



DATA NOTE

# Draft genome assembly and transcriptome sequencing of the golden algae *Hydrurus foetidus* (Chrysophyceae) [version 1; peer review: 1 approved, 1 approved with reservations]

Jon Bråte <sup>1</sup>, Janina Fuss<sup>2</sup>, Kjetill S. Jakobsen<sup>3</sup>, Dag Klaveness<sup>4</sup>

<sup>1</sup>Section for Genetics and Evolutionary Biology (EVOGENE), Department of Biosciences, University of Oslo, Oslo, 0316, Norway

<sup>2</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, 24118, Germany

<sup>3</sup>Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, 0316, Norway

<sup>4</sup>Section for Aquatic Biology and Toxicology (AQUA), Department of Biosciences, University of Oslo, Oslo, 0316, Norway

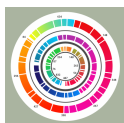
**V1** **First published:** 08 Apr 2019, 8:401 (<https://doi.org/10.12688/f1000research.16734.1>)  
**Second version:** 30 Aug 2019, 8:401 (<https://doi.org/10.12688/f1000research.16734.2>)  
**Latest published:** 07 Oct 2019, 8:401 (<https://doi.org/10.12688/f1000research.16734.3>)

**Abstract**

*Hydrurus foetidus* is a freshwater alga belonging to the phylum Heterokonta. It thrives in cold rivers in polar and high alpine regions. It has several morphological traits reminiscent of single-celled eukaryotes, but can also form macroscopic thalli. Despite its ability to produce polyunsaturated fatty acids, its life under cold conditions and its variable morphology, very little is known about its genome and transcriptome. Here, we present an extensive set of next-generation sequencing data, including genomic short reads from Illumina sequencing and long reads from Nanopore sequencing, as well as full length cDNAs from PacBio IsoSeq sequencing and a small RNA dataset (smaller than 200 bp) sequenced with Illumina. We combined this data with, to our knowledge, the first draft genome assembly of a chrysophyte algae. The assembly consists of 5069 contigs to a total assembly size of 171 Mb and a 77% BUSCO completeness. The new data generated here may contribute to a better understanding of the evolution and ecological roles of chrysophyte algae, as well as to resolve the branching patterns within the Heterokonta.

**Keywords**

*Hydrurus foetidus*, Chrysophyceae, golden algae, genome, transcriptome, Nanopore, PacBio



This article is included in the **Draft Genomes** collection.

**Open Peer Review**

**Reviewer Status** ✓ ✓

	Invited Reviewers	
	1	2
<b>REVISÉD</b> version 3 published 07 Oct 2019	✓ report	
<b>REVISÉD</b> version 2 published 30 Aug 2019	↑	
version 1 published 08 Apr 2019	? report	✓ report

- Daniela Beisser** <sup>id</sup>, University of Duisburg-Essen, Essen, Germany  
**Stephan Majda**, University of Duisburg-Essen, Essen, Germany
- Blake T. Hovde** <sup>id</sup>, Los Alamos National Laboratory, Los Alamos, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Jon Bråte ([jon.brate@ibv.uio.no](mailto:jon.brate@ibv.uio.no))

**Author roles:** **Bråte J:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Fuss J:** Data Curation, Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Jakobsen KS:** Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing; **Klaveness D:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** JB was funded by the Research Council of Norway (grant numbers: 213703 and 240284).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2019 Bråte J *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Bråte J, Fuss J, Jakobsen KS and Klaveness D. **Draft genome assembly and transcriptome sequencing of the golden algae *Hydrurus foetidus* (Chrysophyceae) [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2019, 8:401 (<https://doi.org/10.12688/f1000research.16734.1>)

**First published:** 08 Apr 2019, 8:401 (<https://doi.org/10.12688/f1000research.16734.1>)

## Introduction

Here, we present extensive genome sequencing data, including a multiapproach assembly, as well as transcriptome data of mRNAs and small RNAs of the golden algae *Hydrurus foetidus* (Villars) Trevisan.

There has been considerable interest in the golden algae for many reasons: they are ecologically diverse, important as primary producers (phototrophs) in oligotrophic to dystrophic lakes (Kristiansen, 2005; Nicholls & Wujek, 2015), some are also mixotrophs, phagotrophs or osmotrophs (Kristiansen & Preisig, 2001; Pringsheim, 1963). The chrysophytes span a large range of feeding and nutrient uptake modes (Kristiansen, 2005) and therefore play a significant role in aquatic food webs. Chrysophytes also make up a significant fraction of sequence reads and novel operational taxonomic units in clone libraries from freshwater environmental samples (del Campo & Massana, 2011).

