The family of *CONSTANS*-like genes in the moss *Physcomitrella patens*

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät der
Universität zu Köln

vorgelegt von
Oliver Zobell
aus Wilrijk (Antwerpen), Belgien

Köln, 2005

Diese Arbeit wurde am

Max-Planck-Institut für Züchtungsforschung in Köln

in der Abteilung Entwicklungbiologie der Pflanzen

angefertigt.

Berichterstatter:

Prof. Dr. Ulf-Ingo Flügge

Prof. Dr. George Coupland

Tag der mündlichen Prüfung: 15.02.06

TABLE OF CONTENTS

ABBREVIATIONS

A. tumefaciens	*Agrobacterium tumefaciens*
bHLH	basic Helix Loop Helix
bp	basepair
bZIP	basic-leucine zipper
CaMV	*Cauliflower Mosaic Virus*
ccg	clock-controlled gene
CCT	CO, CO-like and TOC1
cDNA	complementary deoxyribonucleic acid
CO	CONSTANS
COL	CONSTANS-like
CRY1-2	CRYPTOCHROME 1-2
DNA	deoxyribonucleic acid
DSB	double-strand break
E. coli	*Escherichia coli*
EST	expressed sequence tag
FT	FLOWERING LOCUS T
GFP	green fluorescent protein
GI	GIGANTEA
GUS	β-glucuronidase
hnRNA	heterogenous nuclear ribonucleic acid
hpt	hygromycin phosphotransferase
HR	homologous recombination
kb	kilobasepair
Lhcb	light harvesting chlorophyll A/B binding
M1-4	middle domain 1-4
mRNA	messenger ribonucleic acid
Mya	million years ago
NHEJ	non-homologous end joining
NLS	nuclear localisation signal
nos	nopaline synthase
nptII	neomycin phosphotransferase II
PCR	polymerase chain reaction
PEG	polyethylene glycol

PHYA-E	PHYTOCHROME A-E
rDNA	ribosomal deoxyribonucleic acid
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
RT-PCR	reverse transcribed polymerase chain reaction
SAM	shoot apical meristem
sul	sulfadiazine resistance gene
T	terminal domain
Ta	annealing temperature
UTR	untranslated region
WT	wild type
XNF7	XENOPUS NUCLEAR FACTOR 7

1 INTRODUCTION

1.1 *CONSTANS* is a regulator of photoperiodic flowering

A major decision in the life of flowering plants is the switch from vegetative development to reproductive development. *Arabidopsis*, for example, initially produces leaf after leaf, and then switches to producing stalk-borne flowers upon flowering. In order to achieve reproduction in the most favourable conditions, flowering is controlled by both environmental and endogenous stimuli. In *Arabidopsis*, environmental cues include day length, light quality and temperature, whereas endogenous signals include developmental age and the hormone gibberellic acid (Boss *et al.*, 2004). At the molecular level, each type of signal is transmitted through its own distinct flowering pathway, only to converge upon integration by the so-called floral pathway integrators, *FLOWERING LOCUS T* (*FT*), *SUPPRESSOR OF CONSTANS 1* (*SOC1*), and *LEAFY* (*LFY*) (Parcy, 2005). These genes convert the heterogeneous inputs from different flowering pathways into an induction of the so-called meristem identity genes, the action of which results in the conversion of the shoot apical meristem (SAM) from a vegetative, to an inflorescence and finally to a floral meristem (Lohmann and Weigel, 2002).

An important environmental signal that influences flowering is day length, or photoperiod. Because at intermediate and higher latitudes, photoperiod changes dramatically during the course of the year, plants use it to sense seasonal progression and regulate their reproductive strategies accordingly. *Arabidopsis*, for example, is a facultative long-day plant, which means that flowering is very much accelerated under long days, although it will eventually also flower under short days. Over the years, a large number of *Arabidopsis* mutants have been characterised that are impaired in proper day length dependent flowering. Some mutants fail to promote flowering under inductive photoperiod (long days), whereas others promote flowering constitutively, also under non-inductive photoperiod (short days). As the corresponding genes were cloned and their hierarchy of action dissected, it was revealed that a gene named *CONSTANS* (*CO*) occupies the most basal

position in the signalling cascade of the photoperiod flowering pathway (Figure 1). As such, *CO* relays the day length signal to the floral pathway integrators *FT* and *SOC1*, which on turn integrate it with signals from the other flowering pathways. In particular the direct activation of *FT* expression by *CO* is established as the decisive step in the day length dependent transition to flowering.

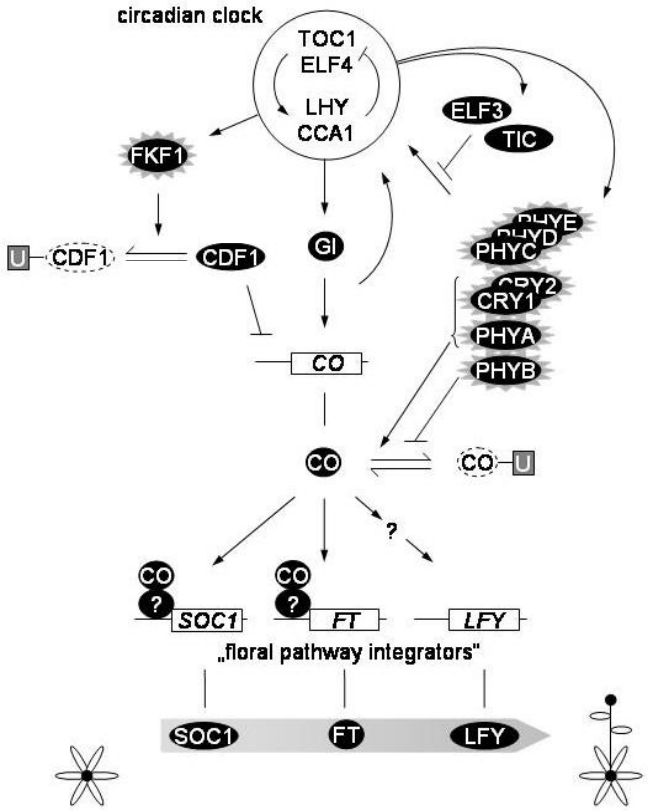


Figure 1 Genetic pathway that controls flowering of *Arabidopsis* in response to photoperiod. Genes are shown as white rectangles with a black line to represent their promoter. Proteins are shown as black circles or ovals; proteins known or believed to perceive light are indicated by an additional grey star. The grey U boxes represent ubiquitin; degradation of that protein is indicated by a dashed line. Arrows represent promotive effects; T-bars represent repressive effects. Adapted from Ausin *et al.* (2005).

Early grafting experiments showed that photoperiod is sensed by the leaves, consistent with their evolutionary function to harvest light for photosynthesis. The expression of *CO* and *FT* specifically overlaps in the vascular tissue of leaves (Takada and Goto, 2003; An *et al.*, 2004). Yet, photoperiod affects flowering and the fate of the SAM; therefore the flowering signal generated by *CO* and *FT* must be transmitted to the SAM. *CO* activates *FT* and promotes flowering cell-autonomously in the phloem of leaves, whereas *FT* promotes flowering in a non cell-autonomous fashion (An *et al.*,

2004; Ayre and Turgeon, 2004). In fact, recent evidence suggests that *FT* mRNA may be part of the mobile floral stimulus (Huang *et al.*, 2005), and that the FT protein acts in the shoot apex through interaction with a bZIP transcription factor called FD (FD) (Abe *et al.*, 2005; Wigge *et al.*, 2005). Nevertheless, the flowering signal generated by *CO* in the leaves is likely to be more complex, because phloem specific expression of *CO* still accelerated flowering in the absence of a functional *FT* allele (An *et al.*, 2004).

Already a long time ago it was recognised that photoperiodism must be associated with the workings of a circadian clock (Bünning, 1936). At the molecular level, the role of the circadian clock becomes apparent in the circadian fluctuations in *CO* transcript abundance (Figure 2). Moreover, several mutations that disrupt circadian clock function also impair proper circadian expression of *CO*, as well as proper day length dependent flowering. These include mutations in genes that constitute the core of the circadian clock, such as *TIMING OF CHLOROPHYLL A/B BINDING PROTEIN 1* (*TOC1*), *EARLY FLOWERING 4* (*ELF4*), *LATE ELONGATED HYPOCOTYL* (*LHY*), and *CIRCADIAN CLOCK ASSOCIATED* (*CCA1*), or genes that are required for regulating the input from the photoreceptors to the clock, such as *EARLY FLOWERING 3* (*ELF3*) (Figure 1) (Hicks *et al.*, 1996; Schaffer *et al.*, 1998; Wang and Tobin, 1998; Somers *et al.*, 2000; Strayer *et al.*, 2000; Doyle *et al.*, 2002). Two genes mediate between the circadian clock and *CO*: *GIGANTEA* (*GI*) and *FLAVIN-BINDING*, *KELCH REPEAT*, *F-BOX 1* (*FKF1*) (Figure 1). *GI* encodes a protein of unknown biochemical function that localises to the nucleus (Huq *et al.*, 2000; Mizoguchi *et al.*, 2005). In the late flowering *gi* mutant, abundance of *CO* mRNA is severely reduced, in long days as well as in short days (Figure 2) (Suarez-Lopez *et al.*, 2001; Mizoguchi *et al.*, 2005). Besides its role as a clock output gene in photoperiodic flowering, *GI* has an additional and unrelated role in circadian clock regulation (Mizoguchi *et al.*, 2005). The other gene, *FKF1*, encodes a multidomain protein. The presence of a functional light-sensing LOV domain indicated that the protein may function as a blue-light receptor (Nelson *et al.*, 2000; Imaizumi *et al.*, 2003), whereas the F-box domain suggested a role in ubiquitinating proteins for degradation (Vierstra, 2003). The latter was confirmed when the FKF1 protein was found to target a repressor of *CONSTANS*, the Dof transcription factor CYCLING DOF FACTOR 1 (CDF1), for degradation, allowing *CO* mRNA levels to rise at the end of the day (Imaizumi

et al., 2005). Together, *GI* and *FKF1* are required to generate the typical circadian rhythm of *CO* mRNA abundance, resulting in high levels of *CO* mRNA coinciding with light in long days, but with darkness in short days (Figure 2).

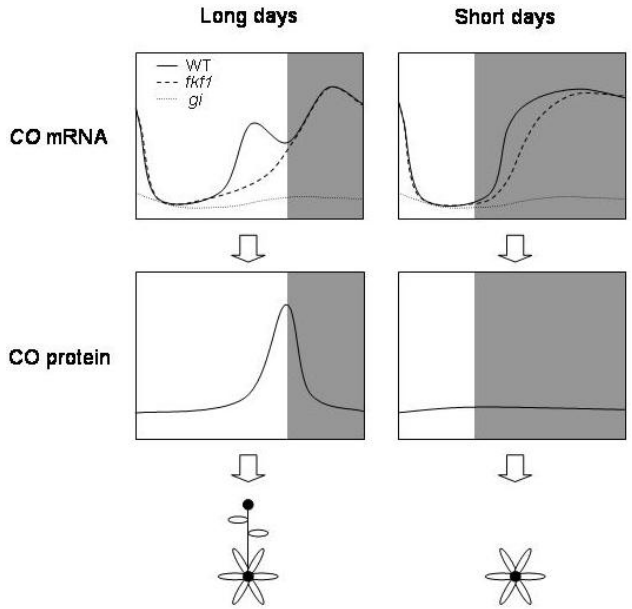


Figure 2 Current working model for the photoperiodic control of flowering in *Arabidopsis*. Top row: circadian profile of *CO* transcript abundance in WT, *fkf1* and *gi* plants. Bottom row: circadian profile of CO protein abundance. Left column: long days. Right column: short days. Adapted from Imaizumi *et al.* (2003), and Searle and Coupland (2004).

The importance of light perception for photoperiodic flowering is illustrated by the flowering time phenotypes of *Arabidopsis* photoreceptor mutants. Specifically, proper photoperiodic flowering depends on functional red/far-red light absorbing phytochromes and blue light absorbing cryptochromes (Koornneef *et al.*, 1991; Halliday *et al.*, 1994; Devlin *et al.*, 1998; Guo *et al.*, 1998). Although photoreceptors play an important role in synchronising the circadian clock to environmental light/dark cycles (Figure 1), it is not through an effect on circadian clock function that they influence flowering time (Suarez-Lopez *et al.*, 2001; Yanovsky and Kay, 2002; Cerdan and Chory, 2003). Instead, they were shown to have a direct, post-transcriptional effect on the stability of the CONSTANS protein. Whereas signalling through PHYTOCHROME B (PHYB) targets CO for degradation by the proteasome, signalling through PHYTOCHROME A (PHYA), CRYPTOCHROME 1 (CRY1) and CRYPTOCHROME 2 (CRY2) stabilises the protein (Figure 1) (Valverde *et al.*, 2004). How these antagonistic effects are balanced is not

entirely clear, but shifts in the light spectrum throughout the day might play a role, as well as distinct circadian cycles of photoreceptor abundance, due to transcriptional (Toth *et al.*, 2001; Sharrock and Clack, 2002; Mockler *et al.*, 2003), or post-transcriptional effects (El-Din El-Assal *et al.*, 2001).

In conclusion, the current data support that flowering is triggered by coincidence of light and a light inducible phase of the circadian cycle, which is provided by *CO* expression. Only when a peak in *CO* expression coincides with light, as is the case in long days but not in short days, will the protein accumulate and act to promote flowering (Figure 2). As such, this confirms one of the classical models of photoperiodic responses, the so-called external coincidence model, at the molecular level (Pittendrigh and Minis, 1964).

1.2 CONSTANS contains two conserved domains

Functionally, the CONSTANS protein is perceived to be a transcription factor, because it quickly and directly activates the expression of several genes, including the floral pathway integrators *FT* and *SOC1* (Samach *et al.*, 2000). This is in agreement with the protein being located in the nucleus (Robson *et al.*, 2001; Valverde *et al.*, 2004). However, there is as yet no demonstration of CONSTANS binding DNA. Hence CO is thought to be recruited to target promoters through interaction with one or more DNA-binding proteins. Although the nature of this DNA-binding protein has remained elusive to date, the two conserved regions of CONSTANS, amino terminal B-box zinc fingers and carboxy terminal CCT (CO, CO-like, and TOC1) domain have both been implicated in protein-protein interactions.

The B-box is a class of zinc finger that was originally identified in a variety of animal proteins, participating in a wide range of cellular processes including regulation of gene expression, differentiation, and/or control of cell growth (Torok and Etkin, 2001). In these proteins, B-boxes are found in conjunction with various other motifs, such as the RING finger, coiled-coil, rfp, or NHL motif (Torok and Etkin, 2001). In *Arabidopsis*, B-box zinc fingers were also found in SALT TOLERANCE (STO) and SALT TOLERANCE HOMOLOGOUS (STH), proteins that lack the CCT domain and that have been suggested to link light signalling with Ca^{2+}-signalling (Lippuner *et al.*, 1996; Holm *et al.*, 2001). Therefore, the B-box zinc finger seems to be an element with an

independent function, which has been combined with other domains in both the animal and plant kingdoms. Although no function has been clearly assigned, in animal proteins it is believed to function in protein-protein interactions, either directly, or indirectly by orienting the true interaction domain (Cao *et al.*, 1997; Cao *et al.*, 1998; Cainarca *et al.*, 1999). Four residues that were shown to bind zinc in the B-box structure of the XENOPUS NUCLEAR FACTOR 7 (XNF7) protein (Borden *et al.*, 1995) are found at these positions in the CONSTANS B-boxes as well. The importance of the B-boxes for CONSTANS function was demonstrated by the fact that five out of seven classical *co* mutant alleles contain mutated B-box residues (Robson *et al.*, 2001). According to the InterPro database of protein domains (www.ebi.ac.uk/interpro), B-box type zinc fingers are also found in the kingdoms of archaea and fungi, indicating very ancient origins of this motif.

The CCT domain on the other hand, is found exclusively in plant proteins, according to the InterPro database (www.ebi.ac.uk/interpro). Because it is found in conjunction with domains other than B-boxes, this domain seems to have an independent function as well. It has been combined with the GATA-type zinc finger, a DNA-binding domain (Reyes *et al.*, 2004; Shikata *et al.*, 2004), as well as with the pseudo-receiver domain of ARABIDOPSIS PSEUDO-RESPONSE REGULATOR 1/TIMING OF CAB 1 (APRR1/TOC1) family proteins (Matsushika *et al.*, 2000; Strayer *et al.*, 2000). Recently, a more diverged version of the motif was found to be conserved in eight unrelated *Arabidopsis* proteins called ACTIVATOR OF SPO^{MIN}::LUC2 (ASML2) family proteins (Masaki *et al.*, 2005). The domain could have a conserved function in nuclear localisation, because it is required for nuclear import of CONSTANS as well as APRR1/TOC1 (Makino *et al.*, 2000; Robson *et al.*, 2001). In addition, the domain has been implicated to be involved in protein-protein interaction (Kurup *et al.*, 2000). The domain's importance for CONSTANS function is reflected by the fact that two further classical *co* alleles map to this region, although one of the two does not impair nuclear localisation, indicating that the domain serves additional purposes (Robson *et al.*, 2001).

1.3 *CONSTANS* is the founder of a plant-specific gene family

To date, *CONSTANS* homologues have been found in several flowering plants, but not in yeast and animals. In *Arabidopsis* as well as in rice, *CONSTANS* is part of a large gene family with seventeen and sixteen members, respectively (Robson *et al.*, 2001; Griffiths *et al.*, 2003). In *Arabidopsis*, the 17-member family of *CO*-like genes consists of three broad clades, referred to as Group 1, Group 2, and Group 3 (Robson *et al.*, 2001; Griffiths *et al.*, 2003). Group 1 genes (*CO* and *COL1-COL5*) have two B-boxes and share the same exon-intron structure (Figure 3). Group 2 genes (*COL6-8* and *COL16*) have only a single B-box and the same exon-intron structure as Group 1 genes. Finally, in Group 3 genes (*COL9-COL15*) the second B-box is replaced by a more divergent zinc-finger, and these genes have a different exon-intron structure altogether. The same three groups were found in the 16-member family of rice, indicating that their evolution predates monocot/dicot divergence (Griffiths *et al.*, 2003).

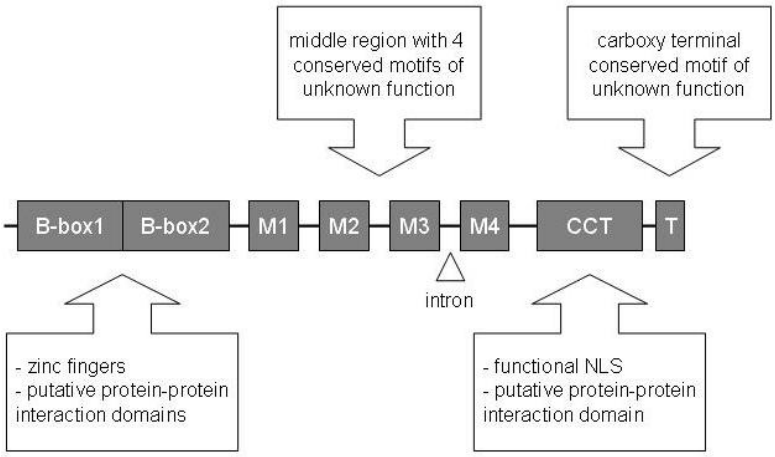


Figure 3 Conserved regions of CO that it shares with other Group 1a CO-like proteins. Grey rectangles depict conserved domains (Robson *et al.*, 2001; Griffiths *et al.*, 2003). The position of the intron is indicated by a white triangle. Feature properties are given in the arrowed boxes.

Group 1 CO-like proteins have additional conserved motifs that are not found in Group 2 or Group 3 proteins, including the so-called T motif, a distinctive motif of six amino acids at the protein's carboxy terminus (Figure 3) (Griffiths *et al.*, 2003). The middle region of CO-like proteins is the most diverged, but comparison of monocot and dicot proteins identified four small conserved motifs that are specific to Group 1 proteins. These motifs are referred to as M1 to M4 (Figure 3), and they helped define a further subdivision within *Arabidopsis* Group 1 genes: *CO*, *COL1* and *COL2* constitute Group 1a and contain middle region motifs M1, M2, M3 and M4, whereas COL3, COL4 and COL5 constitute Group 1c and contain motifs M1 and M4 only (Griffiths *et al.*, 2003). The same subgroups are present in rice and barley, with the same distinctive conserved motifs. However, three additional subgroups of Group 1 *CO*-like genes are described for barley and rice and may be monocot specific (Griffiths *et al.*, 2003).

Apart from *CO* itself, little is known about the function of *CO*-like genes. No function in photoperiodic flowering, or any other function, could be assigned to the two most closely related paralogues in *Arabidopsis*, *COL1* and *COL2* (Ledger *et al.*, 2001). Only for *Arabidopsis COL9* has a function been suggested, most interestingly in photoperiodic flowering (Cheng and Wang, 2005). All the other *Arabidopsis* paralogues have not yet been studied.

Analysis of the rice orthologue of *CONSTANS*, *HEADING DATE 1* (*HD1*), indicated that the *Arabidopsis GI-CO-FT* signalling cascade is generally conserved between the long-day plant *Arabidopsis* and the short-day plant rice (Hayama and Coupland, 2004). The variation in flowering responses is probably achieved by a variation in the effect of coincidence of light on CO protein function; coincidence of light, as signalled through particular photoreceptors, targets CO for degradation in *Arabidopsis*, whereas in rice, it turns CO into a transcriptional repressor (Izawa *et al.*, 2002; Hayama and Coupland, 2004). Nonetheless, the role of *CO* as an integrator of internal circadian rhythms and external factors (light) appears to be conserved between dicotyledonous and monocotyledonous species.

CO homologues from *Brassica*, *Pharbitis*, and ryegrass successfully promoted flowering when expressed in *Arabidopsis* from the *Cauliflower Mosaic Virus* (CaMV) 35S promoter, suggesting that they share the functional properties of the CO protein (Robert *et al.*, 1998; Liu *et al.*, 2001; Martin *et al.*, 2004). In addition, several homologues have been isolated from a variety

of other flowering plants, such as wheat, meadow fescue, poplar, and several legume species (Yuceer *et al.*, 2002; Nemoto *et al.*, 2003; Martin *et al.*, 2004; Hecht *et al.*, 2005). Although present in flowering plants and not in animals in yeast, it is unknown whether *CO* homologues are more widely conserved in the plant kingdom.

1.4 *Physcomitrella patens* as a plant model organism

The non-vascular, multicellular land plant *Physcomitrella patens* is only distantly related to *Arabidopsis* and rice. It is a member of the bryophytes, which are thought to have diverged from the vascular plants about 700 million years ago (Mya) (Hedges *et al.*, 2004). Although all plants have a common life history that involves the alteration of a haploid and a diploid generation, these generations have taken different evolutionary routes during the long separation of bryophyte and vascular plant lineages (Graham *et al.*, 2000). In extant mosses, the haploid gametophyte is the dominant generation, superior to the diploid sporophyte in size as well as complexity, whereas extant angiosperms possess a dominant diploid generation, with unprecedented levels of complexity and tissue differentiation.

The haploid, gametophytic phase of *Physcomitrella* starts with the germination of a haploid spore and consists of two developmental stages with very different morphologies (Figure 4). The first stage, called protonema, is filamentous, arising through division of an apical cell and division of sub-apical cells, resulting in branching. Initially, protonema consists of a single cell type, called chloronema. These cells contain many chloroplasts and seem to be specialised in energy production. After a few days' growth however, they may give rise to a new cell type, called caulonema. Thse cells are longer, divide more rapidly, and contain fewer chloroplasts. Therefore, caulonema cells seem to have differentiated towards a role in vegetative habitat colonisation. The cell types can switch back and forth, and the balance between the two is known to be influenced by light and by phytohormones such as auxin and cytokinin (Cove, 1992). Sub-apical, caulonemal side branch initials may have a developmental fate other than filamentous growth, marking the start of the second stage of gametophyte development. Under the influence of light and phytohormones, the initials may develop into a bud which divides three-

dimensionally and gives rise to the so-called gametophores (Cove, 1992). Gametophores produce leaf-like structures called leaflets, consisting of a single cell layer, and root-like structures called rhizoids. As their name implies, gametophores are the organs that produce the gametes. *Physcomitrella patens* is monoecious; male gametes, or antherozoids, are produced within antheridia, and female gametes, or oogonia, within archegonia on the same gametophore. The male gametes affect fertilisation by swimming to the archegonia through a surface water film. After fertilisation, zygotes are formed, marking the starting point of the next generation, the diploid sporophyte (Figure 4). The *Physcomitrella* sporophyte is small and largely dependent on the gametophytic generation. Nevertheless, it shows some degree of tissue differentiation: sporangium, seta and foot are formed. After maturation, sporangia or spore capsules generally contain ~4000 haploid spores (Cove, 1992). The entire life cycle can be completed in ~3 months in culture.

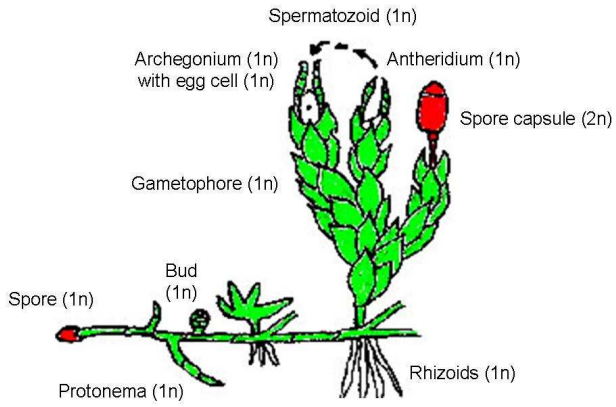


Figure 4 Schematic representation of the *Physcomitrella patens* life cycle. Gametophytic (1n) development starts with the germination of spores and the outgrowth of the filamentous protonema tissue. Bud formation gives rise to the adult moss plant (gametophore), which carries the sex organs (antheridia and archegonia). After fertilisation of the egg cell, the diploid sporophyte develops (spore capsule). Drawing courtesy of Dr. B. Chrost; adapted from Decker and Reski (2004).

Compared with *Arabidopsis*, mosses are small and slow growing plants. This potential problem of limited availability of plant material can readily be overcome thanks to vigorous powers of vegetative regeneration; almost any tissue of *Physcomitrella patens* is capable of regeneration (Cove, 1992). As such, *Physcomitrella* is propagated in the lab by repeated blending and

regeneration of protonemal tissue, generally on a weekly basis. These cultures are grown axenically, on simple solid medium containing only inorganic salts and agar, with the potential of being supplemented with an external carbon source for more vigorous growth.

In spite of the dramatic differences in morphology and life cycle, mosses and vascular plants do share some general themes of plant physiology. These include responsiveness to phytohormones (*i.e.* auxin, cytokinin, abscisic acid) (Cove, 1992; Knight *et al.*, 1995; Imaizumi *et al.*, 2002), phototropic responses (Knight and Cove, 1989), chloroplast movement (Wada *et al.*, 2003; Kasahara *et al.*, 2004), and photomorphogenesis (*i.e.* signalling through phytochromes, cryptochromes, phototropins) (Imaizumi *et al.*, 2002; Kasahara *et al.*, 2004; Mittmann *et al.*, 2004). Furthermore, even underlying molecular mechanisms are sometimes preserved, like the ABA-mediated desiccation stress response network (Knight *et al.*, 1995), or the GLK-mediated regulatory pathway for chloroplast development (Yasumura *et al.*, 2005). Also gene regulation through microRNAs seems to be conserved between several *Physcomitrella* and *Arabidopsis* homologues (Floyd and Bowman, 2004; Arazi *et al.*, 2005), as well as the general mechanisms by which small interfering RNAs (siRNA) cause gene silencing, because RNA interference (RNAi) has successfully been applied in *Physcomitrella* (Bezanilla *et al.*, 2003).

All the differences and similarities between *Arabidopsis* and *Physcomitrella* should be reflected in their genomes. Comparative studies have suggested that the haploid transcriptome of *Physcomitrella* and the genome of *Arabidopsis* largely overlap, and that >90% of the most closely related homologues of *Physcomitrella* transcripts occur in vascular plants (Nishiyama *et al.*, 2003). As yet, *Physcomitrella* has proven instrumental in studying the ancestry of gene families like the MADS-box (Krogan and Ashton, 2000; Henschel *et al.*, 2002; Riese *et al.*, 2005), HD-ZIP (Sakakibara *et al.*, 2001), and KNOX genes (Champagne and Ashton, 2001). Such studies have contributed to the general understanding of these genes as well as the evolutionary development of the corresponding regulatory pathways. Several extensive EST sequencing efforts have been undertaken and have resulted in substantial, publicly available resources for gene discovery (Nishiyama *et al.*, 2003; http://moss.nibb.ac.jp). Furthermore, *Physcomitrella* offers a unique

opportunity among plants to study the function of such genes by gene targeting (Schaefer, 2001).

1.5 Objectives of this work

In the work presented here, the first analysis of *CONSTANS* homologues outside of the realm of flowering plants is conducted, in the distantly related moss species *Physcomitrella patens*. Apart from CO itself, little is known about the function of other *CO*-like genes, therefore the focus is on identifying putative orthologues of *CO*. *Physcomitrella* EST databases are searched in order to identify and isolate the most closely related homologues of *CONSTANS*. Efforts are undertaken to confirm that these genes represent the closest homologues of *CO* in *Physcomitrella*. Functional conservation between the *Physcomitrella* proteins and CO is tested by expression studies in *Arabidopsis*. Because diurnal or circadian regulation of transcription appears to be a conserved feature among *CO*-like genes of flowering plants, it is analysed to what extent this regulation is conserved in moss *CO*-like genes. Finally, the feasibility of gene targeting is exploited in order to investigate the biological function of the *Physcomitrella CO*-like genes.

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemicals, enzymes, oligonucleotides, cloning vectors

Chemicals used for this work were purchased from Sigma-Aldrich (Steinheim), Merck (Darmstadt), Serva (Heidelberg), Duchefa (Haarlem, The Netherlands), Biozym (Hamburg), Roth (Karlsruhe), Eppendorf (Hamburg), FMC Bioproducts (Brussels, Belgium), Gibco BRL (Karslruhe), and Invitrogen (Karlsruhe). Enzymes were purchased from Roche (Penzberg), New England Biolabs (Frankfurt am Main), Fermentas (St. Leon-Rot), Stratagene (Heidelberg), and Invitrogen (Karlsruhe). Oligonucleotides were synthesised at Metabion (Martinsried) and Invitrogen (Karlsruhe). Cloning vectors used were pGEM-T easy (Promega), pBLUESCRIPT SK- (Stratagene), pDONR-201 (Invitrogen), and pJAN33 (Weigel *et al.*, 2003).

2.1.2 Buffers, solutions, media

Standard buffers, solutions and media were prepared as described (Sambrook *et al.*, 1989).

2.1.3 Bacterial strains

E. coli DH5α	supE44 ΔlacU169 (Φ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1
E. coli DH10B	F- mcrA D (mrr -hsdRMS- mcrBC) F80dlacZDM15 DlacX74 endA1 recA1 D (ara, leu) 7697 araD139 galU galK nupG rpsL T1R

E. coli SCS110	rpsL (Strr) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44D (lac-proAB) [F' traD36 proAB lacI q ZΔM15]
A. tumefaciens GV3101:pMP90-RK	(Koncz and Schell, 1986)

2.1.4 Plant materials

Physcomitrella patens	Gransden Wood strain (Ashton and Cove, 1977)
Arabidopsis thaliana	Columbia ecotype (Col-0)
Arabidopsis thaliana	*35S::AtCO* (Col-0) (Samach *et al.*, 2000) (kindly provided by Dr. Wim Soppe)

2.1.5 Database sequences

Accession numbers of protein sequences: AtCO (Q39O57), AtCOL1 (O50055), AtCOL2 (Q96502), AtCOL3 (Q9SK53), AtCOL4 (Q940T9), AtCOL5 (Q9FHH8), AtCOL6 (Q8LG76), AtCOL7 (Q9C9A9), AtCOL8 (Q9M9B3), AtCOL9 (Q9SSE5), AtCOL10 (Q9LUA9), AtCOL11 (O23379), AtCOL12 (Q9LJ44), AtCOL13 (O82256), AtCOL14 (O22800), AtCOL15 (Q9C7E8), AtCOL16 (Q8RWD0), AtSTO (Q96288), AtTOC1 (Q9LKL2), OtCOL (Q5IFM9), PpCOL1 (Q5H7P0), PpCOL2 (Q4W1E9), PpCOL3 (Q4W1E8); accession numbers of *Physcomitrella* EST sequences are given in the text in Table 1.

