The Origin and Evolution of the Plant Cell Surface: Algal Integrin-Associated Proteins and a New Family of Integrin-Like Cytoskeleton-ECM Linker Proteins

Burkhard Becker[1,*], Jean Michel Doan[1], Brandon Wustman[1], Eric J. Carpenter[2], Li Chen[3], Yong Zhang[3], Gane K.-S. Wong[2,3,4], and Michael Melkonian[1]

[1]Biozentrum Köln, Botanical Institute, Universität zu Köln, Germany

[2]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

[3]BGI-Shenzhen, Bei Shan Industrial Zone, Shenzhen, China

[4]Department of Medicine, University of Alberta, Edmonton, Alberta, Canada

*Corresponding author: E-mail: b.becker@uni-koeln.de.

Accepted: May 9, 2015

Abstract

The extracellular matrix of scaly green flagellates consists of small organic scales consisting of polysaccharides and scale-associated proteins (SAPs). Molecular phylogenies have shown that these organisms represent the ancestral stock of flagellates from which all green plants (Viridiplantae) evolved. The molecular characterization of four different SAPs is presented. Three SAPs are type-2 membrane proteins with an arginine/alanine-rich short cytoplasmic tail and an extracellular domain that is most likely of bacterial origin. The fourth protein is a filamin-like protein. In addition, we report the presence of proteins similar to the integrin-associated proteins α-actinin (in transcriptomes of glaucophytes and some viridiplants), LIM-domain proteins, and integrin-associated kinase in transcriptomes of viridiplants, glaucophytes, and rhodophytes. We propose that the membrane proteins identified are the predicted linkers between scales and the cytoskeleton. These proteins are present in many green algae but are apparently absent from embryophytes. These proteins represent a new protein family we have termed gralins for green algal integrins. Gralins are absent from embryophytes. A model for the evolution of the cell surface proteins in Plantae is discussed.

Key words: : gralin, filamin, actinin, Viridiplantae, Rhodophyta, Glaucophyta.

Introduction

Scaly green flagellates (prasinophytes) likely represent the ancestral stock of green algae from which all other green algae (and land plants = embryophytes) evolved (Nakayama et al. 1998; Rodriguez-Ezpeleta et al. 2007; Leliaert et al. 2012). The extracellular matrix (ECM) of most prasinophytes consists of small, mostly submicroscopic scales (Sym and Pienaar 1993). Scales are composed mainly of acidic carbohydrates with no or only little protein (scale-associated proteins [SAPs] $\leq$ 5%) present (Becker et al. 1990, 1994, 1996). Structurally and compositionally similar scales can be found in streptophytes and chlorophytes suggesting that this type of ECM had evolved in the last common ancestor of these two divisions of green plants (Viridiplantae); for example, all prasinophyte scales contain unusual 2-keto-sugar acids (Kdo, Dha, 5OMeKdo, Becker et al. 1991) and one structural scale type, the "pentagonal (square) scale" is present in both streptophytes and chlorophytes (Moestrup 1982).

Several times during evolution of the Viridiplantae, scales and their biosynthetic enzymes have been lost. For example, the sequenced chlorophyte genomes do not contain genes for Kdo-metabolism, and scales are absent from the Chlorophyceae and Trebouxiophyceae. In contrast, within streptophytes, scales are present on zoospores and/or gametes of streptophyte algae, except the conjugating green algae, which lost flagellate cells altogether (Lewis and McCourt 2004). However, the enzymes for Kdo metabolism (and probably Dha metabolism) have been retained in streptophyte algae and embryophytes (Royo et al. 2000; Becker et al. 2001). In the latter, Kdo and Dha are present in rhamnogalacturonan II, a component of cell wall pectin (Becker et al. 1994; Popper and Tuohy 2010).

Although the structure of the polysaccharide component of prasinophyte scales has been studied in some detail (Becker et al. 1995, 1998), no molecular information exists for SAPs. SAP involvement has been suggested in anchoring scales to the plasma membrane and possibly the cytoskeleton, because a special type of scale, termed a hair scale was found to be structurally linked to the axonemal cytoskeleton of flagella. For the flagellar scales of *Scherffelia dubia*, it was shown that most SAPs form two high-molecular mass complexes (Becker et al. 1996). Furthermore, SAPs are glycoproteins (containing high-mannose and complex N-linked glycans) that are cross-linked by disulfide bridges within the two complexes (Becker et al. 1996). Here, we present the first molecular characterization of four different SAPs. Three SAPs are putative membrane proteins representing a new protein family of membrane receptors connecting the cytoskeleton and the ECM analogous to animal integrins. The fourth protein is related to animal filamins, which mediate integrin-actin interaction in animals.

Materials and Methods

Cell Culture

Axenic cultures of *S. dubia* (Perty) Pascher emend. Melkonian et Preisig (CCAC 0019) were grown photoautotrophically in Waris-H medium (McFadden and Melkonian 1986), at 15°C, 70 µmol photon m^{-2} s^{-1} light, with a 14:10 h light:dark cycle. For DNA and RNA extraction, cells were transferred into a 1 l flask containing 400 ml Waris-H medium inside and aerated (0.5 l/min) under gentle agitation.

Mass Spectrometry of SAPs

Flagellar scales were isolated as described in Becker et al. (1996). The yield of protein was generally about 50 µg per 60 l of culture. SAP215, SAP126, SAP116, and SAP98 were isolated by two-dimensional (2D) sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) as described in Becker et al. (1996). Trypsin digestion and sequencing of Trypsin-digested peptides of SAP98, SAP116, SAP126, and SAP215 by time of flight tandem-MS was done at the Gesellschaft für Biologische Forschung (Braunschweig, Germany). Additional peptides for SAP98 and SAP116 were sequenced at the ZMMK (Cologne, Germany). Peptides identified as trypsin or keratin by MASCOT (Matrix Science) search were considered contaminants and discarded.

DNA and RNA Isolation

DNA and RNA were isolated form *S. dubia* cultures grown for 8–14 days (cell density 2–5 × 10^6 cells/ml) using commercial kits according to the manufacturer's protocol. For DNA isolation, the following kits were used: DNEasy Plant Minikit (Qiagen), Wizard genomic DNA purification kit (Promega). The isolated genomic DNA was stored in 10 mM Tris/HCl pH = 8.0 and 1 mM EDTA (TE buffer) at −20°C, at a concentration approximately 100 ng/µl. DNA quality was examined by monitoring absorbance at 260 and 280 nm in a UV spectrophotometer (Eppendorf) and ethidium bromide visualization after agarose gel electrophoresis.