However, chrysophytes have also attracted considerable interest from an evolutionary point of view. They belong to the division (phylum) Heterokonta (Cavalier-Smith, 1986), an immensely diverse group of eukaryotes with many basal branches in the phylogeny still not resolved, despite numerous molecular phylogenetic studies, including multigene phylogenomics (e.g. Burki, 2014; Grossmann *et al.*, 2016; Riisberg *et al.*, 2009; Scoble & Cavalier-Smith, 2014). One reason for this is the presence of cryptic species and many groups with extremely similar morphology (Grossmann *et al.*, 2016). Another reason is the complex evolutionary history of the Heterokonta, including an elaborate plastid evolution (e.g. Kim *et al.*, 2019) and heterotrophic lineages which have lost the plastids altogether (Graupner *et al.*, 2018; Pringsheim, 1963). The lack of genomic or transcriptomic data from many taxa, and even whole orders, which limits the power of multigene phylogenies (Beisser *et al.*, 2017), is yet another motivation for genomic and transcriptomic studies. However, recently, there has been a significant addition of transcriptomic data for chrysophyte taxa (e.g. Beisser *et al.*, 2017; Graupner *et al.*, 2018; Keeling *et al.*, 2014; Kraus *et al.*, 2019; Lie *et al.*, 2017).

*Hydrurus foetidus* is not a typical representative of the golden algae. It is macroscopic and benthic (e.g. Klaveness *et al.*, 2011; Rostafinski, 1882: Tab II, Szklarczyk, 1953), whereas most chrysophytes are microscopic single cells or colonial plankton (Sandgren, 1988; Kristiansen, 2005). Furthermore, *Hydrurus* is native to polar, peri-glacial and alpine rivers in Norway and similar regions around the world (e.g. Klaveness, 2019; Rott *et al.*, 2006; Rott & Schneider, 2014) and can only live in cold waters (2–10 °C) (Bursa, 1934; Kann, 1978). Members of the *Hydrurus* clade may cause colored snow and ice, and may be found on permanent ice sheets (Klaveness *et al.*, 2011; Lutz *et al.*, 2018; Remias *et al.*, 2013).

*Hydrurus* has a number of peculiar morphological characteristics relevant for understanding chrysophyte and heterokont evolution. Although it is multicellular, the cells in the thalli are not physically connected, and under some growth conditions

the cells may slide away from each other in their wall-less polysaccharide tubes, or be released as single-celled swimmers (Klaveness *et al.*, 2011). Other characteristic features, which may be considered primitive for a thallose alga, are contractive vacuoles, often more than one in each cell (Fott, 1959; Klaveness, 2019).

We have assembled a draft genome of *Hydrurus foetidus* using a combination of short-read Illumina sequencing and long-read Nanopore sequencing. The assembly consists of 5069 contigs yielding a total size of 171 Mb and a 77% BUSCO completeness. In addition to the deep genomic sequencing, we have also sequenced full-length poly(A) transcripts using PacBio IsoSeq, as well as sequencing the expressed small RNAs. This extensive dataset will be important, not only for studies of heterokont and chrysophyte evolution but also for elucidating the genetic mechanisms behind cold water adaptation, like the production of polyunsaturated fatty acids (Klaveness, 2017) and the regulation of a complex multicellular lifestyle.

## Materials and methods

### Culturing of *H. foetidus*

The specimen of *Hydrurus foetidus* (Villars) Trevisan (strain G070301) used in this study was sampled from the river at the Finse Alpine Research Center (60°36' N, 07°30' E) in March 2007 and is currently kept in culture at University of Oslo. *H. foetidus* was isolated in an adapted Guillard & Lorenzen's WC (Wright's Chu) medium (Guillard & Lorenzen, 1972) as described by Klaveness & Lindstrøm (2011). To prepare for DNA isolation, the growth of large thalli was promoted by repeated transfer of individual thalli into fresh culture media. Large thalli (0.5–1.0 g wet frozen weight) were collected by removal from the culture medium and immediate transfer to -80°C and storage until further processing. The culture will be deposited in a special culture collection, at the Fraunhofer Culture Collection of Cryophilic Algae (CCCrho).

