MBMG Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 1: e19862
doi: 10.3897/mbmg.1.19862

OPEN ACCESS

Research Article

Functional and phylogenetic analysis of the core transcriptome of Ochromonadales

Nadine Graupner[‡], Jens Boenigk[‡,§], Christina Bock[‡], Manfred Jensen[‡], Sabina Marks[‡], Sven Rahmann[|], Daniela Beisser[‡]

‡ Biodiversity, University of Duisburg-Essen, Essen, Germany
§ Centre for Water and Environmental Research (ZWU), University of Duisburg-Essen, Essen, Germany
| Genome Informatics, University of Duisburg Essen, University Hospital Essen, Essen, Germany

Corresponding author: Nadine Graupner (nadine.graupner@uni-due.de)

Academic editor: Thorsten Stoeck

Received: 27 Jul 2017 | Accepted: 30 Aug 2017 | Published: 20 Sep 2017

Citation: Graupner N, Boenigk J, Bock C, Jensen M, Marks S, Rahmann S, Beisser D (2017) Functional and phylogenetic analysis of the core transcriptome of Ochromonadales. Metabarcoding and Metagenomics 1: e19862. https://doi.org/10.3897/mbmg.1.19862

Abstract

Background

Most protist lineages consist of members with diverging features e.g. different modes of nutrition and adaptations for life in different habitat types and climatic zones. The nutritional mode is particularly variable in chrysophytes and they are therefore an excellent model group to study the core genes and metabolic pathways of a functionally diverse lineage. The objective of our study is the identification of the joint genetic repertoire expressed in closely related chrysophytes as well as the extent of variation on species and strain level. Therefore, we investigated the transcriptomes of six strains belonging to four species of Ochromonadales. We performed analyses on metabolic pathway level as well as on sequence level.

Results

We could identify 1,574 core genes shared between all six investigated strains of Ochromonadales. Most of these core genes were affiliated with the primary metabolism. Phylogenetic analysis of 166 protein-coding core genes supported a close relation of

Poteriospumella lacustris and *Poterioochromonas malhamensis* and resolved for more than 50% of investigated genes the relationship of strains affiliated with the species *P. lacustris*. Further, we found diverging phylogenetic patterns for genes interacting with the environment.

Conclusions

In Ochromonadales, a functionally diverse lineage, the core transcriptome represents only a minor part of the individual transcriptomes. But this small fraction of genes comprises the basal metabolism essential for life in several protist lineages. Phylogenetic analyses of these genes indicate a similar degree of conservation as observed for genes coding for ribosomal proteins.

Keywords

Chrysophyceae, protist, expressed sequence tags (EST), evolutionary ecology, phylotranscriptomics, core genes, metabolic pathways

Introduction

Organisms and their genomes evolved under persistently fluctuating ecological and evolutionary pressures. As a consequence, eukaryotic genomes and their gene content vary considerably in size ranging from 2,000 to 35,000 genes (Koonin 2003). The gene repertoire effectively used by the organisms is represented by their transcriptomes and usually it is a subset of the organisms' gene inventory reflecting the adaptation and acclimatization to their environment (Cox et al. 2012, Caron et al. 2016). The core transcriptome contains many essential genes and pathways of a taxonomic group (Yang et al. 2016) which is particularly interesting in functionally differentiated groups such as the chrysophytes with their different modes of nutrition and adaptations to various environmental factors. Separating genes of the core transcriptome from strain- and taxon-specific genes provides insights into the functional basis as well as ecological adaptation and differentiation. At the same time, gene content as well as gene phylogenies will serve as source for better knowledge of the evolutionary history of the group. Thus, the analysis of transcriptomes is a key to the eco-evolutionary history of the organisms linking phylogenetics with functionality.

Previous investigations of core genes aimed at various research topics spanning from the identification of the minimal gene set necessarily for cellular life (Gibson et al. 2008, Gil et al. 2004, Koonin 2003) and the reconstruction of the evolutionary history of a lineage (Bowler et al. 2008, Hsiang and Baillie 2005, Armbrust et al. 2004) to the identification of group specific core genes as a basis for barcoding approaches (Segata et al. 2012) and for the identification of novel features within lineages (Bowler et al. 2008). Gil et al. (2004) reported a minimal gene set necessary for life for bacteria as low as 250 genes and Koonin (2003) estimated the smallest genomes of free-living organisms to be around 500 genes for bacteria and around 2,000 genes for eukaryotes. Based on whole genomes 60 to 80 genes, mostly affiliated with translation, were so far identified as core genes of Eukarya, i.e. occuring in all lineages of Eukaryotes (Harris et al. 2003, Koonin 2003). Between some lineages or within one group the number of core genes is higher but varies considerably between studies. Bowler et al. (2008) and Radakovits et al. (2012) identified between 1,666 and 3,063 genes that were shared between taxa of different supergroups and Hsiang and Baillie (2005) identified 3,340 genes shared by different fungi. Also the number of genes reported to be exclusive to one lineage varied (e.g. 17 genes in the study of Hsiang and Baillie (2005) were not detectable in other lineages, whereas Bowler et al. (2008) reported around 1,000 diatom specific core genes). However, most above-mentioned studies were based on genome comparisons. For the active part of the genome, i.e. the transcriptome, investigations of core genes are scarce. Estimates on the core transcriptome within a lineage to date indicate a core set of at least 1,400 to 3,000 genes (Koid et al. 2014, Beisser et al. 2017, Liu et al. 2016). Here we focus on the comparative transcriptomics of members of Ochromonadales (Chrysophyceae) aiming at the genetic repertoire essential for chrysophytes.

Further, we ask the question to what extent do gene phylogenies reflect the history, i.e. phylogeny, of the organism and to what extent are these pattern concealed by ecological adaptation. Recent studies (Baldauf et al. 2000, Stoeck et al. 2008) have shown that phylogenies of protein-coding genes can deviate from ribosomal gene phylogenies, and that the inferred phylogenetic position of strains may depend in turn on the choice of the marker gene. Transcriptome-based and genome-based phylogenies have the potential to provide a more detailed view as the great variety of protein coding genes may better reveal the phylogenetic history. Some recent studies from various eukaryotic lineages already used transcriptomic data for protein-coding multigene phylogeny (Wickett et al. 2014, Sun et al. 2014, Wang et al. 2014, Deng et al. 2015, Wodniok et al. 2011, Price and Bhattacharya 2017). However, most previous studies on chrysophytes (Boenigk et al. 2005, Grossmann et al. 2016, Andersen et al. 1999, del Campo and Massana 2011, Scoble and Cavalier-Smith 2014, Škaloud et al. 2013) focused on ribosomal gene phylogenies. Among the few exceptions are e.g. Stoeck et al. (2008) who used alpa-tubulin, beta-tubulin and actin for the multigene phylogeny, and Beisser et al. (2017) who used transcriptomes for a multigene phylogeny of eight protein-coding genes as well as an alignment-free phylogenetic approach of whole transcriptomes. Both studies resulted in a diverging branching order of taxa compared to those of the ribosomal gene phylogenies (Boenigk et al. 2005, Grossmann et al. 2016), highlighting the need for in-depth phylogenetic analyses. Here we analyze the phylogenetic implications of tree topologies for 166 protein-coding genes based on the core transcriptome of chrysophytes affiliated with Ochromonadales.

