Protist, Vol. 167, 526–543, December 2016
http://www.elsevier.de/protis
Published online date 16 September 2016

Protist

ORIGINAL PAPER

The Cytoskeleton of Parabasalian Parasites Comprises Proteins that Share Properties Common to Intermediate Filament Proteins

Harald Preisner[a], Eli Levy Karin[b], Gereon Poschmann[c], Kai Stühler[c], Tal Pupko[b], and Sven B. Gould[a,1]

[a]Institute for Molecular Evolution, Heinrich-Heine-University, Düsseldorf, Germany
[b]Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv, Israel
[c]Molecular Proteomics Laboratory (MPL), BMFZ, Heinrich-Heine-University, Düsseldorf, Germany

Submitted February 10, 2016; Accepted September 2, 2016
Monitoring Editor: George B. Witman

Certain protist lineages bear cytoskeletal structures that are germane to them and define their individual group. Trichomonadida are excavate parasites united by a unique cytoskeletal framework, which includes tubulin-based structures such as the pelta and axostyle, but also other filaments such as the striated costa whose protein composition remains unknown. We determined the proteome of the detergent-resistant cytoskeleton of *Tetratrichomonas gallinarum*. 203 proteins with homology to *Trichomonas vaginalis* were identified, which contain significantly more long coiled-coil regions than control protein sets. Five candidates were shown to associate with previously described cytoskeletal structures including the costa and the expression of a single *T. vaginalis* protein in *T. gallinarum* induced the formation of accumulated, striated filaments. Our data suggests that filament-forming proteins of protists other than actin and tubulin share common structural properties with metazoan intermediate filament proteins, while not being homologous. These filament-forming proteins might have evolved many times independently in eukaryotes, or simultaneously in a common ancestor but with different evolutionary trajectories downstream in different phyla. The broad variety of filament-forming proteins uncovered, and with no homologs outside of the Trichomonadida, once more highlights the diverse nature of eukaryotic proteins with the ability to form unique cytoskeletal filaments.

Key words: Cytoskeleton; intermediate filament proteins; coiled-coils; convergent evolution; Trichomonadida; *Trichomonas vaginalis*.

[1]Corresponding author;
e-mail gould@hhu.de (S.B. Gould).

http://dx.doi.org/10.1016/j.protis.2016.09.001

Introduction

The compartmentalization of the eukaryotic cell and the dynamic nature of the endomembrane system rest upon an elaborate cytoskeleton (Koonin 2010). Actin and tubulin, the major components of the cytoskeleton, are well conserved in sequence and function among all eukaryotic supergroups (Kusdian et al. 2013; Wickstead and Gull 2011). In addition to actin and tubulin, the eukaryotic cytoskeleton includes an extended suite of proteins that form a large range of morphologically distinct filaments that in metazoa are collectively referred to as intermediate filament (IF) proteins (Coulombe and Wong 2004; Goldman et al. 2012; Herrmann and Aebi 2004). Individual IF proteins can assemble into higher-order homo- or heteromultimers through long repetitive core regions (Herrmann et al. 2007). These regions act as a kind of cellular Velcro (Rose et al. 2005), which mediate the formation of dimeric and trimeric coiled-coils. The evolutionary origin and radiation of metazoan IF proteins is not well understood (Herrmann and Strelkov 2011), especially due to a poor conservation in terms of primary sequence. In protists the situation is more involved.

Many proteins have been identified in a variety of protists that share common structural properties, but no homology, with metazoan IF proteins (Fleury-Aubusson 2003; Roberts 1987). An early identified example are the articulins, the most abundant proteins of the euglenoid membrane skeleton that are characterized by a specific and highly-repetitive VPV-motif (Huttenlauch et al. 1998; Marrs and Bouck 1992). Another group of protists, the alveolates, encode the alveolin proteins (Gould et al. 2008). Similar to articulins, these proteins are characterized by a repetitive and charged motif, EKIVEVPV, and the proteins themselves form a filamentous network that supports the alveolar sacs residing underneath the plasma membrane of all alveolates (Anderson-White et al. 2011; Bullen et al. 2009; El-Haddad et al. 2013; Mann and Beckers 2001). A third example are the epiplasmins of ciliates that form a large multigene family. This filament-forming protein family is very diverse, reflecting a rapid evolution fueled by whole genome duplications (Damaj et al. 2009). Epiplasmins are characterized by a conserved central region consisting of a repetitive consensus motif that is [ERK]xx[VILT]EY[VIY], and which is flanked by additional repetitive sequence motifs such as PVQ- and Y-rich domains (Coffe et al. 1996; Damaj et al. 2009). Giardins are filament-forming proteins of the excavate parasite *Giardia lamblia*, which are associated with the microribbons (Crossley and Holberton 1983). This excavate parasite has an elaborate actin cytoskeleton that, surprisingly, is organized in the absence of canonical actin-binding proteins (Paredez et al. 2011). Others include centrin and spasmin (Gogendeau et al. 2008; Vigues et al. 1984) or the tetrins (Brimmer and Weber 2000; Clerot et al. 2001). None of the mentioned protist proteins that form or associate with filaments share any obvious sequence homology with metazoan IFs. These proteins are, however, characterized by repetitive motifs that are biased in terms of their amino acid composition, a property also known from metazoan intermediate filament proteins (Fleury-Aubusson 2003; Gould et al. 2011). Such proteins are usually restricted to a specific sub-cellular structure and often the eukaryotic lineage in which they are found.

Metazoan genomes encode a large variety of filament-forming proteins of the IF family, including the well-known examples lamin, vimentin and keratin (Coulombe and Wong 2004; Fuchs and Weber 1994). It is common perception that IF proteins are restricted to metazoa (Herrmann and Strelkov 2011; Peter and Stick 2015) and that an early duplication and modification of the lamin gene in the bilateria, and the loss of the nuclear localization signal in one of the copies, might be the origin of many extant cytosolic IF proteins in these organisms (Erber et al. 1999; Kollmar 2015; Peter and Stick 2015). The identification of lamin-encoding genes in a range of phylogenetically distant eukaryotic groups (Kollmar 2015; Koreny and Field, 2016) and an earlier hydrophobic cluster analysis (Bouchard et al. 2001), however, indicate an evolutionary early origin of at least some kind of “proto-IF protein” in eukaryotes. A few prokaryotic IF proteins have been identified, too, such as CreS from *Caulobacter*, CfpA and Scc from spirochetes, and AglZ from *Myxococcus xanthus*, but with their evolutionary origin not always fully resolved (Ausmees et al. 2003; Bagchi et al. 2008; Mazouni et al. 2006; Yang et al. 2004; You et al. 1996). True homologs of IF proteins other than lamin appear absent from sequenced protist genomes. Hence, the question about the nature of proteins that support the elaborate cytoskeleton of protists, which in terms of morphological and species diversity outnumber metazoa to a considerable degree, remains open.

The Parabasalia form a monophyletic group within the excavates and they are characterized by unique lineage-specific cytoskeletal structures (Brugerolle 1991; Cepicka et al. 2010; Noda et al. 2012). Parabasalia include a long list of ecologically and medically important protists, which,

among others, include the majority of flagellated gut symbionts of termites (Brune and Dietrich 2015) and human parasites such as *Trichomonas vaginalis* or *Trichomonas tenax* (Hirt and Sherrard 2015; Kusdian and Gould 2014). The parabasalian cytoskeleton typically consists of several characteristic features that include the pelta, the axostyle, the costa and the karyomastigont (Cepicka et al. 2010). Both the axostyle and the pelta are formed by highly organized microtubules (Benchimol 2004; Rosa et al. 2013). The karyomastigont comprises the basal bodies, the microtubule-organizing center, the flagella and the filaments that connect these to the nucleus (Brugerolle 1991). The costa is a rod-like rigid structure exclusively found in the Parabasalia, albeit some parabasalian species appear to lack it (Cepicka et al. 2010). This striated filament (or striated fiber) protects the plasma membrane against the sheering forces generated by the beating of the recurrent flagella that is attached to the undulating membrane (Kulda et al. 1986; Viscogliosi and Brugerolle 1994). In comparison to other parabasal filaments, the costa has been described as the longest and thickest of the striated filaments of *T. vaginalis* (Lee et al. 2009) and *Tritrichomonas foetus* (Rosa et al. 2013). Based on different striation patterns, the costa was divided into two sub-types, the A- and B-type (Honigberg et al. 1971). While morphologically distinct, it appears that the striated filaments of both types of costa are formed by the same proteins that are, however, unique to Trichomonadidae and whose identity remains unknown (Viscogliosi and Brugerolle 1994).

A proteome study of the detergent-resistant pellicle of the ciliate *Tetrahymena* showed that many of the identified proteins shared characteristics with metazoan IF proteins, albeit, again, without any significant sequence similarity to them (Gould et al. 2011). Many of the identified pellicle proteins were united by charged repetitive motifs with a potential to form long coiled-coil regions. Subsequent characterization of individual candidates with similar characteristics from related apicomplexans confirmed the association of such proteins with unique cytoskeletal structures (Katris et al. 2014; Suvorova et al. 2015; Tran et al. 2012). Those studies motivated us to test whether proteins with similar characteristics could be observed for the unique cytoskeleton of Trichomonadidae. If so, it would (i) provide evidence for the presence of cytoskeletal proteins that share properties with metazoan IF proteins in yet another eukaryotic group — next to alveolates — and (ii) offer a more exhaustive set of proteins to study the evolution of cytoskeletal proteins, which form and associate with non actin and tubulin-based filaments.

Here we present the proteomic profiling of the detergent-resistant components of the cytoskeleton of *T. gallinarum* (Fig. 1). We demonstrate that the majority of identified proteins lack homologs outside of the Trichomonadidae and are unique to this group of protists. Based on a transcriptome assembly of *T. gallinarum* and mass spectrometry analyses of its extracted cytoskeleton, we identified homologs of *T. vaginalis* and localized five such homologs in the parasite that invades the human urogenital tract. Cytoskeleton proteins contain significantly more repetitive elements, such as coiled-coil motifs, than cytoskeleton-unrelated proteins. Most importantly, the unique cytoskeletal structure of the parabasalian lineage comprises proteins that share features described for other filament-forming proteins of different protists. Together they share features reminiscent of metazoan IF proteins, but elude themselves from phylogenetic analysis due to virtually no sequence conservation.