### Illumina sequencing of genomic DNA

Six individual thalli were used for the DNA isolation. DNA isolation was performed using the DNeasy Plant Mini Kit from Qiagen (Qiagen Inc., Valencia, CA, US). To ensure efficient lysis and homogenization of the external polysaccharide sheath, a few titanium beads were added to the frozen samples and the tubes were shaken using TissueLyser II machine (Qiagen Inc., Valencia, CA, US) for four minutes. After the addition of the lysis buffer and the RNase, tubes were placed in a thermomixer set at 65°C and 800 rpm with 20 second intervals for 30 mins. After adding Buffer AP2 to the lysate, the incubation was done on ice for 15 minutes to allow for better precipitation of the polysaccharides. Further, the extraction kit protocol was followed as is, until the second elution step. Here we reused the flow through from the previous step elution to avoid excessive dilution of the samples. Afterwards, the samples were de-salted and concentrated by ethanol precipitation and resuspension in 100 µl of Milli Q water. Finally, the samples were concentrated even further by pooling all the samples and freeze drying with a Leybold-Heraeus Lyovac GT2 (Leybold-Heraeus, Köln, Germany).

The isolated and freeze-dried genomic DNA was sent to the [Norwegian Sequencing Center](#) (NSC) at the University of Oslo for library preparation and sequencing. The library was made with 600-700 bp fragment size and sequenced on two lanes of Illumina HiSeq 2500 with 250 bp paired-end reads ([Table 1](#)).

### Nanopore sequencing of genomic DNA

Genomic DNA was isolated from two thalli as described above, except that tissue lysis was done using MagNA Lyser Green Beads (Roche, Penzberg, Germany) and shaken for 15 sec at 4 m/sec and the incubation at 65°C was done for 10 min. In addition, the supernatant (after adding buffer AP2) was run through QiaShredder columns to further homogenize the lysate. The DNA was eluted (twice, but re-using the elution buffer) in 50 µl AE elution buffer. To further clean and concentrate the samples, the samples were pooled and cleaned using the Zymo DNA Clean & Concentrate kit (Zymo Research, CA, US). The sample was double-eluted (as before) in 50 µl kit provided elution buffer.

DNA sequencing was done using the MinION (MIN-101B) sequencer, the R9.5 Flow Cell and following the SQK-LSK108 protocol (version GDE\_9002\_v108\_revT\_18Oct2016) (Oxford Nanopore, Oxford, UK). Approximately 1 µg of starting DNA was used and inspection of the DNA on a 0.7% agarose gel run at 30 volts from 18 hours showed that the majority of the DNA was between 20-30 kbp, but with a long tail of shorter fragments. The sequencing was run using the MinKNOW software (Oxford Nanopore, Oxford, UK; downloaded October 2017) on an iMac and stopped after 36 hours. Base-calling of the raw Nanopore sequence data was done using [Albacore](#) v.2.1.10 (Linux, Python 3.5 version) with default settings. The process was run on the Abel computing cluster at the University of Oslo ([Table 1](#)).

### PacBio transcriptome sequencing

Total RNA was isolated from one frozen thallus using Qiagen RNeasy Plant kit, including a QiaShredder column and lysis using MagNA lyser beads as described above, otherwise following the kit protocol. Isolated RNA was sent to NSC for library preparation and PacBio sequencing. Three size fractions (1-2 kbp, 2-3 kbp and 3-5 kbp) were prepared using the IsoSeq library preparation protocol and sequenced on RSII SMRT cells (Pacific Biosciences, CA, US) ([Table 1](#) and [Table 2](#)).

### Illumina small RNA sequencing

Small RNAs (below 200 bp) were isolated from a frozen thallus using the Sigma mirPremier kit (Sigma-Aldrich, MO, US) following the manufacturer's instructions, but including lysis with MagNA beads as described above. The sample was sent to the NSC for library preparation and sequencing. Sequencing library (up to approx. 40 nt fragment size) was prepared and sequenced with Illumina NextSeq 500 as single-end 75 bp reads ([Table 1](#)).

### Draft genome assembly

The basecalled Nanopore reads were processed with [Porechop](#) v0.2.2 using default parameters to remove sequencing adapters. Next, the reads were filtered with [Nanofilt](#) v2.0.0 ([De Coster et al., 2018](#)) to remove reads shorter than 500 bp and average quality below 9. The filtered reads were further error-corrected with [LoRDEC](#) v0.7 ([Salmela & Rivals, 2014](#)) using the Illumina reads. First the Illumina reads were quality assessed by removing sequencing adapters and bases with an average quality below 20 (average score across 4 bases), in addition to leading and trailing bases with a quality below 20. This was done using [Trimmomatic](#) v0.36 ([Bolger et al., 2014](#)). Then [lordec-correct](#) (options -k 21 -s 3) was run with the trimmed Illumina reads

