Journal of Animal & Plant Sciences, 32(5): 2022, Page: 1261-1273
ISSN (print): 1018-7081; ISSN (online): 2309-8694
https://doi.org/10.36899/JAPS.2022.5.0533

CHLAMYDOMONAS REINHARDTII SEC23 PARALOGS AND THEIR PHYLOGENETIC RELATIONSHIPS

M. Aksoy[1,*], A. R. Grossman[2] and Ö. Musul[1]

[1]Department of Agricultural Biotechnology, Faculty of Agriculture, Akdeniz University, Antalya, Turkiye
[2]Department of Plant Biology, Carnegie Institution for Science, Stanford, USA
[*]Corresponding author's email: Münevver AKSOY maksoy@akdeniz.edu.tr
ORCIDs: M. Aksoy: https://orcid.org/0000-0002-0798-5805; A. R. Grossman: https://orcid.org/0000-0002-3747-5881;
Ö. Musul: https://orcid.org/0000-0002-2456 - 0462

ABSTRACT

C. reinhardtii has two putative SEC23 genes, *CrSEC23A* and *CrSEC23B*. The encoded polypeptides are only ~18.9% identical, suggesting that they might have different functions. It is not clear whether SEC23 paralogs have same or different functions in diverse organisms. Interestingly, our alignment and homology modeling showed that CrSEC23B does not have the conserved SEC24 binding motif (VFR), but instead appears to have an LPA motif in the same position. While LPA might be part of a novel SEC24 binding motif, CrSEC23B might have an alternate function that is either associated with or independent of COPII. Our results also show SEC23 orthologs in various organisms have variations in the putative SEC24 binding motif. Phylogenetic analyses place the SEC23 orthologs into two clusters that we designated group A (conventional; CrSEC23A-like orthologs) and group B (unconventional; CrSEC23B-like orthologs). Our results suggest that many photosynthetic organisms have a divergent SEC23 paralog. This divergence is not seen in animals. We hypothesize that divergent (unconventional) SEC23 paralogs might be the result of gene duplication and divergence that may facilitate specific aspects of trafficking. Since we only identified the B-like proteins in photosynthetic lineages, we hypothesize that B-like proteins may not have been present in the common ancestor involved in the primary endosymbiotic event.

Key words: *Chlamydomonas reinhardtii*, SEC23 paralogs, gene duplication, secretory pathway, COPII

Published first online May 20, 2022 Published final October 05, 2022

INTRODUCTION

The secretory pathway is essential for survival of eukaryotic organisms and many of the proteins that function in this pathway are conserved (Barlowe and Miller, 2013). Proteins that enter the secretory pathway are targeted to the endoplasmic reticulum (ER) through a signal peptide at their amino terminus (Walter *et al*. 1984). Those proteins that passage through the ER are folded, glycosylated (Chen *et al*. 2001; Farquhar *et al*. 1991) and packaged into secretory vesicles that bud from the ER membrane network and fuse with other cellular membrane systems, including those of the Golgi complex. These small membrane vesicles arise through the activity of coat proteins that create membrane deformations, enabling the formation of vesicles that associate with and traffic protein 'cargo' (Kirchhausen, 2000; Bonifacino and Lippincott-Schwartz, 2003; Faini *et al*. 2013). The ER-derived vesicles generated by coat protein complex II (COPII) (Lee *et al*. 2004; Peotter *et al*. 2019) carry cargo from the ER to the Golgi (Barlowe *et al*. 1994; Brandizzi *et al*. 2013), a critical step in protein secretion.

The minimum protein composition of COPII is SEC23, SEC24, SAR1, SEC13 and SEC31 in multimeric associations. COPII formation is initiated by recruitment of the SAR1 GTPase to the ER membrane by the ER-localized SEC12 protein (Nakano *et al*. 1989). GTP bound SAR1 recruits the SEC23-SEC24 complex (Matsuoka *et al*. 1998), with SEC23 functioning as the SAR1 GTPase activating protein (GAP) (Yoshihisa *et al*. 1993). SEC24 is the cargo binding protein (Miller *et al*. 2003). The SAR1/SEC23/SEC24 pre-budding complex then recruits proteins that form the outer layer of the COPII complex, SEC13 and SEC31 (Matsuoka *et al*. 1998). After the release of COPII vesicles from the ER, they tether to and fuse with the Golgi membranes.

The unicellular green alga *Chlamydomonas reinhardtii* is an excellent model organism to study diverse cellular processes including ciliary function, mating, and photosynthesis (Harris *et al*. 2009). It is believed that *C. reinhardtii* retains important characteristics of last eukaryotic common ancestor (LECA) (Cross and Umen, 2015) which makes it an important model organism in phylogenetic studies. Based on the latest genome annotation, the *C. reinhardtii* genome encodes homologs to all key COPII proteins, although the functions of these homologs have not been

experimentally defined. It is not known if two *SEC23* paralogs of *C. reinhardtii* have same or diverse functions. Also in other organisms, there are different opinions on this subject (see discussion for more). In *Arabidopsis thaliana* SEC23 paralogs seem to have different functions (Aboulela *et al.* 2018). In a contrary study, it is suggested that functions of SEC23 paralogs are interchangeable (Khoriaty *et al.* 2018). In this study, we wanted to shed more light on this subject by analyzing *C. reinhardtii* SEC23 paralogs. The purpose of this study was to examine the deduced sequences of *C. reinhardtii* SEC23 proteins concerning conserved motifs, three-dimensional structure, and similarities to SEC23 homologs in other organisms. The results are discussed in the context of both SEC23 function and evolution.

MATERIALS AND METHODS

Protein Sequence Alignment and Domain Prediction: Protein sequences were retrieved through keyword searches in Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html) and NCBI (Altschul *et al.* 1997; https://www.ncbi.nlm.nih.gov/protein). Retrieved sequences were used in Blast searches to identify orthologs in other organisms. *Cyanophora paradoxa* sequences were retrieved from Cyanophora Genome Project site (http://cyanophora.rutgers.edu/cyanophora_v2018/). Alignments of multiple protein sequences were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Pairwise alignments to determine percent identities were performed with Emboss Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). To determine the putative SEC24 binding motifs, protein sequences were aligned with the sequences of proteins for which the crystal structures were solved, such as ScSEC23p. The residues aligning with the VFR motif were considered potential SEC24 binding motifs. The same procedure was used for determination of the catalytic arginine. Domain architecture predictions were made using motif database Pfam (El-Gebali *et al*. 2019, http://pfam.xfam.org/) and domains were drawn with image creator MyDomains (http://prosite.expasy.org/mydomains/).

Homology Modeling of CrSEC23A, CrSEC23B and ScNel1: Homology modeling was performed using the Phyre2 web portal (Kelley *et al*. 2015). The Phyre2 program used chain A (ScSEC23p) of Protein Data Bank structure 1m2o (*S. cerevisiae* Sec23p/Sec24p complex) as a template (Bi *et al*. 2002) to generate the models of CrSEC23A, CrSEC23B and ScNel1p. All the models were generated with 100% confidence. For the CrSEC23A model, 715 amino acid residues (94% of the protein sequence) were used, starting with Asp2 and ending at Glu761. The CrSEC23B model was generated using 664 amino acid residues (67% of the protein sequence), starting at Leu30 and ending at Val914. The ScNel1p model was generated using 98% of the protein sequence. 1m2o structure file (Bi *et al*. 2002) was obtained from Protein Data Bank (https://www.rcsb.org/). 1m2o structure and the models we generated were visualized with Jmol, an open-source Java viewer that represents chemical structures in three-dimensional (http://www.jmol.org/).

