The Evolution of COP9 Signalosome in Unicellular and Multicellular Organisms

Emanuel Barth[1,2], Ron Hübler[1,3,4], Aria Baniahmad[3,*], and Manja Marz[1,2,*]

[1]Bioinformatics/High Throughput Analysis, Friedrich Schiller University, Jena, Germany

[2]FLI Leibniz Institute for Age Research, Jena, Germany

[3]Institute of Human Genetics, Jena University Hospital, Jena, Germany

[4]Applied Systems Biology, Leibniz Institute for Natural Product Research and Infection Biology, Hans-Knöll-Institute (HKI), Jena, Germany

*Corresponding author: E-mail: aria.baniahmad@med.uni-jena.de; manja@uni-jena.de.

Accepted: March 26, 2016

Abstract

The COP9 signalosome (CSN) is a highly conserved protein complex, recently being crystallized for human. In mammals and plants the COP9 complex consists of nine subunits, CSN 1–8 and CSNAP. The CSN regulates the activity of culling ring E3 ubiquitin and plays central roles in pleiotropy, cell cycle, and defense of pathogens. Despite the interesting and essential functions, a thorough analysis of the CSN subunits in evolutionary comparative perspective is missing. Here we compared 61 eukaryotic genomes including plants, animals, and yeasts genomes and show that the most conserved subunits of eukaryotes among the nine subunits are CSN2 and CSN5. This may indicate a strong evolutionary selection for these two subunits. Despite the strong conservation of the protein sequence, the genomic structures of the intron/exon boundaries indicate no conservation at genomic level. This suggests that the gene structure is exposed to a much less selection compared with the protein sequence. We also show the conservation of important active domains, such as PCI (proteasome lid-CSN-initiation factor) and MPN (MPR1/PAD1 amino-terminal). We identified novel exons and alternative splicing variants for all CSN subunits. This indicates another level of complexity of the CSN. Notably, most COP9-subunits were identified in all multicellular and unicellular eukaryotic organisms analyzed, but not in prokaryotes or archaeas. Thus, genes encoding CSN subunits present in all analyzed eukaryotes indicate the invention of the signalosome at the root of eukaryotes. The identification of alternative splice variants indicates possible "mini-complexes" or COP9 complexes with independent subunits containing potentially novel and not yet identified functions.

Key words: signalosome subunits CSN, genomic structure, comparative informatics for plants, fungi, animal kingdom, bacteria.

Introduction

The COP9 signalosome (CSN) complex is a highly conserved protein complex consisting of eight subunits CSN1–CSN8 together with the very recently identified ninth subunit, CSNAP (Rozen et al. 2015). Originally, the CSN complex was identified in 1994 as a photomorphogenic regulator in *Arabidopsis thaliana* mediating light controlled developmental regulation (Wei et al. 1994; Chamovitz et al. 1996; Staub et al. 1996; Karniol et al. 1999; Serino et al. 1999). Later, CSN has also been identified in mammals and invertebrates and hence assumed that the CSN probably exists in almost all multicellular eukaryotes (Wei and Deng 1998; Wei et al. 1998). Functionally, the CSN is associated with enzymatic activity. The CSN functions as an isopeptidase with deneddylation activity removing specifically the covalent NEDD8 modification from cullins (Lyapina et al. 2001) and acting in the ubiquitin-proteasomal pathway of protein degradation. Also other enzymatic function such as phosphorylation activity is associated with the CSN (Bech-Otschir et al. 2001). Although CSN5 is essential for the deneddylation activity, the CSN5 alone does not mediate deneddylation and thus it is suggested that all CSN subunits are required in a complex for the deneddylation activity. Interestingly, CSN5 is suggested to act also as a monomer to bind to transcription factors such as JunD and also CSN subcomplexes have also been described (Kwok et al. 1998; Tomoda et al. 2002; Sharon et al. 2009). These findings indicate deneddylation-independent functions of the CSN. This notion is supported by knocking out different CSN subunits

that results in partial distinct phenotypes (Mundt et al. 2002; Oron et al. 2002). The essential role of CSN subunits for the development of multicellular organisms has been described. In plants using the model organism *Arabidopsis*, the inactivity of individual subunit affects the development of seedlings and meristem (Franciosini et al. 2015). Further, the knockout of individual CSN subunits resulted in very early embryonic lethality in mice, a mammalian model organism (reviewed in: Wei et al. 2008). In mice the knockout of CSN subunits resulted in deregulated key cell cycle factors including the tumor suppressor p53, p27, cyclin E and in *Drosophila* the retinoblastoma factor Rbf1 (Lykke-Andersen et al. 2003; Yan et al. 2003; Menon et al. 2007; Ullah et al. 2007). In *Drosophila* the inactivation or knockdown of CSN subunits results in maintaining the germ line cellular microenvironment and regulates cell fate decisions and the balance between self-renewing function and differentiation (Carreira-Rosario and Buszczak 2014; Pan et al. 2014; Qian et al. 2015). The CSN regulates the expression of genes. CSN subunits physically associate with the ecdysone receptor, which leads to transcriptional repression (Dressel et al. 1999; Huang et al. 2014). The ecdysone signaling to control prepupa-to-pupa transition requires CSN deneddylating activity (Huang et al. 2014). The ecdysone receptor is a ligand-controlled transcription factor and a member of the nuclear hormone receptor family that mediate hormone regulated. These findings suggest that the CSN regulates not only protein stability but may influence the transcription. In line with this, the CSN as a transcription factor has also been reported by the finding that CSN7 interacts with multiple genomic loci to control development in *Drosophila* (Singer et al. 2014). There are indications of a distinct CSN in yeasts. Interestingly, the unicellular organism *Saccharomyces cerevisiae* can survive without a functional CSN (Wee et al. 2002). The deneddylation activity in *S. cerevisiae* is mediated by a CSN5 homolog (Licursi et al. 2014), the homology to the mammalian CSN5 is only about 30%. In other yeast organisms such as in several *Ascomycota*, the CSN is smaller and lacks orthologs for a few CSN subunits, but nevertheless contains a conserved CSN5 (Pick et al. 2012). This indicates that in evolution in the development of the yeast kingdom the CSN has been changed in its composition. Recently, it emerges that the CSN in mammals acts as a regulation modulator of a wide range of different biological processes such as signal transduction, autophagy, circadian rhythm as well as cell and embryonic development. It is assumed that it is also involved in a variety of human cancers since its influence on cell cycle checkpoint control and therefore on cell transformation and tumorgenesis was discovered. Thus, the CSN is with no doubt one of the major key players in eukaryotic developmental and cellular processes.

Besides the examination of the precise functions of the protein complex in different species, efforts were also taken to illuminate the evolution of the CSN complex. Here it was discovered during a database homology search that there exists a one-to-one sequence correspondence between subunits of the CSN and 19S proteasome lid. It is assumed that either the CSN evolved from an ancestral version of lid or that both the CSN and lid evolved from an common ancestor protein complex, so after an duplication event they diverged to their todays clearly distinct complexes (Glickman et al. 1998; Wei et al. 1998; Wei and Deng 1999). As the similarity between the mammal and the plant CSN is very high, but functions of both complexes partially differ, questions arose about their evolution and coevolution of subunits. No precise homology investigations were done to examine differences and similarities on sequence or gene level, with exception to the identification of the PCI (proteasome lid-CSN-initiation factor) and MPN (MPR1/PAD1 amino-terminal) domain in several CSN subunits (Wei et al. 1998). Also less effort was taken to identify the subunits CSN1–CSN8 and CSNAP in other species in addition to *Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *S. cerevisiae*, and *A. thaliana*.