RNA was isolated using Trizol reagent (Invitrogen) and the manufacturer's protocol. Total RNA was finally resuspended in DEPC-H_2O. Purity and integrity were analyzed by UV spectroscopy and agarose gel electrophoresis.

cDNA Synthesis

First-strand cDNA used in reverse transcription polymerase chain reaction (RT-PCR) was synthesized either with Superscript II first-strand synthesis kit (Invitrogen) or Mu-Mlv reverse transcriptase (New England Biolabs), according to the manufacturer's instructions. For RACE-compatible cDNA synthesis, we used the GENERACER kit (Invitrogen) and Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions.

Polymerase Chain Reaction Methods

Standard PCR were run in an MWG Biotech thermal cycler or an Eppendorf Mastercycler using standard protocols. Unless otherwise specified, Taq polymerase (New England Biolabs) was used for Standard PCR.

For amplification with short degenerate primers (12mer), we used Taq beads (Amersham) and the following two-step PCR settings: initial denaturation 94°C for 3 min followed by 30 cycles of denaturation step 94°C for 15 s, annealing step 46°C for 20 s. DNA fragments were selected by visualization in agarose gel with ethidium bromide and purified by agarose gel extraction with QIAEXII beads (Qiagen). The DNA fragments were cloned into pGEM-Teasy vector (Promega).

For inverse PCR, 400 ng of genomic DNA was completely digested (10 µl reaction volume, 1 h at the optimal temperature) with 10 U of each of the following restriction enzymes: *Rsa*I, *Sau*3AI, *Eco*RI, *Bam*HI, and *Bgl*I. The restriction enzymes were heat inactivated, and the digested DNA was circularized by ligation with T4 DNA ligase (Gibco) after a 10X dilution (200 µl reaction volume). The ligase was heat inactivated, and the circularized DNA was purified with ethanol precipitation and resuspended in 20 µl dH_2O (~20 ng/µl). Standard PCR was used to amplify a SAP98 fragment using the following primers: iPCR forward: 5′-GACCAGGGSTTTAACTAC-3′, iPCR reverse: 5′-CAGAGATATGTCGTCRGC-3′. The PCR fragments were purified with a PCR purification kit (SV kit, Promega) and ligated in the pGem-T-Easy vector (Promega) before sequencing.

RACE PCR was performed according to the GENERACER kit (Invitrogen) using the generacer primers and the following gene-specific primer: 5′-ACRGCRGACGACATATCTCTGGAC CAGGGSTTTAACTAC-3′. The optimized PCR settings were 94°C for 3 min, 30x (94°C for 20 s, 65°C for 20 s, and 75°C for 210 s), 75°C for 5 min. The 3′-RACE product was

purified by gel extraction (Wizard SV, Promega), ligated overnight into the pGem-T-easy vector (Promega), and transformed in DH10β electro-competent cells.

Sequence Analysis

BLAST searches were performed either locally or at National Center for Biotechnology Information (NCBI) using the NCBI BLAST program. Proteins were characterized in situ using the InterProScan5 website at EBI and the CBS prediction servers at Center for Biological Sequence Analysis (Denmark, http://www.cbs.dtu.dk/services/, last accessed May 29, 2015). SOAP denovo-Trans generated assemblies were used for Blast searches of the 1KP data (Johnson et al. 2012). Although most of these searches had typical parameters, queries for peptides inferred from mass spectra were run with a maximum *e* value of 1,000, and ambiguities between leucine and isoleucine bases were handled by searching against all combinatoric possibilities from substituting both bases into each ambiguous position in the peptide.

Immunolocalization of SAP98

A DNA fragment representing the C-terminal 484 amino acids was cloned into the pETBlue-2 vector (Novagen) and transformed into Nova blue competent cells. Positive clones were confirmed by sequencing. After induction with IPTG, a protein of the right size was found in inclusion bodies. The protein was isolated by electro elution from a preparative SDS-PAGE as described by Becker et al. (1996) and used for immunization of two rabbits (Davids Biotechnologie, Regensburg, Germany). SDS-PAGE and Western blots were performed as described in Becker et al. (1996). Cells were fixed by adding the same volume of 8% formaldehyde in MT buffer (microtubule-stabilizing buffer: 15 mM HEPES/KOH pH 7, 7.5 mM KCl, 2.5 mM EGTA) to the cell suspension and leaving the cells 30 min on ice. The fixed cells were immobilized on poly-L-lysine coated multiwell slides and incubated with blocking solution (1% bovine serum albumin, 1% fish gelatin, and 0.5% Tween 20 in phosphate-buffered saline [PBS] for 30 min at 37 °C). Slides were washed with PBS (3x), incubated with the primary antibody in blocking solution (1.5 h at 37°C), washed with PBS (3x), and incubated with the secondary antibody in blocking solution for 1 h at 37 °C. After washing with PBS, 4 µl of mounting buffer was added. Slides were stored at 4 °C prior to microscopy. Fluorescence microscopy was performed using a Nikon Eclipse 800 (Nikon GmbH, Düsseldorf, Germany) microscope.

Results

Identification and In Silico Characterization of a SAP of 98 kDa (SAP98)

SAPs were isolated by 2D-gel electrophoresis (under non-reducing/reducing conditions) as described by Becker et al. (1996). SAPs were sequenced by mass spectrometry. Sequences of several peptides were obtained for each of four different SAPs (supplementary table S1, Supplementary Material online). Database searches (NCBI proteins and expressed sequence tags [ESTs]) failed to identify any protein showing similarity to more than a single SAP peptide in its sequence. Standard PCR using degenerate forward and reverse primers for each peptide also failed to isolate cDNA or genomic fragments coding for any SAP. However, amplification of a short fragment of known length representing a single peptide using short oligonucleotide primers was successful in a single case for a SAP98 peptide (fig. 1*A*).

Using inverse PCR and RACE techniques, it was possible to extend the SAP98 sequence (fig. 1*B*) yielding a 2,388 base pair sequence. RT-PCR amplification of the whole fragment from a preparation of total RNA using an outer primer pair confirmed the presence of mRNA with this sequence. Using this sequence, we searched the Thousand Plants (1KP) project (https://sites.google.com/a/ualberta.ca/onekp/, last accessed May 29, 2015) assembly database for *S. dubia*. A single contig contained the full sequence and allowed us to extend the SAP98 sequence for 172 base pairs in the 5′-direction (fig. 1*B*).

