Algal Research 59 (2021) 102440

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Compartmentalization of mRNAs in the giant, unicellular green alga *Acetabularia acetabulum*

Ina J. Andresen [a], Russell J.S. Orr [b], Kamran Shalchian-Tabrizi [a], Jon Bråte [a,c,*]

[a] Centre for Integrative Microbial Evolution (CIME) and Centre for Epigenetics, Development and Evolution (CEDE), Department of Biosciences, University of Oslo, Kristine Bonnevies Hus, Blindernveien 31, 0316 Oslo, Norway
[b] Natural History Museum, University of Oslo, Oslo, Norway
[c] Norwegian Institute of Public Health, Department of Virology, N-0213 Oslo, Norway

ARTICLE INFO

Keywords:
Acetabularia acetabulum
Dasycladales
mRNA
Subcellular RNA localization
Single-cell
Unique molecular indexes

ABSTRACT

Acetabularia acetabulum is a single-celled green alga which has historically been an important model system in cell biology. Here, we attempt to re-introduce *A. acetabulum* as a model system through the investigation of the subcellular localization of mRNAs. *A. acetabulum* is a macroscopic single-celled alga with a highly complex cellular morphology. The cell is gigantic, and the single nucleus, which is located at the basal end, is several centimeters away from the umbrella-shaped tip. To better understand the genetic basis of the morphology of this alga, we aimed to characterize the mRNA composition in four different subcellular regions of adult cells. Analyses of the quantitative gene expression data revealed a high degree of differentially distributed gene products. Gene transcripts involved in photosynthesis were mainly accumulated in the apical cap structure, while the basal end carrying the nucleus was enriched for nuclear processes such as DNA replication and transcription. Cytoskeletal components, intracellular transport and protein translation was present throughout this highly elongated cell. An important observation was that some gene transcripts were distributed throughout the cell while others, often functionally related transcripts, were localized in large pools at distinct subcellular regions. This suggests that post-transcriptional regulation and specific cellular localization of gene products may be important mechanisms behind morphogenesis in this gigantic cell.

1. Introduction

Large and complex morphologies are predominantly found among multicellular organisms such as animals, plants and kelp. However, several single-celled organisms also demonstrate elaborate cellular morphologies, and it has therefore been argued that multicellularity is not a requirement for structural complexity [1–4]. By subcellular compartmentalization of RNA or proteins, single-celled organisms can chamber their bodies into differently shaped subunits, further facilitating development of sophisticated body plans without cellularization [1–3]. The green algae order Dasycladales harbors multiple examples of single-celled species with highly elaborate and complex cellular morphologies. Despite having only a single nucleus, these algae have evolved into numerous macroscopic cell forms, which can grow to several centimeters in length [5,6].

Acetabularia acetabulum is the most studied dasycladalean species [7]. This alga is a predominantly subtropical, marine species found in shallow waters in the Mediterranean Sea, Northern Africa and South-West Asia [5]. Divers often refer to *A. acetabulum* as the "Mermaid's Wineglass" because of its distinctive morphology [8]; the basal end of the cell is a finger-like structure called a rhizoid, which hosts the nucleus as well as anchors the cell to the substrate. The rhizoid is followed by a naked stalk that stretches several centimeters, ending in a concave disc-shaped cap at the apical end (Fig. 1A).

Through his ground-breaking amputation and grafting experiments, Joachim Hämmerling was the first to identify the presence of substances controlling the subcellular morphogenesis of *A. acetabulum* [9–11]. He also showed these substances to be distributed in gradients, with the highest accumulation in the cap and rhizoid [12,13]. These morphogenetic substances were later identified as mRNA [14–17], but the exact types of mRNAs found in the different cellular regions were not assessed.

A few studies have tried to decipher the RNA gradient in *A. acetabulum*. In the 1970's, Naumova et al. [16] observed that the production of RNA remained after the nucleus was removed. They

* Corresponding author at: Norwegian Institute of Public Health, Department of Virology, N-0213 Oslo, Norway.
E-mail address: jon.brate@fhi.no (J. Bråte).

https://doi.org/10.1016/j.algal.2021.102440
Received 28 September 2020; Received in revised form 20 July 2021; Accepted 25 July 2021

therefore hypothesized that the gradient was due to differential metabolism of chloroplast ribosomal RNA along the cell rather than differential transportation of nuclear mRNA. More recently, Vogel et al. [18] examined the expression of 13 housekeeping genes and found differential accumulation of several gene transcripts. Some transcripts seemed to accumulate in the rhizoid, some in the cap, and others were evenly distributed along the cellular body. Serikawa et al. [19] demonstrated that the *A. acetabulum*-specific homeobox-containing gene, *Aaknox*1, shifted localization from being evenly distributed during vegetative growth, to basally accumulated in the final stages of the life cycle. Notably, both the transcripts encoding carbonic anhydrases as well as their translated protein products were shown to be co-localized in the apical regions of adult *A. acetabulum* cells [20], indicating that for at least some genes localization of mRNA is a mechanism for correct localization of the respective protein. Altogether, these results suggest that subcellular localization of mRNA may be an important mechanism for establishment of subcellular structures in *A. acetabulum*.

mRNA localization has also been demonstrated in another gigantic single-celled green algal genus, *Caulerpa* [4,21]. Unlike *A. acetabulum*, these species contain hundreds of nuclei distributed across the entire cell and the localization of mRNA is achieved, at least in part, by differential transcriptional regulation in the different parts of the cell [21]. Obviously, the mechanisms behind the subcellular localization of mRNAs, and ultimately the establishment of the cellular body plan, must differ in *A. acetabulum* and *Caulerpa*. Although these mechanisms are unknown, there are several indications that mRNAs are actively transported along the cytoskeleton in *A. acetabulum*. Kloppstech and Schweiger showed in 1975 that ribosomal and polyadenylated RNA travel with different speeds in *A. acetabulum*. In fact, polyadenylated RNA moved with a speed of 0.2 μm/s, which is much faster than movement through diffusion [22,23]. Further, staining experiments performed on *A. peniculus* showed actin proteins and polyadenylated RNAs to be co-localized, and that treatment with cytochalasin D (which inhibits actin polymerization) lead to a disruption of already established mRNA gradients [24]. These findings suggested RNA transport as both active and specific, and that the cytoskeleton is involved in polar transportation of mRNAs in *Acetabularia* species [18].

Studies on the expression and localization of mRNAs in *A. acetabulum* have thus far only been performed on a restricted number of genes, and it is not known how general this phenomenon is, nor how gene localization is related to the observed cellular RNA gradients. The only large-scale attempt at characterizing the transcriptional profile of *A. acetabularia* was performed on a few hundred transcripts from juvenile and adult cells [25]. Although this study did not examine subcellular regions, it supported that there is extensive transcriptional control as distinct gene expression profiles were detected at different developmental stages. We have in this study characterized the expression profile of all mRNAs in adult *A. acetabulum* cells. To achieve this, we have exploited recent developments in single-cell RNA-seq technology, which allowed us to quantitatively sequence mRNAs from four subcellular regions of the cell: the cap, the upper and lower parts of the stalk, and the rhizoid (Fig. 1A). We hypothesized that subcellular compartmentalization applies to a large number of mRNAs in *A. acetabulum*, and that the distribution of mRNAs may be coupled to the formation of the complex cellular morphology of this alga.

2. Methods

2.1. Culturing Acetabularia acetabulum cells

Adult cells of *A. acetabulum* were grown from hibernating embryonic cells. The embryos were stored in darkness (though exposed to light for 30 min every third months) and cultured in Dasycladales Seawater Medium (prepared using the recipe of UTEX Culture Collection of Algae at The University of Texas at Austin (https://utex.org/products/dasycladales-seawater-medium?variant=30991770976346)). About 20 embryonic cells were distributed into culture flasks (T-25 filter cap tissue culture flasks from Sarstedt, Germany). The flasks were kept in a vertical position and almost completely filled (ca. 60 ml) with Dasycladales Seawater Media. The media was changed biweekly, and the flasks were kept at 20 °C with a 12/12 h light/dark cycle and a light intensity of 45 μmol m^{-2} s^{-1}. Cells were cultured until they reached the adult stage, which was determined by the appearance of fully developed cap structures. The time to reach cell maturation lasted approximately three months, although the exact time varied between cells. Hence, each cell was not cultured for exactly the same amount of time.

2.2. Dissection of cells and RNA isolation

The sampled *A. acetabulum* cells were 5–8 cm long with fully grown caps. No apparent gametes in the gametangia or whorls along the stalk were observed (Fig. 1A). The cells were washed three times in 1× PBS to remove residue from the medium and dissected into four subcellular regions; the "cap" (incision 1–3 mm from the apical tip or just below the cap), "rhizoid" (incision 1–3 mm above the rhizoid), "upper stalk" (the upper half of the stalk) and "lower stalk" (the lower half of the stalk). Dissection was carried out in dry petri dishes to limit cytoplasmic loss, and new sterile scalpels were used between incisions. Subcellular regions from five to eight adult cells were pooled to achieve sufficient RNA quantities. The procedure was repeated to create seven biological replicates for each subcellular region, giving a total of 28 RNA extractions. The subcellular samples were numbered according to which batch of cells (indicated with a number before a punctuation) and which subcellular region (indicated with a number after the punctuation) they originate from: e.g. the samples 19.1, 19.2, 19.3, and 19.4 represent the cap, upper stalk, lower stalk and rhizoid samples from the same batch of individuals (batch 19).