Here, we analyzed 61 eukaryotic and over 2,200 prokaryotic genomes to identify and compare all subunits of CSN to have a closer look on its evolution from a putative common ancestor to the today living species. Additionally, we analyzed and compared the expression of human RNA-Seq data sets. We present a comprehensive overview of the nine CSN subunits in an evolutionary context and claim the signalosome to be invented with the existence of eukaryotes, independently of their cellular complexity.

According to Lozada-Chávez et al. (2011) we distinguish simple (SMOs) and complex multicellular organisms (CMOs). All kind of balls, sheets or filaments of cells are counted as SMOs if they either arise from a single progenitor through mitotic division and keep sticking together (aquatic origin) or if several solitary cells aggregate to form a colony (terrestrial origin). Even though SMOs can form coherent morphology by cell–cell adhesion, they show only limited intercellular signaling and less complex differentiation patterns (Bonner 1998; Wolpert and Szathmáry 2002; Grosberg and Strathmann 2007; Knoll 2011). Nevertheless, differentiation of somatic and reproductive cells is common. Since the first signs of cell differentiation come from fossils of filamentous and mat-forming cyanobacteria-like organisms (Tomitani et al. 2006), SMOs can be found in some eubacterial clades, for example, cyanobacteria, myxobacteria and actinobacteria, but are more common in eukaryotic lineages such as chlorophyceae, dictyostelia and oomycetes (Bonner 1998; Kaiser 2001; Rokas 2008). CMOs on the other side show a diversity of different genes that are involved in processes such as cell-cell and cell matrix adhesion as well as intercellular signaling pathways associated with developmental and cell-death programs. This allows the specialization of cell types and differentiation of multiple tissues mediated by complex regulatory networks. Complex multicellularity is limited to Eukarya and has been the product of both evolutionary innovations and enhancement of genetic material from ancestral unicellular organisms

(King 2004; Floyd and Bowman 2007; King et al. 2008; Rokas 2008; Specht and Bartlett 2009; Cock et al. 2010; Srivastava et al. 2010).

Materials and Methods

Data Sources

We downloaded 61 genomes from National Center for Biotechnology Information (NCBI) (Pruitt et al. 2007) and exons of the eight COP9-subunits of 15 eukaryotes

from Ensemble (release 75) (Flicek et al. 2013) (see supplementary table S1, Supplementary Material online). For isoform identification, we downloaded eight human RNA-Seq data sets from NCBI SRA on to the human genome *GRCh37* (see supplementary table S2, Supplementary Material online). Expression profiles for unicellular plants (*Micromonas pusilla*, *Aureococcus anophagefferens*, *Ostreococcus lucimarinus*) and heterokonts (*Phaeodactylum tricornutum*) are obtained from NCBI's SRA (see supplementary table S6, Supplementary Material online).

CSN Identification

To identify genomic locations of each exon within the genomes, we used tBLASTn (v2.2.1, *E*-value $\leq 10^{-5}$) (Altschul et al. 1990) for homology search. For CSNAP the *E*-Value threshold had to be lowered (*E*-value $\leq 10^{1}$) due to its very small exon sizes and relatively low sequence complexity, especially at its C-terminal end. Overlapping results were merged. If exons were found in mainly consecutive order on the same chromosome/contig, we defined this region to be a homologous gene and aligned the exons individually. If no consecutive exons were obtained, for example, in massively fragmented genomes, we used for each exon independently the best hit. Exon boundaries were automatically ($\leq$ 9 aa) and manually ($\geq$ 10 aa) extended. Final mRNA sequences were aligned, manually inspected and added to the query set. The complete CSN identification was repeated with the new query set until the query set did not change. A validation of the prediction was performed using only CSN sequences of *H. sapiens*, *A. thaliana*, and *D. melanogaster* in order to identify the known COP9-subunits of the other 12 species in the initial search query (see supplementary table S5, Supplementary Material online). We calculated the alignment conservation score by the ratio of blast alignment score and number of amino acids in the alignment.

Alternative Splicing

All reads from the eight RNA-Seq data sets were mapped with TopHat2 (v2.0.11) (Kim et al. 2013) to the human reference genome with –microexon-search. For extraction of splice sites, we used Haarz (v0.1) (Hoffmann et al. 2014) with default parameters. Visualization of mapped reads and splice sites was performed with IGV, Sashimi Plot (v2.0.34) (Thorvaldsdóttir et al. 2012).

Results and Discussion

We examined in silico 25 unicellular, 11 simple multicellular, and 8 CMOs spanning all nonmetazoan eukaryotes. Additionally, we observed 17 metazoan eukaryotes to depict the recent evolution of the signalosome in CMOs. All predicted CSN subunit sequences and locations can be found in supplementary table S3, Supplementary Material online.

Evolutionary Flexibility of Signalosomal Subunits

The single subunits of the signalosome evolved under different selection pressures. The alignment conservation score reveals CSN5 being most conserved, see Table 1, being in line with the fact, that CSN5 harbors the catalytic center of the signalosome (Wei and Deng 2003).