Cell Culture and Strains: *C. reinhardtii* strains CC124 (wild type) and CC5325 was obtained from the Chlamydomonas Resource Center, University of Minnesota, USA. Cells were grown mixotrophically in Tris Acetate Phosphate (TAP) medium (Gorman and Levine, 1965) under a 12 h dark-12 h light period using white fluorescent light bulbs. Light intensity was approximately 40 μmol photons/m^2/sec. Cultures were maintained on solid TAP medium and subcultured once a month in the lab.

RNA Isolation and RT-PCR: For RNA isolation, cells were grown in 100 ml liquid TAP medium on a shaker with 120 rpm. Cells were grown to 2-4x10^6 cells/ml density (which usually takes 2-3 days after inoculation) and 10-20 ml culture was used for RNA isolation.

RNA was isolated using the NucleoSpin RNA Plant kit (Macherey-Nagel, Germany). Cells were dissolved in the lysis buffer and disrupted by passage through a 20-gauge syringe before continuing the isolation according to the instructions provided by the manufacturer. RNA concentrations were measured using Biodrop (Indolab, Utama, Indonesia). 300 ng total RNA was reverse transcribed using SuperScript III reverse transcriptase (Thermo Scientific, USA). 2 μl of cDNA was the substrate for PCR using the Taq polymerase core kit (Qiagen, Germany). Primer sequences were as follows: 23A-Exon10F (GCCTACTTCTACGTGGTGGT) and 23A-3utr-R (TCAAACACTCCGCTACCCAT) for *CrSEC23A*, 23B-Exon10F (CGGTCTGACCTGGAAGACAT) and 23B-Exon11R (TCCCGAGTCACCTCCAGAC) for *CrSEC23B*, CBLP-F (CTTCTCGCCCATGACCAC) and CBLP-R (CCCACCAGGTTGTTCTTCAG) for *CrCBLP*. Product sizes were 765 bp, 209 bp and 105 bp, respectively. We designed the primers to amplify the *CrSEC23A* and *CrSEC23B* transcripts while the *CBLP* primers were identical to those used in published work (Chang *et al*. 2009). PCR products were separated on a 1% agarose gel in TAE buffer and visualized with the UV imager MiniLumi system (DNR Bio-imaging Systems, Israel). In all reactions, a single product (visualized on the agarose gel) was generated. The specificity of the amplified products was determined by sequencing (BM Labosis, Turkey).

Phylogenetic Analysis: Amino acid sequences were aligned and the phylogenetic tree generated using

ClustalW (https://www.genome.jp/tools-bin/clustalw). Alignments and phylogenetic reconstructions were performed using the function "build" of ETE3 v3.1.1 (Huerta-Cepas *et al*. 2016), as implemented on the GenomeNet (https://www.genome.jp/tools/ete/). The phylogenetic tree was constructed using fasttree with slow NNI and MLACC=3 (to make the maximum-likelihood NNIs more exhaustive) (Price *et al*. 2009). Values at nodes are SH-like local support.

RESULTS

***C. reinhardtii* has Two Putative SEC23 Paralogs:** We identified two SEC23 paralogs in *C. reinhardtii* using a keyword search for "SEC23" in the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html#). The CrSEC23A and CrSEC23B paralogs both have similarities to SEC23 proteins in other organisms but show no similarity to other encoded proteins on the *C. reinhardtii* genome (based on BLAST analyses). While CrSEC23A and CrSEC23B share only 18.9% identity, CrSEC23A is 93.8% identical to *Gonium pectorale* SEC23A (GpSEC23A) and CrSEC23B is 54.5% identical to *Volvox carteri* SEC23B (VcSEC23B) (Table 1). Our domain analysis demonstrated that both CrSEC23A and CrSEC23B have domains typical of SEC23 proteins (Figure 1). However, CrSEC23B, ScNel1p and VcSEC23B only have the zinc finger and trunk domains; they are missing the beta sheet (BS), helical (H) and gelsolin repeat (GR) domains (Figure 1).

CrSEC23A and CrSEC23B Both Have the Conserved Catalytic Arginine but Differ in Their Motif Compositions: Both CrSEC23A and CrSEC23B have the conserved catalytic arginine (R722 in ScSEC23p) located in the gelsolin repeat domain present in ScSEC23p (Bi *et al*. 2002). This catalytic arginine appears to be R718 in CrSEC23A and R910 in CrSEC23B. These residues are within the region of the protein containing the gelsolin repeat domain (although CrSEC23B is predicted to not have this domain) and are highlighted with a red font (Figure 2).

Table 1: Similarity of *C. reinhardtii* SEC23A and SEC23B to SEC23 orthologs of various organisms.

Protein	Phytozome* or NCBI Accession Number	Percent Identity to CrSEC23A	Percent Identity to CrSEC23B
CrSEC23A	Cre03.g180850*	Same protein	18.9
CrSEC23B	Cre10.g447350*	18.9	Same protein
GpSEC23A	KXZ56797.1	93.8	19.4
GpSEC23B	KXZ47553.1	18	51.1
VcSEC23A	Vocar.0015s0302.1*	92	18.7
VcSEC23B	Vocar.0070s0024.1*	19.9	54.5
ScSEC23p	QHB12428.1	42.8	18.2
ScNel1p	QHB09024.1	20.9	0.2
AtSEC23A	NP_567217.1	15. 9	18.3
AtSEC23B	NP_563741.1	52.3	14.8
AtSEC23C	NP_179757.1	48.0	16.4
AtSEC23D	NP_565651.1	18.5	14.7
AtSEC23E	NP_189008.1	55.4	15.5
AtSEC23F	NP_193152.2	54.4	14.4
AtSEC23G	NP_568626.1	41.0	13.4
HsSEC23A	CAA65774.1	44.8	16.1
HsSEC23B	CAA65775.1	44.5	16.9
MmSEC23A	NP_001348885.1	45.1	15.9
MmSEC23B	NP_001342308.1	45.3	16.9

Cr: *Chlamydomonas reinhardtii*; Vc: *Volvox carteri;* Sc: *Saccharomyces cerevisiae*; At: *Arabidopsis thaliana*; Hs: *Homo sapiens*; Mm: *Mus musculus.* *: Phytozome accession numbers.