2.1.6 Plasmids

pUC18/sul	B. Reiss, unpublished
pUC18/Hyg	idem
pUC/NPT	idem
pUC18,12,26.5.88	idem
pUC19,13/1,20.11.91	idem

2.1.7 Oligonucleotides

Name	Sequence (5'-3')
5dCO	GGYGGTGWWSTGYYGDGCRG
3dCO	GGMACCACYCCRAAACYTGARTC
5-ATG-AX	ATGCCGAAGCCTTGTGATG
3-Stop-AX	TCAAAAACTTGGAACCACTCC
5-ATG-101/351	ATGCCGAAGTCATGCGATG
3-Stop-101	TCAGAAAGAAGGCACCACTCC
3-Stop-351	TCAACAAGAAGAAGGAACCACC
5-AXRT-II	CATGGACCCTTCGTTTACTAAA
3-AXRT-II	TCCATTTCAGATGACCTTGC
5-101RT-II	CCTTCACAACTGATTTTCATCTG
3-101RT-II	ACTCAACTTGATTGAAGCAAGG
5-351RT-I	ACATCCATTCTGCCAACCC
3-351RT-I	TGTGTGAGAGTAGAAGTGCC
5-AXRT-I	ATGATCTGCTGAAGGGCTG
3-AXRT-I	GGAAATGTCGCTGAGACTG
5-101RT-I	CCACATCCCAAAATACCTACC
3-101RT-I	TCTCCTTGTACCTCATCACTC
pp18for	AGGAATTGACGGAAGGGCAC
pp18rev	GGACATCTAAGGGCATCACA
5-AX180961	TCTCTCGGCGAAGAGCG
3-AX180961	AACTTGGAACCACTCCGAAAC
5-BJ166101	ACCCACAGCATTTCGTGC
3-BJ166101	CAGAAAGAAGGCACCACTCC
5-BJ166351	CAAGATTTGTCAGGTGCGC
3-BJ166351	TCAACAAGAAGAAGGAACCACC
5-AtCO-CCT	GGCTCCTCAGGGACTCACTAC
3-AtCO-CCT	GAATGAAGGAACAATCCCATATC
UP5-101-NotI	GCGGCCGCTTCTAGCACGCATT
UP3-101-XmaI	CCCGGGAGTGCTCAGCACAGACC
DOWN5-101-SacI	GAGCTCACAGTGGATCGGGAAGCTC
DOWN3-101-ApaI	GGGCCCTGGAACAAAAGAAGACTACATC
UP5-AX-NotI	GCGGCCGCGAGTTGGCTCA
UP3-AX-BamHI	GGATCCATACCACGGCACAACACAAC

DOWN5-AX-SacI	GAGCTCGTATTGGAGTCTCCATCGAGC
DOWN3-AX-ApaI	GGGCCCATCAAGAGAGAGGGATTG
UP5-351-NotI	GCGGCCGCTGCCAGACTCTATTAAAG
UP3-351-BamHI	GGATCCTCTCCCGAGGGCTCAGC
DOWN5-351-HindIII	AAGCTTCGGACTCTGATGTGGAGCAG
DOWN3-351-ApaI	GGGCCCTTGGCACGAACCTC
5-SpeI-AX	GACTAGTATGCCGAAGCCTTGTGATG
3-Sac1-AX	CCCGAGCTCTCAAAAACTTGGAACCACTCCG
5-SpeI-101-351	GACTAGTATGCCGAAGTCATGCGATG
3-SacI-101	CCCGAGCTCTCAGAAAGAAGGCACCACTCC
3-SacI-351	CCCGAGCTCTCAACAAGAAGAAGGAACCACCC
5-AX-GW	*GYF*-ATGCCGAAGCCTTGTGAT
3-AX-GW	*GYR*-TCAAAAACTTGGAACCACTCC
5-351and101-GW	*GYF*-ATGCCGAAGTCATGCGAT
3-101-GW	*GYR*-TCAGAAAGAAGGCACCACTCC
3-351-GW	*GYR*-TCAACAAGAAGAAGGAACCACC
5-AX-anti-I	GGGTCAGATCCAAGGAGAGAT
3-AX-anti-I	CGCTGCAACCGCATAAC
5-101-anti-I	GAAATGAACATTGAACAACTTGC
3-101-anti-I	AATTCCTTCTCATTGACATAAGATG
5-101-anti-III	CAGCACAGGAGTCCATTCG
3-101-as-I-bis2	TGCTTGGCTTCGTCAGC
5-351-anti-I	GGGAATGCAAGTATGTGATGAG
3-351-anti-I	GGGCATTACAGAGGCTGG
5-iAX	CGTGGACCTTCCCATCG
3-iAX	CCGAGGATCAAAGGAAGGT
5-i101EcoHind	CGGTCAGGGCAGTTGTCA
3-i101EcoHind	CGCTCGCTTCGCCTGT
5i10-EH-3UTR	CGTCAGGCGCTTTCAACA
3i10-EH-3UTR	GCAGGAGGTTCCGGTAGTG
5-i351	CATGCACCTTTCCGTCG
3-i351	GAATCAAAGGCAGGTTCACC
5-M-AX180961	TTCGAGAGTGCGAGTCCTTT
3-M-AX180961	CTTCGCCGAGAGAGGAAAT
5-M-BJ166101	CTAGCTCCGCAAGAGAAGC
3-M-BJ166101	GCTCTTCTCCGACAGATGAA

5-M-BJ166351	CGTGAGCCAGTTTAGAGGGA
3-M-BJ166351	CTTCTGCGAGGCACGGA
GUS-1877	CTGCATCGGCGAACTGAT
pcrsulout1-2	GCTATTGGTCTCGGTGTCGC
HPT-1167	CTCGATGAGCTGATGCT
NPT-706	AAGCCGGTCTTGTCGATC

GYF = Gateway attB1 extension 5'-GGGGACAAGTTTGTACAAAAAAGCAGGCT-3'
GYR = Gateway attB2 extension 5'-GGGGACCACTTTGTACAAGAAAGCTGGGT-3'

2.2 Methods

2.2.1 Plant growth conditions

Physcomitrella protonema cultures were routinely propagated in growth cabinets at 26°C in continuous light on cellophane covered minimal medium (0,8 g/l $Ca[NO_3]_2$ x $4H_2O$, 0,25 g/l $MgSO_4$ x $7H_2O$, 1ml KH_2PO_4/KOH [250 g/l KH_2PO_4 adjusted to pH 6,5 with KOH], 0,0125 g/l $FeSO_4$ x 7 H_2O, 1 ml alt TES [55 mg $CoSO_4$ x 5 H_2O, 614 mg H_3BO_4, 55 mg $CoCl_2$ x $6H_2O$, 25 mg Na_2MoO_4 x $2H_2O$, 55 mg $ZnSO_4$ x 7 H_2O, 393 mg $MnCl_2$ x 4 H_2O, 28 mg KI, ad 1l with H_2O], 7 g/l agar, autoclaved for 20 min at 121°C) supplemented with 0,5 g/l diammonium tartrate. For in vitro propagation, 4 cultures were collected in water, blended for 90 s with a Miccra homogeniser D8 equipped with a P8 homogeniser tool (ART-Moderne Laborgeräte, Hügelheim), diluted with equal volume water and 1/20th plated. For growth in white light, the following growth cabinets were used: Rumed 1301 from Rubarth Apparate (Laatzen) equipped with Osram L36W/860 Lumilux Daylight lamps; Rumed 1200 and 5001 from Rubarth Apparate (Laatzen) equipped with Osram L36W/11-860 Lumilux Plus Daylight lamps; Percival CU-365/D from CLF Laborgeräte (Emersacker) equipped with Philips F17T8/TL741 17 Watt lamps. Light intensities are given in the text. For growth in red, far-red, and blue light, the following growth cabinet was used: Percival E-30 LED from CLF Laborgeräte (Emersacker). Light intensities are 240 Lux, 206 Lux, and 30 Lux for red light (600-700 nm), far-red light (700-750 nm), and blue light (400-500 nm), respectively.

Arabidopsis was grown on soil in the protected environment of the greenhouse, at 20°C, in long-day conditions of 18 hrs light and 8 hrs darkness (16L:8D) and short-day conditions of 8 hrs light and 16 hrs darkness (8L:16D).

2.2.2 Nucleic acid techniques

2.2.2.1 DNA and RNA isolation

Plasmid DNA was routinely isolated by alkaline lysis method (Birnboim and Doly, 1979); large amounts were isolated using the Plasmid Midi/Maxi Kit from Qiagen (Hilden). Electrophoretic separation of DNA fragments was carried out according to standard procedures (Sambrook *et al.*, 1989). Plant DNA was isolated as described (Markmann-Mulisch *et al.*, 2002). Plant total RNA was prepared as described (Markmann-Mulisch *et al.*, 1999).

Concentration of isolated DNA was determined by electrophoretic comparison with a λ-DNA standard, by standard spectrophotometric measurement (Sambrook *et al.*, 1989), or by fluorometric measurement using the PicoGreen double-stranded DNA quantitation reagent (Molecular Probes), according to the manufacturer's instructions. Concentration of isolated RNA was determined by standard spectrophotometric measurement (Sambrook *et al.*, 1989).

2.2.2.2 Digestion and ligation

Digestion and ligation of DNA fragments with restriction enzymes and ligases, respectively, were carried out according to the manufacturer's instructions and in the provided buffers.

For cloning of *Physcomitrella* transforming constructs pKOcol1, pKOcol2 and pKOcol3, digestions were performed using only minimally required enzyme amounts and incubation times for complete digestion, as indicated by the enzyme manufacturer's instructions. The fragments were separated on a 1% low-melting agarose gel and excised from the gel. Low-melting agarose was melted by incubation at 65°C for 10 min, and a 1 µl aliquot of each

fragment was used in a ligation reaction, which additionally included 1 μl ligase, 1 μl 10x buffer, 1 μl 50% PEG4000 (all Fermentas) and 2 μl H_2O. Ligation was carried out overnight at 15°C.

2.2.2.3 Polymerase chain reaction (PCR) amplification

Standard PCR amplifications were carried out with Taq polymerase (Roche) or with Pfu cloned polymerase (Stratagene) using the following PCR mixture composition and cycling profile.

Components
40,5 μl H_2O
5 μl 10x Buffer
1 μl 10 mM dNTPs mix (dATP, dTTP, dCTP and dGTP)
1 μl 25 μM sense primer
1 μl 25 μM anti-sense primer
1 μl DNA template
0,5 μl polymerase

Thermal profile			
1.	95°C	5 min	
2.	Ta	2 min	
3.	72°C	5 min	
4.	93°C	1 min	
5.	Ta	1 min	n
6.	72°C	2 min	
7.	72°C	10 min	
8.	4°C	∞	

Note: Typically, values for Ta and n were 60°C and 30x, respectively, although they were sometimes optimised for individual reactions.

Long template PCR amplifications were performed with the Expand Long Template PCR system (Roche) according to the manufacturer's instructions.

2.2.2.4 cDNA synthesis

cDNA was synthesised using Superscript II (Invitrogen) and random primers (Invitrogen) according to the manufacturer's instructions, except that the reaction temperature was raised from 37 to 42°C after 10 min.

2.2.2.5 Nucleid acid sequencing

DNA sequences were determined by the MPIZ DNA core facility on Applied Biosystems (Weiterstadt) Abi Prism 377, 3100 and 3730 sequencers using BigDye-terminator v3.1 chemistry. Premixed reagents were from Applied Biosystems (Darmstadt). Oligonucleotides were purchased from Metabion (Martinsried).

2.2.3 Sequence analysis

Standard sequence analysis was performed using components of Vector NTI Suite 9 (Invitrogen). Database searches were routinely carried out using the BLAST algorithm (Altschul et al., 1997) at GenBank (http://www.ncbi.nlm.nih.gov). For the identification of *CO*-like genes from *Physcomitrella patens*, database searches were performed using TBLASTN (Altschul et al., 1997) with the CONSTANS protein sequence as a query sequence at GenBank (release 131.0; http://www.ncbi.nlm.nih.gov), at *Physcobase* (Nishiyama et al., 2003; http://moss.nibb.ac.jp), and at a proprietary EST collection (Rensing et al., 2002b). Sequences were aligned using ClustalW (Thompson et al., 1994) in AlignX, a component of Vector NTI Suite 9 (Invitrogen), and the alignments adjusted manually. Phylogenetic trees were calculated from the alignments using the Neighbour Joining method (Saitou and Nei, 1987); construction and bootstrapping of the trees was performed using programs from the Phylip3.62 software package (Felsenstein, 1989).

2.2.4 Isolation of *PpCOL1*, *PpCOL2* and *PpCOL3*

2.2.4.1 Cloning of genomic and coding DNA gene sequences

The identification and isolation of *PpCOL1*, *PpCOL2*, and *PpCOL3* genes from EST database sequences is described in the text. Initially, incomplete gene sequences of *PpCOL1*, *PpCOL2*, and *PpCOL3* were amplified by PCR from cDNA (Roche Taq polymerase) with degenerate primers 5dCO and 3dCO. Four differently sized PCR products were isolated from an agarose gel by using Qiagen's (Hilden) Gel Extraction Kit, and used as a template in an identical PCR reaction. Three bands were successfully reamplified and cloned in pGEM-T easy according to the manufacturer's instructions. The three genes were named *PpCOL1*, *PpCOL2*, and *PpCOL3*. Full-length genomic and cDNA clones were amplified from genomic DNA or cDNA by PCR (Roche Taq polymerase) with the following primers: 5-ATG-AX and 3-Stop-AX for *PpCOL1*, 5-ATG-101/351 and 3-Stop-101 for *PpCOL2*, and 5-ATG-101/351 and 3-Stop-351 for *PpCOL3*. Fragments from two independent PCR reactions were cloned in pGEM-T easy according to the manufacturer's instructions. The genomic clones were named pGcol1, pGcol2, and pGcol3, for *PpCOL1*, *PpCOL2*, and *PpCOL3*, respectively; the cDNA clones were named pCcol1, pCcol2, and pCcol3, for *PpCOL1*, *PpCOL2*, and *PpCOL3*, respectively. One clone was isolated for each of the two independent PCR reactions, and the inserts were sequenced on both DNA strands. The consensus sequence of the resulting four sequences was calculated using Vector NTI Suite 9 (Invitrogen). Conflicts were resolved by the majority rule; if no majority was found for a nucleotide, it was kept as N for "any nucleotide".

2.2.4.2 Cloning of flanking genomic sequences

The genomic sequences flanking the *PpCOL1*, *PpCOL2* and *PpCOL3* genes were isolated by inverse PCR. *Physcomitrella* genomic DNA was digested with *BamHI* or *EcoRI*. The DNA was re-ligated and used as a template in a long template PCR (Roche Expand Long Template polymerase) according to the manufacturer's instructions. The primers were 5-iAX and 3-iAX for *PpCOL1*, 5-i101EcoHind and 3-i101EcoHind for *PpCOL2* upstream

genomic sequence, 5i10-EH-3UTR and 3i10-EH-3UTR for *PpCOL2* downstream genomic sequence, and 5-i351 and 3-i351 for *PpCOL3*. Fragments from two independent PCR reactions were cloned in pGEM-T easy according to the manufacturer's instructions. One clone was isolated for each reaction and the inserts were sequenced on both strands. Sequences were assembled using Vector NTI Suite 9 (Invitrogen), and the consensus sequence of the four resulting sequences was calculated using the same software. Conflicts were resolved by the majority rule; if no majority was found for a nucleotide, it was kept as N for "any nucleotide".

The gene sequences and flanking genomic sequences of *PpCOL1*, *PpCOL2*, and *PpCOL3* were assembled into continuous contigs using Vector NTI Suite 9 (Invitrogen). The complete contigs were separately amplified from *Physcomitrella* genomic DNA by PCR (Roche Expand Long Template polymerase) using primers 5-AX-anti-I and 3-AX-anti-I for *PpCOL1*, 5-101-anti-I and 3-101-anti-I for *PpCOL2*, and 5-351-anti-I and 3-351-anti-I for *PpCOL3*. Fragments were cloned in pGEM-T easy according to the manufacturer's instructions. The clones were named pGCcol1, pGCcol2, and pGCcol3 for *PpCOL1*, *PpCOL2*, and *PpCOL3*, respectively.

2.2.5 Southern blotting

2.2.5.1 Preparation of probes

The probes used for probing of the *Physcomitrella* genome for additional *CO*-like genes were generated as follows. The *PpCOL1* probe was amplified by PCR (Roche Taq polymerase) from pCcol1 with primers 5-AX180961 and 3-AX180961; the *PpCOL2* probe from pCcol2 with primers 5-BJ166101 and 3-BJ166101; the *PpCOL3* probe from pCcol3 with primers 5-BJ166351 and 3-BJ166351; the *AtCO* probe from *Arabidopsis* cDNA (prepared as described earlier for *Physcomitrella* cDNA) with primers 5-AtCO-CCT and 3-AtCO-CCT. The PCR products were separated on a 1% low-melting agarose gel and the bands cut out. The agarose was melted for 10 min at 65°C, and an aliquot was used for labelling.

The probes used for Southern blot analysis of *Physcomitrella* transformants were generated as follows. The *PpCOL1* probe was obtained by

digestion of pGcol1 with *DraII* and *NotI*, separation of the fragments on a 1% low-melting agarose gel, and cutting out the 1231 bp fragment. The *PpCOL2* probe was obtained by digestion of pGcol2 with *ApaI* and *NotI*, separation of the fragments on a 1% low-melting agarose gel, and cutting out the 1347 bp fragment. The *PpCOL3* probe was obtained by digestion of pGcol3 with *ApaI* and *NotI*, separation of the fragments on a 1% low-melting agarose gel, and cutting out the 915 bp fragment. The agarose was melted for 10 min at 65°C, and equal amounts of each probe were mixed and used for labelling.

The procedures for preparation of radioactively labelled probes were as described (Markmann-Mulisch *et al.*, 2002).

2.2.5.2 Blotting procedures

Genomic DNA was digested with restriction endonucleases, the fragments separated on an agarose gel and alkaline blotted to Zeta-Probe blotting membranes (Bio-Rad). The blot was pre-hybridised in hybridisation buffer (0,25 M sodium phosphate, pH 7, 0,25 M sodium chloride, 7% [w/v] SDS, 1 mM EDTA) and then hybridised in the same buffer containing radioactively labelled probe overnight at 65°C (high stringency), 55°C (medium stringency), or 45°C (low stringency). The membrane was washed once with 2X SSC (0,3 M sodium chloride, 0,03 M sodium citrate, pH 7), 0,1% SDS at 65°C for 10 min, twice with 0,5X SSC, 0,1% SDS at 65°C for 20 min and then exposed to Kodak Biomax MS film with intensifying screens.

2.2.6 Quantitative RT-PCR

2.2.6.1 Analysis of *Physcomitrella* RNA

A 1 µl aliquot of *Physcomitrella* cDNA was amplified by PCR in a mixture with Taq polymerase (Roche), gene-specific sense and anti-sense primers, and *Physcomitrella* 18S rDNA primers and 18S rDNA competimers as internal standard. 18S rDNA primers are pp18for and pp18rev (product length 320 bp). Their sequence identical 18S rDNA competimers are blocked by a 3'-phosphate. The *PpCOL1* gene-specific primers are 5-axRT-II and 3-axRT-II

(product length 455 bp), the *PpCOL2* gene-specific primers are 5-101RT-II and 3-101RT-II (product length 520 bp), and the *PpCOL3* gene-specific primers are 5-351RT-I and 3-351RT-I (product length 520 bp). Primer specificity was confirmed by restriction analysis of the resulting PCR products with gene-specific restriction endonucleases: *NdeI* for *PpCOL1*, *HindIII* for *PpCOL2*, and *PstI* for *PpCOL3*. The cycle numbers were adjusted to the linear range of the PCR reaction. For this, PCR amplification of *Physcomitrella* cDNA was performed according to the scheme below with 25, 27, 29, 31, 33, and 35 cycles. A 5 µl aliquot of the PCR reaction mixture was separated by electrophoresis on a 2,5% low-melting agarose gel (Invitrogen Agarose 1000) in 0,5x TBE buffer, followed by staining with SYBR green (Molecular Probes) and destaining, according to the manufacturer's instructions. Bands were quantified using a Kodak DC290 camera and Kodak 1D Image analysis software. The highest cycle number that resulted in a signal strength that was still in the logarithmic phase was chosen for quantitative RT-PCR analysis. The strength of the 18S rRNA signal was adjusted to the strength of the gene-specific signal, by adding appropiate amounts of 18S rRNA competimers as described (Ambion Quantum kit). The optimised PCR conditions for each gene are given below the thermal profile.

Components
36,5 µl H_2O
5 µl 10x Taq Buffer (Roche)
1 µl 10 mM dNTPs mix (dATP, dTTP, dCTP and dGTP)
1 µl 25 µM sense primer
1 µl 25 µM anti-sense primer
1 µl cDNA template
4 µl 5 µM 18S rDNA primer:competimer (0,3:9,7)
0,5 µl Taq polymerase (Roche)

Thermal profile			
1.	94°C	3 min	
2.	Ta	1 min	
3.	72°C	2 min	
4.	93°C	1 min	
5.	Ta	1 min	n
6.	72°C	2 min	
7.	4°C	∞	

PpCOL1: Ta= 60°C, n=26x; *PpCOL2*: Ta= 64°C, n=24x; *PpCOL3*: Ta= 65°C, n= 26x

2.2.6.2 Analysis of *Arabidopsis* RNA

Analysis was carried out as described previously for *Physcomitrella*, albeit with the following modifications: 5-axRT-I and 3-axRT-I were used as *PpCOL1* gene-specific primers (product length 467 bp); 5-101RT-I and 3-101RT-I were used as *PpCOL2* gene-specific primers (product length 463 bp). The 18S rDNA primer:competimer ratio in the PCR mixture was 1,25:8,75 for *PpCOL1*; 0,4:9,6 for *PpCOL2*; 0,6:9,4 for *PpCOL3*. The annealing temperature (Ta) and cycle number (n) were 68°C and 23x for *PpCOL1*; 65°C and 23x for *PpCOL2*; 65°C and 20x for *PpCOL3*.

2.2.7 Generation of plant transformation constructs

2.2.7.1 Constructs for *Arabidopsis* transformation

The transforming constructs were generated by using Gateway cloning technology (Invitrogen). The coding sequences of *PpCOL1*, *PpCOL2* and *PpCOL3* were amplified by PCR (Stratagene Cloned Pfu polymerase) from pCcol1 with primers 5-AX-GW and 3-AX-GW, from pCcol2 with primers 5-351and101-GW and 3-101-GW, and from pCcol3 with primers 5-351and101-GW and 3-351-GW, respectively. They were integrated into the pDONR-201 entry vector, and then transferred to the pJAN33 destination vector, all

according to the instructions of the Gateway cloning manual (Invitrogen). The inserts were confirmed to be error-free by sequencing. The resulting plasmids, pJAN33-PpCOL1, pJAN33-PpCOL2, and pJAN33-PpCOL3, carry the CaMV 35S promoter in front of the coding sequence of the respective gene.

2.2.7.2 Constructs for *Physcomitrella* transformation

The *PpCOL1* gene-specific gene replacement vector pKOcol1 was constructed by ligation of five fragments. The first fragment, containing the upstream flanking sequence of *PpCOL1*, was amplified by PCR (Roche Expand Long Template polymerase) from pGCcol1 with the primers UP5-AX-NotI and UP3-AX-BamHI, subcloned in pGEM-T easy, and released by digestion with *NotI* and *BamHI*. The second fragment, containing the GUS reporter gene, was released from pUC18,12,26.5.88 by digestion with *BamHI* and *EcoRI*. The third fragment, containing the sulfadiazine resistance gene (*sul*) under control of the CaMV 35S promoter, was released from pUC18/sul by digestion with *EcoRI* and *SacI*. The fourth fragment, containing the downstream flanking sequence of *PpCOL1*, was amplified by PCR (Roche Expand Long Template polymerase) from pGCcol1 with the primers DOWN5-AX-SacI and DOWN3-AX-ApaI, subcloned in pGEM-T easy, and released by digestion with *SacI* and *ApaI*. The fifth fragment was the pBLUESCRIPT SK- vector, linearised by digestion with *NotI* and *ApaI*. Ultimately, clones were characterised by restriction analysis, confirming release of every cloning fragment with the respective restriction endonucleases.

The *PpCOL2* gene-specific gene replacement vector pKOcol2 was constructed by ligation of five fragments. The first fragment, containing the upstream flanking sequence of *PpCOL2*, was amplified by PCR (Roche Expand Long Template polymerase) from pGCcol2 with the primers UP5-101-NotI and UP3-101-XmaI, subcloned in pGEM-T easy, and released by digestion with *NotI* and *XmaCI*. The second fragment, containing the GUS reporter gene, was released from pUC19,13/1,20.11.91 by digestion with *XmaCI* and *HindIII*. The third fragment, containing the hygromycin phosphotransferase gene (*hpt*) under control of the *Agrobacterium nopaline synthase* (*nos*) promoter, was released from pUC18/Hyg by digestion with *HindIII* and *SacI*. The fourth fragment, containing the downstream flanking sequence of *PpCOL2*, was amplified by PCR (Roche Expand Long Template polymerase) from pGCcol2

with the primers DOWN5-101-SacI and DOWN3-101-ApaI, subcloned in pGEM-T easy, and released by digestion with *SacI* and *ApaI*. The fifth fragment was the pBLUESCRIPT SK- vector, linearised by digestion with *NotI* and *ApaI*. The remaining manipulations are as described for pKOcol1, with the only difference that pKOcol2 was finally propagated in *E.coli* strain SCS110.

The *PpCOL3* gene-specific gene replacement vector pKOcol3 was constructed by ligation of five fragments. The first fragment, containing the upstream flanking sequence of *PpCOL3*, was amplified by PCR (Roche Expand Long Template polymerase) from pGCcol3 with the primers UP5-351-NotI and UP3-351-BamHI, subcloned in pGEM-T easy, and released by digestion with *NotI* and *BamHI*. The second fragment, containing the GUS reporter gene, was released from pUC18,12,26.5.88 by digestion with *BamHI* and *EcoRI*. The third fragment, containing the neomycin phosphotransferase II gene (*nptII*) under control of the CaMV 35S promoter, was released from pUC/NPT by digestion with *EcoRI* and *HindIII*. The fourth fragment, containing the downstream flanking sequence of *PpCOL3*, was amplified by PCR (Roche Expand Long Template polymerase) from pGCcol3 with the primers DOWN5-351-HindIII and DOWN3-351-ApaI, subcloned in pGEM-T easy, and released by digestion with *HindIII* and *ApaI*. The fifth fragment was the pBLUESCRIPT SK- vector, linearised by digestion with *NotI* and *ApaI*. The remaining manipulations are as described for pKOcol1.

2.2.8 Transformation and selection procedures

2.2.8.1 Bacterial transformation and selection

Electrocompetent *E.coli* cells were either purchased from the strain's manufacturer, or prepared according to the $RbCl_2$-method (Hanahan, 1983). Electrocompetent *A. tumefaciens* cells were a kind gift from Drs. F. Turck and Y. F. Fu. The cells were stored at -70°C. Transformation and selection procedures were as described (Sambrook *et al.*, 1989).

2.2.8.2 *Arabidopsis* transformation and selection

The plasmids pJAN33-PpCOL1, pJAN33-PpCOL2, and pJAN33-PpCOL3 were transformed into *Agrobacterium tumefaciens* strain GV3101:pMP90-RK. A single colony resistant to 50 mg/l carbenicilline, 25 mg/l kanamycin, 25 mg/l gentamycine and 50 mg/l rifampicine was inoculated for preculture in liquid LB medium (Sambrook *et al.*, 1989) supplemented with the same antibiotics at the same concentrations, apart from the concentration of rifampicine which was reduced to 10 mg/l. The presence of the pJAN33-PpCOL1, pJAN33-PpCOL2, or pJAN33-PpCOL3 plasmid in each preculture was confirmed by PCR amplification (Roche Taq polymerase) from a 1 µl culture aliquot with the primers 5-SpeI-AX and 3-Sac1-AX, 5-SpeI-101-351 and 3-SacI-101, and 5-SpeI-101-351 and 3-SacI-351, respectively. A 1/200 aliquot of preculture was used to inoculate 300 ml YEB medium (Sambrook *et al.*, 1989) supplemented with 100 mg/l carbenicilline. Cultures were grown overnight under shaking at 28°C. When the cultures had reached an OD^{600nm} of ~0.5 60 µl Silwet Copolymer L-77 (OSI Specialties, Düsseldorf) was added, the culture stirred, and then directly used to transform *Arabidopsis* plants as described (Clough and Bent, 1998). Transformed plants were selected on the basis of kanamycin resistance (Hadi *et al.*, 2002) and were self-pollinated to generate T2 populations that segregate the transgenes.

2.2.8.3 *Physcomitrella* transformation and selection

To release the targeting fragments from plasmids pKOcol1, pKOcol2, and pKOcol3, they were digested with *NotI* and *ApaI* prior to transformation. For transformation, protonema was grown on minimal medium supplemented with 0,5 g/l diammonium tartrate and 0,5% glucose, in a Rumed 5001 growth cabinet under conditions of continuous light (light intensity ~250 Lux). 5 day old tissue was harvested, protoplasts isolated and 450.000 protoplasts transformed with 15 µg linearised DNA as described previously (Schaefer and Zryd, 1997). For double transformants, the transformation procedure was scaled up six times, transforming 2,7 x 10^6 protoplasts with 45 µg linearised DNA of each targeting fragment. For triple transformants, the transformation procedure was scaled up 18 times, transforming 8,1 x 10^6 protoplasts with 60

µg linearised DNA of each targeting fragment. Protoplasts were regenerated in liquid minimal medium supplemented with 0,5 g/l diammonium tartrate and 66 g/l mannitol (transformation experiments I, II, IV and V) (Hohe *et al.*, 2004), or embedded in low-melting agarose (transformation experiment III) (Schaefer and Zryd, 1997). Regeneration was carried out in a Rumed 1301 growth cabinet under long-day conditions of 16L:8D (light intensity ~100 Lux). Transformants were selected by tentatively 4 cycles of growth on selective and non-selective media as described (Schaefer *et al.*, 1991), under continuous light conditions in Rumed 5001, Rumed 1200 or Percival CU-365/D growth cabinets (light intensity 250-300 Lux). Selection was carried out on 25 mg/l sulfadiazine for pKOcol1, 15 mg/l hygromycin for pKOcol2, and 50 mg/l G418 for pKOcol3 single transformants during the first round of selection; the concentration of sulfadiazine and hygromocine was increased to 50 mg/l and 25 mg/l, respectively, during consecutive rounds of selection. Double and triple transformants were selected accordingly, by employing the respective combinations of antibiotics. Stable transformants were grown up and total DNA was prepared as soon as sufficient plant material had been obtained.

2.2.9 PCR analysis of *Physcomitrella* transformants

Targeting of the *PpCOL1* locus was analysed by PCR amplification (Roche Taq polymerase) with primers 5-M-AX180961 and 3-M-AX180961 (“gene” PCR), 5-AX-anti-I and GUS-1877 (“5’ targeting” PCR), pcrsulout1-2 and 3-AX-anti-I (“3’ targeting” PCR), and by PCR amplification (Roche Expand Long Template polymerase) with primers 5-AX-anti-I and 3-AX-anti-I (“across locus” PCR). Targeting of the *PpCOL2* locus was analysed by PCR amplification (Roche Taq polymerase) with primers 5-M-BJ166101 and 3-M-BJ166101 (“gene” PCR), 5-101-anti-III and GUS-1877 (“5’ targeting” PCR), HPT-1167 and 3-101-as-I-bis2 (“3’ targeting” PCR), and by PCR amplification (Roche Expand Long Template polymerase) with primers 5-101-anti-III and 3-101-as-I-bis2 (“across locus” PCR). Targeting of the *PpCOL3* locus was analysed by PCR amplification (Roche Taq polymerase) with primers 5-M-BJ166351 and 3-M-BJ166351 (“gene” PCR), 5-351-anti-I and GUS-1877 (“5’ targeting” PCR), NPT-706 and 3-351-anti-I (“3’ targeting” PCR), and by PCR amplification (Roche Expand Long Template polymerase) with primers 5-351-anti-I and 3-

351-anti-I ("across locus" PCR).

2.2.10 Histochemical detection of GUS activity

Physcomitrella tansformants and the untransformed wild type were grown on minimal medium, in a Rumed 5001 growth cabinet (light intensity ~250 Lux) under long-day conditions (16L:8D). Protonema and gametophore tissue were harvested after 19 days. Histochemical reactions with the indigogenic substrate, X-Gluc, were performed by vacuum infiltration for 3 x 5 min in 10ml 100mM NaH_2PO_4 (pH 7,0) with 0,638 mM substrate, followed by overnight incubation at 37°C under continuous shaking. After staining, the material was rinsed twice with 96% ethanol, then photographed.

2.2.11 Phenotypical analysis of transgenic plants

2.2.11.1 Flowering time analysis of *Arabidopsis* transformants

Arabidopsis seeds were distributed on moist filter paper in a petri dish, seed dormancy was broken by incubation for four days in a refrigerator at 4°C, and germination was induced by incubation for four more days in a growth cabinet at 21°C under long-day (16L:8D) or short-day (8L:16D) conditions. After germination, 20 seedlings of each line were transferred to soil and continued growing in the greenhouse under the same day length conditions. The total numbers of leaves (rosette and cauline) was counted until the appearance of the inflorescence bud. Average flowering time and standard deviation were calculated for each line.

2.2.11.2 Growth analysis of *Physcomitrella* disruptants

Routinely propagated *Physcomitrella* material was used for inoculation on on minimal medium supplemented with 0,5 g/l diammonium tartrate, in a Rumed 5001 growth cabinet under conditions of continuous light (light intensity ~250 Lux). 5 day old protonemal tissue was harvested and

protoplasts were isolated as described (Schaefer and Zryd, 1997). Protoplasts were regenerated in liquid minimal medium supplemented with 0,5 g/l diammonium tartrate and 66 g/l mannitol (Hohe *et al.*, 2004), in a Rumed 1301 growth cabinet under long-day conditions of 16L:8D (light intensity ~100 Lux). When protoplasts had regenerated and had started dividing, they were plated on solid minimal medium supplemented with 0,5 g/l diammonium tartrate, and the plates transferred to long-day (16L:8D) growth conditions in a Rumed 5001 growth cabinet (light intensity ~150 Lux). The growth of the cultures was documented photographically.

3 ISOLATION AND CHARACTERISATION OF *PPCOL1*, *PPCOL2* AND *PPCOL3*

3.1 Introduction

Available *Physcomitrella* genomic resources were exploited for the identification of *CO*-like genes. The public *Physcomitrella* EST collections promise to cover a wide range of the genome, because transcript samples originate from various tissue types and growth conditions. The public collection is an ongoing project, and new sequences have been added throughout the course of this study. At the onset, about 50.000 ESTs had been deposited to GenBank (release 131.0, Sep. 2002). During a later stage, a new exhaustive EST sequencing effort culminated in a new *Physcomitrella* EST database: *Physcobase* (Nishiyama *et al.*, 2003; http://moss.nibb.ac.jp). Herein, 85.191 new ESTs were combined with previously available ESTs, amounting to a total of >102.000. In addition, the new database had two major improvements. Firstly, redundant sequences had been substituted by their consensus sequence, facilitating data mining as well as improving sequence quality. Secondly, more than 40.000 cDNA clones had been sequenced from both ends. As a result, almost 40% of the 15.883 putative transcripts of the database contained sequence information from both ends. Finally, an independent, proprietary *Physcomitrella* cDNA collection has been produced (Rensing *et al.*, 2002b). RNAs had been collected from all stages of the life cycle, to produce a sequence database that was estimated to cover the *Physcomitrella* transcriptome to at least 95% (Rensing *et al.*, 2002b; Rensing *et al.*, 2002a). Each of these databases was searched for *CO*-like genes.