**Table 1. Overview of datasets produced in this study.**

Dataset	Description	Accession
Hfoetidus_ACAGTG_L001_R1_001.fastq.gz Hfoetidus_ACAGTG_L001_R1_001.fastq.gz	Genomic DNA sequenced with Illumina HiSeq 2500.	ERR2882522
Hfoetidus_ACAGTG_L002_R1_001.fastq.gz Hfoetidus_ACAGTG_L002_R1_001.fastq.gz	Genomic DNA sequenced with Illumina HiSeq 2500.	ERR3188711
Hydrurus_nanopore_fastq_files.tar.gz	Basecalled Oxford Nanopore reads	ERR2887871
IsoSeq_1-2kb_polished_low_qv_consensus_isoforms.fastq.gz	mRNA sequenced with PacBio SMRT RSII.	ERR2882521
IsoSeq_1-2kb_polished_high_qv_consensus_isoforms.fastq.gz	mRNA sequenced with PacBio SMRT RSII.	ERR2869477
IsoSeq_2-3kb_polished_low_qv_consensus_isoforms.fastq.gz	mRNA sequenced with PacBio SMRT RSII.	ERR2869481
IsoSeq_2-3kb_polished_high_qv_consensus_isoforms.fastq.gz	mRNA sequenced with PacBio SMRT RSII.	ERR2869478
IsoSeq_3-6kb_polished_low_qv_consensus_isoforms.fastq.gz	mRNA sequenced with PacBio SMRT RSII.	ERR2869484
IsoSeq_3-6kb_polished_high_qv_consensus_isoforms.fastq.gz	mRNA sequenced with PacBio SMRT RSII.	ERR2869483
1-Hfo-miRNA_S6_R1_001.fastq.gz	Small RNA sequenced with Illumina NextSeq 500.	ERR2869485
pilon_round3.fasta.gz	Draft genome assembly	ERZ780628

to correct the filtered Nanopore reads. Then the corrected reads were run through *Canu* v1.6 (Koren *et al.*, 2017) for further correction (*canu* -correct with genome Size set to 300 m) and trimming (*canu* -trim) before assembly (*canu* -assemble). The assembly was done with two different corrected error rates, 0.144 and 0.146. The two assemblies were almost identical, but the results from using the corrected error rate of 0.144 were

used further because the total size was slightly larger and also had the largest contig. The *Canu* assembly was then polished using the trimmed Illumina reads (described above) by running three rounds of *Pilon* v1.22 (Walker *et al.*, 2014). The final genome assembly consisted of 5069 contigs with a total length of 171 183 409 nt. The N50 was 43,856 nt and the longest contig of 5,118,963 nt (Table 3).

**Table 2. Summary of the read numbers in the different file types of the IsoSeq data set.**

Library	Size fraction						
	< 1kb	1-2kb	2-3kb	3-4kb	4-5kb	5-6kb	> 6kb
1-2kb_high	7310	31953	17	0	0	0	0
1-2kb_low	903	5908	170	110	89	39	80
2-3kb_high	596	2703	37399	217	443	147	0
2-3kb_low	78	586	7749	116	301	134	215
3-6kb_high	8	13	552	28621	4603	20	388
3-6kb_low	0	6	268	8830	1535	74	418

1-2kb\_high - accession ERR2869477; 1-2kb\_low - accession ERR2882521; 2-3kb\_high - accession ERR2869478; 2-3kb\_low - accession ERR2869481; 3-6kb\_high - accession ERR2869483; 3-6kb\_low - accession ERR2869484.

**Table 3. Statistics of the *Hydrurus foetidus* draft genome assembly.**

Number of contigs > 1000 bp	Largest contig	Contig N50	Assembly size	Estimated genome size <sup>a</sup>	Complete and fragmented BUSCO orthologs <sup>b</sup>
5069	5118963 bp	43856 bp	171 Mb	299.9 Mb	77.2%

<sup>a</sup>The genome size estimation was based on k-mer frequencies on the Illumina data.

<sup>b</sup>BUSCO was run against the Eukaryota dataset.

## Data availability

All *Hydrurus foetidus* datasets produced in this study are available, study accession number PRJEB29405: <https://identifiers.org/ena.embl/PRJEB29405>.

## Author contributions

JB performed DNA and RNA isolation, Nanopore sequencing, genome assembly, prepared and submitted data and wrote the manuscript. JF monitored and performed culturing, isolated DNA and RNA and prepared for Illumina and PacBio sequencing, prepared and submitted data and commented on the manuscript. KSJ planned and monitored the entire project and

commented on the manuscript. DK maintained and developed the cultures and isolation protocol and wrote the manuscript.