The longest open reading frame (ORF, 8–2,251) encodes a peptide of 748 aa. Five of the eight peptides align perfectly with the deduced protein sequence (fig. 1*B*, supplementary fig. S1*A*, Supplementary Material online) and cover the whole sequence, suggesting that the deduced sequence is indeed SAP98. In one peptide each, a W was interpreted as TT and a GA as Q in the mass spectra respectively. The latter are isobars and cannot be distinguished based on mass. The last peptide (VVTVSLPR) cannot be aligned well with the deduced amino acid sequence for SAP98 (fig. 1*B*). The predicted molecular size of the SAP98 peptide is 79.1 kDa. This is smaller than the apparent molecular mass determined from SDS-PAGE (98 kDa) (Becker et al. 1996). However, SAP98 was previously shown to be a glycoprotein containing N-linked glycans and indeed we found two potential N-glycosylation sites in the SAP98 sequence (table 1, fig. 1*C*). Furthermore, InterProScan 5 predicted a type 2 membrane protein with a short 11 aa cytoplasmic domain and a large 716 aa extracellular domain (table 1). A signal peptide (SP) overlapping with the trans-membrane domain (TMD) was also identified (fig. 1*C*). The predicted cytoplasmic domain is rich in alanine (45 mol%) and arginine (16 mol%). No similarities to other functional domains were found. BLASTP searches did not give any significant match (e value $< 10^{-10}$) for any other protein in the NR protein database at NCBI. However, the conserved domain scan detected a bacterial Ig-like domain (group 3, Pfam PF13754) between aa positions 450 and 475. Weak sequence similarity to bacterial cell surface proteins was observed in this region of the protein (aa 390–490, fig. 1*C*; ref|WP_012641554.1, 22 kDa adhesin protein of *Thermomicrobium roseum*, e value = 3e-05;

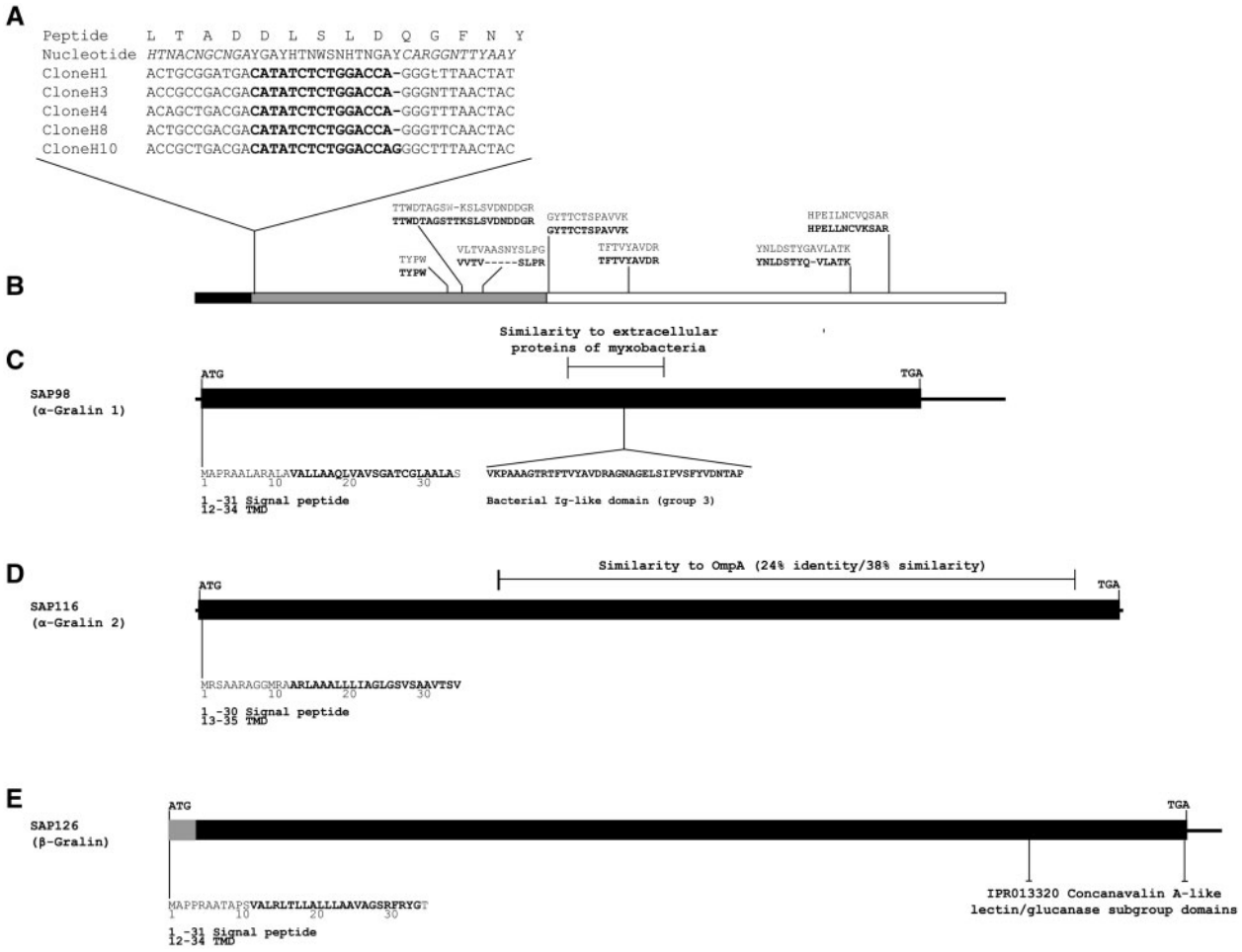

Fig. 1.—Cloning strategy of SAP98 and domain structure of SAP98, SAP116, and SAP126 from *Scherffelia dubia*. (*A*) Isolation of a small PCR fragment, containing the sequence information for a 14 aa peptide from SAP98. The peptide sequence, the corresponding degenerated nucleotide sequence, and the sequences of five clones are given. New sequence information is in bold and the primer region is in italics. (*B*) Extension of the short peptide sequence using inverse PCR (gray) and 3′-RACE (white), and database searches (1KP project, black). The position of peptide sequence obtained by mass spectrometry is given. (*C–E*) Domain structure of SAP98 (C), SAP116 (*D*), and SAP126 (E). The aa sequence of the N-terminal ends including the SP and the TMD is also given. Note: The N-terminal sequence of SAP126 (gray) is from the most similar sequence from *Tetraselmis striata*.

