Daysleeper : from genomic parasite to indispensable gene

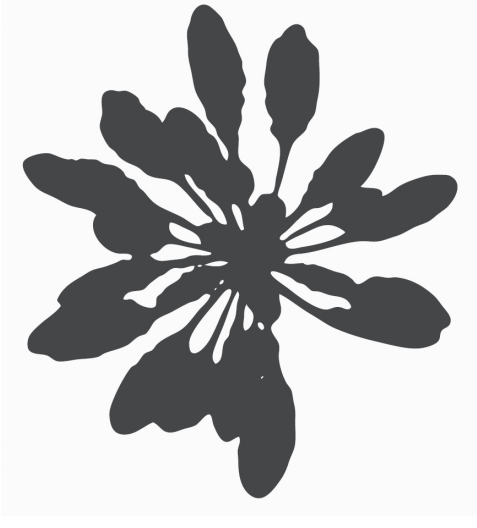
Knip, M.

Chapter 1

DAYSLEEPER: An essential domesticated transposase in higher plants

Marijn Knip

Department of Molecular and Developmental Genetics, Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE Leiden, the Netherlands



General introduction

1

Outline

This thesis describes the evolutionary backgrounds, localization and function of DAYSLEEPER, a protein encoded by the *DAYSLEEPER* gene in the dicotyledonous plant *Arabidopsis thaliana*. Here, an outline will be drawn, putting the subsequent experimental chapters into context.

DAYSLEEPER was described by Bundock and Hooykaas in 2005 [1]. *DAYSLEEPER* was found to be essential in *Arabidopsis thaliana* and closely resemble hAT-transposon superfamily-type transposases. Identified by binding *in vivo* to the promoter of the *Arabidopsis thaliana KU70* DNA-repair gene in yeast, DAYSLEEPER was suggested to function possibly as a transcription factor. *DAYSLEEPER* is a so-called domesticated transposase gene; the process of domestication describes an event whereby a transposon derived sequence loses its transposon-related function and gets taken up by the host as functional genetic material [2]. This process is a recurrent process in evolution: In the last number of years many examples of transposon domestication genes have been described in a great number of species [3]. Many of the domesticated genes fulfill a role in the nucleus, often as transcription factors. Considering transposases already possess a DNA binding domain and a nuclear localization sequence, their evolution into nuclear transcription factors seems logical [4].

DAYSLEEPER is essential in *Arabidopsis thaliana* and is the first domesticated transposase that has been shown to be essential to its host [1]. Plantlets lacking *DAYSLEEPER* display a severe phenotype and eventually die, failing to progress past the early seedling stage [1]. Primary transformants overexpressing *DAYSLEEPER* (*35S::DAYSLEEPER*) grow slowly, display delayed flowering, altered cauline leaves and partial or total sterility [1]. Interestingly, this phenotype increased in severity in the subsequent generation, becoming lethal at the seedling stage in the T_2 generation [1].

A brief introduction to transposons

Transposons are mobile genetic elements that can move through the genome of their host by a “copy-paste” or “cut-paste” manner. Transposons were first discovered in the 1940’s by Barbara McClintock [5]. Since then, transposons have long been viewed as genetic parasites. Transposons were discarded as “junk-DNA”, having no apparent

beneficial function to their hosts at all [6, 7]. This view generally persisted until approximately two decades ago. Nowadays transposons are viewed as being one of the main drivers behind eukaryotic evolution: For example, 50% percent of the human genome and as much as 85% of the maize genome consists of these elements and they have been implicated in numerous ways of shaping the genomic landscape [8, 9]. A generalized view of the structure of a transposable element would be a stretch of genomic DNA, often delimited by flanking repeats and/or target site duplication (TSD). Autonomous elements contain genes between their flanks which are necessary to transposition; non-autonomous elements do not have these (intact) genes, and rely on the factors of other elements to be mobilized. Both classes of transposons described below encompass autonomous and non-autonomous elements. In the following, the "classical" classes of eukaryotic transposons are discussed; several types of elements have been found to date that are not mentioned here and a more elaborate classification of these elements is more common today [10]. However, discussing all different types of transposons falls beyond the scope of this introduction.

Types of transposable elements

Class I elements; retrotransposons

Transposable elements (TE's) from Class I transpose in a copy-paste fashion by creating a RNA intermediate (Figure 1A and 1B). The entire element is transcribed into mRNA, after which a reverse transcriptase generates a DNA copy, which is subsequently inserted in the genome.

a. *LTR Retrotransposons*

These transposons are characterized by Long Terminal direct Repeats (LTR's), which range from a couple of hundred base pairs to an exceptional 25kb [11]. At least two proteins are encoded by autonomous elements: GAG and POL. The *gag* gene encodes a capsid-like protein, whereas *pol* encodes a versatile protein responsible for protease, reverse transcriptase, RNAse H and integrase activities (Figure 1A) [12]. LTR elements may also encode an *env* gene, encoding an envelope protein. Retroviruses contain many of the same proteins (e.g. envelope and capsid proteins), which shows their close relatedness to retroelements. In plant genomes LTR elements can be very abundant. In *Zea mays* the *copia*-like element *Opie-2* has around 100,000 copies in the genome and the gypsy-like element *Huck-2* is present in about double that number of copies