Results

Proteomic Profiling of the *Tetratrichomonas gallinarum* Cytoskeleton

Protocols for the isolation of the detergent-resistant cytoskeleton already exist for *T. gallinarum, T. vaginalis*, *Pentatrichomonas hominis and Tritrichomonas foetus* (de Souza and da Cunha-e-Silva 2003; Rosa et al. 2013; Viscogliosi and Brugerolle 1994). However, even after several rounds of attempts trying to optimize and adapt the protocols to isolate the cytoskeleton of *T. vaginalis*, we were unable to extract a sufficient amount of sample that, based on visual inspection and Western blot analysis, we deemed suitable for electrospray ionization tandem mass spectrometry (ESI-MS) analysis. Similar issues regarding *T. vaginalis* were encountered in previous studies (Viscogliosi and Brugerolle 1994) and it was hence decided to isolate the cytoskeleton of *T. gallinarum* instead.

The isolation of the *T. gallinarum* cytoskeleton leads to a highly-enriched fraction of forked structures that resemble the pelta (at the top of the fork), from which the costa and the axostyle individually emerge (Fig. 1B; Supplementary Material Fig. S1A). In their appearance, these structures look identical to those described by Viscogliosi and Brugerolle (1994). A silver-stained SDS-PAGE of

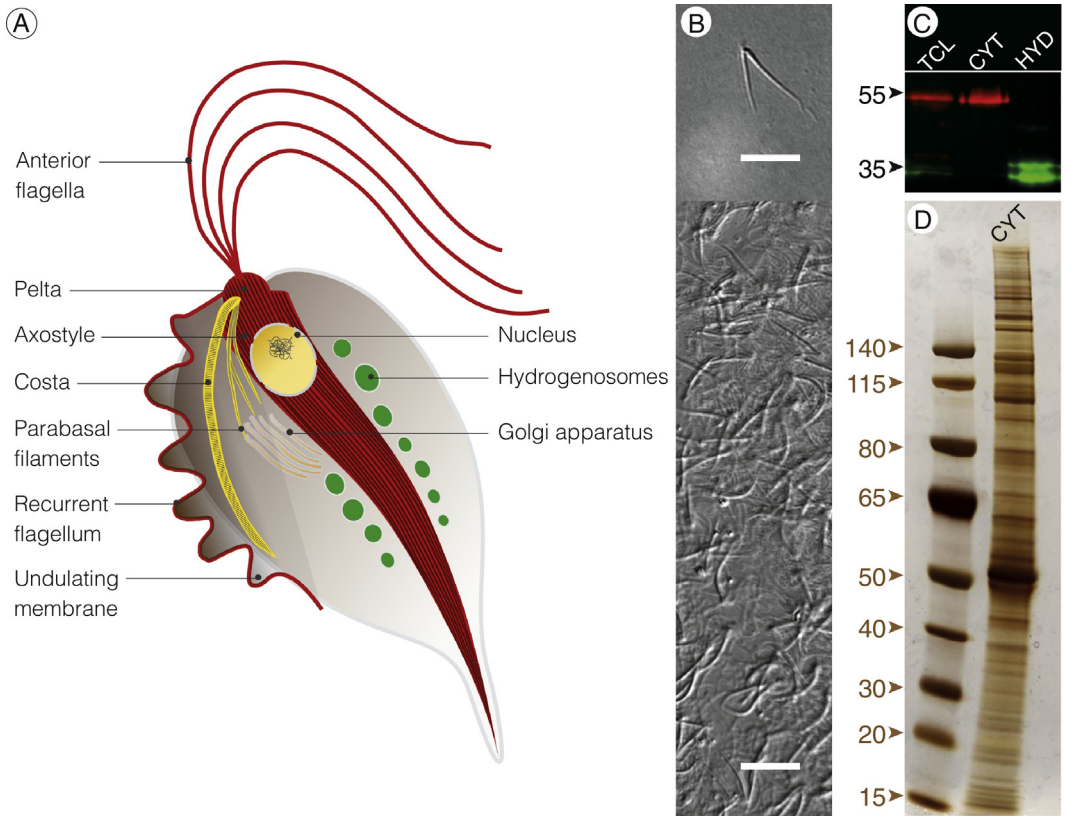

Figure 1. The cytoskeleton of *T. gallinarum* and *T. vaginalis*. **(A)** The cartoon illustrates the pelta-axostyle-costa-karyomastigont complex and its associated structures that represent the main cytoskeletal structures of the parasite. The karyomastigont system in this case comprises nucleus associated filaments and all five kinetosomes. The extracted cytoskeletal fraction **(B)** showing the cytoskeleton was analyzed by the multiplex fluorescent blot **(C)**, separated on the silver stained gel **(D)** and used for mass spectrometry analysis (ESI-MS). The multiplex fluorescent blot visualizes the detection of tubulin (red signal) and SCSalpha (green signal) within the fractions total cell lysate (TCL), cytoskeleton (CYT) and hydrogenosomes (HYD) indicating the absence of hydrogenosomes as a likely contaminant in the cytoskeleton fraction. The two bands of SCSalpha in the hydrogenosomal fraction are frequently observed (Woehle et al. 2014; Zimorski et al. 2013). To verify the data, triplicates were applied for ESI-MS, see Supplementary Material Figure S1. Scale bar: 10 μm. Molecular weights in C and D in kilodalton.

the sample reveals a rich proteinaceous complexity (Fig. 1D). The most abundant protein migrating at 55 kDa most likely represents tubulin with a predicted mass of 50.1 kDa (for TVAG_467840) (Fig. 1D). There are several more intense and distinct bands within the range of 100-140 kDa, whose migration pattern largely matches that of a previous analysis (Viscogliosi and Brugerolle 1994). That study had isolated the cytoskeleton, too, but in the absence of genome data at that time the identity of the proteins remained unknown. The most abundant proteins in *Trichomonas* species are those of the substrate-level phosphorylation pathway of the hydrogenosomes and they are a common source of contamination (Garg et al. 2015; Rada et al. 2011; Twu et al. 2013). A multiplex fluorescent blot confirmed the absence of the hydrogenosomal marker protein SCS (succinyl-CoA synthetase) in the isolated fraction and the presence of tubulin as a marker for the main component of the cytoskeleton (Fig. 1C). By liquid chromatography ESI-MS, three individual replicates of cytoskeleton isolated from three separate *T. gallinarum* cultures were analyzed. For the identification of the proteins through ESI-MS, however, it was first necessary to generate a de novo transcriptome library of *T. gallinarum*, because no sufficient amount of sequence data for this organism were yet available.

RNA sequencing of *T. gallinarum* M3 generated a total of 20,982,889 reads, of which 20,638,776 passed quality filtering. The subsequent assembly yielded 64,756 contigs (N50 length of 694), which were filtered for isoforms and then screened for open reading frames (ORFs). A total of 37,740 ORFs were retrieved and of these, 26,130 share homology to genes (and with 11,268 unique matches) of the sequenced genome of *T. vaginalis* strain G3 (TrichDB, v2.0) (Aurrecoechea et al. 2009).

Using the translated transcriptome assembly of *T. gallinarum* as a source (Supplementary Material Table S1), we identified 582 proteins present with a minimum of two peptides per protein in each of the three replicate cytoskeleton samples analyzed (Supplementary Material Table S2). Homologs for these proteins were identified in *T. vaginalis* using reciprocal best BLAST search. Results were manually curated by filtering them according to their TrichDB annotation to omit obvious cytoskeleton-unrelated candidates such as ribosomal proteins (Supplementary Material Table S2). This yielded 203 potential cytoskeleton proteins in *T. vaginalis*. The 203 proteins were first sorted according to their EuKaryotic Orthologous Groups association (KOGs; Koonin et al. 2004; Tatusov et al. 2003). For 113 proteins no class was available and for an additional 30, the predicted function was either of unknown nature or just a general prediction of function (Supplementary Material Fig. S4). The remaining proteins could be assigned to three major categories: 'cellular processes and signalling' (47 proteins including actin and tubulin), 'information storage and processing' (8 proteins) and 'metabolism' (7 proteins). Note that three proteins were assigned to more than one sub-category (Fig. S4).

Proteins that were distinctly identified with a high number of peptide spectrum matches (PSM) were, as already indicated by the silver stained SDS-PAGE (Fig. 1D), the tubulin beta- (TEGb007706; corresponding to TVAG_467840) and epsilon chain (TEGb007357; corresponding to TVAG_008680) with PSM values of 1,228 and 456, respectively, and a total coverage of 75% for each protein. Next to tubulin, the proteins TEGb005933 (TVAG_339450), TEGb003426 (TVAG_474360), TEGb019317 (TVAG_117060), TEGb017573 (TVAG_030160) and TEGb012599 (TVAG_059360), were detected with PSM values ranging between 111 and 330 and sequence coverage between 63 and 76% (Table 1; Supplementary Material Table S2). To find homologs, a broader BLAST search outside of the Trichomonadidae against RefSeq was performed. For TVAG_339450, TVAG_474360 and TVAG_117060 no significant (cutoff e-value of $\leq 1e^{-10}$) homologs could be found. For TVAG_030160, homologs encoding WD40 domains were identified in a range of organisms and with e-values $\leq 1e^{-138}$. Among them were uncharacterized protist proteins such as TTHERM_01094880 (*Tetrahymena thermophila*; e-value $8e^{-141}$), but also ones of *Angomonas deanei*, a trypanosomatid parasite (EPY27992.1, e-value $2e^{-134}$) or the alga *Chlamydomonas reinhardtii* (XP_001690930; e-value $2e^{-137}$) that are all annotated as "flagella-associated" proteins. Also for TVAG_059360, BLAST hits to proteins from diverse organisms were found (all e-values $\leq 2e^{-150}$), many of which were annotated as "sperm-associated antigen". It was peculiar that all five proteins were predicted to harbor many repetitive elements as predicted by SMART (Letunic et al. 2015) (Fig. 2). In addition, for TVAG_339450, TVAG_474360 and TVAG_117060, coiled-coil regions were widely detected throughout their amino acid sequences by COILS (Lupas et al. 1991) (Supplementary Material Fig. S2). This prompted us to further analyze the sequence characteristics, such as amino acid composition and the potential to encode coiled-coils regions, of the identified cytoskeleton proteins.

Cytoskeleton-associated Proteins Harbor an Elevated Number of Long Coiled-coil Regions

Elongated coiled-coil regions that occupy much of the central region of proteins are a characteristic feature of most metazoan IF proteins (Herrmann et al. 2007; Rose et al. 2005) and, more generally, an enrichment for repetitive charged sequence motifs has been found among cytoskeletal scaffold proteins of excavates and ciliates (Elmendorf et al. 2003; Gould et al. 2011; Kloetzel et al. 2003). We tested whether proteins of our isolated detergent-resistant cytoskeleton might exhibit similar sequence features. We compared our core set of 203 cytoskeleton-associated proteins (CYT) to (i) a set of 301 hydrogenosome-associated proteins (HYD; Garg et al. 2015) and to (ii) a size-equivalent set of 203 randomly selected proteins of *T. vaginalis* (RDM) (Fig. 3; Supplementary Material Fig. S3). Figure 3A shows the distributions of the number of detected coiled-coils per protein in the cytoskeleton and hydrogenosomal sets; these two distributions were found to be statistically different (Mann-Whitney test (MWt), p-value < $2.1e^{-6}$).