3.2 Results

3.2.1 Identification of *CO* homologues in *Physcomitrella*

Physcomitrella EST sequences deposited to GenBank (release 131.0; Sep. 2002), translated in all six reading frames, were searched for sequences showing homology to the CONSTANS protein sequence. In total, 65 5'- and 3'-EST sequences were retrieved. The 5'- and 3'-EST sequences were aligned separately and those with identical, overlapping nucleotide sequences were identified. For each group of redundant sequences, the EST that yielded the most protein sequence information - typically but not necessarily the longest - was selected as a representative (data not shown). As a result, the dataset was reduced to 12 unique 5´-EST sequences and 10 unique 3´-EST sequences (Table 1). Both defining domains of *CO* or a *CO* homologue are located at the extremities of the gene product. Therefore, 5'-ESTs represent *Physcomitrella* transcripts that possess one or two B-boxes, whereas 3'-ESTs represent *Physcomitrella* transcripts that possess a CCT domain. Because in *Arabidopsis*, B-box and CCT domains also occur in unrelated proteins, it was unclear how many ESTs represent true *CONSTANS* homologues. In addition, because *CONSTANS* is a member of a multi-gene family in *Arabidopsis*, transcripts that do contain both domains are not said to be *CONSTANS* homologues. For these reasons, it was decided to focus on those ESTs that are most similar to *CONSTANS*, then to verify whether they correspond to a transcript that contains both domains. Two alignments were generated. The first one included the predicted protein sequences of the 12 unique 5'-EST representatives and the corresponding amino terminal sequences of the 17 *Arabidopsis* CONSTANS paralogues. The second one included the predicted protein sequences of the 10 unique 3'-EST representatives and the corresponding carboxy terminal sequences of the 17 *Arabidopsis* CONSTANS paralogues. From these alignments, phylogenetic trees were constructed and the ESTs that are most closely related to *CONSTANS* were identified. Three 5'-ESTs were found to be most similar to *CONSTANS*: BJ195918, BJ194188 and BJ190646 (data not shown). They encode two canonical B-boxes, just like *CO* and its most closely related paralogues, *COL1* to *COL5*. Besides, three 3'-ESTs

Gene / Putative transcript	Unique representatives of redundant ESTs identified in GenBank		cDNA clones identified in *Physcobase*
	5′-EST	3′-EST	
PpCOL1	BJ195918	BJ167752	pph16d03, pph16d22, pph21b18, pphn25b07, pphn41d16, pphn41f20, pphb10g17, pphb16k16, pphb19d15, ppspm1m15
PpCOL2	BJ194188	BJ166101	pph23i08, pph6n05, pphn20k02, pphb11i03, pphb24j02, pphb28k14, pphb30a03, pphb43g01, pphb5i15, pphb7j16, ppsp13k22, ppsp1c23, ppsp44a18
PpCOL3	BJ190646	BJ166351	pphn10j02, ppsp1n21
4	BJ162754	BJ170770	pph26n04
5	-	BJ170770	pphn23m24, pphn43b23, pphn47n05, pphn50m18
6	BJ178325	BQ827011	pphb22j04
7	-	BQ827011	pphn22b20, pphn27f15, pphn49n19
8	BJ180416	-	pph19e07, pph27d19, pphn48h22, pphb28p10, pphf23d06, pphf17l18, ppsp12i06
9	BJ202256	-	pphn30k18, pphnx45b15, pphf18o23, ppsp22i05
10	BJ191256	-	pphn12e26, pphn12k14, pphn24l20, pphn28j15, pphn31p18, ppsp31e06, ppsp3k16
11	BJ183144	-	pphb36i07
12	BJ174254	-	pphb11m01
13	BJ174595	-	pph25e0, pphn18p07, pphn29a06, pphb12c08, pphf9b14
14	BJ201954	-	pph11a15, pph16j12, pph17k06, pph31m21, pph32j08, pph35j02, pphn39d21, pphn44e15, pphn44g16, pphn44n08, pphb8i01, pphf17h12
15	-	BJ173170	pph29a02, pphn44b15, pphn49f03, pphb32l11, pphb37o04, pphb39i01, pphb4e07, pphf8d19
16	-	-	pphn22e05, pphn29h18
17	-	BI437331	-
18	-	BQ827714	-

Table 1 *Physcomitrella patens* transcripts showing significant homology to the CONSTANS protein sequence. Listed are the accession numbers of unique representatives of redundant ESTs found in the public EST database (Sep. 2002, GenBank Release 131.0), and accession numbers of all cDNA clones identified in *Physcobase* (Jan. 2005) (Nishiyama *et al.*, 2003; http://moss.nibb.ac.jp). The nomenclature of *Physcobase* cDNA clones refers to the different libraries that the clones originate from: pphXXXXX (clone from non-treated library), pphnXXXXX (clone from auxin-treated library), pphbXXXXX (clone from cytokinin-treated library), pphfXXXXX (clone from first protoplast cell division library), and ppspXXXXX (clone from sporophyte library) (http://moss.nibb.ac.jp).

were identified as most similar to *CONSTANS*: BJ167752, BJ166101 and BJ166351 (data not shown). They were the only ESTs to contain the carboxy terminal T motif in addition to the CCT domain, similarly indicating a close relationship with *CO* and *COL1* to *COL5*. Degenerate primers were designed in order to amplify the genes corrseponding to the three 5'-ESTs and three 3'-ESTs. The upstream primer was designed to hybridise to all three 5'-ESTs in a region that corresponds in the protein to the highly conserved B-box domain, whereas the downstream primer was designed to hybridise to all three 3'-ESTs in a region that corresponds to the T motif, which is specific for Group 1 *CO* homologues. Using these two primers, three discrete products were amplified from *Physcomitrella* cDNA, reamplified, cloned, and sequenced. This revealed that the fragments represent three unique genes, and that each corresponds to one of the three 5'- and 3'-EST pairs. In conclusion, three unique genes had been identified that together represent the most closely related *CONSTANS* homologues that are contained in the public EST database. Moreover, the fact that they possess two canonical B-boxes as well as a T motif indicates that they are members of Group 1 of *CO*-like genes, and that they possibly represent *CONSTANS* orthologues. Based on the sequences of these ESTs, new gene-specific primers were designed, and complete coding and genomic sequences were cloned for each gene. The gene corresponding to ESTs BJ195918 and BJ167752 was called *PpCOL1*, the gene corresponding to BJ194188 and BJ166101 was called *PpCOL2*, and the gene corrsponding to BJ190646 and BJ166351 was called *PpCOL3*. *PpCOL1* and *PpCOL2* correspond to a gene and an EST, respectively, that were reported during the course of this work (Griffiths *et al.*, 2003; Shimizu *et al.*, 2004).

The search for *CONSTANS* homologues was later repeated on the enhanced EST database, *Physcobase* (Nishiyama *et al.*, 2003; http://moss.nibb.ac.jp), in the same way as before. This analysis identified seventeen putative transcripts. The nucleotide sequences of the transcripts were aligned with the 5'- and 3'-EST sequences that had been identified in the previous database search. Identical sequences were removed (data not shown). As such, altogether eighteen unique sequences were retained from both databases for further analysis (Table 1).

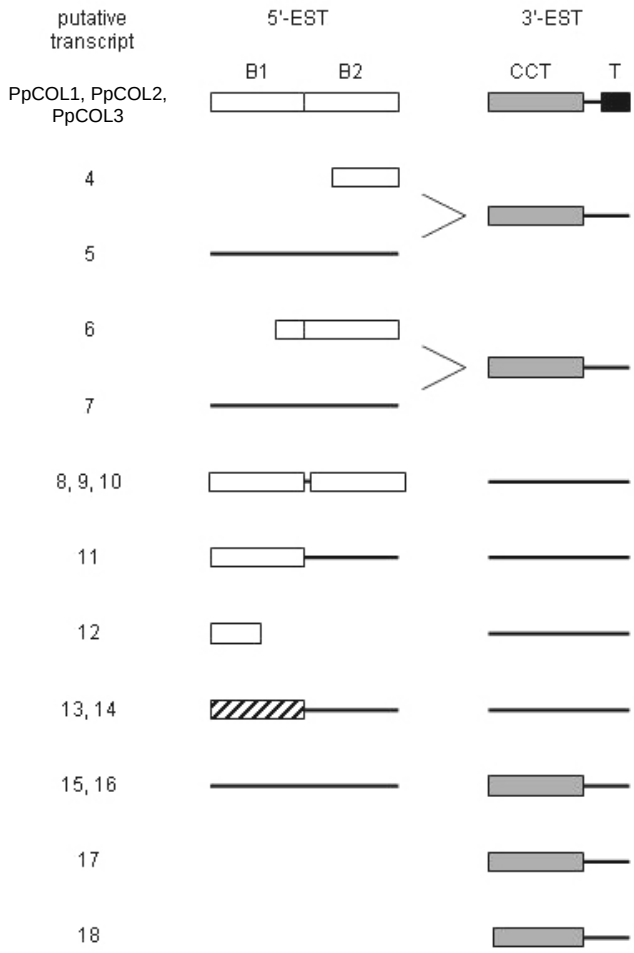


Figure 5 Schematic representation of 18 unique *Physcomitrella* genes and putative transcripts that show homology to the CO protein sequence. Black lines represent relevant sequences showing no homology to CO, whereas boxes represent conserved domains of the CO protein. Canonical CO-like B-box domains are shown as white rectangles, similar but different domains as dashed rectangles. Grey rectangles represent CCT domains, whereas T domains are shown as black rectangles.

For sixteen out of eighteen transcripts, sequence information was available from both ends of the transcript, which allowed distinguishing between those that contain both conserved domains of a *CO*-like gene and those that contain only one of them, or put differently, between those that are likely to represent *CO* homologues, and those that represent unrelated proteins. Three transcripts corresponded to *PpCOL1*, *PpCOL2*, and *PpCOL3*. Two transcripts encoded a B-box and a CCT domain (transcripts 4 and 6) (Figure 5). However, the 3'-ESTs of the latter were also found concatenated with 5'-ESTs that did not encode a B-box (transcripts 5 and 7) (Figure 5).

They either represent splice variants or aberrant transcripts. Several transcripts encoded only a B-box (transcripts 8 to 14) or only a CCT domain (transcripts 15 and 16) (Figure 5). Finally, two transcripts encoded a CCT domain, with no sequence information from the other end of the transcript (transcripts 17 and 18) (Figure 5). Taken together, among eighteen putative transcripts with significant homology to *CONSTANS*, at least five and not more than seven transcripts represent *CO*-like genes, as judged by the presence of a B-box region and a CCT domain.

Finally, the search for *CONSTANS* homologues was repeated in the proprietary *Physcomitrella* EST database (Rensing *et al.*, 2002b). This database is estimated to cover the Physcomitrella transcriptome to at least 95% (Rensing *et al.*, 2002b; Rensing *et al.*, 2002a). The search retrieved *PpCOL1*, *PpCOL2*, and *PpCOL3*, but no more genes that are more closely related to *CONSTANS* (data not shown).

The isolated B-box1, B-box2, B-box1+2, and CCT domain protein sequences encoded by the *Physcomitrella* transcripts were identified. They were aligned with the corresponding motifs from all CO-like proteins of *Arabidopsis* and from unrelated *Arabidopsis* proteins STO and TOC1, to construct the phylogenetic trees shown in Figure 6A-D. The analysis with each separate motif placed the predicted proteins at comparable positions of the tree, and identified *PpCOL1*, *PpCOL2*, and *PpCOL3* as the only Group 1 *CO*-like genes and thus as the genes with highest similarity to *CONSTANS*. The proteins encoded by transcripts 4/5 and 11 grouped more closely to Group 2 CO-like proteins AtCOL6-8 and AtCOL16 (Figure 6A,D), whereas the proteins encoded by transcript 6 (Figure 6B), and transcripts 6/7 and 18 (Figure 6D) grouped more closely to Group 3 CO-like proteins AtCOL9-15. The B-box containing proteins encoded by transcripts 8, 9 and 10 grouped more closely to *Arabidopsis STO* (Figure 6A,B,C). As they were also found to be homologous to STO at the carboxy terminus (data not shown), these transcripts likely represent homologues of *STO*. The proteins encoded by transcripts 15, 16 and 17 grouped more closely to *Arabidopsis TOC1* (Figure 6D), although at least for transcripts 15 and 16 no homology to *TOC1* or any other member of the *APRR1*/*TOC1* gene family was found at the other end of the protein (data not shown). Transcript 12 was shorter than the other transcripts and therefore revealed only part of a canonical B-box, and could

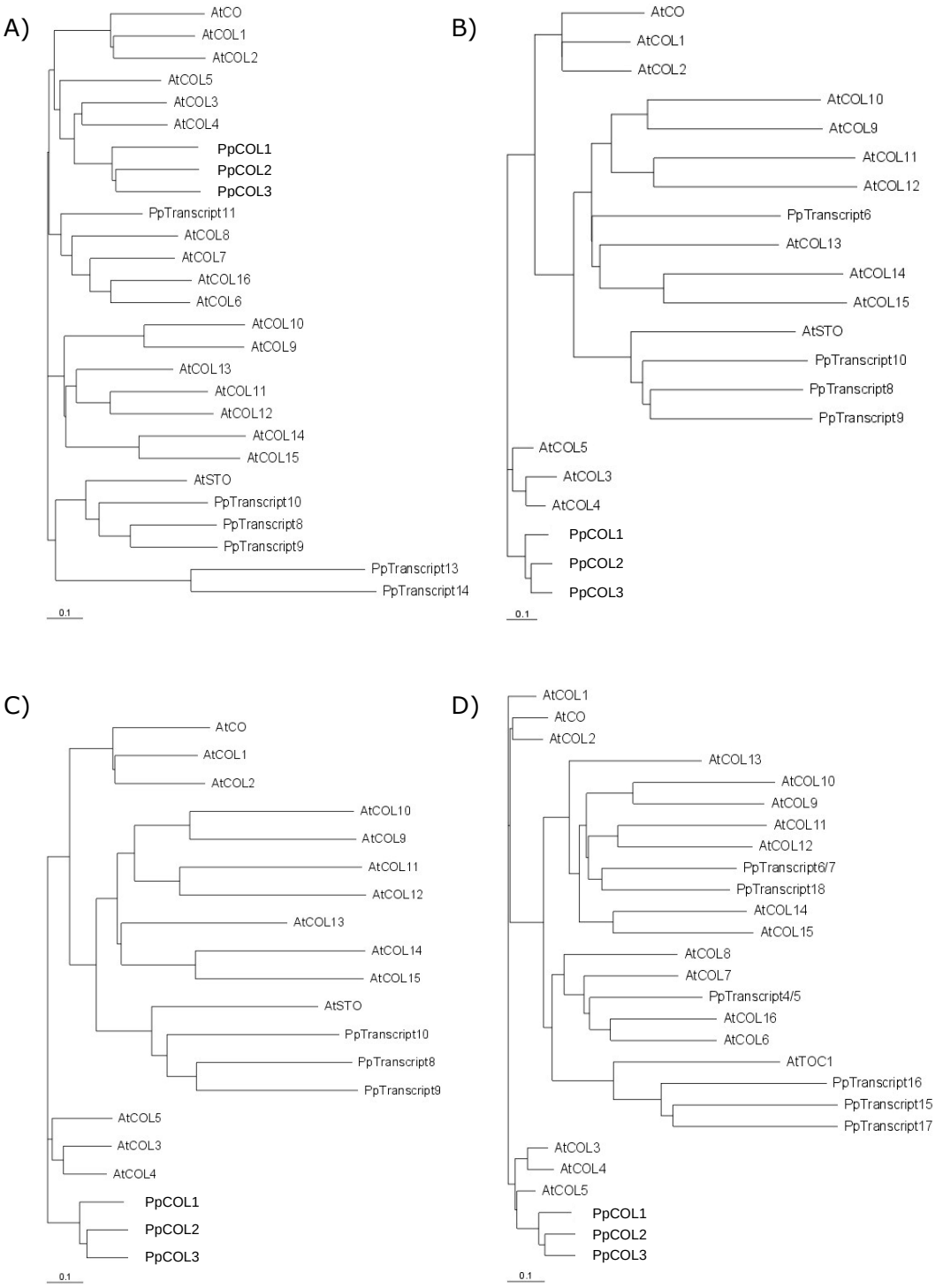


Figure 6 Phylogenetic analysis of isolated domains of CO-like proteins from *Physcomitrella* and *Arabidopsis*. The trees shown are unrooted and based on alignments of (**A**) B-box 1, (**B**) B-box 2, (**C**) B-box 1 and 2, and (**D**) CCT domain protein sequences predicted for CO-like genes from *Physcomitrella* and *Arabidopsis*. The line length indicates genetic distance. The accession numbers of *Physcomitrella* transcripts can be found in Table 1. The accession numbers of *Arabidopsis* sequences are given in the chapter *Materials and methods*.

not be included in the phylogenetic analysis. However, the other end of the protein was not homologous to *CONSTANS* or any other protein in the database (data not shown). Finally, the proteins encoded by transcripts 13 and 14 harbour a motif that is related to, but distinct from a B-box zinc finger motif (Figure 7). The two first, metal-binding cysteine residues of the B-box consensus sequence were absent (Borden *et al.*, 1995; Borden, 1998), although other differently spaced cysteine and histidine residues are present (Figure 7). No similar motifs were found in the PROSITE directory of protein families and domains (http://www.expasy.org/prosite), or in GenBank (release 150.0) by BLAST search (data not shown). Therefore, it possibly represents a novel type of zinc finger.

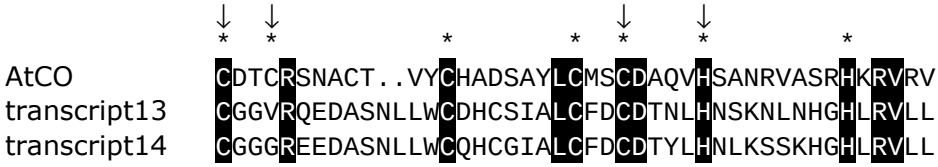
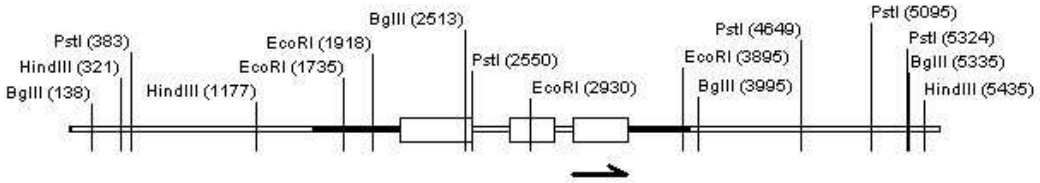


Figure 7 Alignment of the amino acid sequence of B-box1 of AtCO and related domains found in *Physcomitrella* putative transcripts 13 and 14. The consensus cysteine and histidine residues of B-box zinc fingers are indicated by stars, those that have been shown to bind zinc in the NMR structure of the XNF7 B-box (Borden *et al.*, 1995) are additionally indicated by arrows. Identical amino acids are highlighted in black.

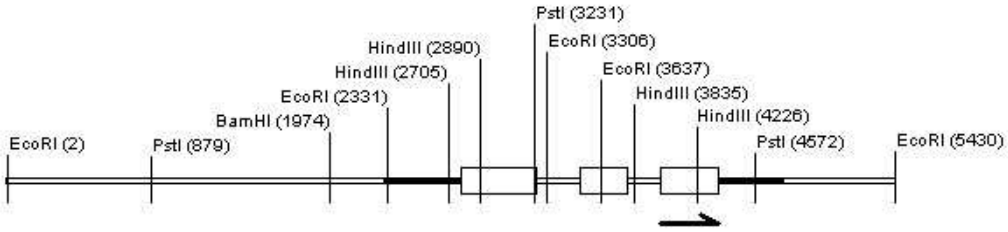
3.2.2 Isolation and characterisation of *PpCOL1*, *PpCOL2*, and *PpCOL3*

The genomic and coding DNA sequences of *PpCOL1*, *PpCOL2*, and *PpCOL3* were aligned to each other and to the original EST sequences to determine the exon-intron structure and the transcript boundaries. Around 5 kb of flanking genomic sequence was obtained for each gene by inverse PCR, information that was also required for targeted gene replacement (see Chapter 5). All sequences were assembled to deduce the gene structures that are shown in Figure 8. The fully annotated genomic sequences are available in GenBank/EMBL with accession numbers AJ890106, AJ890107, and AJ890108 for *PpCOL1*, *PpCOL2* and *PpCOL3*, respectively.

PpCOL1 (5534 bp)



PpCOL2 (5434 bp)



PpCOL3 (5559 bp)

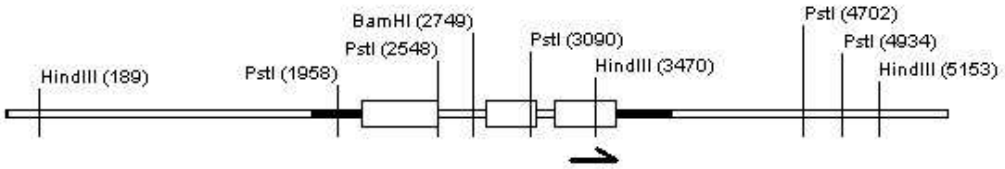


Figure 8 Schematic representation of the genomic sequence of *PpCOL1*, *PpCOL2*, and *PpCOL3*. Coding regions are shown as white rectangles, the 5'- and 3'-untranslated regions, as defined by the ESTs, are shown as black bars. Restriction sites of restriction endonucleases used for Southern blot analysis (Figure 9, Figure 10) are indicated. The regions of the three genes used as probe in the Southern blot analysis are shown as black arrows.

In contrast to *CONSTANS*, which contains one intron, the *PpCOL1*, *PpCOL2*, and *PpCOL3* genes contain two introns. The first intron of *PpCOL1*/*PpCOL2*/*PpCOL3* is found at a location that is not observed for any *CO*-like gene of *Arabidopsis*, shortly after the M1 motif of the middle region (Figure 12, page 55). The second intron is located between the M3 and M4 motifs of the middle region, a position that corresponds to the position of the single intron in *CONSTANS* and other *Arabidopsis* Group 1 and Group 2 *CO*-like genes (Griffiths *et al.*, 2003).

Although the introns of *PpCOL1*, *PpCOL2*, and *PpCOL3* are at highly conserved positions relative to the protein sequence, they differ in length. The first introns of *PpCOL1*, *PpCOL2*, and *PpCOL3* are 243, 269, and 280 bp long, respectively; the second introns are 125, 205 and 115 bp long, respectively. The sequences of the first intron are moderately related to each other (47 to 57% identity), whereas the sequences of the second intron appear to be more diverged, largely due to the more significant size differences between them (24 to 68% identity). In contrast, the nucleotide sequences of the three genes' coding regions are rather uniformly and well related to each other: 74% (*PpCOL1-PpCOL3*), 77% (*PpCOL1-PpCOL2*), and 82% (*PpCOL2-PpCOL3*) sequence identity. The flanking 5'- and 3'-untranslated regions (UTRs) were deduced from the predicted open reading frames of the cDNA sequences, and the sequences were found to be quite diverged, although a block of sequence identity of 32 bp immediately preceding the start codon exists between *PpCOL2* and *PpCOL3*, while the corresponding region is entirely absent from *PpCOL1* (data not shown).

A Southern blot analysis was undertaken to confirm the gene structures and to try to identify additional *CO* homologues. This involved two restriction endonucleases (*EcoRI* and *HindIII*), as well as probes from *PpCOL1*, *PpCOL2*, *PpCOL3*, and *AtCO*. Probes were designed such that the chance of cross-hybridisation was highest, among *PpCOL* genes as well as between *PpCOL* genes and other *CO* homologues. Therefore, probes corresponded roughly to the second exon of *AtCO* and the third exon of *PpCOL* genes (Figure 8), all of which encode the highly conserved CCT domain as well as the T motif. Inclusion of the latter is intended to increase the probes' affinity for Group 1 *CO*-like genes. Shown in Figure 9 are four similar blots, each hybridised at low stringency with a different probe. Hardly any discrete bands could be observed when the blot was hybridised with the *AtCO* probe, even though stringency was low, as can be seen by the appearance of the bands of the λx*PstI* size marker (Figure 9). However, hybridisation with the *PpCOL1*, *PpCOL2* and *PpCOL3* probes indicated considerable cross hybridisation between the three genes (shown as arrowheads in Figure 9). The gene structures shown in Figure 8 were confirmed by the Southern blots. In addition to the gene-specific bands, few additional bands were observed. These either represent more distantly related *CO* homologues, or are the

result of unspecific hybridisation, a possibility that is suggested by the appearance of the bands of the λx*PstI* size marker (Figure 9).

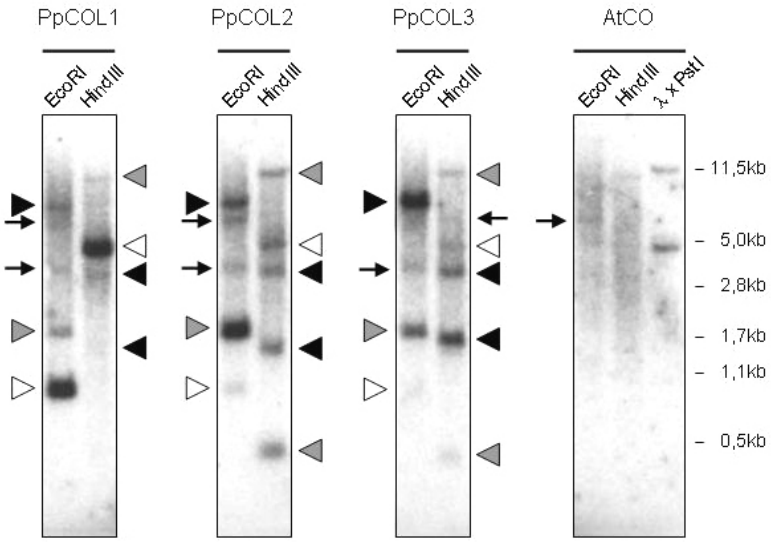


Figure 9 Southern blot analysis. Genomic DNA of *Physcomitrella* was digested with the enzymes indicated and the blot hybridised under low stringency conditions with the probes indicated (*PpCOL1*, *PpCOL2*, *PpCOL3* and *AtCO*). The fragments that were used as *PpCOL1*, *PpCOL2* and *PpCOL3* probes are shown in Figure 8. White, grey and black arrowheads indicate the position of the bands corresponding to *PpCOL1*, *PpCOL2* and *PpCOL3*, respectively. Black arrows indicate bands that do not correspond to any of these genes. The sizes of the DNA size marker (λx*PstI*) are given on the right.

Similar blots were hybridised with one of the three probes, the *PpCOL2* probe, at decreasing stringencies: high, medium, and low stringency. Also, two additional resitriction endonucleases were used (*BglII* and *PstI*). This showed that at high stringency, cross hybridisation was retained between the three *PpCOL* genes, and that no other bands could be observed with the additional resitriction endonucleases (Figure 10). At low stringency (45°C), bands were observed in addition to the bands corresponding to the *PpCOL* genes. However, hybridisation was also observed to fragments of the λx*PstI* size marker, indicating that specificity had been lost at this temperature. When the Southern blot was compared with the original ethidium bromide stained agarose gel that was used for blotting, it was found that all additional bands, in λx*PstI* DNA as well as in *Physcomitrella* genomic DNA, corresponded to positions of high DNA concentration (indicated by stars, Figure 10). Finally,

hybridisation was repeated at medium stringency (55°C). Although most unspecific bands had disappeared at this temperature four faint bands could still be observed (indicated by stars, Figure 10). Comparison with the low-stringency blot and the agarose gel suggests that these bands represent unspecific hybridisation to high concentrations of DNA, although it cannot be excluded that they represent more distantly related *CONSTANS* homologues.

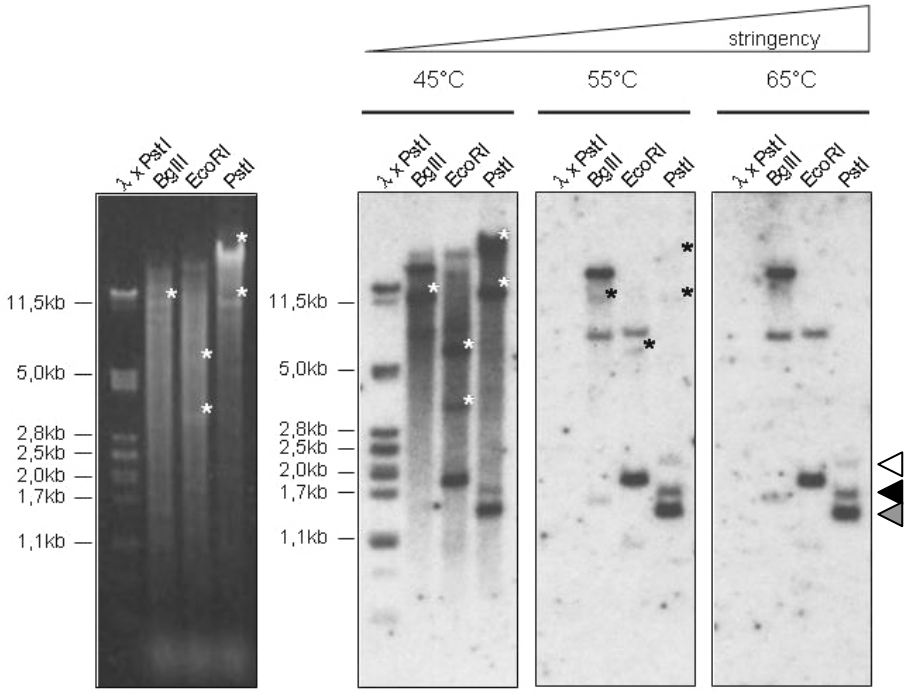


Figure 10 Southern blot analysis. Genomic DNA of *Physcomitrella* was digested with the enzymes indicated and the blot hybridised with the *PpCOL2* probe (shown in Figure 8) under conditions of low (45°C), medium (55°C) and high stringency (65°C). Pictured are the ethidium bromide stained agarose gel before blotting, as well as the resulting Southern blots. Stars indicate presumed unspecific hybridisation in the Southern blots, and the DNA bands that they correspond to in the ethidium bromide stained agarose gel. White, grey and black arrowheads indicate the position of the bands corresponding to *PpCOL1*, *PpCOL2* and *PpCOL3*, respectively, in the rightmost lane (*PstI* digest) of the high stringency blot (65°C). The sizes of the DNA size marker (λx*PstI*) are given on the left of the agarose gel and of the low stringency blot (45°C).

The PpCOL1, PpCOL2, and PpCOL3 predicted protein sequences were aligned with the predicted protein sequences of all *Arabidopsis* CO-like proteins to construct the phylogenetic tree shown in Figure 11. Consistent with the previous analysis of isolated B-box and CCT domains (Figure 6, page

47), PpCOL1, PpCOL2, and PpCOL3 proteins were placed into Group 1. Within this clade, bootstrap support was high for a closer relatedness to Group 1c CO-like proteins than to Group 1a CO-like proteins (Figure 11). Closer inspection of the alignment consistently showed that PpCOL1/PpCOL2/PpCOL3 are more similar to AtCOL3/AtCOL4/AtCOL5, than to AtCO/AtCOL1/AtCOL2 (Figure 12). Thus, PpCOL1, PpCOL2, and PpCOL3 are likely to be representatives of the Group 1c CO-like isoform. Furthermore, a comparison of protein sequence identities shows that the *Physcomitrella* Group 1c proteins are rather similar to one another, whereas the *Arabidopsis* representatives of Group 1a as well as Group 1c are more diverged (Table 2).

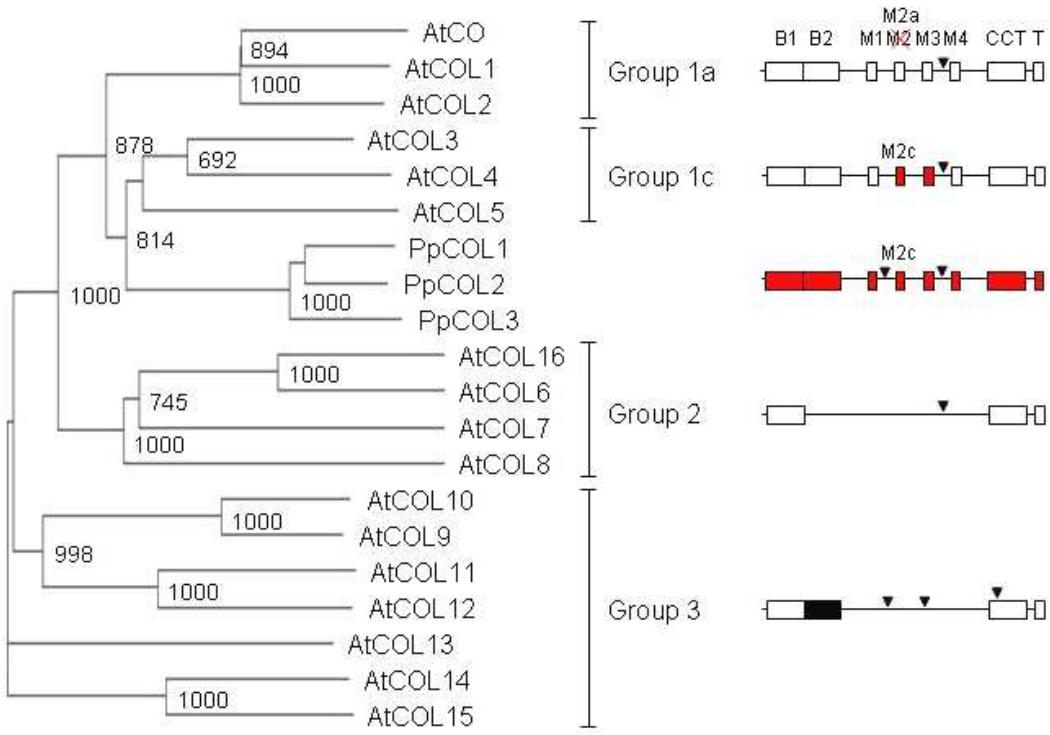


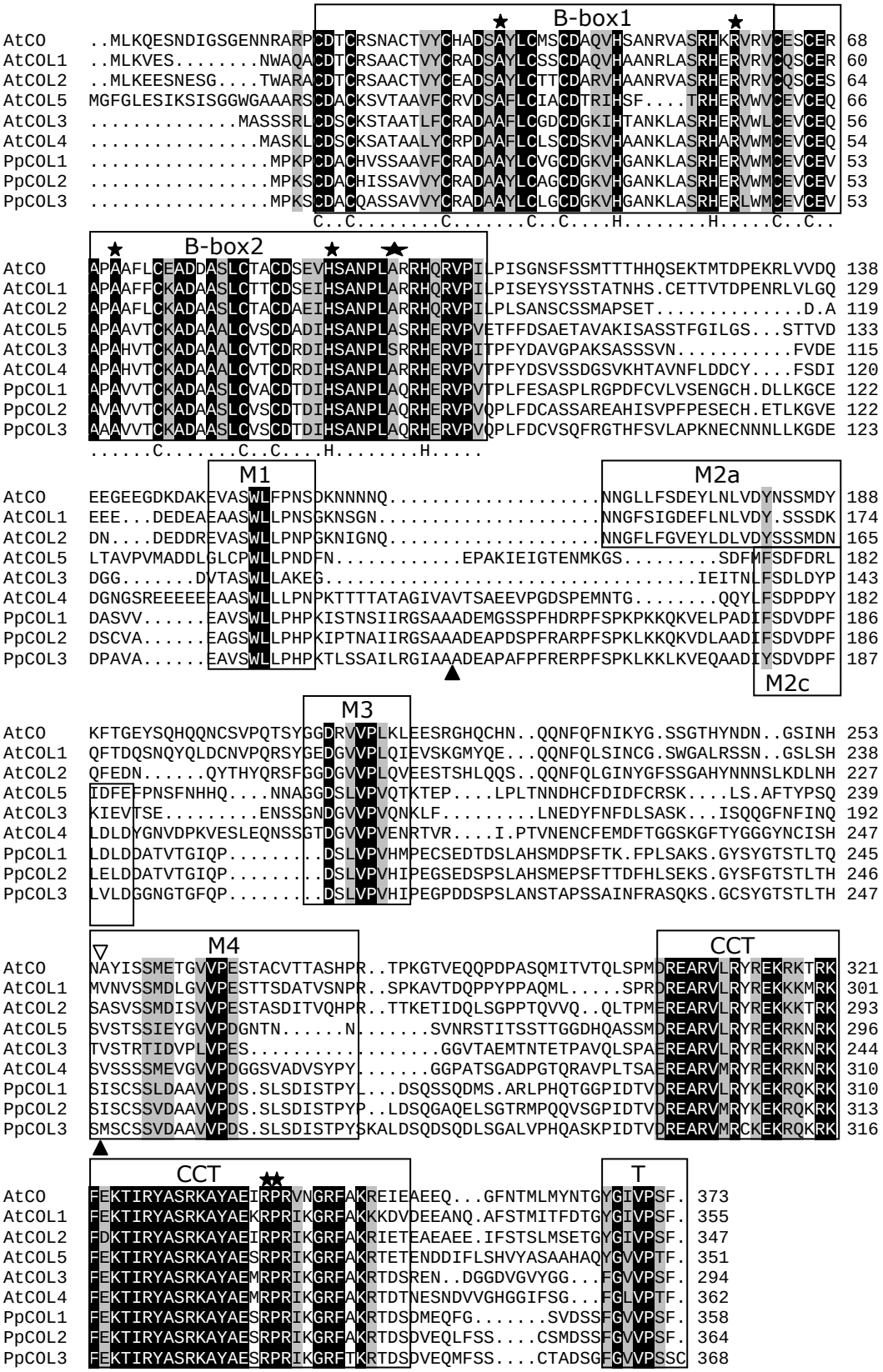
Figure 11 Phylogenetic analysis of CO-like proteins from *Physcomitrella* and *Arabidopsis*. The phylogenetic tree is unrooted and based on the alignment of full length protein sequences predicted for *Physcomitrella* and *Arabidopsis CO*-like genes. Indicated are the bootstrap values obtained from 1000 replicates above a cut off of 690. The domain structures of different groups of CO-like proteins according to Griffiths *et al.* (2003) are shown in black. Additions and changes to the scheme made in this study are shown in red. B1 and B2 are the first and second B-box zinc finger, respectively; the black rectangle indicates a divergent B-box zinc finger. M1 to M4 are conserved middle region motifs. CCT and T are CCT domain and T motif, respectively. Intron positions are shown as solid arrowheads. The accession numbers of the *Arabidopsis* sequences that were used to generate the alignment are given in the chapter *Materials and methods*.