Figure 1

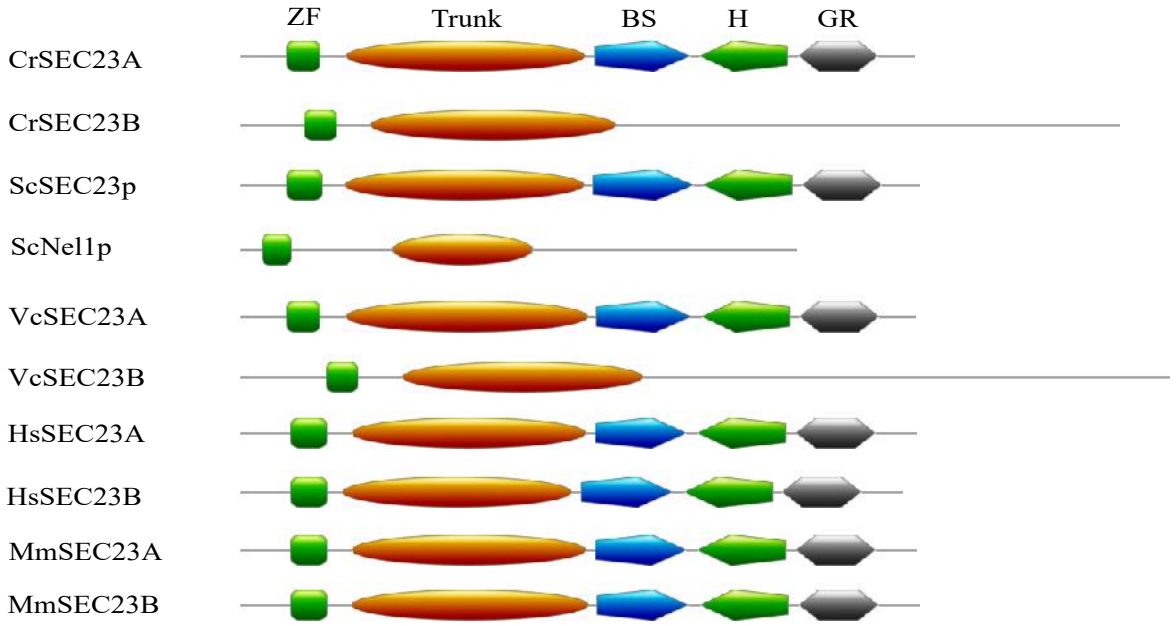

Figure 1. Domain architectures of SEC23 homologs from different organisms. CrSEC23B and VcSEC23B have different domain architectures than the other SEC23 proteins shown. Green square, Zinc Finger (ZF); Orange ellipse, Trunk; Blue pentagon, Beta Sandwich (BS); Green pentagon, Helical (H); Grey hexagon, Gelsolin Repeat (GR). Cr is *Chlamydomonas reinhardtii*, Sc is *Saccharomyces cerevisiae*, Vc is *Volvox carteri*, Hs is *Homo sapiens*, Mm is *Mus musculus*. Scale bar represents 100 amino acids.

A F382L substitution in human SEC23A (HsSEC23A) resulted in the loss of SEC23A function. This lesion segregates with cranio-lenticulo-sutural dysplasia (CLSD), which is characterized by skeletal abnormalities and facial dysmorphisms. Electron microscopy showed gross dilatation of the ER and cytoplasmic mislocalization of SEC31 in people with CLSD (Boyadjiev *et al*. 2006). The F382L substitution in HsSEC23A prevented binding of SEC23A to SEC31, which in turn prevented vesicle formation (Fromme *et al*. 2007; Bi *et al*. 2007). This phenylalanine residue is conserved in both ScSEC23p (F380) and CrSEC23A (F381) (Figure 2, indicated with an arrowhead) but interestingly, CrSEC23B has a leucine residue (L487) that aligns with this phenylalanine (Figure 2), which may mimic the functional limitation observed for the F382L substitution in HsSEC23A and alter the composition of COPII vesicles. The crystal structure of the ScSEC23p/ScSEC24p complex (PDB ID, 1m2o) indicates that ScSec23p binds to ScSec24p through a VFR motif in its trunk domain (Bi *et al*. 2002; Figure 2). Our alignment of ScSec23p, CrSEC23A, and CrSEC23B demonstrates the presence of the VFR motif in CrSEC23A but not in CrSEC23B. Instead, the latter has an LPA motif that aligns with VFR within the trunk domains of ScSec23p and CrSEC23A (Figure 2, highlighted with green background). This finding suggests that CrSEC23B has different amino acids involved in binding to SEC24, or that it doesn't bind SEC24 and is potentially not even a component of the COPII complex. Therefore, we examined other SEC23 homologs for the presence of VFR and LPA motifs. Most of these homologs have the VFR motif or a motif very similar to VFR, e.g. ILR, shown next to accession numbers (Figure 3). VcSEC23B has the LPA motif and one of the *G. pectorale* homologs (which we designated GpSEC23B) has an LPL motif that aligns with the LPA of CrSEC23B (Figure 4A). Alignments of these proteins and a *Chlamydomonas eustigma* ortholog are shown in Figure 4A. The *C. eustigma* ortholog doesn't appear to have either a VFR or LPA/LPL motif.

Figure 2

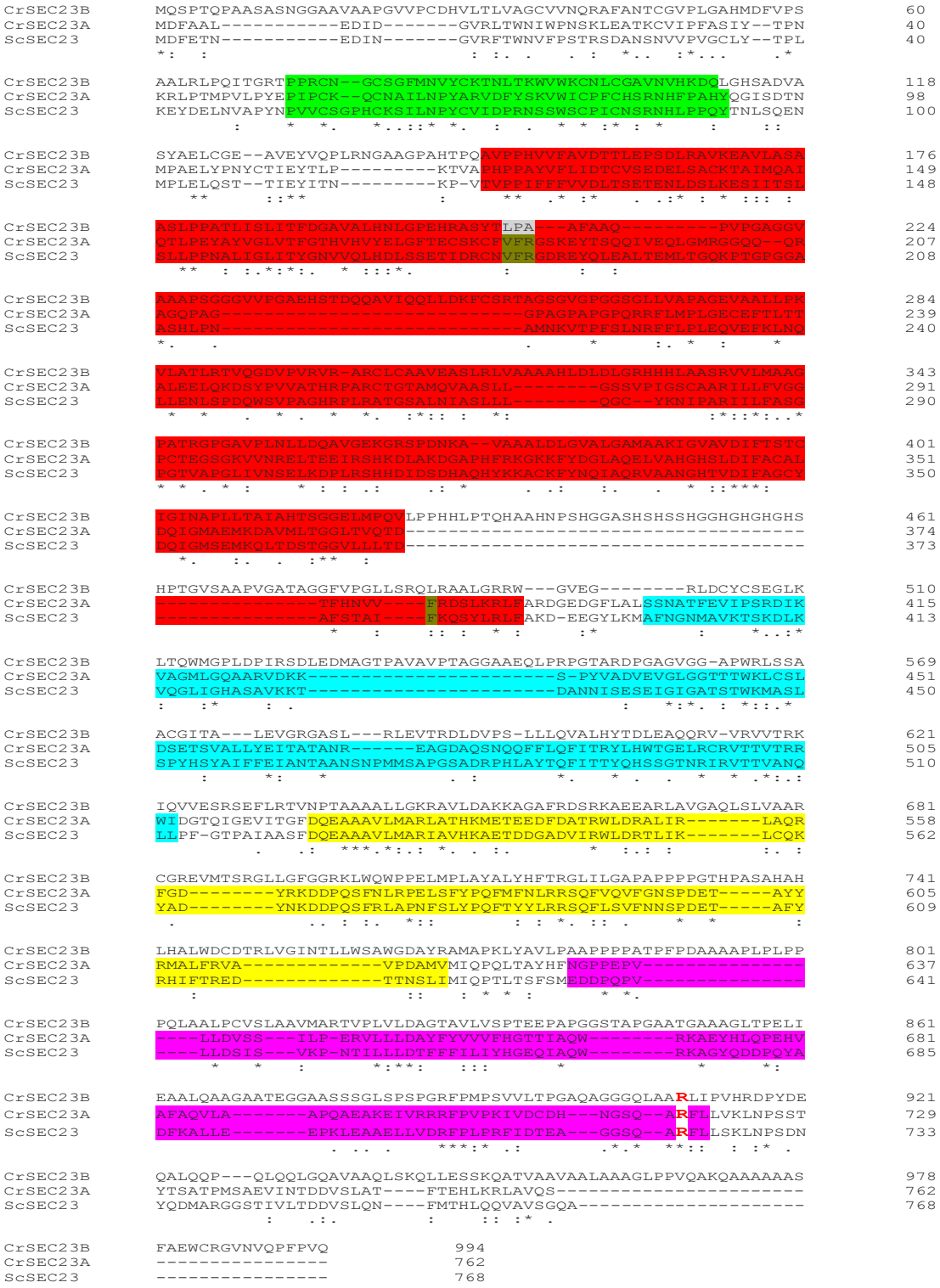

Figure 2. Alignment of *C. reinhardtii* SEC23 paralogs with *S. cerevisiae* ScSEC23p. Domains are highlighted in color as follows; Green, Zink Finger; Red, Trunk; Turquoise, Beta Sandwich; Yellow, Helical; Magenta, Gelsolin Repeat. The SEC24 binding motif VFR is highlighted in dark green and the Fenilalanine important for SEC31 binding is shown with an arrowhead. The arginine (R) of the active site is in red font.