Table 1. Properties of the five cytoskeleton proteins localized and in comparison to actin and tubulin as references. Five candidates were chosen for immunofluorescence localization studies according to the PSM (peptide spectrum matches) value and the molecular weight. The tubulin beta chain (TVAG_467840) and actin (TVAG_172680) were also identified as part of the cytoskeleton proteome. Database values are adopted from TrichDB v2.0. In addition, based on the transcriptome of *T. vaginalis* (Gould et al. 2013), the absolute expression values of the proteins are listed. Cov %, coverage of predicted protein sequence; UP, unique peptides; AA, amino acids; MM, molecular weight; IEP, isoelectric point.

Accession numbers		Annotation	E value	*T. vaginalis*	Mass spectrometry data of *T. gallinarum*			TrichDB annotated properties		
				Expression absolute	Coverage %	UP	PSM	AA	MW (kDa)	IEP
T. gallinarum	***T. vaginalis***									
TEGb005933	TVAG_339450	unknown	0.0	665	65	67	256	977	113.5	4.9
TEGb003426	TVAG_474360	unknown	0.0	2267	65	66	244	1042	119.1	4.8
TEGb019317	TVAG_117060	unknown	0.0	502	70	64	330	878	100.5	4.9
TEGb017573	TVAG_030160	F-box and WD domain protein	0.0	682	76	33	163	605	65.7	6.9
TEGb012599	TVAG_059360	Sperm associated antigen 6	0.0	941	63	30	111	505	55.0	7.1
TEGb007706	TVAG_467840	Tubulin beta chain	0.0	29130	75	27	1228	452	50.1	4.7
TEGb007619	TVAG_172680	Actin	0.0	18485	8	9	24	386	43.0	5.3

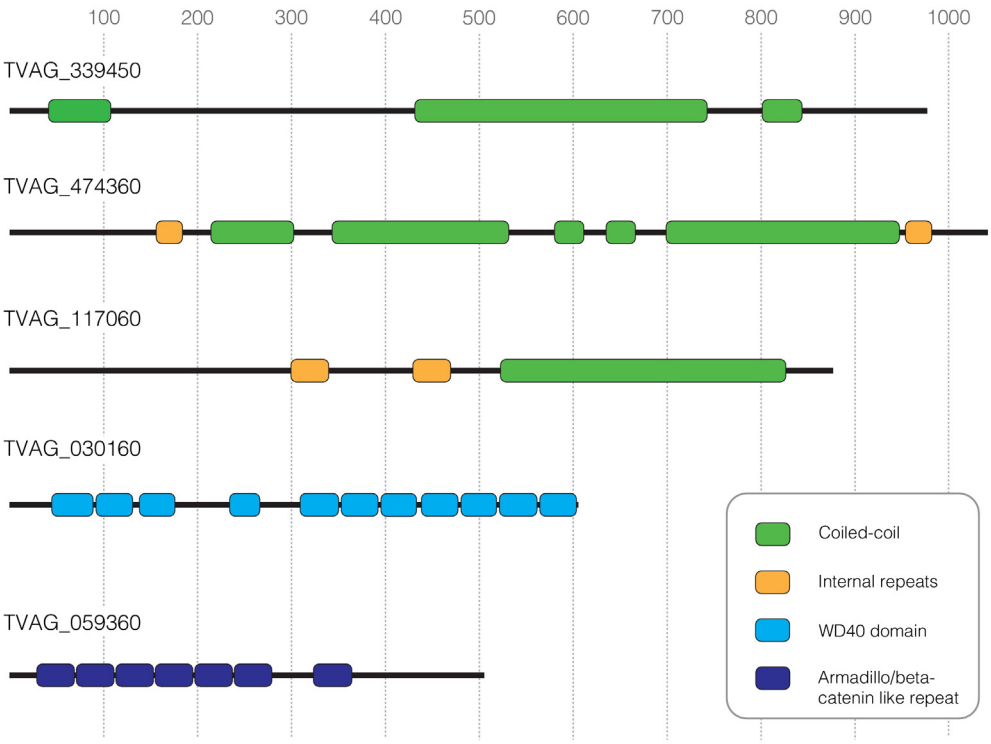

Figure 2. Distribution of recognizable and conserved domains among the five selected *T. vaginalis* proteins. The motifs shown are based on the prediction by the SMART algorithm (Simple Modular Architecture Research Tool; Letunic et al. 2015).

The same was true for the comparison with the random set (Supplementary Material Fig. S3A), which was found to be significant, too (MWt, p-value < $3e^{-12}$). Isolated cytoskeleton proteins also differ with respect to the length of the sequence predicted to form coiled-coils. The length of each coiled-coil unit within the CYT set was found to be significantly increased as compared to the HYD dataset (MWt, p-value < $1.9e^{-5}$, Fig. 3B) and with respect to the RDM dataset (MWt, p-value < 0.022, Fig. S3B). Moreover, CYT proteins were found to contain significantly more repetitive motifs per protein when compared to the HYD dataset (MWt, p-value < 0.0003, Fig. 3C) and the RDM dataset (MWt, p-value < $4.8e^{-5}$, Supplementary Material Fig. S3). To consider the possibility that long proteins tend to have more coiled-coils and repetitive motifs than smaller ones, we conducted an additional analysis and filtered the two control datasets (RDM and HYD) for proteins similar in length to the CYT dataset. Here, the comparison between the CYT and HYD dataset was still significantly different for all three parameters tested: number of coiled-coils (MWt, p-value < 4.8^{-6}), length of coiled-coils (MWt, p-value < 1.3^{-9}) and number of repetitive motifs (MWt, p-value < 0.0001). The re-analysis between the CYT and RDM dataset confirmed the primary result, except for the number of repetitive motifs, which now was no longer found to be significant (number of coiled-coils, MWt, p-value < 1.8^{-10}; length of coiled-coils, MWt, p-value < 0. 0004; number of repetitive motifs, MWt, p-value < 0.73).

Homologous Proteins of *T. vaginalis* Localize to Defined Filamentous Structures

We tested whether homologs of proteins isolated from the *T. gallinarum* cytoskeleton and identified through the ESI-MS analysis are homologous and associated with the cytoskeletal framework of *T. vaginalis*. We chose the five proteins with the highest PSM scores (Table 1) and expressed them as hemagglutinin (HA)-fusion proteins in *T. vaginalis*. The first protein, TVAG_339450, labels a single long filament-like structure that runs along the cell's periphery, suggesting it is a costa-related

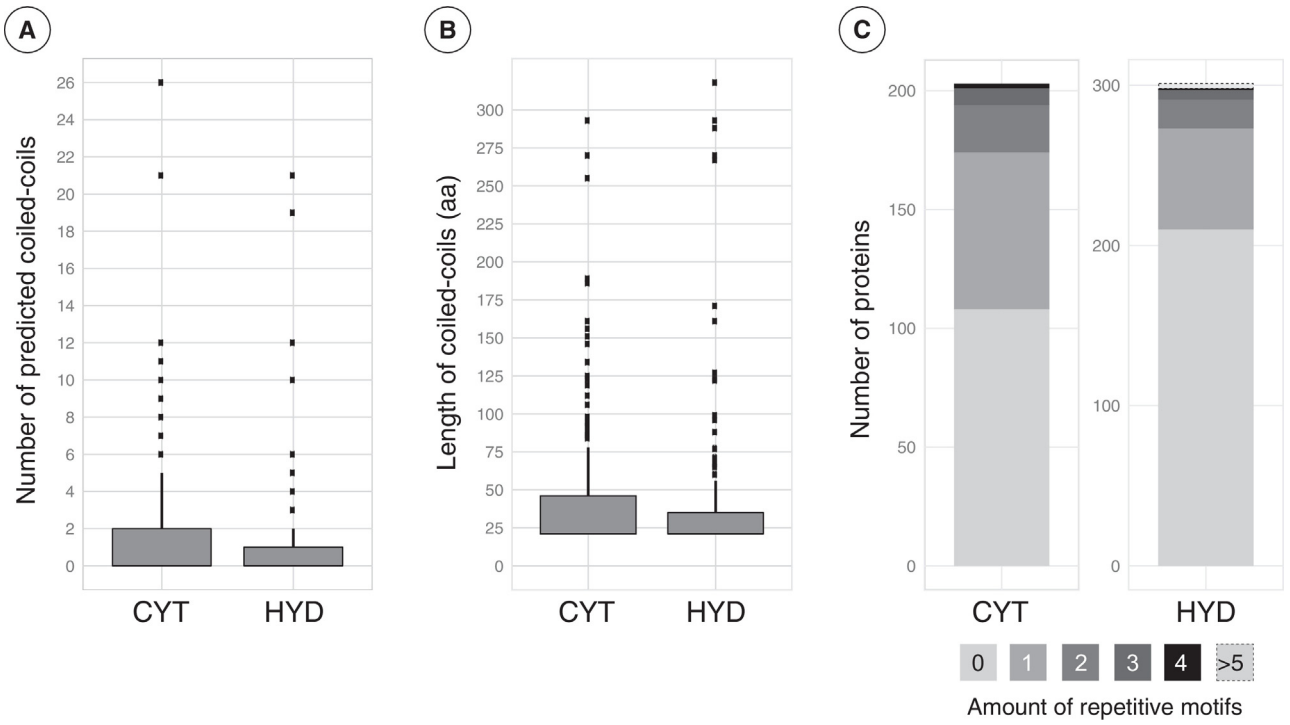

Figure 3. Coiled-coils and other repetitive units are increased among proteins of the cytoskeleton proteome. **(A)** The box blot displays the number of predicted coiled-coils within 203 proteins present in the extracted cytoskeleton (CYT) fraction compared to 301 hydrogenosomal proteins (HYD; based on Garg et al., 2015) as a control. **(B)** The box blot indicates the length of coiled-coils in amino acids (aa) based on the same dataset. **(C)** The bar diagram shows the distribution of the number of repetitive units per protein in the compared datasets. One unit consists of at least 3 repetitive motifs, each being at least 10 amino acids long. Here the number of proteins containing 1, 2, 3 or 4 repetitive motifs is significantly higher in the CYT dataset than in the HYD dataset. In contrast to the dataset of randomly selected proteins (RDM; Fig. S3), the HYD dataset contains proteins harbouring more than five repetitive motifs.

protein (Fig. 4, Supplementary Material Figs S6-S8). As the protein does not co-localize with the tubulin marker, it can be ruled out that the long filament resembles the axostyle or the recurrent flagellum that is associated with the undulating membrane.

