Horticultural Science & Biotechnology 78: 512-517.

Kleine-Vehn J, Dhonukshe P, Swarup R, Bennett M, Friml J (2006) Subcellular trafficking of the *Arabidopsis* auxin influx carrier AUX1 uses a novel pathway distinct from PIN1. Plant Cell 18: 3171-81.

Kobayashi Y, Murata M, Minami H, Yamamoto S, Kagaya Y, Hobo T, Yamamoto A, Hattori T (2005) Abscisic acid-activated SNRK2 protein kinases function in the gene-regulation pathway of ABA signal transduction by phosphorylating ABA response element-binding factors. Plant J 44: 939-49.

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

Kvam VM, Liu P, Si Y (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am J Bot 99: 248-56.

Lackman P, Gonzalez-Guzman M, Tilleman S, Carqueijeiro I, Perez AC, Moses T, Seo M, Kanno Y, Hakkinen ST, Van Montagu MC, Thevelein JM, Maaheimo H, Oksman-Caldentey KM, Rodriguez PL, Rischer H, Goossens A (2011) Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in *Arabidopsis* and tobacco. Proc Natl Acad Sci U S A 108: 5891-6.

Lane WD, Looney NE (1982) A Selective Tissue Culture Medium for Growth of Compact (Dwarf) Mutants of Apple. Theor. Appl. Genet. 61: 219-223.

Lapins KO (1969) Segregation of compact growth types in certain apple seedling progenies. Canad J Plant Sci 49: 765-768.

Lapins KO, Watkins R (1973) Genetics of compact growth habit. Report of East Malling Research Station for 1972 136.

Lee JM, Looney NE (1977) Absisic Acid Levels and Genetic Compaction in Apple Seedlings. Canad J Plant Sci 57: 81-85.

Lester DR, Ross JJ, Smith JJ, Elliott RC, Reid JB (1999) Gibberellin 2-oxidation and the SLN gene of Pisum sativum. Plant J 19: 65-73.

Looney NE, Lane WD (1984) Spur-type growth mutants of McIntosh apple: A review of their genetics, physiology and field performance. Acta Hortic. 146: 31-46.

Martin DN, Proebsting WM, Hedden P (1999) The SLENDER gene of pea encodes a gibberellin 2-oxidase. Plant Physiol 121: 775-81.

McVean G (2007) The structure of linkage disequilibrium around a selective sweep. Genetics 175: 1395-406.

Moriya S, Iwanami H, Kotoda N, Takahashi S, Yamamoto T, Kazuyuki A (2009) Development of a Marker-assisted Selection System for Columnar Growth Habit in Apple Breeding. J. Japan. Soc. Hort. Sci. 78: 279-287.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. PHYSIOLOGIA PLANTARUM 15: 473-497.

Ohmiya A, Tanaka Y, Kadowaki K, Hayashi T (1998) Cloning of genes encoding auxin-binding proteins (ABP19/20) from peach: significant peptide sequence similarity with germin-like proteins. Plant Cell Physiol 39: 492-9.

Pokalyuk C (2012) The effect of recurrent mutation on the linkage disequilibrium under a selective sweep. J Math Biol 64: 291-317.

Reid JB, Davidson SE, Ross JJ (2011) Auxin acts independently of DELLA proteins in regulating gibberellin levels. Plant Signal Behav 6: 406-8.

Romo S, Jimenez T, Labrador E, Dopico B (2005) The gene for a xyloglucan endotransglucosylase/hydrolase from *Cicer arietinum* is strongly expressed in elongating tissues. Plant Physiol Biochem 43: 169-76.

Rose JK, Braam J, Fry SC, Nishitani K (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. Plant Cell Physiol 43: 1421-35.

Santiago J, Rodrigues A, Saez A, Rubio S, Antoni R, Dupeux F, Park SY, Marquez JA, Cutler SR, Rodriguez PL (2009) Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade A PP2Cs. Plant J 60: 575-88.

Sarwar M, Skirvin RM (1997) Effect of thidiazuron and 6-benzylaminopurine on adventitious shoot regeneration from leaves of three strains of 'McIntosh' apple (*Malus* x *domestica* Borkh.) *in vitro* Scientia Horticulturae 86: 95-100.

Sarwar M, Skirvin RM, Kushad M, Norton MA (1998) Selecting dwarf apple (*Malus x domestica* Borkh.) trees in vitro: multiple cytokinin tolerance expressed among three strains of 'McIntosh' that differ in their growth habit under field conditions. Plant Cell, Tissue and Organ Culture 54: 71-76.

Shimizu-Sato S, Tanaka M, Mori H (2009) Auxin-cytokinin interactions in the control of shoot branching. Plant Mol Biol 69: 429-35.

Staswick PE, Serban B, Rowe M, Tiryaki I, Maldonado MT, Maldonado MC, Suza W (2005) Characterization of an *Arabidopsis* enzyme family that conjugates amino acids to indole-3-acetic acid. Plant Cell 17: 616-27.

Suza WP, Staswick PE (2008) The role of JAR1 in Jasmonoyl-L: -isoleucine production during Arabidopsis wound response. Planta 227: 1221-32.

Takahashi T, Gasch A, Nishizawa N, Chua NH (1995) The *DIMINUTO* gene of *Arabidopsis* is involved in regulating cell elongation. Genes Dev 9: 97-107.

Takei K, Yamaya T, Sakakibara H (2004) Arabidopsis *CYP735A1* and *CYP735A2* encode cytokinin hydroxylases that catalyze the biosynthesis of trans-Zeatin. J Biol Chem 279: 41866-72.

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37: 914-39.

Thomas SG, Phillips AL, Hedden P (1999) Molecular cloning and functional expression of gibberellin 2- oxidases, multifunctional enzymes involved in gibberellin deactivation. Proc Natl Acad Sci U S A 96: 4698-703.

Tian F, Stevens NM, Buckler ESt (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proc Natl Acad Sci U S A 106 Suppl 1: 9979-86.

Tian YK, Wang CH, Zhang JS, James C, Dai HY (2005) Mapping *Co*, a gene controlling the columnar phenotype of apple, with molecular markers. Euphytica 145: 181–188.

Tobutt KR (1985) Breeding columnar apples at east malling. Acta Hortic. 159: 63-68.

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagne D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouze P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2010) The genome of the domesticated apple (*Malus* x *domestica* Borkh.). Nat Genet 42: 833-9.

Wang H, Nagegowda DA, Rawat R, Bouvier-Nave P, Guo D, Bach TJ, Chye ML (2011) Overexpression of *Brassica juncea* wild-type and mutant HMG-CoA synthase 1 in *Arabidopsis* up-regulates genes in sterol biosynthesis and enhances sterol production and stress tolerance. Plant Biotechnol J 10: 31-42.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10: 57-63.

Watanabe H, Suzuki A, Komori S, Sato H (2003) Effect of plant growth regulators on shoot growth of a columnar type apple tree. Hort Res 2: 97-100.

Watanabe M, Bessho H, Suzuki A, Komori S (2008) Seasonal Changes of IAA and Cytokinin in Shoots of Columnar Type Apple Trees. Acta Hortic. 774: 75-80.

Watanabe M, Suzuki A, Komori S, Bessho H (2004) Comparison of Endogenous IAA and Cytokinins in Shoots of Columnar and Normal Type Apple Trees. J. Japan. Soc. Hort. Sci. 73: 19-24.

Watanabe M, Suzuki A, Komori S, Bessho H (2006) Effects of Heading-back Pruning on Shoot Growth and IAA and Cytokinin Concentrations at Bud Burst of Columnar-type Apple Trees. J. Japan. Soc. Hort. Sci. 75: 224-230.

Wilson BF (2000) Apical control of branch growth and angle in woody plants. Am J Bot 87: 601-607.

Zhang S, Wei Y, Lu Y, Wang X (2009) Mechanisms of brassinosteroids interacting with multiple hormones. Plant Signal Behav 4(12): 1117-1120.

Zhang Y, Zhu J, Dai HY (2012) Characterization of Transcriptional Differences Between Columnar and Standard Apple Trees Using RNA-Seq. Plant Molecular Biology Reporter 30: 957-965.

Electronic Supplementary Material

Online Resource 1 – Results from DESeq analysis

The table contains 1173 differentially regulated genes with $p < 0.005$ and fold change > 3.0 obtained from DESeq analysis. For every gene, *Malus x domestica* UniGene database ID, *Malus x domestica* genome project ID and *Populus trichocarpa* genome project ID (used for MAPMAN visualization) is provided. Log_2 fold change values are highlighted in green and red representing up or down regulation, respectively. All entries are sorted by base mean values.

This file is included in the electronic supplementary material accompanying this thesis.

Paper 7

Columnar Apple Primary Roots Share Some Features of the Columnar-Specific Gene Expression Profile of Aerial Plant Parts as Evidenced by RNA-Seq Analysis

Romina Petersen[1], Haris Djozgic[1], Benjamin Rieger[1], Steffen Rapp[1], Erwin R. Schmidt[1]

[1] Department of Molecular Genetics, Johannes Gutenberg-University of Mainz

This paper has been submitted to BMC Plant Biology

Abstract

Background

Primary roots (radicles) represent the first visible developmental stages of the plant and are crucial for nutrient supply and the integration of environmental signals. Few studies have analyzed primary roots at a molecular level, and these were mostly limited to *Arabidopsis*. Here we study the transcriptomes of standard type, heterozygous and homozygous columnar primary roots of apple (*Malus* x *domestica*) by RNA-Seq and quantitative real-time PCR. The columnar growth habit is characterized by a stunted main axis and the development of short fruit spurs instead of long lateral branches. This compact growth possesses economic potential because it allows high density planting and mechanical harvesting of the trees. Its molecular basis has been identified as a nested Gypsy-44 retrotransposon insertion; however the link between the insertion and the phenotype as well as the timing of the phenotype emergence are as yet unclear. We extend the transcriptomic studies of columnar tissues to the radicles as the earliest developmental stages and investigate whether homozygous columnar seedlings are viable.

Results

Radicles mainly express genes associated with primary metabolism, growth and development. About 200 genes show differential regulation in a comparison of heterozygous columnar radicles with non-columnar radicles, whereas the comparison of homozygous columnar radicles with non-columnar radicles yields about 300 differentially regulated genes. Genes involved in cellulose and phenylpropanoid biosynthesis, cell wall modification, transcription and translation, ethylene and jasmonate biosynthesis are upregulated in columnar radicles. Genes in the vicinity of the columnar-specific Gypsy-44 insertion experience an especially strong differential regulation: the direct downstream neighbor, *dmr6-like*, is downregulated in heterozygous columnar radicles, but strongly upregulated in columnar shoot apical meristems.

Conclusions

The transcriptomic profile of primary roots reflects their pivotal role in growth and development. Homozygous columnar embryos are viable and form normal radicles under natural conditions, and selection towards heterozygous plants most likely occurs due to breeders' preferences. Cell wall and phytohormone biosynthesis and metabolism experience differential regulation in columnar radicles. Presumably the first step of the differential regulation most likely happens within the region of the retrotransposon insertion and its tissue-specificity suggests involvement of one (or several) tissue-specific regulator(s).

Keywords

primary root – homozygous – columnar – apple – Gypsy-44 – RNA-Seq – DESeq – MapMan – qRT-PCR

Background

Columnar apple trees show a characteristic pillar-like growth habit with a thick, stunted main axis and short lateral fruit spurs [1, 2]. This growth habit could be of potential benefit to apple growers because columnar trees can be planted more closely together and require less pruning than standard type trees [2, 3]. However, none of the columnar cultivars available to date can compete with commercially successful cultivars in terms of fruit quality and disease resistance [2–7]. Columnar growth arose as a spontaneous somaclonal mutation of a McIntosh tree in Canada in 1961 [8, 9]. With one exception [10], all columnar cultivars that have been described so far are heterozygous (hemizygous) for the columnar mutation (*Co*/-) [11]. Whether the lack of homozygous individuals (*Co*/*Co*) is due to a decreased viability of homozygous columnar seeds/seedlings or just the decision of apple growers to preferentially choose non-columnar breeding partners is unclear.