The dissected pieces were transferred to green MagNA Lyser Green Beads (Roche Life Science, Germany), containing lysis buffer (see below), and immediately flash frozen in liquid nitrogen. RNA from three batches (batches 17, 19 and 25) was isolated using the "Single Cell RNA Purification Kit" (Norgen Biotek, Canada) and eluted in 8 μl of elution buffer. RNA from the remaining four batches (batches 26, 27, 45 and 46) was isolated using the "Total RNA Purification Kit" (Norgen Biotek) and eluted in 40 μl of elution buffer. The two different kits were used to establish which would perform better for our samples. In total we had three technical replicates of the Single Cell RNA purification kit and four of the Total RNA Purification Kit. RNA quality and quantity were inspected on the Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Pico kit (Agilent Technologies, Inc., Germany). Variability of the RNA isolation is reported as standard error of the mean.

2.3. Library preparation and sequencing

ERCC RNA Spike-In Mix I (Thermo Fisher Scientific, Massachusetts, USA) was added to each sample before mRNA enrichment according to the manufacturer's protocol. mRNA enrichment was performed using NEXTflex™ Poly(A) Beads (BIOO Scientific Corporation, Texas, USA) prior to library preparation with the NEXTflex™ Rapid Directional qRNA-Seq Library Prep kit for Illumina sequencing. This library preparation kit assigns unique molecular indexes (UMIs), or Stochastic Labels, to both ends of the mRNA fragments after enzymatic fragmentation, but before cDNA synthesis and amplification. This allows for the distinguishing of PCR duplicates and true identical sequences that map to the same loci, ensuring a better quantitative representation of the original number of mRNA fragments in the samples than standard RNA-seq library protocols without UMI-labelling [26]. A total of 30 PCR cycles were run for sample 17.1–4, 25 cycles were run for samples 25.4 and 26.4, and 20 cycles were run for samples 19.1–4, 25.1–3, 26.1–3, 27.1–4, 45.1–4 and 46.1–4 to create libraries of approximately equal concentrations as measured by gel electrophoresis.

The 28 libraries were sequenced on the Illumina HiSeq4000 platform

producing 150 bp paired-end sequences (with an insert size of 350 bp). During library prep, each library was barcoded for multiplex sequencing and libraries from batches 17, 19, 26 and 27 were pooled and sequenced together, as were those from batches 25, 26, 45 and 46. The sequencing was performed at the Norwegian Sequencing Centre (www.sequencing.uio.no) at the University of Oslo. The sequence data have been deposited to the European Nucleotide Archive under the accession nr. PRJEB40460.

2.4. De novo transcriptome assembly and annotation

In order to have a comprehensive transcriptome representing as many potentially expressed genes as possible, the resulting sequences from the 28 adult sequence libraries were assembled together with 20 transcriptome sequence libraries from various developmental stages of *A. acetabulum* (unpublished data generated by our research group). Before assembly, the first nine bases of the 5′ end (the UMIs) of each sequence were removed using Trimmomatic v/0.35 [27]. Then, the 3′-adaptor sequences, and low-quality sequences (phred score < 20) were trimmed. Only sequences longer than 36 bp were retained. An additional trimming with TrimGalore v/0.3.3 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was performed to remove any remaining adaptor sequences. ERCC RNA spike-ins were removed from the dataset by mapping the reads to the known ERCC RNA spike-in sequences using Bowtie 2 v/2.2.9 [28]. A total of 2,095,599,508 paired end reads (1,047,799,754 pairs) were obtained after pre-processing and used for transcriptome assembly. De novo assembly was performed with Trinity v/2.5.1 [29]. To reduce the number of possible mapping sites in downstream analysis, the transcriptome was reduced to the highest expressed isoform for each gene. These isoforms were found by subsampling 10% of the sequenced reads of every sample using the BBMap package [30], mapping them to the Trinity assembly, and further extracting the isoforms with the highest coverage. Transcripts potentially encoded by the chloroplast- and mitochondrial genomes were identified by Megablast (e-value < 0.001) against a database containing chloroplast genomes from 59 published green algae species and a database containing mitochondrial genomes from 24 published green algae species (Tables S1 and S2). Transcripts giving significant hits against the databases where further examined by Megablast against the NCBI Nucleotide collection (nr/nt) database (e-value < 0.001) in order to exclude possible prokaryote contamination. Transcripts with no hits to either the plastid or mitochondrial databases were considered as nuclear encoded. As mRNA enrichment with poly(A) beads does not remove rRNA completely, rRNA transcripts were identified by Megablast against complete or partial 18S, 28S and 5.8S sequences from 20 green algae species (Table S3).

Transcriptome completeness was assessed with BUSCO v3.0 [31] against the Chlorophyta and Eukaryote datasets. Since nuclear genes of *A. acetabulum* have an alternative codon usage, where TGA is the only stop codon, and TAA and TAG instead encode glutamine [32,33], the alternative genetic code (translation table 6: Ciliate, Dasycladaen and Hexamita Nuclear Code) was used during BUSCO evaluation.

TransDecoder v/3.0.0 [34] was used to predict coding regions. Translation table 6 was used to translate nuclear encoded transcripts, translation table 16 (Chlorophycean Mitochondrial Code) was used to translate mitochondrial-encoded transcripts and translation table 1 (Universal Code) was used to translate chloroplast-encoded transcripts. Translation table 1 was also used for transcripts matching both the chloroplast and the mitochondrial database. The minimum peptide length was set to 70 amino acids, and the single best ORF per predicted peptide sequences was further annotated with the eggNOG-mapper v/5.0 [35,36].

2.5. Gene expression quantification, normalization and sample clustering

In order to quantify the gene expression, processed reads were mapped to the transcriptome with Bowtie2 v/2.2.9 [28] and gene count files were generated using dqRNASeq (https://github.com/e-hutchins/dqRNASeq), a Unix script developed for analyzing sequence data obtained with the NEXTflex Rapid Directional qRNA-Seq Library Prep kit. By using the result from the mapping, together with the raw unprocessed reads (which contained the UMIs at the 5′ end; see previous section), the script identifies paired reads with identical start- and stop sites *and* identical UMIs and counts these as one (these were assumed to be PCR duplicates.), while fragments with identical start- and stop sites but different UMIs are counted individually (these were assumed to originate from different RNA fragments.).

The plotCountDepth function in the SCnorm R package [37] was used to calculate and visualize the relationship between sequencing depth and gene counts across samples. For the highest expressed genes (i.e. carrying the most robust signal) there was a positive relationship between sequencing depth and gene expression, and we therefore continued with the normalization procedure in the DESeq2 package [38] and did not normalize against the spike-in ERCC transcripts. Because we used a transcriptome assembled from other developmental stages in addition to the adult cell, the transcriptome was composed of many transcripts with zero, or close to zero, expression in the adult samples analyzed here. These were therefore removed in order to reduce the computational burden. To do this we required transcripts to have a minimum count of 1 (raw count) in at least 4 samples before DESeq2 normalization. Effectively this removed transcripts with zero or close to zero expression, as well as transcripts only present in three out of six replicates from a cell compartment.

To compare the similarity between samples based on the overall variation in transcript abundance, the normalized counts were transformed using the variance stabilizing transformation (VST) function in DESeq2 and the principle components were identified and plotted using the plotPCA function in DESeq2 and the ggplot2 package [39]. To determine the statistical significance of the global differences in transcript abundances between the different subcellular stages, we ran a permutational multivariate analysis of variation (PERMANOVA) [40] using the adonis function in the R-package vegan [41].

2.6. Differential transcript abundance estimation

A Wald test (implemented in DESeq2), performed pairwise between the different subcellular compartments, was used to identify transcripts with differential abundance between at least two subcellular compartments. DESeq2 normalized read counts were used as input for the test. The batch origin of each sample was added as a blocking factor in the test to take into account any potential influence on the gene counts. Transcripts with an adjusted p-value < 0.05 (p-values adjusted for multiple testing using the Benjamini and Hochberg procedure [42]) were considered as significantly differentially abundant (DE). Only nuclear encoded mRNAs were used in the DE test.

A Venn diagram was constructed using the systemPipeR package in R [43] to visualize and compare the DE transcripts of each subcellular compartment. To visualize and plot DE transcripts based on expression levels, the raw counts were converted to CPM's (count per million) followed by TMM normalization (Trimmed Mean of M-values) using the edgeR package in R [44,45]. The mean expression values were further scaled and clustered in a heatmap using the Pheatmap package in R (https://CRAN.R-project.org/package=pheatmap).

2.7. GO-enrichment analysis

GO-terms provided by EggNOG were converted to GO-slim with OmicsBox (https://www.biobam.com/omicsbox), and GO-enrichment analysis on the differentially distributed transcripts unique to the cap, upper stalk, lower stalk and rhizoid were performed using the R package GOseq [46]. In addition, the stalk samples were analyzed together by combining the uniquely differentially distributed transcripts from the

upper- and lower stalk samples, as well as the differentially distributed transcripts shared between them. Lists of the unique transcripts from each subcellular compartment were extracted and inputted into GOseq, together with a list of transcript lengths, to account for any bias introduced from transcript length variation. GO-terms with a false discovery rate (FDR) < 0.05 were considered enriched. The "hit percentage" of each enriched GO-term within a subcellular compartment was calculated as the percentage of differentially distributed transcripts in a given GO-category compared to the number of transcripts in the transcriptome in the same GO-category. The hierarchical organization of the different enriched GO-terms was explored using the Mouse Genome Informatics web page (http://www.informatics.jax.org/vocab/gene_ontology). Heatmaps of GO terms showing the hit percentage of significantly enriched GOs was constructed using the ComplexHeatmap package in R [47] with a suitable number of K-means row-splitting. Row_km_repeats was set to 100.