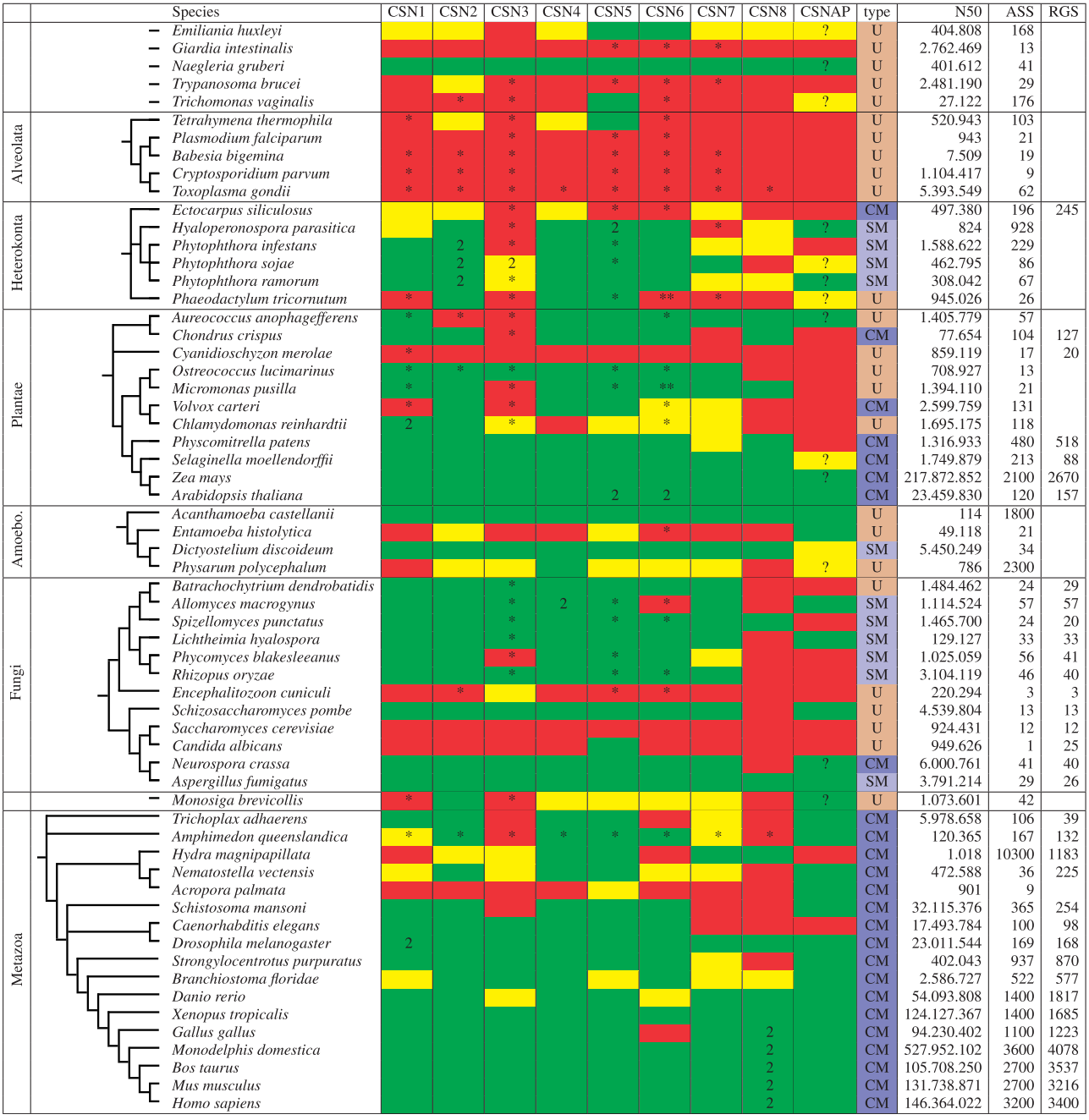
The highly conserved CSN2 is also important for cullin deneddylation activity of the CSN complex, and a shorter isoform, named *Alien*, is involved in nucleosome formation and repression of certain nuclear receptors (Yang et al. 2002; Eckey et al. 2007). CSN3, CSNAP and CSN8 are in relation extremely flexible and the latter one for lower eukaryotes even assumed to be lost (Liu et al. 2013), which we think is questionable, as we clearly identified CSN8 homologs in, for example, *Naegleria gruberi* or *Acanthamoeba castellanii* (fig. 1). CSN8 null mutant experiments showed lethal effects for higher eukaryotes (e.g., *D. melanogaster* or *A. thaliana*) (Oron et al. 2002; Serino and Deng 2003; Wei and Deng 2003). For *C. elegans* it has been shown that CSN-eukaryotic initiation factor (CIF-1) replaces CSN7 and is shared by the CSN and the eIF3 complexes (Luke-Glaser et al. 2007), which might also explain the loss of CSN7 in *S. mansoni*.

Our data also indicate some subunits being at least duplicated in several organisms (e.g., CSN2 in the heterokonts

Table 1
Evolutionary Flexibility of Single Subunits of the Signalosome

Subunit	Alignment Score	No. Alignment Aminoacids	Conservation Score
CSN5	1.241.023	18.436	67.32
CSN2	1.287.595	23.005	55.97
CSN4	723.932	16.502	43.87
CSN1	639.633	17.119	37.36
CSN7	214.604	7.749	27.69
CSN6	270.735	10.601	25.54
CSN3	213.010	11.318	18.82
CSNAP	24.094	1.385	17.40
CSN8	73.710	4.400	16.75

CSN5 has highest selection pressure, whereas CSN3, CSNAP and CSN8 are more flexible.



	Species	CSN1	CSN2	CSN3	CSN4	CSN5	CSN6	CSN7	CSN8	CSNAP	type	N50	ASS	RGS
	Emiliania huxleyi									?	U	404.808	168	
	Giardia intestinalis					*	*	*			U	2.762.469	13	
	Naegleria gruberi									?	U	401.612	41	
	Trypanosoma brucei			*		*	*	*			U	2.481.190	29	
	Trichomonas vaginalis		*	*			*			?	U	27.122	176	
Alveolata	*Tetrahymena thermophila*	*		*			*				U	520.943	103	
Alveolata	*Plasmodium falciparum*			*		*	*				U	943	21	
Alveolata	*Babesia bigemina*	*	*	*		*	*	*			U	7.509	19	
Alveolata	*Cryptosporidium parvum*	*	*	*		*	*	*			U	1.104.417	9	
Alveolata	*Toxoplasma gondii*	*	*	*	*	*	*	*	*		U	5.393.549	62	
Heterokonta	*Ectocarpus siliculosus*			*		*	*				CM	497.380	196	245
Heterokonta	*Hyaloperonospora parasitica*			*		2		*		?	SM	824	928	
Heterokonta	*Phytophthora infestans*		2	*		*					SM	1.588.622	229	
Heterokonta	*Phytophthora sojae*		2	2		*				?	SM	462.795	86	
Heterokonta	*Phytophthora ramorum*		2	*						?	SM	308.042	67	
Heterokonta	*Phaeodactylum tricornutum*	*		*		*	**	*		?	U	945.026	26	
Plantae	*Aureococcus anophagefferens*	*	*	*			*			?	U	1.405.779	57	
Plantae	*Chondrus crispus*			*							CM	77.654	104	127
Plantae	*Cyanidioschyzon merolae*	*									U	859.119	17	20
Plantae	*Ostreococcus lucimarinus*	*	*	*		*	*				U	708.927	13	
Plantae	*Micromonas pusilla*	*		*		*	**				U	1.394.110	21	
Plantae	*Volvox carteri*	*		*			*				CM	2.599.759	131	
Plantae	*Chlamydomonas reinhardtii*	2		*			*				U	1.695.175	118	
Plantae	*Physcomitrella patens*										CM	1.316.933	480	518
Plantae	*Selaginella moellendorffii*									?	CM	1.749.879	213	88
Plantae	*Zea mays*									?	CM	217.872.852	2100	2670
Plantae	*Arabidopsis thaliana*					2	2				CM	23.459.830	120	157
Amoebo.	*Acanthamoeba castellanii*										U	114	1800	
Amoebo.	*Entamoeba histolytica*						*				U	49.118	21	
Amoebo.	*Dictyostelium discoideum*										SM	5.450.249	34	
Amoebo.	*Physarum polycephalum*									?	U	786	2300	
Fungi	*Batrachochytrium dendrobatidis*			*							U	1.484.462	24	29
Fungi	*Allomyces macrogynus*			*	2	*	*				SM	1.114.524	57	57
Fungi	*Spizellomyces punctatus*			*		*	*				SM	1.465.700	24	20
Fungi	*Lichtheimia hyalospora*			*							SM	129.127	33	33
Fungi	*Phycomyces blakesleeanus*			*		*					SM	1.025.059	56	41
Fungi	*Rhizopus oryzae*			*		*	*				SM	3.104.119	46	40
Fungi	*Encephalitozoon cuniculi*		*			*	*				U	220.294	3	3
Fungi	*Schizosaccharomyces pombe*										U	4.539.804	13	13
Fungi	*Saccharomyces cerevisiae*										U	924.431	12	12
Fungi	*Candida albicans*										U	949.626	1	25
Fungi	*Neurospora crassa*									?	CM	6.000.761	41	40
Fungi	*Aspergillus fumigatus*										SM	3.791.214	29	26
	Monosiga brevicollis	*		*						?	U	1.073.601	42	
Metazoa	*Trichoplax adhaerens*										CM	5.978.658	106	39
Metazoa	*Amphimedon queenslandica*	*	*	*	*	*	*	*	*		CM	120.365	167	132
Metazoa	*Hydra magnipapillata*										CM	1.018	10300	1183
Metazoa	*Nematostella vectensis*										CM	472.588	36	225
Metazoa	*Acropora palmata*										CM	901	9	
Metazoa	*Schistosoma mansoni*										CM	32.115.376	365	254
Metazoa	*Caenorhabditis elegans*										CM	17.493.784	100	98
Metazoa	*Drosophila melanogaster*	2									CM	23.011.544	169	168
Metazoa	*Strongylocentrotus purpuratus*										CM	402.043	937	870
Metazoa	*Branchiostoma floridae*										CM	2.586.727	522	577
Metazoa	*Danio rerio*										CM	54.093.808	1400	1817
Metazoa	*Xenopus tropicalis*										CM	124.127.367	1400	1685
Metazoa	*Gallus gallus*								2		CM	94.230.402	1100	1223
Metazoa	*Monodelphis domestica*								2		CM	527.952.102	3600	4078
Metazoa	*Bos taurus*								2		CM	105.708.250	2700	3537
Metazoa	*Mus musculus*								2		CM	131.738.871	2700	3216
Metazoa	*Homo sapiens*								2		CM	146.364.022	3200	3400