The distinction between Group 1a and Group 1c *CO*-like genes is reflected in conserved motifs of the middle region. While Group 1a genes encode M1, M2, M3, and M4 motifs, Group 1c genes only contain M1 and M4 motifs (Griffiths *et al.*, 2003). However, as shown in the alignment of Figure 12, the M3 domain of Group 1a genes is clearly also present in Group 1c proteins of *Arabidopsis* (*AtCOL3*/*AtCOL4*/*AtCOL5*) and *Physcomitrella* (*PpCOL1*/*PpCOL2*/*PpCOL3*). Moreover, although no M2 motif was detected in Group 1c proteins previously, a different motif is detected at this position in the alignment. Thus, while *Arabidopsis* Group 1a proteins possess a clear motif defined as M2, Group 1c proteins of *Arabidopsis* and *Physcomitrella* possess a different one at a similar position (Figure 12). To account for this finding, the M2 motif was redefined as shown in Figure 12; the M2 region found in Group 1a genes was designated M2a and the corresponding motif in Group 1c genes M2c. The similarity of this motif is particularly apparent between PpCOL1/PpCOL2/PpCOL3 and AtCOL4 (Figure 12).

	AtCO	AtCOL1	AtCOL2	AtCOL3	AtCOL4	AtCOL5	PpCOL1	PpCOL2	PpCOL3
AtCO	100	66	63	37	32	34	31	30	28
AtCOL1		100	66	41	36	33	35	35	34
AtCOL2			100	42	36	36	34	35	33
AtCOL3				100	54	47	41	40	37
AtCOL4					100	44	45	46	42
AtCOL5						100	39	39	36
PpCOL1							100	82	76
PpCOL2								100	81
PpCOL3									100

Table 2 Percentage amino acid identity between representatives of *Physcomitrella* and *Arabidopsis* Group 1 CO-like proteins.

Figure 12 Alignment of AtCO, AtCOL1 to AtCOL5, and PpCOL1 to PpCOL3 predicted protein sequences. Boxed regions indicate conserved domains as defined by Griffiths *et al.* (2003) and this study. Amino acids conserved in all sequences are highlighted in black, similar amino acids are highlighted in grey. Conserved cysteine and histidine residues of the B-box zinc finger (Borden, 1998; Robson *et al.*, 2001) are shown below the alignment. Residues affected in *co* mutant alleles (Robson *et al.*, 2001) are indicated by stars above the alignment. Intron positions in the corresponding nucleotide sequences are indicated by a white triangle for AtCO and AtCOL1 to AtCOL5, and by black triangles for PpCOL1, PpCOL2 and PpCOL3. The *Arabidopsis* sequence accession numbers are given in the chapter *Materials and methods*.


B-box1
B-box2
M1
M2a
M2c
M3
M4
CCT
CCT
T

3.2.3 Constitutive expression of *PpCOL* genes in *Arabidopsis*

The most closely related *CO* paralogues of *Arabidopsis*, *COL1* and *COL2*, were found not be able to substitute *CO* function in *Arabidopsis*, in spite of strong sequence conservation (Ledger *et al.*, 2001). Since particularly the B-box regions and the CCT domain are highly similar between the three proteins, it appears that functional specificity might reside in other parts of the protein.

It was investigated whether functional properties are generally conserved between the PpCOL proteins and CONSTANS. Overexpression of the *CONSTANS* gene from the CaMV 35S promoter has a dominant effect on flowering time in *Arabidopsis* (Onouchi *et al.*, 2000), resulting in significantly earlier flowering than the wild type, both under long days and short days (Samach *et al.*, 2000). This characteristic was used to assay functional conservation between *CONSTANS* and the *PpCOL* genes. A transcriptional fusion was constructed between the CaMV 35S promoter on one hand, and the *PpCOL1*, *PpCOL2*, and *PpCOL3* cDNAs on the other. Transgenic *Arabidopsis* plants containing these constructs were generated. The presence of the transgene was confirmed, and the expression level quantified by quantitative RT-PCR, in T1 plants of several independent transgenic lines (Figure 13). For each transgene, the three highest expressing lines were selected (marked by stars in Figure 13). The presence of the transgenic protein could not be confirmed, because no commercially available CONSTANS antibodies were found that crossreact with the PpCOL1, PpCOL2, and PpCOL3 proteins.

Segregating T2 progeny of the selection of highly overexpressing lines was analysed for flowering time under both long-day and short-day conditions. For *35S::PpCOL1*, three lines were analysed (1.41, 1.42, and 1.45), as well as for *35S::PpCOL2* (2.42, 2.43, and 2.44). Only two lines were analysed for *35S::PpCOL3* (3.43 and 3.44), because the germination rate of line 3.41 was too low. The average flowering time was calculated from tentatively 20 segregating plants. Whereas the *AtCO* overexpressing line flowers significantly earlier than the wild type under long days, and dramatically earlier than the wild type under short days, the average flowering time of *PpCOL1*, *PpCOL2* and *PpCOL3* overexpressing lines was similar to that of the wild type, in both long days and short days (Table 3).

Although a strong effect of the transgene on flowering time should become apparent in the average flowering time of a segregating population, particularly under short day conditions, it cannot be excluded that the *35S::PpCOL1*, *35S::PpCOL2* or *35S::PpCOL3* transgene might have a smaller effect on flowering time. However, the fact that, in long days as well as in short days, no 3-to-1 segregation in flowering time could be observed among individuals of a segregating *35S::PpCOL1*, *35S::PpCOL2* or *35S::PpCOL3* population, but rather showed similar variation as a population of wild type individuals, at least suggests that this is not the case (data not shown).

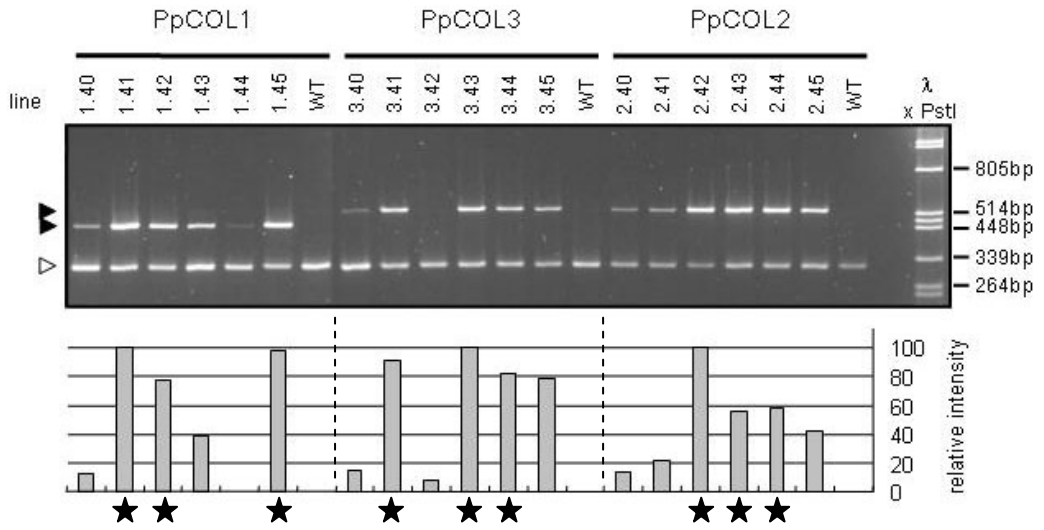


Figure 13 Quantification of *PpCOL1*, *PpCOL2*, and *PpCOL3* expression in transgenic *Arabidopsis* plants by quantitative RT-PCR. The caption above each lane refers to a transgenic T1 plant from which leaf material was harvested and RNA isolated. Plants of lines 1.xx are hemizygous for *35S::PpCOL1*, plants of lines 2.xx are hemizygous for *35S::PpCOL2*, and plants of lines 3.xx are hemizygous for *35S::PpCOL3*. The uppermost caption indicates which gene was analysed (*PpCOL1*, *PpCOL3* or *PpCOL2*). The positions of the internal 18S rRNA standard and the gene-specific product are marked by white and black triangles, respectively. The sizes of the DNA size marker (λx*PstI*) are given on the right. Quantification of transcript abundance, relative to the internal 18S rRNA standard, is shown under the gel picture. The highest relative intensity among lines with the same transgene was taken as 100; others were calculated relative to that. For each transgene, the three lines that have the highest transcript abundance are marked by stars.

Genotype	Total leaf number at flowering	
	Long days	Short days
ColO	10,3 ± 0,6	39,2 ± 2,8
35S::AtCO	4,5 ± 0,5	5,7 ± 0,7
35S::PpCOL1 (1.41)	10,7 ± 1,0	39,2 ± 3,1
35S::PpCOL1 (1.42)	9,6 ± 0,8	38,6 ± 4,1
35S::PpCOL1 (1.45)	9,7 ± 0,7	38,8 ± 5,0
35S::PpCOL2 (2.42)	10,1 ± 1,3	38,5 ± 3,9
35S::PpCOL2 (2.43)	9,6 ± 0,9	40,3 ± 4,1
35S::PpCOL2 (2.44)	9,7 ± 0,8	36,9 ± 2,6
35S::PpCOL3 (3.43)	10,1 ± 1,2	38,4 ± 3,4
35S::PpCOL3 (3.44)	10,2 ± 0,9	39,3 ± 4,2

Table 3 Effect of overexpression of *PpCOL1*, *PpCOL2*, *PpCOL3*, and *AtCO* on flowering time in *Arabidopsis*. Flowering time is represented as the total number of leaves (rosette and cauline) formed on the main shoot; plants that flower later form more leaves. Given are the average value and the standard deviation for a population of 20 individuals. Individuals carrying the *35S::AtCO* transgene are homozygous for the transgene; individuals carrying the *35S::PpCOLx* transgene are segregating. The names of different, independent transgenic lines are given between brackets.

3.3 Conclusions

A search was performed for *CO* in *Physcomitrella*, which involved searching *Physcomitrella* EST libraries for *CO* homologues. For the first time, *CONSTANS*-like genes were identified outside of the realm of flowering plants. At least five and not more than six *CO*-like genes were identified in a database that covers the *Physcomitrella* transcriptome to 60%. One gene belongs to a class of *CO*-like genes called Group 2, two genes belong to Group 3, and three genes belong to Group 1, the class that also *CO* is a member of. The latter three genes were cloned and called *PpCOL1*, *PpCOL2*, and *PpCOL3*. In addition, more than 5kb of flanking genomic sequence was cloned by inverse PCR and the gene structures were confirmed by Southern blotting. Searches of an additional EST database that almost completely covers the *Physcomitrella* transcriptome, and Southern blotting, identified no other Group 1 genes, indicating that *PpCOL1*, *PpCOL2*, and *PpCOL3* represent the most closely related homologues of *CO*. Further analysis revealed that the three *Physcomitrella* homologues are most similar to *AtCOL3*/*AtCOL4*/*AtCOL5*, a class of Group 1 genes closely related to, but distinct from *CO*. These findings

indicate that *CONSTANS* was not present in the last common ancestor of mosses and flowering plants. By sequence alignment, two conserved motifs were identified in the middle regions of PpCOL1/PpCOL2/PpCOL3 and AtCOL3/AtCOL4/AtCOL5, of which the M2c motif was an entirely novel one. Overexpression studies with *PpCOL1*, *PpCOL2*, *and PpCOL3* in *Arabidopsis thaliana* indicated that inherent functional qualities had diverged between CONSTANS and the *Physcomitrella* homologues, consistent with their phylogeny and the observations made with closely related *CO* homologues from *Arabidopsis*.

4 TRANSCRIPTIONAL REGULATION OF *PPCOL1*, *PPCOL2*, AND *PPCOL3*

4.1 Introduction

Circadian regulation of transcription is an essential feature of *CONSTANS* function in *Arabidopsis* (Suarez-Lopez *et al.*, 2001). Interestingly, rhythmic fluctuations in transcript abundance are conserved among *CO*-like genes from flowering plants. The *COL1*, *COL2*, *COL5* and *COL9* genes of *Arabidopsis* all oscillate throughout the day. The cycling profiles of *COL1* and *COL2* are similar to each other, but very different to that of *CO* (Ledger *et al.*, 2001), whereas the cycling profile of *COL9* largely resembles that of *CO* (Cheng and Wang, 2005). *COL5* was found to be expressed diurnally in a study that used microarray technology to identify cycling genes on a genome wide scale (Schaffer *et al.*, 2001). Also, any homologue from any other species that was analysed to date showed diurnal/circadian fluctuations of transcript abundance. The expression profile of the functional orthologue from rice is almost identical, that of the putative orthologue from barley very similar, to the expression profile of *Arabidopsis CO* (Suarez-Lopez *et al.*, 2001; Hayama *et al.*, 2003; Nemoto *et al.*, 2003). Homologues from *Pharbitis*, rice, wheat, and potato on the other hand, display a variety of expression profiles, each with its own characteristic phase, amplitude, and peak width (Liu *et al.*, 2001; Martinez-Garcia *et al.*, 2002; Kim *et al.*, 2003; Nemoto *et al.*, 2003; Shin *et al.*, 2004). Whether this regulatory mechanism is also shown by *CO*-like genes of *Physcomitrella patens* was examined.

The circadian clock is an endogenous timekeeper that allows an organism to keep track of daily and seasonal time. The importance of this mechanism is underscored by its ubiquity; clocks are present in organisms ranging from prokaryotic and eukaryotic microbes to plants, insects and animals (Gillette and Sejnowski, 2005). The clock has been conceptualised as a series of three components: an entrainment pathway that transmits environmental signals to entrain the central oscillator to environmental time; the central oscillator is the core component of the circadian clock that keeps

time and also operates in the absence of environmental cues; output pathways receive temporal information from the central oscillator to regulate rhythmic clock-controlled gene expression and rhythmic biological activity (Bell-Pedersen *et al.*, 2005). An entrainable circadian system is proposed to be advantageous over a purely driven one by providing the possibility of anticipation, that is, preparing physiology before the external environmental changes occur. The ability of the clock to persist in constant conditions is classically regarded as one of the defining characteristics of a circadian rhythm (Roenneberg and Merrow, 2002). In the case of transcriptional regulation, this led to the distinction between genes with a circadian and a diurnal rhythm, because the latter do not persist in constant conditions. In a particular species of cyanobacteria, the circadian clock controls gene expression globally (Liu *et al.*, 1995). In *Arabidopsis*, the number of clock-controlled genes (ccgs) is estimated to be much lower, between 2% (Schaffer *et al.*, 2001) and 6% (Harmer *et al.*, 2000).

4.2 Results

4.2.1 Investigation of diurnal and circadian expression patterns

In order to determine the expression patterns of *PpCOL1*, *PpCOL2*, and *PpCOL3*, mRNA abundance was analysed by quantitative RT-PCR. In our experimental setup, quantification is achieved through an internal 18S rRNA standard. Gene-specific primers were designed to span an intron, such that amplification of possible contaminating genomic DNA will result in a differently sized fragment. The specificity of the primers was confirmed by restriction analysis. The positions of the primers are shown in Figure 21 (page 76).

Young protonema was cultured in a long-day regime (16 hrs of light, 8 hrs of darkness), and RNA was sampled at 4 hr intervals throughout the day. When RT-PCR products were analysed, a third, larger band was observed for *PpCOL1*, *PpCOL2*, and *PpCOL3*, in addition to the gene-specific band and the internal reference band (Figure 14). For each gene, the band corresponded in size to an intron-containing transcript. These fragments did not originate from contaminating genomic DNA, because they were not observed in a mock-

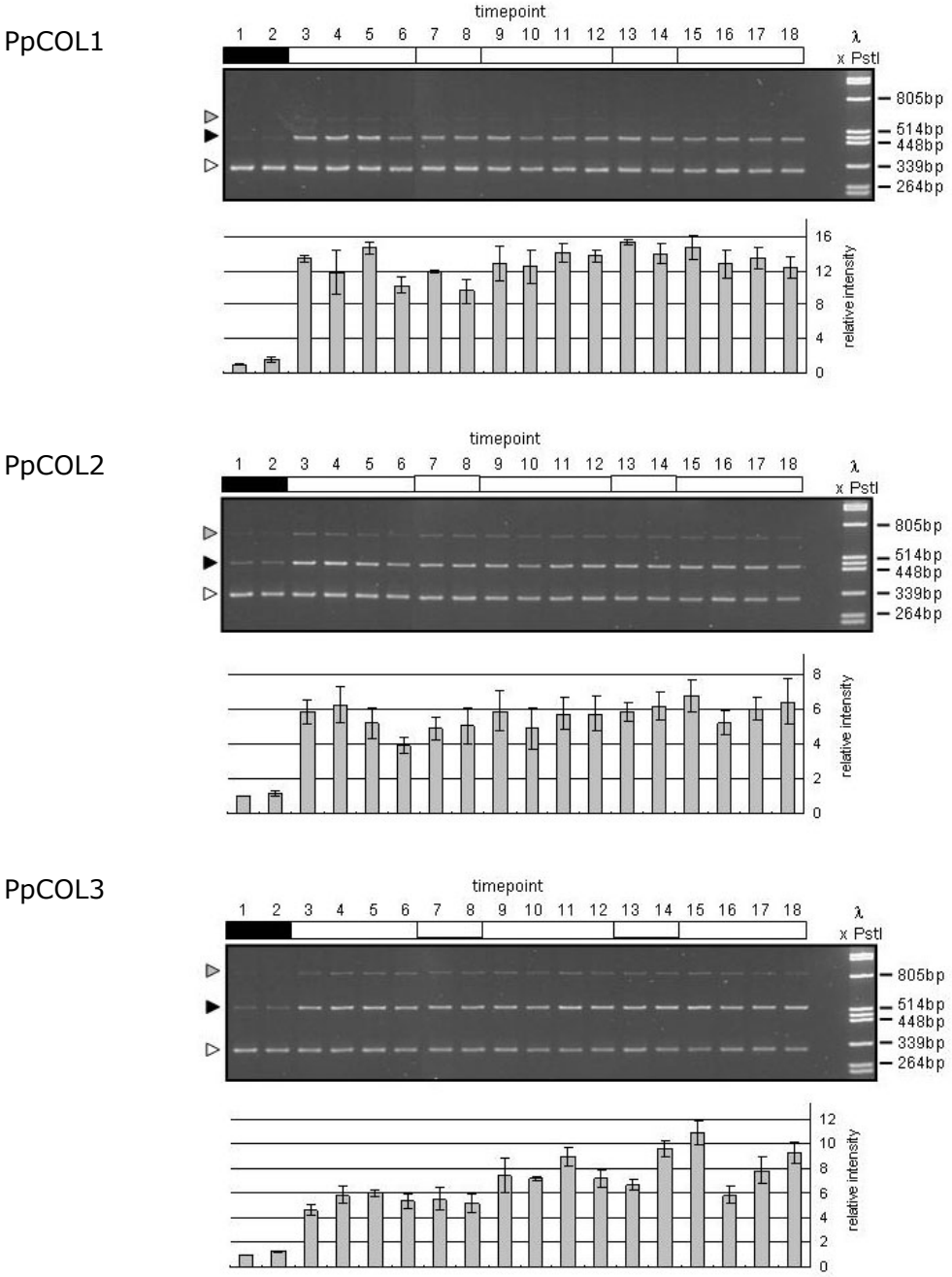


Figure 14 Expression of *PpCOL1*, *PpCOL2*, and *PpCOL3*, under diurnal conditions of light/dark and in continuous light conditions. Expression levels of *PpCOL1*, *PpCOL2*, and *PpCOL3* in protonema were determined by quantitative RT-PCR. Cultures were growing in long days (16 hrs of light, 8 hrs of darkness). The white or black bars at the top indicate light or darkness. Samples were taken every 4 hours, starting 2 hours after dusk (time point 1). Time points 7, 8, 13 and 14 were taken in subjective nights but with lights on. The positions of the internal 18S rRNA standard, the gene-specific product and unspliced gene-specific product are marked by white, black, and grey triangles, respectively. The sizes of the DNA size marker (λx*PstI*) are given on the right. Quantification of transcript abundance, relative to the internal standard, is shown under each gel picture. Intensities are the average of two technical replicates, the values of which are represented by error bars. The lowest intensity (time point 1) was taken as 1; the intensity of every other time point was calculated relative to that.

reaction without reverse transcriptase (data not shown). Therefore, they likely represent a pool of unspliced transcripts. Quantification of the spliced and unspliced fragments in Figure 14 revealed that as much as 15% of the *PpCOL2* and *PpCOL3* mRNA pools might consist of unprocessed transcripts (data not shown). *PpCOL1* appears to be more efficiently spliced (Figure 14).

Under diurnal conditions of light/dark, the PCR product corresponding in size to correctly spliced *PpCOL1*, *PpCOL2*, or *PpCOL3* transcript was found to be present at high levels in all samples exposed to light, and at much lower levels in the dark (Figure 14). The unspliced transcript was found to be cycling in the same way as the spliced transcript, supporting the fact that it does not originate from contaminating genomic DNA. *PpCOL1*, *PpCOL2* and *PpCOL3* all showed very similar patterns of expression. Expression levels strongly increased at dawn and remained high throughout the day. The increase in expression did not start before dawn, but occurred after exposure to light. The increase in expression after light irradiation was twice as high for *PpCOL1* as it was for *PpCOL2* and *PpCOL3* (Figure 14). Quantification of the levels of expression 2 hrs before and 2 hrs after dawn revealed a 12-fold, 5-fold, and 4-fold increase of expression for *PpCOL1*, *PpCOL2*, and *PpCOL3*, respectively (Figure 14).

In order to distinguish between diurnal and circadian regulation of expression, the expression was analysed in free running light conditions. At the end of one day of sampling, the cultures were shifted to continuous light conditions, and sampling was continued for 48 hrs. Again, *PpCOL1*, *PpCOL2*, and *PpCOL3* displayed very similar expression patterns. The expression levels did not decrease in the subjective night when light was present. Instead, expression levels for all three genes remained high throughout the light period (Figure 14). Although fluctuations in transcript could be observed during this time, no circadian periodicity could be identified. Therefore, it was concluded that under conditions of constant light the expression of *PpCOL1*, *PpCOL2*, and *PpCOL3* is arrhythmic.

Other rhythmically expressed genes from *Physcomitrella* were previously shown to become arrhythmic in constant light, and only displayed weak cycling in constant darkness (Aoki *et al.*, 2004; Ichikawa *et al.*, 2004). Therefore, the expression of *PpCOL1*, *PpCOL2* and *PpCOL3* was tested in conditions of extended darkness. Young protonema was cultured, as before, in a long-day regime (16 hrs of light, 8 hrs of darkness). A first RNA sample was

taken in darkness, 2 hours before dawn. At the beginning of day (time = 0), one set of cultures was exposed to light, whereas another set was kept in darkness. Then, samples were taken from each set of cultures, 2 hrs and 6 hrs after actual sunrise (light cultures) or subjective sunrise (dark cultures). The expression was analysed as before, by quantitative RT-PCR (Figure 15). Analysis of the PCR products revealed the presence of faint additional bands that were previously not observed, and that could not be explained. For this reason, the data were not quantified. Consistently, the *PpCOL1*, *PpCOL2* and *PpCOL3* gene-specific bands showed a similar response to light at the beginning of day as they did in the previous experiment (Figure 14). It was observed that the expression of each gene increases also in cultures that are kept in darkness (Figure 15). Whereas expression of *PpCOL2* increased in the absence of light almost as rapidly as it did in the presence of light, the expression of *PpCOL3* was significantly delayed in darkness. *PpCOL1* showed an intermediary response.

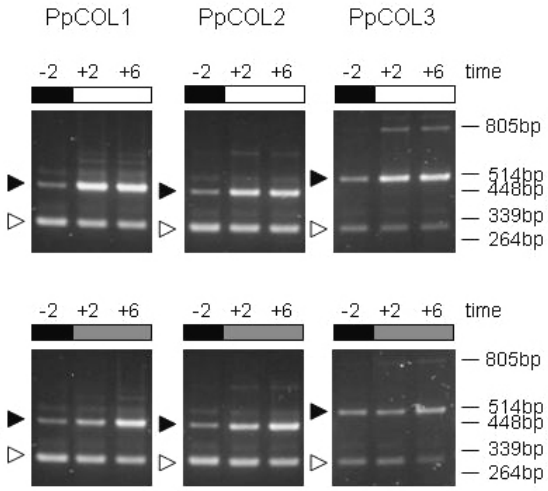


Figure 15 Influence of extended darkness on expression of *PpCOL1*, *PpCOL2*, and *PpCOL3*. Expression levels of *PpCOL1*, *PpCOL2*, and *PpCOL3* in protonema were determined by quantitative RT-PCR. Cultures were growing in long days (16 hrs of light, 8 hrs of darkness). At respective dawn (time = 0), cultures were either irradiated with white light (top row), or kept in darkness (bottom row). The black bar above each picture indicates darkness during nighttime; the white and grey bars represent light and darkness, respectively, during subjective daytime. The positions of the internal 18S rRNA standard and the gene-specific product are marked by white and black triangles, respectively. The sizes of the DNA size marker (λx*PstI*) are given on the right.

4.2.2 Analysis of the effect of light quality on expression

Our findings indicated that although *PpCOL* mRNA levels do rise at dawn in the absence of light, they do so more slowly than if light were present. Therefore, during the first hours of the morning, light has a direct effect on gene expression. Experiments were designed to investigate whether this response depends on light of a particular wavelength. Response to a particular wavelength might indicate a link between *PpCOL* expression and light signalling through a particular photoreceptor. This could provide an indication of the function of *PpCOL1*, *PpCOL2*, and *PpCOL3*, because *Physcomitrella* photoreceptors have established roles in distinct developmental and physiological processes (Imaizumi *et al.*, 2002; Kasahara *et al.*, 2004; Mittmann *et al.*, 2004). Again, young protonema was cultured in a long-day regime (16 hrs of light, 8 hrs of darkness). RNA was sampled 90 min and 30 min before sunrise, and 30 min and 90 min after sunrise (samples 5, 6, 9, and 10; Figure 16A). At the same time points, RNA samples were taken from cultures that had been deprived of light at the subjective morning (samples 7 and 8; Figure 16A), as well as from cultures that were not irradiated with white light at sunrise, but with blue light (samples 3 and 4, Figure 16A), red light (samples 11 and 12, Figure 16A), or far-red light (samples 13 and 14, Figure 16A). The expression was analysed as before, by quantitative RT-PCR (Figure 16B), and the results were quantified (Figure 17).

In line with the previous finding of *PpCOL1* showing the sharpest increase in transcript abundance at dawn (Figure 14), the clearest observations could be made for this gene. Cultures that had been kept in darkness at subjective dawn showed no or hardly any increase in *PpCOL1* expression, up to 90 minutes after subjective sunrise (Figure 17). However, a 4- to 6-fold increase in expression could already be observed half an hour after sunrise in white, blue, red, and far-red light (Figure 17). An hour later, a 6- to 10-fold increase was observed for the different light qualities.

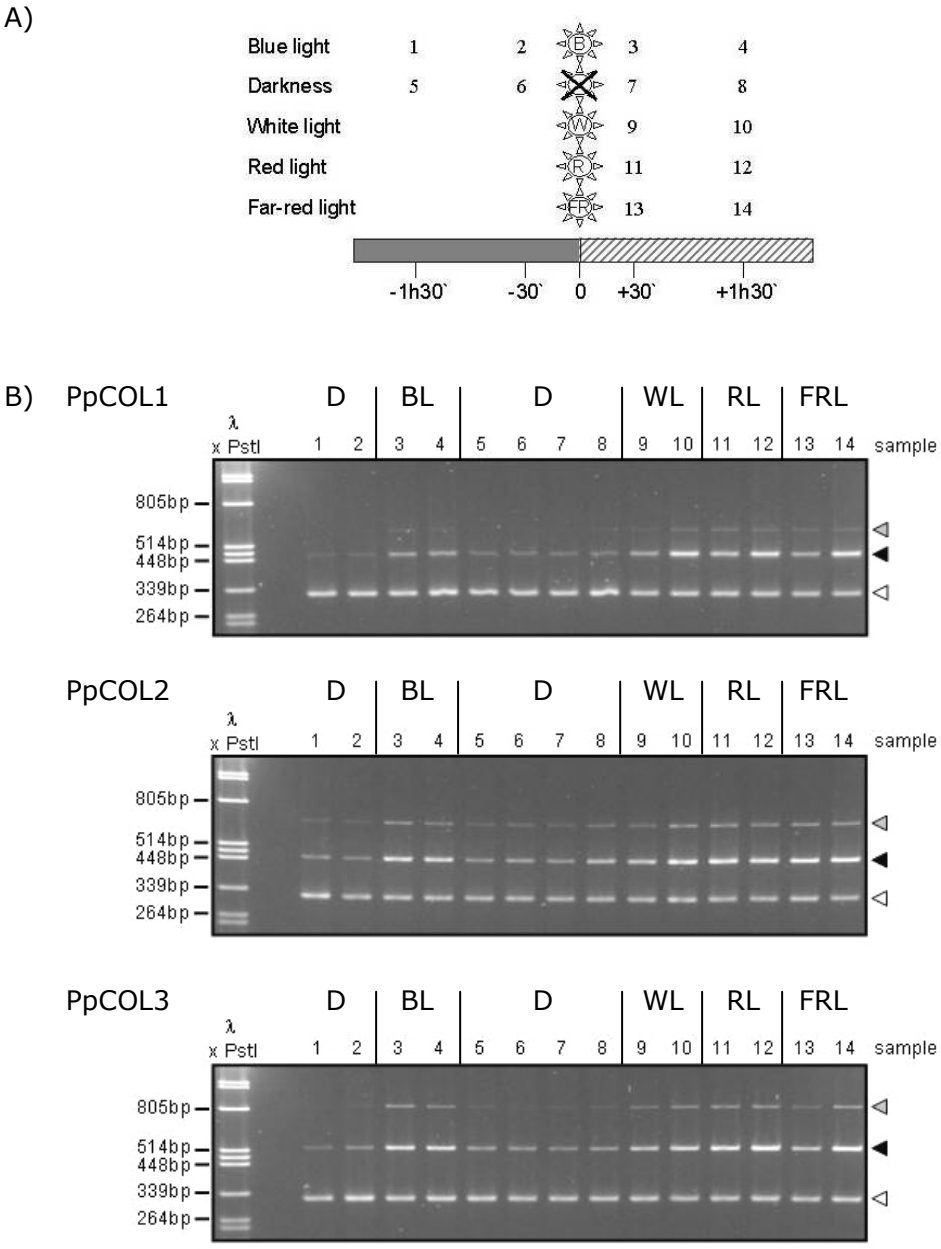


Figure 16 Influence of light and light quality on expression of *PpCOL1*, *PpCOL2*, and *PpCOL3* at dawn. (**A**) Schematic representation of the sampling regime around dawn. The grey bar indicates nighttime, whereas the dashed bar represents subjective day; sampling times are indicated underneath. The time of subjective sunrise is 0. Samples 1, 2, 5 and 6 were taken in darkness during nighttime; other samples were taken during the subjective morning, in the absence of light (samples 7 and 8), and in the presence of blue light (samples 3 and 4), white light (samples 9 and 10), red light (samples 11 and 12), or far-red light (samples 13 and 14). (**B**) Expression levels of *PpCOL1*, *PpCOL2*, and *PpCOL3* in protonema determined by quantitative RT-PCR at the time points and light conditions schematised in (A). D, BL, WL, RL, and FRL indicate darkness, blue light, white light, red light, and far-red light, respectively. Sample numbers correspond to the numbers in (A). The positions of the internal 18S rRNA standard, the gene-specific product, and the unspliced gene-specific product are marked by white, black, and grey triangles, respectively. The sizes of the DNA size marker (λx*PstI*) are given on the left.

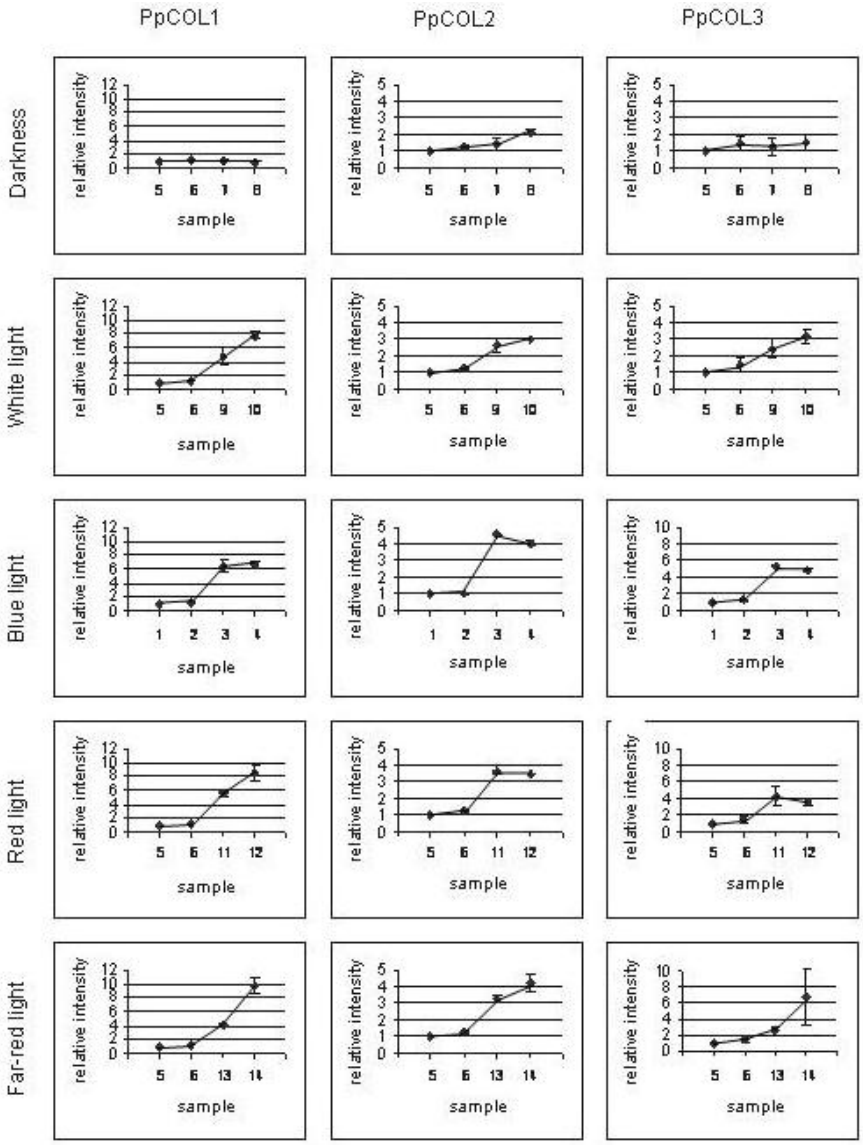


Figure 17 Quantification of transcript abundance shown in Figure 16. Intensities of the gene-specific spliced product are calculated relative to the intensity of the internal 18S rRNA standard. Shown is the average of two technical replicates, the values of which are represented by error bars. For each set of samples, the lowest intensity (sample 1 or sample 5) was taken as 1; the intensity of every other sample was calculated relative to that. Columns represent the transcripts of three different genes: *PpCOL1*, *PpCOL2*, and *PpCOL3*; rows represent the different light conditions at dawn: darkness, white light, blue light, red light, and far-red light. Sample numbers correspond to the numbers in Figure 16.

The data on *PpCOL2* expression indicated that a culture that had been kept in darkness displayed a 2-fold increase in expression 90 minutes after subjective dawn (Figure 17). Also, a culture exposed to white light did not show a significantly stronger increase in *PpCOL2* expression than a dark-kept culture. Both these findings were in agreement with the previous experiment, where expression levels were compared 2 and 6 hours after actual and subjective sunrise (Figure 15). Strikingly, the increase in expression was slightly stronger in cultures irradiated with blue, red, and far-red light, as compared to the white light-exposed cultures (Figure 17).

The *PpCOL3* transcript levels did not rise in the absence of light, whereas a 2- to 3-fold increase could be observed 30 and 90 minutes after irradiation with white light (Figure 17). In cultures kept in blue and red light, a 4- to 5-fold increase was detected already 30 minutes after sunrise, whereas cultures kept in far-red light reached such an increase only about one hour later (Figure 17).