2.8. Annotation of transcripts related to intracellular transport and localization

Transcripts related to cytoskeletal components (*actin*, *tubulin* and related genes) and cytoskeletal motor proteins (*myosins*, *dyneins* and *kinesins*), and poly(A) polymerases, were extracted from the eggNOG annotation. Homologs of genes related to cellular transport (*COP* and *clathrin*) were identified by reciprocal blast using annotated genes in NCBI RefSeq from *Chlamydomonas reinhardtii* as queries (see Table S4 for

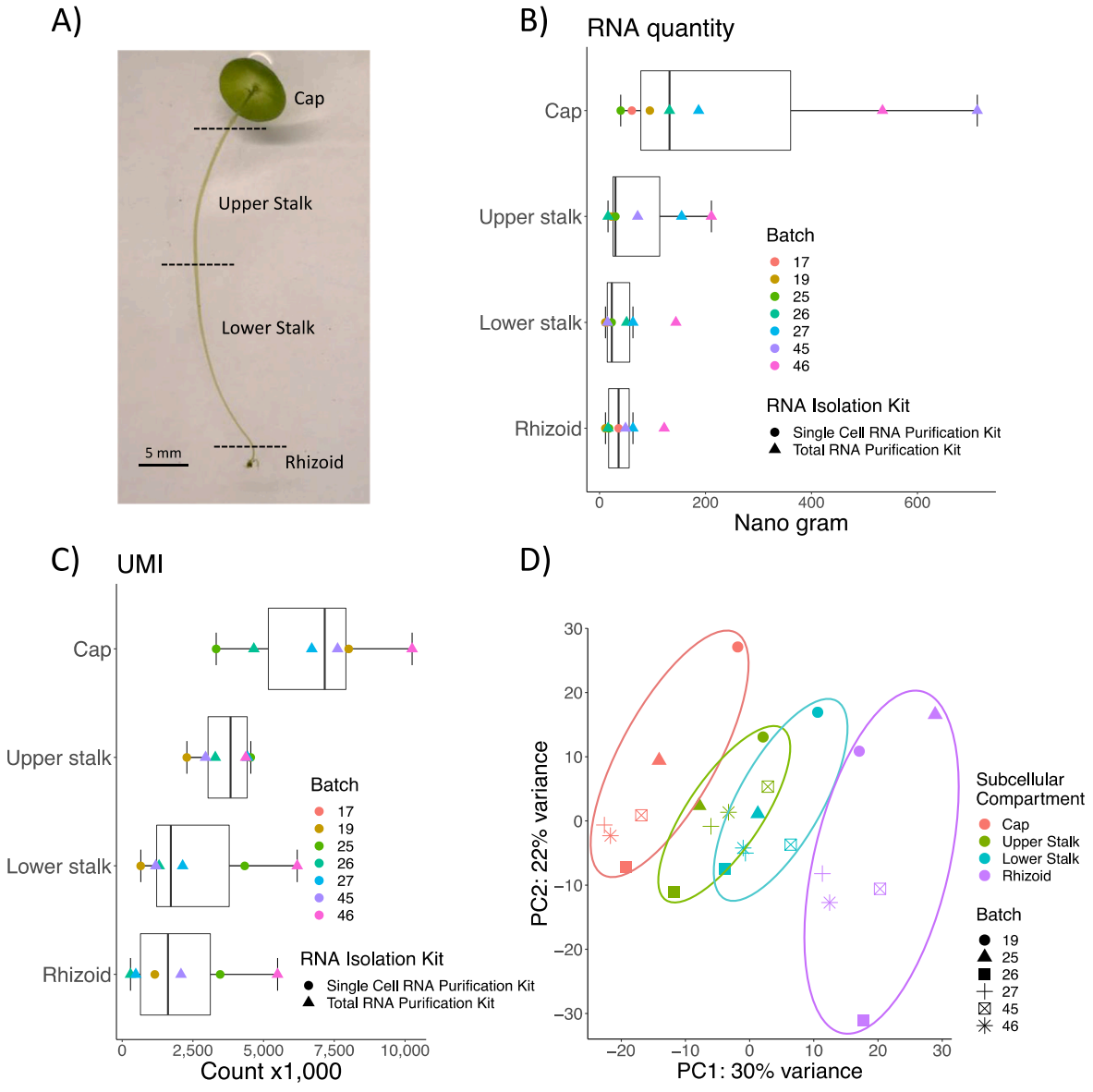

Fig. 1. RNA isolation and sequencing of subcellular compartments of adult *A. acetabulum*. A) Image of an adult cell of *A. acetabulum*. The dashed lines indicate approximate incision sites for separating the different subcellular regions; cap, upper stalk, lower stalk and rhizoid. B) Boxplot showing the total RNA quantity isolated from the different subcellular compartments. The dots represent the individual subcellular samples colored according to which batch the sample originate from (n = 7 for the number of batches), and shaped according to which RNA isolation kit that was used (n = 3 for the Single Cell RNA isolation kit, and n = 4 for the Total RNA isolation kit). C) Boxplot showing the summarized gene expression levels (summarized counts for each compartment when accounting for the Unique Molecular Indexes (UMIs)) of the different subcellular samples. The dots are the same as above, except that Batch 17 is not included (see Methods). Hence, n = 6 for the number of batches. And n = 2 for the Single Cell RNA isolation kit. D) Principal Component Analysis (PCA) of the sample variation based on variance stabilized counts (see Methods). The four subcellular compartments are shown in color, and the different batches which the samples originate from (described in the Methods) are indicated as shapes.

queries) (blastp value cutoff < 0.0001). The resulting hits for the *A. acetabulum* transcriptome were queried against Swissprot using blastp (evalue < 0.0001). Transcripts which did not produce Swissprot hits of the same category as in the first blast search were discarded. Transcripts with a mean TMM-normalized CPM > 1 across all samples were plotted, with standard error, using the R package ggplot2.

2.9. Comparative transcriptomics between A. acetabulum and Caulerpa taxifolia

The *Caulerpa taxifolia* transcriptome [4] was translated into amino acid sequences using Transdecoder v/3.0.0. Orthologous protein sequences between *A. acetabulum* and *C. taxifolia* were identified using Orthofinder v/2.3.3 [48]. RSEM generated gene expression data from six different subcellular compartments of *C. taxifolia* (apical cell section samples: frond apex, pinnules, rachis. Basal cell section samples: frond base, stolon and holdfast) were downloaded from the supplementary datafiles of Ranjan et al. [4]. Counts were rounded to the nearest integer and converted to TMM-normalized counts (as described above). TMM counts from the single-copy orthologs from the different subcellular compartments of *A. acetabulum* and *C. taxifolia* were merged and the differences in sample variation were visualized using the prcomp function in R and the ggplot2 R package.

3. Results

3.1. Subcellular RNA isolation, sequencing and read processing

The highest amount of total RNA was extracted from the cap samples (mean 252 ng ± 100 ng), followed by the upper stalk (mean 76 ng ± 29 ng), the lower stalk (mean 46 ng ± 18 ng), with the lowest amount extracted from the rhizoid (mean 45 ng ± 15 ng) (Table 1 and Fig. 1B). The highest yield of total RNA was obtained using the "Total RNA purification kit" with an elution volume of 40 μl (used for batch 26, 27, 45 and 46), which gave approximately four times more total RNA compared to the "Single Cell RNA purification kit" with an elution volume of 10 μl (used for batch 17, 19 and 25) (Table S5). However, the relative amounts of RNA isolated from the different samples were the same regardless of the isolation kit.

Sequence reads from batch 17 had overall very low-quality scores in addition to high duplicate numbers (mostly from sequencing the adapters). Therefore, very few sequences were retained after filtering and almost no genes were detected in these samples. The samples from batch 17 were therefore discarded from further analyses. Between 14 and 43 million raw read pairs were produced from each of the remaining 24 samples. Quality trimming and removal of unpaired reads after trimming reduced the numbers by 13–25% for the majority of samples, except the rhizoid of batch 26 and 27, where trimming reduced read number by 40 and 49% respectively (Table 1). Though, more than 15 million read pairs remained for these samples.

3.2. Transcriptome assembly and annotation

Assembling reads de novo produced an assembly consisting of 246,083 'genes', or transcripts, with a total of 429,781 different isoforms (Table S6), where the longest transcript was 17,196 bp and the shortest 201 bp. Blasting against algal chloroplast and mitochondrial genomes annotated 389 as chloroplast encoded and 99 transcripts as mitochondrial encoded. Further, 261 transcripts were annotated as prokaryotic or unclassified transcripts as they gave hits against both the chloroplast and mitochondrial databases as well as prokaryotes in NCBI nr database. The remaining 245,334 transcripts were considered to be encoded by the *A. acetabulum* nuclear genome. A total of 114,146 transcripts were predicted as protein coding. Of these, 113,900 belonged to the nuclear encoded genes, 178 to the chloroplast encoded, and 68 to the mitochondrial encoded genes. Analysis using EggNOG resulted in 38,131

Table 1
Total RNA isolation and mRNA sequencing of subcellular fragments of *Acetabularia acetabulum*. The naming of samples is described in the text. Read numbers are given as pairs of reads (single reads were discarded), both before and after trimming. "Mapping rate" describes the percentage of paired reads mapping concordantly (i.e. mapping in the expected orientation relative to each other) to the de novo assembled transcriptome, "total UMI count" is the sum of transcript expression levels for each sample after removing PCR duplicates. "Expressed transcripts" shows the number of transcripts with at least one UMI count. "% of expressed transcripts with count ≤ 2" is the percentage of "low count" transcripts.