Table 1

In Silico Characterization of SAPs

Name (Apparent Size)	Number of aa	Predicted Size	Number of N-Glycosylation Sites and Position[a]	Number, Size, and Location of TMD[b]	SP (Size and Location)[b]	Protein Domains
SAP98 (α1-GRALIN	748	79.1 kDa	2 (324, 538)	1 (12–34)	1–31	None
SAP116 (α2-GRALIN)	963	102.5 kDa	4 (257, 654, 922, 934)	1 (13–35)	1–30	None
SAP 126 (β-GRALIN)	1,074	116.5 kDa	7 (87, 154, 371, 441, 645, 990, 1,010)	0 (1 [12–29]*)	N-terminus missing (1 -28*)	IPR013320
						IPR013320
SAP 215 (filamin-like)	2,483	264.1 kDa	Cytoplasmic	0	Complete no SP	IPR003344
						IPR003344
						IPR017868
						IPR017868
						IPR013783
						IPR013783
						IPR017868

[a]NetN-Glyc 1.0 server.
[b]InterProScan 5.
[c]Based on the most similar sequence from *Tetraselmis striata*.

ref|WP_002638558.1, OmpA domain protein of *Maxococcus* sp., *e* value: 0.015). The more sensitive delta-BLASTP search confirmed this similarity (40–60% query coverage, 24–30% identity, *e* value: 5e-10 to 3e-14).

Subcellular Localization of SAP98

The subcellular localization of SAP126, a member of the large high-molecular weight complex (>2MDa) is known (Becker et al. 1996); however, the localization of the second SAP complex of 190 kDa, which includes SAP98, has not yet been determined. To address this question, the C-terminal part of SAP98 (SAP98-C, 484 amino acids) was expressed in bacteria (not shown). SAP98-C was purified by SDS-PAGE from inclusion bodies and subsequently used for preparation of a polyclonal antibody. On Western blots (fig. 2), the antibody recognized a protein of 98 kDa (80 kDa after deglycosylation with N-Glykosidase F). The apparent molecular mass of the deglycosylated protein (SDS-PAGE) is in very good agreement with the predicted molecular size (79.1 kDa) for SAP98. Immunofluorescence revealed that SAP98 is localized to the flagella (fig. 3). Often the fluorescence was located in two clearly distinguishable rows (fig. 3), suggesting that SAP98 may localize to the basal part of the flagellar hair scales, which are arranged in two opposite rows perpendicular to the plane of flagellar beat (fig. 4*A*). Attempts to localize the protein at the ultrastructural level were unsuccessful.

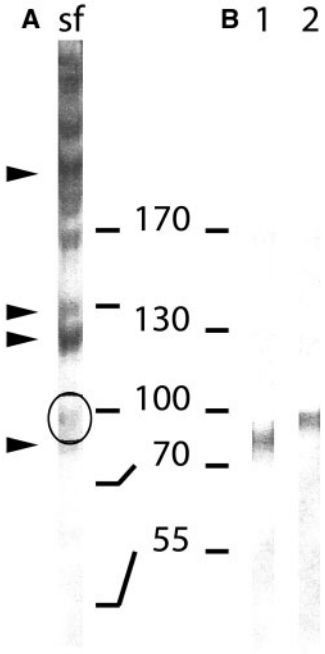

Fig. 2.—SDS-PAGE and Western blot of isolated flagellar scales from *Scherffelia dubia* probed with anti-SAP98. (*A*) Coomassie Brilliant Blue stain of an isolated flagellar scale fraction (sf) run with a 12% SDS-PAGE. Circle, SAP98. The other SAPs for which peptide sequences were obtained are indicated by arrow heads. (*B*) Western blot of the same flagellar scale fraction probed with a polyclonal anti-SAP98 (8% SDS-PAGE). Lane 1 flagellar scale fraction after deglycosylation with N-Glycosidase F and lane 2 isolated flagellar scale fraction. The position of the prestained marker proteins is indicated in between both PAGE.

Identification and In Silico Characterization of Three Additional SAPs

Given that all SAP98 peptide sequences could be found in a single *Scherffelia* contig obtained from the 1KP project database, we searched all contigs of *Scherffelia* in the 1KP database for proteins similar to the peptide sequences that were obtained for the other SAPs. Indeed, the other three SAPs were readily identified. Table 1 summarizes the in silico characterization of these proteins. A more detailed summary displaying the SP, TMD, potential N-glycosylation sites, and showing the alignment of the peptides with each protein sequence can be found in supplementary fig. S1*A–D*, Supplementary Material online. In the following, we will briefly describe the three proteins.

SAP116

Like SAP98, SAP116 is a type 2 membrane glycoprotein (predicted size 102.5 kDa, 963 aa) with four potential N-glycosylation sites in the predicted extracellular domain. The overall amino acid composition is very similar to SAP98 (supplementary table S2, Supplementary Material online, ~35 mol% Ala, Thr, and Ser). No known protein domains could be identified. All 12 MS peptides could be aligned to the deduced amino acid sequence (supplementary fig. S1*B*, Supplementary Material online); however, only five peptides could be aligned perfectly, the others showed at least one mismatch. In one case, a tryptophan was interpreted as a

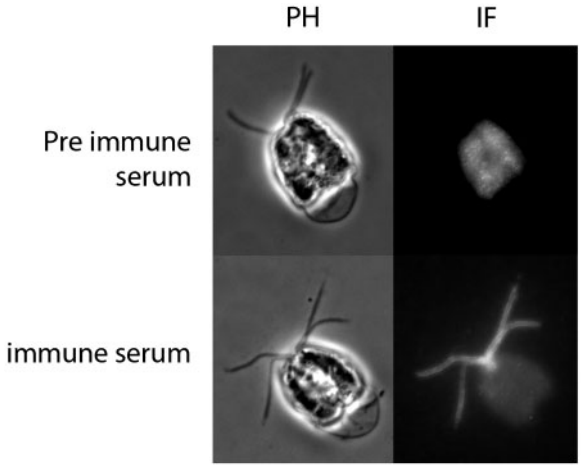

Fig. 3.—Immunofluorescence microscopy of *Scherffelia dubia* cells labeled with anti-SAP98. Only the immune serum shows a flagella labeling. Because of autofluorescence, the chloroplast is visible in both preparations. PH, phase contrast control; IF, immunofluorescence.