Figure 3

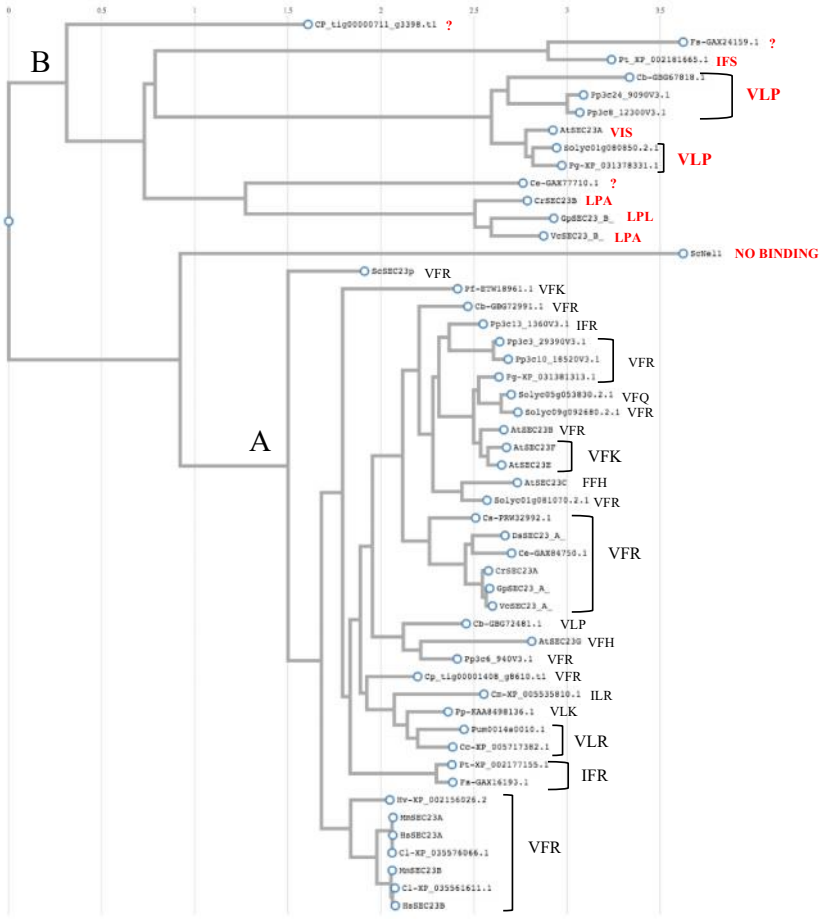

Figure 3. Phylogenetic analysis of SEC23 homologs. Known or putative (deduced from alignments) SEC24 binding motifs are given as three letter amino acid codes next to the name/accession of the proteins. Accession numbers:

Cyanophora Genome Database: *Cyanophora paradoxa* (tig00001408_g8610.t1, tig00000711_g3398.t1).

Phytozome accession numbers: *Physcomitrella patens* (Pp3c13_1360V3.1, Pp3c8_12300V3.1, Pp3c24_9090V3.1, Pp3c3_29390V3.1, Pp3c6_940V3.1, Pp3c10_18520V3.1), *Porphyra umbilicalis* (Pum0014s0010.1), *Solanum lycopersicum* (Solyc05g053830.2.1, Solyc01g081070.2.1, Solyc09g092680.2.1).

NCBI accession numbers: *Canis lupus dingo* (XP_035576066.1, XP_035561611.1) *Chara braunii* (GBG72481.1, GBG72991.1, GBG67818.1) *Chlamydomonas eustigma* (GAX84750.1, GAX77710.1), *Chlorella sorokiniana* (PRW32992.1), *Chondrus crispus* (XP_005717382.1) *Cyanidioschyzon merolae* (XP_005535810.1), *Dunaliella salina* (DsSEC23(A)*: KAF5842832.1), *Fistulifera solaris* (GAX16193.1, GAX24159.1), *Gonium pectorale* (GpSEC23(A)*; KXZ56797.1, GpSEC23(B)*; KXZ47553.1), *Hydra vulgaris* (XP_002156026.2) *Phaeodactylum tricornutum* (XP_002177155.1 , XP_002181665.1), *Plasmodium falciparum* (ETW18961.1), *Porphyridium purpureum* (KAA8498136.1), *Punica granatum* (XP_031381313.1 , XP_031378331.1), *Saccharomyces cerevisiae* Nel1p (QHB09024.1).

*: These proteins are not designated as (A) or (B) isoform in databases but as (A) or (B) in this study according to the phylogenetic relationship with CrSEC23A or CrSEC23B.

?: Residues aligning with VFR of ScSEC23p is very different than VFR and LPA/LPL.

The remaining accession numbers are given in Table 1.

Additional sequence analyses have demonstrated that the VFR motif is absent from SEC23 homologs in several other organisms; *Solanum lycopersicum* paralogs have VFK (Solyc01g081070.2.1), VFR (Solyc05g053830.2.1), VLP (Solyc01g080850.2.1) or VFQ (Solyc09g092680.2.1) in the position of the SEC24 binding motif (Figure 4B, highlighted with green background). We found that each of the red algae analyzed (*Cyanidioschyzon merolae, Chondrus crispus, Porphyra umbilicalis,* and *Porphyridium purpureum*) had

a single SEC23 sequence encoded on their genome. These red algal orthologs have SEC23A-like SEC24 binding motifs, which include ILR, VLR, VLR, and VLK, respectively (Figure 4C, highlighted with green background).