Subcellular compartment	Batch	Total RNA (ng)	Raw reads (PE)	Trimmed reads (PE)	Mapped read pairs	Mapping rate (%)	Total UMI count	Expressed transcripts	% of transcripts with count ≤ 2
Cap	17	61	50,770,695	5,820,120	3,585,755	75	584	290	79
Upper stalk	17	29	82,587,170	11,714,864	7,206,886	73	660	257	68
Lower stalk	17	14	2092	260	166	74	0	0	–
Rhizoid	17	36	38,755,094	18,101,837	11,382,777	75	475	167	51
Cap	19	95	25,526,511	19,934,510	12,521,349	76	8,010,785	62,257	44
Upper stalk	19	22	14,007,423	10,572,421	6,395,304	73	2,285,994	45,187	42
Lower stalk	19	11	18,869,617	14,309,848	8,742,817	74	656,729	30,620	43
Rhizoid	19	11	24,207,912	18,101,837	11,021,291	74	1,156,418	38,275	45
Cap	25	40	22,444,960	19,460,102	11,548,241	72	3,325,240	58,967	45
Upper stalk	25	30	31,098,736	26,710,669	15,574,348	73	4,547,527	57,122	43
Lower stalk	25	23	37,346,096	32,191,425	19,013,248	73	4,334,098	62,677	44
Rhizoid	25	18	37,858,071	32,519,937	19,919,778	73	3,466,028	58,615	48
Cap	26	132	32,643,222	28,358,416	16,777,807	72	4,656,581	63,592	44
Upper stalk	26	16	22,694,589	19,609,743	11,646,681	72	3,291,813	50,613	42
Lower stalk	26	51	37,555,049	32,280,766	19,117,797	73	1,307,014	39,224	43
Rhizoid	26	16	27,307,714	16,273,913	9,951,940	72	293,461	23,183	39
Cap	27	187	27,859,229	21,702,199	13,055,825	72	6,704,550	77,998	45
Upper stalk	27	155	30,596,065	23,362,341	14,172,132	73	4,437,570	59,649	42
Lower stalk	27	63	25,217,910	19,013,147	11,575,387	73	2,137,899	45,753	43
Rhizoid	27	63	29,781,682	15,189,655	9,203,797	73	480,492	33,149	46
Cap	45	713	30,795,845	26,732,918	16,179,143	73	7,615,615	71,764	45
Upper stalk	45	72	28,618,780	24,538,566	14,672,568	74	2,949,290	55,172	47
Lower stalk	45	15	29,851,320	25,462,402	15,279,659	73	1,186,788	36,048	43
Rhizoid	45	49	43,496,924	36,991,784	22,472,939	73	2,078,673	44,095	44
Cap	46	534	43,691,338	37,748,810	22,697,840	72	10,250,494	79,349	44
Upper stalk	46	211	26,586,549	22,842,662	13,930,603	73	4,370,405	61,734	45
Lower stalk	46	144	29,167,985	25,131,292	15,639,747	73	6,189,014	64,045	46
Rhizoid	46	122	28,000,697	24,124,084	14,737,583	72	5,492,434	75,135	48

transcripts with ortholog hits, where 18,779 transcripts were assigned GO-terms.

Assessing the presence of conserved eukaryotic genes in the transcriptome with a BUSCO analysis estimated a ~96% completeness based on a pan-eukaryotic dataset and a ~70% completeness based on a Chlorophyta dataset (Table 2). The pan-eukaryote dataset is smaller than the Chlorophyta dataset (303 genes vs. 2168 genes), which probably explains the differences in the fraction of genes found. Nevertheless, these results indicate that our de novo assembled transcriptome has captured the majority of the expressed genes in *Acetabularia*.

3.3. RNA distribution

For all 24 samples, more than 70% of the trimmed read pairs mapped to the transcriptome (Table 1). There was no correlation between the number of mapped reads and the total UMI count (Table 1), illustrating the extent of PCR duplication in the sequence libraries and the importance of using UMIs. Most of the transcripts were expressed at low levels, with 40–50% of transcripts having a count of two or less.

The total UMI counts follow the same distribution as the amount of isolated total RNA, with highest numbers in the cap and decreasing towards the rhizoid (Fig. 1B and C). As the dissected cap pieces were larger than the other pieces, it is also expected that the cap samples contain more RNA. However, as the same amount of RNA is sequenced from each library, the size of the pieces, nor the amount of isolated total RNA, can explain the higher count values in the cap. This rather indicates a greater diversity, or heterogeneity, of transcripts in the cap.

While the nuclear encoded mRNAs had the same distribution as the total RNA, i.e. decreasing towards the rhizoid (Fig. S1A), ribosomal RNAs were roughly evenly distributed between the different samples, albeit with a few extreme outliers (Fig. S1B). Transcripts presumably originating from the chloroplast were also distributed in an apical-basal gradient (Fig. S1C). Mitochondrial transcripts had the highest counts in the cap and the rhizoid (Fig. S1D), however these genes were much more variable between samples and the counts were also overall much lower and therefore more subject to stochastic variation.

The Principal Component Analysis (PCA) of the count variation between samples showed that the samples largely clustered according to the cell apical-basal axis along PC1 (Fig. 1D). Furthermore, the one-way PERMANOVA showed that the clustering of the different samples was significantly different from each other ($p < 0.05$). This shows that the cap- and the rhizoid are the least similar in transcript composition, while the two stalk compartments overlap and share many of the same transcripts. However, there were a tendency of clustering according to which batch the samples originated from (e.g. batch 19) or which RNA isolation method was used (e.g. batch 19 and 25 vs. batch 26, 27, 25 and 46).

3.4. Differential transcript distribution

Most of the assembled transcripts were expressed at low levels (as expected considering NGS de novo assembly artefacts and wrongly assembled isoforms). 87% of the transcripts had a mean expression of >1 TMM across the samples, and filtering nuclear encoded transcripts with a raw count of one or more in at least four samples retained 82,164 transcripts. Out of these, 13,057 transcripts were identified as significantly differentially distributed between two or more subcellular compartments. Of the differentially distributed transcripts, 2710 transcripts were unique to the cap, 255 and 259 transcripts were uniquely located in the upper- and lower stalk respectively, and 4197 transcripts were uniquely located in the rhizoid (Fig. 2A). 1617 transcripts were enriched in both the cap and the upper stalk, 163 transcripts in both the upper- and lower stalk, and 1234 transcripts were enriched in both the lower stalk and the rhizoid. Visualizing the expression of these differentially distributed transcripts (Fig. 2B) confirms the clustering analysis in that there are two large and distinct pools of enriched transcripts in the cap and the rhizoid, and that these subcellular compartments are the least similar in terms of gene content. The upper- and lower stalk samples have similar expression profiles and share a large number of differentially distributed transcripts. These two compartments also have an overall lower gene expression compared to the cap and rhizoid.

Table 2
BUSCO analysis of the de novo assembled transcriptome of *A. acetabulum*. BUSCOs refer to the genes present in the different databases of the BUSCO software. Two datasets were used in our analysis, one containing 2168 genes conserved across Chlorophyta, and one containing 303 genes conserved across eukaryotes.

	Eukaryote BUSCO hits	Chlorophyta BUSCO hits
Complete BUSCOs	245 (80.9%)	1349 (62.2%)
Single-copy	159 (52.5%)	1047 (48.3%)
Duplicated	86 (28.4%)	302 (13.9%)
Fragmented BUSCOs	45 (14.9%)	177 (8.2%)
Missing BUSCOs	13 (4.2%)	642 (29.6%)
Total BUSCOs searched	303	2168

3.5. GO enrichment

In order to investigate which genetic processes were taking place in the different subcellular compartments, we analyzed the different subcellular pools of nuclear encoded transcripts for the presence of enriched functional categories. As the two stalk samples displayed similar expression patterns, they were analyzed together (referred to as "stalk") to get a clearer picture of the differences between the stalk, cap and the rhizoid. The GO-enrichment analysis resulted in 126 enriched GO-terms in the cap, 134 in the rhizoid, and 57 in the stalk (there were eight enriched GO-terms in the upper stalk and 14 in the lower stalk when analyzed separately) (Table S7).

Nuclear encoded mRNA transcripts accumulating in the cap were enriched for GO-terms related to photosynthesis such as photosynthetic processes, chloroplast components and thylakoid (Fig. 3). General metabolic processes, organization of the plasma membrane and extracellular matrix, development and transport were also enriched in the cap, and to a lesser extent enriched in the rhizoid. No particular processes seemed to be unique to the stalk. However, the GO-term "cytoplasmic chromosome", which was also enriched in the cap, was significantly enriched in the stalk, and GO-terms related to metabolic processes, catalytic activity, cellular organelles and transport was to small degree enriched in the stalk. Nuclear encoded mRNA transcripts accumulating in the rhizoid were enriched for GO-terms related to the nucleus, replication, transcription, and cell motility. In addition, cytoskeleton organization and cell- division and differentiation were enriched in the rhizoid and also to a lesser extent in the cap. Other processes which seemed to be more widely distributed, and which were enriched in both the cap and the rhizoid, were related to transport, translation, ribosome organization, cell wall- and cell membrane organization, development and morphogenesis, and general metabolic and enzymatic processes.