Fig. 1.—Signalosome exists in unicellular organisms. Comparison of COP9 signalosome subunit existence in 61 eukaryotes. Colored columns indicate if the corresponding subunit could be identified to a length of at least 80% (green), at least 40% (yellow) or not (red), in respect to the corresponding alignment. If additionally a close related lid protein was identified it was discriminated and marked (*). If an additional copy of a subunit was found it was marked with the number "2." For CSNAP inconclusive but possible homolog candidates were marked with a question mark. Definition of cellularity of the species is described in the Introduction, based on the publication of Lozada-Chávez et al. (2011). Phylogenetic trees are based on Lozada-Chávez et al. (2011), Cavalier-Smith et al. (2015), Ebersberger et al. (2012), and Federhen (2012). RGS—Real Genome sizes (in Mb)—were obtained from the genome size database projects (Gregory et al. 2007), if available. ASS, the number of nucleotides of the genome assembly (in Mb); U, unicellular; SM, simple multicellular; CM, complex multicellular; N50, the length of the contig containing more than 50% of the nucleotides of the genome assembly when sorting for contig length. Multiple copies are marked with the corresponding number in each column.

Phytophthora). Due to the improvable genome assemblies, we are not able to clearly identify the number of copies per genome. Although we were able to clearly distinguish the signalosome- and the homologous lid-subunits, the possibility of mutual usage of one of the homologous genes cannot be excluded, being for example, very likely for CSN3 of heterokonts.

Macroevolution: Signalosome Exists in Multicellular and Unicellular Organisms

For most of the 25 CMOs, we were able to identify all nine subunits of the signalosome (fig. 1). A notable exception exists for four examined species: *Ectocarpus siliculosus* (heterokont) has either a highly modified signalosome or lost the signalosome. As CSN5 was not found, vague gene candidates for CSN1/2/4/7 may refer to the homologous lid complex. Within plants single subunits were not identified: CSN3/7 in *Chondrus crispus* probably due to an unfinished assembly and CSN1/3/8 in *Volvox carteri* possibly by divergent evolution of these subunits. The recently discovered ninth subunit CSNAP seems to be not conserved in plants as it is in the metazoans, as only two vague *A. thaliana* homologs could be identified in the close *Zea mays* and *Selaginella moellendorffii*. It was already assumed by Rozen et al. (2015) that in plants the conservation may only be maintained in the C-terminal end. This region consists of mainly aspartic acid and phenylalanine showing a relative low sequence complexity, thus making it difficult to identify more homologs if the remaining CSNAP sequence diverged more in plants. For the most basal metazoan *Trichoplax adhaerens* we were not able to identify CSN3/6/8, very likely due to divergent subunits and *Acropora palmata* (Cnidaria, Metazoa), where the entire signalosome has been possibly removed, however, the assembly is one of the worst examined. Simple multicellular fungi and Amoebazoa harbor all or all but CSN8/CSNAP subunits. Simple multicellular heterokonts contain most of the signalosome subunits. The missing CSN6 in *Gallus gallus* is presumably due to the incomplete sequencing data.

One of the most basal examined unicellular Eukaryotes *N. gruberi* contains clearly the previously known eight subunits of the signalosome and a possible CSNAP candidate, leaving little doubt of a functional signalosome in unicellular organisms. For *Emiliania huxleyi*, we were also able to confirm the genomic existence of CSN5/6 and very likely candidates for CSN1/2/4/7/8/CSNAP but not for CSN3. For the unicellular organisms *Giardia intestinalis*, *Trypanosoma brucei*, and *Trichomonas vaginalis* we were not able to annotate the signalosome, although the latter organism shows a clear homolog of the most conserved subunit CSN5. For the latter organism, we were able to identify the homologous subunits 2/3/6 of the lid complex. However, it remains unclear, if these subunits can replace the possibly lost signalosome subunits. We propose Alveolata to have a very diverged signalosome. Most of the subunits were not identified within the genomes, but we found a not rationalizable CSN5 homolog in *Tetrahymena thermophila*. The unicellular heterokont *P. tricornutum* genome contains CSN2/4/5 homologs but no candidates for the other signalosome subunits. The expression profiles of these three subunits are shown in the supplementary material, Supplementary Material online, leaving speculation of possible insertion/expansion of single protein domains.

Although unicellular plant genomes contain not all of the eight subunits, in general there we leave little doubts, that the signalosome is functional. *Cyanidioschyzon merolae* seems to lack all subunits and, although the genome assembly is not too bad, we propose this plant has lost the signalosome during evolution. For *A. anophagefferens*, *O. lucimarinus* and *M. pusilla*, expression profiles of the single subunits are shown (fig. 2 and supplementary table S6, Supplementary Material online).

Within Amoebazoa *A. castellanii* has kept clearly all nine subunits, whereas the existence of the signalosome in *Entamoeba histolytica* and *Physarum polycephalum* remains unclear. Unicellular fungi show also a very diverse evolutionary picture: *Batrachochytrium dendrobatidis* and *Schizosaccharomyces pombe* contain CSN1-7, whereas *Encephalitozoon cuniculi*, *Candida albicans* and *S. cerevisiae* have a highly diverged signalosome-like complex (Maytal-Kivity et al. 2003) or possibly lost their signalosome. Finally, the unicellular *Monosiga brevicollis* has a CSN2 homolog and possibly CSN4–7 and CSNAP homologs. However, whether this organism contains a functional signalosome remains at this point unclear.

The Evolution of CSN2 and CSN5 Reveals Conserved Intron Insertion

We investigated the most conserved and central subunits CSN2 and CSN5 in more detail (fig. 3). Vertebrates show only marginal changes in their intron/exon structure, however, throughout all eukaryotes, we observe a widely varying change for intron and exon length. In general, lower eukaryotes contain less introns than higher eukaryotes.

No correlation between cellularity and number of introns can be observed. However, within the main taxonomic groups (except fungi) we detect a slight trend for less introns in basal organisms.

Interestingly in metazoa, the CSNs seem to have conserved intron insertion sites, recognizable by colors coding for orthologous of human exons in figure 3. Notable is the constant intron insertion between exon 1/2, 2/3, 5/6, and 7/8, which have to be introduced multiple times throughout evolution, considering basal organisms per group containing no or less

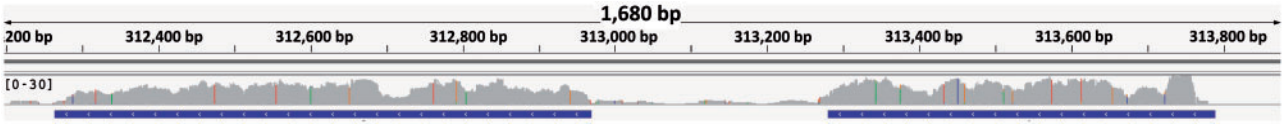


Fig. 2.—The expression profile of CSN4 of the unicellular organism *M. pusilla* (gray, top) covers the predicted homologous region (blue, bottom) of other eukaryotes. The expression of various predicted CSNs in unicellular organisms can be viewed in the supplementary material, Supplementary Material online.