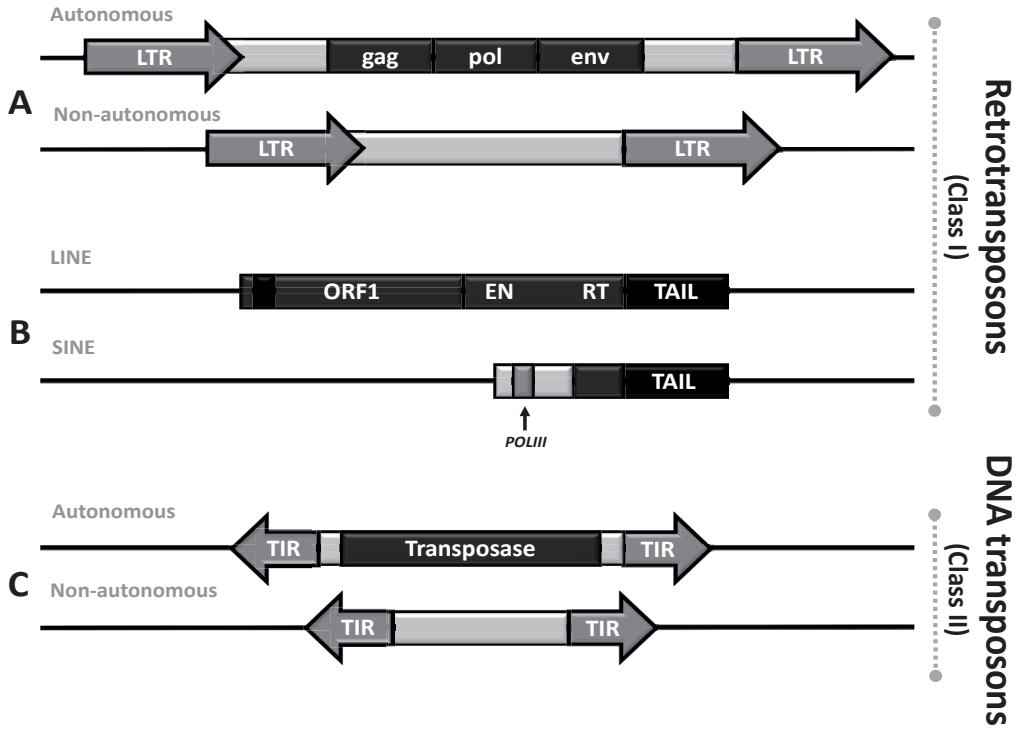


Figure 1. Retrotransposons (Class I) and DNA transposons (Class II) transposable elements, adapted Feschotte *et al.* (2002) [13]. (**A**) Autonomous and non-autonomous class I elements (retrotransposons). Elements are denominated by Long Terminal Repeats (LTR). Autonomous elements encode *GAG*, *POL* and *ENV* proteins that enable the transposition process. (**B**) Non-LTR Retrotransposons. The autonomous elements (Long Interspersed Nuclear Elements, (LINEs)) encode an endonuclease (EN), a reverse transcriptase (RT) and a *GAG*-like protein (ORF1). Short interspersed nuclear elements (SINEs) are non-autonomous and generally possess an internal *POLIII* RNA polymerase promoter. Both SINEs and LINEs elements possess a simple repeat terminus, usually poly(A). (**C**) DNA transposons are characterized by Terminal Inverted Repeats (TIR) and autonomous elements encode a transposase.

(~200,000) [13]. In the much more compact Arabidopsis genome the occurrence of these elements is in the range of tens of copies, but many different LTR families have been shown to be present [14].

b. *Non-LTR Retrotransposons*

As the name aptly indicates, these elements are not flanked by long repeats (Figure 1B). These elements are divided in long and short interspersed nuclear repeats; LINEs

and SINEs respectively. LINEs minimally encode a nuclease and reverse transcriptase. Even though these elements generally create a TSD upon insertion, their sequence is often truncated at the 5'end, hampering their identification in a genome [10, 15]. At their 3' terminus these elements have a poly-AAA tail, a tandem repeat or merely an A-rich region [10]. LINE elements encompass about 20% of the human genome and are found in every kingdom. In plants these elements are usually much less frequently found than LTR elements. Unlike most non-autonomous LTR elements, SINEs are not simply cousins of their autonomous relatives that lost coding capacity. SINEs are between 80-500 base pairs long and do not originate from LINEs. Instead they are retrocopies of polymerase III transcripts [16]. They possess an internal promoter and rely on LINEs to transpose. The 3' termini of these elements are obscure, but can resemble the features of LINEs or consist of the POLIII termination signal.

Class II elements; DNA transposons

Class II elements are generally not very abundant in genomes, but are present in low to moderate numbers; nevertheless they are found in almost all eukaryotes [17]. Instead of reverse transcribing RNA, which is reinserted in the genome, these elements encode a protein which facilitates their excision and insertion in the genome. Two subclasses can be distinguished, based on the number of DNA strands that are cut during the transposition process.

Subclass II contains the so-called rolling circle transposons, which transpose in a replicative manner after cutting a single DNA strand. Since these elements have structural homologies to genes encoding Rep/helicase-like and replication protein A (RPA), it is suggested that they transpose by a rolling circle mechanism [18]. Elements such as Helitrons and Maverick (not found in plants) are subclass II elements [19, 20]. Helitrons have been found frequently in maize and are best described in this organism [21–23]. Subclass II elements will not be further discussed here (for a review on Helitrons see: Kapitonov and Jurka (2007) [20]).

Subclass I consists of several super families of TE's, such as the widespread hAT (hobo, Ac, Tam3), Mutator, CACTA, Harbinger and Mariner elements, but also PiggyBac, which is now widely used as a molecular tool [24]. These elements transpose in a cut-paste fashion. Active Subclass I elements encode a transposase, which recognizes the terminal inverted repeat (TIR) of the element and cuts both strands at both 5'and 3'ends of the element (Figure 1C). The typical length of the TIRs of Class II elements ranges from 2

to 11 base-pairs, although the rare Crypton element, which is found only in fungi, has none [25]. Only Harbinger and CACTA elements have a second ORF, which is necessary for transposition [26–28]. If Class II elements could only move through the genome by a cut-paste fashion, they would have disappeared quickly. These elements can utilize the process of replication to increase their numbers [29]. The elements jump from DNA that has already replicated to a genomic location that has not been replicated yet. Another mode of replicative transposition is by exploiting the gap-repair mechanism of the double strand break in the genome that is formed after excision, thus recreating a copy at the donor site, by using the homologous chromosome or the sister chromatid as a template. This process occurs after an element has transposed and the resulting gap is repaired using the sister chromatid as a repair template, giving rise to a duplication [30].

hAT (*hobo*, Ac and Tam3) transposons are one of the most abundant superfamilies of the class II elements [9]. They can be found in most branches of the tree of life [31]. The Ac element in maize, the first identified mobile element, is a hAT transposon. The mode of transposition of these elements is exemplary for class II elements described above. The transposase of hAT elements contains a BED-zinc finger DNA binding domain at its N-terminus and a dimerization domain (hATC-domain) at its C-terminus [31, 32]. These elements have been found to contain conserved blocks named hAT blocks A to F, of which the latter three overlap with the dimerization domain [31, 33]. A typical hAT transposase is encoded by an approximately 2kb coding sequence [31]. Also, the dimerization domain is suggested to play a role in the transposition process [34, 35].

The influence of TE's on the genomic landscape

The influence of TE's on the genome is vast on almost every level conceivable. Transposons can give rise to duplications, deletions, translocations and inversions of large stretches of DNA or even complete arms of chromosomes [36]. On a smaller scale, one can envisage transposons having an influence on individual genes. Transposons can jump into the coding sequence of a gene, truncating the gene, or adding coding sequence. It is now widely recognized that many common protein domains originate from transposable elements (reviewed in Feschotte (2008) [3]). New splicing variants can arise as a result of a transposon insertion, giving rise to alternative transcripts. Insertion of transposable elements in a promoter region can enhance or reduce gene expression levels or patterns. Many regulatory elements of promoters have been introduced by TE insertion. The human p53 factor has been found to have many binding sites in the

human genome, 30% of these sites consist of insertions of long terminal repeat (LTR) elements [37]. By disturbing boundary elements, or introducing silencing triggers in the vicinity of genes, silencing, or reduced silencing can occur, through altering the epigenetic state of a genomic domain. Small interfering RNA's (siRNA's) can cause silencing of the sequence they target through altering the epigenetic state (reviewed in Kanno and Habu (2011) [38]). Transposons are often targets of siRNA's (reviewed in Feschotte and Prittham (2007) [17]). This integral property of transposons has an effect on the genomic landscape adjacent to the element, because its epigenetic state can be influenced by the nearby TE. Transposons are therefore not the genetic parasites they were thought to be but have a highly dynamic relationship with their hosts [39].