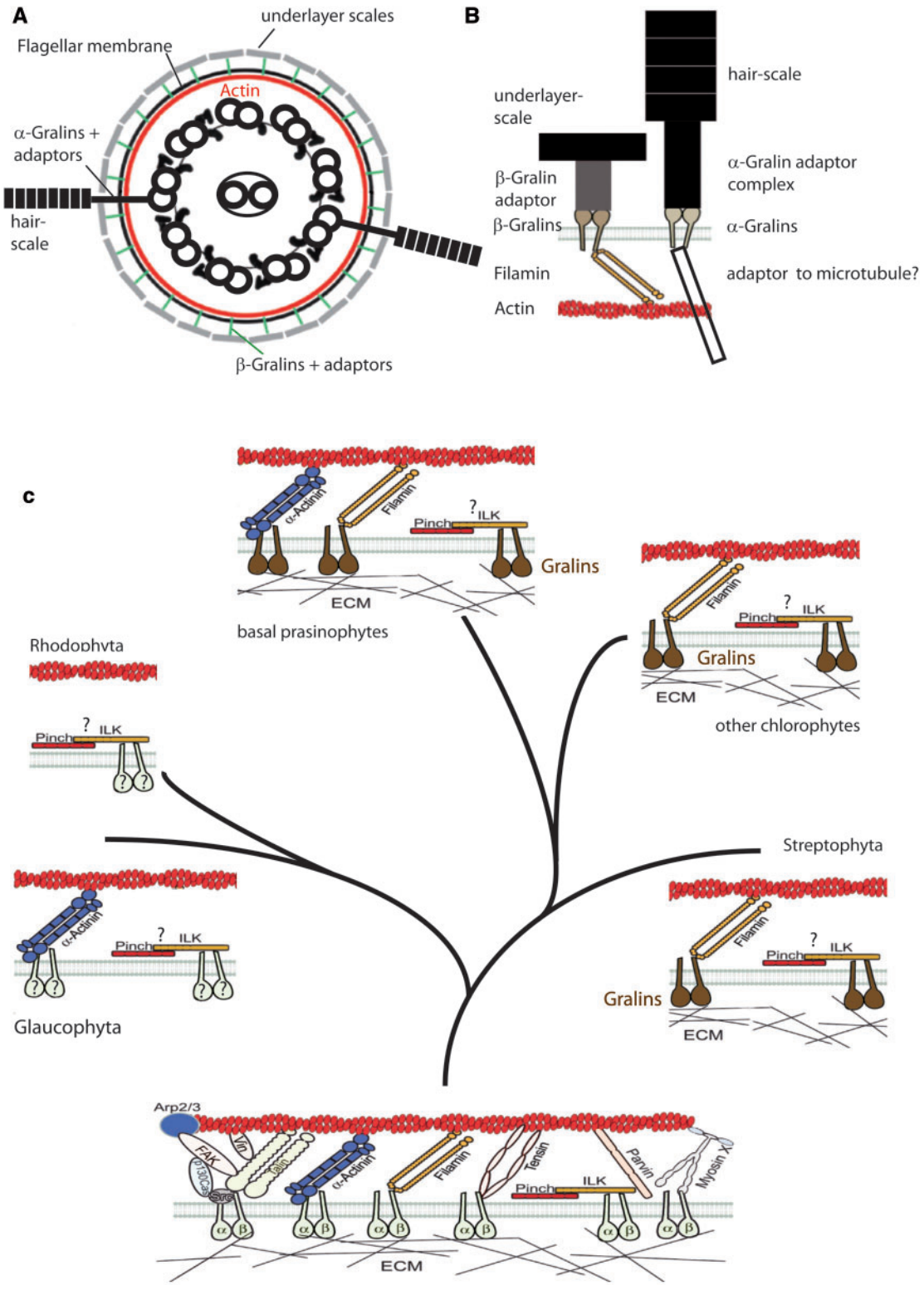

Fig. 4.—Structure and evolution of the prasinophyte cell surface. (*A*) Arrangement of pentagonal and hair scales on the flagellar surface of the Chlorodendrophyceae showing the link between hair scales (only few subunits shown) and the B-microtubules of doublets 4 and 8. (*B*) Cartoon depicting the proposed scale-Gralin-cytoskeleton interactions. (*C*) Evolution of the cell surface in glaucophytes, rhodophytes, and viridiplants excluding embryophytes (in the latter Gralins and filamin were lost). The structure of the ancestral cell surface is based on figure 1 in Roca-Cusachs et al. (2012). Components not found in the Plantae are presented in light colors (e.g., talin, parvin, and myoxin X).

valine-serine dipeptide in the mass spectrum. Like SAP98, SAP116 has only a short N-terminal cytosolic domain rich in alanine and arginine (33 mol% and 25 mol%, fig. 1*D*). The contig shows some similarity to extracellular Omp proteins from Myxobacteria (e.g., ref|WP_011555552.1, cell envelope biogenesis protein OmpA, *e* value 8e-17), which also contain a bacterial Ig-like domain; however, the conserved domain scan did not detect this domain in SAP116. Again the similarity between SAP116 and Omp proteins could be confirmed using delta-BLASTP (query cover up to 92%, identity up to 27%, *e* value up to 5e-24).