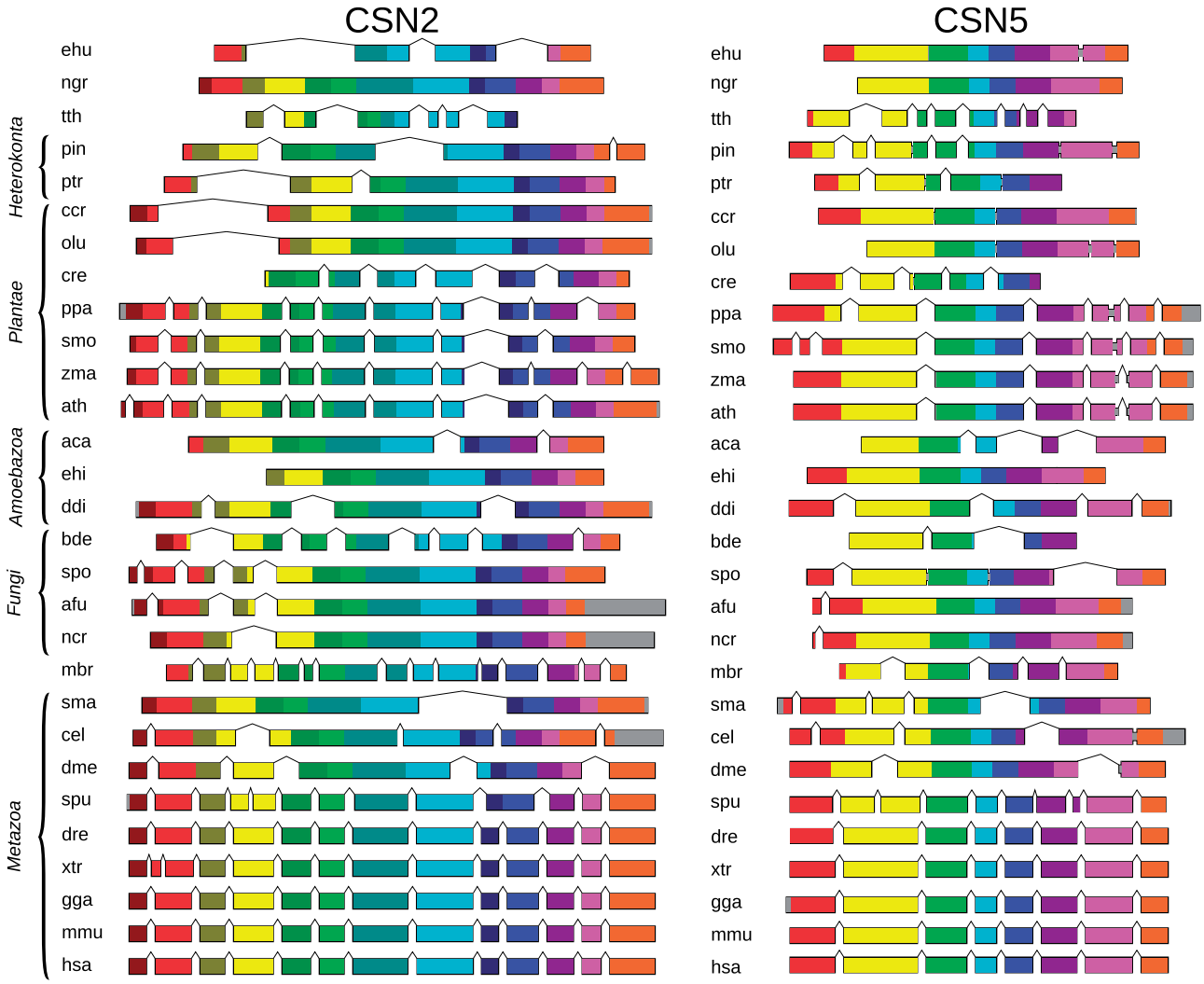


Fig. 3.—Exon–intron structure of *csn2* (left) and *csn5* (right). Homologous exons are displayed as equal colored boxes (adjusted at human), gray boxes indicate unique sequences, incomparable with any other species. Half-sized gray boxes could be either insertions or small intronic sequences. Number of nucleotides retrieved from exons is scaled proportionally in the figure, whereas the length of introns is not comparable. Alignments for all subunits are available in the supplementary material, Supplementary Material online (supplementary table S4, Supplementary Material online). ehu, *E. huxleyi*; ngr, *N. gruberi*; tth, *T. thermophila*; pin, *P. infestans*; ptr, *P. tricornutum*; ccr, *C. crispus*; olu, *O. lucimarinus*; cre, *C. reinhardtii*; ppa, *P. patens*; smo, *S. moellendorffii*; zma, *Z. mays*; ath, *A. thaliana*; aca, *A. castellanii*; ehi, *E. histolytica*; ddi, *D. discoideum*; bde, *B. dendrobatidis*; spo, *S. pombe*; afu, *Aspergillus fumigatus*; ncr, *N. crassa*; mbr, *M. brevicollis*; sma, *S. mansoni*; cel, *C. elegans*; dme, *D. melanogaster*; spu, *S. purpuratus*; dre, *D. rerio*; xtr, *X. tropicalis*; gga, *G. gallus*; mmu, *Mus musculus*; hsa, *H. sapiens*.