3.6. Distribution of genes involved in mRNA compartmentalization

Analyses of the mRNA distribution indicated the presence of subcellular pools of functionally related transcripts. Therefore, we investigated the distribution patterns of transcripts potentially involved in generating this type of distribution.

Gene transcripts related to the cytoskeleton, such as *actin* and *tubulin*, are highly abundant along the entire cell, and not specifically associated with any particular subcellular region (Fig. 4A and B). Transcripts encoding motor proteins moving along the cytoskeleton such as myosin, dynein and kinesin, are also present throughout the cell, although not as evenly distributed as the cytoskeletal components (Fig. 4C–E). *Myosin* transcripts were more abundant in the apical end and decreasing towards the rhizoid. This includes class XIII myosin which have been

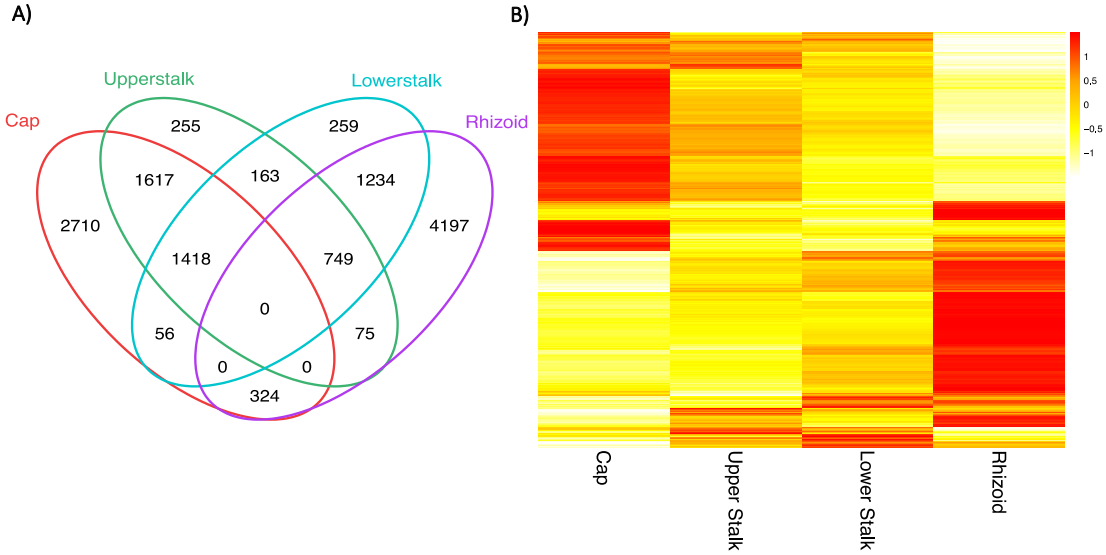

Fig. 2. Differentially distributed transcripts between the subcellular compartments. A) Venn diagram showing the shared and unique number of transcripts that are differentially distributed between the subcellular compartments (DESeq2 adjusted p-values < 0.05). B) Heatmap of the differentially distributed transcripts. Colors represent scaled TMM (Trimmed Mean of M-values) expression values (from −1 to 1). The mean TMM values across the different samples from each subcellular structure are shown.

shown to be involved in organelle transport and tip growth in *A. cliftonii* and enriched in the apical regions of the cell [49]. The same trend was observed also for kinesins, except for two transcripts which were abundant in the rhizoid. Interestingly, these two transcripts have the closest blast hits against kinesin 13 and 14, which are known to move in both directions on the microtubule and can thereby travel in the opposite direction on the microtubules than the other kinesins.

In contrast, the *dyneins* were overall lesser expressed than *myosins* and *kinesins*. The most highly abundant *dynein* was present through the cell in roughly equal amounts, while the rest of the transcripts were most abundant in the rhizoid. Myosins and kinesis generally move towards the plus-ends of the polarized actin microfilaments and microtubules respectively, and thus from the nucleus towards the cell membrane. While dyneins move towards the minus end of microtubules towards the cell interior [50]. Hence, motor proteins moving towards the cell membrane are of a slightly higher abundance in the apical part of the cell (expect for the two kinesin transcripts which have a higher abundance in the basal part of the cell), while motor proteins moving towards the cellular interior are seemingly of a higher abundance in the basal part of the cell (lower stalk and rhizoid).

3.7. Vesicular transport is a fundamental mechanism for intracellular transport of cargo

in eukaryote cells, and has been associated with intracellular transport of RNA [51–53]. Vesicle formation relies on coat proteins, and COPI- COPII- and Clathrin coated vesicles are the main type of vesicles in eukaryote cells. COPI-coated vesicles move from ER to golgi, COPII-coated vesicles move between parts of golgi and retrograde transport from golgi to ER, and Clathrin-coated vesicles move from golgi to the plasma membrane [54]. In our results, two of three *COPI* transcripts were most abundant in the cap and decrease towards the rhizoid, while one *COPII* transcript was most abundant in the rhizoid (Fig. 4F). Two *clathrin* homologs were distributed throughout the cell, one of these transcripts was of noticeable higher abundance than the other (Fig. 4G). This high abundant transcript was also of a slightly higher concentration in the cap and decreasing towards the rhizoid.

Three copies of poly(A) polymerases were expressed in *A. acetabulum*. All three were evenly distributed throughout the cell, however one was higher expressed than the others (Fig. 4H).

3.8. Comparative transcriptomics between Acetabularia acetabulum and Caulerpa taxifolia

Orthology searches between the transcriptomes of *A. acetabulum* and *C. taxifolia* identified 4483 orthogroups represented by at least one transcript from each species. Of these orthogroups, 2120 were single-copy orthologues with a single representative gene from each species. Comparing the different subcellular samples from the two species based on expression dynamics of these single-copy orthologues showed that the samples clustered strongly according to species along PC1 (Fig. 5). However, there was also a slight tendency that apical and basal samples clustered together along PC2, indicating some similarity between apical and basal cell sections between the two species.

4. Discussion

4.1. Apical-basal mRNA gradient in A. acetabulum

A. acetabulum has been used as a model system for cell morphogenesis for decades, and it has been suspected that differential distribution of RNA along the cell axis is an underlying mechanism for its sophisticated morphology [7,9,18,20]. To investigate the distribution of mRNA in adult *A. acetabulum* cells we have performed subcellular mRNA sequencing and functional enrichment analysis. We have tagged each mRNA molecule with unique molecular indexes (UMIs) which allows for true quantification of mRNA by eliminating the effect of amplification bias introduced during library preparation. Despite isolating RNA from subcellular regions, we were able to capture the majority of expressed transcripts in sufficient quantities for RNA sequencing. Although there was some variation between samples, the procedure was repeatable and robust.

Our results demonstrate the presence of RNA throughout the entire cell length. We also found the highest amount of mRNA at the apical end of the cell (the cap) with decreasing concentration towards the basal end (the rhizoid), confirming earlier discoveries of an apical-basal gradient of RNA in *A. acetabulum* [15,55,56]. However, while it has been believed that this gradient is due to different concentrations of RNA encoded by the chloroplasts, and not the nucleus [16], we demonstrate the opposite.