SAP126

SAP 126 is the only SAP for which we most likely do not have the complete protein from *S. dubia*; the N-terminus is apparently missing. However, all peptides can be aligned to the sequence retrieved, representing more than 90% of the protein (supplementary fig. S1*C*, Supplementary Material online). The currently known sequence of *S. dubia* does not contain a TMD or SP. However, the 1KP database contains a putative close orthologous sequence from the related prasinophyte *Tetraselmis striata* (see also below and supplementary fig. S4, Supplementary Material online). The sequence from *T. striata* contains an ORF in which the *Scherffelia* sequence is extended for 28 aa at the N-terminus. The extension contains a SP and a TMD, suggesting that also SAP126 is a type 2 membrane protein (fig. 1*E*). As in SAP98 and SAP116 the 11 aa cytoplasmic domain is rich in alanine and arginine (36 mol% and 9 mol%). In addition, the cytoplasmic domain contains a large amount of proline residues (25 mol%). The predicted protein size for the extended protein is 116.5 kDa, which is in good agreement with the Mr determined for the deglycosylated protein (115 kDa) (Becker et al. 1996). The overall alanine, threonine, and serine content is lower than in SAP98 and SAP116 (~24%, supplementary table S2, Supplementary Material online). InterProScan5 found two potential concanavalin A-like domains in this protein, and the protein contains seven potential N-glycosylation sites (table 1). BLASTP searches revealed low homology to some bacterial proteins (e.g., exported hypothetical protein Photobacterium phosphoreum ANT220; sequence ID: emb|CDN97175.1)|; query cover 9%, *e* value 2e-08). The conserved domain scan of Delta-BLASTP also detected the concanavalin-like domain.

SAP215

The three peptides from SAP215 were found in single contig encoding for an ORF of 2,483 aa (predicted size 264.1 kDa). The protein does not contain a TMD or SP and is therefore most likely cytoplasmic. InterProScan5 predicted domains related to filamin at the N-terminus (aa 152–234) and the c-terminal half of the ORF. The most similar protein in BLASTP analyses using the NCBI protein database were filamin-related proteins from *Micromonas* (*e* value: 1e-017) and *Bathyococcus* (*e* value: 4e-008).

Search for Putative Homologs of SAPs in Other Organisms

The initial BLASTP searches of the NCBI database showed that proteins similar to SAPs are not common in the public databases, except for SAP215, which was readily identified as a filamin-like protein. However, tBLASTN searches of viridiplant EST databases indicated the presence of similar proteins in several green algae. We therefore searched the 1KP database, and all NCBI databases for transcripts similar to the four SAPs. Supplementary tables S3–S5, Supplementary Material online, show some of the BLAST results (percentage of identical [similar] amino acids of the largest aligned region to the protein from *S. dubia,* the *e* value). For several green algae, more than one contig displaying significant similarity to the *S. dubia* SAPs was found. All four SAPs are very well conserved within the class Chlorodendrophyceae (Chlorophyta). Although transcripts similar to SAP98 and SAP116 were found in "advanced" chlorophytes and streptophytes, SAP126 is apparently absent from "advanced" chlorophytes. The sequences from conjugating green algae (Zygnematophyceae) displaying similarity to the *S. dubia* sequence are most likely not homologs of SAPs (except SAP215) as they align only poorly to the other sequences even in the conserved regions (supplementary figs. S2–S4, Supplementary Material online).

Using putative homologs of SAP98, SAP116, SAP126, and SAP215 from *Pyramimonas parkeae*, *Chaetosphaeridium globosum*, or *Micromonas* sp., the list of algae expressing transcripts similar to SAPs could be further extended (not shown). Proteins similar to the filamin-like protein of *Micromonas* were found in all green algal groups, including conjugating green algae (e.g., scaffold-NBYP-2010987-Mesotaenium_kramstai *e* value 8e-53), although no clear homologs of the other SAPs could be found in this group. Using the NCBI database, SAPs were not detected in embryophytes, suggesting that SAPs, scales, and filamin-like proteins were completely lost in land plants. Gralins and filamin-like proteins were also not found in glaucophytes and rhodophytes.

Search for Other Integrin-Associated Proteins in Algae with Primary Plastids

Given the presence of a filamin-like protein in many green algae, we decided to search also for other integrin-associated proteins (integrin, α-actinin, talin, vinculin, paxillin, parvin, pinch, myosin X, and integrin-linked kinase, see also fig. 4*C*) in viridiplants (including the completed embryophyte genomes), rhodophytes, and glaucophytes (supplementary table S6, Supplementary Material online). Putative homologs were detected for α-actinin in glaucophytes and a few basal prasinophytes (supplementary table S6, Supplementary Material online). LIM-domain containing proteins (which

include paxillin and pinch proteins) can be found in all groups (supplementary table S6, Supplementary Material online). Similarly, transcripts/proteins similar to integrin-linked kinase were detected in BLAST analyses using rat ILK (NP_596900.1) as query (>80% query cover to exclude proteins displaying only similarity in the kinase domain) in all investigated groups (not shown).

Discussion

In this study, four SAPs from *S. dubia* have been characterized. Three of these are trans-membrane proteins, suggesting that they are the receptors linking scales (the ECM of prasinophytes) to the plasma membrane and cytoskeleton similar to integrins that link the ECM of animals to the actin cytoskeleton (fig. 4*A* and *B*). SAP126 had previously been shown to be located in a high-molecular weight complex that links one type of scale (i.e., pentagonal scales) to the flagellar membrane (Becker et al. 1996), and in this study, SAP98 was also localized to flagella, suggesting that SAP98 might function together with SAP116 as a second scale receptor, presumably for hair scales (fig. 4*B*). Indeed, these proteins share some properties with integrins. They display a similar size and domain organization as integrins (Campbell and Humphries 2011): a small cytoplasmic domain, a single TMD, and a large extracellular domain. Animal integrins form heterodimers (Campbell and Humphries 2011), and similarly SAP98 and SAP116 are present in a disulphide-linked heterodimeric complex (Becker et al. 1996). However, there are also important differences: 1) SAP98, SAP116, and SAP126 are type 2 membrane proteins, whereas integrins are type 1 membrane proteins, 2) SAPs do not contain any protein domains known to be present in animal integrins, and 3) the cytosolic domain of integrins is longer and contains several conserved binding motifs, for example, for filamin and talin (Anthis and Campbell 2011). None of these motifs can be found in the cytoplasmic domains of the SAPs. Comparison of the predicted sequence for the cytoplasmic tail of SAP126 shows large sequence variability for the cytoplasmic tails in different algae (e.g., supplementary fig. S4, Supplementary Material online, for SAP126).