Domestication

The genes encoded by transposons can be adopted partly or entirely by the host in a process called "domestication". Many examples exist of molecular domestication, which can be found in many species [4]. Arguably the most famous domestication event is the domestication of the ancestral *Transib* DNA-transposon giving rise to the *RAG1* and *RAG2* genes in humans [40]. These genes are responsible for regulating antibody variability, by controlling a process called VDJ-recombination [40]. In plants, proteins encoded by the *FHY3*/*FAR1*-family are an example of domesticated transposases. This gene family is derived from *Mutator* DNA-transposons and plays a transcriptional role in the Phytochrome A signaling pathway for far-red light sensing [41]. Another example is *gary*, which is present in cereal grass genomes [42]. This domesticated transposase is derived from a hAT element and is actively transcribed, although no functionality has been attributed to it yet [42]. Since domestication is essentially a recycling of transposon coding material, domesticated transposases are often involved in processes that make use of domains that were present in the source element. For example: transposases of many different TE families can bind DNA using a common protein domain called "zinc fingers". Many domesticated transposases also bind DNA in their acquired role of transcription factor, chromatin modifier, recombination regulator, insulator, or DNA-replication regulator [4]. The described domesticated genes, *FHY3*/*FAR1* and *RAG1* and 2 are genes that have arisen by adopting the full length of their ancestral transposase, but many domesticated genes are chimaeras: they consist of transposase elements fused to elements derived from other protein families. An example of such a protein is the angiosperm-specific MUSTANG1 [43]. This gene has a transposon derived zinc finger and WRKY/GCM1 domains for DNA binding, but also contains a PB1-domain,

which is normally found in cytoplasmic signaling proteins [44]. Interestingly, relatively few transposase derived genes have been identified in plants compared to the number identified in animal genomes, although the number of transposons in plants is similar or even higher compared to animals [4, 9]. In the human genome, DNA transposons account for ~3% of the genome [9]. A similar percentage (2-3%) of such elements is found in *Arabidopsis thaliana,* while ~8.6% of the maize genome consists of these elements [8, 45].

Retrogenes

Retroelements can cause single duplications of coding genes. These duplications are caused by a promiscuous reverse-transcriptase encoded by the transposable element and may occur when mRNA of a gene is present simultaneously with a reverse-transcriptase. The mRNA is translated into cDNA, which is subsequently inserted in the genome, giving rise to an intronless copy of the gene, transcribed from the source mRNA. These copies are called retrocopies and are ubiquitously present in many genomes [46–48]. These retrocopies are often non-functional, since they have lost their original genomic context. However, retrocopies can acquire new regulatory elements and become actively transcribed functional genes, which are named retrogenes [49–52]. Regulatory elements are often obtained by the insertion of a transposable element upstream of the initial retrocopy [51]. Many retrogenes have been identified to date. The human genome is estimated to contain ~1000 actively transcribed retrocopies, of which ~10% are probably active [53]. In the Arabidopsis genome retrogenes were identified in the *ASK1 (Arabidopsis SKP1*-like*)* gene family [50]. *ASK1-like* genes code for proteins that can be part of an SCF (SKP1-CULLIN1-F-box protein)-complex, which is involved in ubiquitylation of proteins, targeting them for degradation [54]. This gene family has 19 members, of which several have been found to be retrogenes that are expressed [50]. The fact that the encoded proteins are able to interact with F-box proteins and seem to have evolved under purifying selection, suggests that they have acquired functionality [50, 55–57].

Functional context of *DAYSLEEPER*

In the following paragraphs, protein complexes and their function will be discussed that form the context in which *DAYSLEEPER* functions. In this thesis I hope to convince you that DAYSLEEPER interacts with these fundamental protein complexes in the cell and modulates their functions. These complexes include the COP9-signalosome (CSN)

and cullin-RING ligases (CRL's), both part of the ubiquitin proteasome pathway and the ESCRT-III machinery.

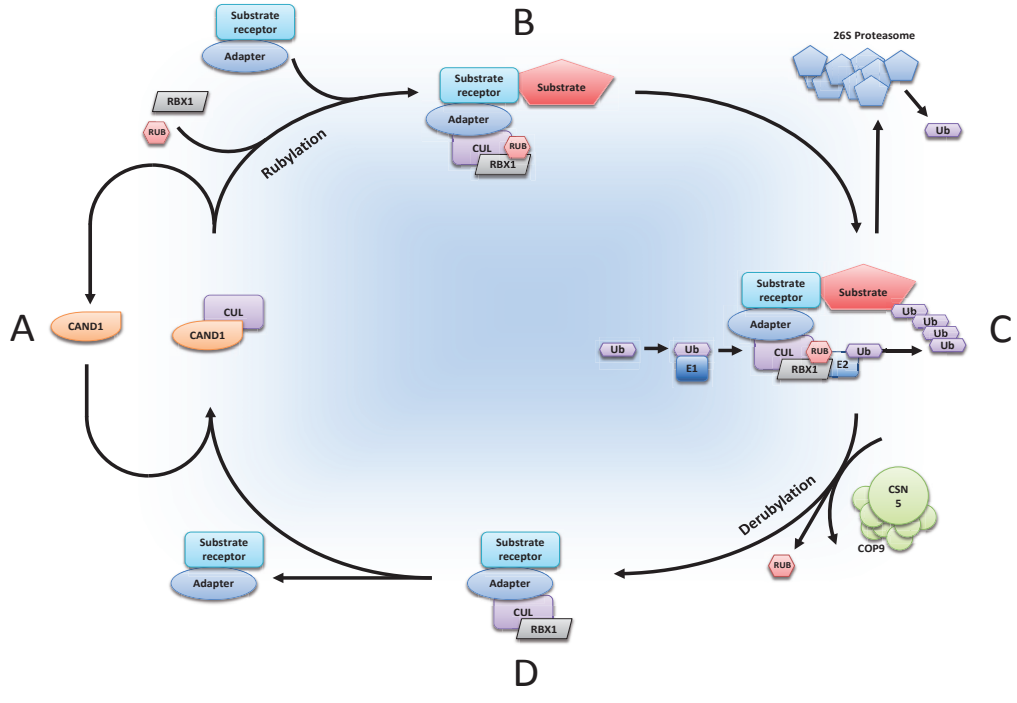


Figure 2. The CRL-mediated proteasome pathway (based on Schwechheimer *et al.* (2010) [76]). CRL-complexes poly-ubiquitylate proteins, which are thereby targeted for degradation. (**A**) CAND1 binds CULLINs and thereby blocks the formation of a CRL complex and the rubylation of the CULLIN subunit [94–96]. (**B**) When CULLINs are not bound to CAND1, CULLINs bind the protein RBX [97]. The RUB-group, attached to the CULLIN, and RBX interact with an UBQ conjugating enzyme (E2). (**C**) This complex will subsequently deposit the UBQ modification on a substrate protein [98, 99]. Polyubiquitylated proteins are targeted for degradation by the 26S proteasome [100]. The E2 enzyme receives UBQ from an UBQ activating enzyme (E1) [59]. CULLIN also binds adaptor and receptor proteins, determining the specificity of the CRL complex. The adaptor, receptor and substrate proteins are CULLIN specific. These complexes are discussed in Hua *et al.* (2011) [61]. The substrate protein completes the structure of the complex. (**D**) The CRL can subsequently be derubylated by the CSN5 subunits of the CSN [61, 72, 73], which inactivates the complex and frees CULLIN to be bound to CAND1 once again, completing the cycle (Reviewed in Hotton *et al.* (2008) [73] and Hua *et al.* (2011) [61]).