## Grant information

JB was funded by the Research Council of Norway (grant numbers: 213703 and 240284).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## Acknowledgements

We thank the [Norwegian Sequencing Centre](#) for all sequence library preparations and sequencing services.

## References

Beisser D, Graupner N, Bock C, *et al.*: **Comprehensive transcriptome analysis provides new insights into nutritional strategies and phylogenetic relationships of chrysophytes.** *PeerJ*. 2017; 5: e2832. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)  
Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina**

**sequence data.** *Bioinformatics*. 2014; 30(15): 2114–2120. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Burki F: **The eukaryotic tree of life from a global phylogenomic perspective.** *Cold Spring Harbor Perspect Biol*. 2014; 6(5): a016147. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

- Bursa A: *Hydrurus foetidus*. Kirch. w Polskich Tatrach. –*Hydrurus foetidus* Kirch. in der Polnischen Tatra. I. Oekologie, Morphologie. II. Phenologie. *Bulletin International de l'Academie Polonaise des Sciences et des Lettres (C lasse des Sciences Mathematiques et Naturelles. Serie B: Sciences Naturelles (I))*, 1934; 69-84 + Pl. 1-2 + 113-31.
- Cavalier-Smith T: The kingdom Chromista: origin and systematics. In: *Progress in phylogenical research*. (Round, F.E. & Chapman, D.J. Eds), Bristol: Biopress Ltd. 1986; 4: 309–347.
- De Coster W, D'Hert S, Schultz DT, et al.: NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*. 2018; 34(15): 2666–2669. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- del Campo J, Massana R: Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist*. 2011; 162(3): 435–448. [PubMed Abstract](#) | [Publisher Full Text](#)
- Fott B: *Algenkunde*. VEB Gustav Fischer Verlag Jena. 1959; 482 S. [Reference Source](#)
- Graupner N, Jensen M, Bock C, et al.: Evolution of heterotrophy in chrysophytes as reflected by comparative transcriptomics. *FEMS Microbiol Ecol*. 2018; 94(4): fty039. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Grossmann L, Bock C, Schweikert M, et al.: Small but Manifold - Hidden Diversity in “*Spumella*-like Flagellates”. *J Eukaryote Microbiol*. 2016; 63(4): 419–439. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Guillard RRL, Lorenzen C: Yellow-green algae with chlorophyllide c. *J Phycol*. 1972; 8(1): 10–14. [Publisher Full Text](#)
- Kann E: Systematik und Ökologie der Algen der Österreichischer Bergbäche. *Arch Hydrobiol*. 1978; Suppl 53(4): 405–643.
- Keeling PJ, Burki F, Wilcox HM, et al.: The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol*. 2014; 12(6): e1001889. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kim JI, Shin H, Škaloud P, et al.: Comparative plastid genomics of Synurophyceae: inverted repeat dynamics and gene content variation. *BMC Evol Biol*. 2019; 19(1): 20. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Klaveness D: *Hydrurus foetidus* (Chrysophyceae)—an inland macroalga with potential. *J Appl Phycol*. 2017; 29(3): 1485–1491. [Publisher Full Text](#)
- Klaveness D: *Hydrurus foetidus* (Chrysophyceae) – an update and request for observations. *Algae*. 2019; 34(1): 1–5. [Publisher Full Text](#)
- Klaveness D, Bråte J, Patil V, et al.: The 18S and 28S rDNA identity and phylogeny of the common lotic chrysophyte *Hydrurus foetidus*. *Eur J Phycol*. 2011; 46(3): 282–291. [Publisher Full Text](#)
- Klaveness D, Lindström EA: *Hydrurus foetidus* (Chromista, Chrysophyceae): A large freshwater chromophyte alga in laboratory culture. *Phycol Res*. 2011; 59(2): 105–112. [Publisher Full Text](#)