Summarising, the results for *PpCOL1* were clearest, indicating a strong transcriptional response to light of every quality. Results for *PpCOL2* and *PpCOL3* were at times hard to interpret, due to low amplitudes of the responses. However, expression levels of both genes appeared to be higher in blue, red, and far-red light than in darkness. Therefore it can be concluded that at least no support was found for a transcriptional response of these genes to light of a particular wavelength.

4.3 Conclusions

The expression patterns of *CO*-like genes of *Arabidopsis* and other flowering plant species consistently show circadian or diurnal regulation. All three *Physcomitrella CO* homologues were found to be diurnally expressed as well, with a broad peak in mRNA abundance during the day and a trough during the night. The observed diurnal rhythm appears to result mainly from a direct responsiveness to light signalling, because mRNA levels increase rapidly at dawn and remain high when cultures are irradiated with light during subjective night. However, when cultures are deprived of light at dawn, mRNA levels increase also in the absence of light, indicating an underlying regulation by the circadian clock. The acute response to light at dawn was dissected into

responses to particular wavelengths. The expression of each gene responded to white light, blue light, red light, and far-red light, suggesting responsiveness to integrated light signalling, rather than light signalling through a particular photoreceptor. Although the expression of *PpCOL1*, *PpCOL2* and *PpCOL3* showed largely similar responses to light and darkness, and to light of different qualities, small variations were observed in the speed and the amplitude of these responses. Expression of *PpCOL1*, for example, showed the strongest light response at dawn, with a 12-fold increase in mRNA levels compared to a 4- to 5-fold increase in *PpCOL2* and *PpCOL3* mRNA levels. Expression of *PpCOL2* on the other hand was found to be more responsive to an endogenous timekeeper than the other two genes, as *PpCOL2* mRNA abundance increased more rapidly in extended darkness.

5 INACTIVATION OF *PPCOL1*, *PPCOL2*, AND *PPCOL3*

5.1 Introduction

Physcomitrella is unique among plants in that the rate of homologous recombination is high enough to make gene targeting a feasible approach to study gene function (Schaefer et al., 1991; Schaefer and Zryd, 1997). The targeting efficiency in *Physcomitrella* is significantly higher than in the classical plant model, *Arabidopsis*, making *Physcomitrella* an attractive alternative plant model organism for gene functional studies. In fact, efficiencies can be compared with those observed in yeast, a finding that tempted enthusiasts to call *Physcomitrella* the "new green yeast" (Schaefer and Zryd, 1997).

The double-strand breakage (DSB) repair machinery has been implicated in the regulation of homologous recombination in plants, as well as in animals and yeast (Dudas and Chovanec, 2004; Puchta, 2005). Double-strand breakage is a type of DNA damage that can be repaired via two pathways, either by obtaining instructions from the sister or homologous chromosome, a process which is called homologous recombination (HR), or via joining of ends that do not share sequence similarity, a process which is called non-homologous end-joining (NHEJ) (Paques and Haber, 1999). The preferential use of the HR pathway over the NHEJ pathway is generally considered to result in efficient gene targeting (Reiss, 2003).

Two general strategies for the generation of *Physcomitrella* gene disruptants have been reported (Schaefer, 2001). Firstly, a gene of interest has been targeted with an insertion vector, based on the insertion of a circular molecule through a single HR event (Figure 18A) (Schaefer and Zryd, 1997; Hofmann et al., 1999; Mittmann et al., 2004). Secondly, a gene of interest has been targeted with a replacement vector, which is based on the insertion of a linear molecule through a double HR event. For the latter strategy, a distinction can be made between two different designs. If the regions of homology between the genome and the targeting construct are located within the gene of interest, a double HR event will result in a gene insertion (Figure

18B) (Imaizumi *et al.*, 2002; Mittmann *et al.*, 2004; Thelander *et al.*, 2004; Tanahashi *et al.*, 2005; Yasumura *et al.*, 2005). On the other hand, if the regions of homology are located outside of the gene of interest, a double HR event will result in gene removal (Figure 18C) (Lee et al., 2005). Targeting with a replacement vector is currently the strategy of choice, since it results in a modification without sequence duplication (Hohe *et al.*, 2004).

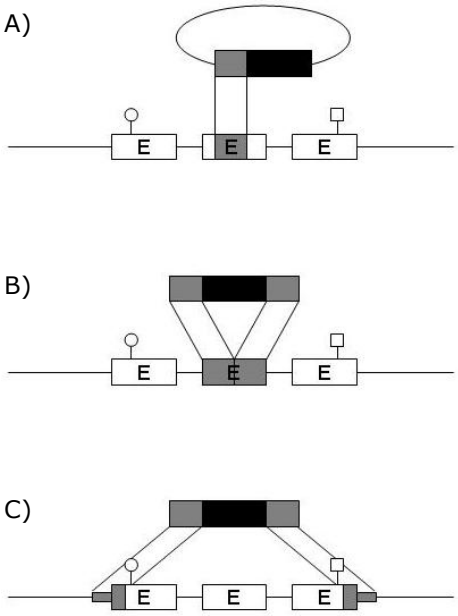


Figure 18 Gene targeting designs in *Physcomitrella*. The targeting strategy with an insertion vector (Schaefer, 2001) results in gene disruption due to a single HR event (**A**). The targeting strategy with a replacement vector (Schaefer, 2001) results in gene disruption (**B**) or gene removal (**C**) due to a double HR event; the difference between the two is the location of homologous regions. Represented is in each case the introduced targeting construct (above) and the genomic locus of the gene of interest (below). Exons of the gene of interest are shown as rectangles marked with E. The circle and square indicate putative start and stop codon, respectively. Regions of homology where HR takes place are shown in grey; an HR event is symbolised by interconnecting lines. The black rectangle indicates the recombinant selection cassette. The circle in (A) represents unlinearised vector sequence.

It has been observed that gene targeting in *Physcomitrella* might involve different types of targeting events (Kamisugi *et al.*, 2005). "One-end gene targeting" is the result of an HR event at one end of the construct accompanied by a NHEJ event at the other (HR/NHEJ). True allele replacement occurs by two HR events (HR/HR): this may involve insertion of multiple copies of the targeting construct or single-copy allele replacement. In fact, it was recently found that in *Physcomitrella* not more than 7% of a population of transformants consists of single-copy allele replacements (Yasumura *et al.*, 2005). Both "one-end gene targeting" and true allele replacement may additionally be accompanied by non-targeted insertions of the targeting construct (NHEJ/NHEJ) (Kamisugi *et al.*, 2005). It has to be

noted that in the case of gene targeting through gene removal (Figure 18C), which is the strategical gene targeting design employed in this study, one-end gene targeting events will not result in gene removal (HR/NHEJ in Figure 19). Only a HR/HR event will result in gene removal (Figure 19).

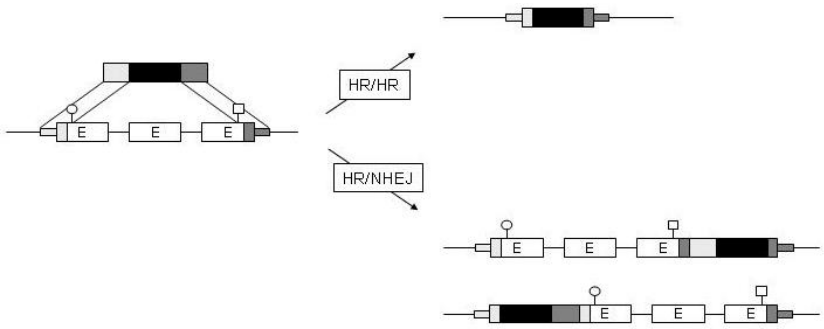


Figure 19 Different outcomes of targeting with a replacement vector (Schaefer, 2001), in the case of a gene removal design. A double HR event (HR/HR) will result in gene removal, whereas a combination of HR and NHEJ (HR/NHEJ) will not. A combination of NHEJ and NHEJ, resulting in random integration in the genome, is not depicted. Exons of the gene of interest are shown as white rectangles marked with E. The circle and square indicate putative start and stop codon, respectively. Regions of homology at the 5'- and 3'-end are shown as light grey and dark grey rectangles, respectively. The black rectangle indicates the recombinant selection cassette.

Different selection cassettes have been used for the selection of *Physcomitrella* transformants. Genes that were used include those that confer resistance to geneticin (*nptII*), hygromycin (*hph*, *aphIV*), zeocine (*zeo*) and sulphonamide (*sul*), driven by the CaMV 35S promoter or the nopaline synthase (nos) promoter (Cove, 2005). Also reporter genes commonly used in seed plants function well in *Physcomitrella*: the *uidA* (GUS) gene and the gene coding for green fluorescent protein (GFP) (Cove, 2005). The transformation procedure in *Physcomitrella* gives rise to two classes of antibiotic-resistant transformants: stable and unstable transformants. Stable transformants have the transgenic DNA inserted in the genome. Unstable transformants lose the antibiotic-resistance phenotype when selection is relaxed, suggestive of extrachromosomal replication of the transformed DNA (Ashton *et al.*, 2000). It has been reported that after two rounds of selection and relaxation more than 98% of surviving transformants are stable (Schween *et al.*, 2002). Others

routinely use four rounds of selection and relaxation to obtain stably transformed lines (Cove, 2000).

5.2 Results

5.2.1 Targeting strategy used for inactivation of *PpCOL* genes

A replacement vector was used, such that the target gene starting with the start codon would be replaced either completely (*PpCOL1* and *PpCOL2*), or almost entirely (*PpCOL3*), with a GUS reporter gene and a selectable marker. The selectable marker is expressed from the 35S plant viral promoter or the T-DNA nos promoter. Allele replacement will result in the GUS reporter gene being inserted in the original genomic context of the target gene in such a way as it should be expressed from the regulatory sequences of the target gene (Figure 21).

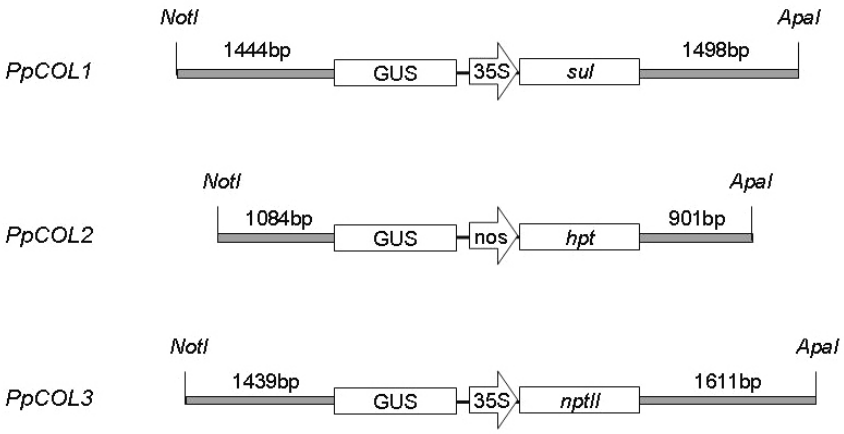


Figure 20 Targeting constructs for *PpCOL1*, *PpCOL2*, and *PpCOL3*. Grey bars represent the flanking regions of homology where crossover will take place, resulting in gene replacement. The sizes of these regions are indicated. The coding region of the GUS reporter gene and the coding regions of the selectable marker genes (*sul*, *hpt*, and *nptII*) are shown as white rectangles. Promoters (35S, *nos*) are shown as block arrows. *NotI* and *ApaI* are the restriction endonucleases that were used to excise the targeting constructs from the cloning vector to enable transformation with linear fragments.

The large degree of sequence identity between the PpCOL1, PpCOL2, and PpCOL3 proteins suggested that functional redundancy might compromise the study of gene function in single gene disruptants. Therefore attempts were made to generate double and triple disruptant lines. Previously, multiple disruptants were generated by successive single-gene targeting experiments (Imaizumi *et al.*, 2002; Thelander *et al.*, 2004; Tanahashi *et al.*, 2005; Yasumura *et al.*, 2005). However, a novel procedure of simultaneous targeting of multiple loci was attempted, as this would be significantly less time-consuming (Hohe *et al.*, 2004).

The constructs that were used to target *PpCOL1*, *PpCOL2* and *PpCOL3* are shown in Figure 20. The up- and downstream flanking sequences of the genes, where HR events between the targeting construct and the genome will take place, were obtained from the *Physcomitrella* genome by inverse PCR (Figure 8, page 49). In order to allow simultaneous selection of targeting of two or three *PpCOL* genes, three different selectable markers were used for targeting of *PpCOL1*, *PpCOL2*, and *PpCOL3*. The three markers were the sulfadiazine resistance gene (*sul*), the hygromicin phosphotransferase gene (*hpt*), and the neomycin phosphotransferase II gene (*nptII*), respectively. The flanking regions of homology that were used were as large as possible, because targeting efficiency strongly depends on homology length (Kamisugi *et al.*, 2005).

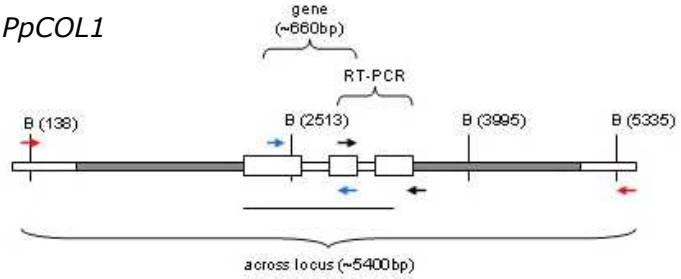
5.2.2 Generation of transformants

The transformation method of polyethylene glycol (PEG)-mediated DNA transfer to *Physcomitrella* protoplasts was used (Schaefer and Zryd, 1997), with the constructs depicted in Figure 20. Transformation experiments were carried out with every targeting construct individually, with every combination of two constructs and with all three constructs. The scale of the transformation experiment used for targeting with single constructs was increased 6 times and 18 times for double and triple targeting experiments, respectively (see Materials and Methods). For double transformations the lines were selected simultaneously with two selectable agents, and for triple transformations with three agents. After transformation, repeated rounds of selection and relaxation were applied in order to retain stably transformed

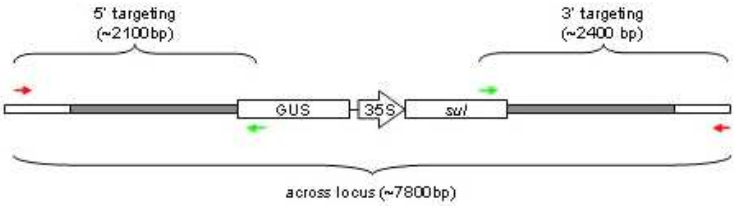
lines. Plants surviving the second round of selection were considered as putative stable transformants. Five transformation experiments resulted in a total of 125 stably transformed lines (Table 4, page 78). The single transformation experiments resulted in 4, 10, and 6 stable transformants for *PpCOL1*, *PpCOL2*, and *PpCOL3*, respectively. Double transformation experiments resulted in 21 stably transformed lines for *PpCOL2* and *PpCOL3*, 16 for *PpCOL1* and *PpCOL2*, and 28 for *PpCOL1* and *PpCOL3*. The transformation experiment with all three targeting constructs resulted in 40 stably transformed lines. DNA was isolated from each transformant as soon as enough plant material was obtained. For different transformants, this was after two, three, or four rounds of selection (Table 4, page 78), depending on the timing of the experiments and the growth of the cultures.

Figure 21 Genomic loci of *PpCOL1* (A), *PpCOL2* (B), and *PpCOL3* (C), before targeting (WT locus) and after targeting (altered locus). Grey bars represent the regions of homology that had been cloned into the respective targeting construct (Figure 20); crossover will take place in these regions. White rectangles represent *PpCOL1*, *PpCOL2* and *PpCOL3* coding regions in the wild type loci. In targeted loci, white rectangles represent the GUS reporter gene and the selectable marker gene (*sul*, *hpt*, and *nptII*); block arrows represent the promoters (35S, nos) of the selectable marker genes. The probes that were used in the Southern blot analysis of transformants are shown as a black line under the WT locus of each gene. All *BglII* restriction sites that are present in wild type loci are marked with B; their respective positions on the sequence are shown between brackets. Primers are shown as arrows. Black primers were used in the quantitative RT-PCR analysis of Figure 27. Coloured primers were used in the PCR analysis of transformed lines (Figure 24, Figure 25B, Figure 26). Primers in blue were used to verify removal of the target gene ("gene"). Primers in red were used to verify whether the WT locus was altered ("across locus"). Primers in green were used in combination with red primers as depicted, to verify homologous recombination at the 5'-end ("5'-targeting") and at the 3'-end ("3'-targeting") of the target gene. Accolades depict all the primer combinations that were used in a PCR reaction, as well as the names by which they are referred to in the text; the expected product size for each primer combination is shown between brackets.

(A) *PpCOL1*

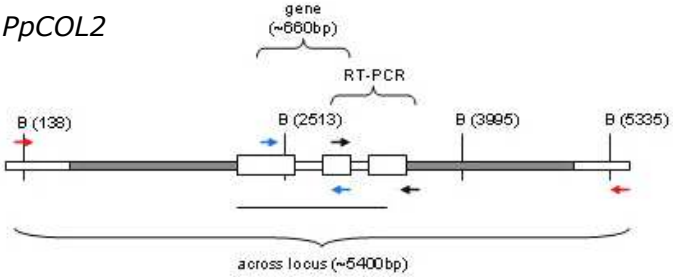


WT locus

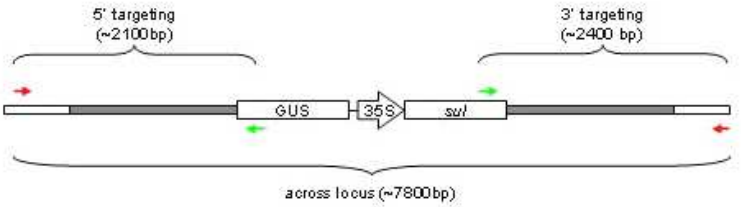


altered locus

(B) *PpCOL2*

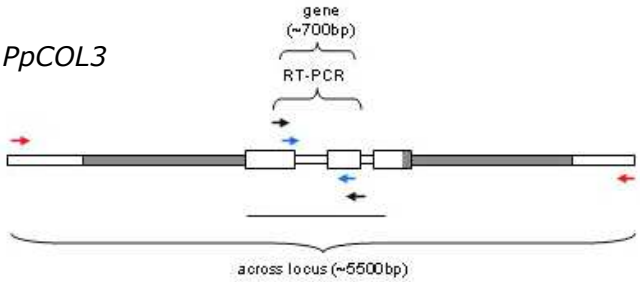


WT locus

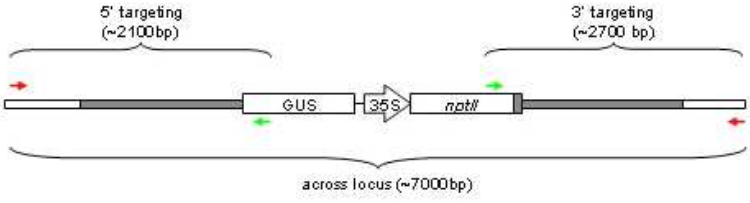


altered locus

(C) *PpCOL3*



WT locus



altered locus

5.2.3 High-throughput screen for *PpCOL* mutants

5.2.3.1 Strategy

A Southern blot-based procedure was devised as a high-throughput screen to test the structure of the targeted genes in the transformants. Probes were used that hybridise to the gene sequence that is targeted for removal (shown as a black line under WT loci in Figure 21). This approach allows identification of two different targeting events: a shifted band reflects a "one-end gene targeting" event, whereas disappearance of the band corresponds to allele replacement, either through single-copy or multi-copy insertion (Kamisugi *et al.*, 2005). By using the restriction endonuclease *BglII* for digestion of genomic DNA, wild type copies of the three genes appear as differently sized bands (~2,0 kb and ~1,5 kb for *PpCOL1*; >11,5 kb for *PpCOL2*; ~8,0 kb for *PpCOL3*). This approach enables three probes to be used together and therefore multiple targeting events can be analysed simultaneously (Figure 22, Figure 23, Figure 25A). Wild type DNA was always included as control. The Southern blots were used as a first, high-throughput screening method to eliminate transformants that certainly did not harbour disruptants. Whenever the Southern blot data could not be unambiguously interpreted, additional analysis was carried out by PCR, by using gene-specific primers to verify removal of the target gene from the genome (blue arrows in Figure 21). This PCR is hereafter referred to as *"gene"* PCR.

Table 4 Summary of results of high-throughput screen for *PpCOL* mutants from single, double, and triple targeting experiments. The transformation experiment, the number of rounds of selection and relaxation that the line had gone through at the time of molecular analysis, and the type of molecular analysis that was performed (Southern blot or "gene" PCR) are shown. If a field is grey, the gene was not targeted in that experiment; if a field is blank, the analysis was not performed. Observations are given for individual loci (*PpCOL1*, *PpCOL2*, and *PpCOL3*). + indicates unsuccessful targeting (Southern) or unsuccessful allele replacement ("gene" PCR); ? indicates ambiguity in interpretation; - indicates successful targeting and/or allele replacement (Southern) or successful allele replacement ("gene" PCR).

transformant	transf. expt.	# sel. rounds	Southern blot			"gene" PCR		
			PpCOL1	*PpCOL2*	*PpCOL3*	*PpCOL1*	*PpCOL2*	*PpCOL3*
COL1-2	IV	3	-					
COL1-6	IV	3	+					
COL1-7	IV	3	-					
COL1-8	IV	2	+					
COL2-1	IV	3		-				
COL2-2	IV	3		-				
COL2-3	IV	3		-				
COL2-4	IV	3		-				
COL2-5	IV	3		-				
COL2-6	IV	3		-				
COL2-8	IV	3						
COL2-9	IV	3		+				
COL2-10	IV	2		-				
COL2-29	IV	2		-				
COL3-11	IV	3			+			
COL3-23	IV	3			+			
COL3-83	IV	3			+			
COL3-98	IV	3			+			
COL3-111	IV	3			-			
COL3-112	IV	3			-			
COL2/3-1	I	4		+	+			
COL2/3-2	I	4		+	-			
COL2/3-3	I	4		-	+			
COL2/3-4	I	4		?	-		+	
COL2/3-5	I	4		?	-		+	
COL2/3-6	III	3		+	+			
COL2/3-7	III	3		+	-			
COL2/3-8	III	3		+	-			
COL2/3-11	III	3		?	+			
COL2/3-13	III	3		+	+			
COL2/3-14	III	3		?	+			
COL2/3-18	V	2		?	+			
COL2/3-19	V	2		?	-		+	
COL2/3-20	V	2		?	+			
COL2/3-21	V	2		?	+			
COL2/3-22	V	2		?	-		+	
COL2/3-23	V	2		?	+			
COL2/3-24	V	2		?	+			
COL2/3-27	V	2		?	+			
COL2/3-28	V	2		+	+			
COL2/3-30	V	2		?	+			
COL1/2-1	I	4	?	+				
COL1/2-2	III	3	-	+				
COL1/2-3	III	3	+	-				
COL1/2-4	III	3	+	+				
COL1/2-6	V	2	+	-				
COL1/2-7	V	2	+	+				
COL1/2-8	V	2	+	+				
COL1/2-9	V	2	+	-				
COL1/2-10	V	2	+	?				
COL1/2-12	V	2	+	-				
COL1/2-14	V	2	?	+				
COL1/2-15	V	2	-	+				
COL1/2-16	V	2	+	-				
COL1/2-17	V	2	+	+				
COL1/2-18	V	2	?	+				
COL1/2-19	V	2	+	?				

transformant	transf. expt.	# sel. rounds	Southern blot			"gene" PCR		
			PpCOL1	*PpCOL2*	*PpCOL3*	*PpCOL1*	*PpCOL2*	*PpCOL3*
COL1/3-1	I	4	?		-	+		
COL1/3-2	I	4	?		-	+		
COL1/3-3	III	3	?		+			
COL1/3-4	III	3	?		-	+		
COL1/3-5	III	3	?		-	+		
COL1/3-6	III	3	?		-	+		
COL1/3-7	III	3	?		-	+		
COL1/3-8	III	3	?		+			
COL1/3-9	III	3	?		-	+		
COL1/3-10	III	3	?		+	+		
COL1/3-13	III	3	?		+	+		
COL1/3-14	III	3				+		+
COL1/3-15	V	2				+		+
COL1/3-17	V	2				+		+
COL1/3-18	V	2				+		+
COL1/3-19	V	2				+		+
COL1/3-20	V	2				+		+
COL1/3-21	V	2				+		+
COL1/3-22	V	2				-		+
COL1/3-23	V	2				+		-
COL1/3-24	V	2				+		-
COL1/3-25	V	2				+		+
COL1/3-25	V	2				+		+
COL1/3-26	V	2				+		+
COL1/3-27	V	2				+		+
COL1/3-28	V	2				+		+
COL1/3-29	V	2				+		+
COL1/3-30	V	2				+		+
COL1/2/3-1	II	4	+	+	-			
COL1/2/3-2	II	4	+	+	+			
COL1/2/3-5	II	4	+	-	+			
COL1/2/3-6	II	4	+	+	+			
COL1/2/3-17	II	4	-	+	+			
COL1/2/3-18	II	4	+	+	+			
COL1/2/3-22	II	4	+	?	-			
COL1/2/3-26	II	4	-	+	+			
COL1/2/3-27	II	4	-	+	+			
COL1/2/3-30	II	4	+	+	+			
COL1/2/3-32	II	4	+	+	+			
COL1/2/3-33	II	4	?	?	-		+	
COL1/2/3-40	II	4	-	?	+			
COL1/2/3-41	II	4	+	?	+			
COL1/2/3-46	II	4	+	+	+			
COL1/2/3-50	II	4	?	?	-		+	
COL1/2/3-51	II	4	+	-	+			
COL1/2/3-52	II	4	+	+	+			
COL1/2/3-54	II	4	?	+	+			
COL1/2/3-64	II	4	?	+	+			
COL1/2/3-65	II	4	?	+	+			
COL1/2/3-69	II	4	?	?	-		+	
COL1/2/3-70	II	4	?	+	-			
COL1/2/3-71	II	4	-	?	+			
COL1/2/3-72	II	4	?	+	+			
COL1/2/3-76	II	4	+	+	+			
COL1/2/3-83	II	4	+	+	+			
COL1/2/3-89	II	4	?	?	+			
COL1/2/3-92	II	4	?	+	+			

transformant	transf. expt.	# sel. rounds	Southern blot			"gene" PCR		
			PpCOL1	*PpCOL2*	*PpCOL3*	*PpCOL1*	*PpCOL2*	*PpCOL3*
COL1/2/3-96	II	4	+	+	+			
COL1/2/3-97	II	4	?	+	+			
COL1/2/3-112	II	4	?	-	+			
COL1/2/3-117	II	4	?	?	+			
COL1/2/3-121	II	4	?	?	+			
COL1/2/3-122	II	4	+	+	+			
COL1/2/3-125	II	4	?	+	+			
COL1/2/3-127	II	4	+	+	+			
COL1/2/3-128	II	4	+	+	+			
COL1/2/3-135	II	4	+	-	+			
COL1/2/3-136	II	4	+	-	+			

5.2.3.2 Analysis

One-hundred-and-seven stable transformants were analysed by Southern blotting, followed by PCR analysis in case of ambiguity. The remaining 18 transformants were analysed by PCR only. The findings of these analyses are summarised in Table 4.

All 40 stable triple transformants were analysed by Southern blotting (Figure 22). This analysis indicated that 13 transformants had not undergone recombination at any of the three loci, reflecting random integration of all three targeting constructs through NHEJ/NHEJ events (Table 4). Sixteen transformants were found not to be altered at two loci, with the third locus either altered or not clearly interpretable (Table 4). Another 8 transformants were not altered at one locus, with the other two loci either altered or not clearly interpretable (Table 4). The remaining 3 lines possessed one altered locus, with remaining ambiguity concerning the targeting of the other two loci (Table 4). Therefore they were additionally analysed by PCR for gene removal at one of the latter two loci. This revealed that in each of these lines, at least the *PpCOL2* gene had not been removed by targeted gene replacement (Figure 24). In conclusion, the Southern analysis indicated that at least 16 lines possibly represent single gene disruptants, whereas no evidence was found for the presence of transformants that had 2 or 3 target genes altered.

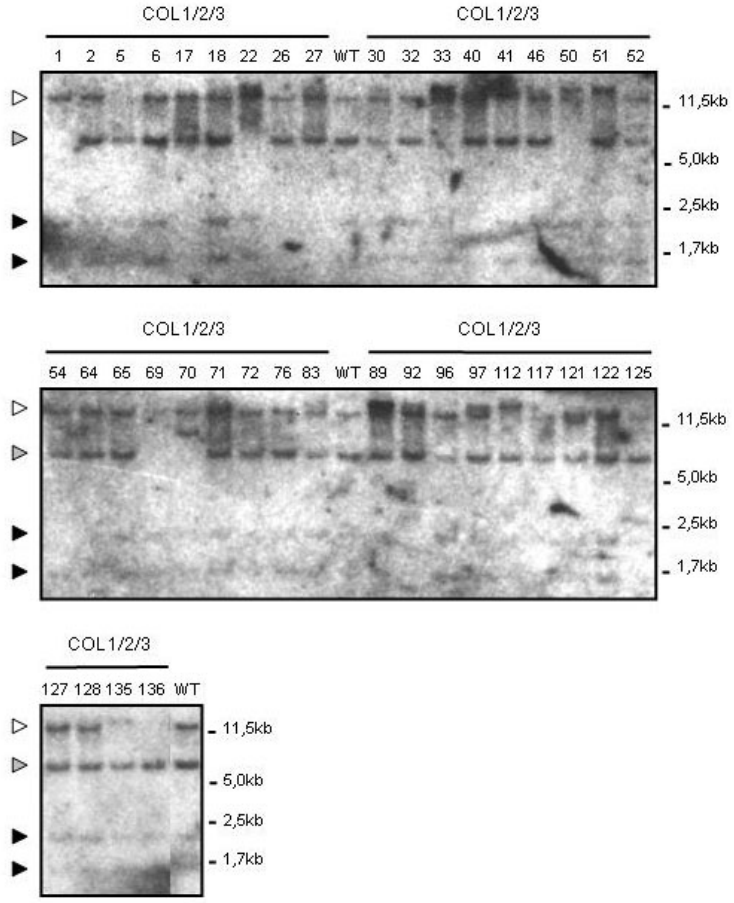


Figure 22 Southern blot analysis of stable triple transformants. Genomic DNA was digested with *BglII*. The blots were hybridised, under stringent conditions, with *PpCOL1*, *PpCOL2*, and *PpCOL3* probes (shown in Figure 21) simultaneously. The caption above each lane corresponds to a stably transformed line listed in Table 4; WT refers to the untransformed wild type strain. Positions that correspond to unaltered *PpCOL1*, *PpCOL2*, and *PpCOL3* loci are indicated by black, white, or grey triangles, respectively. The sizes of the DNA size marker (λx*PstI*) are given on the right.

Forty-eight out of 65 stable double transformants were analysed by Southern blotting (Figure 23). Among these, 8 lines were found not to be altered at any of the two loci (Table 4). Twenty-nine lines were not altered at one locus, with the other locus either altered or not clearly interpretable (Table 4). The remaining 11 lines included lines that possessed one altered locus with remaining ambiguity concerning the targeting of the other locus, as well as lines that had not been analysed by Southern blotting (Table 4). These were analysed for target gene removal by "gene" PCR. This indicated that in each of these lines, at least one of the targeted genes had not been removed

by targeted gene replacement (Figure 24). In conclusion, no double transformant was found that had both targeted loci altered, whereas as many as 25 lines were identified that possibly represent single gene disruptants.

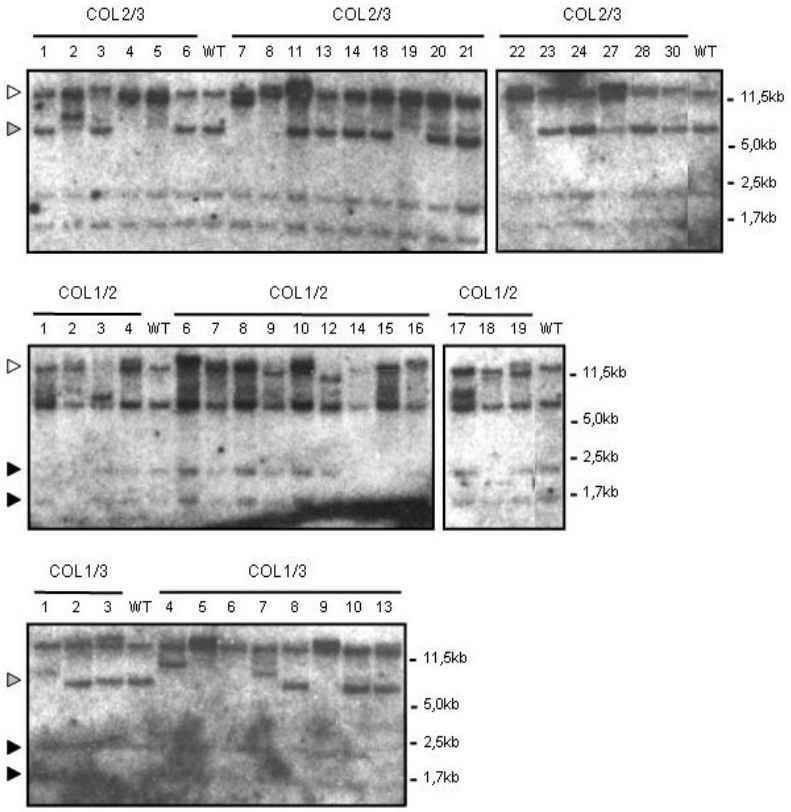


Figure 23 Southern blot analysis of stable double transformants. Genomic DNA was digested with *BglII*. The blots were hybridised, under stringent conditions, with *PpCOL1*, *PpCOL2*, and *PpCOL3* probes (shown in Figure 21) simultaneously. The caption above each lane corresponds to a stably transformed line listed in Table 4; WT refers to the untransformed, wild type strain. Positions that correspond to unaltered *PpCOL1*, *PpCOL2*, and *PpCOL3* loci are indicated by black, white, or grey triangles, respectively. The sizes of the DNA size marker (λx*PstI*) are given on the right.

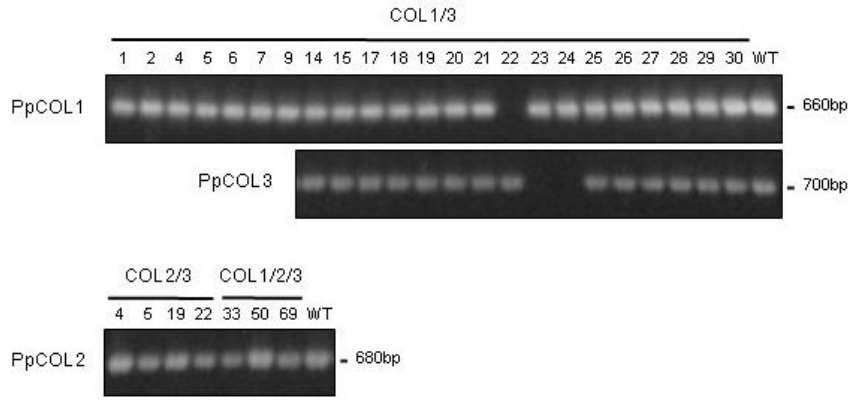


Figure 24 PCR analysis of stably transformed lines from double and triple transformation experiments. A positive PCR signal reflects the presence of the target gene. The primer combinations that were used are depicted as blue arrows in Figure 21 (dubbed "gene"). The gene that was analysed is given on the left of each picture. Approximate product sizes are given on the right of each picture. The caption above each lane corresponds to a stably transformed line listed in Table 4; WT refers to the untransformed, wild type strain. Product sizes are given on the right of each picture.

