



De novo Assembly and Annotation of the Antarctic Alga *Prasiola crispa* Transcriptome

Evelise L. Carvalho^{1†}, Lucas F. Maciel^{1†}, Pablo E. Macedo¹, Filipe Z. Dezordi¹, Maria E. T. Abreu¹, Filipe de Carvalho Victória², Antônio B. Pereira², Juliano T. Boldo¹, Gabriel da Luz Wallau³ and Paulo M. Pinto^{1*}

¹ Applied Proteomics Laboratory, Federal University of Pampa, São Gabriel, Brazil, ² Núcleo de Estudos da Vegetação Antártica, Federal University of Pampa, São Gabriel, Brazil, ³ Department of Entomology, Aggeu Magalhães Institute (IAM), Recife, Brazil

Keywords: RNA-seq, Trebouxiophyceae, Prasiolales, transcriptome, extreme environments, anti-freeze proteins

INTRODUCTION

The Antarctic, located on the South Pole of the Earth and isolated from other continents by the Atlantic, Pacific and Indian oceans. It is considered a continent with severe environmental conditions for the development of life, thus limiting the Antarctic fauna and flora to specific organisms that have survival adaptation mechanisms (Jackson and Seppelt, 1997). The average annual precipitation of Antarctic is only 200 mm with winds of 327 km/h and temperatures below -90°C have already been recorded (Martínez-Rosales et al., 2012). The total area is 14,000,000 km², with 98–99.7% covered by snow and ice, with layers averaging 1.6 km thick (Convey et al., 2008; Martínez-Rosales et al., 2012). In addition, the ozone hole over the Antarctic region, first described in the 1980s, causes a high rate of ultraviolet radiation over the region, which is intensified by the ice-generated reflection (Kuttippurath and Nair, 2017; Marizcurrena et al., 2017).

Among the algae present on the Antarctic ice-free areas, *Prasiola crispa* (Lightfoot) Kützing is the most commonly found organism. *P. crispa* is a green macroalga belonging to the Trebouxiophyceae class and is among the most important primary producers in the Antarctic territory. *P. crispa* occurs in hydro-terrestrial soils, in the supralittoral zones of the maritime and continental Antarctica, where they form large and green carpets on the humid soil. *P. crispa* is found close to bird populations, mainly adjacent to penguin colonies, where the soil is rich in guano, a substrate with a high incidence of uric acid and nitrogen compounds (Kováčik and Pereira, 2001).

The morphology of these algae varies from uniseriated filaments to stalks in the form of a tape, expanded blades or packages as colonies, which are characterized by a large phenotypic plasticity related to environmental factors (Rindi et al., 2007).

During the course of the seasons, *P. crispa* needs to tolerate extreme environments, such as repeated freeze and thaw cycles, physiological drought, salinity stress, and high levels of UV radiation (Jacob et al., 1992; Jackson and Seppelt, 1997). However, the genes associated with these adaptive characteristics in *P. crispa* remain unknown. Therefore, to better understand the genetic and metabolic adaptations that allow this organism to survive in harsh environments, we sequenced its transcriptome.

Transcriptomes represent all the expressed fractions of genomes and are a viable alternative to understand and characterize genome wide genetic information of organisms since it simplifies genetic analyses, as compared to whole genome sequencing (Riesgo et al., 2012).

High-throughput sequencing of transcriptomes (RNA-Seq) has provided new routes to study the genetic and functional information stored within any organism at an unprecedented scale and speed. Transcriptome approaches have been employed in a large number of the studies involving

OPEN ACCESS

Edited by:

Philipp Kapranov,
Huaqiao University, China

Reviewed by:

Sergio Verjovski-Almeida,
University of São Paulo, Brazil
Peter G. Zaphiropoulos,
Karolinska Institute (KI), Sweden

*Correspondence:

Paulo M. Pinto
paulopinto@unipampa.edu.br

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Molecular Biosciences

Received: 26 September 2017

Accepted: 05 December 2017

Published: 08 January 2018

Citation:

Carvalho EL, Maciel LF, Macedo PE, Dezordi FZ, Abreu MET, Victória FdC, Pereira AB, Boldo JT, Wallau GdL and Pinto PM (2018) De novo Assembly and Annotation of the Antarctic Alga *Prasiola crispa* Transcriptome. *Front. Mol. Biosci.* 4:89. doi: 10.3389/fmolb.2017.00089

non-model organisms, which normally lack reference genomes (Ekblom and Galindo, 2011; Haas et al., 2013).

Among these organisms are the algae group. The available data consists of organisms belonging to different phylum, such as Prymnesiophyte (Koid et al., 2014), Chlorophyta (Rismani-Yazdi et al., 2011), Haptophyta (Talarski et al., 2016), Stramenopiles (Im et al., 2015), and Rhodophyta (Shuangxiu et al., 2014).