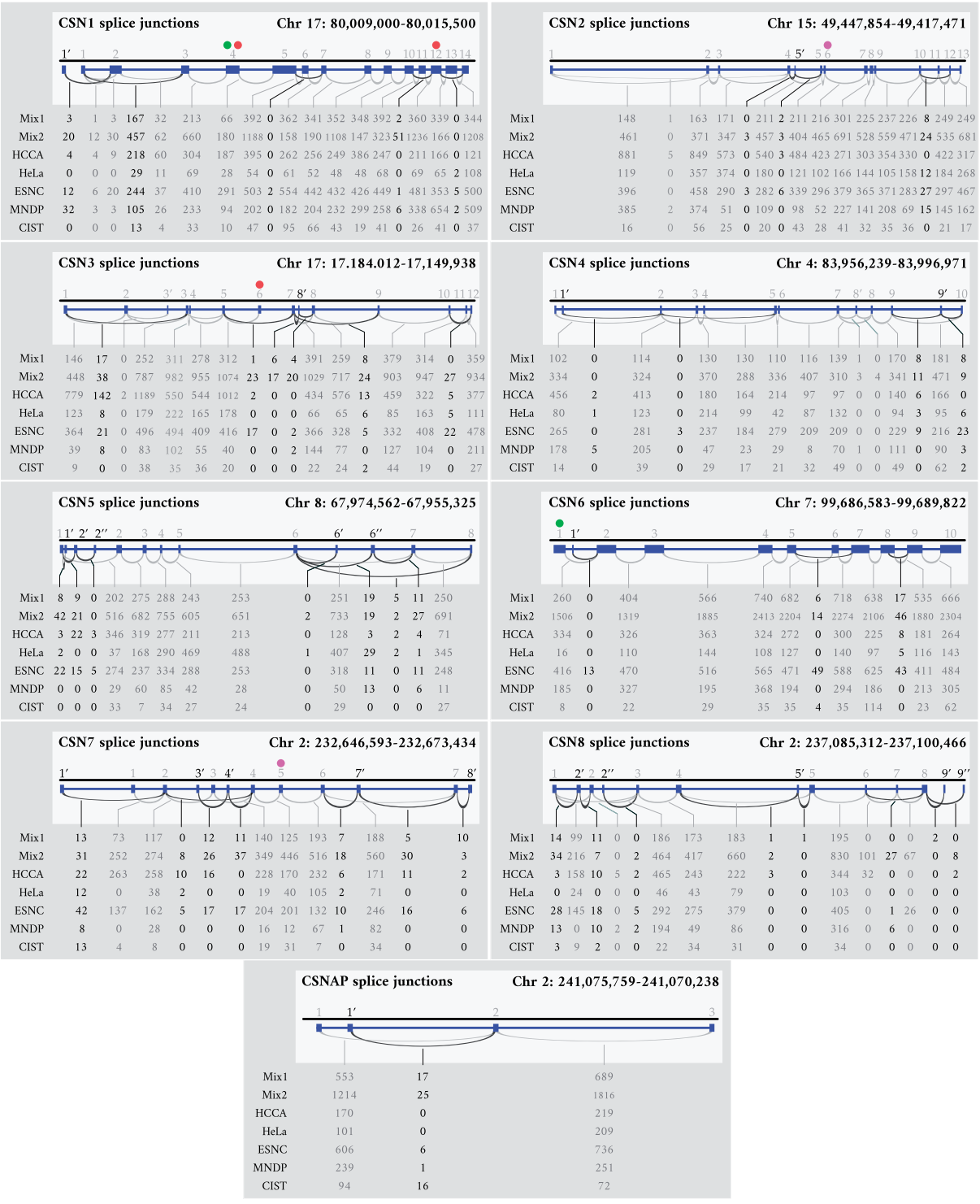


Fig. 4.—Alternative splicing of human CSN subunits. Blue boxes indicate exons whereas lines between them are splice junctions. Known junctions are colored gray and putative new junctions are colored black. The same applies for the number labels of the exons. Distances between exons were drawn proportionally in respect to their genomic locations. The tables underneath contain the amount of mapped splitted reads supporting a certain splice junction, where columns contain the number of reads supporting a specific junction and rows all supporting reads belonging to a specific RNA-Seq library. Exons marked with colored dots contain a base exchange at a specific position (green: G to T; red: G to A; purple: A to G). Mix1, Universal Human Reference RNA 1; Mix2, Universal Human Reference RNA 2; HcCa, hepatocell-carcinoma; HeLa, cervical-carcinoma; ESNC, ES-derived neural progenitor cells; MNDP, motoric neuron from reprogrammed dental pulp; CIST, cortical ischemic stroke tissue.

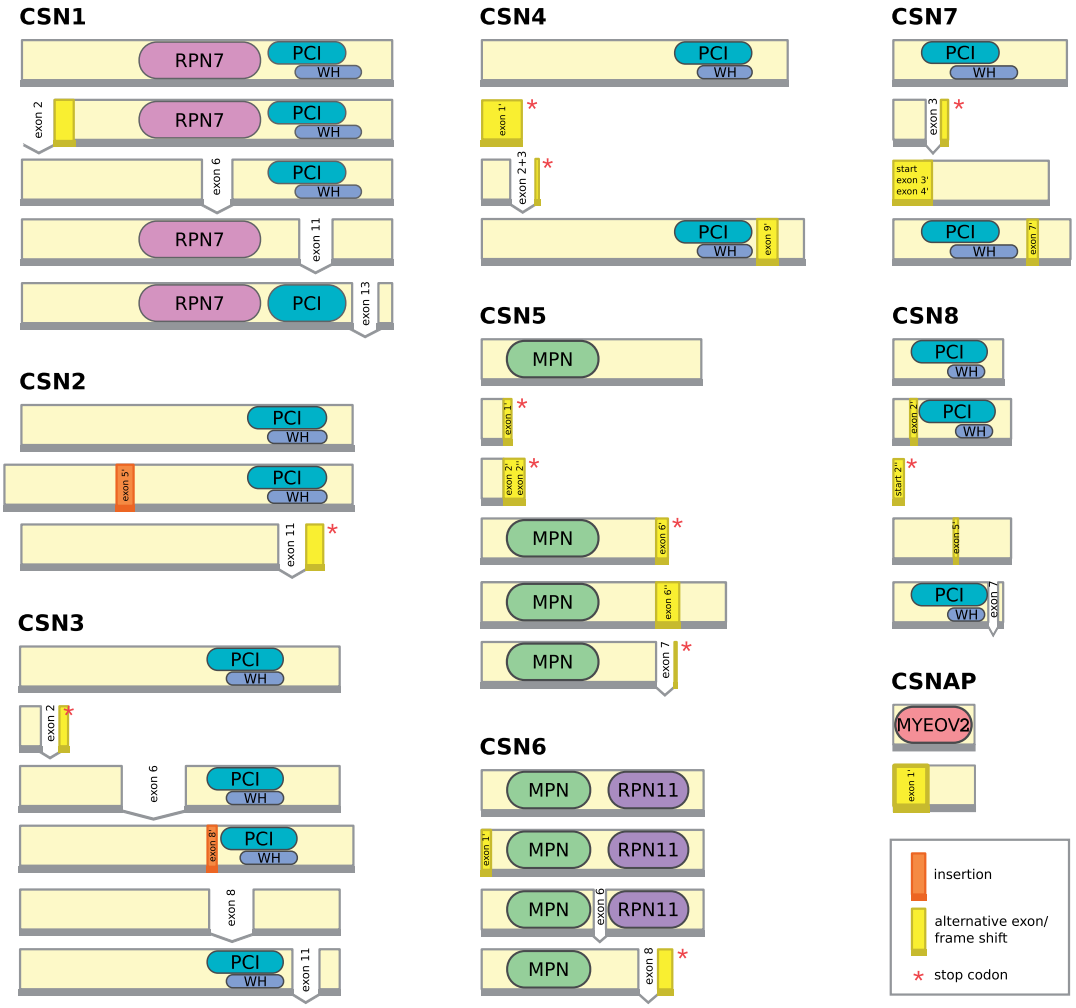


Fig. 5.—Consequences of the predicted alternative splicing points for the CSN proteins in human. For each subunit, the common protein isoform is placed on the top of the stack of predicted alternative isoforms. Yellowish boxes indicate the CSN proteins, where their left and right ends are the 5′- and 3′-ends of the corresponding proteins, respectively. Colored ellipses within the protein boxes depict the different domains present in the CSN proteins (MYEOV2—myeloma overexpressed 2). A small orange or yellow box indicates an insertion of an aminoacid sequence or a changed sequence due to alternative exons and a resulting frameshift, respectively, whereas gaps in a protein box indicate skipped exons. Alternative isoforms showing a premature stop codon as a consequence of a frameshift are marked with a small red asterisk. All isoforms and their features sizes are depicted proportionally to each other.

introns. However, at this point the driving factors for intron insertions at specific positions remain unaquainted.

Identification of Novel Exons and Splice Variants

Recently, the crystal structure of the human CSN complex has been published, giving insights into the composition and three-dimensional interaction of the CSN subunits (Lingaraju et al. 2014). The carboxy-termini of each CSN subunit determines the functions by the MPN, PCI and winged-helix (WH) domains are localized. The PCI domains are characterized by helical repeats followed by a WH domain each.