All 20 single transformants were analysed by Southern blotting as well (Figure 25A). Among the 4 *PpCOL1* transformants, 2 lines were found to possess an altered *PpCOL1* locus (lines COL1-2 and COL1-7), because they lacked the 1,5 kb and 2 kb fragments characteristic of *PpCOL1*. Among 9 *PpCOL2* transformants, 8 were found to possess an altered *PpCOL2* locus (lines COL2-1 to -6, COL2-10 and COL2-29), as they lacked the >11,5 kb fragment characteristic of *PpCOL2*, whereas one contained an unaltered *PpCOL2* locus (line COL2-9). Finally, 2 out of 6 *PpCOL3* transformants possessed an altered *PpCOL3* locus (lines COL3-111 and COL3-112), because they lacked the 8 kb fragment characteristic of *PpCOL3*.

5.2.4 Identification of single disruptants

As it was clear from the Southern blot analysis that multiple gene disruptants were not present among the multiple transformants, a more thorough PCR-based analysis of gene targeting events was focused on the identification of single gene disruptants. Therefore, this analysis was carried out on all single transformants, and on a few selected multiple transformants

that were identified as likely single gene disruptants in the Southern blot analysis (COL1/2-2, COL1/3-6, COL1/2/3-5 and COL1/2/3-135; Table 4).

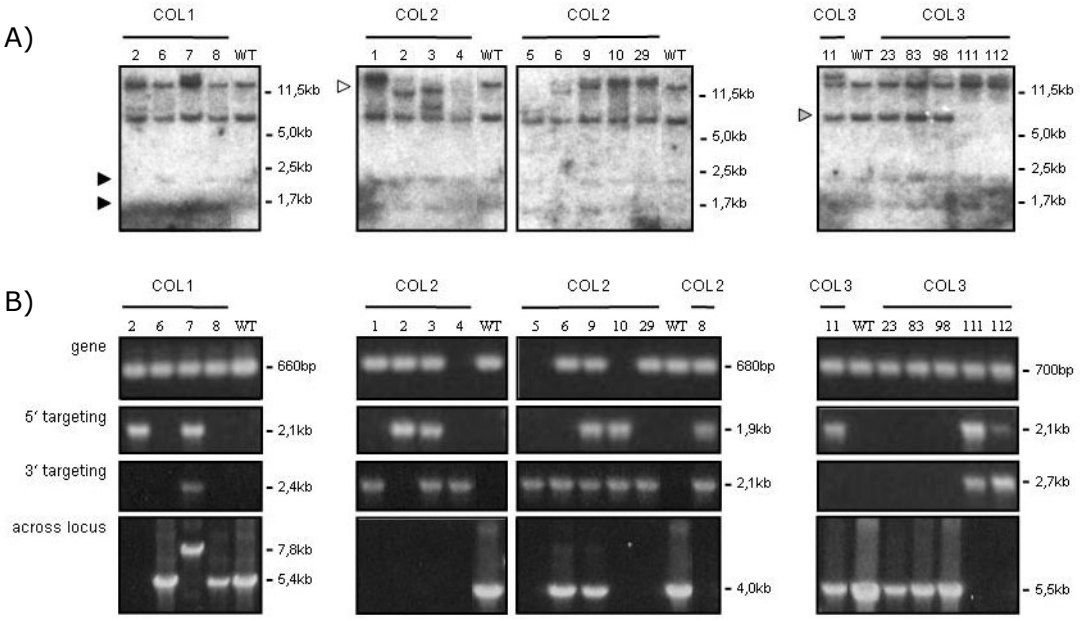


Figure 25 Analysis of stable single transformants by Southern blot and by PCR. (**A**) Southern blot analysis. Genomic DNA of each transformed line was digested with *BglII*. Each blot was hybridised, under stringent conditions, with *PpCOL1*, *PpCOL2*, and *PpCOL3* probes (shown in Figure 21) simultaneously. The caption above each lane corresponds to a stably transformed line listed in Table 4; WT refers to the untransformed, wild type strain. In each case, the position of the wild type band that corresponds to a targeted gene is indicated by a triangle. Positions that correspond to unaltered *PpCOL1*, *PpCOL2*, and *PpCOL3* loci are indicated by black, white, or grey triangles, respectively. The sizes of the DNA size marker (λx*PstI*) are given on the right. (**B**) PCR analysis. The name on the left of each row of pictures refers to the primer combination (Figure 21) that was used in the corresponding PCR reactions. In each case, primers were used that are specific for the targeted gene of that line. The caption above each lane corresponds to a transformed line given in Table 4; WT refers to the untransformed, wild type strain. Product sizes are given on the right of each picture.

5.2.4.1 Strategy

A thorough PCR-based analysis was designed that used different combinations of primers, adequately positioned on the targeting construct and on the target locus, allowing the probing of distinct aspects of a gene targeting event. Figure 21 (page 76) illustrates the *PpCOL1*, *PpCOL2* and

PpCOL3 loci, before and after a single-copy allele replacement. The positions of the primers employed in the analysis are indicated, as well as their resulting PCR products. Gene-specific primers are used to verify gene removal (dubbed "*gene*" PCR; blue arrows in Figure 21). Another set of primers, annealing outside of the regions of homology where HR will take place, is used to reveal an HR event at the target locus (dubbed "*across locus*" PCR; red arrows in Figure 21). If the targeting construct integrated randomly through a NHEJ/NHEJ event, the DNA of the transformant will support amplification of a product that is identical in size to the product obtained with wild type DNA. If the locus was altered, the PCR will support amplification of a larger product of a particular size in the case of an HR/HR event with single-copy allele replacement (Figure 21), or no product at all in the case of an HR/HR event with multiple-copy insertions, because the locus then likely got too large to support PCR amplification. Possibly, a one-end gene targeting event (NHEJ/HR) might also result in a locus that is too large to support PCR amplification, even by long-template PCR. Finally, two other sets of primers separately reveal correct HR events on each side of the gene. Primers that anneal to either side of the marker cassette (the GUS gene or the selectable marker gene; green arrows in Figure 21) are combined with primers that anneal outside of the region of homology (the same were used in the "across locus" PCR; red arrows in Figure 21). These PCRs are hereafter referred to as *"5′ targeting"* PCR and *"3′ targeting"* PCR (Figure 21). Finally, ultimate and necessary proof of a gene disruptant was the absence of the target gene transcript, as analysed by quantitative RT-PCR (primers are shown as black arrows in Figure 21).

5.2.4.2 Analysis

5.2.4.2.1 PpCOL1 disruptants

The Southern blot data suggested that the *PpCOL1* locus might have been altered and removed in lines COL1-2 and COL1-7 (Figure 25A). However, both lines supported amplification of the *PpCOL1* gene ("gene"; Figure 25B), and of the *PpCOL1* transcript (Figure 27), indicating that the

gene was present and expressed. In line COL1-2, an HR event had occurred at the 5'-end ("5'-targeting"; Figure 25B), resulting in the failed support for PCR amplification of the target locus ("across locus"; Figure 25B). Together with the finding that the 3'-end had not been altered ("3'-targeting"; Figure 25B), it fits with a one-end gene targeting event. As shown in Figure 19 on page 73, such an event does not result in gene removal. In line COL1-7, HR events had occurred at the 5'-end, as well as the 3'-end of the *PpCOL1* gene ("5'-targeting" and "3'-targeting"; Figure 25B), resulting in single-copy allele replacement ("across locus", Figure 25B). The phenomena of break-induced recombination and gene conversion (Haber, 1999) could explain the fact that the *PpCOL1* gene is nevertheless still present and expressed. However, further investigations of this matter were not considered, as they go beyond the goal of the analysis, which was to identify gene disruptants. Therefore, our focus shifted to the analysis of double transformants for the identification of an unambiguous *PpCOL1* disruptant, which could then be used for the analysis of *PpCOL1* gene function. Line COL1/2-2 was found to exhibit all the PCR-based diagnostics of a *PpCOL1* disruptant line. The *PpCOL1* gene had been removed ("gene"; Figure 26) through a double HR event ("5'-targeting" and "3'-targeting"; Figure 26) with multi-copy allele replacement ("across locus"; Figure 26), resulting in absence of expression of the mRNA of the gene (Figure 27). This line was used for the phenotypical analysis of *PpCOL1* gene function, which is the subject of section 5.2.5.

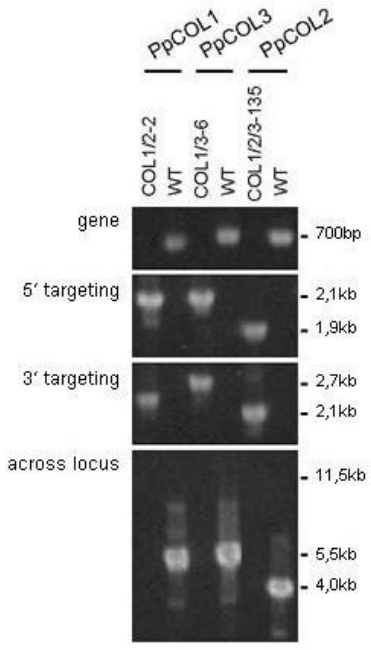


Figure 26 PCR analysis of a selection of stable double and triple transformants. The name on the left of each picture refers to the primer combination (Figure 21) that was used in the corresponding PCR reactions. The top captions refer to the gene analysed in the corresponding lanes (*PpCOL1*, *PpCOL2*, and *PpCOL3*). The caption directly above each lane corresponds to a transformed line given in Table 4; WT refers to the untransformed, wild type strain. Product sizes are given on the right of each picture.

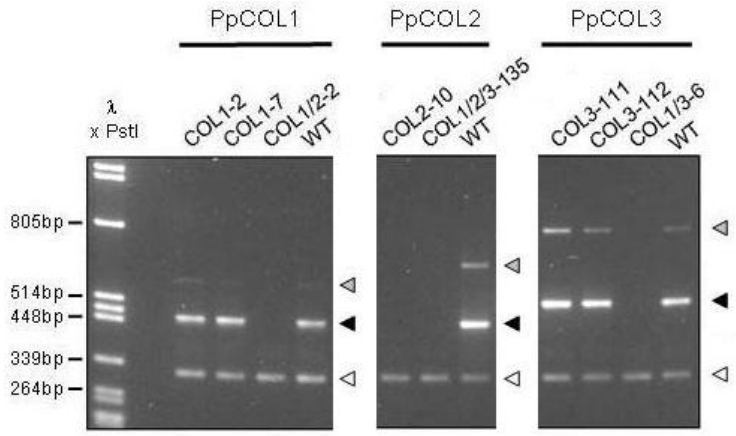


Figure 27 Quantitative RT-PCR analysis of targeted genes in protonema of stably transformed lines. The top captions refer to the gene analysed in the corresponding lanes (*PpCOL1*, *PpCOL2*, and *PpCOL3*). The caption directly above each lane corresponds to a transformed line given in Table 4; WT refers to the untransformed, wild type strain. The primers that were used are shown as black arrows in Figure 21. The positions of the 18S rRNA product, the gene-specific product, and putatively unspliced gene-specific product are marked by white, black, and grey triangles, respectively. The sizes of the DNA size marker (λx*PstI*) are given on the left.

5.2.4.2.2 PpCOL2 disruptants

The Southern blot analysis of 9 *PpCOL2* single transformants had indicated that as many as 8 of these might have been altered at the *PpCOL2* locus (Figure 25A). However, PCR analysis of all *PpCOL2* transformants revealed that only 3 lines failed to support amplification of the *PpCOL2* gene: lines COL2-4, COL2-5 and COL2-10 ("gene"; Figure 25B). In line COL2-10, HR events had occurred at both ends of the *PpCOL2* gene ("5'-targeting" and "3'-targeting"; Figure 25B), resulting in multi-copy allele replacement ("across locus"; Figure 25B). Lines COL2-4 and COL2-5 appeared to have undergone an HR event at the 3'-end of the *PpCOL2* gene ("3'-targeting"; Figure 25B), but not at the 5'-end ("5'-targeting"; Figure 25B). However, this does not reflect a one-end gene targeting event, as this would not have resulted in the observed removal of the *PpCOL2* gene (Figure 19, page 73). It seems that instead of an HR event, a larger deletion occurred at the 5'-end that includes at least part of the *PpCOL2* gene, resulting in a failed "gene" PCR, as well as the region upstream of the *PpCOL2* gene that includes the binding site of the forward primer of the "5'-targeting" PCR (red arrow in Figure 21, page 76). In fact, similar growth aberrations were observed for lines COL2-4 and COL2-5.

Both lines formed colonies that were significantly reduced in size because cells are unable to expand as in the wild type (Figure 28). However, this phenotype was not observed in two other lines, COL2-10 and COL1/2/3-135, both of which were shown to have lost the *PpCOL2* gene and transcript (Figure 26 and Figure 27). Therefore, the growth phenotype of lines COL2-4 and COL2-5 was not connected to the lack of *PpCOL2* gene function, but probably the result of a larger chromosomal deletion at the 5'-end of the target gene. It has previously been observed that transformation-induced growth aberrations are common occurrences in *Physcomitrella* transformation experiments, possibly due to such deletion events (U. Markmann-Mulisch and B. Reiss, personal communication). The PCR analysis further revealed that the remaining lines from the *PpCOL2* targeting experiment still possess the *PpCOL2* gene, possibly due to one-end gene targeting events (*e.g.* lines COL2-1, COL2-2, COL2-29; Figure 25B), break-induced recombination or gene conversion (*e.g.* lines COL2-3, COL2-6, COL2-8; Figure 25B), or random integration events (e.g. line COL2-9; Figure 25B). Again, further investigation of these lines was beyond the goal of this analysis, and only line COL2-10 was retained for phenotypical analysis of *PpCOL2* gene function, which is the subject of section 5.2.5.

5.2.4.2.3 PpCOL3 disruptants

The Southern blot data indicated that lines COL3-111 and COL3-112 might represent *PpCOL3* disruptants (Figure 25A). However, both lines supported amplification of the *PpCOL3* gene ("gene"; Figure 25B), and of the *PpCOL3* transcript (Figure 27), indicating that the gene was present and expressed. As in line COL1-7, HR events had occurred at both ends of the target gene ("5'-targeting" and "3'-targeting"; Figure 25B), this time resulting in multi-copy allele replacement ("across locus", Figure 25B). The phenomena of break-induced recombination and gene conversion (Haber, 1999) could explain the fact that the *PpCOL3* gene is still present and expressed. Again, further investigations of this matter were not considered. Instead, transformants from the double and triple targeting experiments were screened for PpCOL3 disruptants. Line COL1/3-6 was found to exhibit all the PCR-based diagnostics of a *PpCOL3* disruptant line. The *PpCOL3* gene had been removed ("gene"; Figure 26) through a double HR event ("5'-targeting"

and "3'-targeting"; Figure 26) with multi-copy allele replacement ("across locus"; Figure 26), resulting in absence of expression of the mRNA of the gene (Figure 27). This line was used for the phenotypical analysis of *PpCOL3* gene function, as described in the next section.

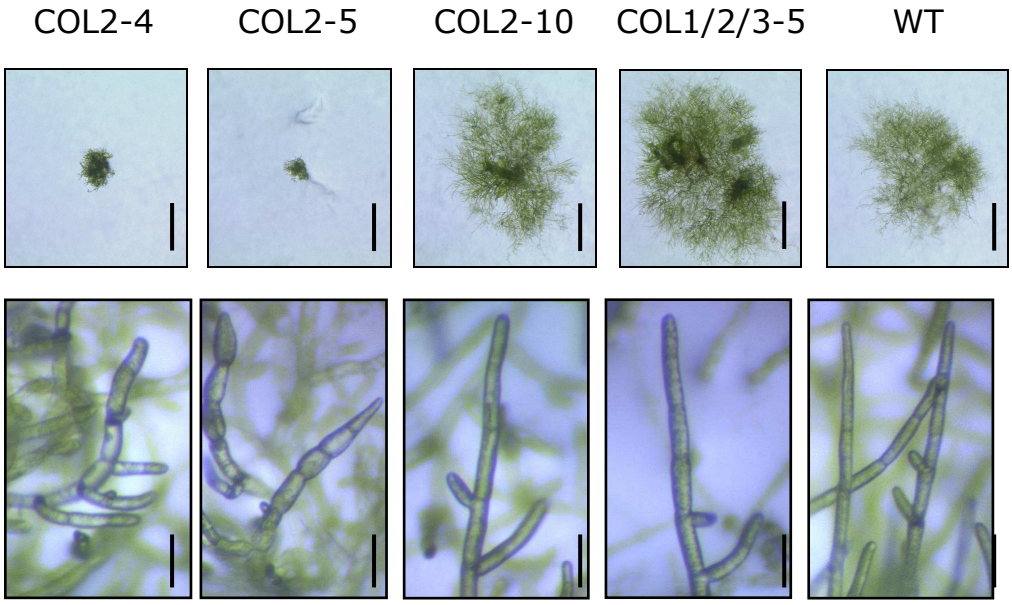


Figure 28 Growth of four potential *PpCOL2* disruptant lines, compared to the wild type. Shown are colonies (top row) and chloronema cells (bottom row) from cultures that had been growing in a long-day regime (16L:8D) for 19 days. Scale bars: 2mm (top row); 125nm (bottom row).

5.2.5 Phenotypical analysis of disruptants

The single disruptant lines COL1/2-2 (*PpCOL1* disruptant), COL2-10 (*PpCOL2* disruptant), and COL1/3-6 (*PpCOL3* disruptant) were used in a phenotypical analysis. Cultures were compared with the wild type in order to look for general growth aberrations. Furthermore, given the importance of light signalling in the transcriptional regulation of *PpCOL1*, *PpCOL2*, and *PpCOL3*, it was decided to analyse the disruptant lines for an easily tractable light response, namely phototropism of caulonema and chloronema cells. Both cell types show distinct phototropic responses (Cove *et al.*, 1978; Cove, 1992). Caulonemal filaments spread from the central part of the culture,

growing fast, and at right angle to, or at a small angle away from the direction of light. Almost all caulonemal sub-apical cells of such a filament divide and give rise to side branch initials. The majority of these develop into chloronemal filaments which show a positive phototropic response.

Cultures were initiated from a common starting point, the regenerating protoplast stage. When a protoplast starts dividing, it grows into a protonemal filament, much like a germinating spore. Growth of the cultures was observed during four weeks' growth on standard medium, in long days (16L:8D). Each disruptant line was found to progress through gametophyte development much like the wild type did, from regenerating protoplasts, to protonemal filaments, and finally to the stage of gametophores (Figure 29A). The morphology of mature gametophores appeared normal, with normally developing leaflets and rhizoids (Figure 29B). The *PpCOL1*, *PpCOL2*, and *PpCOL3* disruptant lines were concluded not to be affected in development of the gametophyte, which represents the dominant generation in mosses. After 19 days' growth, both protonemal cell types were present and displayed their characteristic phototropic responses, in the wild type as well as in disruptant lines. Caulonemal filaments spread from the central part of the culture, growing slightly away from the light, whereas chloronemal side-branches are formed only on one side, as they grow towards the light (Figure 30, Figure 31). Therefore, the phototropic light response of protonemal filaments appeared unaffected by the lack of *PpCOL1*, *PpCOL2*, or *PpCOL3* function.

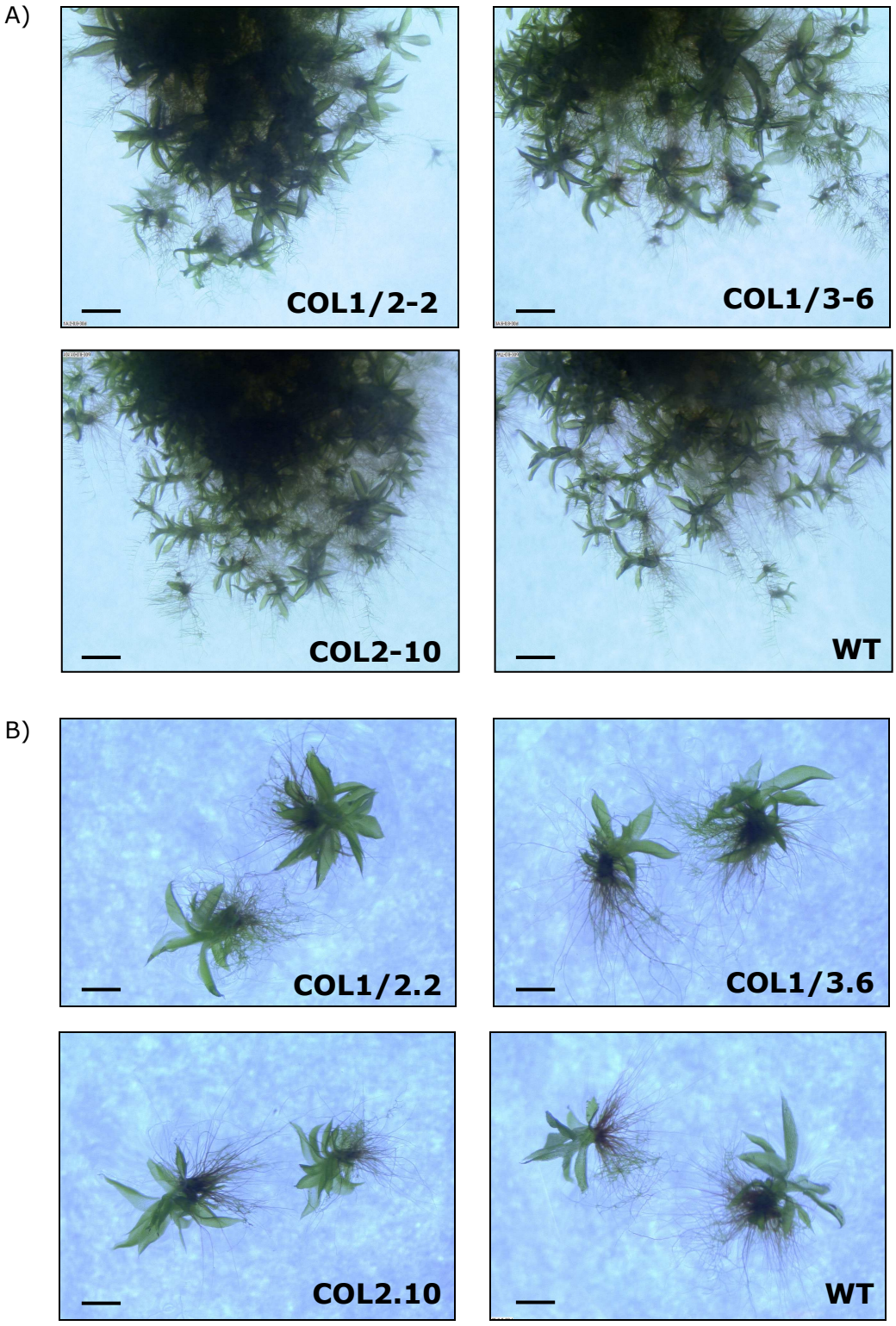


Figure 29 Broad culture morphology (**A**) and isolated gametophores (**B**) of disruptant lines of *PpCOL1* (COL1/2-2), *PpCOL2* (COL2-10), and *PpCOL3* (COL1/3-6), and of the wild type after 30 days' growth on standard medium in long days (16L:8D). Scale bars: 2mm in (A), 1mm in (B).

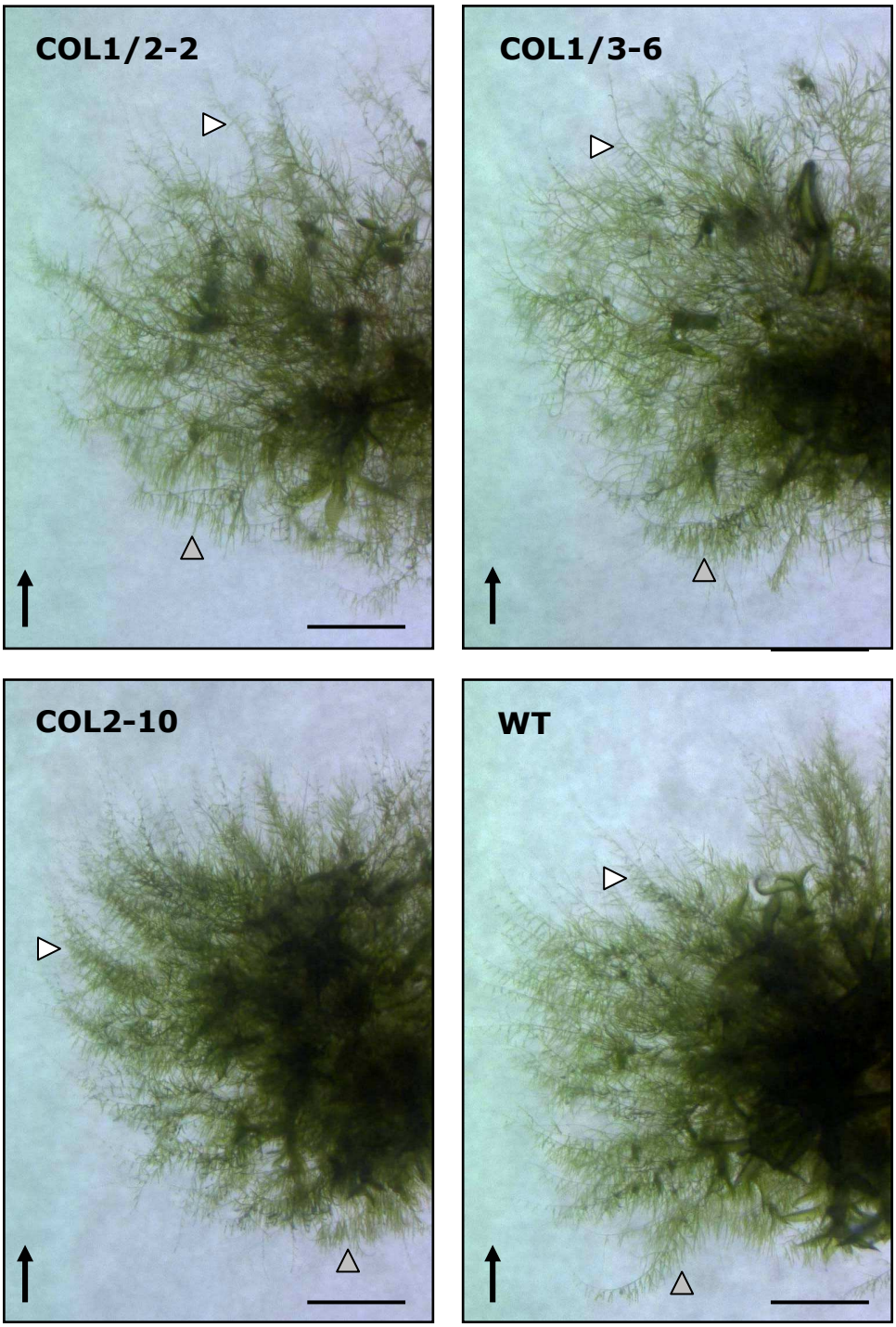


Figure 30 Phototropic responses of disruptant lines of *PpCOL1* (COL1/2-2), *PpCOL2* (COL2-10), *PpCOL3* (COL1/3-6), and of the wild type. Black arrows indicate the direction of light. White arrowheads indicate the slightly negative phototropism of caulonemal filaments; grey arowheads indicate the positive phototropism of chloronemal filaments. Shown are colonies from cultures that had been growing in a long-day regime (16L:8D) for 19 days. Scale bars: 2mm

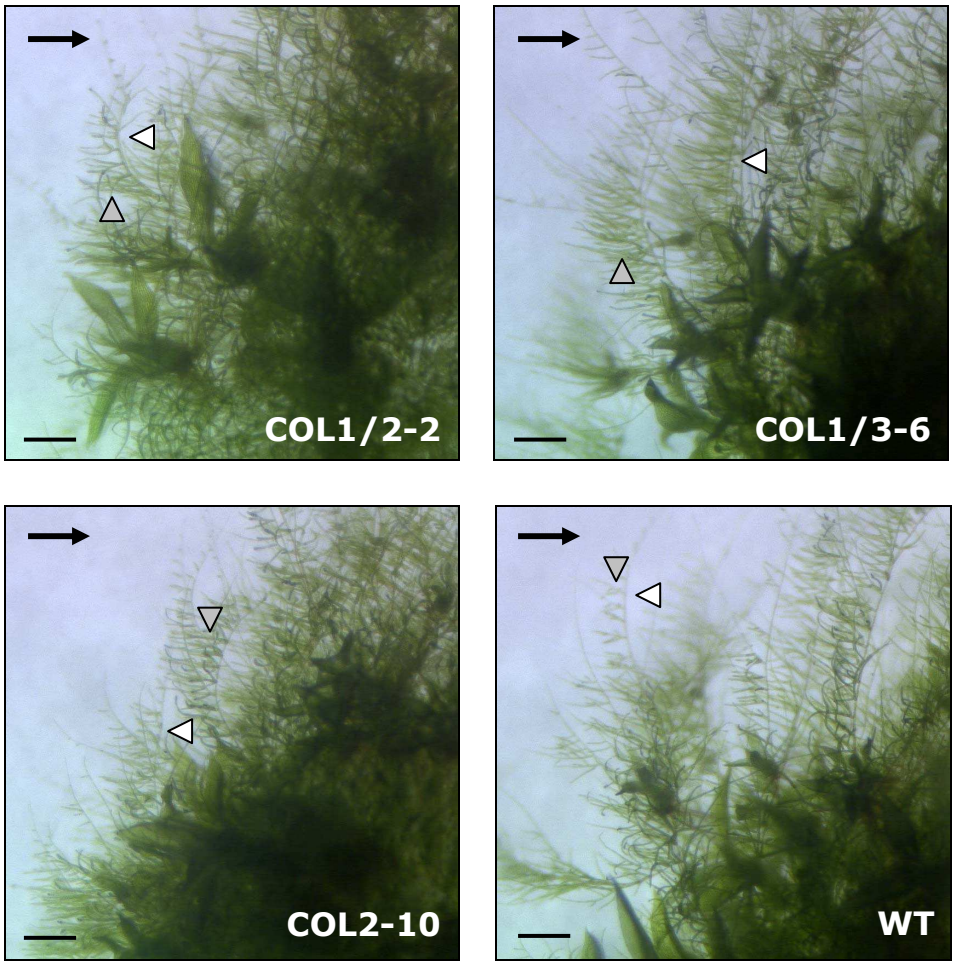


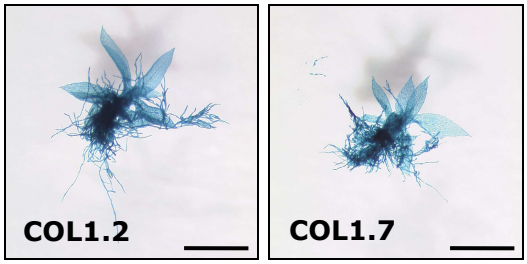
Figure 31 Phototropic responses of disruptant lines of *PpCOL1* (COL1/2-2), *PpCOL2* (COL2-10), *PpCOL3* (COL1/3-6), and of the wild type. Black arrows indicate the direction of light. White arrowheads indicate the slightly negative phototropism of caulonemal filaments; grey arowheads indicate the positive phototropism of chloronemal filaments. Shown are colonies from cultures that had been growing in a long-day regime (16L:8D) for 19 days. Scale bars: 0,5mm

5.2.6 Analysis of spatial expression patterns of *PpCOL* genes

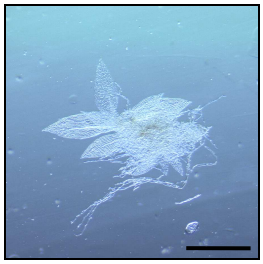
The generation of disruptants was designed in such a way that the entire coding sequence of *PpCOL1* and *PpCOL2*, and almost the entire coding sequence of *PpCOL3*, would be replaced by the GUS reporter gene (Figure 21, page 76). In the PCR analysis of single disruptants, several lines had been identified that supported amplification in the "5'-targeting" PCR, indicating that the GUS gene had been inserted at the target locus and may therefore be assumed to be expressed in a similar way to the target gene (Figure 21, page 76). These lines included two *PpCOL1* transformants (COL1-2 and COL1-7), five *PpCOL2* transformants (COL2-2, COL2-3, COL2-8, COL2-9 and COL2-10), and three *PpCOL3* transformants (COL3-11, COL3-111 and COL3-112) (Figure 25B, page 85). These lines were used for histochemical detection of GUS activity. Protonemata and gametophores were harvested during the day from 19 day old cultures grown in long days with 16 hrs of light and 8 hrs of darkness. The GUS expression patterns were the same in all lines, with a strong GUS signal found throughout protonema filaments and gametophores (Figure 32). The untransformed wild type was devoid of any GUS activity.

As discussed earlier, most lines probably contain multiple copies of the GUS gene. Nevertheless, it seems likely that the GUS expression pattern does reflect the true pattern of expression of the wild type genes. Firstly, each transformation construct contained between 1.000 and 1.500 bp of promoter sequence in front of the GUS gene, therefore ectopically integrated copies likely included these promoting elements as well. Secondly, the same ubiquitous expression pattern was observed for all twenty single transformants (data not shown), whereas if the expression was due to genomic flanking sequences derived from other genes then the expression of GUS should have varied between lines. Finally, the RT-PCR analysis confirmed the expression of *PpCOL1*, *PpCOL2,* and *PpCOL3* in protonemata, where GUS expression was detected, and another study confirmed the expression of *PpCOL1* in gametophores (Shimizu *et al.*, 2004).

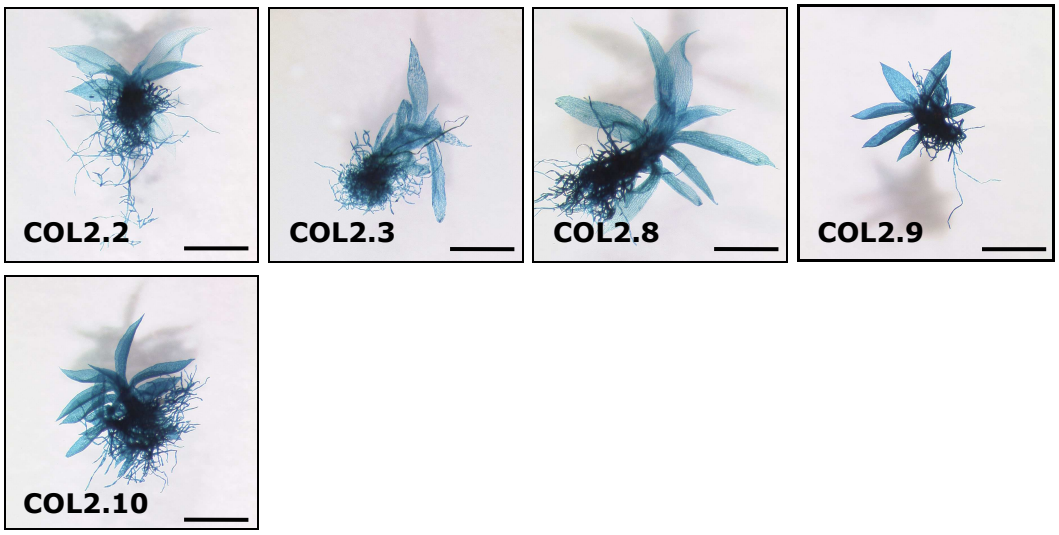
A) *PpCOL1*



D) Wild type



B) *PpCOL2*



C) *PpCOL3*

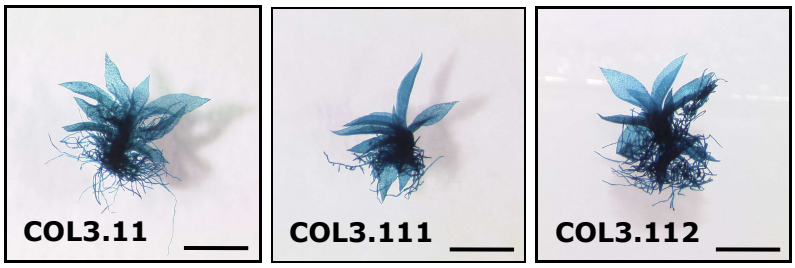


Figure 32 Histochemical detection of GUS activity in selected *PpCOL1* transformants (**A**), *PpCOL2* transformants (**B**), *PpCOL3* transformants (**C**), and the untransformed wild type (**D**). Shown are protonemata and gametophores from cultures that had been growing in a long-day regime (16L:8D) for 19 days. Scale bars: 1mm.