We used six strains affiliated with four species within the order Ochromonadales (Chrysophyceae), i.e. *Poteriospumella lacustris* (JBC07, JBM10, JBNZ41), *Poterioochromonas malhamensis* (DS), *Spumella vulgaris* (199hm) and *Pedospumella encystans* (JBMS11), which were part of the overarching transcriptome study of Beisser et al. (2017) that focused on differences based on the nutritional modes. In contrast to the former study we here focus on the implications of gene phylogenies and functional implications in particular of genes affiliated with the core transcriptome. The strains reflect the broad spectrum of characteristic features typical for Ochromonadales as they originate from different habitat types i.e. freshwater and soil, perform diverging modes of nutrition, i.e. mixotrophy and heterotrophy, originate from different climatic zones, i.e. arctic to temperate zones and at least one strain is able to produce cysts. We are interested in what they have in common, despite the differences mentioned above and aim at analyzing the active core genes on functional and phylogenetic level.

Material and methods

Cultivation of strains and RNA isolation

We focused on strains affiliated with the order Ochromonadales. For our study we used six strains affiliated with four species: *Poteriospumella lacustris*, *Poterioochromonas malhamensis*, *Spumella vulgaris* and *Pedospumella encystans*. For further information regarding origin, mode of nutrition and culturing conditions see Table 1. All cultures were harvested in exponential growth phase via centrifugation at 3,000 g for 5 to 10 minutes at 20°C and RNA isolation was performed under sterile conditions using TRIzol (Life Technologies, Paisley) (see Beisser et al. 2017). Quality and quantity of RNA was assessed using the Nanodrop2000 spectrometer (Thermo Fischer Scientific Inc.) and by the sequencing provider (Eurofins MWG, Ebersberg, Germany).

Table 1.
Overview of species regarding to phylogeny, geographic & habitat origin and mode of nutrition as well as their cultivation conditions.

	Poteriospumella lacustris	***Poterioochromonas malhamensis***	***Spumella vulgaris***	***Pedospumella encystans***
Strain	JBC07, JBM10, JBNZ41	DS	199hm	JBMS11
18S clade	C3	C3	C2	C1
Geographical origin	China, Austria, New Zealand	Austria	Arctic	Austria
Habitat origin	freshwater	freshwater	freshwater	soil
Mode of nutrition	heterotroph	mixotroph	heterotroph	heterotroph
Media	IB + 3g/l nutrient broth, soytone & yeast extract	IB + 3g/l nutrient broth, soytone & yeast extract	IB + bacteria (Listonella pelagia PG5)	IB + bacteria (Listonella pelagia PG5)
Temperature	15°C	15°C	15°C	15°C

Light:dark-cycle	16 : 8	16 : 8	16 : 8	16 : 8
Illumination	75 - 100 µE	75 - 100 µE	75 - 100 µE	75 - 100 µE

Sequencing, assembly and annotation

The transcriptome sequences were generated in an overarching study of 18 chrysophyte strains, published in Beisser et al. (2017); see methodological details therein. The raw and assembled sequences are available at the European Nucleotide Archive (ENA) under accession number PRJEB13662. Construction of cDNA-libraries, poly-A selection and paired-end sequencing on the Illumina HiSeq2000 platform were performed by a commercial service (Eurofins MWG, Ebersberg, Germany).

Base quality of raw sequence reads was checked using the FastQC software (v0.10.1; Andrews 2015) and preprocessed and trimmed by Cutadapt (v1.3; Martin 2011). The de novo assemblies were carried out using Trinity (release 2013-11-10; Grabherr et al. 2011) with default parameters. Gene identification and functional annotation to genes and pathways were conducted using the similarity search tool RAPsearch2 (v2.15; Zhao et al. 2012) with E-value $< 10^{-5}$ against the KEGG database (release 2014-06-23; Kanehisa and Goto 2000).

Within the KEGG database genes are associated with orthologous groups and thus assigned to KEGG Orthology (KO) identifiers. In the following we use the term *gene* for the annotated orthologous gene of the considered transcript.

Comparative analysis of core, shared and exclusive genes

The assigned KO identifiers were used to determine the core transcriptome, constituted by the intersection of the KO identifier sets between all strains, shared transcripts between several species and exclusive transcripts of single strains using the R package Vennerable (3.0; Swinton 2009). The percentage of shared KO identifiers was calculated with regard to the smaller transcriptome. Furthermore, we compared the strains based on sequence similarity of identified orthologous proteins in the six-frame translated transcriptomes using the software proteinortho (v5.15; Lechner et al. 2011) with default parameters (except a decreased coverage of 30% between BLAST alignments).

Pathway analysis

Orthologous genes (KOs) assignable to KEGG pathways of the KEGG BRITE functional hierarchy level A, B and C (Kanehisa and Goto 2000) were considered. The six strains were compared based on the presence-absence of orthologous genes for these hierarchical levels and on the completeness of the main reaction steps within pathways (KEGG modules; Kanehisa et al. 2017).

Expression analysis

Gene expression counts of all six strains were compared. Therefore, transcript expression values were obtained with the tool eXpress (v1.3.1; Roberts and Pachter 2013), and gene (KO) expression was calculated by summing over the effective counts of the transcripts. Differences in sequencing depth were accounted for by Total Count normalization. Hierarchical cluster analysis (Ward) combined with principle coordinate analysis (PCoA) was performed with the R-package vegan (v2.3-3; Oksanen et al. 2016). For the PCoA the Sørensen distance (presence-absence) and the Bray-Curtis distance (KO expression) were used. All figures were created by using vegan or ggplot2 (Wickham 2009) in R.

Phylogenetic analysis

To create sequence alignments the pairwise alignments to KEGG gene sequences were used as a reference. All transcripts aligning to the same gene were truncated to the minimum overlapping region of at least 100bp and combined in a fasta file. The sequence alignments were constructed with the MAFFT software (7.164b; Katoh and Standley 2013) and manually checked and edited with GeneDoc (v2; Nicholas and Nicholas 1997). Only transcripts that were found in all six strains were considered for phylogenetic analyses which were performed in R with the package phangorn (v1.99-12; Schliep 2011). A model test was used to find the best substitution model, then maximum-likelihood and bootstrap analyses were calculated. All resulting phylogenetic trees were visualized with TreeGraph2 (Stöver and Müller 2010).

Data resources

European Nucleotide Archive (ENA) under accession number PRJEB13662

Results

General aspects - Transcriptome sizes and functional assignment

Sequencing of the chrysophyte strains resulted in 13.8 to 19.4 million read pairs, which were assembled into 24,783 to 58,003 transcripts (Table 2). The high quality of the assemblies is reflected by a high percentage of reads that were remappable to the assembled transcripts (92.8% on average) and an average N50 value, characterizing the contig length, of 1,190 nucleotides.

Estimates based on the number of Trinity components yielded a maximum of 20,441 to 38,883 genes. The transcriptomes of the three strains of *Poteriospumella lacustris* displayed a similar number of estimated genes (20,441 to 21,629 Trinity components) and similar functional annotations despite diverging sequencing depths. This finding and the completeness of several KEGG modules (main reaction steps in KEGG pathways), e.g.

glycolysis, citrate cycle, oxidative phosphorylation and nucleotide biosynthesis, denoted a sufficient sequencing depth in all samples.

Table 2.