The PCI domains build by their WH subdomains an open ring formation. Interestingly, the short three-stranded beta-sheets in each WH subdomain are oligomerized edge-to-edge in the order CSN7–CSN4–CSN2–CSN1–CSN3–CSN8. Although our conservation score suggests that CSN3 and CSN8 are least conserved, both, CSN3 and CSN8, seem to interact not only at their carboxy-terminal part but also in their amino-terminal part in humans (Lingaraju et al. 2014). This is supported by our data for CSN8 across all investigated species: Small but highly conserved regions can be found at the beginning of the proteins N-terminal and the very end of the C-terminal part. For CSN3 a similarly high conserved region can be found at its C-terminal for all species except fungi. Many investigated species lack by in silico identification the N-terminal and the conserved region can be observed in the

higher metazoa and plantae as well as the basal *N. gruberi* (see supplementary table S4, Supplementary Material online). In line with our findings, the crystal structure analysis suggests the absence of CSN3/8 can be tolerated whereas the lack of CSN1, 2, 4, 6 or 7 strongly disfavors CSN5 incorporation and CSN complex formation. The data of Pick et al. (2012) suggest that the deletions of the C-terminal helices have a pronounced effect on the CSN integrity, which is confirmed by the crystal structure. This is also supported by our data: The helical parts in the corresponding CSN proteins are among the most conserved parts over all species (see supplementary table S4, Supplementary Material online). When examining the transcripts of each of the eight subunits of the human signalosome, we were able to identify previously unknown isoforms. Novel exons and exonic jumps are specially described for each of the seven RNA-Seq data of human samples in figure 4. Analysis of alternative splice products of each CSN suggests the existence of CSN variants that lack part or the entire MPN, PCI, and WH domains (see fig. 5). Also the RPN domains of CSN 1 and 6 could be deleted by splice variants. This strongly indicates that splice variants of CSN subunits with deletion of important integrative structures may exist. The lack of functional PCI and WH domains suggests that shorter isoforms lack the required domain to be incorporated into the CSN and suggest the existence of sub- or mini-complexes that may interfere with the CSN complex or individual uncomplexed CSN subunits. In line with this, it was reported that CSN2 might have several isoforms analyzing various mouse tissues (Tenbaum et al. 2003).

Conclusions

The highly conserved CSN complex is present in all kind of unicellular and multicellular eukaryotes in plant, fungal, and animal kingdom for which a common ancestor is suggested. None of the 2,200 noneukaryotes comprises fragments of one of the CSNs. This indicates that the CSN has been invented at the root of eukaryotes. To date, the major function of the CSN complex has been shown to regulate stem cells, development and cell cycle, therewith multicellular organisms depend on a functional CSN. This work leaves speculations of further functions in unicellular and multicellular organisms. The CSN has been speculated to be originated from the lid subcomplex and evolved in parallel with the ancient protein complexes (Wei et al. 1998; Wei and Deng 1999). Although homologous factors of the 20S proteasome are known in archaeas and bacteria (De Mot et al. 1999), subunits of the lid-containing 19S proteasome are not known to exist in prokaryotes or archaeas.

The high conservation of the CSN subunits can be mainly detected on protein level, rather than on nucleotide level, suggesting the function of COP9 to be essential for life. The identified novel exons and splice variants may allow the construction of signalosome-like complexes which may lead to a changed interaction pattern with other factors, to a modulation of protein turnover or allow the generation of various mini-complexes. The unicellular yeast *S. cerevisiae* indeed has a 19S subcomplex belonging to its 26S proteasome. However, when compared with the mammalian CSN subunits it showed distinct less similarity than other eukaryotic species (Maytal-Kivity et al. 2003).

We included also the recently identified ninth CSN subunit, namely CSNAP, in our conservation investigations. Clear homologs in the *Metazoa*, *Amoebozoa*, and some fungi have been easily identified. However, in plants it seems like the CSNAP sequence has differed more compared with the conservation of the other subunits and their regarding homologs. Therefore, it is not clear whether CSNAP is lost in some eukaryotic species and replaced by its eIF3 or lid counterpart as a shared subunit, similar to CSN7 in *C. elegans*.

With this comprehensive in silico overview of the signalosome we open the perspective to find more functions of the signalosome additionally in unicellular organisms in the future.

Supplementary Material

Supplementary figure S1 and tables S1–S6 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work has been partially financed by Carl Zeiss Stiftung (Manja Marz) and was funded by the Thuringian country programme ProExzellenz (RegenerAging - FSU-I-03/14) of the Thuringian Ministry for Research (TMWWDG) (Emanuel Barth).

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

Bech-Otschir D, et al. 2001. COP9 signalosome-specific phosphorylation targets p53 to degradation by the ubiquitin system. EMBO J. 20(7):1630–1639.

Bonner JT. 1998. The origins of multicellularity. Integr Biol Issues News Rev. 1(1):27–36.

Carreira-Rosario A, Buszczak M. 2014. A competitive cell fate switch. Dev Cell 31(3):261–262.

Cavalier-Smith T, et al. 2015. Multigene phylogeny resolves deep branching of Amoebozoa. Mol Phylogenet Evol. 83:293–304.

Chamovitz DA, et al. 1996. The COP9 complex, a novel multisubunit nuclear regulator involved in light control of a plant developmental switch. Cell 86(1):115–121.

Christmann M, et al. 2013. Control of multicellular development by the physically interacting deneddylases DEN1/DenA and COP9 signalosome. PLoS Genet. 9(2):e1003275.

Cock JM, et al. 2010. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature 465(7298):617–621.

De Mot R, Nagy I, Walz J, Baumeister W. 1999. Proteasomes and other self-compartmentalizing proteases in prokaryotes. Trends Microbiol. 7(2):88–92.

