Protist, Vol. 170, 125697, December 2019
http://www.elsevier.de/protis
Published online date 25 October 2019

Protist

ORIGINAL PAPER

Protist Genome Reports

The Draft Genome of the Small, Spineless Green Alga *Desmodesmus costato-granulatus* (Sphaeropleales, Chlorophyta)

Sibo Wang[a,b,2], Linzhou Li[a,c,2], Yan Xu[d,e], Barbara Melkonian[f], Maike Lorenz[g], Thomas Friedl[g], Morten Petersen[b], Sunil Kumar Sahu[a,e], Michael Melkonian[f,1], and Huan Liu[a,b,1]

[a]BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China
[b]Department of Biology, University of Copenhagen, Copenhagen, Denmark
[c]Department of Biotechnology and Biomedicine, Technical University of Denmark, Copenhagen, Denmark
[d]BGI Education Center, University of Chinese Academy of Sciences, Beijing, China
[e]State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, China
[f]University of Duisburg-Essen, Campus Essen, Faculty of Biology, Universitätsstr. 2, 45141 Essen, Germany
[g]Department 'Experimentelle Phykologie und Sammlung von Algenkulturen', University of Göttingen, Nikolausberger Weg 18, 37073 Göttingen, Germany

Submitted October 9, 2019; Accepted October 21, 2019

Desmodesmus costato-granulatus **(Skuja) Hegewald 2000 (Sphaeropleales, Chlorophyta) is a small, spineless green alga that is abundant in the freshwater phytoplankton of oligo- to eutrophic waters worldwide. It has a high lipid content and is considered for sustainable production of diverse compounds, including biofuels. Here, we report the draft whole-genome shotgun sequencing of *D. costato-granulatus* strain SAG 18.81. The final assembly comprises 48,879,637 bp with over 4,141 scaffolds. This whole-genome project is publicly available in the CNSA (https://db.cngb.org/cnsa/) of CNGBdb under the accession number CNP0000701.**

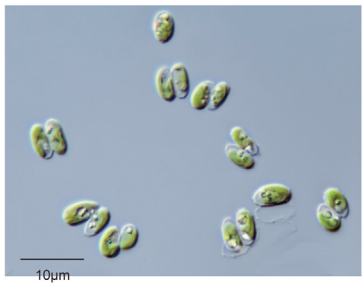
Key words: Scenedesmaceae; genome; comparative genomics.

Coccoid green algae that form flat coenobia consisting of 2-32 elongate cells, arranged in one or two rows, and reproducing asexually by autospores that assemble into a new coenobium within the parental cell, have traditionally been classified in genus

[1]Corresponding authors;
[2]Equal contribution.
e-mails michael.melkonian@uni-koeln.de (M. Melkonian), liuhuan@genomics.cn (H. Liu).

https://doi.org/10.1016/j.protis.2019.125697

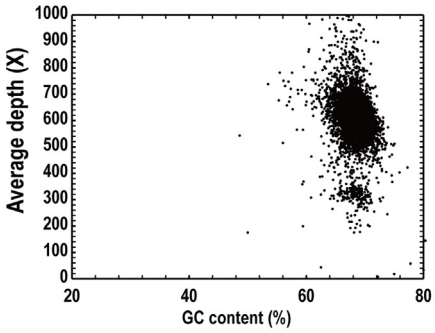
A



B

Feature	*Desmodesmus costato-granulatus*
Genome assembly size (bp)	48,879,637
Scaffold N50 (bp)	196,681
Contig N50 (bp)	15,731
% N	10.17
% G + C	68
% BUSCO Assembly-Viridiplantae	80.4
Number of genes	9,985
Average gene size (bp)	2,512
Known repeats (bp)	1,589,936
Unknown repeats (bp)	380,388