Overview statistics of *Poteriospumella lacustris* (JBC07, JBM10, JBNZ41), *Poterioochromonas malhamensis* (DS), *Spumella vulgaris* (199hm) and *Pedospumella encystans* (JBMS11) including transcriptome size, assembly quality and annotation.

	***P. lacustris* (JBC07)**	***P. lacustris* (JBM10)**	***P. lacustris* (JBNZ41)**	***P. malhamensis* (DS)**	***S. vulgaris* (199hm)**	***P. encystans* (JBMS11)**
No. read pairs (million)	13.8	19.4	15.8	18.8	13.9	14.2
Reads after quality control (%)	92.5	91.7	92.4	94.3	92.4	93
No. transcripts	24,783	26,330	27,754	39,537	58,003	40,532
N50	1,155	1,246	1,275	1,405	983	1,077
Remapped reads (%)	97.3	97	91.6	95.1	89.5	86.1
Estimated no. of protein-coding genes	20,441	20,515	21,629	30,189	38,883	28,497
No. KEGG hits (E-value < 10^{-5})	6,619	6,784	7,025	9,556	10,711	9,893
No. unique KEGG orthologs	2,248	2,265	2,265	2,620	2,694	2,652
No. KEGG orthologs assignable to pathways	1,367	1,389	1,378	1,599	1,635	1,591
No. assigned KEGG pathways	243	246	244	247	259	257

Between 45.2% and 56.7% of all predicted genes (Trinity components) could be assigned to a gene in the KEGG database. The removal of redundant hits resulted in 2,249 to 2,695 unique KEGG orthologous genes of which 1,367 to 1,635 could be assigned to 243 to 259 pathways (Table 2). The identified genes showed a similar distribution to the five KEGG categories among all six strains (Fig. 1A). Approximately 38% of the genes matched the KEGG category “Metabolism”, 24% “Genetic Information Processing”, 10% “Environmental Information Processing”, 11% “Cellular Processes” and 18% “Organismal Systems”.

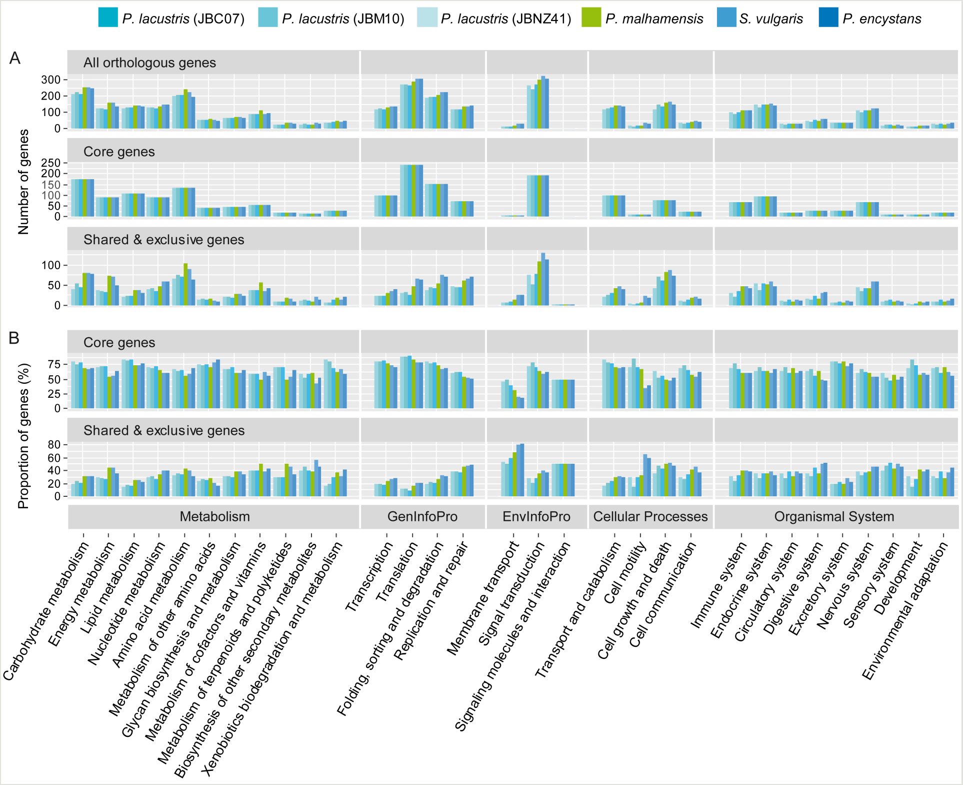

Figure 1.

Functional assignment of non-redundant KEGG orthologous genes of *Poteriospumella lacustris* (JBC07, JBM10, JBNZ41), *Poterioochromonas malhamensis* (DS), *Spumella vulgaris* (199hm) and *Pedospumella encystans* (JBMS11) to 31 pathway groups of the KEGG categories “Metabolism”, “Genetic Information Processing” (GenInfoPro), “Environmental Information Processing” (EnvInfoPro), “Cellular Processes” and “Organismal Systems”. **A** All functional hits per pathway group and strain were summarized for the whole transcriptomes, for the core transcriptome of all six strains and for the shared and exclusive genes. **B** The percentage proportion of genes per pathway group and strain was calculated for the core genes and the shared and exclusive genes.

Comparison of strains

A PCoA based on presence-absence data of KOs clearly separated all species (Fig. 2A), whereas the PCoA based on gene expression data placed *Spumella vulgaris* closer to *Pedospumella encystans* (Fig. 2B). All three strains of *Poteriospumella lacustris* clustered together supporting their affiliation with the same species. Within the *P. lacustris* cluster relatively small differences were identified compared to the other species. The relationship between the three strains differed slightly between the analysis of presence-absence data and of the relative read abundance (i.e. expression level): PCoA based on the expression level indicated a closer relationship between the strains JBC07 and JBM10, whereas PCoA based on presence-absence data did not indicate a closer intraspecific relation of these strains. Naturally, the presence-absence data varies less than the gene expression data and small differences in expression due to culturing conditions and cellular states cannot be totally excluded. We tried to oppose these effects by applying comparable conditions and extracting the RNA at the exponentially growth phase for all cultures.

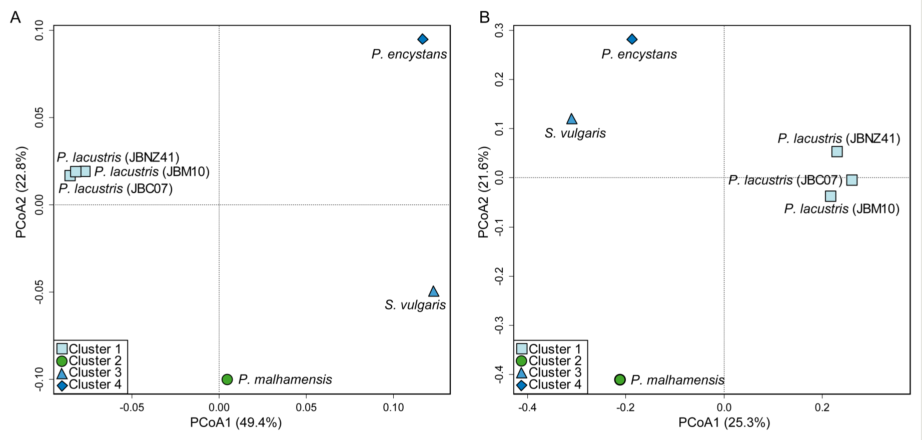

Figure 2.