Figure'4

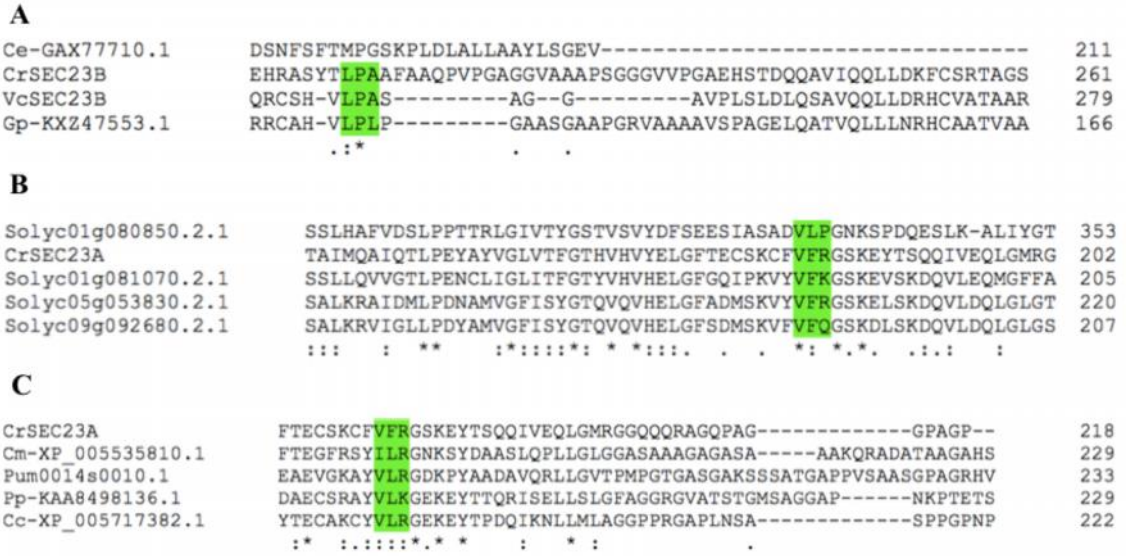

Figure 4. SEC24 binding motifs of SEC23 orthologs of various organisms. A. Alignment of CrSEC23B (CrSEC23B; Cre10.g447350), and putative orthologs in *Chlamydomonas eustigma* (GAX77710.1), *Volvox carteri* (VcSEC23B; Vocar.0070s0024.1) and *Gonium pectorale* (GpSEC23B; KXZ47553.1). B. Alignment of CrSEC23A (Cre03.g180850) and homologs in *Solanum lycopersicum* (Solyc05g053830.2.1, Solyc01g081070.2.1, Solyc09g092680.2.1, Solyc01g080850.2.1). C. Alignment of red algal species, *Chondrus crispus* (XP_005717382.1), *Cyanidioschyzon merolae* (XP_005535810.1), *Porphyra umbilicalis* (Pum0014s0010.1) and *Porphyridium purpureum* (KAA8498136.1).

A phylogeny of the various SEC23 proteins from various organisms and that indicates the sequence of the putative SEC24 binding motifs are shown in Figure 3.

The alignments that we analyzed also identified the arginine potentially critical for catalytic activity (Figure 2, plus additional alignments not shown); all of the homologs analyzed in this study have a putative catalytic arginine which suggests SEC23-like function.

***C. reinhardtii* SEC23 Orthologs are Phylogenetically Distant:** Phylogenetic analyses were used to understand more about the relatedness of CrSEC23A and CrSEC23B and their relationship with other SEC23 orthologs. From these analyses we observed that SEC23 orthologs cluster into two groups. Based on their relationship with CrSEC23A or CrSEC23B, we designated them group A (conventional, A-like orthologs) and group B (unconventional, B-like orthologs) (Figure 3). Interestingly all the A-like orthologs that we analyzed had all five of the SEC23-associated domains while the B-like orthologs never had the gelsolin repeat domain and some only had the zinc finger and trunk domains (Figure 1, plus additional analysis not shown). While CrSEC23A and CrSEC23B are phylogeneticaly distant from each other, they each have closely related homologs in *D. salina, V. carteri* and *G. pectoral*e (Figure 3). Interestingly, species of Rhodophyta and Opisthokonta only have paralogs in group A. We were able to detect at least one A-like and one B-like ortholog in all Chlorophyta examined except for *Chlorella sorokiniana* and *Dunaliella salina* (Figure 3); *Chlorella variabilis* and *Chlorella Beijerinck* 1890 also only had A-like SEC23As (data not shown). Not all of the SEC23 homologs analyzed have a VFR SEC24 binding motif, although in many cases the motif present in the same position of the protein is similar to VFR (VFK, IFR, VFQ, VFH, ILR), similar to the LPA motif (LPL), or different than both (IFS, VIS, FFH, VLP) (Figure 3). ScNel1p which was reported to not bind SEC24 (Kodera *et al*. 2014), appears to be distantly related to the other SEC23 proteins (Figure 3). These results suggest that in addition to CrSEC23A and CrSEC23B, other species also have SEC23-like sequences with potentially distinct functions.

Homology Modeling of *C. reinhardtii* SEC23 Paralogs Suggest a Similar Structure to the *S. cerevisiae* Ortholog: The VFR motif in the Protein Database, structure 1m2o, is on a beta strand of the trunk domain

(Bi *et al*. 2002), as shown in Figure 5A and Figure 6 (visualization of 1m2o with Jmol viewer). Although the protein alignments suggest that the LPA motif is present in the CrSEC23B trunk domain (Figure 2), we generated a homology model of both CrSEC23A and CrSEC23B and determined the position of the motif in the three-dimensional model. The Phyre2 program modelled both CrSEC23A and CrSEC23B with 100% confidence using 1m2o as the template. The CrSEC23A (Figure 5C) and CrSEC23B (Figure 5D) models are very similar to that of ScSec23p (compare Figure 5A with Figure 5C and 5D). The VFR in CrSEC23A is in the same region of the molecule as VFR in ScSec23p (compare the position of these residues in Figure 5A and Figure 5C). The LPA motif of CrSEC23B is also in this region, (compare the position of these residues in Figure 5A and 5D). The position of the catalytic arginine (Arg722) in ScSec23p (Figure 5A) and the putative catalytic arginine of CrSEC23A (Arg718) (Figure 5C) and CrSEC23B (Arg910) (Figure 5D) are indicated in the figures in spacefill models and white color. The ScNel1p three-dimensional model also resembles the structure of ScSec23p (Figure 5B). However, CrSEC23B and ScNel1p three-dimensional models are slightly different, especially concerning the position of the putative SEC24 binding motifs (INE in ScNel1p). These results suggest that SEC23A and SEC23B form similar three-dimensional structures with the VFR and LPA motifs located at the same position in both the *S. cerevisiae* and *C. reinhardtii* proteins. Replacement of the VFR by LPA might (1) change the binding affinity of CrSEC23B to CrSEC24, (2) allow it to interact with other proteins associated with the vesicle coat, or (3) confer functionality to the protein that is not associated with COPII activity. This motif may have a similar role in VcSEC23B and the GpSEC23B. The presence and similar positions of the catalytic arginine in CreSEC23A and CrSEC23B structures suggest that both might have GAP activity (see discussion).