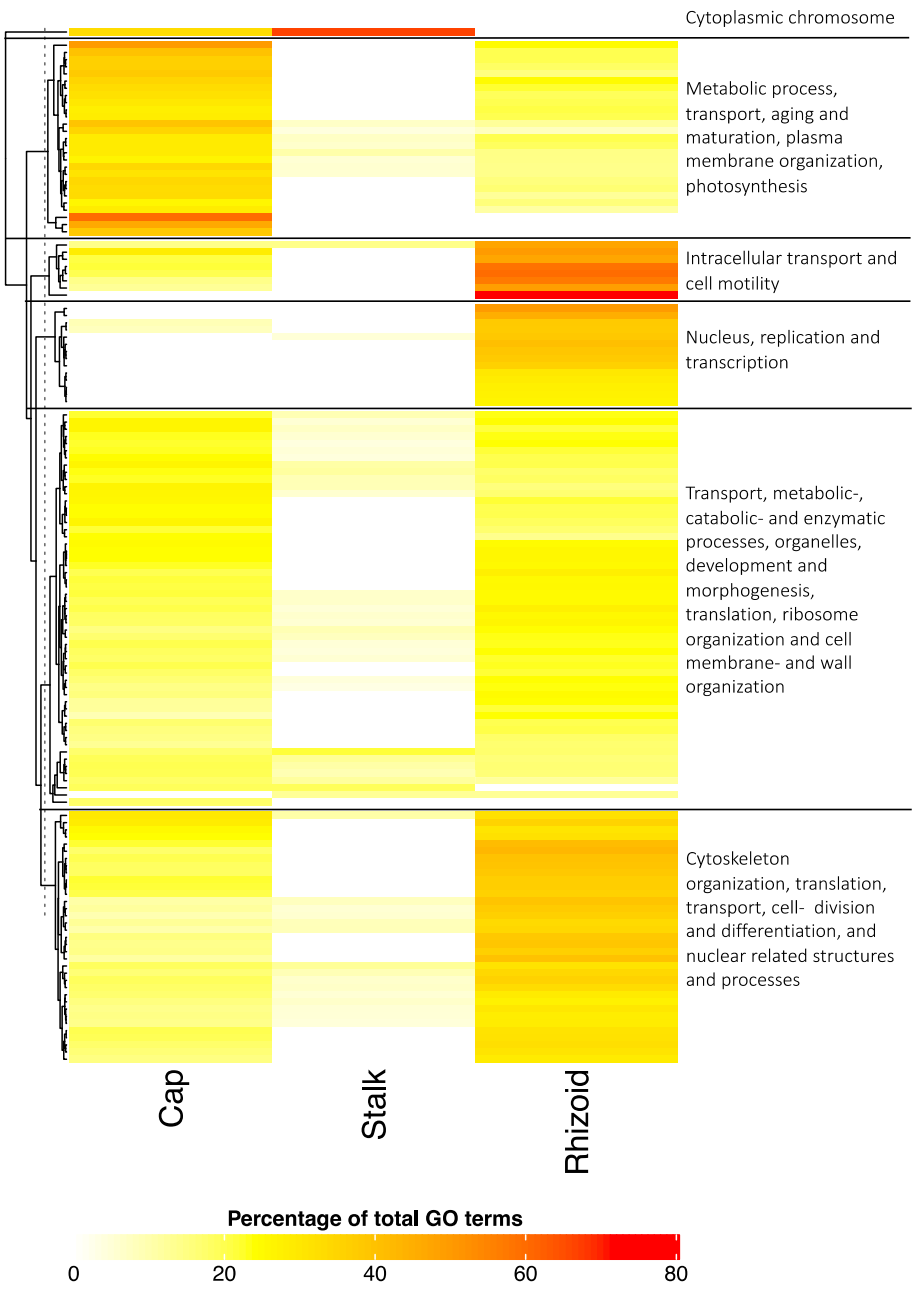

Fig. 3. GO-enrichment analysis of transcripts differentially distributed between subcellular compartments. Heatmap of the enriched GO-terms (FDR < 0.05) among the differentially distributed transcripts in each subcellular compartment (note that the stalk samples are analyzed together). The colors indicate the percentage of differentially distributed transcripts annotated with a given GO-category compared to the total number of transcripts in the same GO-category. All GO-categories (Biological Process, Cellular Compartment and Molecular Function) are shown together. A GO-term not significantly enriched in a subcellular compartment is set to 0% (hence shown in white color). The most prevalent GO-terms have been simplified and highlighted on the right side of the heatmap. See Table S7 for a full description of the GO-enrichment results.

4.2. Localized pools of transcripts support subcellular mRNA compartmentalization

A long-standing question has been whether the observed gradient of mRNA in *A. acetabulum* is homogeneous in transcript composition, or whether there are distinct pools of transcripts along the cell. Hämmerling's grafting experiments suggested the existence of local determinants of morphogenesis [9,57], and Dumais et al. [7] speculated that mRNAs would either be distributed throughout the cell, or localized to the apical or basal ends. Our results show that while some gene transcripts are distributed evenly across the cell, a large part are located to different subcellular compartments. Although we have only studied the expressed mRNAs, and it is not entirely certain how their distribution reflects protein levels, it has been shown that RNA-seq is able to capture relative protein differences, especially for differentially expressed transcripts [58]. We also found that the differentially distributed pools of transcripts are composed of functionally related transcripts. Transcripts related to photosynthesis are co-localized and accumulate in the apical end of the cell, while transcripts related to nuclear processes co-localized in the basal end. This pattern shows that the RNA gradient is not a homogeneous mix of gene transcripts, which confirms that mechanisms to ensure specific and functional RNA localization must be in place in *A. acetabulum*.

There were overall fewer transcripts localized in the stalk. The stalk is mainly filled with a central vacuole, with only a thin layer of cytoplasm covering it [7], leaving very little room for other subcellular structures or pools of transcripts. We therefore assume that there are very few processes occurring exclusively in the stalk, and that the structure might even function as a physical barrier between the cap and the rhizoid, limiting the mixing of molecules between compartments, but alternatively allowing transport between them.

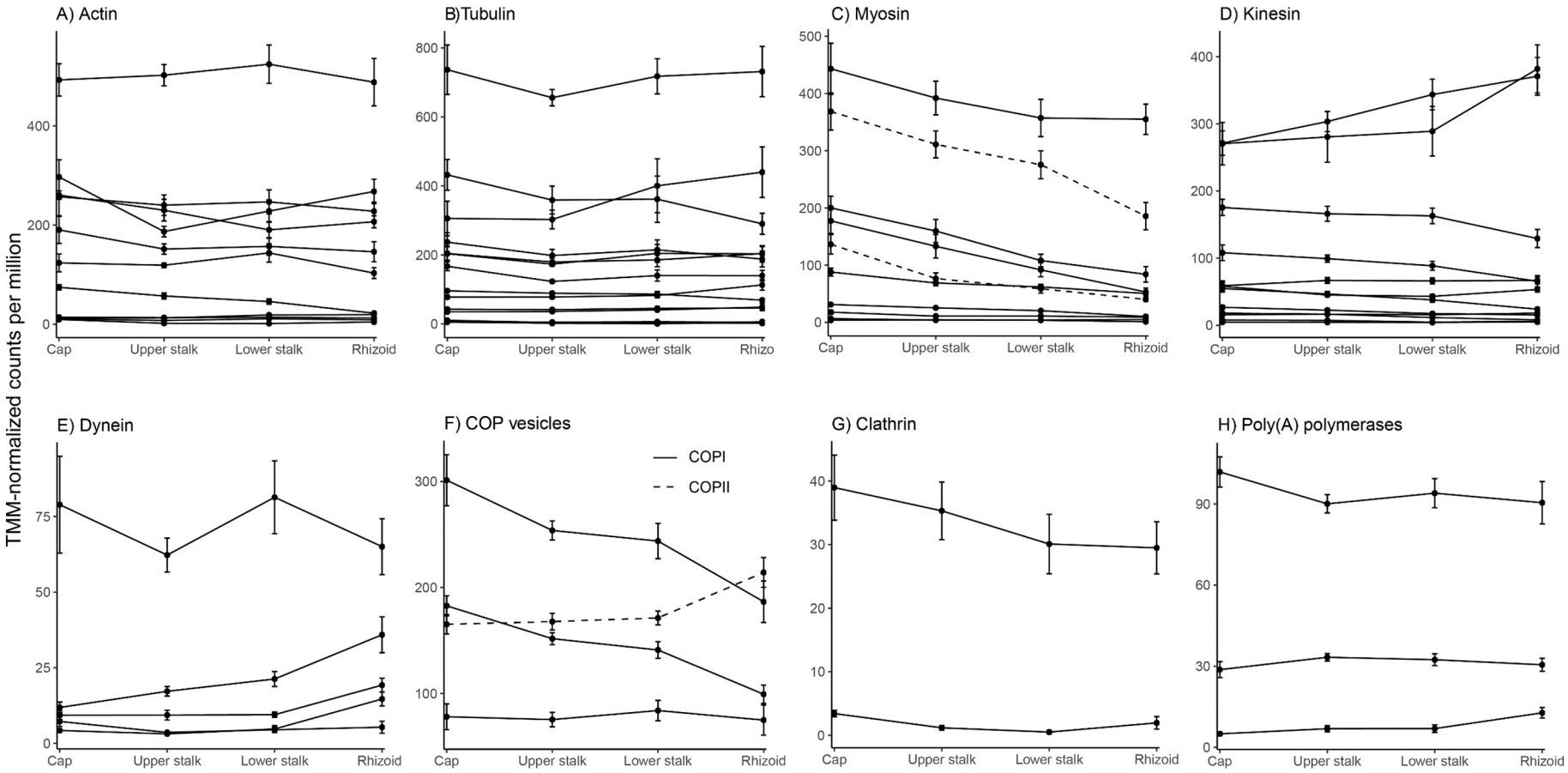

Fig. 4. Subcellular distribution of transcripts associated with transport and RNA localization. The expression levels of homologs of *actin* (A), *tubulin* (B), *myosin* (C), *kinesin* (D), *dynein* (E), *COP* vesicle genes (F), *clathrin* (G), and poly(A) polymerases (H) in the different subcellular sections are plotted. The dashed lines in C) indicate homologs of Class XIII myosin identified in *Acetabularia cliftonii*. In F), solid lines indicate *COPI* homologs and the dashed line indicate a *COPII* homolog. Expression values are shown on the y-axis as TMM-normalized counts per million, with error bars representing standard error of the mean (n = 6). Only transcripts with a mean expression value >1 across all samples are shown.