- Koren S, Walenz BP, Berlin K, et al.: Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017; 27(5): 722–736. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kraus D, Chi J, Boenigk J, et al.: Putatively asexual chrysophytes have meiotic genes: evidence from transcriptomic data. *PeerJ*. 2019; 6: e5894. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kristiansen J: *Golden Algae. A Biology of Chrysophytes*. A.R.G. Gantner Verlag K.G. Rugell, Lichtenstein. 2005; 167. [Reference Source](#)
- Kristiansen J, Preisig HR: *Encyclopedia of Chrysophyte Genera. Bibliotheca Phycologica*. Bd. 110. J. Cramer, Berlin, Stuttgart. 2001; 260. [Reference Source](#)
- Lie AA, Liu Z, Terrado R, et al.: Effect of light and prey availability on gene expression of the mixotrophic chrysophyte, *Ochromonas* sp. *BMC Genomics*. 2017; 18(1): 163. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lutz S, McCutcheon J, McQuaid JB, et al.: The diversity of ice algal communities on the Greenland Ice Sheet as revealed by oligotyping. *Microb Genom*. 2018; 4(3): 1–10. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nicholls KH, Wujek DE: Chrysophyceae and Phaeothamniophyceae. In: Wehr, J.D., Sheath, R.G. & Kociolek, J. Patrick (Ch. 12) *Freshwater Algae of North America*. Ecology and Classification. Second Edition. Elsevier/Academic Press, San Diego. 2015; 537–586. [Publisher Full Text](#)
- Pringsheim EG: *Farblose Algen. Ein Beitrag zur Evolutionsforschung*. Gustav Fischer Verlag, Stuttgart. 1963; 471 S. [Reference Source](#)
- Remias D, Jost S, Boenigk J, et al.: *Hydrurus*-related golden algae (Chrysophyceae) cause yellow snow in polar summer snowfields. *Phycol Res*. 2013; 61(4): 277–285. [Publisher Full Text](#)
- Riisberg I, Orr RJ, Kluge R, et al.: Seven gene phylogeny of heterokonts. *Protist*. 2009; 160(2): 191–204. [PubMed Abstract](#) | [Publisher Full Text](#)
- Rostafinski J: *Hydrurus i jego pokrewienczo*. Monografia. (Rzecz czytana na posiedzeniu Wydziału matem.-przyrodn. Akad. Um. w Krakowie dnia 20 Czerwca 1881 r.). Osobne odbicie z Rozpraw Akad. umiej., Wydż. matem.- przyr., tom X. Z tablica II. 58-86 (1-29) + Résumé: Hydrurus und seine Vervandschaft. Eine Monographie. (Vortragen in der math.-natur. Classe der Akademie der Wissenschaften in Krakau am 30 Juli 1881) (31-34) 1882. [Reference Source](#)
- Rott E, Cantonati M, Füreder L, et al.: Benthic algae in high altitude streams of the Alps – a neglected component of the aquatic biota. *Hydrobiol*. 2006; 562(1): 195–216. [Publisher Full Text](#)
- Rott E, Schneider SC: A comparison of ecological optima of soft-bodied benthic algae in Norwegian and Austrian rivers and consequences for river monitoring in Europe. *Sci Total Environ*. 2014; 475: 180–186. [PubMed Abstract](#) | [Publisher Full Text](#)
- Sandgren CD: *The Ecology of Chrysophyte Flagellates: their Growth and Perennation Strategies as Freshwater Phytoplankton*. In: Sandgren, C.D. (ed.) *Growth and reproductive strategies of freshwater phytoplankton* (Ch. 2). Cambridge University Press, New York. 1988; 442. [Reference Source](#)
- Salmela L, Rivals E: LoRDEC: accurate and efficient long read error correction. *Bioinformatics*. 2014; 30(24): 3506–3514. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Scoble JM, Cavalier-Smith T: Scale evolution in Paraphysomonadida (Chrysophyceae): Sequence phylogeny and revised taxonomy of *Paraphysomonas*, new genus *Clathromonas*, and 25 new species. *Eur J Protistol*. 2014; 50(5): 551–592. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Szklarczyk C: *Observacje nad morfologią i biologią Hydrurus foetidus* Kirch. W Ojcowie. *Acta Soc Bot Poloniae*. 1953; 22(2): 397–410 + TABLICA 1. [Reference Source](#)
- Walker BJ, Abeel T, Shea T, et al.: Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014; 9(11): e112963. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status:



Version 1

Reviewer Report 12 August 2019

<https://doi.org/10.5256/f1000research.18291.r49312>

© 2019 Hovde B. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Blake T. Hovde** 

Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM, USA

Review of Draft Genomes: Draft genome assembly and transcriptome sequencing of the golden algae *Hydrurus foetidus* (Chrysophyceae):

This data note represents a significant contribution to the algal genomics research community in couple different ways. This project represents the first publicly available draft genome of a chrysophyte (golden algae) to date, making it a critical contribution to the study of this particular clade of algae and will be a much-utilized resource for other chrysophyte genome analyses as they emerge. Because this is the first representative of Chrysophyceae to be sequenced it will be a de facto resource, which leads to the next important contribution of the note – which is the importance of the modern sequencing technologies used in development of this genome and transcriptome assembly. The tools used are likely to be the gold-standard tools of *de novo* genome and transcriptome sequencing and assembly. The use of Pacific Biosciences “Isoseq” long read, full length transcriptome sequencing approach will allow this team to generate high quality gene annotations, a great advantage for a new reference such as this genome. Additionally, the use of long read technologies (either Nanopore or PacBio) will be used to generate higher quality genome assemblies with fewer, longer contigs and scaffolds.

While broad strokes are painted about the contribution of this draft genome to phylogenetic (or phylogenomic) analysis and the importance of the algae as primary producers it is unclear what additional analyses this team will be doing to utilize this genome assembly. Next steps may be to perform biosynthetic pathway analyses and phylogenetic analysis of the Chrysophyte class or higher level algal/protist analysis. It would be beneficial if the authors could point to some specific follow-ups.

I was unable to locate the genome assembly using accession number ERZ780628 as listed in Table 1. Though, I was able to find the assembly under NCBI accession number: GCA\_900617105.1

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and materials provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Algal genomics, genome editing, metabolic engineering, genome sequencing technology

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 29 Aug 2019

**Jon Bråte**, University of Oslo, Oslo, Norway

We thank the reviewer for the critical reading. We have updated the accession number to the genome assembly (now GCA\_900617105.1) and this should work both on EBI and NCBI.

**Competing Interests:** The authors declare no competing interests.

Reviewer Report 09 May 2019

<https://doi.org/10.5256/f1000research.18291.r47028>

© 2019 Beisser D et al. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Daniela Beisser**

Department of Biodiversity, University of Duisburg-Essen, Essen, Germany

**Stephan Majda**

Department of Biodiversity, University of Duisburg-Essen, Essen, Germany

Please find below the comments to the article “Draft genome assembly and transcriptome sequencing of the golden algae *Hydrurus foetidus* (Chrysophyceae)” by Jon Bråte, Janina Fuss, Kjetill S. Jakobsen and Dag Klaveness.

The authors present in the article draft genome data for the chrysophyte *Hydrurus foetidus*. *H. foetidus* was chosen because it is a special chrysophyte which is, in contrast to most other chrysophytes, macroscopic, but cells can slide apart under some growth conditions, benthic and thrives under cold conditions in polar and high alpine regions.

The genome was sequenced using Illumina and Nanopore sequencing, in addition mRNA reads sequenced with PacBio and small RNA reads sequenced with Illumina are provided. The genome was



assembled in a hybrid approach, resulting in a length of 171 Mb and 5,069 contigs with a BUSCO completeness of 77%.

DNA extraction, sequencing and assembly generation are described in detail, but some information is missing, which is listed below.

Overall, the work is technically sound and should be indexed after minor corrections.

#### Minor comments:

1. Abstract: "We combined this data with, to our knowledge, the first draft genome assembly of a chrysophyte algae". The data is not combined with the draft genome, maybe it should be "We combined this data to create, to our knowledge, the first draft genome assembly of a chrysophyte algae". Further, there are two draft genomes listed at JGI for *Ochromonas* and *Paraphysomonas* species ([https://genome.jgi.doe.gov/portal/OchCCMStandDraft\\_FD/OchCCMStandDraft\\_FD.info.html](https://genome.jgi.doe.gov/portal/OchCCMStandDraft_FD/OchCCMStandDraft_FD.info.html), [https://genome.jgi.doe.gov/portal/ParimpEvaluation\\_FD/ParimpEvaluation\\_FD.info.html](https://genome.jgi.doe.gov/portal/ParimpEvaluation_FD/ParimpEvaluation_FD.info.html)). They cannot be accessed without registration, but perhaps they should be cited.
2. Abstract: "...belonging to the phylum Heterokonta." Recent classification of Eukaryotes (Adl *et al* 2005, 2012, 2018)<sup>1,2,3</sup> place the Chrysophyceae into the Stramenopiles which themselves belong to the SAR supergroup. The current classification should be adapted.
3. Abstract: An assembly of 171 Mb was obtained, is this size expected? Is there other data available which suggests this genome size? On the other hand, the k-mer based genome size estimation suggests a much larger genome.
4. Introduction: "multiapproach assembly", should better be described as hybrid assembly.
5. Culturing of *H. foetidus*: The introduction mentions plastid reduction and heterotrophy in chrysophyte species. *H. foetidus* is called alga, so I assume it is phototrophic, but it would be good to state that somewhere explicitly. Otherwise one would wonder if the cultures were axenic.
6. Concerning the sequencing: Was there quality control performed by the sequencing center (RIN values etc.) before sequencing? Which library preparation protocols were used for the different sequencing data? Was the RNA rRNA-depleted before sequencing? What was the average sequence quality before and after filtering? An overview table of all the sequencing data including quality and number of reads/yield should be added.
7. Nanopore sequencing of genomic DNA: "Albacore v2.1.10 (Linux, Python 3.5 version)" Please provide a reference for Albacore. The information that it was run under Linux with Python 3.5 is probably not necessary.
8. Next line: Table 1 is not the correct reference for the sentence.
9. Table 2: It is not described in the text or caption what "\_high" and "\_low" means.
10. Table 3: The genome size estimation based on k-mers and BUSCO are not described in the methods section. Maybe the GC content could also be added to the table.