The molecular cause of the columnar phenotype has recently been identified as the insertion of a Gypsy-44 long terminal repeat (LTR) retrotransposon into the LTR of another retrotransposon on chromosome 10 [10]. Wolters et al. [12] detected a smaller columnar-specific insertion at the same position, which is the solo-LTR of Gypsy-44 and thus most likely represents an artefact (unpublished data). The Gypsy-44 insertion is probably responsible for the upregulation of a nearby gene encoding a 2OG-Fe(II) oxygenase (also called *downy mildew resistance 6-like* (*dmr6-like*)) of unknown function in apical meristems and axillary buds of columnar trees, whereas leaves do not show any expression of *dmr6-like* [10, 12]. Overexpression of *dmr6-like* in *Arabidopsis thaliana* led to a columnar-like phenotype [12]. In addition to the upregulation of *dmr6-like*, a significant increase of the expression level of other genes within the retrotransposon vicinity has been shown [10, 13]. A change of the overall gene expression pattern of the columnar plants is the consequence. The shoot apical meristems and leaves of columnar apple trees show a differential regulation of defense-associated genes, genes involved in secondary metabolism such as terpene and phenylpropanoid synthesis, as well as genes related to auxin and jasmonate synthesis and signaling [10, 13, 14]. Since a reliable detection of the columnar growth habit

is only possible after about two to three years, it is as yet unclear at which developmental time point the gene expression patterns leading to the formation of the columnar habit are established. Up to now, it is also not known whether and how the gene expression pattern of roots is affected by the *Co* mutation. Even the phenotype of own roots of columnar apple trees has never been analyzed, which is probably due to the fact that the vast majority of columnar trees are grown as scions on non-columnar rootstocks.

Germination and radicle emergence are the first developmental steps towards the formation of a new plant. The radicle is important for anchorage, nutrient and water supply of the plantlet as well as for the perception and integration of a multitude of environmental signals such as gravity or pest attacks. Its tip has even been described as the "brain" of the plant by Charles Darwin [15] (cited in [16]). Despite their crucial regulatory and pioneering role in development, little research has been conducted on primary roots at the molecular level (e.g. deciphering their transcriptome profiles) and those were mainly limited to the model plants *Arabidopsis thaliana* and *Zea mays* (for example [17–21]). Only one publication has dealt explicitly with the transcriptome of adult roots of poplar [22]. With regard to apple, whereas seeds have been the subject of several studies mostly owing to their deep and well-pronounced dormancy [23–28], research interest seems to fade significantly when they finish germination. Apple seed dormancy can be overcome by cold treatment (stratification) for 60 – 90 days depending on the cultivar and environmental conditions [23, 28]. After this time, the radicle protrudes the testa as the primary root. Primary roots of the dicotyledonous model plant *Arabidopsis* consist of four longitudinal sections: the root cap (columella) at the tip, followed by the zone of division, the elongation zone and the differentiation zone [29]. A cross section of the root reveals a radial organization of different cell layers: epidermis, cortex, endodermis and the vascular cylinder encompassing the pericycle, protoxylem, protophloem and procambium [30, 31]. This radial symmetry is established at the root apical meristem, a small set of cells near the root tip surrounding the mitotically less active quiescent center [32].

In this study we analyzed and compared the transcriptomes of heterozygous columnar, homozygous columnar and non-columnar primary apple roots. Our aims were 1) to gather general information about the gene expression profile of this poorly studied plant tissue, 2) to analyze whether homozygous columnar seedlings exist and are viable, 3) to determine whether the columnar-specific gene expression profile observed in aerial plant parts can already be observed at the earliest stages of development in the root and 4) to further investigate how the Gypsy-44 insertion might be linked to the formation of the growth phenotype. For this purpose, apple seeds were subjected to stratification and radicles were harvested when they had reached a length of about 3 cm. Half of the radicle was used for DNA isolation and subsequent genotyping, while the remaining half containing the root tip was used for RNA isolation followed by Illumina sequencing. The transcriptomic reads were assembled and contigs were subjected to Basic Local Alignment Search Tool (BLAST) searches [33] as well as Blast2GO analyses [34, 35] to gain a comprehensive view of the genes expressed in primary apple roots in general. Furthermore, the reads were mapped to the apple draft genome [36] and individual gene expression levels (normalized read counts) were compared between columnar and non-columnar radicles, with a special focus on the genes in the vicinity of Gypsy-44, the *Co* mutation. Gene expression patterns in the *Co* target region were compared across primary roots, leaves and shoot apical meristems by additional quantitative real-time PCRs (qRT-PCRs). Similarities and differences in the gene

expression patterns of the different tissues were found. Our data make a substantial contribution to the understanding of the development of the columnar growth habit and primary root function in general.

Results

Homozygous Columnar Apple Seedlings Are Viable

To investigate whether homozygous columnar apple seedlings show reduced viability or phenotypic effects compared to standard-type seedlings in early developmental stages, seeds obtained from apples of the heterozygous columnar cultivar 'Procats 28' (P28) that had been subjected to open pollination were germinated. As the trees were grown surrounded by other columnar apple varieties, the chance of pollination by a columnar father was high. After about 12 weeks of incubation at 4 °C, the germination rate of the seeds was approximately 80 %. No obvious phenotypic differences could be observed between individual radicles. The radicle genotype with regard to the presence of the columnar-specific Gypsy-44 transposable element (TE), which most likely represents the original *Co* mutation [10], was determined via PCRs. The diagnostic PCR assays as established by Otto et al. discriminate unambiguously between the non-columnar, the heterozygous columnar and the homozygous columnar genotype [10]. Of 119 seedlings subjected to genotyping, 40 seedlings were detected to be non-columnar, 59 showed a heterozygous columnar genotype, while the remaining 20 seedlings carried the columnar-specific Gypsy-44 insertion homozygously. These results could be confirmed by PCR assays using our indel-based markers I2_3_M1 and H1_M1 that are tightly linked to the *Co* mutation [37]. In total, a genotype ratio of non-columnar : heterozygous columnar : homozygous columnar seedlings of 2 : 3 : 1 (Fig. 1) was detected. This proves that homozygous columnar apple embryos are viable and most likely germinate at normal ratios.

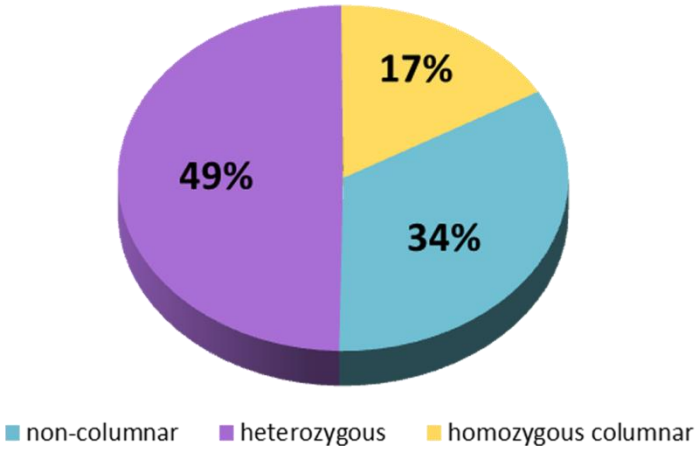


Figure 1 – Genotypes of primary roots
The genotype of 119 apple primary roots was determined by marker PCRs with regard to the presence of the Gypsy-44 retrotransposon insertion.

Primary Roots Mainly Express Genes for Growth, Development and Signaling

We sequenced RNA extracted from one primary root of each genotype and obtained about 118 million, 104 million and 126 million reads for the non-columnar, the heterozygous columnar and the homozygous columnar sample, respectively [EMBL: PRJEB6212] (Table 1). In order to investigate biological replicates, in a second approach three radicles of each genotype were pooled prior to RNA isolation and were subjected to Illumina sequencing, yielding about 40 million, 28 million and 67 million reads for the non-columnar, the heterozygous columnar and the homozygous columnar sample, respectively [EMBL: PRJEB6212]. Illumina reads were assembled in the CLC Assembly Cell using different k-mer sizes, and assemblies yielding the highest N50 value were used for downstream analysis. For the first datasets comprising more than 100 million sequences, the highest N50 values are obtained for k = 17, for the smaller replicate datasets for k = 18. Table 1 summarizes the assembly results. In the larger datasets, about 44,000 – 49,000 contigs are produced, whereas assemblies of the smaller datasets yield about 28,000 – 38,000 contigs. N50 values are in the range of 1,200 – 1,500 bp. With no mismatches allowed, more than 40 % of the trimmed reads of each dataset could be mapped back to the corresponding contigs.

Table 1 – Results of assemblies
Illumina sequences were assembled and key values for contigs are indicated. The percentage of reads that was successfully mapped back to the contigs as a reference is shown in the last column.

	# raw seqs	k-mer size	# contigs	GC content (%)	longest contig	shortest contig	N50	reads mapped back (%)
PR (-/-)	118400555	17	48023	43	9994	200	1484	63
PR (Co/-)	104464241	17	44197	44	10381	200	1455	59
PR (Co/Co)	125731385	17	48932	43	16563	200	1487	60
PRs (-/-)	40261320	18	32471	45	11741	200	1288	43
PRs (Co/-)	27708416	18	28493	45	13248	200	1242	40
PRs (Co/Co)	67322394	18	37527	44	8970	200	1329	52

In order to find out how many and which genes are represented in the datasets, BLAST searches were conducted with all contigs against the annotated apple genes (MDPs), *Malus* x *domestica* expressed sequence tags (ESTs), *Malus* x *domestica* unigenes and the SwissProt/UniProtKB database (Table 2 and Table 3). A maximum of 89 % of the sequences match to a homolog in the MDP list, whereas only up to 60 % of the contigs have a match against the SwissProt/UniProtKB database. Only up to 19 contigs have the same hit in the MDP database (meaning they most likely represent fragments of the same gene), indicating that about 89 % of the contigs represent different genes. By contrast, up to 266 contigs yield the same SwissProt/UniProtKB hit, thereby reducing the number of individual genes detected to about 11,500 (44 % of the contigs created).

Replicate datasets were subjected to Blast2GO analysis. Of the sequences yielding BLAST hits against SwissProt/UniProtKB, 51 % – 63 % were successfully annotated. About 86,000 – 100,000 Level 2 gene ontology (GO) terms were assigned (Additional File 1), half of which are associated with growth and development (categories “metabolic process”, “growth”, “developmental process” and “cellular component organization or biogenesis”) or

with the integration and reaction to environmental signals (categories “response to stimulus”, “biological regulation” and “signaling”). This is in line with the major roles of primary roots.

Table 2 – Results of BLASTx searches against MDPs and Malus ESTs
The number and percentage of contigs yielding a hit and the number of different hits are indicated. The percentage of different hits refers to the total number of contigs and thus represents an estimate for the percentage of contigs representing individual genes.

	MDPs		Malus ESTs	
	# contigs with hits (% of total contigs)	# different hits (% of contigs with hits)	# contigs with hits (% of total contigs)	# different hits (% of contigs with hits)
PR (-/-)	40010 (83)	25197 (63)	38094 (79)	31484 (83)
PR (Co/-)	37666 (85)	23966 (64)	35816 (81)	30183 (84)
PR (Co/Co)	40110 (82)	25100 (63)	38208 (78)	31583 (83)
PRs (-/-)	28377 (87)	19456 (69)	26746 (82)	23621 (88)
PRs (Co/-)	25278 (89)	17481 (69)	23550 (83)	21262 (90)
PRs (Co/Co)	32067 (85)	21350 (67)	30071 (80)	26012 (87)

Table 3 – Results of BLASTx searches against Malus Unigene and SwissPro/UniProtKB
The number and percentage of contigs yielding a hit and the number of different hits are indicated. The percentage of different hits refers to the total number of contigs and thus represents an estimate for the percentage of contigs representing individual genes.