The presence of a filamin-like protein in isolated SAP preparations further corroborates the similarity of SAPs to animal integrins, although an identifiable filamin-binding motif was not found in SAP98, SAP116, or SAP126. Either the filamin-like protein uses a different binding motif, or the true filamin receptor has not yet been found. It is therefore concluded that SAP98, SAP116, and SAP126 are not orthologs of animal integrins but could have replaced integrins functionally in the common ancestor of the Viridiplantae. The novel protein family is here termed Gralins (green algal integrins). SAP98 and SAP116 are overall more similar to each other than to SAP126 and are thus named α-Gralins (α1-Gralin [SAP98] and α2-Gralin [SAP116]), and SAP126 is termed β1-Gralin. We predict that future work will identify a β2-Gralin. 2D-PAGE revealed another SAP of approximately 116 kDa in the high-molecular complex that also contained SAP126 (Becker et al. 1996), which might be a good candidate for β2-Gralin.

The extracellular domains of α-Gralins are most likely of bacterial origin, as they show similarity to extracellular and Omp proteins from bacteria, whereas the extracellular domain of β-Gralins could be derived from a carbohydrate-binding domain (concanavalin A-like domain). Gralins and filamin-like proteins are well conserved in the Chlorodendrophyceae, but similar proteins occur also in other chlorophytes and in streptophyte algae but not in embryophytes (neither in red algae nor in glaucophytes). This is in agreement with earlier observations that have shown that embryophytes lack true homologs of classical adhesion proteins of animal cells including integrin, talin, vinculin, filamin, α-actinin, and tensin (Hussey et al. 2002, see also Baluska et al. 2003 for a discussion of the plant cytoskeleton–plasma membrane–cell wall continuum).

Some animal cell surface molecules or domains evolved prior to the evolution of Metazoa (e.g., integrins, fibrillin, and c-peptide domain of collagen, summarized by Ozbek et al. 2010). For example, integrins and many of its intracellular-associated proteins (fig. 4*C*) are found in the unicellular apusomonads (Sebé-Pedrós et al. 2010). Preliminary BLAST searches of the 1KP assembly database and the NCBI databases using the apusomonad or rat integrin, α-actinin, talin, vinculin, paxillin, parvin, pinch, and ilk sequences (Sebé-Pedrós et al. 2010) revealed transcripts similar to α-actinin, in basal prasinophytes and in glaucophytes. Proteins containing LIM-domains were found in all groups investigated, with many of them most similar to paxillin. However, it seems unlikely that these are true homologs as in general the conserved regions were restricted to the LIM domains. Similarly, we could detect transcripts/proteins similar to ilk in all groups investigated, but again it is currently not clear whether these are true homologs. No proteins similar to integrin, talin, parvin, and vinculin were detected. Given the presence of some putative integrin-associated proteins in green algae, red algae, and glaucophytes, it seems very likely that the ECM-Integrin-cytoskeleton system was present in the last common ancestor of algae with primary plastids (i.e., the Plantae, sometimes referred to as Archaeplastida). Although the monophyly of Plantae has been widely accepted, some recent reports challenged this conclusion (e.g., Kim and Graham 2008; Hampl et al. 2009); the branching pattern among the three groups of Plantae, however, is still controversial (e.g., Jackson and Reyes-Prieto 2014; Rockwell et al. 2014). The following discussion is based on the notion that Plantae are monophyletic but can be easily adapted to other phylogenetic scenarios. Based on the results presented here, the following scenario about the evolution of the plant cell surface is conceivable (fig. 4*C*). Upon establishment of primary plastids by endocytobiosis of a cyanobacterial cell, the ancient animal-like cell

surface was eventually transformed to the cellulose-type cell wall system of plants. The first green algae apparently used bacterial components (bacterial IgG fold in Gralins [this study], Kdo metabolism [see Introduction], and cellulose synthase [Salerno and Curatti 2003]) to evolve the typical cellulose wall of plants, a process which might have been driven by the excess carbohydrate generated by photosynthesis and using components of the ECM present in the photosynthesizing cyanobacterial and/or other bacterial endosymbiont(s) (Baum 2013). This transformation presumably took place in steps. The cell surface of the early diverging prasinophytes is already covered by complex, acidic carbohydrates forming scales; however, these organisms still use an integrin-like receptor (Gralins) and filamin (complemented with α-actinin and possibly paxillin-like and ilk-like proteins in some early-branching taxa) to connect the ECM to the cytoskeleton. Gralins were lost in "advanced" chlorophytes, and among the streptophytes in the Zygnematophyceae and embryophyte land plants together with the loss/transformation of scales, in the former accompanied by the loss of flagellate cells. The embryophytes later also lost filamin completely. Gralins and filamins are also absent in red algae and glaucophytes, which, however, have retained some other integrin-associated proteins, that is, α-actinin (glaucophytes) and possibly paxillin-like and ilk-like proteins (glaucophytes and rhodophytes), suggesting that the modification of the cell surface started before these lineages separated from the Viridiplantae.

An interesting question is why have some basal prasinophytes retained so many integrin-associated proteins? We note that the same genera reported here to have retained a large number of integrin-signaling components have recently been shown to be mixotrophic (able to phagocytose bacteria in addition to phototrophic nutrition; Bell and Laybourn-Parry 2003; Maruyama and Kim 2013; McKie-Krisberg and Sanders 2014), apparently the only known mixotrophic taxa among the Viridiplantae. The integrin-signaling pathway is also involved in phagocytosis (Dupuy and Caron 2008). Thus, it seems plausible that integrin-associated proteins have been retained in these basal prasinophytes to enable them to phagocytose. Further work is necessary to address the functional role of Gralins and associated proteins to test this hypothesis.

Supplementary Material

Supplementary tables S1–S6 and figures S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

J.D. and B.W. isolated the SAP proteins. All PCR experiments were done by J.D. Expression of SAP98 and immunofluorescence analysis were done by B.B. Sequencing within the 1KP project was done by L.C. and Y.Z. BLAST searches were done by B.B and E.C. B.B., E.C., and M.M. analyzed the data. The study was conceived by B.B., M.M., and G.W. B.B. and M.M. wrote the manuscript. The manuscript was read and approved by all authors. The authors thank Friederike Goebbels and Stefanie Junkermann for technical help with the expression, purification, and analysis of SAP98. We thank Dr H. Conradt for the initial mass spectometric analyses of SAPs. This work was supported by a fellowship from the "Graduate School of Functional Genetics" (funded by the state of North-Rhine Westphalia) to J.M.D. The 1000 Plants (1KP) initiative, led by G.K.-S.W., was supported by the Alberta Ministry of Enterprise and Advanced Education, Alberta Innovates Technology Futures, Innovates Centre of Research Excellence, Musea Ventures, and BGI Shenzhen.