C



D

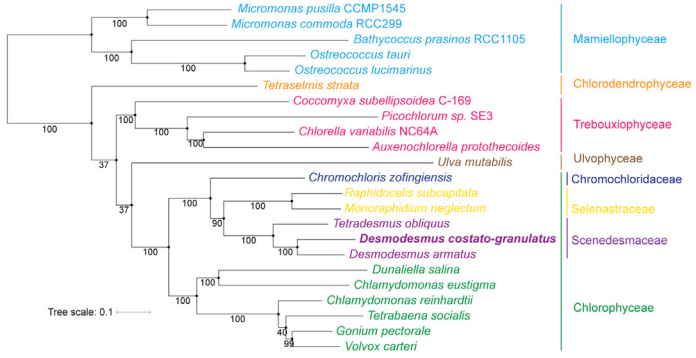


Figure 1. Genome Features and phylogenomic analysis of *D. costato-granulatus.* (**A**) Light micrograph (Nomarski Interference Contrast) of *D. costato-granulatus;* (**B**) The table displays the summary statistics of the *D. costato-granulatus* genome. (**C**) GC-depth plot showing the distribution between the GC content and the average reads mapping depth. The slide window is 500bp. (**D**) The phylogenetic tree was constructed by using the maximum-likelihood method in RAxML based on a concatenated sequence alignment of 76 single-copy genes (25,277 amino acid positions; numbers on branches refer to % bootstrap values; in total 500 bootstrap iterations).

Scenedesmus (Komárek and Fott 1983). Originally a subgenus (Hegewald 1979), *Desmodesmus* is characterized by an additional, outer sporopolleninic cell wall layer with protrusions such as granulations and/or ribs (Hegewald 1997). Molecular phylogenetic analyses confirmed the monophyly of *Desmodesmus* and raised it to genus status (An et al. 1999). *Desmodesmus* is one of the most abundant coccoid green algal genera in fresh to brackish waters globally, and is also the most species-rich taxon within the subfamily Desmodesmoideae (Guiry and Guiry 2019; Hegewald and Braband 2017). Most of the *Desmodesmus* species have a convoluted taxonomic history with many synonymic names (Hegewald 2000) and that also applies to *D. costato-granulatus* (Skuja) Hegewald (Hegewald et al. 1994; Vanormelingen et al. 2007). Although most *Desmodesmus* species have one or several spines (or dents) on the cells, species with small, spineless, two-celled coenobia, such as *D. costato-granulatus* (Fig. 1A) are also common, from oligotrophic to eutrophic waters. The alga may form blooms in spring and is then an important link in the food web (Casper 1985). A strain of *D. costato-granulatus* has been used as feedstock for rotifers (Schlüter and Groeneweg 1981) and more recently, *D. costato-granulatus* was tested as a potential source for algal biofuel (Kodihalli et al. 2018). The draft nuclear genome of *D. costato-granulatus* (strain SAG 18.81) has been established in the frame of the 10 KP project (Cheng et al., 2018), a phylodiverse genome sequencing plan.

An axenic culture of *D. costato-granulatus* (SAG18.81; Sammlung von Algenkulturen, University of Göttingen, Germany) was grown in 3N BBM +V culture medium (https://www.ccap.ac.uk/media/documents/3N_BBM_V.pdf) in aerated Erlenmeyer flasks at 40 μmol photons m^{-2} s^{-1} in a 14:10 hr L/D cycle up to a volume of 1.5 L. The dense culture was harvested by centrifugation (400 × *g,* 10 min), the pellet rapidly transferred into liquid nitrogen and stored at −80 °C until freeze-drying. The DNA extraction was done on the freeze-dried samples using the CTAB method (Sahu et al. 2012). During all steps of cultivation until nucleic acid extraction, axenicity was monitored by sterility tests as well as light microscopy. Light microscopy was performed with

a Leica DMLB light microscope using a PL-APO 100/1.40 objective, an immersed condenser N.A. 1.4 and a Metz Mecablitz 32 Ct3 flash system.

The genomic DNA was used to construct an stLFR (single-tube long fragment reads) genomics library and sequenced by the BGISEQ-500 platform. Single tube long fragment read (stLFR) technology allows efficient WGS, haplotyping, and contig scaffolding at comparatively lower cost. It adds the same barcode sequence to sub-fragments of the original DNA molecule (DNA co-barcoding) and generates long reads using second generation sequencing instruments like BGISEQ (Wang et al. 2019). Paired-end reads of 100bp in length were generated using the BGISEQ-500 sequencer. For assembly, the low quality and duplicated reads were first filtered using SOAPfilter with default parameters (v2.2), and the barcodes of the stLFR reads were changed to 10X Genomics type. Then, the raw reads were directly used for genome assembly by using the Supernova software (v2.1.1) according to the manufacturer's protocol (Weisenfeld et al. 2017). The total genome size was estimated as 46.6 Mb by using the K-mer method (Supplementary Material Fig. S1). The sequenced DNA reads were assembled to 48.9 Mb with 4,141 scaffolds covering 104.9% of the predicted genome size with a scaffold N50 reaching 196 kb (Fig. 1B). The *D. costato-granulatus* genome size is similar to that of "*Scenedesmus quadricauda*" LWG002611 (65.35 Mb; Dasgupta et al. 2018) and some sequenced Selenastraceae (*Chromochloris zofingiensis*, *Monoraphidium neglectum* and *Raphidocelis subcapitata*), but two other *Desmodesmus* species, *D. armatus* (115 Mb; https://www.ncbi.nlm.nih.gov/assembly/GCA_007449985.1) and *D.* obliquus (108 Mb; Carreres et al. 2017) display significantly larger genome sizes compared to *D. costato-granulatus*. An even larger genome size has been reported for another strain of *D. obliquus* (208 Mb; Starkenburg et al. 2017). We also checked the GC bias, which describes the relationship between GC content and read coverage across a genome (Fig. 1C). No contamination was identified.