The second candidate, TVAG_474360, has a more complex localization pattern. It was observed to cluster towards the apical end of the nucleus (i.e. in the region of the pelta) and from there to form or associate with filaments that appeared to envelope the nucleus (Fig. 4; Supplementary Material Fig. S6). In addition, TVAG_474360 was found to cluster inside the cytosol and at the end of the axostyle, perhaps as a result of overexpression, although this was not observed for any other construct. Also, the native expression of TVAG_474360 is the highest among the selected candidates (Table 1). In any case, the pelta-associated labeling is resilient and also observed among the isolated cytoskeletons (Supplementary Material Fig. S7), suggesting it is associated with the centrosome-like structure of *T. vaginalis* also known as the atractophore (Bricheux et al. 2007). Intriguingly, the heterologous expression of TVAG_474360 in *T. gallinarum* induces the formation of additional striated filaments. A single straight and thick rod-like structure traverses the cells centrally, which protrudes from the anterior- and posterior ends of the cell (Fig. 5A). Transmission electron microscopy of the mutants reveal this structure to consist of an accumulation of individual striated and thinner filaments that resemble, but occur independently, of the actual wildtype forms of the costa and the parabasal filaments (Fig. 5B).

The third protein, TVAG_117060, labels a fork-like structure just beneath the pelta and in close proximity of the nucleus and could possibly

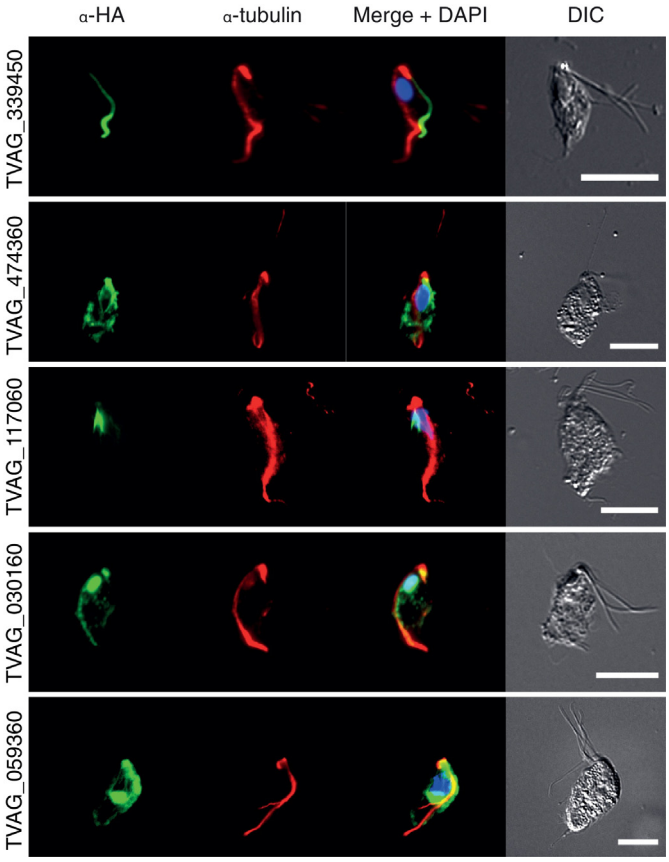

Figure 4. Immunofluorescence localization in *T. vaginalis* of the five cytoskeleton candidate proteins. All proteins were expressed as hemagglutinin-fusion proteins (anti-HA, green) and co-localized with tubulin (red) that labels in particular the axostyle. Scale bar: 10 μm.

represent two parabasal filaments of the parasite (Fig. 4; Supplementary Material Fig. S6). The fourth protein, TVAG_030160, co-localizes with the axostyle, spanning from the anterior pelta end to almost the posterior end, but is excluded from the axostyle's most terminal end that protrudes the cell. Furthermore, it appears to accumulate in the pelta region and somehow envelope the nucleus (Fig. 4; Supplementary Material Fig. S6). Conspicuously, Western blot analysis indicated that this protein is present in two different forms, with a faint signal at 66 kDa (in accordance with its mass prediction of 55.7 kDa based on the TrichDB annotated ORF) and with a dominant signal of about 110 kDa, suggesting it may dimerize (Supplementary Material Fig. S5). The fifth protein, TVAG_059350, does not co-localize with the axostyle and appears to form some sort of filaments around the nucleus and additionally some kind of thin filaments that span the entire length of the cell (Fig. 4; Supplementary Material Fig. S6).

Discussion

Specialized cytoskeletal structures are often unique to individual protist groups. The apomorphic and eponymous structure of the Parabasalia is the parabasal apparatus. It is part of the more intricate cytoskeletal scaffold that includes the pelta, axostyle, parts of the karyomastigont system and several other accessory rootlet filaments. A main component of the eukaryotic cytoskeleton is tubulin, but microtubules — together with other well-known accessory proteins we also identified — account only for the composition of the flagella, pelta and axostyle. Our study, based on the proteomic profiling of the detergent-resistant cytoskeleton of *T. gallinarum* and subsequent verification in *T. vaginalis*, identified dozens of proteins of (previously) unknown function that share some features characteristic for metazoan IFs, albeit sharing no sequence homology with the latter.

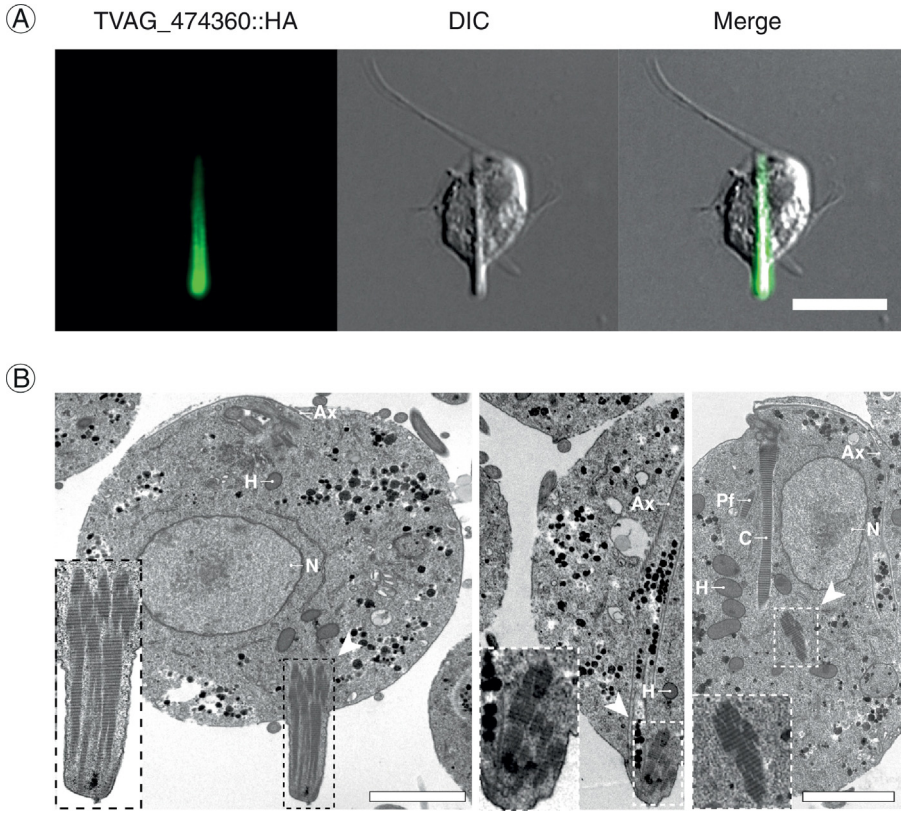

Figure 5. Heterologous expression of TVAG_474360 in *T. gallinarum*. **(A)** Immunofluorescence microscopy images show the localization of the hemagglutinin (HA)- tagged TVAG_474360 (green). Scale bar: 10 μm. (**B**) Transmission electron microscopy images of longitudinal sections through three exemplary cells. The cells show additional striated filaments (highlighted by the dashed boxes and indicated by arrow heads) that together aggregate to form a single thick rod. The striated pattern of the additional filaments looks similar to that of the costa (C) and the parabasal filaments (Pf), but they differ in width and were not observed to co-localize with either of the two naturally-occurring filaments. Axostyl, Ax; Nucleus, N; Hydrogenosomes, H. Scale bar: 1 μm.

Proteome profiling of the *T. gallinarum* cytoskeleton revealed 203 proteins that share reciprocal best BLAST hit homology to proteins of *T. vaginalis*. Among these were homologs of well-known cytoskeleton proteins such as actin, tubulin, centrin and dynein, all of which had high peptide spectrum matches (PSM; Supplementary Material Table S2), but also many new potential cytoskeletal protein candidates with previously unknown function. This observation provides indirect support for the reliability of the protocol used for the isolation of the cytoskeleton. For the five proteins of the cytoskeleton with the highest PSM values next to the canonical cytoskeletal proteins (listed in Table 1), we could show them to label filamentous structures that are, or associate with, the *T. vaginalis* cytoskeleton.

TVAG_339450 is likely a major, detergent-resistant component of the costa. As a rod-like structure, the costa can be clearly distinguished from the recurrent flagellum of the undulating membrane that constitutes the typical microtubule organization (Benchimol et al. 2000; Delgado-Viscogliosi et al. 1996; Rosa et al. 2013). TVAG_339450 does not co-localize with tubulin, the main component of pelta, axostyle and flagella (Fig. 4; Supplementary Material Fig. S6). In comparison to the parabasal filaments, the costa has been described as the longest and thickest striated filament of *T. vaginalis* (Lee et al., 2009) and *T. foetus* (Rosa et al., 2013). This is in line with the observed labeling pattern of TVAG_339450 (Fig. 4; Supplementary Material Fig. S6). Western blotting confirmed the predicted mass of around 118 kDa

for TVAG_339450 (Fig. S5), which is furthermore coherent with a previous study that generated antibodies against the isolated cytoskeleton of *T. vaginalis* (Viscogliosi and Brugerolle 1994). The antibodies decorated the isolated costa and a band of 118 kDa in the accompanied Western blot analysis. This very confined localization to the costa is also observed for our HA-tagged TVAG_339450 among the isolated cytoskeleton structures (Supplementary Material Fig. S7). The main function of the costa, a semi-rigid rod, is the stabilization of the undulating membrane to which the recurrent flagellum is attached (Cepicka et al. 2010; Rosa et al. 2013). Proteins known to be specialized in buffering shearing forces and to protect against mechanical stress are IF proteins (Goldman et al. 2012; Herrmann et al. 2007). IF proteins can differ with regard to their sequence and biochemical properties, but are generally united by a common architecture that is defined through long coiled-coil motifs (Coulombe and Wong 2004; Herrmann and Aebi 2004). The same is true for TVAG_339450, which shows this costa protein to unite some features that are characteristic for metazoan IF proteins, albeit not being homologous to any known protein of the metazoan IF family.