PCoA based on all identified KEGG orthologous genes of *Poteriospumella lacustris* (JBC07, JBM10, JBNZ41), *Poterioochromonas malhamensis* (DS), *Spumella vulgaris* (199hm) and *Pedospumella encystans* (JBMS11). **A** PCoA based on the presence-absence of the respective genes. **B** PCoA based on the relative expression level of the respective genes.

Genes that were responsible for the grouping of strains were present in several but not in all species (around 800 genes) or exclusive to one species (181 to 307 genes). *Poteriospumella lacustris* had the lowest number and *Spumella vulgaris* the highest number of exclusive genes. Most of these partly shared and exclusive genes were affiliated with “Signal transduction”, “Amino acid metabolism”, “Cell growth and death”, “Carbohydrate metabolism”, “Energy metabolism” and “Replication & repair” (Fig. 1A).

The pairwise comparison showed that the percentage of shared KEGG orthologous genes (Fig. 3) between the transcriptomes of different species ranged from 78% (for *Pedospumella*, *Spumella* and *Poterioochromonas*) to 88% (for *Poterioochromonas* and *Poteriospumella*), reflecting the close relationship of the latter two taxa. The intraspecific variation was smaller, i.e. the transcriptomes of the three investigated strains of *P. lacustris* shared 92% to 93% of their genes. The shared transcriptome of this species comprised 2,004 genes.

Sequence comparisons based on all transcripts (instead of the annotated ones only) revealed that the majority of transcripts was exclusive to one strain. However, the strains of *Poteriospumella lacustris* shared approximately 50% of their transcripts with each other also indicating their close relation. The number of core transcripts was comparable to those of the annotated part.

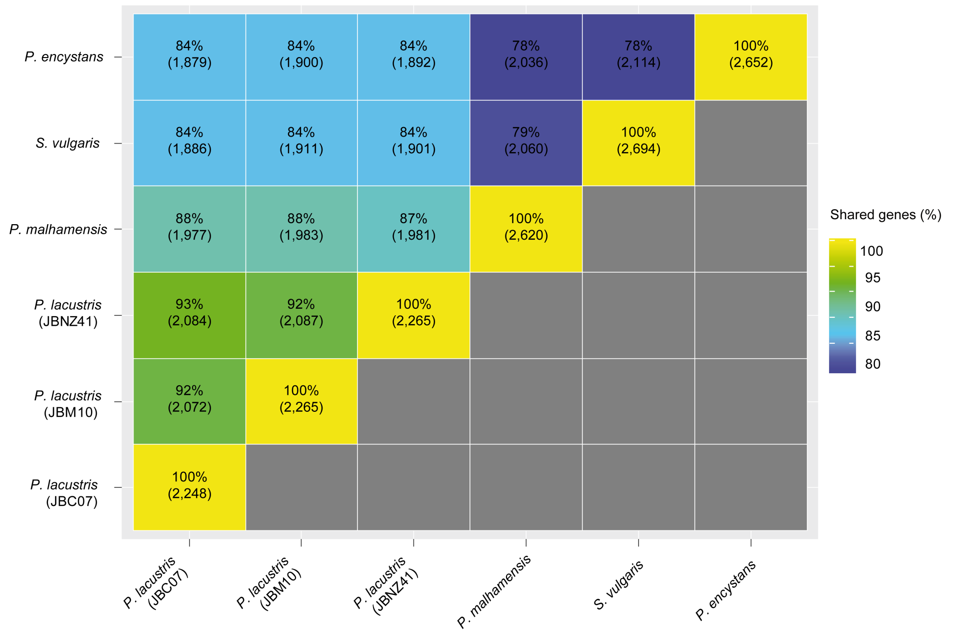

Figure 3.

Pairwise comparison of the proportion of shared KEGG orthologous genes of *Poteriospumella lacustris* (JBC07, JBM10, JBNZ41), *Poterioochromonas malhamensis* (DS), *Spumella vulgaris* (199hm) and *Pedospumella encystans* (JBMS11). The values in each matrix element are provided as the percentage of shared KEGG orthologous genes in relation to the number of KEGG orthologous genes in the smaller transcriptome. The absolute number of shared KEGG orthologous genes is given in brackets.

Core transcriptome

The core transcriptome derived from the annotated part of the transcriptomes of the investigated strains of Ochromonadales comprised 1,574 KEGG orthologous genes. Of these, we were able to assign 1,017 to one or more KEGG pathways (Fig. 1A). Thus 65% of all orthologous genes of the core transcriptome of the six strains could be assigned to KEGG pathways.

The core transcriptome comprised a large number of genes involved in basic cell metabolism: Roughly 38% of the genes of the core transcriptome were assigned to “Metabolism” and 27% of the genes were assigned to “Genetic Information Processing” (Fig. 1A). Particularly genes affiliated with ribosome and ribosome biogenesis and RNA transport within “Translation”, spliceosome within “Transcription” as well as RNA degradation, protein processing and proteasome within “Folding, sorting and degradation” were almost exclusively found in the core transcriptome. Similarly, pathways of the “Lipid metabolism”, especially steroid biosynthesis, fatty acid degradation and elongation as well as sphingolipid metabolism and glycerophospholipid metabolism were part of the core transcriptome. Apart from that, the following pathways were characteristic for the core transcriptome although several genes were found in the shared and exclusive parts of the transcriptomes: The “Carbohydrate metabolism” e.g. glycolysis and citrate cycle, the main reaction ways of the purine and pyrimidine metabolism, pathways of the “Energy metabolism” e.g. oxidative phosphorylation, various pathways in the “Amino acid metabolism” e.g. valine, leucine & isoleucine degradation, the main reaction ways of DNA replication, mismatch repair and nucleotide excision repair as well as various pathways of “Signal transduction” (Fig. 1B). In contrast, genes participating in “Cell motility” as well as “Membrane transport” were scarce in the core transcriptome and more characteristic for the shared and exclusive parts of transcriptomes (Fig. 1B).

Genes of the core transcriptome that were highly expressed in all strains were affiliated with the KEGG category “Genetic Information Processing” coding for various ribosomal proteins of the large and small subunit of ribosomes and the elongation factor affiliated with RNA transport. But also genes like ubiquitin c, heat shock proteins, calmodulin and solute carrier proteins affiliated with various signaling pathways as well as the F-type H^+-transporting ATPase affiliated with the oxidative phosphorylation were highly expressed.