The ubiquitin proteasome pathway: Cullins

The ubiquitin proteasome pathway is responsible for the covalent modification of proteins by the addition of ubiquitin (UBQ) (Figure 2) [58]. Ubiquitin is a 76 amino-acid protein that is conserved from yeast to humans and is widely used as a modifier of

other proteins. Once a protein is mono-ubiquitylated, it can become poly-ubiquitylated by repetition of the catalytic pathway [59]. Poly-ubiquitylation targets a protein to be degraded by the 26S proteasome; mono-ubiquitylation, however, does not [60]. The ubiquitylation process is performed by an UBQ activating enzyme, an UBQ conjugating enzyme and a UBQ ligase, named E1, E2 and E3, respectively. The E3 UBQ ligase can be either a monomeric protein or a multimeric complex that transfers UBQ from the E2 to the substrate [58]. An important class of multimeric E3 ligases in plants possesses a cullin-protein (CUL) as scaffolding subunit. In Arabidopsis, 3 types of cullins have been described, namely CUL1, CUL3 and CUL4 [61]. Complexes with CUL2 and CUL5 have not been reported in plants until now [61], but a complex that is structurally similar is present in the form of the Anaphase-Promoting Complex/Cyclosome (APC/C), which is involved in DNA duplication and cell division [62]. Cullin-RING ligase (CRL) complexes are involved in the regulation of numerous processes in the plant, which include hormone and light perception, regulation of the cell-cycle and transcription and response to biotic and abiotic challenges [58, 61, 63]. Cullins are able to bind substrate adaptor proteins with their N-terminus, whereas their C-terminus can bind the RING-finger protein RBX1, and an RBX1-associated E2 protein [64]. In the SCF (for SKP1-like-CUL1-F-Box)-complex, CUL1 binds an ASK (Arabidopsis Skp1-like protein) and an F-Box protein, which in turn offers substrate specificity. CUL3 binds a BTB/POZ-protein for substrate specificity in a complex known as BCR (for BTB-domain protein, CUL3 and a RING protein) [65]. CUL4 binds the DAMAGED DNA BINDING1 (DDB1) and a variable substrate receptor (X-box protein) in a protein complex referred to as DCX (for DDB1-CUL4A-X-box) [66]. Figure 2 depicts the ubiquitylation process performed by a generalized ubiquitin ligase complex.

The APC/C is based on the distant CUL relative APC2 and the RING finger protein APC11, but consists of 12 subunits (13 in yeast) in total, making it the largest E3 UBQ ligase described thus far [62, 67]. The APC/C has been studied extensively for regulating cell-cycle transitions and sister chromatid separation. It promotes progression and exit from mitosis by mediating proteolysis of different cell-cycle regulators, including PDS1 and CYCLIN B (reviewed in Pesin *et al.* (2008) [68] and Marrocco *et al.* (2009) [69]).

The ubiquitin proteasome pathway: COP9-signalosome

The COP9-signalosome (CSN) was first identified in Arabidopsis, but is a highly conserved protein complex in eukaryotes [70, 71]. The CSN is a multi-protein complex in the ubiquitin-proteasome pathway (Figure 2). The complex is typically comprised of 8 CSN subunits and interacts with CRL's. The CSN modulates the function of the

ubiquitin ligases by derubylation/deneddylation of the cullin protein in plants (Figure 2). The CSN5 subunits of the CSN-complex mediate the derubylation [72]. Rubylation and derubylation control the assembly and disassembly of CRL's (Figure 2) [61, 72, 73], allowing recycling of the subunits of the complex. It is widely believed that the subunits of the CSN-complex are evolutionary derived from subunits of an ancestral 26S proteasome, because of conserved structural similarities between the 26S proteasome and the CSN-complex [71]. Deletion of any one of the subunits of the CSN in Arabidopsis leads to the same seedling growth arrest phenotype, which correlates with an accumulation of rubylated cullins [74, 75]. Interestingly, it has been shown that *csn* mutants accumulate DNA damage [74]. It is as of yet unknown what causes this accumulation and what the precise nature of the DNA damage is [74, 76].

The Endosomal Sorting Complex Required for Transport-III (ESCRT-III) complex

The Endosomal Sorting Complex Required for Transport-III (ESCRT-III) complex is one of 4 ESCRT complexes, namely -I, -II, -III and- 0. ESCRT-0 is only present in *Opisthokonta* (i.e. animals, fungi and a number of protists), whereas the other three are conserved in all eukaryotes [77, 78]. These complexes work sequentially and in conjunction to sort mono-ubiquitylated cargo-proteins into the endosomal vesicles of multi-vesicular bodies (MVB's) [79, 80]. The ESCRT-machinery has also been shown to play a role in cytokinesis and to be recruited by several viruses, including HIV (reviewed in [81, 82]). However, the ESCRT-III subunit Snf7p in yeast has also been reported to influence gene expression of several genes in anaerobic conditions [83], indicating that the role of ESCRT-complex subunits might extend beyond the classical roles of these complexes. ESCRT-0, -I and -II are cytosolic complexes that are transiently recruited to endosome membranes, whereas ESCRT-III is formed on the endosomal membrane [79, 84]. The ESCRT-III machinery consists of VPS2, SNF7, VPS20 and VPS24, which together facilitate the budding of vesicles from the membrane [85–87]. A schematic overview of the assembly and disassembly of the ESCRT-III subunits is depicted in Figure 3. In Arabidopsis, for all subunits of the canonical ESCRT-III complex, two homologous genes are present, with the exception of VPS2 which is encoded by three homologous genes [87]. Until 10 years ago the ESCRT-III machinery was hardly studied in Arabidopsis, but in the last decade this has changed, although much is still unknown. It has been suggested that the ESCRT-III machinery in Arabidopsis functions in conjunction with VPS40, VPS60 and dynamins [88], since these proteins were found to interact with VPS2.2. Members of the ESCRT-machinery were initially identified as chromatin-modifying proteins (CHMP),