	Malus Unigene		SwissProt/UniProt KB	
	# contigs with hits (% of total contigs)	# different hits (% of contigs with hits)	# contigs with hits (% of total contigs)	# different hits (% of contigs with hits)
PR (-/-)	31593 (66)	17379 (55)	26221 (55)	11467 (44)
PR (Co/-)	29806 (67)	16883 (57)	24992 (57)	11238 (45)
PR (Co/Co)	31636 (65)	17288 (55)	26044 (53)	11404 (44)
PRs (-/-)	22545 (69)	14437 (64)	19102 (59)	9821 (51)
PRs (Co/-)	19923 (70)	13398 (67)	17111 (60)	9253 (54)
PRs (Co/Co)	25278 (67)	15373 (61)	21343 (57)	10331 (48)

Differential Gene Expression in Columnar versus Non-Columnar Radicles is Widespread

For evaluation of differential gene expression, reads were mapped against the annotated Golden Delicious genome [36], total read counts for each MDP were extracted, and differential gene expression was statistically analyzed in DESeq [38]. Highly expressed (base mean > 100), significantly up- or downregulated genes (fold change > 2 or < 0.5 and p-value < 0.05) were visualized in MapMan [39]. In order to evaluate whether replicate datasets showed consistency for gene expression values, Pearson correlation coefficients were determined for all library comparisons. All libraries show a high degree of correlation, with Pearson correlation coefficients ranging from 0.80 – 0.98 (Table 4).

The comparison of non-columnar versus heterozygous columnar primary roots identifies 194 significantly differentially expressed genes (Additional File 2). Visualization in MapMan (Fig. 2) shows that many genes are upregulated in heterozygous columnar roots when

compared with non-columnar roots. This applies to genes involved in cellulose synthesis, cell wall modification and degradation, glycolysis, the oxidative pentose phosphate pathway, the biosynthesis of phenylpropanoids, starch and fatty acids and the mitochondrial electron transport. On the other hand, single genes involved in lipid and starch degradation and photorespiration are downregulated in the heterozygous columnar radicles (Fig. 2A). Some genes associated with stress reactions such as ethylene biosynthesis and signaling (Fig. 2B) are also more highly expressed in heterozygous columnar than in non-columnar primary roots. This also holds true for genes encoding the plastidic ribosomal proteins (Fig. 2C).

The comparison of non-columnar versus homozygous columnar primary roots yields 269 significantly differentially expressed genes (Additional File 3) and most of them are induced in homozygous columnar radicles (Fig. 3). Single genes encoding components of the carbohydrate metabolism, cellulose synthesis, cell wall modification and degradation, the oxidative pentose phosphate pathway, fermentation, starch synthesis and ascorbate/glutathione metabolism are more highly expressed in the homozygous columnar radicles than in the non-columnar radicles (Fig. 3A). On the other hand, single genes involved in fatty acid synthesis and lipid degradation, glycolysis, the tricarboxylic acid cycle, the mitochondrial electron transport chain and the Calvin cycle show lower expression in the homozygous columnar roots than in the non-columnar primary roots. With regard to genes involved in stress reactions (Fig. 3B), abscisic acid, ethylene and jasmonate-associated genes, genes encoding Myb transcription factors and peroxidases as well as genes involved in the maintenance of the redox state are upregulated. Genes linked to transcription and translation are also induced, and this holds true for nuclear as well as plastidic genes (Fig 3C).

Table 4 – Pearson correlation coefficients of RNA-Seq libraries
Correlation coefficients were calculated based on total read counts obtained for individual MDPs. Shading indicates the degree of correlation, dark green indicating perfect correlation and white the lowest correlation of these comparisons.

	PR (-/-)	PRs (-/-)	PR (Co/-)	PRs (Co/-)	PR (Co/Co)	PRs (Co/Co)
PR (-/-)	1.00	0.84	0.80	0.83	0.85	0.86
PRs (-/-)		1.00	0.91	0.98	0.95	0.96
PR (Co/-)			1.00	0.93	0.93	0.89
PRs (Co/-)				1.00	0.95	0.96
PR (Co/Co)					1.00	0.92
PRs (Co/Co)						1.00

When heterozygous and homozygous primary root samples are compared with each other, a similar picture emerges as for the analysis of non-columnar versus homozygous columnar radicles (Additional File 4). 264 genes show significant differential regulation (Additional File 5) and of these, most genes are upregulated in the homozygous columnar primary roots. These genes are linked to primary and secondary metabolism (Fig. 4A), stress reactions (Fig. 4B) as well as transcription and translation (Fig. 4C). The only exceptions are some genes involved in glycolysis and mitochondrial electron transport, which show a lower expression in homozygous columnar radicles than in heterozygous columnar radicles.

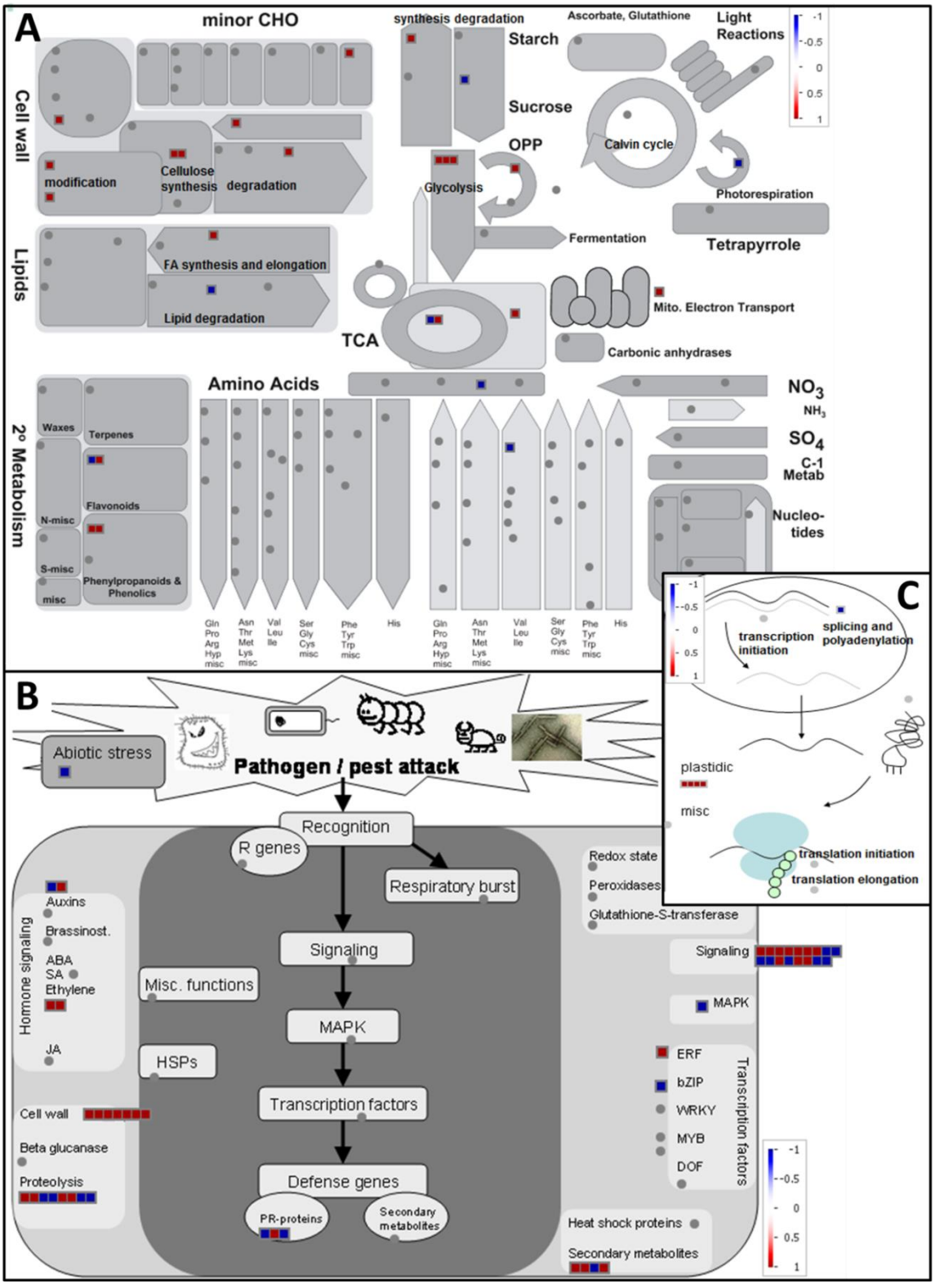


Figure 2 – Differential gene expression in heterozygous compared with non-columnar primary roots

LOG2 fold changes of significantly differentially expressed genes (normalized to the non-columnar sample) were imported and visualized in MapMan for the heterozygous columnar sample with regard to a metabolism overview (A), pathogen/pest attack (B) and transcription and translation (C). Genes upregulated in heterozygous columnar radicles are shown as red boxes, downregulated genes are shown as blue boxes.