Phylogenetic inference

Phylogenetic analyses were performed for 166 orthologous genes (Suppl. material 1) of the core transcriptome. The alignment length was between 120 and 1,431 nucleotides. We identified four main topologies to which 144 genes could be assigned, in the following denoted as topology A-D with decreasing frequency. They all had in common that the strains of *Poteriospumella lacustris* always clustered together. The phylogenetic position of *Poterioochromonas malhamensis* differed depending on the analyzed gene, i.e. the topologies mainly differed in the position of *Poterioochromonas malhamensis*:

In topology A (Fig. 4A) obtained for 76 investigated genes, *Poterioochromonas malhamensis* separates *Poteriospumella lacustris* from a clade formed by *Spumella vulgaris* and *Pedospumella encystans*. In topology B (Fig. 4B) obtained for 36 genes, *Poterioochromonas malhamensis* form a clade with *Spumella vulgaris* and is separated by *Pedospumella encystans* from *Poteriospumella lacustris*. In topology C obtained for 19 genes (Fig. 4C), *Poterioochromonas malhamensis* branch within *Poteriospumella lacustris* whereas *Spumella vulgaris* and *Pedospumella encystans* form another clade. In topology D obtained for 13 genes (Fig. 4D), *Poterioochromonas malhamensis* and *Pedospumella encystans* group together and *Spumella vulgaris* separates this clade from *Poteriospumella lacustris*. For several of these genes, a second transcript variant (with sequence variation in several positions) is present for *Poterioochromonas malhamensis* which always clustered with *Poteriospumella lacustris*, i.e. for 26 genes of topology A, for 14 genes of topology B and for one gene of topology D. In other strains we also identified transcript variants for some genes. Sequence deviation of these variants was small and therefore did not affect tree topology. Only for a very few genes (which were always affiliated with very different tree topologies not depicted in Fig. 4) variants deviated more strongly.

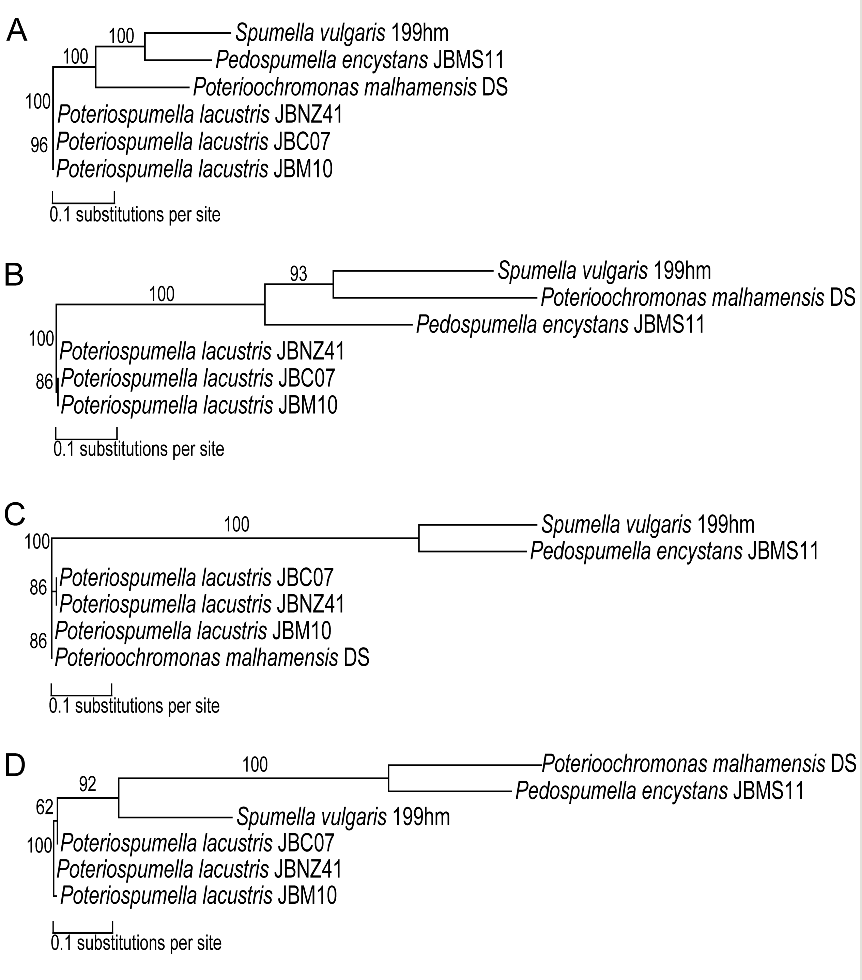
A
100
100
100
96
Spumella vulgaris 199hm
Pedospumella encystans JBMS11
Poterioochromonas malhamensis DS
Poteriospumella lacustris JBNZ41
Poteriospumella lacustris JBC07
Poteriospumella lacustris JBM10
0.1 substitutions per site
B
93
100
100
86
Spumella vulgaris 199hm
Poterioochromonas malhamensis DS
Pedospumella encystans JBMS11
Poteriospumella lacustris JBNZ41
Poteriospumella lacustris JBC07
Poteriospumella lacustris JBM10
0.1 substitutions per site
C
100
100
86
86
Spumella vulgaris 199hm
Pedospumella encystans JBMS11
Poteriospumella lacustris JBC07
Poteriospumella lacustris JBNZ41
Poteriospumella lacustris JBM10
Poterioochromonas malhamensis DS
0.1 substitutions per site
D
100
92
62
100
Poterioochromonas malhamensis DS
Pedospumella encystans JBMS11
Spumella vulgaris 199hm
Poteriospumella lacustris JBC07
Poteriospumella lacustris JBNZ41
Poteriospumella lacustris JBM10
0.1 substitutions per site

Figure 4.

Main topologies (unrooted phylogenetic trees; values at the nodes indicate statistical support >50% estimated by maximum likelihood method with 1,000 replicates) of 166 investigated KEGG orthologous genes of the core transcriptome of *Poteriospumella lacustris* (JBC07, JBM10, JBNZ41), *Poterioochromonas malhamensis* (DS), *Spumella vulgaris* (199hm) and *Pedospumella encystans* (JBMS11). For each topology one gene is exemplarily illustrated. **A** most frequent tree topology (obtained for 76 genes), **B** second most frequent tree topology (obtained for 36 genes), **C** third most frequent tree topology (obtained for 19 genes), and **D** fourth most frequent tree topology (obtained for 13 genes).

The phylogenetic relation between the three strains of *Poteriospumella lacustris* was resolved by approximately 58% of the investigated genes; approximately 53% of these indicated the closest relationship between the strains JBC07 and JBNZ41. Further, approximately 36% of these genes indicated a close relationship between strain JBM10 and *Poterioochromonas malhamensis*.

The affiliation to a distinct pathway is known for 117 of the genes included in the phylogenetic analysis (Fig. 5). These genes belonged mainly to “Metabolism” and “Genetic Information Processing”, whereby especially “Translation” is represented by many genes that followed the topology A. Genes that did not fit to any of the previously described topologies are mainly assigned to signaling pathways of “Environmental Information Processing”.

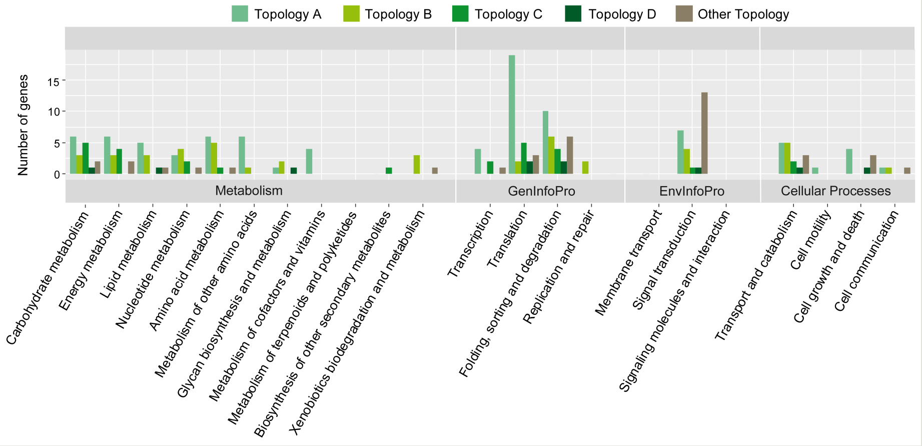

Figure 5.