The coiled-coil containing cytoskeleton proteins TVAG_474360 and TVAG_117060 did not localize to the costa, but to other filamentous structures (Fig. 4; Supplementary Material Figs S6, S8). In the case of TVAG_117060 these might be the parabasal filaments, although the localization in isolated cytoskeleton structures is more punctuate in close proximity to the pelta (Supplementary Material Fig. S7). For TVAG_474360 the situation is more complicated. In several independent experiments the protein always associated with the pelta and several defined filaments embracing the entire nucleus, almost resembling a thin ring, suggesting it could act as a crosslinker between the microtubule-based cytoskeleton and a structure of unknown nature, a function similar to that reported for some IF-proteins (Eckert et al. 1982; Kalnins et al. 1985). In some cases, such crosslinking IF proteins were observed in the close proximity of the nucleus (Goldman et al. 1985; Trevor et al. 1995) and a localization similar observed to that of TVAG_474360. With a similar pattern TVAG_030160 co-localized with the pelta, most of the axostyle and accumulated in close proximity of the nucleus (Fig. 4; Supplementary Material Figs S6, S8). This protein also appears to form a stable dimer (Supplementary Material Fig. S5), but this observation requires an analysis dedicated to the protein in question. In contrast, TVAG_059360 does neither co-localize with the axostyle nor the nucleus distinctly. Instead, it appears net-like around the nucleus and branches from there through the cell (Fig. 4; Supplementary Material Figs S6, S8). The latter two proteins are both not predicted to form long coiled-coils, but WD40 or ARM domains (armadillo/beta-catenin like repeat), respectively (Fig. 2). The functions of these two domains are rather versatile, but both are also found in proteins associated with the cytoskeleton (Stirnimann et al. 2010; Tewari et al. 2010).

Based on their association with the nucleus, TVAG_474360, TVAG_030160 and TVAG_059360, might be relevant for the positioning of the nucleus and its anchoring to the karyomastigont system (Brugerolle 1991; Cepicka et al. 2010), but a dedicated study to unravel their function, and the reason for the potential dimerization of TVAG_030160, is required. In any case, the heterologous expression of the coiled-coil protein TVAG_474360 in *T. gallinarum*, and the formation of many additional striated filaments as a result (Fig. 5), suggest this protein (and its homologs) to be responsible for the formation of striated filaments in Trichomonadidae. Most likely this protein itself forms those striated filaments, but we cannot rule out it only acts as an accessory protein that recruits others.

Independent of their function within the cytoskeleton, there is something peculiar about a few dozen of the proteins identified. Based on a search algorithm specific for charged repeat motif proteins that were found enriched among a ciliate pellicle (Gould et al. 2011), we found a similar tendency among the proteins of the cytoskeleton albeit not as dominant (Fig. 3A; Supplementary Material Fig. S3A). Such motifs, sometimes predicted to also form long coiled-coils, have been routinely identified among structural scaffolding proteins of protists that form filaments, thin and thick, and these include: the euglenoid articulins (Marrs and Bouck 1992), the autoantigen I/6 of *Trypanosoma brucei* (Detmer et al. 1997), SF-assemblin of the basal apparatus of green algae and many other protists (Weber et al. 1993), the FAZ1 (a Flagellum Attachment Zone related protein) of *T. brucei* (Vaughan et al. 2008), the alveolins of *T. thermophila* (El-Haddad et al. 2013; Gould et al. 2011), the H49/calpain protein of *T. cruzi* (Galetovic et al. 2011) and the 477 kDa centrosome-associated protein of *T. vaginalis* (Bricheux et al. 2007). This is to name just a few and plenty more must be present (Dawson and Paredez 2013; Fleury-Aubusson 2003; Roberts 1987). While coiled-coil containing proteins are

involved in diverse processes, long and centrally-located coiled-coil regions are characteristic for IF proteins (Herrmann and Strelkov 2011; Mason and Arndt 2004; Rose et al. 2005) and exactly these were found enriched among proteins of the isolated cytoskeleton (Fig. 3B; Supplementary Material Fig. S3B).

The identification of the first IF proteins stems from research on the cytoskeleton of vertebrates. The first intermediate filaments were found in muscle cells of chick embryos (Ishikawa et al. 1968). Later the name 'intermediate filaments' became more commonly used to indicate their general, but not obligate, intermediate width between actin filaments and microtubules (Fuchs and Weber 1994). Ever since, the definition of IF proteins orbits around those families that have been identified in vertebrates. Now consider two things: First, already among vertebrates the IF protein families are diverse and not ubiquitously present, also because of the flexible architecture of coiled-coils that can deviate a lot from the canonical IF heptad repeat pattern (Hicks et al. 1997; Holberton et al. 1988; Mason and Arndt 2004). Second, vertebrates make up only a very small portion of extant eukaryotic diversity (Baldauf 2008; Burki 2014). It is quite possible that vertebrate IF protein families represent only a small fraction of the proteins that can form filamentous structures in eukaryotes, and only those that originate from an ancient duplication of lamin genes are recognizable as a family of homologous proteins inside metazoa. IF protein evolution is difficult to track across phylogenetically distant groups (Bouchard et al. 2001; Fleury-Aubusson 2003; Gould et al. 2011). If so, the two case examples of the alveolate pellicle proteome (Gould et al. 2011) and the cytoskeleton proteome of an excavate (this study) — two phylogenetically independent supergroups and independent of opisthokont metazoans — suggest that we can expect to find a similar diversity among the majority, if not all, individual eukaryotic supergroups.

Our data show that several of the identified detergent-resistant cytoskeleton proteins share features that are considered a trade mark of metazoan IF proteins. This includes the length of individual coiled-coil forming motifs, the number of total coiled-coil sequences and the cytoskeleton-associated scaffolding nature of the localization pattern we observe, that is their function inside the eukaryotic cell. More than 50% (113) of the 203 proteins identified have no significant sequence similarity (E-value of $10e^{-10}$) to proteins of any other organism outside of the Trichomonadida and none to canonical metazoan IF protein families. The presented data cannot claim that it has uncovered dozens of new IF proteins similar to those of metazoa per se, but it encourages to dig deeper and it provides a source from where to start. Considering the morphological complexity of protists, it is hard to imagine how these single cells would realize their many unique scaffolding structures, if not also with the support of proteins that are analog to metazoan IFs. Future studies will now need to characterize individual candidates and for instance demonstrate that some of these proteins can form filaments autonomously in vitro and induce such in other heterologous systems like yeast.

We conclude that many of the discussed cytoskeleton proteins share properties described for metazoan IFs, but that are at the same time specific to the parabasalian lineage. They evolved either rapidly, the primary sequence no longer serving as a reliable source to screen for phylogenetic relationships, or independently. A combination of the two options is plausible and would complicate the matter of identification and classification even further. Rapid evolution for the coiled-coil forming domains of such proteins has been observed for syntenic genes of apicomplexan parasites (Gould et al. 2011), and the limited amount of interactions with other cytosolic proteins is thought to lift the sequence constrains of such proteins in general (Fleury-Aubusson 2003). On the contrary, if the many cytoskeleton proteins of these phylogenetically distant eukaryotic groups are of multiple evolutionary origins, then they provide an excellent example of extensive convergent evolution. Either way, such observations require us to discuss the restrictive use of the term IF protein only for metazoa.

Methods

Cell cultivation: *Trichomonas vaginalis* FMV-1 (kindly provided by M. Benchimol, University Santa Ursula, Rio de Janeiro, Brazil) was cultivated in Tryptone Yeast extract maltose Medium (TYM) containing 2.22% (w/v) tryptose, 1.11% (w/v) yeast extract, 15 mM maltose, 9.16 mM L-cysteine, 1.25 mM L(+) ascorbic acid, 0.77 mM KH_2PO_4, 3.86 mM K_2HPO_4, 10% (v/v) heat-inactivated horse serum, 0.71% (v/v) iron solution [= 1% (w/v) $Fe(NH_4)_2(SO_4)$ x $6H_2O$, 0.1% (w/v) 5-sulfosalicylacid)] adjusted to pH 6.2 and incubated anoxically at 37 °C. *Tetratrichomonas gallinarum* M3 isolated from the caecum of *Melaegris gallopavo* (kindly provided by Prof. Tachezy, Department of Parasitology, Charles University of Prague, Czech Republic) was cultivated anoxically in TYM with pH 7.2 at 37 °C.

Cytoskeleton extraction of *T. gallinarum*: Based on different cytoskeleton extraction protocols (de Souza and da Cunha-e-Silva 2003; Palm et al. 2005; Viscogliosi and Brugerolle 1994), five hundred ml cell culture of *T. gallinarum* with approx. 7 x 10^6 cell/ml were collected by centrifugation at

1,500 x *g* for 10 min at room temperature (RT). The cell pellet was washed twice in Ringer solution: 0.12 M NaCl, 3.5 mM KCl, 2.0 mM $CaCl_2$, 2.5 mM $NaHCO_3$, pH 7.2 with centrifugations at 999 x *g*, 10 min, RT. Washed cells were then resuspended in 30 ml ice-cold Triton solution: 10 mM Tris base, 2 mM EDTA, 2 mM DTT, 1 mM ATP, 2 mM $MgSO_4$, 200 mM KCl, 1.5% Triton X-100, pH 7.8 including one 1x complete Mini protease inhibitor tablet (Roche). The solution was vortexed strongly for 2 min, incubated on ice for 2 min and this was repeated three times. After cell lysis was confirmed via light microscopy the cell extract was centrifuged at 277 x *g* for 15 min, RT. For increasing the output and purity of the isolated cytoskeleton complex, a sucrose gradient centrifugation was applied. Therefore, the resulting pellet was transferred on top of 1 ml 2 M sucrose and centrifuged at 1,000 x *g*, 15 min, 4 °C resulting in an upper band and a lower band. The upper band was transferred on top of 1 ml 1.5 M sucrose and centrifuged at 10,000 x *g*, 15 min, 4 °C. The pellet was transferred on top of a four-step gradient containing 1 M, 1.5 M, 2 M and 2.5 M sucrose (from top to bottom) within a 2 ml tube and centrifuged at 19,000 x *g*, 1 h, 4 °C. This led to three different pellets along the tube. The purest and most highly concentrated cytoskeleton fraction could be obtained from the upper-most pellet in the interface of 1 M and 1.5 M sucrose. This pellet was washed twice in PBS with centrifugations at 12,000 x *g*, 5 min, 4 °C, yielding a highly concentrated cytoskeleton fraction as evident through microscopy and Western blotting (Fig. 1).