5.3 Conclusions

Gene targeting was exploited to inactivate the *PpCOL* genes in *Physcomitrella patens*, and to replace the target gene with a GUS reporter gene. A total of 107 transformants was generated. Although a newly described procedure for the generation of double and triple gene disruptants was unsuccessful, several single gene disruptants were generated for *PpCOL1*, *PpCOL2*, and *PpCOL3*. In each line, abolition of the *PpCOL* gene and gene function was confirmed by PCR and RT-PCR. Expression analysis with the GUS reporter gene indicated that *PpCOL1*, *PpCOL2* and *PpCOL3* are expressed throughout the gametophyte stage, which is the dominant generation in mosses. However, none of the disruptants was found to display changes in development of the gametophyte, as they progressed normally from the single-celled protoplast stage, to the filamentous protonema stage, and finally to the three-dimensional gametophore stage, without any observable changes in morphology. Furthermore, lack of *PpCOL1*, *PpCOL2*, or *PpCOL3* function had no effect on the phototropic response of protonemal filaments. Possibly, the genes have redundant functions and inactivation of all three genes may be necessary to elucidate their function.

6 DISCUSSION

6.1 The evolution and divergence of *CO*-like genes

CO-like genes have been found in plants, but not in animals and yeast. In this study, *CO* homologues were found for the first time outside of the realm of flowering plants. Three *CO* homologues were isolated from the moss *Physcomitrella patens*: *PpCOL1*, *PpCOL2*, and *PpCOL3*. Analysis revealed that *PpCOL1*, *PpCOL2* and *PpCOL3* represent Group 1 genes, one of three classes of *CO*-like genes that also *CO* belongs to. Based on overall protein sequence homology and the presence of conserved motifs in the region between B-boxes and CCT domain, a further subdivision is made within Group 1 between Group 1a and Group 1c, both of which are represented by three genes in *Arabidopsis* (Griffiths *et al.*, 2003). Phylogenetic analysis indicated that *PpCOL1*, *PpCOL2*, and *PpCOL3* are members of Group 1c of *CO*-like genes, and as such more closely related to *AtCOL3*/*AtCOL4*/*AtCOL5* than to *AtCO*/*AtCOL1*/*AtCOL2*.

No more genes as closely related to *CO* as *PpCOL1*, *PpCOL2*, and *PpCOL3* were detected in *Physcomitrella* by Southern blotting, or analysis of EST libraries that are estimated to cover the transcriptome to at least 95% (Rensing *et al.*, 2002a). Therefore, Group 1a genes are likely to be absent from *Physcomitrella* and probably evolved only in the lineage leading to extant angiosperms, from Group 1c genes that were present in the progenitor of bryophytes and tracheophytes. Thus, *CO* itself appeared later in evolution than Group 1c genes *AtCOL3*, *AtCOL4*, and *AtCOL5*, and was likely recruited by flowering plants to take on a conserved role in photoperiodic flowering (Griffiths *et al.*, 2003; Hayama *et al.*, 2003). These findings indicate that *Arabidopsis* Group 1c genes *COL3*, *COL4* and *COL5* should be considered for functional studies, as they promise to throw light on more anciently diverged functions of *CO*-like genes than the function of *CO* orthologues in photoperiodic flowering. The functions of Group 1c genes have possibly been retained from the last common ancestor of bryophytes and tracheophytes.

Previous analysis of *CO*-like genes concluded that they evolve rapidly, particularly in the middle region between B-box zinc fingers and CCT domain (Lagercrantz and Axelsson, 2000). In agreement with this, it was found that the middle region of PpCOL1/PpCOL2/PpCOL3 were almost equally dissimilar to the middle region of AtCO/AtCOL1/AtCOL2 as to the middle region of AtCOL3/AtCOL4/AtCOL5 (data not shown), in spite of the closer phylogenetic relationship with the latter. Nevertheless, conserved peptide stretches were found in the middle region and at the carboxy terminus, most of which had already been identified in *Arabidopsis*, rice and barley Group 1 proteins (Griffiths *et al.*, 2003). The fact that these motifs have been retained in even more distantly related homologues, adds additional weight to the functional relevance of these regions. Between 7 and 25 residues long, they are probably too short to constitute an independently folding structural unit (Doolittle, 1995), as the ~43 residue long B-box zinc finger does (Borden *et al.*, 1995) and possibly also the 47 residue long CCT domain. Rather, they may represent a site of post-translational modification, or of cofactor binding.

All conserved motifs of the middle region (M1 to M4) and carboxy terminus (T) of Group 1 CO-like proteins were originally not noticed in sequence alignments of *Arabidopsis* paralogues (Robson *et al.*, 2001), only to be recognised upon inclusion of orthologues from two monocot species, rice and barley (Griffiths *et al.*, 2003). Originally, motifs M2 and M3 were not found in Group 1c *CO*-like genes of *Arabidopsis*, rice, and barley (Griffiths *et al.*, 2003). However, in this study the M3 motif was detected in Group 1c genes of *Arabidopsis* and *Physcomitrella*, and the entirely novel M2c motif was identified and found to be conserved between *Arabidopsis* and *Physcomitrella* Group 1c genes. Together, this illustrates that an even higher resolution can be obtained with respect to motif detection when orthologues from more distantly related species are compared, as expected from the fact that increased sequence erosion of functionally unconstrained regions makes regions that are constrained even more apparent. Such findings may provide important experimental leads for subsequent functional analyses.

The identification of the conserved M2c motif and the overall higher degree of amino acid identity strongly suggest a closer functional relationship of PpCOL1/PpCOL2/PpCOL3 to AtCOL3/AtCOL4/AtCOL5 than to CO or any other CO-like protein. Consistently, the *PpCOL* overexpression studies in *Arabidopsis* showed that the CONSTANS protein acquired distinct functional

features as it evolved towards a specialised function in photoperiodic flowering. Although the only conserved motifs that are recognised to diverge between Group 1a genes and Group 1c genes are the M2a and M2c motifs, the functional specificity of CO likely resides in other parts of the protein, because even the closely related Group 1a genes *AtCOL1* and *AtCOL2* failed to complement *CO* function (Ledger *et al.*, 2001).

In addition to *PpCOL1*, *PpCOL2*, and *PpCOL3*, one other *Physcomitrella CO*-like gene was identified that is likely to be a member of Group 3, because orthology to this clade was found separately in the B-box region and the CCT domain. Furthermore, the presence of one more Group 3 gene and one Group 2 gene has been implicated by orthology of isolated B-box and CCT domains. These findings suggest that the major isoforms of *Arabidopsis CO*-like genes, Groups 1, 2 and 3, are all present in *Physcomitrella*. Thus, the three isoforms seem to predate the bryophyte/tracheophyte divergence and to have retained distinctive B-box and CCT domain characteristics. The existence of common classes of *CO*-like genes in bryophytes and tracheophytes clearly suggests that *CO*-like genes have an ancient origin. In fact, when sequence databases were searched, an mRNA from the green alga *Ostreococcus tauri* was found that encodes a B-box near its amino terminus and a CCT domain near its carboxy terminus. Although phylogenetically, the predicted protein sequence of this gene, called *OtCOL*, clustered within the Group 2 clade of *Arabidopsis* CO-like proteins (data not shown), no sequence homology was found outside the B-box and CCT domain regions, and also the exon-intron structure was different (Figure 33). Therefore, instead of reflecting an orthologous relationship, the clustering with Group 2 genes might be the result of a shared single B-box region, as opposed to Group 1 and Group 3 genes which encode two B-boxes. It is possible that the B-box and CCT domain combination was the result of convergent evolution, in which case organisms from both lineages evolved proteins with the same combination of domains independently. Alternatively, relationships have been blurred by excessive sequence change outside of functionally constrained regions. Unlike the green algal *CO*-like gene, orthology of the *PpCOL* genes to distinct classes of *Arabidopsis CO*-like genes could readily be identified. Besides, alignment of *Arabidopsis* and *Physcomitrella CO*-like orthologues led to the discovery of the conserved M2c motif, as described previously, whereas comparison of monocot and dicot orthologues had failed to do so (Griffiths *et al.*, 2003).

Together, these findings suggest that, relative to *Arabidopsis*, *Physcomitrella* occupies a phylogenetically unique position for the study of gene ancestry, as well as for the detection of conserved peptide motifs of potential functional importance. The fact that >66% of *Arabidopsis* genes have homologues in *Physcomitrella* gametophytes (Nishiyama *et al.*, 2003) indicates that *Physcomitrella* may give such insights for many other genes and gene families.

In addition to *CO* homologues, three putative *Physcomitrella STO* homologues were identified in this study. Phylogenetic analysis revealed that the B-boxes of *Arabidopsis CO*- and *STO*-homologues were orthologous to the B-boxes of *Physcomitrella CO*- and *STO*-orthologues, respectively. Therefore, distinct B-box zinc fingers had already become locked in both classes of proteins before the divergence of bryophytes and tracheophytes, and have retained orthologous sequence resemblances since, probably due to functional constraints. In addition, *Physcomitrella* transcripts were identified that possess a CCT domain that is more similar to the CCT domain of TOC1 than to that of CO, suggesting that also the divergence of this domain is of ancient origin. Together, these findings further corroborate the notion that B-box zinc fingers and CCT domains are evolutionary mobile modules with independent functions. Furthermore, they indicate that the modular shuffling of B-box zinc fingers and CCT domains has been exploited in the plant kingdom long before the separation of bryophytes and vascular plants.

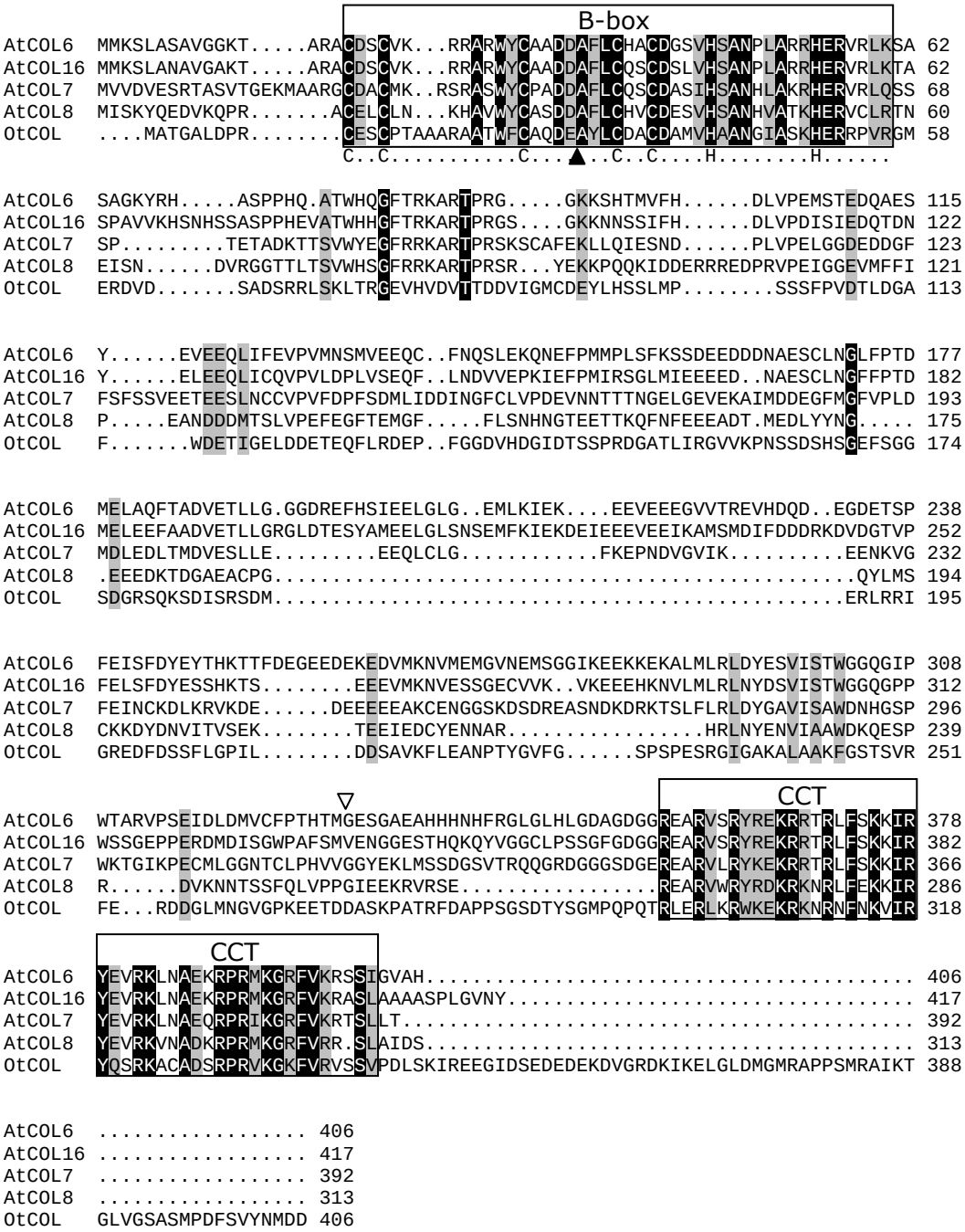


Figure 33 Alignment of predicted protein sequences of *AtCOL6*, *AtCOL7*, *AtCOL8*, *AtCOL16*, and *OtCOL*. Conserved domains are boxed. Amino acid residues conserved in all sequences are highlighted in black, similar amino acids are marked in grey. Conserved cysteine and histidine residues of the B-box zinc finger (Borden, 1998; Robson *et al.*, 2001) are shown below the alignment. Intron positions in the corresponding nucleotide sequences are indicated by a white triangle for AtCOL6, AtCOL7, AtCOL8 and AtCOL16, and by a black triangle for OtCOL. The sequence accession numbers are given in the chapter *Materials and methods*.

Group 1 and Group 2 *CO*-like genes of *Arabidopsis* contain one intron at a highly conserved position relative to the protein sequence, whereas Group 3 genes contain three introns at different positions. The three *Physcomitrella* Group 1 genes contain two introns, the second of which has a position relative to the protein sequence that is similar to the position of the single intron of *Arabidopsis* Group 1 genes. Therefore, this intron was likely inherited from a common ancestral Group 1 gene. The second intron of *PpCOL1*, *PpCOL2*, and *PpCOL3* was either gained specifically in the lineage leading to *Physcomitrella* (Babenko *et al.*, 2004), or lost in the lineage leading to *Arabidopsis*. Possibly, the introns of *PpCOL* genes have a regulatory function, because the expression analysis by quantitative RT-PCR identified transcripts that are likely to represent unspliced *PpCOL1*, *PpCOL2* and *PpCOL3* mRNA. Whereas unspliced *PpCOL1* transcripts were found to be less abundant, unspliced *PpCOL2* and *PpCOL3* transcripts might represent as much 15% of the total mRNA pools of those genes. Although strictly speaking the analysis identified the presence of the first intron in unspliced *PpCOL3* transcripts, and the presence of the second intron in unspliced *PpCOL1* and *PpCOL2* transcripts, it seems likely that also the latter will have retained the first intron. For each gene, the first intron introduces an in-frame stop codon that would prematurely terminate translation, resulting in proteins that contain the two B-boxes but are truncated between motifs M1 and M2c. In theory, such proteins could be functional, because B-boxes appear to constitute independently functional modules, as described previously. The unspliced transcripts may reside in the nucleus as an RNA intermediate, for example, of slowly processed hnRNA, without being translated. Alternatively, they may represent transcripts present in the cytoplasm in association with ribosomes, in the process of being translated. It is known that intron retention is a major phenomenon in *Arabidopsis* and in humans; about 2-3% of *Arabidopsis* transcripts contain retained introns (Ner-Gaon *et al.*, 2004), and the rate of intron retention is similar in humans (Kan *et al.*, 2002). In *Arabidopsis*, at least a subset of unspliced transcripts was associated with ribosomes, indicating that intron retention is a functional alternative transcript form (Ner-Gaon *et al.*, 2004). Whereas no evidence was found for intron retention in the *CO* transcript, the intron from a likely *CO* orthologue, *PnCO* from *Pharbitis nil*, is not efficiently spliced either (Liu *et al.*, 2001). The *PnCO* gene contains a single intron, at a similar position as the single intron of *AtCO* and the second

intron of *PpCOL1*, *PpCOL2*, and *PpCOL3*. As many as 16 out of 18 *PnCO* cDNA clones isolated from a cDNA library contained this intron, indicating that the unspliced form, theoretically resulting in a truncated protein, constitutes a major fraction of the *PnCO* mRNA pool (Liu *et al.*, 2001). Together with our findings, this suggests that intron retention in *CO*-like transcripts of *Physcomitrella* and *Pharbitis nil* either represents a shared regulatory mechanism of alternative splicing, or reflects a slow processing of hnRNA in general (or of *CO*-like hnRNA in particular) in both organisms. The fact that to date no other report was made of intron retention in *Physcomitrella* seems to disfavour a general inefficiency of hnRNA processing in this organism.

6.2 Transcriptional regulation of *PpCOL* genes

Our data demonstrated diurnal regulation of *PpCOL1*, *PpCOL2*, and *PpCOL3* expression, with a broad peak in expression during the day and a trough during the night. The expression of all three genes showed a rapid response to light. The first increase in transcript abundance was observed within 30 minutes after dawn, and maximum levels were reached within 2 hours, for *PpCOL1*, *PpCOL2*, as well as *PpCOL3*. Differences were observed between the three genes in the strength of the light response, which was about twice as strong for *PpCOL1* as it was for *PpCOL2* and *PpCOL3*.

The strong and rapid increase in *PpCOL1*, *PpCOL2*, and *PpCOL3* transcript abundance at dawn, together with the absence of any anticipatory increase shortly before dawn, suggests a direct response to light. Nevertheless, the pattern could also be the result of circadian clock regulation. The most reliable diagnostic feature of circadian rhythms is that they persist under constant conditions (Johnson, 2001). In continuous light conditions, expression levels of *PpCOL1*, *PpCOL2* and *PpCOL3* remained high and became arrhythmic, consistent with a direct and positive responsiveness to light. However, in extended darkness, expression levels increased, albeit more slowly than they did in the presence of light. Taken together, the data indicate that the diurnal expression pattern of *PpCOL1*, *PpCOL2*, and *PpCOL3* is achieved mainly by a direct response to light signalling, which overlies a weaker response to an endogenous timekeeper that becomes apparent only in

the absence of light. A recent publication confirmed these findings for *PpCOL1* (Shimizu *et al.*, 2004).

It is well established that rhythmic expression under control of the circadian clock is an essential feature of *CO* function (Suarez-Lopez *et al.*, 2001). However, microarray analyses have indicated that under particular conditions, *CO* expression can also display a direct and acute response to light, namely in seedlings that have been kept in darkness for several days and then exposed to light (Tepperman *et al.*, 2001; Jiao *et al.*, 2003; Tepperman *et al.*, 2004). Therefore, *CO* appears to display an underlying acute response to light that becomes apparent only in the absence of entraining light/dark cycles. Interestingly, the *PpCOL* genes seem to show the opposite dependency on light and circadian clock signals.

The convergence of light and circadian clock signalling at the transcriptional level has been studied in considerable detail for the *Arabidopsis* circadian marker gene *Lhcb1*1*. The promoter of this gene shows an acute response to red light that is mediated by phytochromes, as well as regulation by the circadian clock (Anderson *et al.*, 1997). However, phytochrome and clock signalling are unlikely to converge only at the promoter level, but rather form a network upstream of the promoter (Anderson *et al.*, 1997). In fact, it is established that a close association between light input and the oscillator itself is a common feature of circadian systems (Devlin, 2002). The output from the clock feeds back on the input, by modulating the responsiveness to photoreceptor signalling, a mechanism which is referred to as gating. In *Arabidopsis*, gating has been shown to be mediated by the *ELF3* gene (Figure 1, page 10). Whereas normally, *Lhcb1*1* expression levels oscillate in constant light and in constant darkness (Millar and Kay, 1996), in the *elf3* mutant, rhythmicity is maintained in constant darkness (Covington *et al.*, 2001), but lost in constant light (Hicks *et al.*, 1996). Thus *ELF3* is required for inhibiting light induction of *Lhcb1*1* expression during darkness.

Although the molecular identity of a putative *Physcomitrella* circadian clock is unknown, a few clock-controlled genes have been analysed to date. Interestingly, the expression of each clock-controlled gene, *PpLhcb2*, *PpSig5*, and *psbD*, only showed significant rhythmicity in constant darkness, not in constant light (Aoki *et al.*, 2004; Ichikawa *et al.*, 2004). This is in contrast with their respective *Arabidopsis* homologues, the transcript levels of which continue oscillating in constant light for several cycles (Millar and Kay, 1996;

Nakahira *et al.*, 1998; Morikawa *et al.*, 1999). Now, our analysis offers three additional examples of *Physcomitrella* clock-controlled genes whose expression reveals circadian control only in conditions of constant darkness. This is again in contrast with related genes in *Arabidopsis*, *CO*, *COL1*, *COL2*, and *COL9*, which all continue cycling in constant light conditions (Ledger *et al.*, 2001; Suarez-Lopez *et al.*, 2001; Cheng and Wang, 2005). Taken together, these observations possibly reflect fundamental differences in the responses of the *Physcomitrella* and *Arabidopsis* clocks to light. Interestingly, all the above mentioned *Physcomitrella* clock-controlled genes, including *PpCOL1*, *PpCOL2* and *PpCOL3*, respond to constant light and dark conditions in a similar way as *Lhcb1*1* and another circadian marker gene, *COLD-CIRCADIAN RHYTHM-RNA BINDING 2* (*CCR2*), in the *Arabidopsis elf3* mutant. Possibly, the *Physcomitrella* clock resembles the clock of the *Arabidopsis elf3* mutant in its inability to modulate light signalling in constant light conditions. However, other clock-controlled genes will have to be investigated to confirm this. Homologues of the *Arabidopsis CCR2* gene would be good candidates, because in *Arabidopsis* expression of the gene shows a robust circadian rhythm in both constant light and constant darkness (Kreps and Simon, 1997; Strayer *et al.*, 2000).

The light perceiving photoreceptors of *Physcomitrella* have been cloned and were shown to belong to the same three major classes of photoreceptors that are found in *Arabidopsis*: the red/far-red light perceiving phytochromes and the blue/UV-A light perceiving cryptochromes and phototropins. Each of the *Physcomitrella* photoreceptors has been functionally studied by gene targeting. Phytochromes were implicated in mediating phototropism, polarotropism and chloroplast movement (Mittmann *et al.*, 2004). Phototropins were shown to be involved in chloroplast movement (Kasahara *et al.*, 2004). Cryptochromes were found to regulate many steps in moss development, *e.g.* branching of protonema filaments and gametophore development, partly by controlling auxin signal transduction (Imaizumi *et al.*, 2002). Taken together, the same classes of photoreceptors that regulate light-dependent processes in *Arabidopsis* are also key regulators of light-dependent developmental and physiological processes in *Physcomitrella*.

Experiments were designed to dissect the light induction of *PpCOL* gene expression into responses to light of different wavelengths. This revealed that light signalling to the transcriptional control of *PpCOL1*, *PpCOL2*, and *PpCOL3*

is unlikely to depend on a photoreceptor activated by particular wavelengths of light, but rather on an integrated light signal. Further investigation of this matter may include analysing the expression of *PpCOL* genes in the different *Physcomitrella* photoreceptor mutants.

Previously, global gene expression analysis in *Arabidopsis* found massive light-regulated reprogramming of the transcriptome (Jiao *et al.*, 2005), whereby light perceived through distinct photoreceptors regulates distinct signalling pathways, as well as overlapping pathways (Ma *et al.*, 2001; Jiao *et al.*, 2003; Tepperman *et al.*, 2004). Direct interaction of photoreceptors (Ahmad *et al.*, 1998; Mas *et al.*, 2000) has been suggested to be one of the means to achieve regulation of a shared transcription cascade (Jiao *et al.*, 2003). In addition, *Arabidopsis* transcription factors like the HYPOCOTYL 5 (HY5) bZIP protein and the PHYTOCHROME INTERACTING FACTOR 3 (PIF3) bHLH were shown to regulate the expression of large sets of genes by light (Casal and Yanovsky, 2005). By analogy, our data suggest that *PpCOL1*, *PpCOL2* and *PpCOL3* may act in a light signalling transcription cascade, downstream of a transcription factor that integrates signalling of red, far-red and blue light. Remarkably, microarray analyses indicated that in dark-grown seedlings, *CO* expression also shows an acute response to light of any wavelength (Tepperman *et al.*, 2001; Jiao *et al.*, 2003; Tepperman *et al.*, 2004).

Every *CO*-like gene from a flowering plant that has been analysed to date displayed diurnal or circadian fluctuations in transcript abundance (Ledger *et al.*, 2001; Liu *et al.*, 2001; Hayama *et al.*, 2002; Martinez-Garcia *et al.*, 2002; Kim *et al.*, 2003; Nemoto *et al.*, 2003; Shin *et al.*, 2004). Together with the findings for *PpCOL1*, *PpCOL2* and *PpCOL3*, it is suggested that circadian/diurnal regulation of transcription is a generally conserved feature of *CO*-like genes. These observations are consistent with the notion that *CO*-like genes may have widely conserved roles in light signal transduction, and that during flowering plant evolution *CO* has taken on a specialised light signalling role controlling floral induction. Because *CO*-like genes were found in green algae, mosses, and flowering plants, but not in animals and yeast, the ancestral role of a *CO*-like gene may have been in light signalling to the control of a process that is restricted to photosynthetic organisms.

6.3 Functional analysis of *PpCOL* genes

CO is the only one of six *Arabidopsis* Group 1 *CO*-like genes for which a mutant has been reported to date (Koornneef *et al.*, 1991; Putterill *et al.*, 1995). The closely related *COL1* and *COL2* genes of *Arabidopsis* have been the subject of studies of overexpression, but not of inactivation (Ledger *et al.*, 2001). In the work presented here, the feasibility of gene targeting was exploited to generate disruptants of *PpCOL1*, *PpCOL2*, and *PpCOL3*. As such, mutants have been generated of every Group 1 gene present in *Physcomitrella*. This illustrates the value of *Physcomitrella* for functional studies of genes that belong to larger gene families, as gene families generally tending to be smaller in *Physcomitrella* than in *Arabidopsis* (Rensing *et al.*, 2002a).

The function of *PpCOL1*, *PpCOL2*, and *PpCOL3* was analysed in the respective disruptant lines, COL1/2-2, COL2-10, and COL1/3-6. The lack of *PpCOL1*, *PpCOL2*, or *PpCOL3* gene function did not result in any observable developmental defects in the haploid gametophyte stage. Cultures progressed normally through gametophyte development, and no morphological defects could be observed. The transcriptional analysis had indicated that the expression of *PpCOL* genes is induced by blue, red, as well as far-red light. Whereas *Physcomitrella* phytochrome and phototropin disruptants are not affected in development (Kasahara *et al.*, 2004; Mittmann *et al.*, 2004), abolition of blue light signalling through cryptochromes was shown to result in severe developmental defects, namely side branch induction and gametophore development (Imaizumi *et al.*, 2002). Such aberrations were not detected in *PpCOL* disruptants, consistent with the notion that these genes act in a light signalling cascade that can be activated by any wavelength of light.

Furthermore, the *PpCOL* disruptants were analysed for an easily tractable light response, namely the phototropic light response of protonemata. It was found that chloronemal filaments grow away from the light, and caulonemal filaments at slight angle away from the light, in disruptant cultures as well as in the wild type.

It is unlikely that there is residual gene activity after gene targeting, since gene targeting was designed in such a way that the targeted gene starting with the start codon would be replaced either completely (*PpCOL1*

and *PpCOL2*), or almost entirely (*PpCOL3*). Furthermore, RT-PCR analysis confirmed that the mRNA was absent in each of the disruptants.

Possibly, *PpCOL1*, *PpCOL2* and *PpCOL3* play a role in more subtle light responses that are elicited by light of different wavelengths. For example, chloroplast movements have been reported to be mediated by red light as well as blue light, through phytochrome and phototropin ligh signalling (Kasahara *et al.*, 2004; Mittmann *et al.*, 2004). However, the regulatory link between *PpCOL* gene expression on one hand, and phytochrome or phototropin light signalling on the other, should be established first, for example by analysis of *PpCOL* expression in the respective photoreceptor mutants, before embarking on such technically demanding phenotypical analyses. Finally, as the analysis of sporophyte development did not fit in the time frame of this work, a role in the sporophyte cannot be excluded. Nevertheless, since each of the genes was found to be expressed throughout the gametophyte stage, they may also be functional in the gametophyte.

Another possibility is that the *PpCOL* genes have redundant functions, in agreement with the high degree of sequence conservation and the very similar expression patterns. To address the problem of functional redundancy, attempts were undertaken to inactivate two or three *PpCOL* genes simultaneously, as recently reported (Hohe *et al.*, 2004). A total of 95 double and triple transformant lines was analysed, and targeting could be observed at single loci, but not at multiple loci. Even double disruptants were not obtained, which was surprising since our setup directly selected for double targeting, whereas the setup by Hohe *et al.* (2004) used the same selection marker for both targeting events. The double targeting efficiencies in this study were below 1,5%, in contrast to the rate of 5% reported by Hohe *et al.* (2005). Until this matter is further investigated, it can only be speculated that differences in the transformation procedure (*e.g.* preculturing of plant material for protoplast isolation) or in the intrinsic qualities of the target loci might be responsible for these discrepancies.

Whereas the *Arabidopsis* Group 1a and Group 1c proteins show only moderate sequence identity (~65% and ~50%, respectively), the Group 1c proteins from *Physcomitrella* possess a high degree of sequence identity (~80%). This suggests that PpCOL1, PpCOL2, and PpCOL3 experienced higher functional constraints than related homologues from *Arabidopsis*, or that they are of more recent origin. Interestingly, similar observations were made for

another well-studied family of transcription factors, the MADS-box gene family. While MIKC*-type MADS-box genes from *Arabidopsis* show only moderate sequence similarity, the MIKC*-type genes from *Physcomitrella* possess a high degree of sequence identity (Riese *et al.*, 2005). At the nucleotide level, *PpCOL1*, *PpCOL2*, and *PpCOL3* are rather uniformly and well related to each other (74-82%). And although some degree of homology exists in the 5' and 3' untranslated leader sequences, this does not continue beyond the transcript boundaries, nor is it observed in intron sequences. These observations disfavour a relatively recent evolutionary origin of *PpCOL1*, *PpCOL2*, and *PpCOL3*, but instead suggest that their strong sequence conservation is the product of functional constraints.

It is generally believed that expression divergence is a major reason for conserving duplicated genes (Blanc and Wolfe, 2004; Duarte *et al.*, 2005; Moore and Purugganan, 2005). However, no differences could be observed in the spatial expression pattern of *PpCOL* genes in the gametophyte, by analysis with the GUS reporter gene. Furthermore, the quantitative RT-PCR analysis revealed that at least in protonemata also the temporal expression patterns of the three genes are broadly similar, with smaller variations in the kinetics of expression. Therefore, other functional constraints seem to have been at work.

To date, several functional studies in *Physcomitrella* have identified pairs of very similar, functionally redundant genes. Examples include homologues of *LEAFY* (*LFY*) (Tanahashi *et al.*, 2005), *SNF1-RELATED KINASE 1* (*SNRK1*) (Thelander *et al.*, 2004), *GOLDEN2-LIKE* (*GLK*) (Yasumura *et al.*, 2005), *CRYPTOCHROME 1* (*CRY1*) (Imaizumi *et al.*, 2002) and *PHOTOTROPIN 1-2* (*PHOT1-2*) (Kasahara *et al.*, 2004). Besides, phylogenetic analysis of homologues of the widely conserved gene *RAD51* has prompted the suggestion of a rather recent duplication event for at least part of the *Physcomitrella* genome (Markmann-Mulisch *et al.*, 2002). The analysis of increasing amounts of genomic sequence has revealed that large-scale duplication events are common occurrences in the evolution of plant genomes (Blanc and Wolfe, 2004; Wessler and Carrington, 2005). And although the vast majority of duplicate genes suffer degenerative mutations that destine them for extinction (Presgraves, 2005), it has been suggested that genes that function in haploid tissue, like the gametophyte of *Physcomitrella patens*, might experience additional selective pressure to maintain duplicate copies,

particularly if the genes are essential for cell function (Yasumura *et al.*, 2005). Although this might explain the retention of functionally redundant doublets, like the homologues mentioned earlier, it does not explain the retention of functionally redundant triplets, like the *PpCOL* genes. Alternatively, only two *PpCOL* genes might have overlapping and redundant functions, with the third one having a divergent function. However, the fact that all three genes are rather uniformly related to each other at the nucleotide as well as at the amino acid level does not support this possibility. Simultaneous inactivation of two or three *PpCOL* genes may be necessary to answer these questions of functional redundancy, and of function altogether.

7 SUMMARY - ZUSAMMENFASSUNG

7.1 Summary

The *CONSTANS* (*CO*) gene plays a central role in the regulation of flowering time in *Arabidopsis*, and is the founding member of a family of 17 *CO* homologues. *CO* and *CO* homologues have been found in flowering plants, but not in yeast and animals. To address the question of the origin of *CO*, this gene family was analysed in the moss *Physcomitrella patens*, a phylogenetically distant organism.

In *Arabidopsis* and rice, three classes of *CO* homologues exist. The same three classes were found in *Physcomitrella*, suggesting that this gene family has ancient origins in the plant kingdom. In *Arabidopsis*, *CO* and 5 other genes belong to Group 1. Since only three Group 1 genes were identified in *Physcomitrella*, the family of *CO* homologues appears to be smaller in *Physcomitrella* than in *Arabidopsis*, in agreement with observations made with other gene families. Further analysis demonstrated that the *Physcomitrella* Group 1 genes are most similar to *Arabidopsis* Group 1 genes *COL3*/*COL4*/*COL5*, which are closely related to, but distinct from *CO*. An essential feature of *CO* function in *Arabidopsis* is a circadian controlled rhythm of transcript abundance. The three closely related *Physcomitrella* Group 1 genes have diurnal expression patterns that are distinct from the pattern of *CO* expression, and that are mainly caused by direct light induction. Distinct diurnal expression patterns are also observed for *CO* homologues that are not involved in control of flowering time. Consistently, the *Physcomitrella CO* homologues are unable to promote flowering upon expression in *Arabidopsis*. Together, the findings indicate that the *CO* branch of Group 1 genes does not exist in *Physcomitrella*. The role of *CO* in flowering time control was possibly derived from an ancestral function of Group 1 genes in light signal transduction.

The function of the three *Physcomitrella CO* homologues was studied by exploiting the feasibility of gene targeting. A disruptant was generated for each Group 1 *CO* homologue in *Physcomitrella*, whereas in *Arabidopsis* only *CO* has been inactivated to date. Phenotypical analysis of the disruptants revealed no developmental defects, nor an alteration of the phototropic growth response. The high degree of sequence conservation between the three genes and the similar expression patterns suggest redundancy. Therefore, simultaneous inactivation of all three genes may be necessary to elucidate their function.

7.2 Zusammenfassung

Das *CONSTANS* Gen (*CO*) spielt eine zentrale Rolle in der Regulation der Blühzeit von *Arabidopsis*. Es war das erste identifizierte Gen von einer Familie von 17 *CO*-homologen Genen. Das *CO* Gen und *CO*-homologe Gene wurden in Blütenpflanzen, aber nicht in Hefe oder Säugetieren nachgewiesen. Um die Herkunft des *CO* Gens herauszufinden, wurde die *CO* Genfamilie in dem Moos *Physcomitrella patens*, einem phylogenetisch weit von *Arabidopsis* entfernten Organismus, analysiert.