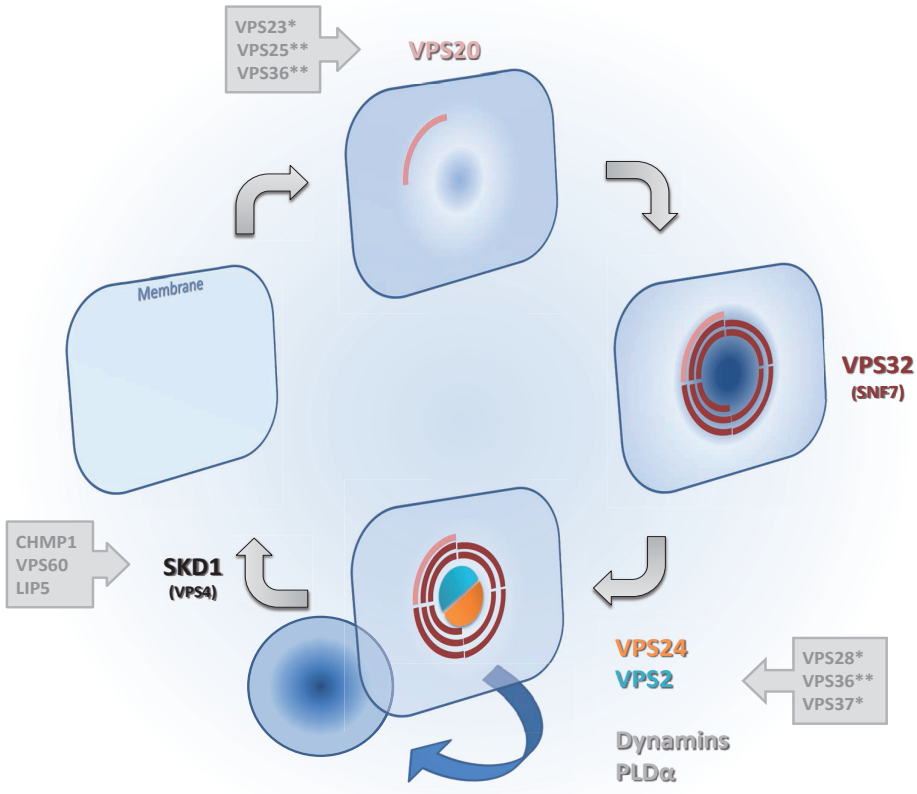


Figure 3. ESCRT-III machinery in Arabidopsis. The assembly and disassembly of the ESCRT-III machinery based on Wollert *et al.* (2009) [101]. VPS20 accumulates on the membrane, followed by polymers of VPS32 (SNF7). Finally, VPS2 and VPS24 complete the complex, snaring off a vesicle. SKD1 (VPS4) initiates the disassembly of the complex. ESCRT-III subunits have been shown to interact with each other [85], but have also been shown to interact with subunits of other ESCRT-complexes. These interactions can be found in the grey boxes and were reported by Shahriari *et al.* (2010)/(2011) [85, 91] and Ibl *et al.* (2012) [88]. VPS23, VPS28 and VPS37 are ESCRT-I proteins (*****), whereas VPS25 and VPS36 belong to the ESCRT-II complex (******) [85, 91]. CHMP1, LIP5 and VPS60 are proteins that regulate the association of SKD1 to endosomes [77, 102–104]. Arabidopsis contains 2 *CHMP1* genes, one *LIP5* gene and several putative *VPS60* genes [85, 105]. The three *VPS2 A. thaliana* genes encode slightly different proteins, that all interact with different ESCRT-complex subunits [85]. Ibl *et al.* (2012) identified many interaction partners of VPS2.2, among which several dynamin proteins and PLDα1 [88]. Ibl *et al.* (2012) suggest a model in which PLDα1 and dynamins supports ESCRT-III mediated membrane narrowing [88].

which was later changed into “Charged Multivesicular body Proteins”. CHMP1 and CHMP2 proteins in mammalian cells have both been linked to chromatin remodeling [89, 90]. Interestingly VPS2.2 was found to bind histone protein HTA6 and several RNA-binding proteins, which suggests that VPS2.2 is also involved in transcriptional control, or at least has nuclear functions [88]. VPS4 homolog *AtSKD1* has been found to be

important for vacuolar maintenance, but might also play a role in the regulation of the cell-cycle [91]. It was shown that VPS2.1 is essential for plant development; VPS2.2 and VPS2.3 mutants display a short root phenotype and seem to have a disorganized cellular structure in their root, likely caused by problems with cytokinesis [88, 92, 93].

Synopsis

In this thesis, the evolutionary background of *DAYSLEEPER* will be investigated as well as its expression pattern and the localization of the DAYSLEEPER protein. The function of *DAYSLEEPER* has been investigated by identifying interaction partners and studying its potential functionality in relation to the ESCRT-III and CRL and CSN complexes. Finally a transcriptomics approach was used to identify potential targets of *DAYSLEEPER* and identify the functional implications of *DAYSLEEPER* overexpression using a high-throughput method.

Experimental Chapters

Chapter 2: The SLEEPER genes: A transposase-derived angiosperm-specific gene family

This chapter contains an analysis of the evolutionary backgrounds of *DAYSLEEPER* and proposes a model for its domestication. Furthermore, functionally conserved orthologous genes from several different species are described and conserved features of the *SLEEPER* protein family designated. It was found that *SLEEPER*s from rice and grapevine cause a delayed growth and flowering phenotype when expressed in Arabidopsis. We found that *SLEEPER*-genes are conserved in angiosperms, but not in gymnosperms and beyond.

Chapter 3: *DAYSLEEPER* is expressed predominantly in meristems and is both nuclear and vesicular localized

This chapter describes the expression pattern of *DAYSLEEPER* using *DAYSLEEPER*-promoter::*gusA* fusions and Q-RT-PCR, both in mature plants, as well as during development. The cellular localization of DAYSLEEPER is studied using confocal microscopy, in protoplasts, as well as *in planta*. We found that DAYSLEEPER resides in the nucleus as well as in vesicular structures *in planta* and in protoplasts. Using marker-constructs, we determined DAYSLEEPER-positive vesicular structures to be constituents

of the trans-golgi network, multivesicular bodies (MVB's) and late endosomes. It is also shown that DAYSLEEPER, like hAT-transposases, has a functional dimerization domain.

Chapter 4: The influence of *DAYSLEEPER* on the *Arabidopsis thaliana* transcriptome

In this chapter the results of a Serial Analysis of Gene Expression (SAGE) experiment coupled to Illumina GAIIx ® sequencing are discussed. Plants that have an inducible *DAYSLEEPER* gene were used to study the effects of *DAYSLEEPER* overexpression on the *A. thaliana* transcriptome. After 24 hours of *DAYSLEEPER* induction, many genes in various functional fields were differentially expressed, compared to control samples. We found an overrepresentation of genes involved in ubiquitin-related processes, RNA-metabolism and RNA-meditated silencing genes and plant defense, but also found genes involved in regulating the timing of flowering, circadian rhythm, chromatin modifiers and light/dark-response related genes. We hypothesize that DAYSLEEPER plays a role in the regulation of COP9-signalosome and cullin-RING-ligase activities, thereby modulating protein ubiquitylation in a broad sense.

Chapter 5: The DAYSLEEPER interactome

This chapter describes the *in vivo* interaction of DAYSLEEPER with other proteins that were previously identified in a yeast two-hybrid screen, by Bi-molecular Fluorescence Complementation (BiFC) in Arabidopsis protoplasts. We found that DAYSLEEPER homodimerizes and interacts with the *NPH3*-family protein NRL8, the COP9-signalosome subunit CSN5A, the exosome subunit RRP6A and the 3 homologs of the ESCRT-III machinery subunit VPS2. We found that *daysleeper* mutant plants have aberrant formation of cell-layers in the root-tip, have disturbed auxin distribution and display problems maintaining the root meristems. We also found that like *csn*-mutants, *daysleeper* mutants accumulate DNA-damage.