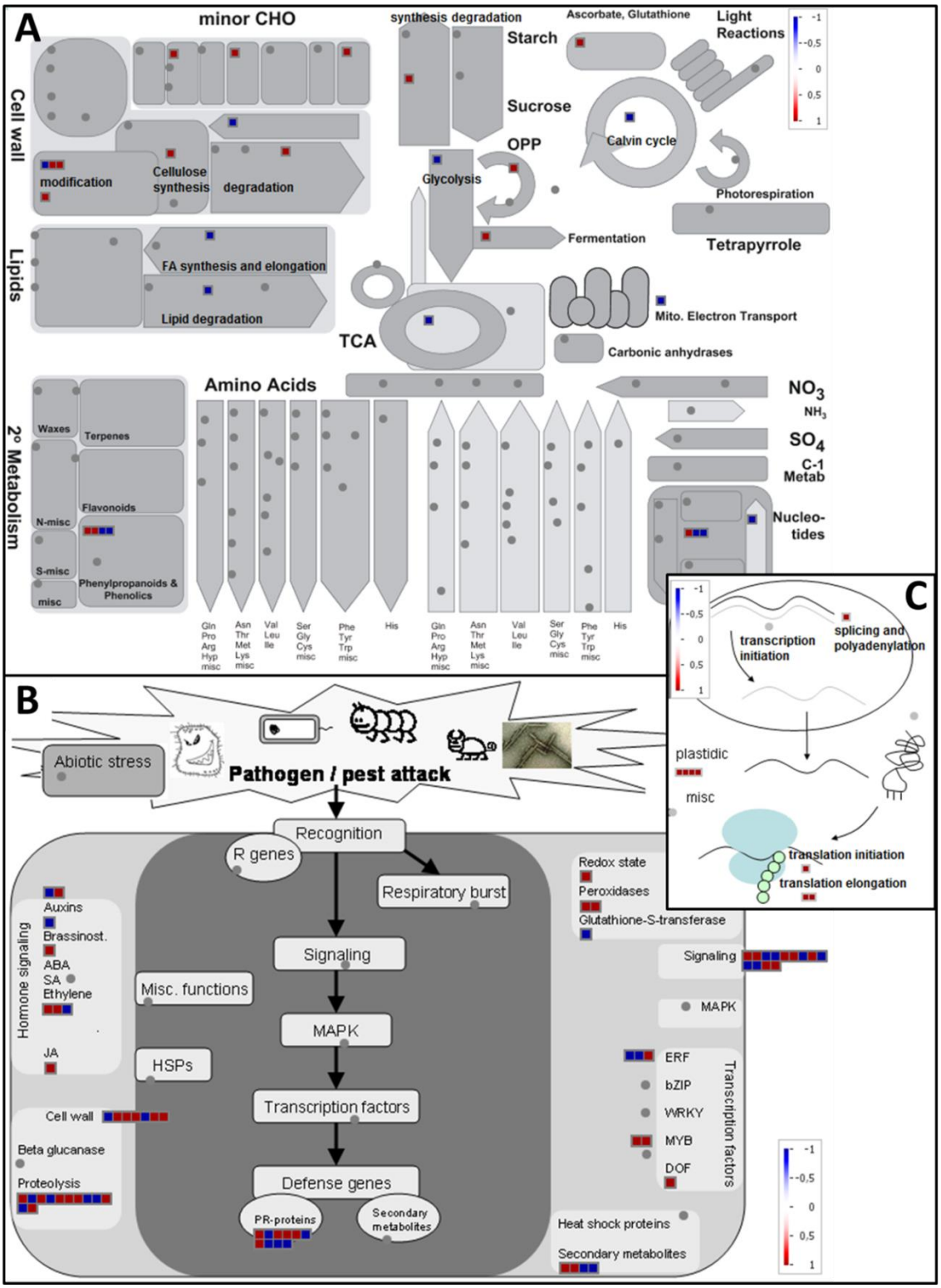
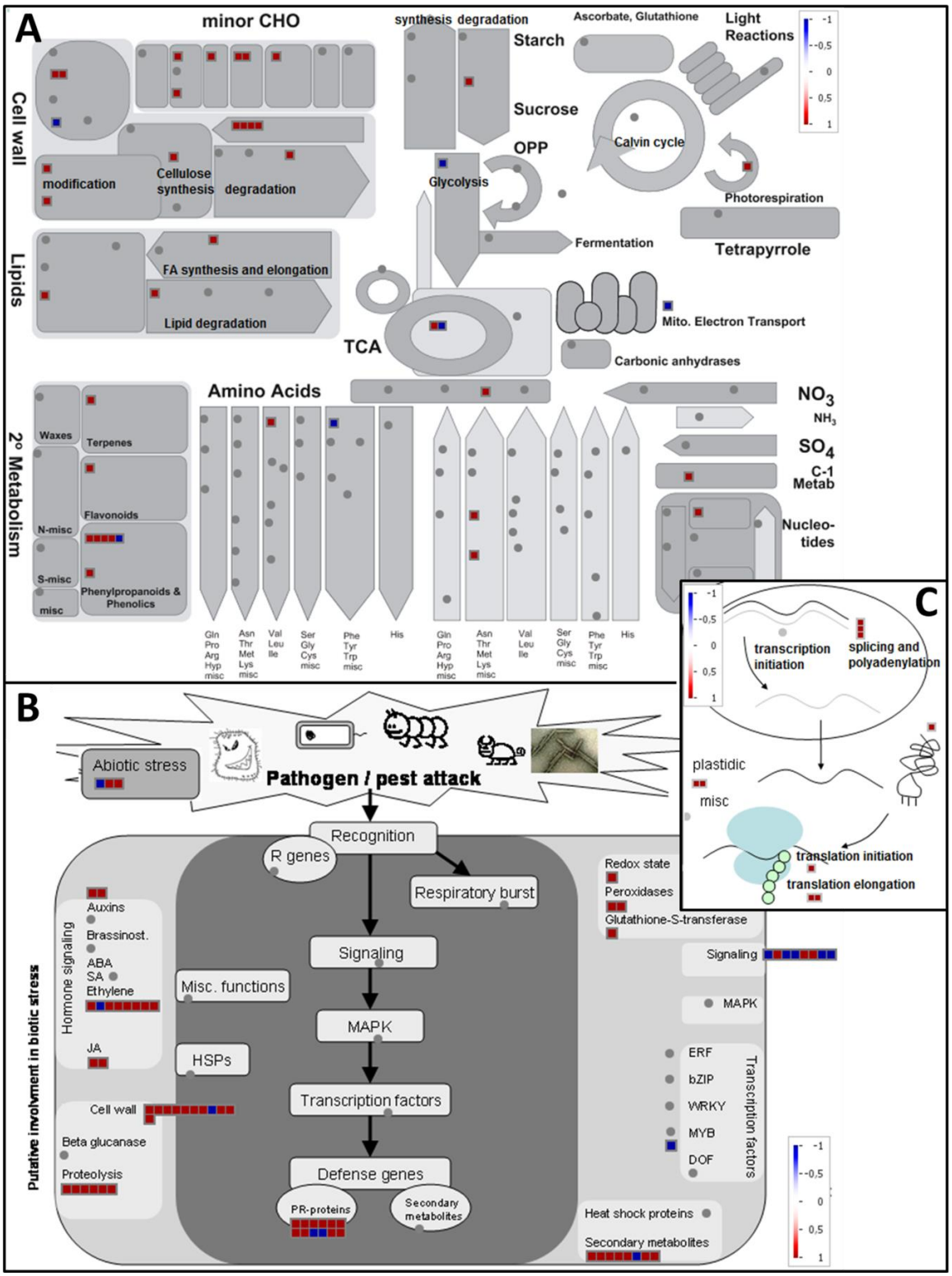


Figure 3 – Differential gene expression in homozygous compared with non-columnar primary roots

LOG2 fold changes of significantly differentially expressed genes (normalized to the non-columnar sample) for the homozygous columnar sample were imported and visualized in MapMan with regard to a metabolism overview (A), pathogen/pest attack (B) and transcription and translation (C). Genes upregulated in homozygous columnar radicles are shown as red boxes, downregulated genes are shown as blue boxes.



Additional file 4 (png) – Figure of differentially expressed genes in homozygous columnar compared with heterozygous primary roots.
LOG2 fold changes of significantly differentially expressed genes (normalized to the heterozygous sample) were imported and were visualized in MapMan for the homozygous columnar sample with regard to a metabolism overview (A), pathogen/pest attack (B) and transcription and translation (C). Genes upregulated in homozygous columnar radicles are shown as red boxes, downregulated genes are shown as blue boxes.

Genes Downstream of Gypsy-44 Are Downregulated in Radicles but Upregulated in Aerial Organs of Columnar Plants

In order to identify the link between the columnar-specific Gypsy-44 insertion and the columnar growth habit (most likely corresponding to the first, causal level of differential gene expression) we performed detailed expression analyses of all genes in the vicinity of the Gypsy-44 insertion. Eight genes that had previously been found to be transcribed and were annotated on our BAC metacontig based on the transcriptomic data [10] as well as Gypsy-44 itself were used for in-depth RNA-Seq and qRT-PCR analyses (Table 5, Fig. 5). Fold changes were calculated for the primary root datasets as well as for three RNA-Seq Illumina datasets each, generated from shoot apical meristems of A14 and P28 ([13, 40], [EMBL: PRJEB2506]) and two transcriptomic datasets each, obtained from the total RNA of a McIntosh and a Wijcik leaf ([10], [EMBL: PRJEB1902]). If all genotypes of one tissue displayed read counts of 0 or 1 for a specific gene, this gene was considered not to be expressed (numeric read counts and fold changes can be found in Additional file 6). The results are shown in the upper part of Fig. 5. *At1g08530-like* and *MDP0000934866* (*At1g06150-like*) show similar expression levels for columnar and non-columnar varieties in all samples analyzed. For *MDP0000927091* (*Autophagy9-like*) and *5NG4-like*, similar levels of transcription are reached in all but the SAM3 dataset, in which they are induced in the columnar sample when compared with the non-columnar sample. Fold changes are less consistent between different tissues and across biological replicates for *MDP0000912172* (*PP2C15-like*), *MDP0000163720* (*ACC1-like*) and Gypsy-44 itself. For the former two genes, this might be caused by the generally lower expression level making fold changes prone to fluctuation. The most striking results are obtained for *dmr6-like* and *MDP0000927098* (*ATL5K-like*), the first two protein coding genes that follow downstream of the Gypsy-44 insertion: they are downregulated in the primary root samples and strongly upregulated in the shoot apical meristem of columnar varieties. These effects are most pronounced for *dmr6-like* in the shoot apical meristem, where no or only basal transcription (0 reads or 1 read) occurs in non-columnar A14, while expression is 30-, 53- and 56-times higher in the three biological replicates of P28.

Gypsy-44 itself, its direct neighboring genes and all other genes within the *Co* target region showing interesting differential gene expression were subjected to qRT-PCRs in order to verify the RNA-Seq results (Fig. 5, lower part). Overall, the qRT-PCR results are mostly consistent with the RNA-Seq results. However, fold changes show less variation across biological replicates in the qRT-PCRs than in the RNA-Seq data. Fold changes of around 1 are obtained for *At1g08530-like* and *MDP0000934866* (*At1g06150-like*) for all tissues with the exception of a 2.5-fold induction of *MDP0000934866* in one of the shoot apical meristem samples. Gypsy-44 is upregulated in the shoot apical meristem, but downregulated in leaves and heterozygous primary roots. *MDP0000163720* (*ACC1-like*) is induced in shoot apical meristems of P28 and slightly induced in heterozygous and homozygous primary roots of one replicate dataset. With regard to the two genes downstream of Gypsy-44, downregulation of *MDP0000927098* (*ATL5K-like*) in columnar primary roots is corroborated, whereas its upregulation in the shoot apical meristem is less pronounced in the qRT-PCR than in the RNA-Seq data. In the qRT-PCR results for *dmr6-like*, downregulation is only detected when heterozygous radicles are compared with non-columnar radicles, but not when homozygous columnar radicles are compared with non-columnar radicles. Strong upregulation is detected in the Wijcik leaf when compared with

the McIntosh leaf. The very strong induction of *dmr6-like* in the shoot apical meristem samples is validated by a 41-fold and 202-fold induction in the columnar shoot apical meristem replicates.

In order to identify any conserved cis-regulatory sequences of *dmr6-like* whose effect might be impaired by the insertion of Gypsy-44, we conducted comparative sequence analyses among the Rosaceae species apple, pear (*Pyrus communis*), peach (*Prunus persica*), Chinese plum (*Prunus mume*) and strawberry (*Fragaria vesca*). In these species, the genes flanking the Gypsy-44 insertion are microcolinear (with an inverse orientation on linkage group 2 of *Fragaria*), enabling the analysis of conserved non-coding sequences (CNSs) in the intergenic region. Remarkably, the intergenic region between *At1g08530-like* and *dmr6-like* or their orthologs has a size of about 33 kb in *Malus* (without the additional 8.2 kb of Gypsy-44), 6.8 kb in *Pyrus*, 1.7 kb in *Fragaria* and 1.4 kb in *Prunus*, suggesting that it has served as a popular target of TE insertions in the Pyreae and especially in the *Malus* lineage. Within this region, one CNS of about 400 bp showing two peaks in the identity plot can be observed about 31 kb upstream of *dmr6-like* in *Malus* (Fig. 5). The sequence of the second peak region yields a hit against a class II TE of the Mariner group in a CENSOR BLAST search [41], whereas no information can be obtained from database searches for the first part, suggesting that it might contain sequences conserved owing to their importance for gene regulation.

In conclusion, there is evidence that Gypsy-44 influences the expression level of its direct downstream gene, *dmr6-like*, and most likely also some genes located further downstream, possibly via impairment of the function of a CNS.