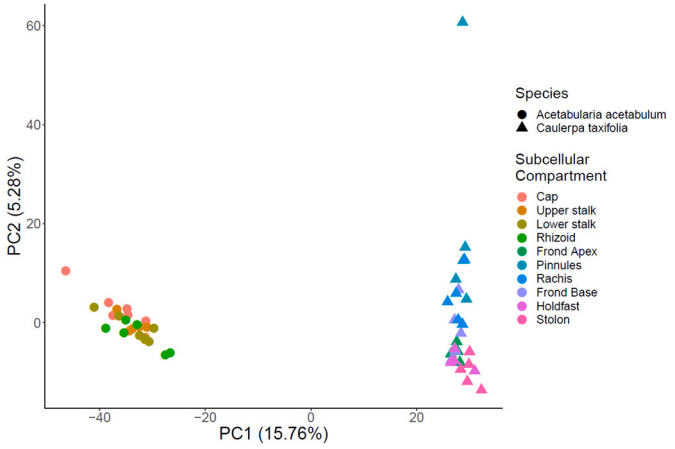

Fig. 5. Comparison of the expression profile of gene orthologues between subcellular regions of *A. acetabulum* and *C. taxifolia*. Principal Component Analysis (PCA) of the sample variation based on TMM-normalized counts per million (see Methods) of single-copy orthologues between *A. acetabulum* (circles) and *C. taxifolia* (triangles). Colors represent different subcellular regions. Frond Apex, Pinnules and Rachis are apical regions of *C. taxifolia*, while Holdfast, Stolon and Frond Base are basal regions.

4.3. Active mRNA transport is likely the main mechanism for establishment of cell polarity

RNA can be distributed around a cell either by passive diffusion from the nucleus, or by active transport along the cytoskeleton [59]. Studies tracking the movement of radioactively labelled RNA have shown that mRNA travels faster in *A. acetabulum* than what is possible by diffusion alone [22], suggesting active transport of mRNA. Also, simply the size of the cell, with the nucleus and the cap separated by several centimeters, puts obvious demands on active intracellular transport. A highly sophisticated and extensively developed cytoskeleton has been overserved in *A. acetabulum*, with large tracks of actin filaments running the entire length of the cell [60]. Experiments by Mine et al. [24] showed that inhibiting actin polymerization with cyclohalasin D disrupts the established mRNA gradients, indicating an association between mRNA and the cytoskeleton. As predicted, we find that transcripts encoding the main cytoskeletal components such as actin and tubulin are uniformly distributed throughout the cell. Furthermore, we see that transcripts encoded by both *clathrin* and *COP* genes, as well as transcripts encoding different homologs of bidirectional cytoskeletal motor proteins, are distributed throughout the cell. This observation suggests that these types of vesicular transport systems are active in the entire cell and should be investigated further through protein localization studies.

4.4. mRNA stabilization and post-transcriptional control

That some transcripts are evenly distributed while others are localized to subcellular regions, implies that the cell is able to distinguish between which mRNAs should be transported where. In addition, mechanisms for selective stabilization and degradation of mRNAs at different locations in the cell are likely in place. While actin microfilaments are present throughout the cell during the entire life cycle of *A. acetabulum*, microtubules do not appear until the final stages of development where they serve as transport tracks in the cap [60,61]. This is interesting as tubulin genes are expressed earlier and distributed throughout the cell, suggesting tubulin mRNAs are stabilized and stored in the cytoplasm and further protected from degradation. mRNAs located in the cell tip have been estimated to be at least three days old because of the long distance they must travel from the nucleus [25]. That mRNA is long-lived in *A. acetabulum* cells is supported by experiments showing that development and morphogenesis can continue for days, and even weeks, after amputation of the nucleus [62,63], and that radioactive RNAs exist in *A. acetabulum* cells long after treatment with radioactive labelled UTP [64,65]. It is also interesting that we find transcripts encoding polyadenylation proteins distributed throughout the cell, as editing, shortening and elongation of the poly-A-tail is an important mechanism for translational control [66,67].

4.5. The cap is the main morphogenetic and metabolic structure

The GO-enrichment analysis suggests a higher level of catalytic- and metabolic activity in the cap compared to the rhizoid, indicating the cap as metabolically more diverse and active. We also see the greatest diversity of expressed transcripts in the cap, and the highest overall RNA content. These findings agree with earlier observations that the cap has the highest morphogenetic capacity, or developmental potential, as it can regenerate both whorls of hair in addition to the entire cap structure after rhizoid dissection [20,60]. Conversely, the nucleus is predominantly responsible for production of mRNAs and replication.

4.6. Ortholog comparison indicates little genetic homology between subcellular regions of A. acetabulum and Caulerpa taxifolia

Caulerpa taxifolia is another Chlorophyte alga with many similarities to *A. acetabulum*, most notably they are both gigantic single-celled species with highly complex cellular morphologies with clearly distinguished apical and basal ends. However, unlike *A. acetabulum*, *C. taxifolia* is a syncytium with hundreds, or even thousands, of nuclei scattered throughout the cell. Ranjan et al. [4] characterized the gene expression patterns of the different subcellular sections of *C. taxifolia* and found that they contained unique expression profiles, similar to that observed in *A. acetabulum*. However, comparing the expression profile of single-copy orthologs between the two species shows that the subcellular sections are much more similar within species than between species, hence there is little homology at the genetic level. Nevertheless, comparing the functional annotations of these genes indeed shows some similarities. Both species are enriched for nuclear components and DNA-related processes, such as DNA replication and transcription in the basal region, as well as displaying a higher catalytic activity in the apical part [4]. Therefore, although morphologically similar cell regions of these two species contain largely non-orthologous mRNAs, they share overall similar genetic functions. However, as *A. acetabulum* only has a single nucleus, while *C. taxifolia* is a syncytium, these similarities must be an evolutionary convergence on a functional level in the two species as the underlying gene regulation is different; post-transcriptional regulation is probably necessary to achieve mRNA compartmentalization in *A. acetabulum*, while transcriptional regulation is more dominating in *C. taxifolia* [4,21].

4.7. Potential of subcellular RNA-seq on Acetabularia cells

A. acetabulum has been an important model in cell biology for studying the link between transcriptional control and cell morphogenesis since the 1930s. Here, we have attempted to re-introduce *A. acetabulum* as a model system by developing RNA-seq methodology to investigate subcellular localization of mRNA. The data presented here illustrates the usefulness of being able to isolate and sequence mRNA from subcellular regions, which is possible because of the size and elongated polarized shape of *A. acetabulum* cells. Though similar investigations can be performed on other macroscopic green algal cells, such as *C. taxifolia* [4,21], *Acetabularia* only has a single nucleus. In this respect it better resembles a general eukaryote cell and is a more appropriate model for understanding general cell developmental mechanisms.

Statement of informed consent, human/animal rights

No conflicts, informed consent, or human or animal rights are applicable to this study.

CRediT authorship contribution statement

Ina J. Andresen: Data curation, Formal analysis, Investigation, Visualization, Methodology, Writing - original draft, Writing - review & editing. **Russell J. S. Orr**: Supervision, Writing - review & editing. **Kamran Shalchian-Tabrizi**: Conceptualization, Funding acquisition, Project administration, Supervision, Methodology, Writing - review & editing. **Jon Bråte**: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are grateful to Professor William Martin at the Heinrich-Heine-Universität in Düsseldorf, Germany for providing us with embryos of *A. acetabulum*. We thank the Norwegian Sequencing Centre (www.sequencing.uio.no) for library preparation and sequencing, and Uninett/Sigma2 for providing a high-performance computing service. This work has been funded by a grant from the Centre for Integrative Microbial Evolution (CIME) at UiO to KS-T, and by the Norwegian Research Council through a grant to JB (grant nr. 240284 and 306571). We also thank the reviewers for valuable comments and suggestions to improve this text.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.algal.2021.102440.