References

1. Bundock P, Hooykaas P: **An Arabidopsis hAT-like transposase is essential for plant development.** *Nature* 2005, **436**:282–4.

2. Volff J-N: **Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes.** *Bioessays* 2006, **28**:913–22.

3. Feschotte C: **Transposable elements and the evolution of regulatory networks.** *Nat. Rev. Genet.* 2008, **9**:397–405.

4. Sinzelle L, Izsvák Z, Ivics Z: **Molecular domestication of transposable elements: from detrimental parasites to useful host genes.** *Cell. Mol. Life Sci.* 2009, **66**:1073–93.

5. McClintock B: **The origin and behavior of mutable loci in maize.** *Proc. Natl. Acad. Sci. U.S.A.* 1950, **36**:344–55.

6. Faulkner GJ, Carninci P: **Altruistic functions for selfish DNA.** *Cell Cycle* 2009, **8**:2895–900.

7. Orgel LE, Crick FH: **Selfish DNA: the ultimate parasite.** *Nature* 1980, **284**:604–7.

8. Schnable PS, Ware D, Fulton RS, et al.: **The B73 maize genome: complexity, diversity, and dynamics.** *Science* 2009, **326**:1112–5.

9. Lander ES, Linton LM, Birren B, et al.: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860–921.

10. Seberg O, Petersen G: **A unified classification system for eukaryotic transposable elements should reflect their phylogeny.** *Nat. Rev. Genet.* 2009, **10**:276.

11. Neumann P, Pozárková D, Macas J: **Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced.** *Plant Mol. Biol.* 2003, **53**:399–410.

12. Sabot F, Schulman a H: **Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome.** *Heredity* 2006, **97**:381–8.

13. Feschotte C, Jiang N, Wessler SR: **Plant transposable elements: where genetics meets genomics.** *Nat. Rev. Genet.* 2002, **3**:329–41.

14. Le QH, Wright S, Yu Z, Bureau T: **Transposon diversity in Arabidopsis thaliana.** *Proc. Natl. Acad. Sci. U.S.A.* 2000, **97**:7376–81.

15. Biedler J, Tu Z: **Non-LTR retrotransposons in the African malaria mosquito, Anopheles gambiae: unprecedented diversity and evidence of recent activity.** *Mol. Biol. Evol.* 2003, **20**:1811–25.

16. Goodier JL, Kazazian HH: **Retrotransposons revisited: the restraint and rehabilitation of parasites.** *Cell* 2008, **135**:23–35.

17. Feschotte C, Pritham EJ, Cédric Feschotte and Ellen J. Pritham: **DNA transposons and the evolution of eukaryotic genomes.** *Annu. Rev. Genet.* 2007, **41**:331–68.

18. Yang L, Bennetzen JL: **Structure-based discovery and description of plant and animal Helitrons.** *Proc. Natl. Acad. Sci. U.S.A.* 2009, **106**:12832–7.

19. Pritham EJ, Putliwala T, Feschotte C: **Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses.** *Gene* 2007, **390**:3–17.

20. Kapitonov VV, Jurka J: **Helitrons on a roll: eukaryotic rolling-circle transposons.** *Trends Genet.* 2007, **23**:521–9.

21. Lal SK, Giroux MJ, Brendel V, Vallejos CE, Hannah LC: **The maize genome contains a helitron insertion.** *Plant Cell* 2003, **15**:381–91.

22. Feschotte C, Pritham EJ: **A cornucopia of Helitrons shapes the maize genome.** *Proc. Natl. Acad. Sci. U.S.A.* 2009, **106**:19747–8.

23. Yang L, Bennetzen JL: **Distribution, diversity, evolution, and survival of Helitrons in the maize genome.** *Proc. Natl. Acad. Sci. U.S.A.* 2009, **106**:19922–7.

24. VandenDriessche T, Ivics Z, Izsvák Z, Chuah MKL: **Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells.** *Blood* 2009, **114**:1461–8.

25. Goodwin TJD, Butler MI, Poulter RTM: **Cryptons: a group of tyrosine-recombinase-encoding DNA**

transposons from pathogenic fungi. *Microbiology (Reading, Engl.)* 2003, **149**:3099–109.

26. Sinzelle L, Kapitonov VV, Grzela DP, Jursch T, Jurka J, Izsvák Z, Ivics Z: **Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes.** *Proc. Natl. Acad. Sci. U.S.A.* 2008, **105**:4715–20.

27. Yang G, Zhang F, Hancock CN, Wessler SR: **Transposition of the rice miniature inverted repeat transposable element mPing in Arabidopsis thaliana.** *Proc. Natl. Acad. Sci. U.S.A.* 2007, **104**:10962–7.

28. Frey M, Reinecke J, Grant S, Saedler H, Gierl A: **Excision of the En/Spm transposable element of Zea mays requires two element-encoded proteins.** *EMBO J.* 1990, **9**:4037–44.

29. Greenblatt IM, Brink RA: **Twin Mutations in Medium Variegated Pericarp Maize.** *Genetics* 1962, **47**:489–501.

30. Nassif N, Penney J, Pal S, Engels WR, Gloor GB: **Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair.** *Mol. Cell. Biol.* 1994, **14**:1613–25.

31. Rubin E, Lithwick G, Levy a a: **Structure and evolution of the hAT transposon superfamily.** *Genetics* 2001, **158**:949–57.

32. Kempken F, Windhofer F: **The hAT family: a versatile transposon group common to plants, fungi, animals, and man**. *Chromosoma* 2001, **110**:1–9.

33. Yamashita D, Komori H, Higuchi Y, Yamaguchi T, Osumi T, Hirose F: **Human DNA replication-related element binding factor (hDREF) self-association via hATC domain is necessary for its nuclear accumulation and DNA binding.** *J. Biol. Chem.* 2007, **282**:7563–75.

34. Michel K: **The C-terminus of the Hermes transposase contains a protein multimerization domain**. *Insect Biochemistry and Molecular Biology* 2003, **33**:959–970.

35. Essers L, Adolphs RH, Kunze R: **A highly conserved domain of the maize activator transposase is involved in dimerization.** *Plant Cell* 2000, **12**:211–24.

36. Weil CF: **Too many ends: aberrant transposition.** *Genes Dev.* 2009, **23**:1032–6.

37. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D: **Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53.** *Proc. Natl. Acad. Sci. U.S.A.* 2007, **104**:18613–8.

38. Kanno T, Habu Y: **siRNA-mediated chromatin maintenance and its function in Arabidopsis thaliana.** *Biochim. Biophys. Acta* 2011, **1809**:444–451.

39. Levin HL, Moran JV: **Dynamic interactions between transposable elements and their hosts**. *Nature Reviews Genetics* 2011, **12**:615–627.