Dressel U, et al. 1999. Alien, a highly conserved protein with characteristics of a corepressor for members of the nuclear hormone receptor superfamily. Mol Cell Biol. 19(5):3383–3394.
Ebersberger I, et al. 2012. A consistent phylogenetic backbone for the fungi. Mol Biol Evol. 29(5):1319–1334.
Eckey M, Hong W, Papaioannou M, Baniahmad A. 2007. The nucleosome assembly activity of NAP1 is enhanced by Alien. Mol Cell Biol. 27(10):3557–3568.
Federhen S. 2012. The NCBI taxonomy database. Nucleic Acids Res. 40(D1):D136–D143.
Flicek P, et al. 2013. Ensembl 2014. Nucleic Acids Res. 1196. 42:D749–D755.
Floyd SK, Bowman JL. 2007. The ancestral developmental tool kit of land plants. Int J Plant Sci. 168(1):1–35.
Franciosini A, et al. 2015. The COP9 SIGNALOSOME is required for post-embryonic meristem maintenance in *Arabidopsis thaliana*. Mol Plant 8(11):1623–1634.
Glickman MH, et al. 1998. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. Cell 94(5):615–623.
Gregory TR, et al. 2007. Eukaryotic genome size databases. Nucleic Acids Res. 35(suppl 1):D332–D338.
Grosberg RK, Strathmann RR. 2007. The evolution of multicellularity: a minor major transition? Annu Rev Ecol Evol Syst 38:621–654.
Hoffmann S, et al. 2014. A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection. Genome Biol. 15(2):R34.
Huang YC, Lu YN, Wu JT, Chien CT, Pi H. 2014. The COP9 signalosome converts temporal hormone signaling to spatial restriction on neural competence. PLoS Genet. 10(11):e1004760.
Kaiser D. 2001. Building a multicellular organism. Annu Rev Genet. 35:(1):103–123.
Karniol B, Malec P, Chamovitz DA. 1999. *Arabidopsis* FUSCA5 encodes a novel phosphoprotein that is a component of the COP9 complex. The Plant Cell 11(5):839–848.
Kim D, et al. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14(4):R36.
King N. 2004. The unicellular ancestry of animal development. Dev Cell 7(3):313–325.
King N, et al. 2008. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451(7180):783–788.
Knoll AH. 2011. The multiple origins of complex multicellularity. Annu Rev Earth Planet Sci. 39:217–239.
Kwok SF, et al. 1998. Arabidopsis homologs of a c-Jun coactivator are present both in monomeric form and in the COP9 complex, and their abundance is differentially affected by the pleiotropic cop/det/fus mutations. Plant Cell 10(11):1779–1790.
Licursi V, et al. 2014. The COP9 signalosome is involved in the regulation of lipid metabolism and of transition metals uptake in *Saccharomyces cerevisiae*. FEBS J. 281(1):175–190.
Lingaraju GM, et al. 2014. Crystal structure of the human COP9 signalosome. Nature 512(7513):161–165.
Liu C, et al. 2013. COP9 signalosome subunit Csn8 is involved in maintaining proper duration of the G1 phase. J Biol Chem. 288(28):20443–20452.
Lozada-Chávez I, Stadler PF, Prohaska SJ. 2011. "Hypothesis for the modern RNA world": a pervasive scoding RNA-based genetic regulation is a prerequisite for the emergence of multicellular complexity. Orig Life Evol Biosph. 41(6):587–607.
Luke-Glaser S, et al. 2007. CIF-1, a shared subunit of the COP9/signalosome and eukaryotic initiation factor 3 complexes, regulates MEL-26 levels in the *Caenorhabditis elegans* embryo. Mol Cell Biol. 27(12):4526–4540.
Lyapina S, et al. 2001. Promotion of nedd8-cul1 conjugate cleavage by cop9 signalosome. Science 292(5520):1382–1385.
Lykke-Andersen K, et al. 2003. Disruption of the COP9 signalosome Csn2 subunit in mice causes deficient cell proliferation, accumulation of p53 and cyclin E, and early embryonic death. Mol Cell Biol. 23(19):6790–6797.
Maytal-Kivity V, Pick E, Piran R, Hofmann K, Glickman MH. 2003. The COP9 signalosome-like complex in *S. cerevisiae* and links to other PCI complexes. Int J Biochem Cell Biol. 35(5):706–715.
Menon S, et al. 2007. COP9 signalosome subunit 8 is essential for peripheral T cell homeostasis and antigen receptor-induced entry into the cell cycle from quiescence. Nat Immunol. 8(11):1236–1245.
Mundt KE, Liu C, Carr AM. 2002. Deletion mutants in COP9/signalosome subunits in fission yeast *Schizosaccharomyces pombe* display distinct phenotypes. Mol Biol Cell 13(2):493–502.
Oron E, et al. 2002. COP9 signalosome subunits 4 and 5 regulate multiple pleiotropic pathways in *Drosophila melanogaster*. Development 129(19):4399–4409.
Pan L, et al. 2014. Protein competition switches the function of COP9 from self-renewal to differentiation. Nature 514:233–236.
Pick E, et al. 2012. The minimal deneddylase core of the COP9 signalosome excludes the Csn6 MPN-domain. PLoS One 7(8):e43980–e43980.
Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35(suppl 1):D61–D65.
Qian Y, Ng CL, Schulz C. 2015. CSN maintains the germline cellular microenvironment and controls the level of stem cell genes via distinct CRLs in testes of *Drosophila melanogaster*. Dev Biol. 398(1):68–79.
Rokas A. 2008. The origins of multicellularity and the early history of the genetic toolkit for animal development. Annu Rev Genet. 42:235–251.
Rozen S, et al. 2015. CSNAP Is a Stoichiometric Subunit of the COP9 Signalosome. Cell Rep. 13(3):585–598.
Serino G, Deng XW. 2003. The COP9 signalosome: regulating plant development through the control of proteolysis. Annu Rev Plant Biol. 54(1):165–182.
Serino G, et al. 1999. *Arabidopsis* cop8 and fus4 mutations define the same gene that encodes subunit 4 of the COP9 signalosome. Plant Cell 11(10):1967–1979.
Sharon M, et al. 2009. Symmetrical modularity of the COP9 signalosome complex suggests its multifunctionality. Structure 17(1):31–40.
Singer R, et al. 2014. Drosophila COP9 signalosome subunit 7 interacts with multiple genomic loci to regulate development. Nucleic Acids Res. 42:9761–9770.
Specht CD, Bartlett ME. 2009. Flower evolution: the origin and subsequent diversification of the angiosperm flower. Annu Rev Ecol Evol Syst. 40:217–243.
Srivastava M, et al. 2010. The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466(7307):720–726.
Staub JM, Wei N, Deng XW. 1996. Evidence for FUS6 as a component of the nuclear-localized COP9 complex in Arabidopsis. Plant Cell 8(11):2047–2056.
Tenbaum SP, et al. 2003. Alien/CSN2 gene expression is regulated by thyroid hormone in rat brain. Dev Biol. 254(1):149–160.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 14:178–192.
Tomitani A, Knoll AH, Cavanaugh CM, Ohno T. 2006. The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives. Proc Natl Acad Sci U S A. 103(14):5442–5447.

Tomoda K, et al. 2002. The cytoplasmic shuttling and subsequent degradation of p27Kip1 mediated by Jab1/CSN5 and the COP9 signalosome complex. J Biol Chem. 277(3):2302–2310.
Ullah Z, Buckley MS, Arnosti DN, Henry RW. 2007. Retinoblastoma protein regulation by the COP9 signalosome. Mol Biol Cell 18(4):1179–1186.
Wee S, Hetfeld B, Dubiel W, Wolf DA. 2002. Conservation of the COP9/signalosome in budding yeast. BMC Genet. 3(1):15.
Wei N, Deng XW. 1998. Characterization and purification of the mammalian COP9 complex, a conserved nuclear regulator initially identified as a repressor of photomorphogenesis in higher plants. Photochem Photobiol. 68(2):237–241.
Wei N, Deng XW. 1999. Making sense of the COP9 signalosome: a regulatory protein complex conserved from Arabidopsis to human. Trends Genet. 15(3):98–103.
Wei N, Deng XW. 2003. The COP9 signalosome. Annu Rev Cell Dev Biol. 19(1):261–286.
Wei N, Chamovitz DA, Deng XW. 1994. Arabidopsis COP9 is a component of a novel signaling complex mediating light control of development. Cell 78(1):117–124.
Wei N, et al. 1998. The COP9 complex is conserved between plants and mammals and is related to the 26S proteasome regulatory complex. Curr Biol. 8(16):919–924.
Wolpert L, Szathmáry E. 2002. Multicellularity: evolution and the egg. Nature 420(6917):745–745.
Yan J, et al. 2003. COP9 signalosome subunit 3 is essential for maintenance of cell proliferation in the mouse embryonic epiblast. Mol Cell Biol. 23(19):6798–6808.
Yang X, et al. 2002. The COP9 Signalosome inhibits p27 *kip1* degradation and impedes G1-S phase progression via deneddylation of SCF Cul1. Curr Biol. 12(8):667–672.

Associate editor: Yoshihito Niimura