P. crispa represents the first organism of the Prasiolales order with an available transcriptome since, until this work, the mitochondrial and plastid genomes were the only molecular data available for this species (Carvalho et al., 2015, 2017). Therefore, the purpose of this study was to sequence the transcriptome of *P. crispa*. The identification of transcripts will help to identify genes that are responsible for organism survival in this environment, as well as assisting in future genetic, phylogenetic, and biotechnological studies of *P. crispa* and other Antarctic organisms.

EXPERIMENTAL DESIGN, MATERIALS, AND METHODS

Algae Collection

P. crispa was collected in areas near the Arctowski Polish Station Region, Admiralty Bay, King George Island (61°50′–62°15′ S and 57°30′–59°00′W), Antarctic. The collection was carried out in the Antarctic summer, in January of 2014 Austral summer, with temperature ranging from 0.5 to 2.0°C. The samples were maintained in RNAlater[®] (Sigma-Aldrich, USA) until RNA extraction.

Total RNA Extraction and RNA-Seq

Total RNA was extracted from three pools of samples using an RNAqueous[®]-Micro Total RNA Isolation Kit (Thermo Fisher Scientific Inc., USA) according to the manufacturer's instructions. The RNA-Seq library was prepared using random primers. The transcriptome was sequenced by Macrogen Service using the Solexa-Illumina HiSeq 2000 next-generation sequencing platform device according to the manufacturer's instructions. A paired-end reads with a read size of ~100 bp separated by insert size of 300 bp was employed.

The BioProject ID of our data is PRJNA329112, and the BioSample accession number is SAMN05392062. All raw reads were deposited into the Sequencing Read Archive (SRA) of NCBI with accession number SRR5754271. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFTS00000000.

De novo Transcriptome Assembly

Raw reads from data sets were filtered to remove the adapter sequences, and low quality reads with Fastx-toolkit (quality cut-off = 30) (Gordon and Hannon, 2010) and Trimmomatic v 0.36 using default parameters (Bolger et al., 2014). Next, we used Trinity version r2014-07-17 (Ekblom and Galindo, 2011) as Bruijn graph assembler with 25 kmer size. Due to the sequencing of a complex sample extracted from the Antarctic soil, we expected some amount of bacterial and fungal contamination. Therefore, in order to remove such contaminants, we performed

a tblastx with default parameters (Altschul et al., 1990) searching against all of the NCBI nucleotide non-redundant database and recovered all contigs in which the best blast hit occurred with algae and plant sequences. Next, we used Bowtie2 (Langmead and Salzberg, 2012) with default parameters to recover only the reads that mapped against those *P. crispa* contigs.

Functional Annotation

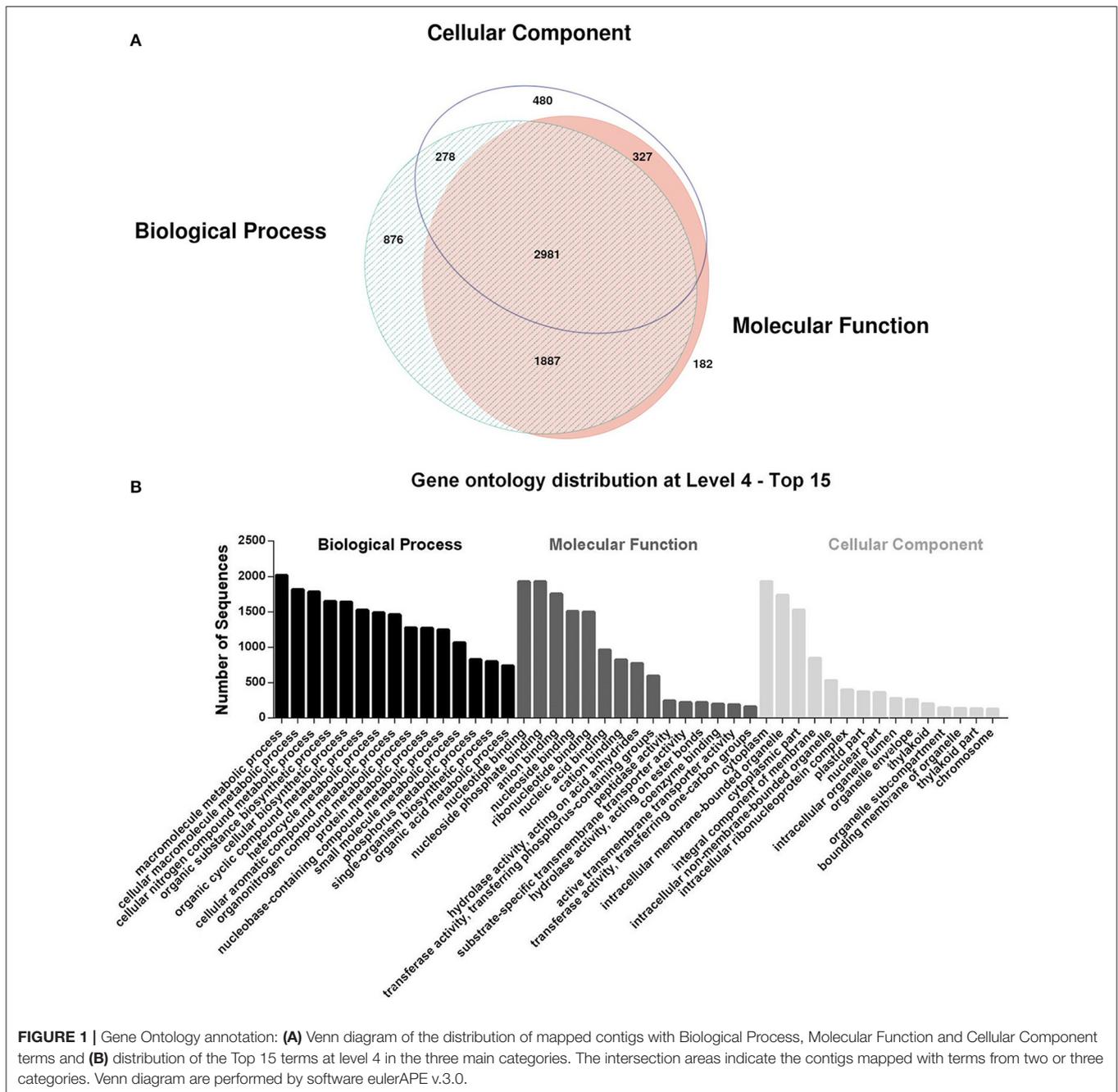
The assembled and recovered contigs were searched against the NCBI protein non-redundant database using the BLASTX algorithm; the *E*-value cut-off was set at 1e-10. Genes were tentatively identified based on the best hits against known sequences. Blast2GO (Conesa and Götz, 2008) was used for mapping and annotation, associating Gene Ontology (GO) terms and predicting their function. Assembled and annotated transcriptome is publicly available on Figshare at: <https://figshare.com/s/a60ae8a0445d547b9360>.