40. Kapitonov VV, Jurka J: **RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons.** *PLoS Biol.* 2005, **3**:e181.

41. Hudson ME, Lisch DR, Quail PH: **The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway.** *Plant J.* 2003, **34**:453–71.

42. Muehlbauer GJ, Bhau BS, Syed NH, Heinen S, Cho S, Marshall D, Pateyron S, Buisine N, Chalhoub B, Flavell AJ: **A hAT superfamily transposase recruited by the cereal grass genome.** *Mol. Genet. Genomics* 2006, **275**:553–63.

43. Cowan RK, Hoen DR, Schoen DJ, Bureau TE: **MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms.** *Mol. Biol. Evol.* 2005, **22**:2084–9.

44. Terasawa H, Noda Y, Ito T, Hatanaka H, Ichikawa S, Ogura K, Sumimoto H, Inagaki F: **Structure and ligand recognition of the PB1 domain: a novel protein module binding to the PC motif.** *EMBO J.* 2001, **20**:3947–56.

45. Zhang X, Wessler SR: **Genome-wide comparative analysis of the transposable elements in the related species Arabidopsis thaliana and Brassica oleracea.** *Proc. Natl. Acad. Sci. U.S.A.* 2004, **101**:5589–94.

46. Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, Liu D, Zhang J, Vang S, Lu Z, Wong GK-S, Long M, Wang J: **High rate of chimeric gene origination by retroposition in plant genomes.** *Plant Cell* 2006, **18**:1791–802.

47. Baertsch R, Diekhans M, Kent WJ, Haussler D, Brosius J: **Retrocopy contributions to the evolution of the human genome.** *BMC Genomics* 2008, **9**:466.

48. Fu B, Chen M, Zou M, Long M, He S: **The rapid generation of chimerical genes expanding protein diversity in zebrafish.** *BMC Genomics* 2010, **11**:657.

49. Pace JK, Feschotte C: **The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage.** *Genome Res.* 2007, **17**:422–32.

50. Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, DePamphilis CW: **Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth.** *Plant J.* 2007, **50**:873–85.

51. Vaknin K, Goren A, Ast G: **TEs or not TEs? That is the evolutionary question.** *J. Biol.* 2009, **8**:83.

52. Huang C-J, Lin W-Y, Chang C-M, Choo K-B: **Transcription of the rat testis-specific Rtdpoz-T1 and -T2 retrogenes during embryo development: co-transcription and frequent exonisation of transposable element sequences.** *BMC Mol. Biol.* 2009, **10**:74.

53. Vinckenbosch N, Dupanloup I, Kaessmann H: **Evolutionary fate of retroposed gene copies in the human genome.** *Proc. Natl. Acad. Sci. U.S.A.* 2006, **103**:3220–5.

54. Willems AR, Schwab M, Tyers M: **A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin.** *Biochim. Biophys. Acta* 2004, **1695**:133–70.

55. Gagne JM, Downes BP, Shiu S-H, Durski AM, Vierstra RD: **The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis.** *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**:11519–24.

56. Risseeuw EP, Daskalchuk TE, Banks TW, Liu E, Cotelesage J, Hellmann H, Estelle M, Somers DE, Crosby WL: **Protein interaction analysis of SCF ubiquitin E3 ligase subunits from Arabidopsis.** *Plant J.* 2003, **34**:753–67.

57. Takahashi N, Kuroda H, Kuromori T, Hirayama T, Seki M, Shinozaki K, Shimada H, Matsui M: **Expression and interaction analysis of Arabidopsis Skp1-related genes.** *Plant Cell Physiol.* 2004, **45**:83–91.

58. Smalle J, Vierstra RD: **The ubiquitin 26S proteasome proteolytic pathway.** *Annu Rev Plant Biol* 2004, **55**:555–90.

59. Pickart CM, Cohen RE: **Proteasomes and their kin: proteases in the machine age.** *Nat. Rev. Mol. Cell Biol.* 2004, **5**:177–87.

60. Huang TT, D'Andrea AD: **Regulation of DNA repair by ubiquitylation.** *Nat. Rev. Mol. Cell Biol.* 2006, **7**:323–34.

61. Hua Z, Vierstra RD: **The cullin-RING ubiquitin-protein ligases.** *Annu Rev Plant Biol* 2011, **62**:299–334.

62. Peters J-M: **The anaphase-promoting complex: proteolysis in mitosis and beyond.** *Mol. Cell* 2002, **9**:931–43.

63. Lechner E, Achard P, Vansiri A, Potuschak T, Genschik P: **F-box proteins everywhere.** *Curr. Opin. Plant Biol.* 2006, **9**:631–8.

64. Gray WM: **Role of the Arabidopsis RING-H2 Protein RBX1 in RUB Modification and SCF Function**. *THE PLANT CELL ONLINE* 2002, **14**:2137–2144.

65. Xu L, Wei Y, Reboul J, Vaglio P, Shin T-H, Vidal M, Elledge SJ, Harper JW: **BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3.** *Nature* 2003, **425**:316–21.

66. Wertz IE, O'Rourke KM, Zhang Z, Dornan D, Arnott D, Deshaies RJ, Dixit VM: **Human De-etiolated-1 regulates c-Jun by assembling a CUL4A ubiquitin ligase.** *Science* 2004, **303**:1371–4.

67. Yoon H-J, Feoktistova A, Wolfe B a, Jennings JL, Link AJ, Gould KL: **Proteomics analysis identifies new**

components of the fission and budding yeast anaphase-promoting complexes. *Curr. Biol.* 2002, **12**:2048–54.

68. Pesin J a, Orr-Weaver TL: **Regulation of APC/C activators in mitosis and meiosis.** *Annu. Rev. Cell Dev. Biol.* 2008, **24**:475–99.

69. Marrocco K, Thomann A, Parmentier Y, Genschik P, Criqui MC: **The APC/C E3 ligase remains active in most post-mitotic Arabidopsis cells and is required for proper vasculature development and organization.** *Development* 2009, **136**:1475–85.

70. Wei N, Deng XW: **COP9: a new genetic locus involved in light-regulated development and gene expression in arabidopsis.** *Plant Cell* 1992, **4**:1507–18.

71. Schwechheimer C: **The COP9 signalosome (CSN): an evolutionary conserved proteolysis regulator in eukaryotic development.** *Biochim. Biophys. Acta* 2004, **1695**:45–54.

72. Gusmaroli G, Feng S, Deng XW: **The Arabidopsis CSN5A and CSN5B subunits are present in distinct COP9 signalosome complexes, and mutations in their JAMM domains exhibit differential dominant negative effects on development.** *Plant Cell* 2004, **16**:2984–3001.

73. Hotton SK, Callis J: **Regulation of cullin RING ligases.** *Annu Rev Plant Biol* 2008, **59**:467–89.

74. Dohmann EMN, Levesque MP, De Veylder L, Reichardt I, Jürgens G, Schmid M, Schwechheimer C: **The Arabidopsis COP9 signalosome is essential for G2 phase progression and genomic stability.** *Development* 2008, **135**:2013–22.