In *Arabidopsis* und in Reis gibt es drei Gruppen von *CO*-homologen Genen. Dass dieselben drei Gruppen auch in *Physcomitrella* nachgewiesen werden konnten, deutet auf einen sehr alten Ursprung der *CO* Genfamilie im Pflanzenreich hin. In *Arabidopsis* gehören *CO* und fünf andere Gene zur Gruppe 1. Da in *Physcomitrella* nur drei Gene der Gruppe 1 identifiziert werden konnten, ist vermutlich auch die Familie der *CO*-homologen Gene, wie schon für andere Genfamilien beobachtet, in *Physcomitrella* kleiner als in *Arabidopsis*. Weitere Untersuchungen zeigten, dass die *Physcomitrella* Gruppe 1 Gene eine höhere Homologie zu *Arabidopsis* Gruppe 1 Genen *COL3*/*COL4*/*COL5* haben. Diese Gene sind zwar mit *CO* eng verwandt, aber unterscheiden sich von *CO*. Ein wichtiges Merkmal der *CO* Funktion in *Arabidopsis* ist das circadian regulierte Transkriptvorkommen. Die drei Gruppe 1 Gene von *Physcomitrella* zeigen diurnale Expressionsmuster die sich vom Expressionsmuster von *CO* unterscheiden, und die hauptsächlich durch direkte Lichtinduktion hervorgerufen werden. *CO*-homologe Gene die nicht an der

Kontrolle der Blühzeit beteiligt sind, zeigen ebenfalls unterschiedliche diurnale Expressionsmuster. Damit übereinstimmend löst auch die Expression der *CO*-homologen Gene aus *Physcomitrella* in *Arabidopsis* kein Blühen aus. Zusammengenommen deuten die Ergebnisse daraufhin, dass der *CO*-Zweig der Gruppe 1 Gene in *Physcomitrella* nicht vorhanden ist. Die Rolle von *CO* in der Blühzeitkontrolle ist vermutlich auf eine ältere Funktion der Gruppe 1 Gene in der Lichtsignalweiterleitung zurückzuführen.

Um die Funktion der drei zu *CO* homologen Gene von *Physcomitrella* zu untersuchen, wurde die Methode des Gen-Targeting verwendet. Dazu wurde in *Physcomitrella* jedes der *CO*-homologen Gene der Gruppe 1 einzeln ausgeschaltet, während in *Arabidopsis* bisher ausschließlich *CO* inaktiviert wurde. Bei der Analyse des Phänotyps konnten weder Entwicklungsdefekte noch eine Änderung der phototropen Wachstumsantwort detektiert werden. Die hohe Konserviertheit der Gene und ähnliche Expressionsmuster deuten auf eine redundante Funktion der drei Gene hin. Um die Funktion der drei *CO*-homologen Gene herauszufinden, könnte es notwendig sein, alle drei Gene gleichzeitig zu inaktivieren.

8 LITERATURE

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

Ahmad, M., Jarillo, J.A., Smirnova, O., and Cashmore, A.R. (1998). The CRY1 blue light photoreceptor of Arabidopsis interacts with phytochrome A in vitro. Mol Cell 1, 939-948.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

An, H., Roussot, C., Suarez-Lopez, P., Corbesier, L., Vincent, C., Pineiro, M., Hepworth, S., Mouradov, A., Justin, S., Turnbull, C., and Coupland, G. (2004). CONSTANS acts in the phloem to regulate a systemic signal that induces photoperiodic flowering of Arabidopsis. Development 131, 3615-3626.

Anderson, S.L., Somers, D.E., Millar, A.J., Hanson, K., Chory, J., and Kay, S.A. (1997). Attenuation of phytochrome A and B signaling pathways by the Arabidopsis circadian clock. Plant Cell 9, 1727-1743.

Aoki, S., Kato, S., Ichikawa, K., and Shimizu, M. (2004). Circadian expression of the PpLhcb2 gene encoding a major light-harvesting chlorophyll a/b-binding protein in the moss *Physcomitrella patens*. Plant & Cell Physiology 45, 68-76.

Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., and Baulcombe, D.C. (2005). Cloning and characterization of micro-RNAs from moss. Plant J 43, 837-848.

Ashton, N.W., and Cove, D.J. (1977). The isolation and preliminary characterization of auxotrophic and analogue resistant mutants in the moss *Physcomitrella patens*. Molecular & General Genetics 154, 87-95.

Ashton, N.W., Champagne, C.E.M., Weiler, T., and Verkoczy, L.K. (2000). The bryophyte *Physcomitrella patens* replicates extrachromosomal transgenic elements. New Phytologist 146, 391-402.

Ayre, B.G., and Turgeon, R. (2004). Graft transmission of a floral stimulant derived from CONSTANS. Plant Physiol 135, 2271-2278.

Babenko, V.N., Rogozin, I.B., Mekhedov, S.L., and Koonin, E.V. (2004). Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic Acids Res 32, 3724-3733.

Bell-Pedersen, D., Cassone, V.M., Earnest, D.J., Golden, S.S., Hardin, P.E., Thomas, T.L., and Zoran, M.J. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nat Rev Genet 6, 544-556.

Bezanilla, M., Pan, A., and Quatrano, R.S. (2003). RNA interference in the moss Physcomitrella patens. Plant Physiol 133, 470-474.

Birnboim, H.C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7, 1513-1523.

Blanc, G., and Wolfe, K.H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell 16, 1667-1678.

Borden, K.L. (1998). RING fingers and B-boxes: zinc-binding protein-protein interaction domains. Biochem Cell Biol 76, 351-358.

Borden, K.L., Lally, J.M., Martin, S.R., O'Reilly, N.J., Etkin, L.D., and Freemont, P.S. (1995). Novel topology of a zinc-binding domain from a protein involved in regulating early Xenopus development. Embo J 14, 5947-5956.

Boss, P.K., Bastow, R.M., Mylne, J.S., and Dean, C. (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. Plant Cell 16 Suppl, S18-31.

Bünning, E. (1936). Die endogene Tagesrhythmik als Grundlage der photoperiodischen Reaktion. Ber Dtsch Bot Ges 54, 590-607.

Cainarca, S., Messali, S., Ballabio, A., and Meroni, G. (1999). Functional characterization of the Opitz syndrome gene product (midin): evidence for homodimerization and association with microtubules throughout the cell cycle. Hum Mol Genet 8, 1387-1396.

Cao, T., Borden, K.L., Freemont, P.S., and Etkin, L.D. (1997). Involvement of the rfp tripartite motif in protein-protein interactions and subcellular distribution. J Cell Sci 110 (Pt 14), 1563-1571.

Cao, T., Duprez, E., Borden, K.L., Freemont, P.S., and Etkin, L.D. (1998). Ret finger protein is a normal component of PML nuclear bodies and interacts directly with PML. J Cell Sci 111 (Pt 10), 1319-1329.

Casal, J.J., and Yanovsky, M.J. (2005). Regulation of gene expression by light. Int J Dev Biol 49, 501-511.

Cerdan, P.D., and Chory, J. (2003). Regulation of flowering time by light quality. Nature 423, 881-885.

Champagne, C.E.M., and Ashton, N.W. (2001). Ancestry of KNOX genes revealed by bryophyte (*Physcomitrella patens*) homologs. New Phytologist 150, 23-36.

Cheng, X.F., and Wang, Z.Y. (2005). Overexpression of COL9, a CONSTANS-LIKE gene, delays flowering by reducing expression of CO and FT in Arabidopsis thaliana. Plant J 43, 758-768.

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

Cove, D. (2000). The moss, *Physcomitrella patens*. Journal of Plant Growth Regulation 19, 275-283.

Cove, D. (2005). The moss physcomitrella patens. Annu Rev Genet 39, 339-358.

Cove, D.J. (1992). Regulation of development in the moss, *Physcomitrella patens*. (Berlin: Springer).

Cove, D.J., Schild, A., Ashton, N.W., and Hartmann, E. (1978). Genetic and physiological studies of the effect of light on the development of the moss, *Physcomitrella patens*. Photochemistry and Photobiology 27, 249-254.

Covington, M.F., Panda, S., Liu, X.L., Strayer, C.A., Wagner, D.R., and Kay, S.A. (2001). ELF3 modulates resetting of the circadian clock in Arabidopsis. Plant Cell 13, 1305-1315.

Devlin, P.F. (2002). Signs of the time: environmental input to the circadian clock. J Exp Bot 53, 1535-1550.

Devlin, P.F., Patel, S.R., and Whitelam, G.C. (1998). Phytochrome E influences internode elongation and flowering time in Arabidopsis. Plant Cell 10, 1479-1487.

Doolittle, R.F. (1995). The multiplicity of domains in proteins. Annu Rev Biochem 64, 287-314.

Doyle, M.R., Davis, S.J., Bastow, R.M., McWatters, H.G., Kozma-Bognar, L., Nagy, F., Millar, A.J., and Amasino, R.M. (2002). The ELF4 gene

controls circadian rhythms and flowering time in Arabidopsis thaliana. Nature 419, 74-77.
Duarte, J.M., Cui, L., Wall, P.K., Zhang, Q., Zhang, X., Leebens-Mack, J., Ma, H., Altman, N., and Depamphilis, C.W. (2005). Expression Pattern Shifts Following Duplication Indicative of Subfunctionalization and Neofunctionalization in Regulatory Genes of Arabidopsis. Mol Biol Evol.
Dudas, A., and Chovanec, M. (2004). DNA double-strand break repair by homologous recombination. Mutat Res 566, 131-167.
El-Din El-Assal, S., Alonso-Blanco, C., Peeters, A.J., Raz, V., and Koornneef, M. (2001). A QTL for flowering time in Arabidopsis reveals a novel allele of CRY2. Nat Genet 29, 435-440.
Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164-166.
Floyd, S.K., and Bowman, J.L. (2004). Gene regulation: ancient microRNA target sequences in plants. Nature 428, 485-486.
Gillette, M.U., and Sejnowski, T.J. (2005). Physiology. Biological clocks coordinately keep life on time. Science 309, 1196-1198.
Graham, L.E., Cook, M.E., and Busse, J.S. (2000). The origin of plants: body plan changes contributing to a major evolutionary radiation. Proc Natl Acad Sci U S A 97, 4535-4540.
Griffiths, S., Dunford, R.P., Coupland, G., and Laurie, D.A. (2003). The evolution of CONSTANS-like gene families in barley, rice, and *Arabidopsis*. Plant Physiology 131, 1855-1867.
Guo, H., Yang, H., Mockler, T.C., and Lin, C. (1998). Regulation of flowering time by Arabidopsis photoreceptors. Science 279, 1360-1363.
Haber, J.E. (1999). DNA recombination: the replication connection. Trends Biochem Sci 24, 271-275.
Hadi, M.Z., Kemper, E., Wendeler, E., and Reiss, B. (2002). Simple and versatile selection of Arabidopsis transformants. Plant Cell Reports 21, 130–135.
Halliday, K.J., Koornneef, M., and Whitelam, G.C. (1994). Phytochrome B and at Least One Other Phytochrome Mediate the Accelerated Flowering Response of Arabidopsis thaliana L. to Low Red/Far-Red Ratio. Plant Physiology 104, 1311-1315.
Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J Mol Biol. 166, 557-580.
Harmer, S.L., Hogenesch, J.B., Straume, M., Chang, H.S., Han, B., Zhu, T., Wang, X., Kreps, J.A., and Kay, S.A. (2000). Orchestrated transcription of key pathways in Arabidopsis by the circadian clock. Science 290, 2110-2113.
Hayama, R., and Coupland, G. (2004). The molecular basis of diversity in the photoperiodic flowering responses of *Arabidopsis* and rice. Plant Physiology 135, 677-684.
Hayama, R., Izawa, T., and Shimamoto, K. (2002). Isolation of rice genes possibly involved in the photoperiodic control of flowering by a fluorescent differential display method. Plant Cell Physiol 43, 494-504.
Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature 422, 719-722.
Hecht, V., Foucher, F., Ferrandiz, C., Macknight, R., Navarro, C., Morin, J., Vardy, M.E., Ellis, N., Beltran, J.P., Rameau, C., and Weller, J.L. (2005). Conservation of Arabidopsis flowering genes in model legumes. Plant Physiol 137, 1420-1434.

Hedges, S.B., Blair, J.E., Venturi, M.L., and Shoe, J.L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol 4.

Henschel, K., Kofuji, R., Hasebe, M., Saedler, H., Münster, T., and Theißen, G. (2002). Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella patens*. Molecular Biology and Evolution 19, 801-814.

Hicks, K.A., Millar, A.J., Carre, I.A., Somers, D.E., Straume, M., Meeks-Wagner, D.R., and Kay, S.A. (1996). Conditional circadian dysfunction of the Arabidopsis early-flowering 3 mutant. Science 274, 790-792.

Hofmann, A.H., Codon, A.C., Ivascu, C., Russo, V.E., Knight, C., Cove, D., Schaefer, D.G., Chakhparonian, M., and Zryd, J.P. (1999). A specific member of the Cab multigene family can be efficiently targeted and disrupted in the moss Physcomitrella patens. Mol Gen Genet 261, 92-99.

Hohe, A., Egener, T., Lucht, J.M., Holtorf, H., Reinhard, C., Schween, G., and Reski, R. (2004). An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, Physcomitrella patens. Curr Genet 44, 339-347.

Holm, M., Hardtke, C.S., Gaudet, R., and Deng, X.W. (2001). Identification of a structural motif that confers specific interaction with the WD40 repeat domain of Arabidopsis COP1. Embo J 20, 118-127.

Huang, T., Bohlenius, H., Eriksson, S., Parcy, F., and Nilsson, O. (2005). The mRNA of the Arabidopsis gene FT moves from leaf to shoot apex and induces flowering. Science 309, 1694-1696.

Huq, E., Tepperman, J.M., and Quail, P.H. (2000). GIGANTEA is a nuclear protein involved in phytochrome signaling in Arabidopsis. Proc Natl Acad Sci U S A 97, 9789-9794.

Ichikawa, K., Sugita, M., Imaizumi, T., Wada, M., and Aoki, S. (2004). Differential expression on a daily basis of plastid sigma factor genes from the moss Physcomitrella patens. Regulatory interactions among PpSig5, the circadian clock, and blue light signaling mediated by cryptochromes. Plant Physiol 136, 4285-4298.

Imaizumi, T., Kadota, A., Hasebe, M., and Wada, M. (2002). Cryptochrome light signals control development to suppress auxin sensitivity in the moss *Physcomitrella patens*. The Plant Cell 14, 373-386.

Imaizumi, T., Tran, H.G., Swart, T.E., Briggs, W.R., and Kay, S.A. (2003). FKF1 is essential for photoperiodic-specific light signalling in *Arabidopsis*. Nature 426, 302-306.

Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A., and Kay, S.A. (2005). FKF1 F-box protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. Science 309, 293-297.

Izawa, T., Oikawa, T., Sugiyama, N., Tanisaka, T., Yano, M., and Shimamoto, K. (2002). Phytochrome mediates the external light signal to repress FT orthologs in photoperiodic flowering of rice. Genes Dev 16, 2006-2020.

Jiao, Y., Ma, L., Strickland, E., and Deng, X.W. (2005). Conservation and Divergence of Light-Regulated Genome Expression Patterns during Seedling Development in Rice and Arabidopsis. Plant Cell.

Jiao, Y., Yang, H., Ma, L., Sun, N., Yu, H., Liu, T., Gao, Y., Gu, H., Chen, Z., Wada, M., Gerstein, M., Zhao, H., Qu, L.J., and Deng, X.W. (2003). A genome-wide analysis of blue-light regulation of Arabidopsis

transcription factor gene expression during seedling development. Plant Physiol 133, 1480-1493.
Johnson, C.H. (2001). Endogenous timekeepers in photosynthetic organisms. Annu Rev Physiol 63, 695-728.
Kamisugi, Y., Cuming, A.C., and Cove, D.J. (2005). Parameters determining the efficiency of gene targeting in the moss Physcomitrella patens. Nucleic Acids Res 33, e173.
Kan, Z., States, D., and Gish, W. (2002). Selecting for functional alternative splices in ESTs. Genome Res 12, 1837-1845.
Kasahara, M., Kagawa, T., Sato, Y., Kiyosue, T., and Wada, M. (2004). Phototropins mediate blue and red light-induced chloroplast movements in Physcomitrella patens. Plant Physiol 135, 1388-1397.
Kim, S.-J., Moon, J., Lee, I., Maeng, J., and Kim, S.-R. (2003). Molecular cloning and expression analysis of a CONSTANS homologue, PnCOL1, from *Pharbitis nil*. Journal of Experimental Botany 54, 1879-1887.
Knight, C.D., and Cove, D.J. (1989). The genetic analysis of tropic responses. Environ Exp Bot 29, 57-70.
Knight, C.D., Sehgal, A., Atwal, K., Wallace, J.C., Cove, D.J., Coates, D., Quatrano, R.S., Bahadur, S., Stockley, P.G., and Cuming, A.C. (1995). Molecular Responses to Abscisic Acid and Stress Are Conserved between Moss and Cereals. Plant Cell 7, 499-506.
Koncz, C., and Schell, J. (1986). The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. Mol Gen Genet 204, 383-396.
Koornneef, M., Hanhart, C.J., and van der Veen, J.H. (1991). A genetic and physiological analysis of late flowering mutants in Arabidopsis thaliana. Mol Gen Genet 229, 57-66.
Kreps, J.A., and Simon, A.E. (1997). Environmental and genetic effects on circadian clock-regulated gene expression in Arabidopsis. Plant Cell 9, 297-304.
Krogan, N.T., and Ashton, N.W. (2000). Ancestry of plant MADS-box genes revealed by bryophyte (*Physcomitrella patens*) homologues. New Phytologist 147, 505-517.
Kurup, S., Jones, H.D., and Holdsworth, M.J. (2000). Interactions of the developmental regulator ABI3 with proteins identified from developing Arabidopsis seeds. Plant J 21, 143-155.
Lagercrantz, U., and Axelsson, T. (2000). Rapid evolution of the family of CONSTANS LIKE genes in plants. Mol Biol Evol 17, 1499-1507.
Ledger, S., Strayer, C., Ashton, F., Kay, S.A., and Putterill, J. (2001). Analysis of the function of two circadian-regulated CONSTANS-LIKE genes. The Plant Journal 26, 15-22.
Lee, K.J., Sakata, Y., Mau, S.L., Pettolino, F., Bacic, A., Quatrano, R.S., Knight, C.D., and Knox, J.P. (2005). Arabinogalactan Proteins Are Required for Apical Cell Extension in the Moss Physcomitrella patens. Plant Cell.
Lippuner, V., Cyert, M.S., and Gasser, C.S. (1996). Two classes of plant cDNA clones differentially complement yeast calcineurin mutants and increase salt tolerance of wild-type yeast. Journal of Biological Chemistry 271, 12859-12866.
Liu, J., Yu, J., McIntosh, L., Kende, H., and Zeevaart, J.A.D. (2001). Isolation of a CONSTANS ortholog from *Pharbitis nil* and its role in flowering. Plant Physiology 125, 1821-1830.

Liu, Y., Tsinoremas, N.F., Johnson, C.H., Lebedeva, N.V., Golden, S.S., Ishiura, M., and Kondo, T. (1995). Circadian orchestration of gene expression in cyanobacteria. Genes Dev 9, 1469-1478.

Lohmann, J.U., and Weigel, D. (2002). Building beauty: the genetic control of floral patterning. Dev Cell 2, 135-142.

Ma, L., Li, J., Qu, L., Hager, J., Chen, Z., Zhao, H., and Deng, X.W. (2001). Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. Plant Cell 13, 2589-2607.

Makino, S., Kiba, T., Imamura, A., Hanaki, N., Nakamura, A., Suzuki, T., Taniguchi, M., Ueguchi, C., Sugiyama, T., and Mizuno, T. (2000). Genes encoding pseudo-response regulators: insight into His-to-Asp phosphorelay and circadian rhythm in Arabidopsis thaliana. Plant Cell Physiol 41, 791-803.

Markmann-Mulisch, U., Reiss, B., and Mulisch, M. (1999). Cell type-specific gene expression in the cell cycle of the dimorphic ciliate *Eufolliculina uhligi*. Molecular & General Genetics 262, 390-399.

Markmann-Mulisch, U., Hadi, M.Z., Koepchen, K., Alonso, J.C., Russo, V.E.A., Schell, J., and Reiss, B. (2002). The organization of *Physcomitrella patens* RAD51 genes is unique among eukaryotic organisms. Proceedings of the National Academy of Sciences of the United States of America 99, 2959-2964.

Martin, J., Storgaard, M., Andersen, C.H., and Nielsen, K.K. (2004). Photoperiodic regulation of flowering in perennial ryegrass involving a CONSTANS -like homolog. Plant Mol Biol 56, 159-169.

Martinez-Garcia, J.F., Virgos-Soler, A., and Prat, S. (2002). Control of photoperiod-regulated tuberization in potato by the *Arabidopsis* flowering-time gene CONSTANS. Proceedings of the National Academy of Sciences of the United States of America 99, 15211-15216.

Mas, P., Devlin, P.F., Panda, S., and Kay, S.A. (2000). Functional interaction of phytochrome B and cryptochrome 2. Nature 408, 207-211.

Masaki, T., Tsukagoshi, H., Mitsui, N., Nishii, T., Hattori, T., Morikami, A., and Nakamura, K. (2005). Activation tagging of a gene for a protein with novel class of CCT-domain activates expression of a subset of sugar-inducible genes in Arabidopsis thaliana. Plant J 43, 142-152.

Matsushika, A., Makino, S., Kojima, M., and Mizuno, T. (2000). Circadian waves of expression of the APRR1/TOC1 family of pseudo-response regulators in Arabidopsis thaliana: insight into the plant circadian clock. Plant Cell Physiol 41, 1002-1012.

Millar, A.J., and Kay, S.A. (1996). Integration of circadian and phototransduction pathways in the network controlling CAB gene transcription in Arabidopsis. Proc Natl Acad Sci U S A 93, 15491-15496.

Mittmann, F., Brucker, G., Zeidler, M., Repp, A., Abts, T., Hartmann, E., and Hughes, J. (2004). Targeted knockout in *Physcomitrella* reveals direct actions of phytochrome in the cytoplasm. Proceedings of the National Academy of Sciences of the United States of America 101, 13939-13944.

Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., Mouradov, A., Fowler, S., Kamada, H., Putterill, J., and Coupland, G. (2005). Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in Arabidopsis. Plant Cell 17, 2255-2270.

Mockler, T., Yang, H., Yu, X., Parikh, D., Cheng, Y.C., Dolan, S., and Lin, C. (2003). Regulation of photoperiodic flowering by Arabidopsis photoreceptors. Proc Natl Acad Sci U S A 100, 2140-2145.

Moore, R.C., and Purugganan, M.D. (2005). The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol 8, 122-128.

Morikawa, K., Ito, S., Tsunoyama, Y., Nakahira, Y., Shiina, T., and Toyoshima, Y. (1999). Circadian-regulated expression of a nuclear-encoded plastid sigma factor gene (sigA) in wheat seedlings. FEBS Lett 451, 275-278.

Nakahira, Y., Baba, K., Yoneda, A., Shiina, T., and Toyoshima, Y. (1998). Circadian-regulated transcription of the psbD light-responsive promoter in wheat chloroplasts. Plant Physiol 118, 1079-1088.

Nelson, D.C., Lasswell, J., Rogg, L.E., Cohen, M.A., and Bartel, B. (2000). FKF1, a clock-controlled gene that regulates the transition to flowering in Arabidopsis. Cell 101, 331-340.

Nemoto, Y., Kisaka, M., Fuse, T., Yano, M., and Ogihara, Y. (2003). Characterization and functional analysis of three wheat genes with homology to the CONSTANS flowering time gene in transgenic rice. The Plant Journal 36, 82-93.

Ner-Gaon, H., Halachmi, R., Savaldi-Goldstein, S., Rubin, E., Ophir, R., and Fluhr, R. (2004). Intron retention is a major phenomenon in alternative splicing in Arabidopsis. Plant J 39, 877-885.

Nishiyama, T., Fujita, T., Shin, I.T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A., Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y., and Hasebe, M. (2003). Comparative genomics of Physcomitrella patens gametophytic transcriptome and Arabidopsis thaliana: implication for land plant evolution. Proc Natl Acad Sci U S A 100, 8007-8012.

Onouchi, H., Igeno, M.I., Perilleux, C., Graves, K., and Coupland, G. (2000). Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among Arabidopsis flowering-time genes. Plant Cell 12, 885-900.

Paques, F., and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 63, 349-404.

Parcy, F. (2005). Flowering: a time for integration. Int J Dev Biol 49, 585-593.

Pittendrigh, C.S., and Minis, D.H. (1964). The entrainment of circadian oscillations by light and their role as photoperiodic clocks. Am Nat 108, 261-295.

Presgraves, D.C. (2005). Evolutionary genomics: new genes for new jobs. Curr Biol 15, R52-53.

Puchta, H. (2005). The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. J Exp Bot 56, 1-14.

Putterill, J., Robson, F., Lee, K., Simon, R., and Coupland, G. (1995). The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80, 847-857.

Reiss, B. (2003). Homologous recombination and gene targeting in plant cells. Int Rev Cytol 228, 85-139.

Rensing, S.A., Rombauts, S., Van de Peer, Y., and Reski, R. (2002a). Moss transcriptome and beyond. Trends Plant Sci 7, 535-538.

Rensing, S.A., Rombauts, S., Hohe, A., Lang, D., Duwenig, E., Rouze, P., Van de Peer, Y., and Reski, R. (2002b). The transcriptome of the moss *Physcomitrella patens*: comparative analysis reveals a rich source of new genes. http://www.plant-biotech.net/Rensing_et_al_transcriptome2002.pdf.

Reyes, J.C., Muro-Pastor, M.I., and Florencio, F.J. (2004). The GATA family of transcription factors in *Arabidopsis* and rice. Plant Physiology 134, 1718-1732.

Riese, M., Faigl, W., Quodt, V., Verelst, W., Matthes, A., Saedler, H., and Munster, T. (2005). Isolation and characterization of new MIKC*-Type MADS-box genes from the moss Physcomitrella patens. Plant Biol (Stuttg) 7, 307-314.

Robert, L.S., Robson, F., Sharpe, A., Lydiate, D., and Coupland, G. (1998). Conserved structure and function of the *Arabidopsis* flowering time gene CONSTANS in *Brassica napus*. Plant Molecular Biology 37, 763-772.

Robson, F., Costa, M.M.R., Hepworth, S.R., Vizir, I., Pnieiro, M., Reeves, P.H., Putterill, J., and Coupland, G. (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. The Plant Journal 28, 619-631.

Roenneberg, T., and Merrow, M. (2002). Life before the clock: modeling circadian evolution. J Biol Rhythms 17, 495-505.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology & Evolution 4, 406-425.

Sakakibara, K., Nishiyama, T., Kato, M., and Hasebe, M. (2001). Isolation of homeodomain-leucine zipper genes from the moss *Physcomitrella patens* and the evolution of homeodomain-leucine zipper genes in land plants. Molecular Biology & Evolution 18, 491-502.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., and Coupland, G. (2000). Distinct roles of CONSTANS target genes in reproductive development of Arabidopsis. Science 288, 1613-1616.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. (NY: Cold Spring Harbor Laboratory Press).

Schaefer, D., Zryd, J.P., Knight, C.D., and Cove, D.J. (1991). Stable transformation of the moss Physcomitrella patens. Mol Gen Genet 226, 418-424.

Schaefer, D.G. (2001). Gene targeting in Physcomitrella patens. Curr Opin Plant Biol 4, 143-150.

Schaefer, D.G., and Zryd, J.P. (1997). Efficient gene targeting in the moss *Physcomitrella patens*. The Plant Journal 11, 1195-1206.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., and Wisman, E. (2001). Microarray analysis of diurnal and circadian-regulated genes in Arabidopsis. Plant Cell 13, 113-123.

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carre, I.A., and Coupland, G. (1998). The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. Cell 93, 1219-1229.

Schween, G., Hohe, A., Koprovova, A., and Reski, R. (2002). High-throughput PCR screen of 15,000 transgenic *Physcomitrella* plants. Plant Mol Biol Rep 20, 43-47.

Sharrock, R.A., and Clack, T. (2002). Patterns of expression and normalized levels of the five Arabidopsis phytochromes. Plant Physiol 130, 442-456.

Shikata, M., Matsuda, Y., Ando, K., Nishii, A., Takemura, M., Yokota, A., and Kohchi, T. (2004). Characterization of Arabidopsis ZIM, a member of a novel plant-specific GATA factor gene family. J Exp Bot 55, 631-639.

Shimizu, M., Ichikawa, K., and Aoki, S. (2004). Photoperiod-regulated expression of the PpCOL1 gene encoding a homolog of CO/COL proteins in the moss *Physcomitrella patens*. Biochemical and Biophysical Research Communications 324, 1296-1301.

Shin, B.-S., Lee, J.-H., Lee, J.-H., Jeong, H.-J., Yun, C.-H., and Kim, J.-K. (2004). Circadian regulation of rice (*Oryza Sativa* L.) CONSTANS-like gene transcripts. Molecular Cells 17, 10-16.

Somers, D.E., Schultz, T.F., Milnamow, M., and Kay, S.A. (2000). ZEITLUPE encodes a novel clock-associated PAS protein from Arabidopsis. Cell 101, 319-329.

Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Mas, P., Panda, S., Kreps, J.A., and Kay, S.A. (2000). Cloning of the *Arabidopsis* clock gene TOC1, an autoregulatory response regulator homolog. Science 289, 768-771.

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

Takada, S., and Goto, K. (2003). Terminal flower2, an Arabidopsis homolog of heterochromatin protein1, counteracts the activation of flowering locus T by constans in the vascular tissues of leaves to regulate flowering time. Plant Cell 15, 2856-2865.

Tanahashi, T., Sumikawa, N., Kato, M., and Hasebe, M. (2005). Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss Physcomitrella patens. Development 132, 1727-1736.

Tepperman, J.M., Zhu, T., Chang, H.S., Wang, X., and Quail, P.H. (2001). Multiple transcription-factor genes are early targets of phytochrome A signaling. Proc Natl Acad Sci U S A 98, 9437-9442.

Tepperman, J.M., Hudson, M.E., Khanna, R., Zhu, T., Chang, S.H., Wang, X., and Quail, P.H. (2004). Expression profiling of phyB mutant demonstrates substantial contribution of other phytochromes to red-light-regulated gene expression during seedling de-etiolation. Plant J 38, 725-739.

Thelander, M., Olsson, T., and Ronne, H. (2004). Snf1-related protein kinase 1 is needed for growth in a normal day-night light cycle. The EMBO Journal 23, 1900-1910.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22, 4673–4680.

Torok, M., and Etkin, L.D. (2001). Two B or not two B? Overview of the rapidly expanding B-box family of proteins. Differentiation 67, 63-71.

Toth, R., Kevei, E., Hall, A., Millar, A.J., Nagy, F., and Kozma-Bognar, L. (2001). Circadian clock-regulated expression of phytochrome and cryptochrome genes in Arabidopsis. Plant Physiol 127, 1607-1616.

Valverde, F., Mouradov, A., Soppe, W., Ravenscroft, D., Samach, A., and Coupland, G. (2004). Photoreceptor regulation of CONSTANS protein in photoperiodic flowering. Nature 303, 1003-1006.

Vierstra, R.D. (2003). The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci 8, 135-142.

Wada, M., Kagawa, T., and Sato, Y. (2003). Chloroplast movement. Annu Rev Plant Biol 54, 455-468.

Wang, Z.Y., and Tobin, E.M. (1998). Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. Cell 93, 1207-1217.

Weigel, M., Varotto, C., Pesaresi, P., Finazzi, G., Rappaport, F., Salamini, F., and Leister, D. (2003). Plastocyanin is indispensable for photosynthetic electron flow in Arabidopsis thaliana. J Biol Chem 278, 31286-31289.

Wessler, S.R., and Carrington, J.C. (2005). The consequences of gene and genome duplication in plants. Curr Opin Plant Biol 8, 119-121.

Wigge, P.A., Kim, M.C., Jaeger, K.E., Busch, W., Schmid, M., Lohmann, J.U., and Weigel, D. (2005). Integration of spatial and temporal information during floral induction in Arabidopsis. Science 309, 1056-1059.

Yanovsky, M.J., and Kay, S.A. (2002). Molecular basis of seasonal time measurement in *Arabidopsis*. Nature 419, 308-312.

Yasumura, Y., Moylan, E.C., and Langdale, J.A. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. Plant Cell 17, 1894-1907.

Yuceer, C., Harkess, R.L., Land Jr. , S.B., and Luthe, D.S. (2002). Structural and developmental regulation of *CONSTANS-LIKE* genes isolated from *Populus deltoides*. Plant Science 163, 615-625.

9 APPENDIX

9.1 Acknowledgments

I would like to express my gratitude to Prof. Dr. George Coupland for the opportunity to work in his department and for his precious advice and support.

My gratefulness goes to Dr. Bernd Reiss for his support and supervision.

For financial support, I thank the Deutsche Forschungsgemeinschaft (Graduiertenkolleg Molekulare Analyse von Entwicklungsprozessen) and the Max-Planck-Gesellschaft.

I am particularly grateful to Dr. Andrea Lauri and Dr. Bozena Chrost for critical reading of the thesis, to Vanpatiënten & Vanderpleegster for last-minute corrections, to Bozena Chrost for translations, and to Henrike Gebhardt for translations and more.

I am thankful to past and present members of the Reiss and Coupland groups for always being available for questions and advice, and to Dr. Kurt Stueber for valuable help with DNA sequence and phylogenetic analysis.

I thank the friends and colleagues that provided pleasure during business, particularly Kenneth Berendzen, Andrea Busch, Maria Cartolano, Bozena Chrost, Markus Garzón, Hugo Konijn, Andrea Lauri, Dean Ravenscroft, Iain Searle and Edelgard Wendeler. Finally, I thank Frau Esser for 5 precious days.

9.2 Eidesstattliche Erklärung

Ich versichere, daß die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Dr. Bernd Reiss und Prof. Dr. George Coupland betreut worden.

Teilpublikation: Zobell, O., Coupland, G., and Reiss, B. (2005). The family of *CONSTANS*-like genes in *Physcomitrella patens*. Plant Biol (Stuttg) 7, 266-275

Oliver Zobell

9.3 Lebenslauf

Persönliche Angaben:

Name:	Oliver Zobell
Anschrift:	Kolibriweg 14, 50829 Köln
Geburtsdatum:	24.03.1977
Geburtsort:	Wilrijk (Antwerpen), Belgien
Staatsangehörigkeit:	belgisch
Familienstand:	unverheiratet

Schulbildung:

1983-1989	Besuch der *Basisschool* in Schoten, Belgien
1989-1995	Besuch der *Middelbare school* Onze-Lieve-Vrouwe-College in Antwerpen, Belgien
Juni 1995	Abschluss: allgemeine Hochschulreife

Studium und Beruf:

1996-1998	*Kandidaat* Studium Biologie an der Rijksuniversiteit Antwerpen, Belgien
1998-2000	*Licentiaat* Studium Biotechnologie an der Rijksuniversiteit Gent, Belgien; Diplomarbeit unter der Anleitung von Prof. Dr. Ann Depicker in der Abteilung *Moleculaire Genetica* der Rijksuniversiteit Gent
5. Juli 2000	Verleihung des akademischen Grades *Licentiaat in de Biotechnologie*
1. Sep. 2000-15. Feb. 2002	Wissenschaftlicher Mitarbeiter in der Arbeitsgruppe von Prof. Dr. Ed Hurt am Biochemie Zentrum der Universität Heidelberg

Promotion:

Seit 1. April 2002	Dissertation im Fach Botanik an der Universität zu Köln; Experimentelle Arbeit durchgeführt am Max-Planck-Institut für Züchtungsforschung in Köln, unter Anleitung von Dr. Bernd Reiss und Prof. Dr. George Coupland