References

[1] D.R. Kaplan, W. Hagemann, The relationship of cell and organism in vascular plants - are cells the building-blocks of plant form, Bioscience 41 (1991) 693–703.
[2] D.R. Kaplan, The relationship of cells to organisms in plants - problem and implications of an organismal perspective, Int. J. Plant Sci. 153 (1992) S28–S37.
[3] K.J. Niklas, E.D. Cobb, D.R. Crawford, The evo-devo of multinucleate cells, tissues, and organisms, and an alternative route to multicellularity, Evol Dev 15 (2013) 466–474.
[4] A. Ranjan, B.T. Townsley, Y. Ichihashi, N.R. Sinha, D.H. Chitwood, An intracellular transcriptomic atlas of the giant coenocyte Caulerpa taxifolia, PLoS Genet. 11 (2015), e1004900.
[5] S. Berger, Photo-atlas of Living Dasycladales, Carnets de Géologie, Brest, 2006.
[6] S. Berger, Dasycladaceae - a family of giant unicellular algae ideal for research, Nato Adv Sci I a-Lif 195 (1990) 3–19.
[7] J. Dumais, K. Serikawa, D.F. Mandoli, Acetabularia: a unicellular model for understanding subcellular localization and morphogenesis during development, J. Plant Growth Regul. 19 (2000) 253–264.
[8] S. Berger, Photo-atlas of Living Dasycladales, Carnets de Geologie, 2006.
[9] J. Hämmerling, On substances determining the development in Acetabularia mediterranea, their spatial and temporal distribution and origin, Roux Arch Dev Biol 131 (1934) 1–81.
[10] J. Hämmerling, On genome effects and formation ability in Acetabularia, Roux Arch Dev Biol 132 (1934) 424–462.
[11] J. Hämmerling, Nucleo-cytoplasmic relationships in the development of Acetabularia, in: G.H. Bourne, J.F. Danielli (Eds.), International Review of Cytology, Academic Press, 1953, pp. 475–498.
[12] J. Hämmerling, Physiological development and genetic foundation of the form creation in the shielding algae Acetabularia, Naturwissenschaften 32 (1934) 829–836.
[13] J. Hämmerling, Nucleo-cytoplasmic interactions in Acetabularia and other cells, Annu. Rev. Plant Physiol 14 (1) (1963) 65–92, https://doi.org/10.1146/annurev.pp.14.060163.000433.
[14] K. Kloppstech, Messenger-type RNA from Acetabularia, Protoplasma 91 (1977) 221–228.
[15] E. Baltus, J.E. Edstrom, M. Janowski, R. Tencer Hanocqqu.J, J. Brachet, Base composition and metabolism of various RNA fractions in Acetabularia Mediterranea, P Natl Acad Sci USA 59 (1968) 406.
[16] L.P. Naumova, E.K. Pressman, L.S. Sandakchiev, Gradient of RNA distribution in cytoplasm of Acetabularia-Mediterranea, Plant Sci Lett 6 (1976) 231–235.
[17] E. Garcia, A.C. Dazy, Spatial-distribution of poly(a)+ RNA and protein-synthesis in the stalk of Acetabularia-Mediterranea, Biol. Cell. 58 (1986) 23–29.
[18] H. Vogel, G.E. Grieninger, K.H. Zetsche, Differential messenger RNA gradients in the unicellular alga Acetabularia acetabulum. Role of the cytoskeleton, Plant Physiol 129 (2002) 1407–1416.
[19] K.A. Serikawa, D.F. Mandoli, Aaknox1, a kn1-like homeobox gene in Acetabularia acetabulum, undergoes developmentally regulated subcellular localization, Plant Mol. Biol. 41 (1999) 785–793.
[20] K.A. Serikawa, D.M. Porterfield, D.F. Mandoli, Asymmetric subcellular mRNA distribution correlates with carbonic anhydrase activity in Acetabularia acetabulum, Plant Physiol. 125 (2001) 900–911.
[21] A. Arimoto, K. Nishitsuji, Y. Higa, N. Arakaki, K. Hisata, C. Shinzato, N. Satoh, E. Shoguchi, A siphonous macroalgal genome suggests convergent functions of homeobox genes in algae and land plants, DNA Res. 26 (2019) 183–192.
[22] K. Kloppstech, H.G. Schweiger, Polyadenylated RNA from Acetabularia, Differentiation 4 (1975) 115–123.
[23] K. Kloppstech, H.G. Schweiger, 80 S ribosomes in Acetabularia major. Distribution and transportation within the cell, Protoplasma 83 (1975) 27–40.
[24] I. Mine, K. Okuda, D. Menzel, Poly(A)(+) RNA during vegetative development of Acetabularia peniculus, Protoplasma 216 (2001) 56–65.
[25] I.M. Henry, M.D. Wilkinson, J.M. Hernandez, Z. Schwarz-Sommer, E. Grotewold, D.F. Mandoli, Comparison of ESTs from juvenile and adult phases of the giant unicellular green alga Acetabularia acetabulum, BMC Plant Biol. 4 (2004) 3.
[26] M. Toloue, J. Risinger, A molecular indexing for improved RNA-Seq, J. Biomol. Tech. (25) (2014) S12. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4162288/.
[27] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[28] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359.
[29] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q.D. Zeng, Z.H. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–U130.
[30] B. Bushnell, BBMap Short-read Aligner, and Other Bioinformatics Tools, 2015.
[31] F.A. Simao, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (2015) 3210–3212.
[32] S.U. Schneider, M.B. Leible, X.P. Yang, Strong homology between the small subunit of ribulose-1,5-bisphosphate carboxylase oxygenase of 2 species of Acetabularia and the occurrence of unusual codon usage, Mol. Gen. Genet. 218 (1989) 445–452.
[33] T.H. Jukes, Neutral changes and modifications of the genetic code, Theor. Popul. Biol. 49 (1996) 143–145.
[34] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M.D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis, Nat. Protoc. 8 (2013) 1494–1512.
[35] J. Huerta-Cepas, K. Forslund, L.P. Coelho, D. Szklarczyk, L.J. Jensen, C. von Mering, P. Bork, Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper, Mol. Biol. Evol. 34 (2017) 2115–2122.
[36] J. Huerta-Cepas, D. Szklarczyk, D. Heller, A. Hernandez-Plaza, S.K. Forslund, H. Cook, D.R. Mende, I. Letunic, T. Rattei, L.J. Jensen, C. von Mering, P. Bork, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (2019) D309–D314.
[37] R. Bacher, L.F. Chu, N. Leng, A.P. Gasch, J.A. Thomson, R.M. Stewart, M. Newton, C. Kendziorski, SCnorm: robust normalization of single-cell RNA-seq data, Nat. Methods 14 (2017) 584.
[38] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.
[39] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016.
[40] M.J. Anderson, A new method for non-parametric multivariate analysis of variance, Austral Ecology 26 (2001) 32–46.
[41] J. Oksanen, R. Kindt, P. Legendre, B. O'Hara, G.L. Simpson, P. Solymos, M.H. H. Stevens, H. Wagner, vegan: Community Ecology Package, 2008.
[42] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B 57 (1995) 289–300.
[43] T.W.H. Backman, T. Girke, systemPipeR: NGS workflow and report generation environment, BMC Bioinformatics 17 (2016) 388.
[44] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[45] D.J. McCarthy, Y.S. Chen, G.K. Smyth, Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation, Nucleic Acids Res. 40 (2012) 4288–4297.

[46] M.D. Young, M.J. Wakefield, G.K. Smyth, A. Oshlack, Gene ontology analysis for RNA-seq: accounting for selection bias, Genome Biol. 11 (2010).
[47] Z.G. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data, Bioinformatics 32 (2016) 2847–2849.
[48] D.M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy, Genome Biol. 16 (2015).
[49] O. Vugrek, H. Sawitzky, D. Menzel, Class XIII myosins from the green alga Acetabularia: driving force in organelle transport and tip growth? J. Muscle Res. Cell Motil. 24 (2003) 87–97.
[50] B. Alberts, A. Johnson, J. Lewis, R. Martin, K. Roberts, P. Walter, Molecular Biology of the Cell, 4 ed., Garland Science, New York, 2002.
[51] J. Skog, T. Wurdinger, S. van Rijn, D.H. Meijer, L. Gainche, M. Sena-Esteves, W. T. Curry, B.S. Carter, A.M. Krichevsky, X.O. Breakefield, Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers, Nat. Cell Biol. 10 (2008) 1470–U1209.
[52] E. Basyuk, T. Galli, M. Mougel, J.M. Blanchard, M. Sitbon, E. Bertrand, Retroviral genomic RNAs are transported to the plasma membrane by endosomal vesicles, Dev. Cell 5 (2003) 161–174.
[53] C.T. Roberts Jr., P. Kurre, Vesicle trafficking and RNA transfer add complexity and connectivity to cell-cell communication, Cancer Res. 73 (2013) 3200–3205.
[54] N. Gomez-Navarro, E.A. Miller, COP-coated vesicles, Curr. Biol. 26 (2016) R54–R57.
[55] J. Hämmerling, Studien zum Polaritätsproblem, Zoologische Jahrbücher 56 (1936) 441–483.
[56] G. Werz, Kernphysiologische Untersuchungen an Acetabularia, Planta 46 (1955) 113–153.
[57] J. Hämmerling, Dreikernige Transplantate zwischen Acetabularia crenulata und mediterranea I, Zeitschrift Naturforschung Teil A 1 (1946) 337–342.
[58] A. Koussounadis, S.P. Langdon, I.H. Um, D.J. Harrison, V.A. Smith, Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system, Sci. Rep. 5 (2015) 10775.
[59] D. St Johnston, Moving messages: the intracellular localization of mrnas, Nat Rev Mol Cell Bio 6 (2005) 363–375.
[60] D. Menzel, Cell differentiation and the cytoskeleton in Acetabularia, New Phytol. 128 (1994) 369–393.
[61] D. Menzel, Visualization of cytoskeletal changes through the life-cycle in Acetabularia, Protoplasma 134 (1986) 30–42.
[62] H. Stich, W. Plaut, The effect of ribonuclease on protein synthesis in nucleated and enucleated fragments of Acetabularia, J. Biophys. Biochem. Cytol. 4 (1958) 119–121.
[63] V.G. Yasinovski, T.N. Zubarev, N.P. Rogatykh, I.V. Yanushevich, Kinetics of synthesis and distribution of the morphogenetic substances in whole cells and in anucleate fragments of Acetabularia mediterranea, in: S. Bonotto, V. Kefeli, S. Puiseux-Dao (Eds.), Developmental Biology of Acetabularia, Elsevier/North_ Holland Biomedical Press, Amsterdam, Holland New York, NY Oxford, England, 1979, pp. 65–70.
[64] K. Kloppstech, H.G. Schweiger, Stability of poly(A)(+)RNA in nucleate and anucleate cells of Acetabularia, Plant Cell Rep. 1 (1982) 165–167.
[65] H.-G. Schweiger, Transcription of the nuclear genome of Acetabularia, in: L. Bogorad, J.H. Weil (Eds.), Nucleic Acids and Protein Synthesis in Plants, Springer US, Boston, MA, 1977, pp. 65–83.
[66] M. Wickens, Forward, backward, how much, when: mechanisms of poly(A) addition and removal and their role in development, Semin. Dev. Biol. 3 (1992) 399–412.
[67] R. Aphasizhev, RNA uridylyltransferases, Cell. Mol. Life Sci. 62 (2005) 2194–2203.