Figure 5

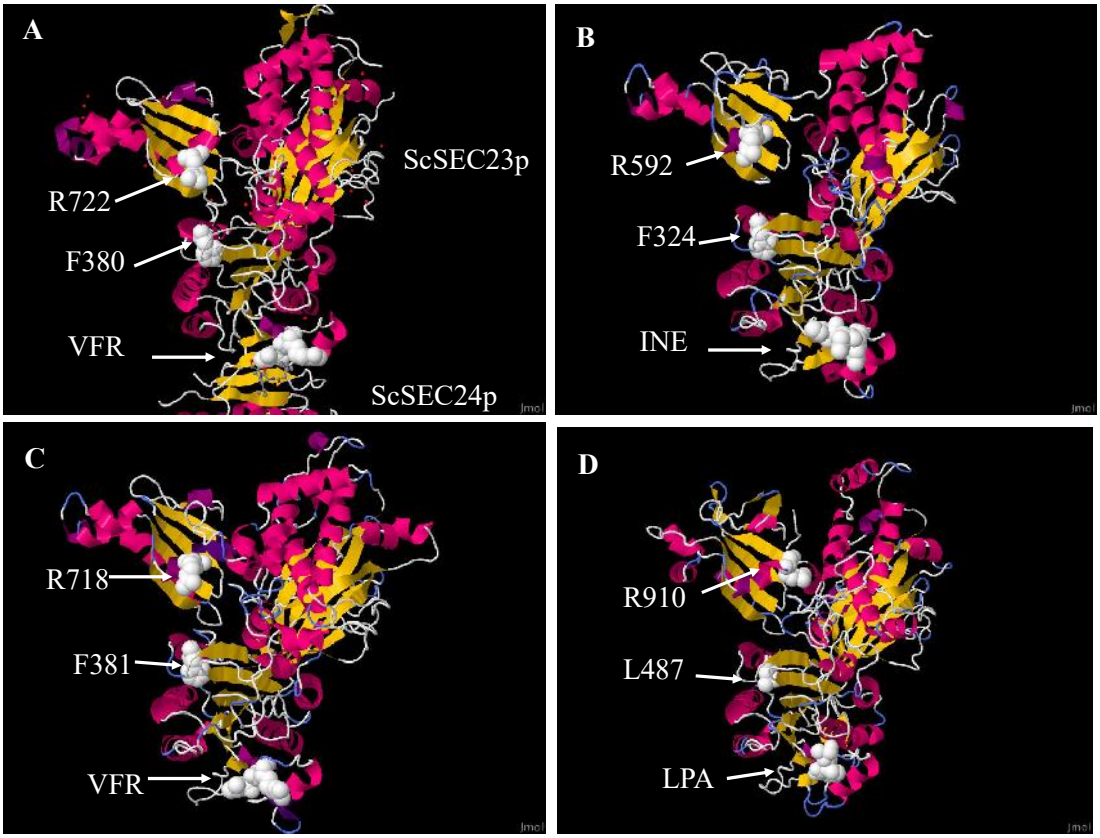

Figure 5. Homology modeling of CrSEC23 paralogs and ScNel1p. A. Yeast Sec23p/SEC24 complex; Jmol view of PDB structure 1m2o (Bi *et al*. 2002). Conserved FLP motif of SEC24 which binds to VFR of SEC23 is shown in ball and stick model (see Figure 6 for magnification of this region). B. Jmol view of ScNel1p 3D model. INE motif (I141, N142, E143) is highlighted. C. Jmol view of CrSEC23A 3D model D. Jmol view of CrSEC23B 3D model. White arrows identify the positions of the catalytic arginines, SEC24 binding motifs, and conserved Phe residue involved in SEC31 binding. All the conserved residues are highlighted in spacefill model in white color.

Figure'6

Figure 6: Conserved residues of ScSEC23p and ScSEC24p (Jmol view of PDB ID 1m2o). A. Catalytic arginine (R722) and a phenylalanine residue (F380) involved in binding to Sec31 are shown. Residues V181, F182 and R183 of ScSec23p are involved in binding to residues F385, L386 and P387 of ScSec24p. B. Enlargement of Sec23/Sec24 binding region shown in A (region shown with a square).

Both *CrSEC23A* and *CrSEC23B* are Expressed : To determine if the two *SEC23* genes of *C. reinhardtii* are expressed, we used RT-PCR. The primers used for the analysis of *CrSEC23A* anneal to exon 10 and the 3' UTR, generating a product of 765 bp (Figure 7). The primers used for RT-PCR analysis of *CrSEC23B* anneal to exons 10 and 11, generating a product of 209 bp (Figure 7). Sequencing analysis showed that both RT-PCR products had the expected sequence, demonstrating expression of both genes in *C. reinhardtii.* We also checked the results of published RNA-seq analyses of *C. reinhardtii* (González -Ballester *et al*. 2010; Castruita *et al*. 2011) which also demonstrated expression of *CrSEC23A* (Cre03.g180850) and *CrSEC23B* (Cre10.g4473500) (http://genomes.mcdb.ucla.edu/CreSulfur/). These genes appear to be constitutively expressed under a variety of growth conditions (see discussion).

Figure 7

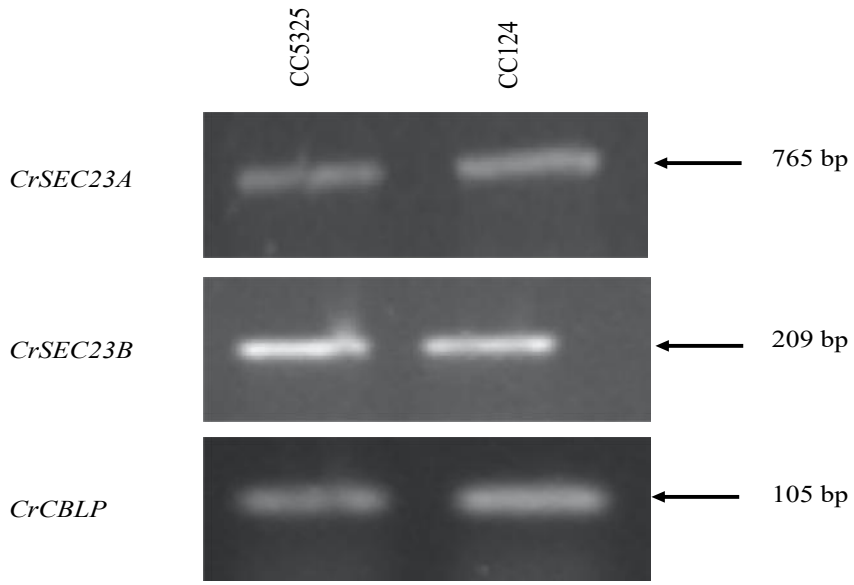

Figure 7. RT-PCR analysis of Cr*SEC23A,* Cr*SEC23B* and Cr*CBLP* transcripts to examine expression in CC5325 and CC124 (wt) strains of *C. reinhardtii.*

DISCUSSION

COPII is involved in generating vesicles from the ER membranes and carrying cargo from the ER to the Golgi (Yoshihisa *et al*. 1993; Matsuoka *et al*. 2001; Barlowe and Miller, 2013). The SEC23/SEC24 subcomplex and SAR1 form the first layer of the COPII coat. The *S. cerevisiae* genome contains a single SEC23 (ScSEc23p) homolog that associates with the COPII complex while a novel paralog (ScNel1p) has been identified that does not associate with this complex but still has in-vitro GAP (GTPase activating protein) activity (Kodera *et al*. 2014). There are seven and two SEC23 paralogs encoded in the *A. thaliana* (Aboulela *et al*. 2018) and *H. sapiens* (Khoriaty *et al*. 2018) genomes, respectively. We found two putative SEC23 paralogs, CrSEC23A (Cre03.g180850) and CrSEC23B (Cre10.g4473500), that are encoded on the genome of the unicellular model alga *C. reinhardtii*.

To obtain more information about the *C. reinhardtii* SEC23 proteins, we analyzed them using various bioinformatics resources. The two *C. reinhardtii SEC23* genes are located on different chromosomes; *CrSEC23A* on chromosome 3 and *CrSEC23B* on chromosome 10. While CrSEC23A and CrSEC23B amino acid sequences are only 18.9% identical to each other, they have similarities to orthologs in other organisms (Table 1). The domain architecture of CrSEC23A and CrSEC23B are different; CrSEC23B lacks the beta sandwich, helical, and gelsolin repeat domains while CrSEC23A has all five of the domains found in most SEC23 proteins of other organisms (Figure 1). The difference in domain architecture of CrSEC23A and CrSEC23B is also observed in SEC23 orthologs of other photosynthetic organisms (Aboulela *et al*. 2018). Aboulela *et al*. showed that SEC23A and SEC23D paralogs of *A. thaliana* have different domain architectures and are phylogenetically distant to other *A. thaliana* paralogs. This same study also showed that these two paralogs are involved in male gamete formation; other *A. thaliana* paralogs play no role in this process, suggesting distinct functions of the distantly related SEC23 paralogs in other systems. Interestingly, human and mouse genomes each contain two SEC23 paralogs which are highly identical and phylogenetically close (Figure 3). While human paralogs are 80% identical, mouse paralogs are 84.9% identical (data not shown). Experimental evidence shows that the functions of the mouse paralogs are interchangeable (Khoriaty *et al*. 2018).