Table 5 – Genes annotated on the BAC metacontig

Eight genes were found to be expressed in at least one of the tissues investigated and were annotated on the BAC metacontig. Their name, position on the BAC metacontig and on chromosome 10 of the Golden Delicious (GD) draft genome sequence [36] as well as the possible function (according to BLAST searches) are indicated. bHLH, basic helix-loop-helix

Gene Name	Position on BAC Metacontig (strand)	Position on Chr 10 of GD Genome	Probable Function
At1g08530-like	49164 – 52233 (+)	18743425 – 18746497	unknown
dmr6-like	93949 – 96036 (+)	18814292 – 18815452	defense reaction
MDP0000927098 (ATL5K-like)	103945 – 104770 (+)	18853768 – 18854596	ubiquitination
MDP0000927091 (Autophagy9-like)	119780 – 126200 (-)	18832988 – 18836788	recycling of cell components
5NG4-like	140776 – 142998 (-)	18862782 – 18864480	auxin-induced transporter
MDP0000912172 (PP2C15-like)	145267 – 146261 (+)	18866768 – 18867762	serine/threonine phosphatase
MDP0000934866 (At1g06150-like)	169780 – 175840 (+)	18884175 – 18886818	bHLH transcription factor
MDP0000163720 (ACC1-like)	184078 – 185923 (-)	18905698 – 18907541	ethylene biosynthesis

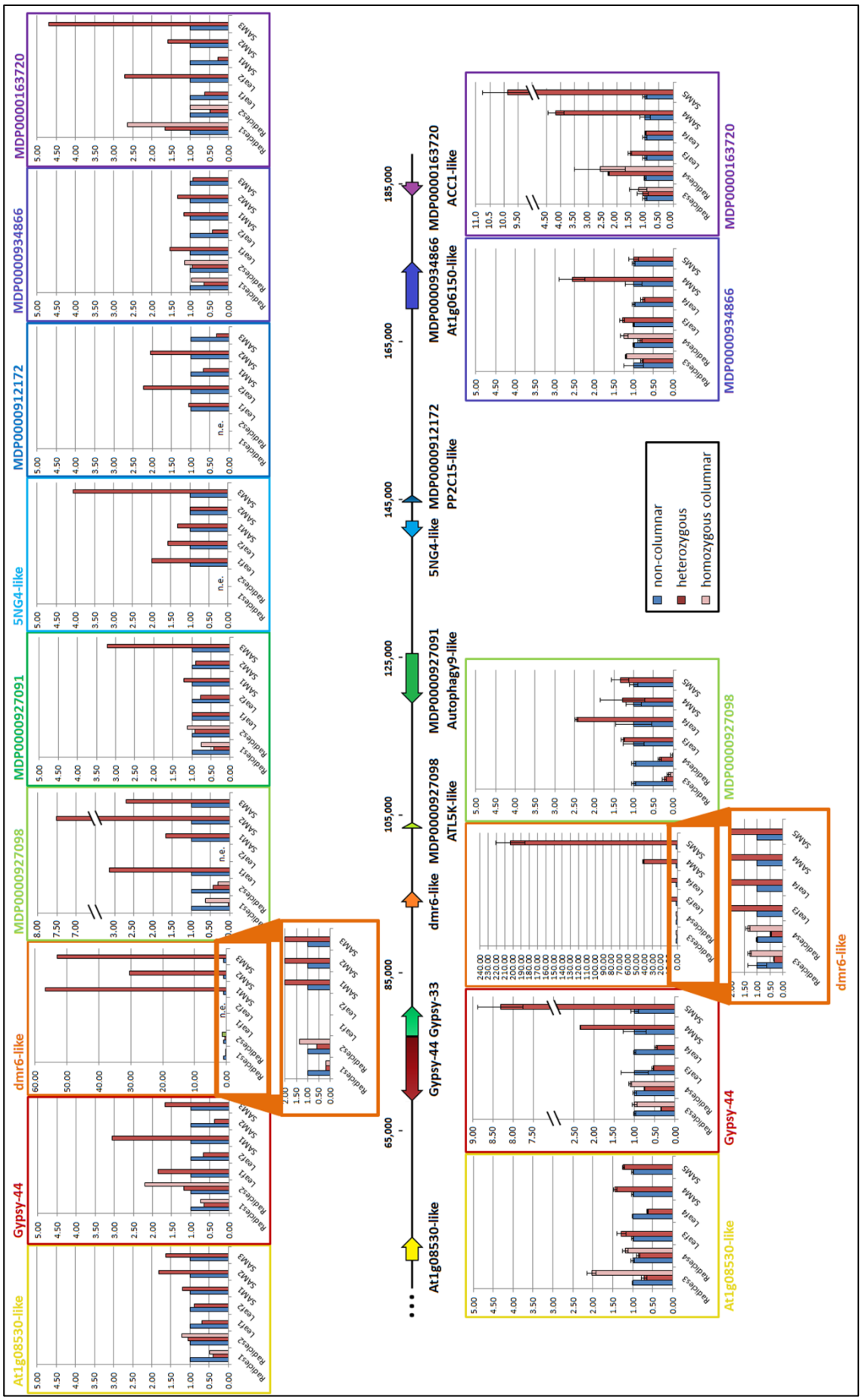

At1g08530-like
Gypsy-44
dmr6-like
MDP0000927098
MDP0000927091
5NG4-like
MDP0000912172
MDP0000934866
MDP0000163720
At1g08530-like
Gypsy-44
dmr6-like
MDP0000927098
MDP0000927091
Autophagy9-like
5NG4-like
MDP0000912172
PP2C15-like
MDP0000934866
At1g06150-like
MDP0000163720
ACC1-like
non-columnar
heterozygous
homozygous columnar
65,000
85,000
105,000
125,000
145,000
165,000
185,000

Figure 4 – Differential gene expression in the *Co* target region
Fold changes of RNA-Seq samples (upper panel) and n-fold expression of qRT-PCR experiments (lower panel) in different tissues were calculated for up to eight genes located in the vicinity of Gypsy-44 as well as for Gypsy-44 itself in at least two biological replicates. Positions on the metacontig are given. Blue bars indicate fold changes for the non-columnar sample (normalized to 1), red and light pink bars indicate fold changes in the heterozygous and homozygous columnar sample, respectively, for a comparison with the non-columnar sample. Two parallel lines on a bar and the y axis represent a broken axis. Absolute read counts of 0 or 1 in both genotypes of a tissue are considered no expression (n.e.). Error bars in qRT-PCR bar chart represent standard deviations of three technical replicates. The two orange boxes below the two panels of graphs signify a zoom-in on the *dmr6-like* graphs within the region 0 – 2.

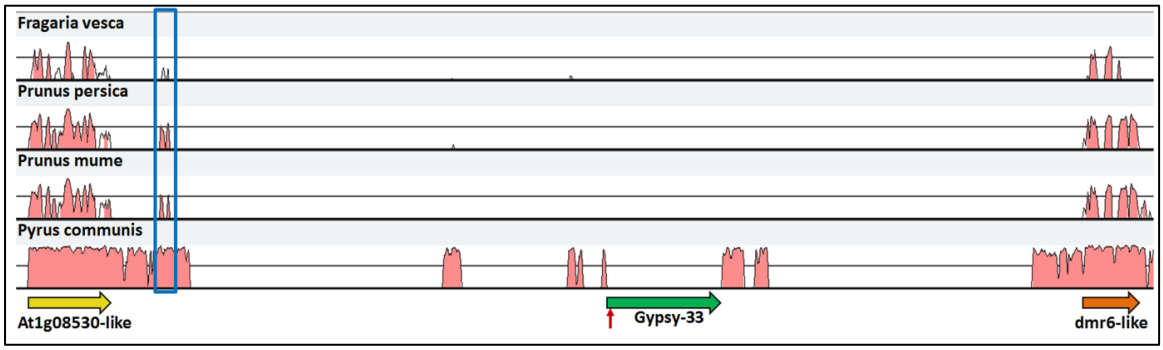


Figure 5 – Conserved regions within the Gypsy-44 region
An mVISTA plot of sequence identity between the *Malus* x *domestica* sequence spanning the region from *At1g08530-like* to *dmr6-like* (x axis) and different Rosaceae species indicates sequence conservation in exon regions and one CNS present in all species investigated (blue box). Other conserved regions in Pyrus correspond to TEs that probably inserted before the divergence of pear and apple. The red arrow marks the Gypsy-44 insertion site.

Discussion

Genotyping, Sequencing and Assemblies

So far, only heterozygous columnar cultivars have been described in the available literature, with one exception of a homozygous cultivar at Geisenheim University [10]. Therefore, Meulenbroek et al. [4] suggested that *Co* or a linked gene might negatively influence the fitness of pollen, seeds or early seedlings. However, crosses between two columnar apple trees have been shown to yield up to 75 % columnar progeny [4, 42], which is in accordance with the result of dominant Mendelian inheritance comprising 25 % homozygous and 50 % heterozygous columnar plants. In our genotyping experiments with the progeny of a heterozygous columnar cultivar we detected 17 % homozygous columnar radicles. Hence, homozygous columnar primary roots are clearly viable. They did not show any deviating phenotypic differences either. The genotype ratio of non-columnar : heterozygous columnar : homozygous columnar of 2 : 3 : 1 is not really surprising because seeds were obtained by open pollination.

In this study, we detected that apple primary roots express between 9,000 and 32,000 distinct genes (represented as contigs yielding different BLAST hits). The number of genes identified as expressed is dependent on the dataset and the criteria used to distinguish

individual genes. Assembling the smaller replicate datasets produced less contigs than assembling the larger datasets because increased sequencing depth facilitates detection of a higher number of transcripts. Regarding the total number of active genes in primary roots, we consider the numbers obtained from BLAST searches against the Malus Unigene dataset, around 17,000, to be the best estimate for those genes that are assembled. MDP and EST most likely show redundancy, listing alleles or fragments of one gene as individual genes, and the SwissProt/UniProtKB probably does not contain all the apple-specific genes. However, considering that not all contigs matched to a Malus Unigene entry and that not all reads were assembled into contigs, the actual number of active genes is probably significantly higher than estimated. There is still a lot of discussion about the real number of genes in apple. Within its genomic sequence, almost 58,000 genes have been anchored, which is the highest gene number reported for plants so far, and this has even been considered an underestimate [36]. However, pear only has about 42,000 genes, and when the apple genome is re-assembled filtering out overlapping genes in apple chromosomes that might correspond to alleles instead of individual genes, the gene number drops down to 45,293 [43]. This would be more consistent with gene numbers of other close relatives such as peach (27,852) [44] and strawberry (34,809) [45]. Newcomb et al. [46] conducted one of the first exhaustive EST analyses in apple and found about 43,000 non-redundant sequences, which they considered to be approximately half of the apple genes. By contrast, Allan et al. [47] assumed the apple EST dataset of 68,599 sequences from databases to be an overestimate. On the other hand, based on EST analyses by Sanzol [48], 68 % of the apple genes fall into families with a mean copy number of 4.6 owing to several genome duplications, and the members of one family can have high sequence similarity but still represent different genes rather than alleles. Hence, the exact gene number remains debatable.

With regard to gene function, the well-known non-model-organism problem of an unsatisfying proportion of genes being successfully annotated is encountered [49]. Further efforts would be necessary to assign possible functions to a higher number of contigs. However, of those genes that were annotated, the majority had the expected function in growth, development and signaling, which is in agreement with the results obtained from transcriptome analyses of the poplar root and the *Arabidopsis* pericycle, where the most highly expressed genes were found to be involved in protein synthesis, metabolism, cellular communication and signal transduction [20, 22].