The transcriptome of *T. gallinarum*: RNA-Seq reads were obtained using Illumina sequencing based on *T. gallinarum* M3 RNA (NCBI, accession SRA318841), which was isolated as described for *T. vaginalis* (Woehle et al. 2014). A quality filtering step was applied to the reads so that the first nine nucleotide (nt) positions were rejected according to a FastQC analysis (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) that showed low quality for the first 9 base calls. Subsequently, only reads with a minimum of 25 nt were retained. In addition, all reads containing 25% of low quality bases (25% of all bases with values $\leq$ Q15) identified by a self-written Perl script were rejected as well. The reads were assembled via *Trinity assembler* (version r20131110) (Grabherr et al. 2011). From all assembled contigs only the longest isoform of a candidate was selected. Open reading frames (ORFs) were identified and translated into the corresponding amino acid (aa) sequences by *getorf* from EMBOSS 6.6.0 (Rice et al. 2000) and a self-written Perl script was used to select only the longest ORF per candidate. To define an ORF, only stop codons were considered (option-find 0). Furthermore, only sequences with a minimum of 100 aa as a minimum for protein identification were used. For those sequences the best matches with *T. vaginalis* annotated genes were determined by using the BLAST program (version 2.2.28) (Altschul et al. 1997) in combination with the database TrichDB (version 1.3) (Aurrecoechea et al. 2009) based on an e value cutoff at $\leq 1e^{-10}$.

Protein identification by liquid chromatography electrospray ionization tandem mass spectrometry: Cytoskeletal fractions of *T. gallinarum* were separated in a polyacrylamide gel ($\sim$ 4 mm running distance). Protein containing bands from the silver stained gel were cut out, destained, reduced, alkylated with iodoacetamide and digested with trypsin (1:50 w/w Serva, Heidelberg, Germany) overnight at 37 °C as described (Poschmann et al. 2014). After that, resulting peptides were extracted from the gel and subjected to liquid chromatography in 0.1% trifluoroacetic acid.

Before mass spectrometric peptide identification, peptides were separated by an Ultimate 3000 Rapid Separation liquid chromatography system (Thermo Scientific, Dreieich, Germany) on an analytical column (Acclaim PepMapRSLC, 2 μm C18 particle size, 100 Å pore size, 75 μm inner diameter, 25 cm length, Thermo Scientific, Dreieich, Germany) over a two h gradient as described earlier (Hartwig et al. 2014). Using a nano electrospray ionization source, peptides were transferred to an Orbitrap Elite high resolution mass spectrometer (Thermo Scientific, Bremen, Germany) operated in positive mode with capillary temperature set to 275 °C and a source voltage of 1.4 kV. The orbitrap analyzer of the instrument was used for survey scans over a mass range from 350 – 1,700 m/z. A resolution of 60,000 (at 40 m/z) was used and the target value for the automatic gain control was set to 1,000,000 and the maximum fill time to 200 ms. Fragment spectra of the 20 most intense 2+ and 3+ charged peptide ions (minimal signal intensity 500) were recorded in the linear ion trap part of the instrument after collision induced dissociation based fragmentation using an available mass range of 200-2,000 m/z and at a resolution of 5,400 (at 400 m/z). A maximal fill time of 300 ms and an automatic gain control target value of 10,000 were used for the analysis of peptide fragments and already fragmented ions were excluded from fragmentation for 45 seconds.

Protein identification from mass spectrometric data was carried out using the MASCOT search engine (version 2.4.1, Matrix Science, London, UK) embedded in the Proteome Discoverer environment (version 1.4.1.14, Thermo Scientific, Dreieich, Germany) with standard parameters for spectrum selection. Searches were carried out in a *T. gallinarum* specific database containing 37,740 ORFs (obtained from the transcriptome analysis) with tryptic cleavage specificity allowing a maximum of one missed cleavage site. The precursor mass tolerance was set to 10 ppm, the fragment mass tolerance to 0.4 Da, carbamidomethyl at cysteine as static modification and methionine oxidation and N-terminal acetylation as variable modification. For peptide evaluation, the percolator node was used with standard parameters (strict target false discovery rate 1%, validation based on q-value). Only peptides passing the "high confidence" filter (1% false discovery rate) were used for protein assembly and only proteins reported with a minimum of two peptides were considered. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Vizcaino et al. 2014) with the dataset identifier *PRIDE*: PXD003212.

Protein datasets: Out of all the proteins identified by mass spectrometry, we retained only those proteins supported by at least two peptides. This resulted in a dataset of 582 proteins (Supplementary Material Table S2). Homologous proteins in *T. vaginalis* were next identified using reciprocal blast: best BLASTp hit with 25% identity and a minimum e value of $\leq 1e^{-10}$, without low complexity filter against the latest version of TrichDB (version 2.0, March 2015) (Aurrecoechea et al. 2009). This resulted in 271 pairs of putative orthologous proteins. Subsequently, this list was manually filtered to exclude cytoskeleton unrelated proteins, such as several ribosomal subunits. This filtering was performed based on TrichDB annotations. A final set of 203 cytoskeleton associated proteins was obtained.

For statistical purposes we established two control datasets. The first was composed of 301 hydrogenosomal proteins. This set was based on manually filtering the 359 proteins comprising the core hydrogenosomal proteome (Garg et al. 2015) to exclude hydrogenosomal unrelated proteins, such as diverse ribosomal subunits. The second set was obtained by randomly selecting 203 proteins from TrichDB. In an additional run, the random control datasets was normalized for their length to match those of the cytoskeleton dataset.

The number and length of coiled-coils within proteins were detected by NCOILS (Lupas et al. 1991). Repetitive motifs were

detected by RADAR (Heger and Holm 2000). A repetitive motif was considered to be at least three repetitive segments, each with a minimal length of ten amino acids, as used before for identifying cytoskeleton related proteins (Gould et al. 2011). Functional annotation was assigned to the 203 cytoskeleton associated cytoskeleton proteins, based on the KOG database (Koonin et al. 2004; Tatusov et al. 2003). Protein domains of the five candidate proteins (TVAG_339450, TVAG_474360, TVAG_117060, TVAG_030160 and TVAG_059360) were analyzed using SMART (Letunic et al., 2015). Coiled-coil motifs of TVAG_339450, TVAG_474360, TVAG_117060 were identified using COILS (Lupas et al. 1991).

Cloning and localization analysis: The gene sequences of all analyzed proteins were amplified by specific primers (Supplementary Material Table S3) with a proof-reading polymerase. PCR products were ligated into the expression vector pTagVag2 (Zimorski et al. 2013) to label the proteins with a C-terminal hemagglutinin (HA) tag. The final expression constructs were verified by standard sequencing. Transfection using electroporation (Delgadillo et al. 1997) was done using approx. 2.5 x 10^8 cells of *T. vaginalis* and 30 μg expression vector. After four h of anoxic incubation at 37 °C, geneticin (G418) was added to a final concentration of 100 μg/ml to initiate selection.

Protein probes were separated through a 8% or a 12% SDS-PAGE and blotted onto a Porablot nitrocellulose membrane (Macherey-Nagel) via the standard preset of the Trans-Blot Turbo Transfer System (BioRad). Membranes were blocked in blocking buffer: 5% milk powder + Tris-Buffered Saline (TBS) for 1 h, RT. Membranes of chemiluminescence blots were incubated with a primary monoclonal mouse anti-HA antibody (Sigma-Aldrich H9658) in blocking buffer (1:5,000) over night (ON) at 4 °C. After three washes, each 10 min in TBS-T (TBS + 0.1% Tween 20), membranes were incubated with a secondary anti-mouse and horseradish peroxidase-conjugated antibody (produced in rat, ImmunoPure, Pierce, ThermoFisher Scientific) in TBS-T (1:10,000)) for a minimum of 1 h at RT, followed by three washes of 10 min. Prior to detection, membranes were treated with WesternBright ECL spray (Advansta). The chemiluminescence reaction was analyzed with the ChemiDoc MP System (BioRad). Membranes of multiplex fluorescent blots were incubated overnight with primary antibodies (anti-HA (produced in mouse, H9658, Sigma-Aldrich) and anti-SCS (produced in rabbit, Eurogentec)) in blocking buffer (1:2,000) at 4 °C. After five washes for 5 min in TBS-T, membranes were incubated with secondary antibodies [(Alexa Fluor®488 goat anti-Rabbit IgG H+L, A-11008); Alexa Fluor®594 donkey anti-Mouse IgG H+L, A-21203 (ThermoFischer Scientific)] in TBS-T (1:2,000) for 2 h at RT followed by six washes of 5 min prior to immunofluorescence detection using the ChemiDoc MP System (BioRad).

For immunofluorescence microscopy a 12 ml culture of *T. vaginalis* with approximately 7 x 10^6 cell/ml was centrifuged at 914 x g for 8 min, RT. The cell pellet was washed twice in 1 ml Phosphate Buffer Saline (PBS) and centrifuged at 550 x g for 2 min, RT. The latter centrifugation setup was then maintained throughout the procedure. Cells were fixed and permeablized in 1 ml fix-perm-solution (3.5% paraformaldehyd, 0.5% Triton X-100 in PBS) and incubated for 20 min on a tube rotator, RT. Cells were collected by centrifugation and resuspended in 1 ml blocking-PBS (0.1% BSA in PBS) and blocked on a tube rotator for 20 min, RT. Cells were collected by centrifugation and resuspended in 0.5 ml primary antibody solution [(anti-HA (produced in rabbit, H6908, Sigma-Aldrich) and anti-tubulin (IG10, produced in mouse, from Bricheux et al. 2007) diluted in blocking-PBS, 1:1,000)] and incubated on a tube rotator, 2 h, RT. Cells were washed twice in PBS, resuspended in 0.5 ml secondary antibody solution [(Alexa Fluor®488 goat anti-Rabbit IgG H+L, A-11008 and Alexa Fluor®594 donkey anti-Mouse IgG H+L, A-21203, ThermoFischer Scientific) diluted in blocking-PBS 1:1,000)] and incubated on a tube rotator, 2 h, RT, in the dark. Cells were washed twice in PBS. Washed cells were mounted in DAPI solution (Fluoroshield, Sigma) plus PBS, 1:1. Imaging was conducted via a Nikon ECLIPSE Ti immunofluorescence microscope.