Three types of repeats (DNA transposon elements, retrotransposon elements, tandem repeats) were identified in the genome of *D. costato-granulatus.* DNA transposons and retrotransposon elements were identified using MITE-hunter and LTRharvest, respectively. RepeatModeler (version 1.0.8) was used to search for other repeats using the *denovo* approach. In addition, the RepeatMasker was also used for a custom library consisting of Repbase and a de novo predicted repetitive elements.

For gene prediction, we used de novo gene prediction methods. Briefly, the program GeneMark-ES (version 4.32) was used to analyze the genome by utilizing unsupervised training and the genes having an incomplete gene structure were removed from the predicted gene sets. The final gene set was evaluated with two approaches. The BUSCO (version 3; Waterhouse et al. 2018) core eukaryotic gene-mapping approach was used to determine the gene set completeness. For the functional annotation of genes, BLASTP ($1e-5$) was used against several known databases, such as SwissProt, KEGG, COG. InterProScan (using data from Pfam, PRINTS, SMART, ProDom and PROSITE) was used to identify protein motifs and protein domains of the predicted gene set. Gene Ontology information was obtained through Blast2go (version 2.5.0).

The genomes of *D. costato-granulatus* were compared with selected algal genomes. These algal genomes were used to define orthogroups (using OrthoFinder, version 1.1.8). Single copy orthogroups (i.e. gene families with only one gene member per species) were used to construct phylogenetic trees based on maximum likelihood. For phylogenetic tree construction, RAxML was used with the CAT + GTR amino acid substitution model. The online tool iTOL was applied to edit and display the final phylogenetic tree.

We predicted 9,985 protein-coding genes. The genome completeness analyses revealed that about 80.4% of the assembled genome captured the Viridiplantae BUSCO v10 dataset. Repetitive elements account for 3.6% of the genome. The long terminal repeats (LTRs) and long interspersed elements (LINEs) were dominant among these repeats, representing 167 kb and 901 kb, respectively (Supplementary Material Table S1). Of the predicted protein-coding genes, 5,808 (58.2%) were assigned putative functions using the SwissProt database, and 5,361 (53.6%) were assigned to KEGG orthologs. A phylogenetic tree inferred from concatenated 76 nuclear-encoded, single-copy genes supported the position of *D. costato-granulatus* within Sphaeropleales and a sister-group relationship of Scenedesmaceae with Selenastraceae in the Chlorophyceae (Fig. 1D).

Based on KEGG pathway mapping, we annotated and mapped 136 level-3 pathways which belong to 22 level-2 metabolic pathways. A summary of the findings is presented in Figure 2A. The largest number of sequences were those associated with global and overview maps (1,195, 12.0%),