Evaluation of Differential Gene Expression

DESeq was used to evaluate differential gene expression since it has been found to perform well in a comparison of DESeq, edgeR, baySeq and a method employing a two-stage Poisson model [50]. The comparison of gene expression in columnar and non-columnar radicles yields a lower number of significantly differentially expressed genes than the comparison of gene expression in columnar and non-columnar shoot apical meristems [13]. Differential expression was detected for 200 – 300 and more than 600 genes in primary roots and shoot apical meristems, respectively, despite a more stringent definition of significant differential expression in the shoot apical meristem study of Krost et al. [13]. This might indicate that gene expression of columnar and non-columnar varieties is more similar in underground than in aerial organs. Alternatively, the difference in gene activity might be smaller in early developmental stages than at a higher age of the plants, which would

explain the fact that the columnar growth habit can only be reliably detected after about 2 – 3 years [51, 52]. However, interpretation of the differential expression data is hampered by the fact that the genetic background of the individual radicles used for RNA extraction is unclear due to the open pollination of the flowers. While the mother plant is P28, the father could be any of the dozens of different varieties grown on the field. Additionally we have no knowledge about recombination that might have occurred within the gametes. We chose this material because we wanted to find out whether homozygous columnar individuals occur under natural field conditions. However, for gene expression analyses seeds obtained from targeted crossings would be much more favorable. Moreover, radicles consist of distinct developmental regions in the longitudinal direction and radicles were halved without detailed prior investigation so that different developmental zones most likely harboring different gene expression patterns might have been grouped to each RNA isolation sample. In future studies, the radicles should be subdivided into their individual zones by microscopic control (possibly after fluorescence *in situ* hybridization against marker genes of the distinct zones, similarly to [17]) and the gene expression should be investigated separately for each zone.

Nevertheless, the pattern of differential gene expression in primary roots is in concordance with that of aerial plant parts. The differential expression of genes involved in cell wall biosynthesis and modification observed for heterozygous and homozygous columnar primary roots has already been described for shoot apical meristems of columnar trees [14]. In addition, an increased transcription of genes associated with jasmonate signaling and biosynthesis has been found for shoot apical meristems and leaves of columnar varieties [10, 13, 14], and elevated ethylene signaling and biosynthesis seems to occur not only in columnar radicles, but also in leaves of columnar Wijcik compared with non-columnar McIntosh [10]. Therefore, alterations in cell boundary components and an increased level of jasmonates and ethylene might represent features of columnar apple trees that are already established in the early developmental stages and across different organs of the plantlet.

The First Level of Differential Gene Expression Occurs in the Gypsy-44 Vicinity

In a previous study [10] we found that columnar growth is most likely caused by the insertion of the Gypsy-44 LTR retrotransposon on chromosome 10 at 18.8 Mb, in accordance with the target region determined for the *Co* gene by Moriya et al. [53] and Baldi et al. [52]. Since it is a nested retrotransposon insertion, the link between its presence and the formation of the columnar phenotype remains to be unraveled. In this study, gene expression of Gypsy-44 itself and its neighboring genes was found to differ in columnar compared with non-columnar varieties in the individual plant organs. The expression of Gypsy-44 is induced in the shoot apical meristem. Its nearest downstream neighbor gene, *dmr6-like*, shows very strong (up to 200-fold) upregulation in shoot apical meristems and downregulation in the primary root samples of columnar individuals. The next gene downstream of *dmr6-like*, *MDP0000927098* (*ATL5K-like*), and one other downstream gene, *MDP0000163720* (*ACC1-like*), are also induced in shoot apical meristems of columnar P28. In contrast, expression of the upstream neighbor of Gypsy-44, *At1g08530-like*, remains unchanged. This indicates that Gypsy-44 influences the gene expression pattern of the downstream genes, especially *dmr6-like*, which has already been discussed as a possible candidate gene for the mediation of the columnar phenotype [10, 12]. In addition to its

induction in axillary meristems [12] and shoot apical meristems [10], *dmr6-like* is induced in our Wijcik leaf qRT-PCR samples, which were generated from RNA of the newly developing leaves near the top of the tree, but it shows no expression in the Wijcik leaf RNA-Seq data, which were obtained from RNA of young but fully developed leaves. We therefore conclude that *dmr6-like* is upregulated in all aerial organs of columnar trees that are important for the development of the plant organs showing differences between columnar and non-columnar plants: the shoot apical meristem, which is responsible for the formation of the short internodes and short lateral branches, the axillary meristem, which does not develop into lateral branches but instead produces short spurs, and the young leaves, which form a thicker palisade parenchyma with a high chlorophyll content [54]. In contrast, this upregulation does not occur in organs that have finished their development, such as older leaves, as well as in primary roots, which are less significant for the growth habit of the aerial organs; instead, a reverse regulation seems to take place on the below-ground organs. Wolters et al. [12] have already overexpressed *dmr6-like* in *Arabidopsis* and found the plants to have shorter internodes and smaller branches, strongly suggesting that *dmr6-like* is involved in the formation of the columnar growth habit. Our results support this hypothesis, but also raise further questions, e.g. how the tissue specificity of this induction is achieved.

Possible effects of the Gypsy-44 insertion like alterations in the methylation pattern or the allocation of enhancers and/or silencers have already been discussed [10]. The tissue specificity of this effect suggests that the Gypsy-44 insertion with a size of 8.2 kb might move a tissue-specific silencer of *dmr6-like* that is normally active in apical and axillary meristems and young developing leaves to a distance from the gene that abolishes silencer function, resulting in upregulation of *dmr6-like*. Alternatively, an enhancer being active in those tissues crucial for development might be moved into spatial proximity of *dmr6-like* due to an altered chromosome architecture (e.g. looping of the Gypsy-44 containing region) and thereby cause induction in tissues where *dmr6-like* is normally not expressed. We identified a candidate for a CNS that might have a role in gene regulation at 31.4 kb distance upstream of *dmr6-like*. This is reminiscent of the situation for the *vgt1* locus in maize, a CNS positioned 70 kb downstream of *ZmRap2.7*. *ZmRap2.7* is a repressor of flowering and its expression is downregulated in varieties that carry a miniature inverted repeat TE (MITE) insertion within *vgt1*, resulting in early flowering [55]. However, in our case the TE insertion does not occur at the CNS site, but downstream of it. This might still influence its function due to the increased distance between the CNS and *dmr6-like*. The fact that the region crucial for the formation of columnar growth, including *dmr6-like*, is conserved across several Rosaceae species might provide an opportunity for the creation not only of new columnar apple cultivars, but also of columnar pear and plum varieties by genetic engineering in future.

In *Arabidopsis*, *dmr6* loss-of-function mutants show enhanced expression of a subset of defense-associated genes, suggesting a control different from that in apple [56]. Interestingly, the truncated protein of the *dmr6-1* mutant is constitutively expressed in the shoot apical meristem and leaf primordia of *Arabidopsis* [56], in line with the expression profile of *dmr6-like* in columnar apple trees. The regulatory pathways influenced by columnar growth downstream of the *dmr6-like* upregulation still await disclosure. Since *dmr6-like* is associated with downy mildew resistance in *Arabidopsis* [56] it might trigger a defense response leading to a diseased-looking growth habit. On the other hand, members of the 2OG Fe(II) oxygenase family are involved in the biosynthesis of ethylene, gibberellins

and flavonoids, the latter of which modulate polar auxin transport, so that upregulation of *dmr6-like* might directly influence phytohormone concentrations [12, 57–60]. In apple, silencing of the chalcone synthase, the first committed enzyme in flavonoid biosynthesis, leads to decreased concentrations of flavonoids and causes altered cellular organization and cell wall composition, increased auxin transport and a reduction of plant size due to shortened internodes [61]. Specifically some genes encoding indole-3-acetic acid (IAA) transporters show upregulation in shoot apical meristem of P28 compared with A14, in line with an altered auxin transport capacity [13].

Furthermore, besides the induction of *dmr6-like*, other gene expression changes in the Gypsy-44 region or the upregulation of Gypsy-44 itself in shoot apical meristems might play a role in phenotype generation. Induction of *MDP0000163720* in columnar shoot apical meristems, which encodes a homolog of Aminocyclopropane-1-carboxylate oxidase, might lead to increased levels of ethylene. Since ethylene crosstalks with many other phytohormones such as auxin, cytokinin, abscisic acid, jasmonates and salicylic acid (for example [62–66]), this could contribute to the altered hormonal balance in columnar trees.

Conclusions

Open pollination of heterozygous columnar apple trees yield non-columnar, heterozygous columnar and homozygous columnar seedlings in a genotype ratio of 3 : 2 : 1. This suggests that homozygous columnar plants are viable within their early life stages and the lack of described homozygous columnar varieties is probably due to breeders' selection criteria. Apple primary roots display a transcriptome profile typical for a developing and actively growing tissue. About 200 – 300 genes are differentially regulated in columnar compared with non-columnar radicles, and these are involved in cell wall synthesis and modification, stress reactions and phytohormone signaling and biosynthesis. The molecular cause of columnar growth, the Gypsy-44 insertion, leads to upregulation of its nearest downstream gene, *dmr6-like*, in all tissues that are crucial to the development of the visible plant growth habit. The role of *dmr6-like* in pathogen resistance possibly results in upregulation of defense reaction in aerial plant organs leading to a phenotype reminiscent of a diseased plant. Alternatively or additionally, the upregulation of *dmr6-like* might lead to alterations in flavonoid content causing altered auxin transport, which would be in line with changes in auxin signaling and metabolism that have been detected in previous studies. The Gypsy-44 insertion is the causative event, then leading to upregulation of *dmr6-like* as a second step. In columnar primary roots, however, *dmr6-like* is downregulated, so that the activity of a tissue-specific transcriptional regulator might play a pivotal role. We are confident that our data will contribute to the elucidation of the whole signaling cascade causing columnar growth in apples.

Methods

Plant Material

For primary root transcriptome studies, apples from the heterozygous columnar cultivar P28 obtained by open pollination on a field with many columnar apple trees were collected in August 2011 at Geisenheim University. Seeds were extracted and dried at room temperature for about 1 – 2 weeks.

For qRT-PCRs, shoot tips of columnar P28 and non-columnar 'A14-190-93K' (A14) sampled on September 29, 2009 and very young apical leaves of 'McIntosh' and 'McIntosh Wijcik' sampled on July 19, 2013 were snap frozen in dry ice or liquid nitrogen, respectively, and stored at -80 °C until further use.

Seed germination

Dried seeds were swabbed with 70 % ethanol to roughly remove contamination and were placed in Petri dishes equipped with towel-paper that had been moistened with tap water. The petri dishes were wrapped in tinfoil and were stored in the dark at 4 °C for about 12 weeks. When the radicles reached a length of about 2 – 3 cm they were snap frozen in liquid nitrogen, cut off the seed, cut in half and stored at -80 °C until further use.

DNA Isolation

The upper half of each radicle was subjected to DNA extraction according to a manual protocol. One radicle was transferred into 500 µL Cetyltrimethyl ammonium bromide (CTAB) extraction buffer, to which 168 µL 8 M urea and a spatula tip of polyvinylpyrrolidone (PVPP) were added. The radicle was thoroughly homogenized with a micropestle followed by incubation at 65 °C for 30 min. After centrifugation (10 min at 13,000 rpm and room temperature) the supernatant was purified by three chloroform isoamyl alcohol (24:1) extractions and subsequent ethanol precipitation.

Primary Root Genotyping

In order to investigate the genotype of each radicle with regard to the presence of the Gypsy-44 retrotransposon associated with the columnar growth habit, PCRs were performed with primers spanning the left border, the right border or the whole TE insertion as described by Otto et al. [10]. Results were double checked by PCRs with our indel-based markers I2_3_M1 and H1_M1 [37]. All primers were purchased from Invitrogen™ (Invitrogen, Darmstadt, Germany). PCRs were carried out in 50 µL reaction volume containing at least 10 ng of radicle DNA and 1 U Go-Taq polymerase (Promega, Madison, United States) as follows: initial denaturation at 94 °C for 5 min followed by 40 amplification cycles of denaturation at 94 ° C for 30 sec, annealing at 58.4 °C or 60 °C for 30 sec and elongation at 72 °C for 30 sec. A final elongation step at 72 °C for 10 min was included. PCR products were analyzed on horizontal 2 % agarose gels.