75. Gusmaroli G, Figueroa P, Serino G, Deng XW: **Role of the MPN subunits in COP9 signalosome assembly and activity, and their regulatory interaction with Arabidopsis Cullin3-based E3 ligases.** *Plant Cell* 2007, **19**:564–81.

76. Schwechheimer C, Isono E: **The COP9 signalosome and its role in plant development.** *Eur. J. Cell Biol.* 2010, **89**:157–62.

77. Winter V, Hauser M-T: **Exploring the ESCRTing machinery in eukaryotes.** *Trends Plant Sci.* 2006, **11**:115–23.

78. Leung KF, Dacks JB, Field MC: **Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage.** *Traffic* 2008, **9**:1698–716.

79. Babst M, Katzmann DJ, Estepa-Sabal EJ, Meerloo T, Emr SD: **Escrt-III: an endosome-associated heterooligomeric protein complex required for mvb sorting.** *Dev. Cell* 2002, **3**:271–82.

80. Babst M, Katzmann DJ, Snyder WB, Wendland B, Emr SD: **Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting at the multivesicular body.** *Dev. Cell* 2002, **3**:283–9.

81. Martin-Serrano J, Neil SJD: **Host factors involved in retroviral budding and release.** *Nat. Rev. Microbiol.* 2011, **9**:519–31.

82. Neto H, Gould GW: **The regulation of abscission by multi-protein complexes.** *J. Cell. Sci.* 2011, **124**:3199–207.

83. Snoek ISI, Tai SL, Pronk JT, Yde Steensma H, Daran J-M: **Involvement of Snf7p and Rim101p in the transcriptional regulation of TIR1 and other anaerobically upregulated genes in Saccharomyces cerevisiae.** *FEMS Yeast Res.* 2010, **10**:367–84.

84. Wollert T, Hurley JH: **Molecular mechanism of multivesicular body biogenesis by ESCRT complexes.** *Nature* 2010, **464**:864–9.

85. Shahriari M, Richter K, Keshavaiah C, Sabovljevic A, Huelskamp M, Schellmann S: **The Arabidopsis ESCRT protein-protein interaction network.** *Plant Mol. Biol.* 2011, **76**:85–96.

86. Otegui MS, Spitzer C: **Endosomal functions in plants.** *Traffic* 2008, **9**:1589–98.

87. Spitzer C, Schellmann S, Sabovljevic A, Shahriari M, Keshavaiah C, Bechtold N, Herzog M, Müller S, Hanisch F-G, Hülskamp M: **The Arabidopsis elch mutant reveals functions of an ESCRT component in cytokinesis.** *Development* 2006, **133**:4679–89.

88. Ibl V, Csaszar E, Schlager N, Neubert S, Spitzer C, Hauser M-T: **Interactome of the plant-specific ESCRT-III component AtVPS2.2 in Arabidopsis thaliana.** *J. Proteome Res.* 2012, **11**:397–411.

89. Stauffer DR, Howard TL, Nyun T, Hollenberg SM: **CHMP1 is a novel nuclear matrix protein affecting chromatin structure and cell-cycle progression.** *J. Cell. Sci.* 2001, **114**:2383–93.

90. Hodges E, Redelius JS, Wu W, Höög C: **Accelerated discovery of novel protein function in cultured human cells.** *Mol. Cell Proteomics* 2005, **4**:1319–27.

91. Shahriari M, Keshavaiah C, Scheuring D, Sabovljevic A, Pimpl P, Häusler RE, Hülskamp M, Schellmann S: **The AAA-type ATPase AtSKD1 contributes to vacuolar maintenance of Arabidopsis thaliana.** *Plant J.* 2010, **64**:71–85.

92. Katsiarimpa A, Anzenberger F, Schlager N, Neubert S, Hauser M-T, Schwechheimer C, Isono E: **The Arabidopsis Deubiquitinating Enzyme AMSH3 Interacts with ESCRT-III Subunits and Regulates Their Localization.** *Plant Cell* 2011, **23**:3026–3040.

93. Müller S, Fuchs E, Ovecka M, Wysocka-Diller J, Benfey PN, Hauser MT: **Two new loci, PLEIADE and HYADE, implicate organ-specific regulation of cytokinesis in Arabidopsis**. *Plant Physiol.* 2002, **130**:312–324.

94. Goldenberg SJ, Cascio TC, Shumway SD, Garbutt KC, Liu J, Xiong Y, Zheng N: **Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases.** *Cell* 2004, **119**:517–28.

95. Liu J, Furukawa M, Matsumoto T, Xiong Y: **NEDD8 modification of CUL1 dissociates p120(CAND1), an inhibitor of CUL1-SKP1 binding and SCF ligases.** *Mol. Cell* 2002, **10**:1511–8.

96. Zheng J, Yang X, Harrell JM, Ryzhikov S, Shim EH, Lykke-Andersen K, Wei N, Sun H, Kobayashi R, Zhang H: **CAND1 binds to unneddylated CUL1 and regulates the formation of SCF ubiquitin E3 ligase complex.** *Mol. Cell* 2002, **10**:1519–26.

97. Kamura T, Conrad MN, Yan Q, Conaway RC, Conaway JW: **The Rbx1 subunit of SCF and VHL E3 ubiquitin ligase activates Rub1 modification of cullins Cdc53 and Cul2.** *Genes Dev.* 1999, **13**:2928–33.

98. Kawakami T, Chiba T, Suzuki T, Iwai K, Yamanaka K, Minato N, Suzuki H, Shimbara N, Hidaka Y, Osaka F, Omata M, Tanaka K: **NEDD8 recruits E2-ubiquitin to SCF E3 ligase.** *EMBO J.* 2001, **20**:4003–12.

99. Wu K, Chen A, Pan ZQ: **Conjugation of Nedd8 to CUL1 enhances the ability of the ROC1-CUL1 complex to promote ubiquitin polymerization.** *J. Biol. Chem.* 2000, **275**:32317–24.

100. Hershko A, Ciechanover A: **The ubiquitin system.** *Annu. Rev. Biochem.* 1998, **67**:425–79.

101. Wollert T, Wunder C, Lippincott-schwartz J, Hurley JHJH, H J: **Membrane scission by the ESCRT-III complex.** *Nature* 2009, **458**:172–7.

102. Hurley JH, Hanson PI: **Membrane budding and scission by the ESCRT machinery: it's all in the neck**. *Nature Reviews Molecular Cell Biology* 2010, **11**.

103. Rue SM, Mattei S, Saksena S, Emr SD: **Novel Ist1-Did2 complex functions at a late step in multivesicular body sorting**. *Molecular Biology of the Cell* 2008, **19**:475–484.

104. Lata S, Schoehn G, Solomons J, Pires R, Göttlinger HG, Weissenhorn W: **Structure and function of ESCRT-III.** *Biochem. Soc. Trans.* 2009, **37**:156–60.

105. Spitzer C, Reyes FC, Buono R, Sliwinski MK, Haas TJ, Otegui MS: **The ESCRT-related CHMP1A and B proteins mediate multivesicular body sorting of auxin carriers in Arabidopsis and are required for plant development.** *Plant Cell* 2009, **21**:749–66.

1