For TEM cells were pelleted at 1000 g for 10 min and washed three times with PBS [0.8% (w/v) NaCl, 0.02% (w/v) KCl, 0.144% Na2- HPO4, 0.024% (w/v) KH2PO4; pH 7.4]. After fixation over night at 4 °C in fixation buffer [2.5% (v/v) glutaraldehyde in 0.1 M Na-cacodylate buffer pH 7.3] the cells were washed four times for 10 min with 0.1 M Na-cacodylate buffer pH 7.3. Post fixation was done within two hours incubation with 2% (w/v) osmiumtetroxide diluted in 0.1 M Na-cacodylate buffer pH 7.3 containing 0.8% (w/v) potassium ferricyanide III. The cells were washed again four times like before and were then dehydrated by incubation with 50% (v/v), 70% (v/v), 80% (v/v), 90% (v/v) and absolute acetone (15 min each). Impregnation was done overnight in 1:1 acetone–epon mixture. The samples were polymerized in pure epon within 48 h at 60 °C. Ultra-thin sections of embedded samples were collected on formvar coated nickel grids (400 square mesh) and contrasted with subsequent incubation with saturated uranyl acetate solution and with 1% lead citrate for 5 min. Pictures were obtained using a Zeiss CEM 902 operated at 80 kV equipped with a wide-angle Dual Speed 2K CCD camera (TRS, Moorenweis Germany).

Acknowledgements

ELK is a fellow of the Edmond J. Safra Center for Bioinformatics at the Tel-Aviv University. TP and SBG wish to thank the German-Israeli Foundation for joint funding (grant ID I-1207-264.13/2012) and SBG and KS the Deutsche Forschungsgemeinschaft for additional support (CRC1208). HP and SBG wish to thank Ramona Stach and Maria Handrich for their help in the laboratory. We also thank Jan Tachezy and Marlene Benchimol for providing the strains.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.protis.2016.09.001

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**:3389–3402

Anderson-White BR, Ivey FD, Cheng K, Szatanek T, Lorestani A, Beckers CJ, Ferguson DJ, Sahoo N, Gubbels

MJ (2011) A family of intermediate filament-like proteins is sequentially assembled into the cytoskeleton of *Toxoplasma gondii*. Cellul Microbiol **13**:18–31

Aurrecoechea C, Brestelli J, Brunk BP, Carlton JM, Dommer J, Fischer S, Gajria B, Gao X, Gingle A, Grant G, Harb OS, Heiges M, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Miller JA, Morrison HG, Nayak V, Pennington C, Pinney DF, Roos DS, Ross C, Stoeckert CJ Jr, Sullivan S, Treatman C, Wang H (2009) GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. Nucleic Acids Res **37**:D526–D530

Ausmees N, Kuhn JR, Jacobs-Wagner C (2003) The bacterial cytoskeleton: an intermediate filament-like function in cell shape. Cell **115**:705–713

Baldauf SL (2008) An overview of the phylogeny and diversity of eukaryotes. J Syst Evol **46**:263–273

Bagchi S, Tomenius H, Belova LM, Ausmees N (2008) Intermediate filament-like proteins in bacteria and a cytoskeletal function in *Streptomyces*. Mol Microbiol **70**:1037–1050

Benchimol M (2004) Trichomonads under microscopy. Microsc Microanal **10**:528–550

Benchimol M, Diniz JA, Ribeiro K (2000) The fine structure of the axostyle and its associations with organelles in *Trichomonads*. Tissue Cell **32**:178–187

Bouchard P, Chomilier J, Ravet V, Mornon JP, Vigues B (2001) Molecular characterization of the major membrane skeletal protein in the ciliate *Tetrahymena pyriformis* suggests n-plication of an early evolutionary intermediate filament protein subdomain. J Cell Sci **114**:101–110

Bricheux G, Coffe G, Brugerolle G (2007) Identification of a new protein in the centrosome-like "atractophore" of *Trichomonas vaginalis*. Mol Biochem Parasitol **153**:133–140

Brimmer A, Weber K (2000) The cDNA sequences of three tetrins, the structural proteins of the *Tetrahymena* oral filaments, show that they are novel cytoskeletal proteins. Protist **151**:171–180

Brugerolle G (1991) Flagellar and cytoskeletal systems in amitochondrial flagellates: Archamoeba, Metamonada and Parabasalia. Protoplasma **164**:70–90

Brune A, Dietrich C (2015) The gut microbiota of termites: Digesting the diversity in the light of ecology and evolution. Annu Rev Microbiol **69**:145–166

Bullen HE, Tonkin CJ, O'Donnell RA, Tham WH, Papenfuss AT, Gould S, Cowman AF, Crabb BS, Gilson PR (2009) A novel family of apicomplexan glideosome-associated proteins with an inner membrane-anchoring role. J Biol Chem **284**:25353–25363

Burki F (2014) The eukaryotic tree of life from a global phylogenomic perspective. Cold Spring Harb Persp Biol **6**:a016147

Cepicka I, Hampl V, Kulda J (2010) Critical taxonomic revision of parabasalids with description of one new genus and three new species. Protist **161**:400–433

Clerot JC, Iftode F, Budin K, Jeanmaire-Wolf R, Coffe G, Fleury-Aubusson A (2001) Fine oral filaments in *Paramecium*: a biochemical and immunological analysis. J Eukaryot Microbiol **48**:234–245

Coffe G, Le Caer JP, Lima O, Adoutte A (1996) Purification, in vitro reassembly, and preliminary sequence analysis of epiplasmins, the major constituent of the membrane skeleton of *Paramecium*. Cell Motil Cytoskeleton **34**:137–151

Coulombe PA, Wong P (2004) Cytoplasmic intermediate filaments revealed as dynamic and multipurpose scaffolds. Nat Cell Biol **6**:699–706

Crossley R, Holberton DV (1983) Characterization of proteins from the cytoskeleton of *Giardia Lamblia*. J Cell Sci **59**:81–103

Damaj R, Pomel S, Bricheux G, Coffe G, Vigues B, Ravet V, Bouchard P (2009) Cross-study analysis of genomic data defines the ciliate multigenic epiplasmin family: strategies for functional analysis in *Paramecium tetraurelia*. BMC Evol Biol **9**:125

Dawson SC, Paredez AR (2013) Alternative cytoskeletal landscapes: cytoskeletal novelty and evolution in basal excavate protists. Curr Opin Cell Biol **25**:134–141

de Souza W, da Cunha-e-Silva NL (2003) Cell fractionation of parasitic protozoa: a review. Mem Inst Oswaldo Cruz **98**:151–170

Delgadillo MG, Liston DR, Niazi K, Johnson PJ (1997) Transient and selectable transformation of the parasitic protist *Trichomonas vaginalis*. Proc Natl Acad Sci USA **94**:4716–4720

Delgado-Viscogliosi P, Brugerolle G, Viscogliosi E (1996) Tubulin post-translational modifications in the primitive protist *Trichomonas vaginalis*. Cell Motil Cytoskeleton **33**:288–297

Detmer E, Hemphill A, Muller N, Seebeck T (1997) The *Trypanosoma brucei* autoantigen I/6 is an internally repetitive cytoskeletal protein. Eur J Cell Biol **72**:378–384

Eckert BS, Daley RA, Parysek LM (1982) Assembly of keratin onto Ptk1 cytoskeletons - Evidence for an intermediate filament organizing center. J Cell Biol **92**:575–578

El-Haddad H, Przyborski JM, Kraft LGK, McFadden GI, Waller RF, Gould SB (2013) Characterization of TtALV2, an essential charged repeat motif protein of the *Tetrahymena thermophila* membrane skeleton. Eukaryot Cell **12**:932–940

Elmendorf HG, Dawson SC, McCaffery M (2003) The cytoskeleton of *Giardia lamblia*. Int J Parasitol **33**:3–28

Erber A, Riemer D, Hofemeister H, Bovenschulte M, Stick R, Panopoulou G, Lehrach H, Weber K (1999) Characterization of the *Hydra* lamin and its gene: A molecular phylogeny of metazoan lamins. J Mol Evol **49**:260–271

Fleury-Aubusson A (2003) Novel cytoskeletal proteins in protists": Introductory remarks. J Eukaryot Microbiol **50**:3–8

Fuchs E, Weber K (1994) Intermediate filaments - Structure, dynamics, function, and disease. Annu Rev Biochem **63**:345–382

Galetovic A, Souza RT, Santos MRM, Cordero EM, Bastos IMD, Santana JM, Ruiz JC, Lima FM, Marini MM, Mortara RA, da Silveira JF (2011) The repetitive cytoskeletal protein H49 of *Trypanosoma cruzi* Is a calpain-like protein located at the flagellum attachment zone. PloS ONE **6(11)**:e27634

Garg S, Stoelting J, Zimorski V, Rada P, Tachezy J, Martin W, Gould S (2015) Conservation of transit peptide-independent

protein import into the mitochondrial and hydrogenosomal matrix. Genome Biol Evol **7**:2716–2726

Gogendeau D, Klotz C, Arnaiz O, Malinowska A, Dadlez M, de Loubresse NG, Ruiz F, Koll F, Beisson J (2008) Functional diversification of centrins and cell morphological complexity. J Cell Sci **121**:65–74

Goldman R, Goldman A, Green K, Jones J, Lieska N, Yang HY (1985) Intermediate filaments - Possible functions as cytoskeletal connecting links between the nucleus and the cell-surface. Ann NY Acad Sci **455**:1–17

Goldman RD, Cleland MM, Murthy SNP, Mahammad S, Kuczmarski ER (2012) Inroads into the structure and function of intermediate filament networks. J Struct Biol **177**:14–23

Gould SB, Tham WH, Cowman AF, McFadden GI, Waller RF (2008) Alveolins, a new family of cortical proteins that define the protist infrakingdom Alveolata. Mol Biol Evol **25**: 1219–1230

Gould SB, Woehle C, Kusdian G, Landan G, Tachezy J, Zimorski V, Martin WF (2013) Deep sequencing of *Trichomonas vaginalis* during the early infection of vaginal epithelial cells and amoeboid transition. Int J Parasitol **43**:707–719

Gould SB, Kraft LGK, van Dooren GG, Goodman CD, Ford KL, Cassin AM, Bacic A, McFadden GI, Waller RF (2011) Ciliate pellicular proteome identifies novel protein families with characteristic repeat motifs that are common to alveolates. Mol Biol Evol **28**:1319–1331

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol **29**:644–652

Hartwig S, Goeddeke S, Poschmann G, Dicken HD, Jacob S, Nitzgen U, Passlack W, Stühler K, Ouwens DM, Al-Hasani H, Knebel B, Kotzka J, Lehr S (2014) Identification of novel adipokines differential regulated in C57BL/Ks and C57BL/6. Arch Physiol Biochem **120**:208–215