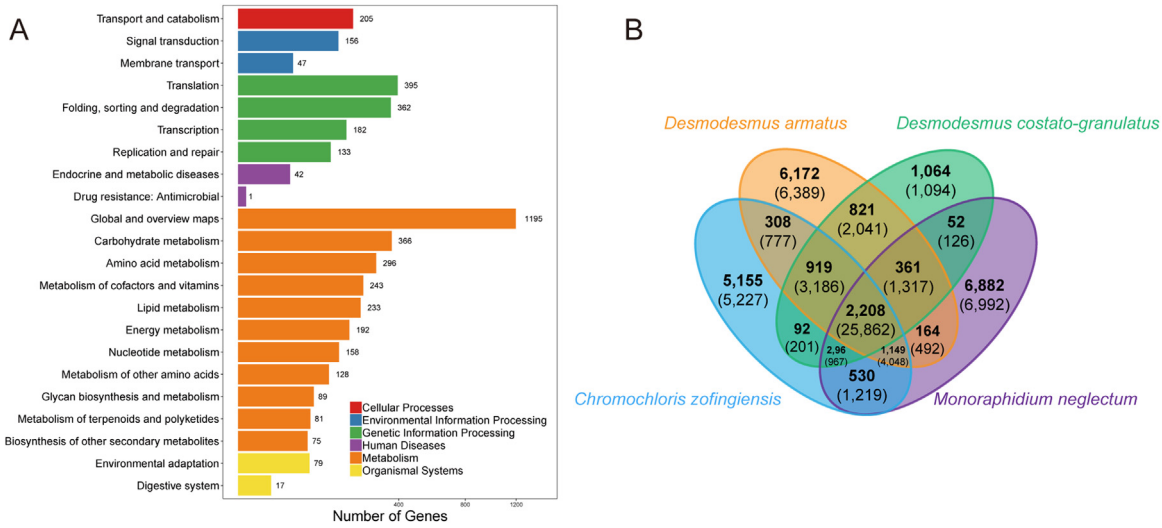


Figure 2. Comparative analysis of the genomes of *D. costato-granulatus* and other Sphaeropleales. (**A**) KEGG pathway mapping of annotated *D. costato-granulatus* genes. (**B**) Venn diagram showing unique and shared orthogroups among *D. costato-granulatus*, *Chromochloris zofingiensis*, *Monoraphidium neglectum* and *Desmodesmus armatus*. Gene numbers are given in parentheses.

followed by sequences that are involved in translation (395, 4.0%), carbohydrate metabolism (366, 3.7%), and "folding, sorting and degradation" (362, 3.7%).

Nearly 2,208 orthogroups of *D. costato-granulatus* are shared with three other Sphaeropleales used in our genome comparisons (Fig. 2B). Of the 1,094 genes unique to *D. costato-granulatus*, most are involved in ubiquitin-mediated proteolysis, pyrimidine metabolism, and starch and sucrose metabolism. To further compare the gene complement between *D. costato-granulatus* and three Sphaeropleales, we performed KEGG enrichment of unique genes between *D. costato-granulatus* and each Sphaeropleales. 2,388 genes were unique to *D. costato-granulatus* when performing the gene family clustering between *D. costato-granulatus* and *D. armatus*, and these unique genes were found to be enriched in various categories of the biological process like purine metabolism, ubiquitin-mediated proteolysis, and monoterpenoid biosynthesis. 6,552 unique genes could be identified when comparing the gene family clustering between *D. costato-granulatus* and *M. neglectum*, and according to the KEGG enrichment display, these unique genes mainly participate in taurine and hypotaurine metabolism, ubiquitin mediated proteolysis, and starch and sucrose metabolism. Gene family clustering between *D. costato-granulatus* and *C. zofingiensis* showed 4,632 proteins unique to *D. costato-granulatus*, which were highly enriched in several categories of biological process namely plant hormone signal transduction, taurine and hypotaurine metabolism, and starch and sucrose metabolism.

The genome data of *D. costato-granulatus* provide a foundation for future genetic and genomic research on an important component of the freshwater phytoplankton. To the best of our knowledge this is the first draft genome of a spineless member of the genus *Desmodesmus*.

Data Availability

The data supporting the findings of this study are available in the CNSA (https://db.cngb.org/cnsa/) of CNGBdb under the accession number CNP0000701.

Acknowledgements

We thank Gerd Günther (http://www.mikroskopia.de/index.html), who took microscopic images of *D. costato-granulatus* strain SAG 18.81. This work is part of the 10KP project led by BGI-Shenzhen and China National GeneBank.

Appendix A. Supplementary Data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.protis.2019.125697.