The crystal structure of ScSec23p shows the position of the highly conserved VFR motif in the trunk domain, which is involved in binding SEC24 (Bi *et al*. 2002; Figure 6). While CrSEC23A has the conserved VFR motif (V182, F183 and R184), in CrSEC23B it appears to be replaced by a LPA motif (L209, P210 and A211) (Figure 2; Figure 5). We determined which SEC23 homologs in other organisms have the VFR, LPA or other motifs in the region associated with SEC24 binding. Most of the orthologs have the VFR motif (Figure 3), but in VcSEC23B it appears to be replaced by LPA and in one of the *G. pectorale* orthologs (GpSEC23B in this study) by LPL (Figure 3, Figure 4A).

According to the three-dimensional models, both SEC23 paralogs in *C. reinhardtii* have a very similar conformation to that of ScSec23p (Figure 5). In the PDB structure 1m2o, the configuration of the SEC23/SEC24 complex resembles a bow-tie while that of SEC23 alone has a Y shape (Bi *et al*. 2002). Modeling the SEC23 paralogs in *C. reinhardtii* suggests that both attain a Y shape (Figure 5). Interestingly, the LPA motif of CrSEC23B is at the same position within a beta strand of the trunk domain as that of the VFR motif in CrSEC23A, however the position of the motif looks slightly different (compare position of the motifs in Figure 5). This raises the question of whether CrSEC23B binds to SEC24, to another protein associated with COPII or is not associated with COPII.

Both *C. reinhardtii* SEC23 paralogs have the catalytic arginine (R718 in CrSEC23A and R910 in CrSEC23B) that is required to activate GTPase activity of SAR1 (Figure 2, Figure 5). Recent findings for *S. cerevisiae* ScNel1p, a SEC23 paralog that is not a component of COPII, shows that this protein still has in-vitro GAP activity for Sar1p (Kodera *et al*. 2014). ScNel1p also doesn't have a VFR motif, but instead has an INE motif that aligns with VFR of ScSec23p (Kodera *et al*. 2014). The INE motif of ScNel1p also is located in the same region of the three-dimentional model, however, its position looks slightly different (Figure 5B). Hence, SEC23 orthologs that do not bind SEC24 might also be present in other organisms, *e.g*. potentially CrSEC23B. The orthologs that we analyzed all have the catalytic arginine, but not all of them have a VFR motif; some have a motif similar to VFR while a variety of other potential motifs appear to occur (Figure 3). Therefore, it is likely that certain SEC23 paralogs in other organisms, especially those missing the VFR motif, are not involved in binding SEC24 and may even not be part of the COPII complex.

For most organisms that we analyzed there was more than one SEC23 paralog, with some similar to CrSEC23A (A-like orthologs) and others more similar to CrSEC23B (B-like orthologs). Interestingly all four Rhodophyta species and animal species that we analyzed only had A-like orthologs; while the Rhodophyta only had a single *SEC23* gene, human, mouse and dog had two *SEC23* genes which encode nearly identical polypeptides. Interestingly *H. vulgaris* had a single SEC23 paralog which is A-like (Figure 3).

RT-PCR analysis showed that the *C. reinhardtii* SEC23 paralogs are expressed under our standard growth

conditions (Figure 7). RNA-seq studies that focused on the responses of *C. reinhardtii* to various growth conditions, including deficiencies for sulfur (González - Ballester *et al*. 2010), copper (Castruita *et al*. 2011), nitrogen (Schmollinger *et al*. 2014) and phosphorus (Bajhaiya *et al*. 2016), showed little difference in expression of SEC23A and SEC23B under the nutrient-limited and nutrient-replete conditions. However, in response to H_2O_2 treatment *CrSEC23A* expression was elevated by almost 2-fold and *CrSEC23B* by over 2- fold (Blaby *et al*. 2015). The transcripts from both genes also increased by 1.5-2.0-fold following exposure of the cells to oxidative stress elicited by rose bengal treatment (Ma *et al*. 2020). Furthermore, in response to a light-dark diel cycle (Strenkert *et al*. 2019), *CrSEC23A* and *CrSEC23B* exhibited differential expression, with an increase in *CrSEC23A* expression of almost ten-fold and *CrSEC23B* expression of five-fold during the light period, and with overall expression of *CrSEC23A* higher than that of *CrSEC23B*. These results suggest that the demand for CrSEC23A and CrSEC23B function and the trafficking of vesicles may track faster growth, which occurs during the light phase of the diel cycle, and potentially also reflects the need to repair cell structures during oxidative stress.

These results suggest that B-like SEC23 proteins may be the result of a gene duplication and later divergence, but this duplication did not occur in Rhodophyta and the B-like sequence may have been lost in some Chlorophyta. Interestingly, in animals a gene duplication appears to have occurred, but the two genes did not diverge into A- and B-like sequences. Alternatively, genes encoding B-like proteins may have been present in the common ancestor involved in the primary endosymbiotic event, but the B-like sequence was lost in Rhodophyta, some species of Chlorophyta (Chlorella and Dunaliella) and animals (human, mouse, dog and hydra).

Conclusion: Our results show that *C. reinhardtii* has two *SEC23* paralogs (SEC23A and SEC23B) and the putative polypeptide sequences have low identity. Phylogenetic analysis placed the two paralogs in two clusters. Interestingly, many photosynthetic organisms have a divergent paralog. Interestingly, this divergence is not seen in animal species. These results suggest that divergent paralogs might have different functions; functions of the paralogs may not be interchangeable. Functional and mutational analysis should be done to understand more about this divergence which will be the focus of our future work.

Acknowledgements: This work was supported by BAP Coordination, Akdeniz University (grant number FBA-2018-4156 and FYL-2019-4901 to MA). ARG was supported by the Carnegie Institution for Science. The authors would like to thank the faculty members of the Department of Agricultural Biotechnology, Akdeniz University.

REFERENCES

Aboulela, M., T. Nakagawa, A. Oshima, K. Nishimura and Y. Tanaka (2018). The Arabidopsis COPII components, AtSEC23A and AtSEC23D, are essential for pollen wall development and exine patterning. J. Exp. Bot. 69: 1615–1633.

Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang J, Z. Zhang and W. Miller (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research 25(17): 3389–3402.

Bajhaiya, A.K., A.P. Dean, L.A.H Zeef, R.E. Webster and J.K. Pittman (2016). PSR1 is a global transcriptional regulator of phosphorus deficiency responses and carbon storage metabolism in Chlamydomonas reinhardtii. Plant Physiology 170: 1216–1234.

Barlowe, C., L. Orci, T. Yeung, M. Hosobuchi, S. Hamamoto, N. Salama, M.F. Rexach, M. Ravazzola, M. Amherdt and R. Schekman (1994). COPII: A membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell 77(6): 895–907.