RNA Extraction

The distal halves (tips) of radicles were used for RNA extraction. For the initial Illumina RNA-Seq, one radicle of each genotype was subjected to total RNA isolation. For replicate datasets, three frozen radicles of one genotype were pooled prior to RNA extraction. Total

RNA was isolated using the innuPREP Plant RNA Kit (Analytik Jena, Jena, Germany) following the manufacturer's instructions. A DNase I (Fermentas, St. Leon-Roth, Germany) digestion step on the column was included to remove genomic DNA. RNA concentration and integrity were assessed with Bioanalyzer RNA Nano or Pico chips (Agilent, Santa Clara, USA).

For qRT-PCRs with leaf and shoot tip material, samples were transferred to liquid nitrogen, disrupted with a mortar and pestle and RNA was isolated using the innuPREP Plant RNA kit (Analytik Jena, Jena, Germany) following the manufacturer's instructions.

Illumina Sequencing

For RNA-Seq library construction 5 µg of high-quality total RNA of each primary root sample were sent to GENterprise Genomics (Mainz, Germany). Subsequently, 100 bp paired-end runs were conducted on the Illumina HiSeq 2000 or Illumina HiSeq 2500 (Illumina, San Diego, USA) of the IMSB Mainz.

Assembly of Illumina Data

Prior to assembly, Illumina raw data were trimmed with an in-house Perl script to remove low quality stretches. 6 bp at the 5' end and 5 bp at the 3' end were cut off, and bases with a Phred score < 20 or called as "N" were removed. After trimming, only sequences with a minimum length of 30 bp were maintained. All remaining sequences were assembled with the CLC Assembly Cell (CLC bio, Aarhus, Denmark) using different k-mer sizes and a bubble size of 300. The assemblies yielding the highest N50 value were used for BLAST searches. To evaluate these assemblies, trimmed reads were mapped back to contigs with Bowtie [67], allowing no mismatches.

BLAST searches

Contigs generated by assembling were subjected to BLAST searches [33] against 1) MDPs representing the genes annotated in the Golden Delicious genome sequence [36] as extracted from the Genome Database of Rosaceae (http://www.rosaceae.org/), 2) the NCBI Malus expressed sequence tag (EST) dataset, 3) the Malus Unigene v5.0 dataset representing filtered ESTs assembled with CAP3 [68], available at the Genome Database of Rosaceae (http://www.rosaceae.org/), and 4) the NCBI SwissProt/UniProtKB database, all databases downloaded on February 3, 2014. BLAST results were filtered and analyzed at an E-Value cutoff of 10^{-5}.

Blast2GO

Results of BLAST searches with the replicate datasets against SwissProt/UniProtKB were imported into CLC Genomics Workbench v6.5 (CLC bio, Aaarhus, Denmark) and analyzed with the Blast2GO plugin using default parameters [34, 35]. Following GO Slim reduction, combined graphs of the GO term biological process were created and were converted to pie charts of the second-level terms.

RNA-Seq Analysis

Illumina raw sequencing data were imported into CLC Genomics Workbench v6.5 (CLC bio, Aaarhus, Denmark) as paired-end reads and were subjected to trimming and RNA-Seq mappings against our BAC metacontig of the transposon region or the Golden Delicious

reference genome as previously described [10]. A similarity of at least 98 % was demanded for at least 90 % of the read length and using the "include broken pairs" counting scheme. For visualization of differential gene expression between two datasets each, total gene read counts for each MDP were extracted from RNA-Seq mapping results and analyzed in DESeq [38] with default parameters and including the biological replicates. Only genes with base mean > 100, fold change > 2 (or < 0.5) and p-value < 0.05 were maintained, and genes were ranked according to their p-values. Subsequently, the LOG2 fold changes were loaded into MapMan [39] and mapped against the *Malus* x *domestica* mapping file.

qRT-PCRs

Genes showing interesting differential regulation were subjected to qRT-PCR analysis in order to verify the bioinformatics results. Primers were designed to span the last intron of each gene and were purchased from Invitrogen (Darmstadt, Germany). A list of primers can be found in Additional File 7. 800 ng of total RNA extracted from three pooled primary roots of each genotype (RIN > 7) and 200 ng of human MDA-MB468 RNA serving as exogenous control were converted to cDNA using SuperScript III Reverse Transcriptase (Invitrogen, Darmstadt, Germany) and 100 µM oligo$(dT)_{18}$VN primer according to the manufacturer's instructions. In order to compare results of different tissues, the same procedure was applied to RNA extracted from shoot tips (shoot apical meristems) of A14 and P28 and to RNA isolated from young leaves of McIntosh and McIntosh Wijcik. All qRT-PCR experiments were carried out in a 7500 Fast Real-Time PCR System using Power SYBR Green PCR Master Mix (Applied Biosystems, Carlsbad, USA). Reactions were conducted in triplicate and two biological replicates were analyzed.

Conserved Sequence Analysis

Orthologs were retrieved from pear, peach, Chinese plum (Mei) and strawberry based on BLASTx searches yielding PCP006734.1 and PCP039177.1 (pear), ppa009607m and ppa014841m (peach), Pm020608 and Pm020609 (Chinese plum) and mrna08066.1-v1.0-hybrid and mrna08065.1-v1.0-hybrid (strawberry) as the best hits for At1g08530-like and dmr6-like, respectively. The corresponding genomic sequences were extracted from GDR and were aligned in mVISTA (http://genome.lbl.gov/vista/index.html) [69] with parameters "calculated window" of 100 bp, "minimum consensus width" 100 bp and "consensus identity" 70 %.

List of Abbreviations

ACC1, 1-aminocyclopropane-1-carboxylate synthase 1; BLAST, Basic Local Alignment Search Tool; CLE40, Clavata3/Embryo surrounding region 40; *Co*, *Columnar*; CNS, conserved non-coding sequence; CTAB, cetyltrimethyl ammonium bromide; *dmr6*, *downy mildew resistant 6*; EST, expressed sequence tag; GO, Gene Ontology; IAA, indole-3 acetic acid; LTR, long terminal repeat; MDP, *Malus* x *domestica* protein; PIN, Pinformed; PP2C, protein phosphatase 2C; PVPP, polyvinylpyrrolidone; qRT-PCR, quantitative real-time polymerase chain reaction; WOX5, Wuschel-related homeobox 5

Competing Interest

The authors declare no competing interest.

Authors' contributions

RP carried out differential gene expression analysis and qRT-PCRs and drafted the manuscript. HD participated in RNA-Seq analysis. BR conducted assemblies and BLAST searches. BR and SR analyzed assemblies and provided bioinformatics support. ERS initiated the project and supervised the project throughout. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the Institutes of Botany and Pomology of the Geisenheim University for providing Procats 28 apples, Clemens Krost for advice on DESeq and qRT-PCR analysis and Ramona Petersen for linguistic proofreading. We are grateful to Oliver Dautermann, Inessa Penner and Nina Seiwert for help with genotyping and RNA extraction. This work was supported by grants of the Federal Ministry of Agriculture and Nutrition (no. 511-06.01-28-1-43.042-07 and no. 313-06.01-28-1-43.042-07).

References

1. Lapins KO: Segregation of Compact Growth Types in Certain Apple Seedling Progenies. *Can J Plant Sci* 1969, 49:765–768.
2. Tobutt KR: Breeding Columnar Apple Trees at East Malling. *Acta Hortic* 1985, 159:63–68.
3. Gelvonauskienë D, Gelvonauskis B, Sasnauskas A: Impact of Rootstock on Columnar Apple Tree Growth in a Nursery. *Sci Work Lith Inst Hortic Lith Univ Agric* 2006, 25:51–56.
4. Meulenbroek J, Verhaegh J, Janse J: Inheritance Studies with Columnar Type Trees. *Acta Hortic* 1998, 484:255–258.
5. Moriya S, Iwanami H, Kotoda N, Takahashi S, Yamamoto T, Abe K: Development of a Marker-Assisted Selection System for Columnar Growth Habit in Apple Breeding. *J Japanese Soc Hortic Sci* 2009, 78:279–287.
6. Lauri PE, Lespinasse JM: The Relationship between Cultivar Fruiting-Type and Fruiting Branch Characteristics in Apple Trees. *Acta Hortic* 1993, 349:259–263.
7. Petersen R, Krost C: Tracing a key player in the regulation of plant architecture: the columnar growth habit of apple trees (*Malus* × *domestica*). *Planta* 2013, 238:1–22.
8. Fisher D V: Spur-type Strains of McIntosh for High Density Plantings. *Bristish Columbia Fruit Grow Assoc Quaterly Rep* 1969, 14:3–10.
9. Fisher D V: The "Wijcik Spur Mclntosh." *Fruit Var J* 1995, 49:212–213.
10. Otto D, Petersen R, Brauksiepe B, Braun P, Schmidt ER: The columnar mutation ("*Co* gene") of apple (*Malus* × *domestica*) is associated with an integration of a *Gypsy*-like retrotransposon. *Mol Breed* 2014, 33:863–880.
11. Tian Y-K, Wang C-H, Zhang J-S, James C, Dai H-Y: Mapping *Co*, a Gene Controlling the Columnar Phenotype of Apple, with Molecular Markers. *Euphytica* 2005, 145:181–188.

12. Wolters PJ, Schouten HJ, Velasco R, Si-Ammour A, Baldi P: Evidence for regulation of columnar habit in apple by a putative 2OG-Fe(II) oxygenase. *New Phytol* 2013, 200:993–999.
13. Krost C, Petersen R, Lokan S, Brauksiepe B, Braun P, Schmidt ER: Evaluation of the hormonal state of columnar apple trees (*Malus* × *domestica*) based on high throughput gene expression studies. *Plant Mol Biol* 2013, 81:211–220.
14. Krost C, Petersen R, Schmidt ER: The Transcriptomes of Columnar and Standard Type Apple Trees (*Malus* × *domestica*) – a Comparative Study. *Gene* 2012, 498:223-230.
15. Darwin CR (assisted by DF): *The Power of Movement in Plants*. London: John Murray; 1880.
16. Baluska F, Mancuso S, Volkmann D, Barlow PW: The "root-brain" hypothesis of Charles and Francis Darwin: Revival after more than 125 years. *Plant Signal Behav* 2009, 4:1121–1127.
17. Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN: A Gene Expression Map of the *Arabidopsis* Root. *Science* 2003, 302:1956–1960.
18. Nawy T, Lee J-Y, Colinas J, Wang JY, Thongrod SC, Malamy JE, Birnbaum K, Benfey PN: Transcriptional Profile of the *Arabidopsis* Root Quiescent Center. *Plant Cell* 2005, 17:1908–1925.
19. Brady SM, Orlando DA, Lee J-Y, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN: A High-Resolution Root Spatiotemporal Map Reveals Dominant Expression Patterns. *Science* 2007, 318:801–806.
20. Dembinsky D, Woll K, Saleem M, Liu Y, Fu Y, Borsuk LA, Lamkemeyer T, Fladerer C, Madlung J, Barbazuk B, Nordheim A, Nettleton D, Schnable PS, Hochholdinger F: Transcriptomic and Proteomic Analyses of Pericycle Cells of the Maize Primary Root. *Plant Physiol* 2007, 145:575–588.
21. Hoecker N, Lamkemeyer T, Sarholz B, Paschold A, Fladerer C, Madlung J, Wurster K, Stahl M, Piepho H-P, Nordheim A, Hochholdinger F: Analysis of nonadditive protein accumulation in young primary roots of a maize (Zea mays L.) F1-hybrid compared to its parental inbred lines. *Proteomics* 2008, 8:3882–3894.
22. Kohler A, Dalaruelle C, Martin D, Encelot N, Martin F: The poplar root transcriptome: analysis of 7000 expressed sequence tags. *FEBS Lett* 2003, 542:37–41.
23. Eichholtz DA, Robitaille HA, Herrmann KM: Protein Changes during the Stratification of *Malus domestica* Borkh. Seed. *Plant Physiol* 1983, 72:750–753.
24. Bogatek R, Dziewanowska K, Lewak S: Hydrogen cyanide and embryonal dormancy in apple seeds. *Physiol Plant* 1991, 83:417–421.
25. Gniazdowska A, Dobrzyńska U, Babańczyk T, Bogatek R: Breaking the apple embryo dormancy by nitric oxide involves the stimulation of ethylene production. *Planta* 2007, 225:1051–1057.
26. Gniazdowska A, Krasuska U, Debska K, Andryka P, Bogatek R: The beneficial effect of small toxic molecules on dormancy alleviation and germination of apple embryos is due to NO formation. *Planta* 2010, 232:999–1005.
27. Gniazdowska A, Krasuska U, Bogatek R: Dormancy removal in apple embryos by nitric oxide or cyanide involves modifications in ethylene biosynthetic pathway. *Planta* 2010, 232:1397–1407.
28. Dębska K, Krasuska U, Budnicka K, Bogatek R, Gniazdowska A: Dormancy removal of apple seeds by cold stratification is associated with fluctuation in H_2O_2, NO production and protein carbonylation level. *J Plant Physiol* 2013, 170:480–488.
29. Petricka JJ, Winter CM, Benfey PN: Control of *Arabidopsis* root development. *Annu Rev Plant Biol* 2012, 63:563–590.
30. Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S, Roberts K, Scheres B: Cellular organisation of the *Arabidopsis thaliana* root. *Development* 1993, 119:71–84.
31. Van den Berg C, Weisbeek P, Scheres B: Cell fate and cell differentiation status in the *Arabidopsis* root. *Planta* 1998, 205:483–491.
32. Van den Berg C, Willemsen V, Hendriks G, Weisbeek P, Scheres B: Short-range control of cell differentiation in the *Arabidopsis* root meristem. *Nature* 1997, 390:287–289.
33. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic Local Alignment Search Tool. *J Mol Biol* 1990, 215:403–410.
34. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005, 21:3674–3676.
35. Conesa A, Götz S: Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. *Int J Plant Genomics* 2008, 2008:619832.
36. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D,

Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, et al.: The Genome of the Domesticated Apple (*Malus* × *domestica* Borkh.). *Nat Genet* 2010, 42:833–839.
37. Otto D, Petersen R, Krost C, Schmidt ER, Brand R, Brauksiepe B, Braun P: Molecular Characterization of the *Co* Gene Region in *Malus* × *domestica*. *Acta Hortic*, under review.
38. Anders S, Huber W: Differential Expression Analysis for Sequence Count Data. *Genome Biol* 2010, 11:R106.
39. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M: Mapman: a User-Driven Tool to Display Genomics Data Sets onto Diagrams of Metabolic Pathways and Other Biological Processes. *Plant J* 2004, 37:914–939.
40. Krost C, Petersen R, Schmidt ER: The Transcriptomes of Columnar and Standard Type Apple Trees (*Malus* × *domestica*) - a Comparative Study. *Gene* 2012, 498:223–230.
41. Kohany O, Gentles AJ, Hankus L, Jurka J: Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 2006, 7:474.
42. Lapins KO: Inheritance of Compact Growth Type in Apple. *J Am Soc Hortic Sci* 1976, 101:133–135.
43. Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, Chen NJ, Nishio T, Xu X, Cong L, Qi K, Huang X, Wang Y, Zhao X, Wu J, Deng C, Gou C, Zhou W, Yin H, Qin G, Sha Y, Tao Y, Chen H, Yang Y, Song Y, Zhan D, et al.: The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res* 2013, 23:396–408.
44. Initiative TIPG: The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* 2013, 45:487–494.
45. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen R V, Allan AC, Michael TP, Setubal JC, Celton J-M, Rees DJG, Williams KP, Holt SH, Ruiz Rojas JJ, Chatterjee M, et al.: The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* 2011, 43:109–116.
46. Newcomb RD, Crowhurst RN, Gleave AP, Rikkerink EHA, Allan AC, Beuning LL, Bowen JH, Gera E, Jamieson KR, Janssen BJ, Laing WA, Mcartney S, Nain B, Ross GS, Snowden KC, Souleyre EJF, Walton EF, Yauk Y-K: Analyses of Expressed Sequence Tags from Apple. *Plant Physiol* 2006, 141:147–166.
47. Allan AC, Crowhurst R, Gleave A, Newcomb R, Schaffer R: Apple Functional Genomics. In *Genet Genomics Rosaceae*. Edited by Folta KM, Gardiner SE.; 2009:121–142.
48. Sanzol J: Dating and functional characterization of duplicated genes in the apple (*Malus domestica* Borkh.) by analyzing EST data. *BMC Plant Biol* 2010, 10:87.
49. Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, Harper JF, Zhu J-K, Cushman JC, Gollery M, Girke T: Annotating genes of known and unknown function by large-scale coexpression analysis. *Plant Physiol* 2008, 147:41–57.
50. Kvam VM, Liu P, Si Y: A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am J Bot* 2012, 99:248–56.
51. Blazek J: Segregation and General Evaluation of Spur Type or Compact Growth Habits in Apples. *Acta Hortic* 1992, 317:71–79.
52. Baldi P, Wolters PJ, Komjanc M, Viola R, Velasco R, Salvi S: Genetic and physical characterisation of the locus controlling columnar habit in apple (*Malus* × *domestica* Borkh.). *Mol Breed* 2012, 31:429–440.
53. Moriya S, Okada K, Haji T, Yamamoto T, Abe K: Fine Mapping of *Co*, a Gene Controlling Columnar Growth Habit Located on Apple (*Malus* × *domestica* Borkh.) Linkage Group 10. *Plant Breed* 2012, 131:641-647.
54. Zhang Y, Dai HY: Comparison of Photosynthetic and Morphological Characteristics, and Micro-Structure of Roots and Shoots, between Columnar Apple and Standard Apple Trees of Hybrid Seedlings. *Phyton* 2011, 80:119–125.
55. Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev E V, Svitashev S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski J-A, Tingey S V, Miao G-H, Phillips RL, Tuberosa R: Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci U S A* 2007, 104:11376–81.
56. Van Damme M, Huibers RP, Elberse J, Van den Ackerveken G: *Arabidopsis DMR6* encodes a putative 2OG-Fe(II) oxygenase that is defense-associated but required for susceptibility to downy mildew. *Plant J* 2008, 54:785–93.
57. Prescott AG, John P: DIOXYGENASES: Molecular Structure and Role in Plant Metabolism. *Annu Rev Plant Physiol Plant Mol Biol* 1996, 47:245–271.

58. Brown DE, Rashotte AM, Murphy AS, Normanly J, Tague BW, Peer WA, Taiz L, Muday GK: Flavonoids Act as Negative Regulators of Auxin Transport in Vivo in *Arabidopsis*. *Plant Physiol* 2001, 126:524–535.
59. Peer WA, Murphy AS: Flavonoids and auxin transport: modulators or regulators? *Trends Plant Sci* 2007, 12:556–563.
60. Buer CS, Imin N, Djordjevic MA: Flavonoids: New Roles for Old Molecules. *J Integr Plant Biol* 2010, 52:98–111.
61. Dare AP, Tomes S, Jones M, McGhie TK, Stevenson DE, Johnson RA, Greenwood DR, Hellens RP: Phenotypic changes associated with RNA interference silencing of chalcone synthase in apple (*Malus* × *domestica*). *Plant J* 2013, 74:398–410.
62. Chae HS, Cho YG, Park MY, Lee MC, Eun MY, Kang BG, Kim WT: Hormonal cross-talk between auxin and ethylene differentially regulates the expression of two members of the 1-aminocyclopropane-1-carboxylate oxidase gene family in rice (*Oryza sativa* L.). *Plant Cell Physiol* 2000, 41:354–362.
63. Stepanova AN, Yun J, Likhacheva AV, Alonso JM: Multilevel interactions between ethylene and auxin in *Arabidopsis* roots. *Plant Cell* 2007, 19:2169–2185.
64. El-Showk S, Ruonala R, Helariutta Y: Crossing paths: cytokinin signalling and crosstalk. *Development* 2013, 140:1373–1383.
65. Leon-Reyes A, Spoel SH, De Lange ES, Abe H, Kobayashi M, Tsuda S, Millenaar FF, Welschen RAM, Ritsema T, Pieterse CMJ: Ethylene Modulates the Role of NONEXPRESSOR OF PATHOGENESIS-RELATED GENES1 in Cross Talk Between Salicylate and Jasmonate Signaling. *Plant Physiol* 2009, 149:1797–1809.
66. Ghassemian M, Nambara E, Cutler S, Kawaide H, Kamiya Y, McCourt P: Regulation of abscisic acid signaling by the ethylene response pathway in *Arabidopsis*. *Plant Cell* 2000, 12:1117–1126.
67. Langmead B, Trapnell C, Pop M, Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009, 10:R25.
68. Huang X, Madan A: CAP3: A DNA Sequence Assembly Program. *Genome Res* 1999, 9:868–877.
69. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I: VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 2004, 32(Web Server issue):W273–W279.

Additional files

These files are included in the electronic supplementary material accompanying this thesis.

Additional file 1 (png) – Pie charts of level 2 GO terms assigned to contigs of primary root transcriptome assemblies.
Assemblies of primary root transcriptome Illumina reads were subjected to BLAST searches against SwissProt/UniProtKB and subsequent Blast2GO analysis. Pie charts
of level 2 GO terms are shown for data obtained from three pooled non-columnar (A), heterozygous columnar (B) and homozygous columnar (C) primary roots.

Additional file 2 (xlsx) – Table of significantly differentially expressed genes in heterozygous compared with non-columnar primary roots
Total gene reads of the two biological replicate data were imported into DESeq and analysed for differential expression. Only genes with base mean > 100, fold change > 2 and p-value < 0.05 are displayed.

Additional file 3 (xlsx) – Table of significantly differentially expressed genes in homozygous columnar compared with non-columnar primary roots
Total gene reads of the two biological replicate data were imported into DESeq and analyzed for differential expression. Only genes with base mean > 100, fold change > 2 and p-value < 0.05 are displayed.

Additional file 5 (xlsx) – Table of significantly differentially expressed genes in homozygous columnar compared with heterozygous primary roots

Total gene reads of the two biological replicate data were imported into DESeq and analyzed for differential expression. Only genes with base mean > 100, fold change > 2 and p-value < 0.05 are displayed.

Additional file 6 (xlsx)– Table of total read counts and fold changes of genes and Gypsy-44 within the *Co* target region.
Total Read counts of RNA-Seq data are displayed.

Additional file 7 (xlsx) – Primer List
Primers used for marker PCRs and qRT-PCR.

Electronic Supplementary Material

The electronic supplementary material (attached as a CD) contains the following files:

- **Paper 3:**
  - Supplementary Table 2

- **Paper 4:**
  - Additional file 1
  - Additional file 2
  - Additional file 4
  - Additional file 5
  - Additional file 7
  - Additional file 8

- **Paper 6:**
  - Online Resource 1

- **Paper 7:**
  - Additional file 1
  - Additional file 2
  - Additional file 3
  - Additional file 5
  - Additional file 6
  - Additional file 7

- a pdf version of this thesis

Curriculum Vitae

Dipl. Biol. Romina Petersen

Acknowledgements

Eidesstattliche Erklärung

Hiermit erkläre ich, die vorliegende Dissertation selbstständig und ohne fremde Hilfe angefertigt zu haben. Weiterhin versichere ich, örtlich übernommene Ausführungen anderer Autoren und an Gedankengänge Anderer anlehnende eigene Formulierungen entsprechend gekennzeichnet und die Quellen zitiert zu haben.

---------- ----------

(Ort, Datum) (Romina Petersen)