Literature Cited

Anthis NJ, Campbell ID. 2011. The tail of integrin activation. Trends Biochem Sci. 36:191–198.

Baluska F, Samaj J, Wojtaszek P, Volkmann D, Menzel D. 2003. Cytoskeleton-plasma membrane-cell wall continuum in plants. Emerging links revisited. Plant Physiol. 133:482–491.

Baum D. 2013. The origin of primary plastids: a pas de deux or a ménage à trois? Plant Cell 25:4–6.

Becker B, Becker D, Kamerling JP, Melkonian M. 1991. 2-Keto-sugar acids in green flagellates - a chemical marker for prasinophycean scales. J Phycol. 27:498–504.

Becker B, Feja N, Melkonian M. 2001. Analysis of expressed sequence tags (ESTs) from the scaly green flagellate *Scherffelia dubia* Pascher emend. Melkonian et Preisig. Protist 152:139–147.

Becker B, Lommerse JPM, Melkonian M, Kamerling JP, Vliegenthart JFG. 1995. The structure of an acidic trisaccharide component from a cell wall polysaccharide preparation of the gren algae *Tetraselmis striata* Butcher. Carbohydr Res. 267:313–321.

Becker B, Marin B, Melkonian M. 1994. Structure, composition, and biogenesis of prasinophyte cell coverings. Protoplasma 181:233–244.

Becker B, Melkonian M, Kamerling JPP. 1998. The cell wall (theca) of *Tetraselmis striata* (Chlorophyta): macromolecular composition and structural elements of the complex polysaccharides. J Phycol. 34:779–787.

Becker B, Perasso L, Kammann A, Salzburg N, Melkonian M. 1996. Scale-associated glycoproteins of *Scherffelia dubia* (Chlorophyta) form high-molecular-weight complexes between the scale layers and the flagellar membrane. Planta 199:503–510.

Becker D, Becker B, Satir P, Melkonian M. 1990. Isolation, purification, and characterization of flagellar scales from the green flagellate *Tetraselmis striata* (Prasinophyceae). Protoplasma 156:103–112.

Bell EM, Laybourn-Parry J. 2003. Mixotrophy in the antarctic phytoflagellate, *Pyramimonas gelidicola* (Chlorophyta: Prasinophyceae). J Phycol. 39:644–649.

Campbell ID, Humphries MJ. 2011. Integrin structure, activation, and interactions. Cold Spring Harb Perspect Biol. 3:1–14.

Dupuy AG, Caron E. 2008. Integrin-dependent phagocytosis—spreading from microadhesion to new concepts. J Cell Sci. 121:1773–1783.

Hampl V et al. 2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups." Proc Natl Acad Sci U S A. 106:3859–3864.

Hussey PJ, Allwood EG, Smertenko AP. 2002. Actin-binding proteins in the *Arabidopsis* genome database: properties of functionally distinct plant actin-depolymerizing factors/cofilins. Philos Trans R Soc Lond B Biol Sci. 357:791–798.
Jackson CJ, Reyes-Prieto A. 2014. The mitochondrial genomes of the glaucophytes *Gloeochaete wittrockiana* and *Cyanoptyche gloeocystis*: multilocus phylogenetics suggests a monophyletic archaeplastida. Genome Biol Evol. 6:2774–2785.
Johnson MTJ et al. 2012. Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS One 7:e50226.
Kim E, Graham LE. 2008. EEF2 analysis challenges the monophyly of archaeplastida and chromalveolata. PLoS One 3:e2621.
Leliaert F et al. 2012. Phylogeny and molecular evolution of the green algae. CRC Crit Rev Plant Sci. 31:1–46.
Lewis LA, McCourt RM. 2004. Green algae and the origin of land plants. Am J Bot. 91:1535–1556.
Maruyama S, Kim E. 2013. A modern descendant of early green algal phagotrophs. Curr Biol. 23:1081–1084.
McFadden GI, Melkonian M. 1986. Use of Hepes buffer for microalgal culture media and fixation for electron microscopy. Phycologia 25:551–557.
McKie-Krisberg ZM, Sanders RW. 2014. Phagotrophy by the picoeukaryotic green alga *Micromonas*: implications for Arctic Oceans. ISME J. 8:2151.
Moestrup Ø. 1982. Phycological reviews 7. Phycologia 21:427–528.
Nakayama T et al. 1998. The basal position of scaly green flagellates among the green algae (chlorophyta) is revealed by analyses of nuclear-encoded SSU rRNA sequences. Protist 149:367–380.
Ozbek S, Balasubramanian PG, Chiquet-Ehrismann R, Tucker RP, Adams JC. 2010. The evolution of extracellular matrix. Mol Biol Cell. 21:4300–4305.
Popper ZA, Tuohy MG. 2010. Beyond the green: understanding the evolutionary puzzle of plant and algal cell walls. Plant Physiol. 153:373–383.
Roca-Cusachs P, Iskratsch T, Sheetz MP. 2012. Finding the weakest link—exploring integrin-mediated mechanical molecular pathways. J Cell Sci. 125:3025–3038.
Rockwell NC, Lagarias JC, Bhattacharya D. 2014. Primary endosymbiosis and the evolution of light and oxygen sensing in photosynthetic eukaryotes. Front Ecol Evol. 2:1–13.
Rodriguez-Ezpeleta N, et al. 2007. Phylogenetic analyses of nuclear, mitochondrial, and plastid multigene data sets support the placement of Mesostigma in the Streptophyta. Mol Biol Evol. 24:723–731.
Royo J, Gímez E, Hueros G. 2000. CMP–KDO synthetase: a plant gene borrowed from Gram-negative eubacteria. Trends Genet. 16:432–433.
Salerno GL, Curatti L. 2003. Origin of sucrose metabolism in higher plants: when, how and why? Trends Plant Sci. 8:63–69.
Sebé-Pedrós A, Roger AJ, Lang FB, King N, Ruiz-Trillo I. 2010. Ancient origin of the integrin-mediated adhesion and signaling machinery. Proc Natl Acad Sci U S A. 107:10142–10147.
Sym SD, Pienaar RN. 1993. The class Prasinophyceae. In: Raund FE, Chapman DJ, editors. Progress in phycological research. Bristol (United Kingdom): Biopress Ltd. p. 281–376.

Associate editor: Geoff McFadden