References

An SS, Friedl T, Hegewald E (1999) Phylogenetic relationships of *Scenedesmus* and *Scenedesmus*-like coccoid green algae as inferred from ITS-2 rDNA sequence comparisons. Plant Biol **1**:418–428

Carreres BM, de Jaeger L, Springer J, Barbosa MJ, Breuer G, van den End EJ, Kleinegris DMM, Schäffers I, Wolbert EJH, Zhang H, Lamers PP, Draaisma RB, Martins dos Santos VAP, Wijffels RH, Eggink G, Schaap PJ, Martens DE (2017) Draft genome sequence of the oleaginous green alga *Tetradesmus obliquus* UTEX 393. Genome Announc **5**, e01449-16

Casper SJ (ed) Lake Stechlin: A Temperate Oligotrophic Lake (Vol. 58). Junk Publ., Dordrecht, 553 p

Cheng S, Melkonian M, Smith SA, Brockington S, Archibald JM, Delaux P-M, Li F-W, Melkonian B, Mavrodiev EV, Fu WSY, Yang H, Soltis DE, Graham SW, Soltis PS, Liu X, Xu X, GK-S Wong (2018) 10KP: A phylodiverse genome sequencing plan. GigaScience **7**:1–9

Dasgupta CN, Nayaka S, Toppo K, Singh AK, Deshpande U, Mohapatra A (2018) Draft genome sequence and detailed characterization of biofuel production by oleaginous microalga *Scenedesmus quadricauda* LWG002611. Biotechnol Biofuels **11**:308

Guiry MD, Guiry GM (2019) AlgaeBase. World-wide electronic publication. National University of Ireland, Galway, searched on 21 October 2019 http://www.algaebase.org

Hegewald E (1979) Eine neue Unterteilung der Gattung *Scenedesmus* Meyen. Nova Hedwigia **30**:343–376

Hegewald E (1997) Taxonomy and phylogeny of *Scenedesmus*. Algae **12**:235–246

Hegewald E (2000) New combinations in the genus *Desmodesmus* (Chlorophyceae, Scenedesmaceae). Arch Hydrobiol Suppl Algol Stud **96**:1–18

Hegewald E, Braband A (2017) A taxonomic revision of *Desmodesmus* serie *Desmodesmus* (Sphaeropleales, Scenedesmaceae). Fottea **17**:191–208

Hegewald E, Krienitz L, Schnepf E (1994) Studies on *Scenedesmus costato-granulatus* Skuja. Nova Hedwigia **59**:97–128

Kodihalli MH, Sriapareddy T, Haniyambadi BMK, Yelagiri MP (2018) Screening of potential microalgae species from different natural environment for biodiesel production. Int J Microbiol Res **10**:1052–1057

Komárek J, Fott B (1983) Chlorophyceae (Grünalgen). Ordnung: Chlorococcales. In Huber-Pestalozzi (ed) Das Phytoplankton des Süßwassers. Systematik und Biologie. 7.Teil, 1. Hälfte. E. Schweizerbart'sche, Verlagsbuchhandlung, Stuttgart, 1044 p

Sahu SK, Thangaraj M, Kathiresan K (2012) DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. ISRN Mol Biol **2012**:205049

Schlüter M, Groeneweg J (1981) Mass production of freshwater rotifers on liquid wastes: I. The influence of some environmental factors on population growth of *Brachionus rubens* Ehrenberg 1838. Aquaculture **25**:17–24

Starkenburg SR, Polle JEW, Hovde B, Daligault HE, Davenport KW, Huang A, Neofotis P, McKie-Krisberg Z (2017) Draft nuclear genome, complete chloroplast genome, and complete mitochondrial genome for the biofuel/bioproduct feedstock species *Scenedesmus obliquus* strain DOE0152z. Genome Announc **5**:e00617-17

Vanormelingen P, Hegewald E, Braband A, Kitschke M, Friedl T, Sabbe K, Vyverman W (2007) The systematics of a small spineless *Desmodesmus* species, *D. costato-granulatus* (Sphaeropleales, Chlorophyceae), based on ITS2 rDNA sequence analyses and cell wall morphology. J Phycol **43**:378–396

Wang O, Chin R, Cheng X, Wu M, Mao Q, Tang J, Sun Y, Anderson E, Lam HK, Chen D, Zhou Y, Wang L, Fan F, Zou Y, Xie Y, Zhang RY, Drmanac S, Nguyen D, Xu C, Villarosa C, Gablenz S, Barua N, Nguyen S, Tian W, Liu JS, Wang J, Liu X, Qi X, Chen A, Wang H, Dong Y, Zhang W, Alexeev A, Yang H, Wang J, Kristiansen K, Xu X, Drmanac R, Peters B (2019) Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. Genome Res **29**:798–808

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol **35**:543–548

Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB (2017) Direct determination of diploid genome sequences. Genome Res **27**:757–767