Heger A, Holm L (2000) Rapid automatic detection and alignment of repeats in protein sequences. Proteins **41**: 224–237

Herrmann H, Aebi U (2004) Intermediate filaments: Molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. Annu Rev Biochem **73**:749–789

Herrmann H, Strelkov SV (2011) History and phylogeny of intermediate filaments: Now in insects. BMC Biology **9**:16

Herrmann H, Bar H, Kreplak L, Strelkov SV, Aebi U (2007) Intermediate filaments: from cell architecture to nanomechanics. Nat Rev Mol Cell Biol **8**:562–573

Hicks MR, Holberton DV, Kowalczyk C, Woolfson DN (1997) Coiled-coil assembly by peptides with non-heptad sequence motifs. Fold Des **2**:149–158

Hirt RP, Sherrard J (2015) *Trichomonas vaginalis* origins, molecular pathobiology and clinical considerations. Curr Opin Infect Dis **28**:72–79

Holberton D, Baker DA, Marshall J (1988) Segmented alpha-helical coiled-coil structure of the protein giardin from the *Giardia* cytoskeleton. J Mol Biol **204**:789–795

Honigberg BM Mattern CFT, C, Daniel WA (1971) Fine structure of the mastigont system in *Tritrichomonas foetus* (Riedmueller). J Protozool **18**:183-198

Huttenlauch I, Peck RK, Stick R (1998) Articulins and epiplasmins: two distinct classes of cytoskeletal proteins of the membrane skeleton in protists. J Cell Sci **111**:3367–3378

Ishikawa H, Bischoff R, Holtzer H (1968) Mitosis and intermediate-sized filaments in developing skeletal muscle. J Cell Biol **38**:538–555

Kalnins VI, Subrahmanyan L, Fedoroff S (1985) Assembly of glial intermediate filament protein Is initiated in the centriolar region. Brain Res **345**:322–326

Katris NJ, van Dooren GG, McMillan PJ, Hanssen E, Tilley L, Waller RF (2014) The Apical Complex Provides a Regulated Gateway for Secretion of Invasion Factors in *Toxoplasma*. PloS Pathog **10(4)**:e1004074

Kloetzel JA, Baroin-Tourancheau A, Miceli C, Barchetta S, Farmar J, Banerjee D, Fleury-Aubusson A (2003) Plateins: A novel family of signal peptide-containing articulins in euplotid ciliates. J Eukaryot Microbiol **50**:19–33

Kollmar M (2015) Polyphyly of nuclear lamin genes indicates an early eukaryotic origin of the metazoan-type intermediate filament proteins. Sci Rep **5**:10652

Koonin EV (2010) The origin and early evolution of eukaryotes in the light of phylogenomics. Genome Biol **11**:209

Koonin EV, Fedorova ND, Jackson JD, Jacobs AR, Krylov DM, Makarova KS, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Rogozin IB, Smirnov S, Sorokin AV, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. Genome Biol **5**:R7

Koreny L, Field MC (2016) Ancient eukaryotic origin and evolutionary plasticity of nuclear lamina. Genome Biol Evol, http://dx.doi.org/10.1093/gbe/evw087, In press

Kulda J, Nohýnková E, Ludvík J (1986) Basic structure and function of the trichomonad cell. Acta Univ Carol Biol **30**:181–198

Kusdian G, Gould SB (2014) The biology of *Trichomonas vaginalis* in the light of urogenital tract infection. Mol Biochem Parasitol **198**:92–99

Kusdian G, Woehle C, Martin WF, Gould SB (2013) The actin-based machinery of *Trichomonas vaginalis* mediates flagellate-amoeboid transition and migration across host tissue. Cellul Microbiol **15**:1707–1721

Lee KE, Kim JH, Jung MK, Arii T, Ryu JS, Han SS (2009) Three-dimensional structure of the cytoskeleton in *Trichomonas vaginalis* revealed new features. J Electron Microsc **58**:305–313

Letunic I, Doerks T, Bork P (2015) SMART: recent updates, new developments and status in 2015. Nucleic Acids Res **43**:D257–D260

Lupas A, Vandyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science **252**:1162–1164

Mann T, Beckers C (2001) Characterization of the subpellicular network, a filamentous membrane skeletal component in the parasite *Toxoplasma gondii*. Mol Biochem Parasitol **115**:257–268

Marrs JA, Bouck GB (1992) The two major membrane skeletal proteins (articulins) of *Euglena gracilis* define a novel class of cytoskeletal proteins. J Cell Biol **118**:1465–1475

Mason JM, Arndt KM (2004) Coiled coil domains: Stability, specificity, and biological implications. ChemBioChem **5**:170–176

Mazouni K, Pehau-Arnaudet G, England P, Bourhy P, Saint Girons I, Picardeau M (2006) The scc spirochetal coiled-coil protein forms helix-like filaments and binds to nucleic acids generating nucleoprotein structures. J Bacteriol **188**:469–476

Noda S, Mantini C, Meloni D, Inoue JI, Kitade O, Viscogliosi E, Ohkuma M (2012) Molecular phylogeny and evolution of Parabasalia with improved taxon sampling and new protein markers of actin and elongation factor-1 alpha. PloS ONE **7(1)**:e29938

Palm D, Weiland M, McArthur AG, Winiecka-Krusnell J, Cipriano MJ, Birkeland SR, Pacocha SE, Davids B, Gillin F, Linder E, Svard S (2005) Developmental changes in the adhesive disk during *Giardia* differentiation. Mol Biochem Parasitol **141**:199–207

Paredez AR, Assaf ZJ, Sept D, Timofejeva L, Dawson SC, Wang CJR, Cande WZ (2011) An actin cytoskeleton with evolutionarily conserved functions in the absence of canonical actin-binding proteins. Proc Natl Acad Sci USA **108**:6151–6156

Peter A, Stick R (2015) Evolutionary aspects in intermediate filament proteins. Curr Opin Cell Biol **32**:48–55

Poschmann G, Seyfarth K, Besong Agbo D, Klafki HW, Rozman J, Wurst W, Wiltfang J, Meyer HE, Klingenspor M, Stühler K (2014) High-fat diet induced isoform changes of the Parkinson's disease protein DJ-1. J Proteome Res **13**:2339–2351

Rada P, Dolezal P, Jedelsky PL, Bursac D, Perry AJ, Sedinova M, Smiskova K, Novotny M, Beltran NC, Hrdy I, Lithgow T, Tachezy J (2011) The core components of organelle biogenesis and membrane transport in the hydrogenosomes of *Trichomonas vaginalis*. Plos One **6(9)**:e24428

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. Trends Genet **16**:276–277

Roberts TM (1987) Fine (2-5-Nm) Filaments - New types of cytoskeletal structures - Invited review. Cell Motil Cytoskeleton **8**:130–142

Rosa ID, de Souza W, Benchimol M (2013) High-resolution scanning electron microscopy of the cytoskeleton of *Tritrichomonas foetus*. J Struct Biol **183**:412–418

Rose A, Schraegle SJ, Stahlberg EA, Meier I (2005) Coiled-coil protein composition of 22 proteomes - differences and common themes in subcellular infrastructure and traffic control. BMC Evol Biol **5**:66

Stirnimann CU, Petsalaki E, Russell RB, Muller CW (2010) WD40 proteins propel cellular networks. Trends Biochem Sci **35**:565–574

Suvorova ES, Francia M, Striepen B, White MW (2015) A Novel Bipartite Centrosome Coordinates the Apicomplexan Cell Cycle. PloS Biology **13(3)**:e1002093

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics **4**:41

Tewari R, Bailes E, Bunting KA, Coates JC (2010) Armadillo-repeat protein functions: questions for little creatures. Trends Cell Biol **20**:470–481

Tran JQ, Li C, Chyan A, Chung L, Morrissette NS (2012) SPM1 Stabilizes Subpellicular Microtubules in *Toxoplasma gondii*. Eukaryot Cell **11**:206–216

Trevor KT, Mcguire JG, Leonova EV (1995) Association of vimentin intermediate filaments with the centrosome. J Cell Sci **108**:343–356

Twu O, de Miguel N, Lustig G, Stevens GC, Vashisht AA, Wohlschlegel JA, Johnson PJ (2013) *Trichomonas vaginalis* exosomes deliver cargo to host cells and mediate host:parasite interactions. PloS Pathog **9(7)**:e1003482

Vaughan S, Kohl L, Ngai I, Wheeler RJ, Gull K (2008) A repetitive protein essential for the flagellum attachment zone filament structure and function in *Trypanosoma brucei*. Protist **159**:127–136

Viguès B, Méténier G, Sénaud J (1984) The sub-surface cytoskeleton of the ciliate *Polyplastron multivesiculatum*: isolation and major protein components. Eur J Cell Biol **35**:336–342

Viscogliosi E, Brugerolle G (1994) Striated fibers in trichomonads - Costa proteins represent a new class of proteins forming striated roots. Cell Motil Cytoskeleton **29**:82–93

Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, Dianes JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolome S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol **32**:223–226

Weber K, Geisler N, Plessmann U, Bremerich A, Lechtreck KF, Melkonian M (1993) Sf-assemblin, the structural protein of the 2-nm filaments from striated microtubule-associated fibers of algal flagellar roots, forms a segmented coiled-coil. J Cell Biol **121**:837–845

Wickstead B, Gull K (2011) The evolution of the cytoskeleton. J Cell Biol **194**:513–525

Woehle C, Kusdian G, Radine C, Graur D, Landan G, Gould SB (2014) The parasite *Trichomonas vaginalis* expresses thousands of pseudogenes and long non-coding RNAs independently from functional neighbouring genes. BMC Genomics **15**:906

Yang R, Bartle S, Otto R, Stassinopoulos A, Rogers M, Plamann L, Hartzell P (2004) AglZ is a filament-forming coiled-coil protein required for adventurous gliding motility of *Myxococcus xanthus*. J Bacteriol **186**:6168–6178

You Y, Elmore S, Colton LL, Mackenzie C, Stoops JK, Weinstock GM, Norris SJ (1996) Characterization of the cytoplasmic filament protein gene (cfpA) of *Treponema pallidum* subsp. *pallidum*. J Bacteriol **178**: 3177–3187

Zimorski V, Major P, Hoffmann K, Bras XP, Martin WF, Gould SB (2013) The N-terminal sequences of four major hydrogenosomal proteins are not essential for import into hydrogenosomes of *Trichomonas vaginalis*. J Eukaryot Microbiol **60**:89–97