Functional assignment of core genes used for phylogeny analyses of *Poteriospumella lacustris* (JBC07, JBM10, JBNZ41), *Poterioochromonas malhamensis* (DS), *Spumella vulgaris* (199hm) and *Pedospumella encystans* (JBMS11). The KEGG category “Organismal Systems” was not illustrated as no core genes were affiliated to this category.

Discussion

The chrysophyte core transcriptome - the essential minority

The comparative transcriptomic approach allowed insights into the genetic repertoire transcribed by Ochromonadales showing that the transcriptomes differed considerably between strains. Between 20,441 and 38,883 estimated protein coding genes were identified. This is in the lower and mid-range of previously reported values for gene estimates based on transcriptomes (Koid et al. 2014: 26,000 - 46,000, prymnesiophytes (Haptophyta); Di Dato et al. 2015: 18,000 - 20,000, diatoms (Stramenopiles); Liu et al. 2016: 24,000 - 27,000, chrysophytes (Stramenopiles); Herraiz et al. 2016 : 75,000, *Solanum* sp. (Embryophyta); Lin et al. 2014: 36,000 - 40,000, Spalacidae (Eumetazoa)) whereby genome based studies of related lineages e.g. diatoms (Bowler et al. 2008; 10,402 genes *P. tricornutum*, 11,776 genes *T. pseudonana*) or eustigmatophytes (Radakovits et al. 2012; 8,892 genes) indicate that estimates for the predicted genes are possibly overestimated in transcriptome-based studies. In our study 1,574 of all estimated genes were shared by all strains, i.e. represent the core transcriptome, whereas the majority of genes was specific to a single or only a few strains. This seems low but is similar in size to core transcriptomes reported for prymnesiophytes (Koid et al. 2014: 1,433 core genes) and diatoms (Di Dato et al. 2015: 1,370 core genes); all these estimates are based on functionally assignable genes. Sequence based methods for the identification of core genes are independent from the necessity of functional assignment and resulted in previous studies in approximately 3,000 core genes (Liu et al. 2016, Koid et al. 2014). In our study we also tested such a sequence based method but in contrast to the results of Liu et al. (2016) and Koid et al. (2014) we could not identify a higher number of core genes. Genome based studies of related lineages also range from 1,666 to 3,063 core genes (Bowler et al. 2008, Radakovits et al. 2012). These studies focused on a broader taxonomic group, so that the number of core genes within one lineage is presumably higher. A lower number of core genes detectable in transcriptomes compared to genomes might be due to the fact that not all genes are expressed (Dong and Chen 2013).

The majority of genes could not be affiliated with a distinct function. Koid et al. (2014) already pointed out that genes which are not part of the basal metabolism or more specific to distinct taxonomic groups often cannot be identified with respect to their function. Further, Fuhrman (2009) assumed that amongst the non-core genes with unknown function are those genes that are responsible for niche adaption.

Accordingly, roughly 65% of the identified orthologous genes are part of the core transcriptome in our study. A significant fraction of these core genes, was affiliated with the primary metabolism, reflecting the general importance of these metabolic pathways irrespective of phylogeny, nutritional mode and origin. They were affiliated with metabolic pathways including translation and ribosomal biogenesis, transcription and protein processing as well as with pathways affiliated with “Carbohydrate metabolism”, “Lipid metabolism”, “Nucleotide metabolism”, “Amino acid metabolism” and “Signal transduction”. This corresponds well with transcriptomic studies of other taxa which revealed a similar number and affiliation of genes of the core transcriptome with metabolic pathways (prymnesiophytes 1,433 core genes: Koid et al. 2014; diatoms 1,370 core genes: Di Dato et al. 2015). This proves the general importance of these pathways for taxa affiliated with various groups of Eukaryota. Further, a core set of around 1,400 genes can thus be considered to reflect the basic “active” genetic repertoire of various protist lineages. This is in accordance with estimates of Koonin (2003) reporting that the smallest genome of free-living organisms presumably comprise around 2,000 genes for eukaryotes.

Gene function interferes with phylogenetic signals in tree topologies

We calculated alignment-based phylogenetic trees of the six investigated strains for 166 core genes. The analyses resulted in four different tree topologies. The most frequent and the third frequent topology confirmed the close relation of *Poterioochromonas malhamensis* and *Poteriospumella lacustris* inferred from SSU rRNA gene sequences (both species are members of the C3-clade: Boenigk et al. 2005, Grossmann et al. 2016). This result may have been expected for genes coding for ribosomal proteins as these genes have slow evolutionary rates (e.g. Drummond et al. 2005). But several genes that follow this phylogenetic pattern were affiliated with functional genes within the KEGG category metabolism. This indicates a similar evolutionary pressure (and potentially extend of conservation) of these genes compared to those coding for ribosomal proteins. It is striking that phylogenies deviating from the main topologies often have several transcript variants and are particularly frequent in genes coding for proteins which interact with the environment. It seems likely that such proteins respond more strongly to environmental selection pressures and therefore deviate from the pattern observed for ribosomal genes and genes affiliated with the primary metabolism.

Furthermore, our analyses helped to resolve the relationship between closely related strains affiliated with one species. More than 50% of the investigated protein-coding genes showed sequence variations between the three strains JBC07, JBM10 and JBNZ41, which all belong to the species *Poteriospumella lacustris*. Earlier studies of these strains based on a multigene phylogeny of the protein-coding genes alpha-tubulin, beta-tubulin and actin (Stoeck et al. 2008), indicated a close relationship between the Chinese strain JBC07 and the New Zealand strain JBNZ41. The same result was obtained with our transcriptome-wide analysis of multiple gene tree-topologies. This relationship seems sensible given the geographic origins of the different strains. In contrast, an alignment-free approach (Beisser et al. 2017) and a multi gene phylogeny based on protein-coding genes and rDNA fragments (Stoeck et al. 2008) suggested a closer relationship between the strains from China (JBC07) and Austria (JBM10). This latter grouping of strains was reflected by PCoA of gene expression levels in our study. It has been shown by Drummond et al. (2005) that highly expressed genes evolve more slowly regardless of their function but rather due to the cost of protein misfolding. Whole transcriptome-based phylogenies include more genes with a low expression level, in contrast to single gene phylogenies that are mostly base on highly expressed but more conserved genes and therefore might display a diverging phylogeny. Genotype-environment interactions that affect fitness may thus result in diverging phylogenies for genes with higher evolutionary rates.

Conclusions

The present study reveals the interplay of functionality and phylogeny of the core transcriptome of Ochromonadales. We could demonstrate that the core transcriptome of Ochromonadales with its 1,574 genes represents only a small proportion of the transcriptomes but it comprises the genes affiliated with the primary metabolism. We assume that roughly 1,400 genes represent the basic “active” genetic repertoire of various protist lineages. Furthermore, we performed phylogenetic analyses of 166 protein-coding core genes. Most of the investigated genes coding for ribosomal genes or metabolism confirmed the close relation of *Poterioochromonas malhamensis* and *Poteriospumella lacustris* known from SSU rRNA gene phylogenies. Genes interacting with the environment largely show diverging phylogenetic patterns presumably due to a stronger impact of ecological selection pressures. Furthermore, we demonstrated the strength of comparative transcriptomics for the analysis of intraspecific and interspecific variation. Both, orthologous gene content analysis (PCoA) and phylogenetic analyses for several genes lead to congruent results of the relationship of Ochromonadales supporting the robustness of our results.