#### References

1. Adl SM, Simpson AG, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup O, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, Taylor MF: The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J Eukaryot Microbiol.* **52** (5): 399-451 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, Brown MW, Burki F, Dunthorn M, Hampel V, Heiss A, Hoppenrath M, Lara E, Le Gall L, Lynn DH, McManus H, Mitchell EA, Mozley-Stanridge SE, Parfrey LW, Pawlowski J, Rueckert S, Shadwick L, Shadwick L, Schoch CL, Smirnov A, Spiegel FW: The

revised classification of eukaryotes. *J Eukaryot Microbiol.* 2012; **59** (5): 429-93 [PubMed Abstract](#) | [Publisher Full Text](#)

3. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, Agatha S, Berney C, Brown MW, Burki F, Cárdenas P, Čepička I, Chistyakova L, Del Campo J, Dunthorn M, Edvardsen B, Eglit Y, Guillou L, Hampel V, Heiss AA, Hoppenrath M, James TY, Karnkowska A, Karpov S, Kim E, Kolisko M, Kudryavtsev A, Lahr DJG, Lara E, Le Gall L, Lynn DH, Mann DG, Massana R, Mitchell EAD, Morrow C, Park JS, Pawlowski JW, Powell MJ, Richter DJ, Rueckert S, Shadwick L, Shimano S, Spiegel FW, Torruella G, Youssef N, Zlatogursky V, Zhang Q: Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. *J Eukaryot Microbiol.* 2019; **66** (1): 4-119 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the rationale for creating the dataset(s) clearly described?**

Yes

**Are the protocols appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and materials provided to allow replication by others?**

Partly

**Are the datasets clearly presented in a useable and accessible format?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** High-throughput sequencing analysis

**We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.**

Author Response 29 Aug 2019

**Jon Bråte**, University of Oslo, Oslo, Norway

We thank the reviewer for the critical reading. Here are our responses to the minor comments:

1. We have rephrased the sentence and it no longer states that we present the first chrysoophyte draft genome assembly.
2. We have rephrased the first sentence of the abstract. And the first time Heterokonta now is mentioned in the introduction we have also added that Adl et al. uses Stramenopila and added the reference. However, we prefer the name Heterokonta as it has seniority to Stramenopila, and is still used by for instance Cavalier-Smith (e.g. Ruggiero et al. 2015; A Higher Level Classification of All Living Organisms. Plos One: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4418965/>).
3. The k-mer estimation was done on the Illumina data alone. While the assembly was done on the Nanopore reads alone (after error correcting with the Illumina reads). Also, k-mer based estimations of genome size can be inaccurate, especially since we know nothing about the ploidy level or the amount of genetic variation between chromosome copies. Therefore we don't think that this discrepancy is particularly large or surprising.

4. Corrected.

5. We have added a sentence explicitly mentioning that *H. foetidus* is photosynthetic to the methods section under the paragraph describing the culture procedure.

6. The PacBio transcriptome sequencing was performed by selecting for polydenylated transcripts (information now added to the manuscript), and for the smallRNA sequencing the totalRNA was size fractionated to remove ribosomal RNA. Information about concentration, integrity and library preparation protocols have been added for the Illumina DNA sequencing under the "Illumina sequencing of genomic DNA" section. Information about concentration and integrity of the DNA for Nanopore sequencing has been added to the section "Nanopore sequencing of genomic DNA". In our opinion, there is not much added value in showing the average sequence quality before trimming/filtering of the different datasets because the various quality thresholds used for the DNA sequences used in the assembly is described. For the small RNA and the transcriptome data we have not performed any filtering or assembly and the raw data is provided.

7. There are different versions of Albacore for different versions of Python. So, we believe it is more specific to keep this information. Albacore is not published (and it is not developed any more) but we added a reference to Oxford Nanopore who developed the software.

8. Corrected.

9. The information has been added.

10. The information has been added.

**Competing Interests:** The authors have no competing interests.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

F1000Research