Barlowe, C.K. and E.A. Miller (2013). Secretory protein biogenesis and traffic in the early secretory pathway. Genetics 193: 383–410.

Brandizzi, F. and C. Barlowe (2013). Organization of the ER-Golgi interface for membrane traffic control. Nature Reviews Molecular Cell Biology 14(6): 382–392.

Blaby, I.K., C.E. Blaby-Haas, M.E. Pérez-Pérez, S. Schmollinger, S. Fitz-Gibbon *et al*. (2015). Genome-wide analysis on Chlamydomonas reinhardtii reveals the impact of hydrogen peroxide on protein stress responses and overlap with other stress transcriptomes. Plant J. 84: 974–988.

Bonifacino, J.S. and J. Lippincott-Schwartz (2003). Coat proteins: Shaping membrane transport. Nature Reviews Molecular Cell Biology 4: 409–414.

Boyadjiev, S.A., J.C. Fromme, J. Ben, Chong SS, C. Nauta *et al*. (2006). Cranio-lenticulo-sutural dysplasia is caused by a SEC23A mutation leading to abnormal endoplasmic-reticulum-to-Golgi trafficking. Nature Genetics 38: 1192–1197.

Bi, X., R.A. Corpina and J. Goldberg (2002). Structure of the Sec23/24–Sar1 pre-budding complex of the COPII vesicle coat. Nature 419(6904): 271–277.

Bi, X., J.D. Mancias and J. Goldberg (2007). Insights into COPII Coat Nucleation from the Structure of Sec23•Sar1 Complexed with the Active

Fragment of Sec31. Developmental Cell 13: 635–645.

Castruita, M., D. Casero, Karpowicz SJ, Kropat J, A. Vieler *et al*. (2011). Systems biology approach in Chlamydomonas reveals connections between copper nutrition and multiple metabolic steps. The Plant Cell 4: 1273–1292.

Chang, C.W., J.L. Moseley, D. Wykoff and A.R. Grossman (2005). The LPB1 gene is important for acclimation of Chlamydomonas reinhardtii to phosphorus and sulfur deprivation. Plant Physiology 138(1): 319–329.

Chen, X., C. VanValkenburgh, H. Liang, H. Fang and N. Green (2001). Signal Peptidase and Oligosaccharyltransferase Interact in a Sequential and Dependent Manner within the Endoplasmic Reticulum. J. of Biological Chemistry 276: 2411–2416.

Cross, R. and J.G Umen (2015). The Chlamydomonas Cell Cycle. Plant J. 82: 370–92.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A *et al*. (2019). The Pfam protein families database in 2019. Nucleic Acids Research 47: D427–D432.

Faini, M., R. Beck, F.T. Wieland FT and J.A. Briggs (2013). Vesicle coats: Structure, function, and general principles of assembly. Trends in Cell Biology 23(6): 279–288.

Farquhar, R., N. Honey, S.J. Murant, P. Bossier, L. Schultz, D. Montgomery, R.W. Ellis, R.B. Freedman and M.F. Tuite (1991). Protein disulfide isomerase is essential for viability in Saccharomyces cerevisiae. Gene 108: 81–89.

Fromme, J.C., M. Ravazzola, S. Hamamoto, M. Al-Balwi, W. Eyaid W, S.A. Boyadjiev, P. Cosson, R. Schekman and L. Orci (2007). The Genetic Basis of a Craniofacial Disease Provides Insight into COPII Coat Assembly. Developmental Cell 13: 623–634.

González-Ballester, D., D. Casero, S. Cokus, M. Pellegrini, S.S. Merchant and A.R. Grossman (2010). RNA-Seq Analysis of Sulfur-Deprived Chlamydomonas Cells Reveals Aspects of Acclimation Critical for Cell Survival. The Plant Cell 22(6): 2058–2084.

Gorman, D.S. and R.P. Levine (1965). Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of Chlamydomonas reinhardi. Proceedings of the National Academy of Sciences of the United States of America 54(6): 1665–1669.

Harris, H.E., D.B. Stern DB and G.B. Witman (2009). The Chlamydomonas Sourcebook. 2nd Ed. Cambridge, MA, USA: Academic Press.

Huerta-Cepas, J., F. Serra and P. Bork (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Molecular Biology and Evolution 33(6): 1635–1638.

Kelley, L.A., S. Mezulis, C.M. Yates, M.N. Wass and M.J.E. Sternberg MJE (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10: 845–858.

Khoriaty, R., G.G. Hesketh, A. Bernard, A.C. Weyand, D. Mellacheruve *et al*. (2018). Functions of the COPII Gene Paralogs SEC23A and SEC23B Are Interchangeable in Vivo. Proceedings of the National Academy of Sciences of the United States of America 115 (33): E7748–57.

Kirchhausen, T. (2000). Three ways to make a vesicle. Nature Reviews Molecular Cell Biology 1(3): 187-198.

Kodera, C., T. Yorimitsu and K. Sato (2014). Sec23 homolog Nel1 is a novel GTPase-activating protein for Sar1 but does not function as a subunit of the coat protein complex II (COPII) coat. J. Biological Chemistry 289(31): 21423–21432.

Lee, M.C.S., E.A. Miller, J. Goldberg,L. Orci and R. Schekman (2004). Bi-directional protein transport between the ER and Golgi. Annual Review of Cell and Developmental Biology 20: 87–123.

Ma, X., B. Zhang, R. Miao, X. Deng X, Y. Duan *et al*. (2020). Transcriptomic and physiological responses to oxidative stress in a chlamydomonas reinhardtii glutathione peroxidase mutant. Genes 11(463).

Matsuoka, K., L. Orci, M. Amherdt, S.Y. Bednarek, S. Hamamoto, R. Schekman and T. Yeung (1998). COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes. Cell 9: 263–275.

Miller, E.A., T.H. Beilharz, P.N. Malkus, M.C.S. Lee, S. Hamamoto, L. Orci and R. Schekman (2003). Multiple cargo binding sites on the COPII subunit Sec24p ensure capture of diverse membrane proteins into transport vesicles. Cell 114: 497–509.

Nakano, A. and M. Muramatsu (1989). A novel GTP-binding protein, sar1p, is involved in transport from the endoplasmic reticulum to the Golgi apparatus. J. Cell Biology 109: 2677–2691.

Peotter, J., W. Kasberg, I. Pustova and A. Audhya (2019). COPII mediated trafficking at the ER/ERGIC interphase. Traffic 20: 491–503.

Price, M.N., P.S. Dehal and A.P. Arkin AP (2009). Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. Molecular Biology and Evolution 26(7): 1641–1650.

Schmollinger, S., T. Mühlhaus, N.R. Boyle, I.K. Blaby, D. Casero *et al*. (2014). Nitrogen-sparing

mechanisms in Chlamydomonas affect the transcriptome, the proteome, and photosynthetic metabolism. The Plant Cell 26: 1410–1435.

Strenkert, D., S. Schmollinger, S.D. Gallaher, P.A. Salomé, S.O. Purvine *et al*. (2019). Multiomics resolution of molecular events during a day in the life of Chlamydomonas. Proceedings of the National Academy of Sciences of the United States of America 116(6): 2374–2383.

Walter, P., R. Gilmore and G. Blobel (1984). Protein translocation across the endoplasmic reticulum. Cell 38(1): 5–8.

Yoshihisa, T., C. Barlowe and R. Schekman (1993). Requirement for a GTPase-activating protein in vesicle budding from the endoplasmic reticulum. Science 259(5100): 1466–8.