Acknowledgements

We thank the projects DFG Projekt BO 3245/17 and BO 3245/14 for financial support. Further, we thank Susann Chamrad for technical assistance.

Funding program

DFG Projekt BO 3245/17 and BO 3245/14.

Conflicts of interest

The authors declaire no conflict of interests.

References

- Andersen Ra, Van de Peer Y, Potter D, Sexton JP, Kawachi M, LaJeunesse T (1999) Phylogenetic analysis of the SSU rRNA from members of the Chrysophyceae . Protist 150 (March): 71-84. https://doi.org/10.1016/S1434-4610(99)70010-6
- Andrews S (2015) FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
- Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou SG, Allen AE, Apt KE, Bechner M, Brzezinski MA, Chaal BK, Chiovitti A, Davis AK, Demarest MS, Detter JC, Glavina T, Goodstein D, Hadi MZ, Hellsten U, Hildebrand M, Jenkins BD, Jurka J, Kapitonov VV, Kroger N, Lau WWY, Lane TW, Larimer FW, Lippmeier JC, Lucas S, Medina M, Montsant A, Obornik M, Parker MS, Palenik B, Pazour GJ, Richardson PM, Rynearson TA, Saito MA, Schwartz DC, Thamatrakoln K, Valentin K, Vardi A, Wilkerson FP, Rokhsar DS (2004) The genome of the diatom Thalassiosira pseudonana: Ecology, evolution, and metabolism. Science 306 (5693): 79-86. https://doi.org/10.1126/science.1101156
- Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290 (5493): 972-977. https://doi.org/10.1126/science.290.5493.972
- Beisser D, Graupner N, Bock C, Wodniok S, Grossmann L, Vos M, Sures B, Rahmann S, Boenigk J (2017) Comprehensive transcriptome analysis provides new insights into nutritional strategies and phylogenetic relationships of chrysophytes. PeerJ 5 https://doi.org/10.7717/peerj.2832
- Boenigk J, Pfandl K, Stadler P, Chatzinotas A (2005) High diversity of the 'Spumella-like' flagellates: An investigation based on the SSU rRNA gene sequences of isolates from

habitats located in six different geographic regions. Environmental Microbiology 7 (5): 685-697. https://doi.org/10.1111/j.1462-2920.2005.00743.x

- Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, Rayko E, Salamov A, Vandepoele K, Beszteri B, Gruber A, Heijde M, Katinka M, Mock T, Valentin K, Verret F, Berges JA, Brownlee C, Cadoret JP, Chiovitti A, Choi CJ, Coesel S, De Martino A, Detter JC, Durkin C, Falciatore A, Fournet J, Haruta M, Huysman MJJ, Jenkins BD, Jiroutova K, Jorgensen RE, Joubert Y, Kaplan A, Kroger N, Kroth PG, La Roche J, Lindquist E, Lommer M, Martin-Jezequel V, Lopez PJ, Lucas S, Mangogna M, McGinnis K, Medlin LK, Montsant A, Oudot-Le Secq MP, Napoli C, Obornik M, Parker MS, Petit JL, Porcel BM, Poulsen N, Robison M, Rychlewski L, Rynearson TA, Schmutz J, Shapiro H, Siaut M, Stanley M, Sussman MR, Taylor AR, Vardi A, von Dassow P, Vyverman W, Willis A, Wyrwicz LS, Rokhsar DS, Weissenbach J, Armbrust EV, Green BR, Van De Peer Y, Grigoriev IV (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nature 456 (7219): 239-244. https://doi.org/10.1038/nature07410
- Caron D, Alexander H, Allen A, Archibald J, Armbrust EV, Bachy C, Bell C, Bharti A, Dyhrman S, Guida S, Heidelberg K, Kaye J, Metzner J, Smith S, Worden A (2016) Probing the evolution, ecology and physiology of marine protists using transcriptomics. Nature Reviews Microbiology 15 (1): 6-20. https://doi.org/10.1038/nrmicro.2016.160
- Cox MM, Doudna JA, O'Donnell M (2012) Molecular Biology: Principles and Practice. W. H. Freeman and Company, New York..
- del Campo J, Massana R (2011) Emerging Diversity within Chrysophytes, Choanoflagellates and Bicosoecids Based on Molecular Surveys. Protist 162 (3): 435-448. https://doi.org/10.1016/j.protis.2010.10.003
- Deng H, Zhang GQ, Lin M, Wang Y, Liu ZJ (2015) Mining from transcriptomes: 315 single-copy orthologous genes concatenated for the phylogenetic analyses of Orchidaceae . Ecology and Evolution 5 (17): 3800-3807. https://doi.org/10.1002/ece3.1642
- Di Dato V, Musacchia F, Petrosino G, Patil S, Montresor M, Sanges R, Ferrante MI (2015) Transcriptome sequencing of three Pseudo-nitzschia species reveals comparable gene sets and the presence of Nitric Oxide Synthase genes in diatoms. Scientific reports 5 (April): 12329-12329. https://doi.org/10.1038/srep12329
- Dong ZC, Chen Y (2013) Transcriptomics: Advances and approaches. Science China-Life Sciences 56 (10): 960-967. https://doi.org/10.1007/s11427-013-4557-2
- Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH (2005) Why highly expressed proteins evolve slowly. Proc Natl Acad Sci USA 102 (40): 14338-43. https://doi.org/10.1073/pnas.0504070102
- Fuhrman JA (2009) Microbial community structure and its functional implications. Nature 459 (7244): 193-9. https://doi.org/10.1038/nature08058
- Gibson D, Benders G, Andrews-Pfannkoch C, Denisova E, Baden-Tillson H, Zaveri J, Stockwell T, Brownley A, Thomas D, Algire M, Merryman C, Young L, Noskov V, Glass J, Venter JC, Hutchison C, Smith H (2008) Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome. Science 319 (5867): 1215-1220. https://doi.org/10.1126/science.1151721
- Gil R, Silva F, Peretó J, Pereto J (2004) Determination of the Core of a Minimal Bacterial Gene Set Determination of the Core of a Minimal Bacterial Gene Set. Microbiology and Molecular Biology Reviews 68 (3): 518-537. https://doi.org/10.1128/MMBR.68.3.518

- Grabherr M, Haas B, Yassour M, Levin J, Thompson D, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren B, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29 (7): 644-52. https://doi.org/10.1038/nbt.1883
- Grossmann L, Bock C, Schweikert M, Boenigk J (2016) Small but Manifold – Hidden Diversity in “Spumella-like Flagellates”. Journal of Eukaryotic Microbiology 63 (4): 419-439. https://doi.org/10.1111/jeu.12287
- Harris JK, Kelley ST, Spiegelman GB, Pace NR (2003) The genetic core of the universal ancestor. Genome research 13 (3): 407-412. https://doi.org/10.1101/gr.652803
- Herraiz F, Blanca J, Ziarsolo P, Gramazio P, Plazas M, Anderson G, Prohens J, Vilanova S (2016) The first de novo transcriptome of pepino (Solanum muricatum): assembly, comprehensive analysis and comparison with the closely related species S. caripense, potato and tomato. BMC Genomics 17 (1): . https://doi.org/10.1186/s12864-016-2656-8
- Hsiang T, Baillie D (2005) Comparison of the Yeast Proteome to Other Fungal Genomes to Find Core Fungal Genes. Journal of Molecular Evolution 60 (4): 475-483. https://doi.org/10.1007/s00239-004-0218-1
- Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28 (1): 27-30. https://doi.org/10.1093/nar/28.1.27
- Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 45 https://doi.org/10.1093/nar/gkw1092
- Katoh K, Standley D (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular biology and evolution 30 (4): 772-780. https://doi.org/10.1093/molbev/mst010
- Koid A, Liu Z, Terrado R, Jones A, Caron D, Heidelberg K (2014) Comparative transcriptome analysis of four prymnesiophyte algae. Plos One 9 (6): . https://doi.org/10.1371/journal.pone.0097801
- Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nature Reviews Microbiology 1 (2): 127-136. https://doi.org/10.1038/nrmicro751
- Lechner M, Findeiß S, Steiner L, Marz M, Stadler P, Prohaska S (2011) Proteinortho: Detection of (Co-)orthologs in large-scale analysis. BMC Bioinformatics 12 (1): 124-124. https://doi.org/10.1186/1471-2105-12-124
- Lin GH, Kun Wang ENFZJSSGXD, Zhao H (2014) Transcriptome sequencing and phylogenomic resolution within Spalacidae (Rodentia). BMC Genomics 15: 32-32. https://doi.org/10.1186/1471-2164-15-32
- Liu Z, Campbell V, Heidelberg K, Caron D (2016) Gene expression characterizes different nutritional strategies among three mixotrophic protists. Fems Microbiology Ecology 92 (7): 1-11. https://doi.org/10.1093/femsec/fiw106
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17 (1): 10-12. https://doi.org/10.14806/ej.17.1.200
- Nicholas K, Nicholas H (1997) GeneDoc: a tool for editing and annotating multiple sequence alignments. http://www.psc.edu/biomed/genedoc/.
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O’Hara R, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2016) vegan:

Community ecology package. R package version 2.4.-1. URL: https://CRAN.R-project.org/package=vegan

- Price DC, Bhattacharya D (2017) Robust Dinoflagellata phylogeny inferred from public transcriptome databases. J Phycol 53 (3): 725-729. https://doi.org/10.1111/jpy.12529
- Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC (2012) Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropis gaditana. Nat Commun 3 https://doi.org/10.1038/ncomms1688
- Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Meth 10 (1): 71-73. https://doi.org/10.1038/nmeth.2251
- Schliep KP (2011) phangorn: phylogenetic analysis in R. Bioinformatics 27 (4): 592-593. https://doi.org/10.1093/bioinformatics/btq706
- Scoble JM, Cavalier-Smith T (2014) Scale evolution in Paraphysomonadida (Chrysophyceae): Sequence phylogeny and revised taxonomy of Paraphysomonas, new genus Clathromonas, and 25 new species. European Journal of Protistology 50 (5): 551-592. https://doi.org/10.1016/j.ejop.2014.08.001
- Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Meth 9 (8): 811-814. https://doi.org/10.1038/nmeth.2066
- Škaloud P, Kristiansen J, Škaloudová M (2013) Developments in the taxonomy of silica-scaled chrysophytes - from morphological and ultrastructural to molecular approaches. Nordic Journal of Botany 31 (4): 385-402. https://doi.org/10.1111/j.1756-1051.2013.00119.x
- Stoeck T, Jost S, Boenigk J (2008) Multigene phylogenies of clonal Spumella-like strains, a cryptic heterotrophic nanoflagellate, isolated from different geographical regions. International Journal of Systematic and Evolutionary Microbiology 58 (3): 716-724. https://doi.org/10.1099/ijs.0.65310-0
- Stöver B, Müller K (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11 https://doi.org/10.1186/1471-2105-11-7
- Sun J, Wang L, Wu S, Wang X, Xiao J, Chi S, Liu C, Ren L, Zhao Y, Liu T, Yu J (2014) Transcriptome-wide evolutionary analysis on essential brown algae (Phaeophyceae) in China. Acta Oceanologica Sinica 33 (2): 13-19. https://doi.org/10.1007/s13131-014-0436-3
- Swinton J (2009) Venn diagrams in R with the Vennerable package. https://rdrr.io/rforge/Vennerable/f/inst/doc/Venn.pdf.
- Wang G, Sun J, Liu G, Wang L, Yu J, Liu T, Chi S, Liu C, Guo H, Wang X, Wu S (2014) Comparative analysis on transcriptome sequencings of six Sargassum species in China. Acta Oceanologica Sinica 33 (2): 37-44. https://doi.org/10.1007/s13131-014-0439-0
- Wickett N, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker M, Burleigh JG, Gitzendanner M, Ruhfel B, Wafula E, Der J, Graham S, Mathews S, Melkonian M, Soltis D, Soltis P, Miles N, Rothfels C, Pokorny L, Shaw AJ, DeGironimo L, Stevenson D, Surek B, Villarreal JC, Roure B, Philippe H, dePamphilis C, Chen T, Deyholos M, Baucom R, Kutchan T, Augustin M, Wang J, Zhang Y, Tian Z, Yan Z, Wu X, Sun X, Wong GK, Leebens-Mack J (2014) Phylotranscriptomic analysis of the origin and early diversification of land plants. Proceedings of the National Academy

of Sciences of the United States of America 111 (45): 4859-68. https://doi.org/10.1073/pnas.1323926111

- Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York.
- Wodniok S, Brinkmann H, Glöckner G, Heidel A, Philippe H, Melkonian M, Becker B (2011) Origin of land plants: Do conjugating green algae hold the key? BMC Evolutionary Biology 11 (1): 104-104. https://doi.org/10.1186/1471-2148-11-104
- Yang X, Li Y, Zang J, Li Y, Bie P, Lu Y, Wu Q (2016) Analysis of pan-genome to identify the core genes and essential genes of Brucella spp. Molecular Genetics and Genomics 291 (2): 905-912. https://doi.org/10.1007/s00438-015-1154-z
- Zhao Y, Tang H, Ye Y (2012) RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. BIOINFORMATICS APPLICATIONS NOTE 28 (1): 125-126. https://doi.org/10.1093/bioinformatics/btr595

Supplementary material

Suppl. material 1: Functional and phylogenetic analysis of the core transcriptome of Ochromonadales doi

Authors: Nadine Graupner, Jens Boenigk, Christina Bock, Manfred Jensen, Sabina Marks, Sven Rahmann, Daniela Beisser

Data type: Table of genes used for phylogeny

Brief description: KEGG orthologous genes of the core trancriptome of the herein investigated Ochromonadales (*Poteriospumella lacustris* strains JBC07, JBM10, JBNZ41; *Poterioochromonas malhamensis* DS; *Spumella vulgaris* 199hm; *Pedospumella encystans* JBMS11) used for phylogenetic analyses.

Filename: Suppl1_Genes_for_phylogenetic_analyses_final.pdf - Download file (37.42 kb)