After a stringent filtering process, the processed reads were assembled into 17,201 contigs. Statistics of the assembly are summarized in **Table 1**. CD-HIT-EST version 4.6.8, 2017-06-21 (Fu et al., 2012) was used for clustering of assembled transcripts with the default parameters with sequence identity threshold set to 95%, in order to indicate the number of unigenes. The clustering reduced the number of transcripts marginally by 3.5%. Our analysis showed that 93% of the unigenes are represented by only one isoform. The metrics of *P. crispa* were compared with others transcriptomes of the organism from the Trebouxiophyceae class, including *Chlorella minutissima* (Yu et al., 2016), *Trebouxia gelatinosa* (Carniel et al., 2016), *Coccomyxa subellipsoidea* (Peng et al., 2016), *Chlorella sorokiniana* (Li et al., 2016), *Botryococcus braunii* (Xu et al., 2015), and was found to have the lowest number of total reads. In relation to the number of contigs, *P. crispa* is in an intermediary position, with *C. sorokiniana* being the organism with the highest number of contigs (63,811) and *C. subellipsoidea* having the lowest number of contigs (9,409). More information on the metrics of transcriptomes is given in Supplementary Table S1.

The search of these contigs against the NCBI protein non-redundant database with BLASTX demonstrates that 8,980 (52.19%) sequences had at least one hit. The mapping of the sequences against the GO database retrieved 7,009 sequences mapped, and all assigned GO terms were classified into three

TABLE 1 | Summary of *Prasiola crispa* assembly.

Attributes	Value
Total raw reads	42,978,976
Total processed reads	5,233,428
Number of contigs	17,201
Total length (pb)	13,127,645
N50	1,036
GC (%)	49.66
Average size (pb)	763.19
Size range (pb)	200–12,802
Contigs > 1 kb	3,973 (22.05%)



main categories: cellular component, molecular function, and biological process. The distribution of sequences mapped to the three different categories and the top 15 GO terms are represented in **Figures 1A,B**, respectively.

Thus, taking the comparative numbers of contigs, mean length, the clustering process and sequencing depth and coverage into account its likely that our assembly comprises a representative number of transcripts but which were partially reconstructed. Moreover, it is important to note that although we have used stringent blast parameters to remove contaminations, some contigs can still come from contamination. Therefore,

experimental validation to confirm *P. crispa* origin of some contigs is warranted.

Data Validation and Quality Control

The reading quality of the data of this transcriptomic analysis was evaluated through FastQC software (Babraham Bioinformatics) [RRID:SCR_014583]. The paired-end reads results were merged using MultiQC (<http://multiqc.info>) and are shown in Supplementary Figure S1. Per base quality phred scores range from 32.78 to 40.06, indicating base call accuracies of >99.9% (Supplementary Figure S1A). Per sequence quality

shows that 99.62% of reads had a mean phred score of 30 or above (Supplementary Figure S1B) and per base N content was low, with a maximum value 0.18% (Supplementary Figure S1C).

Re-use Potential

Through the data of this transcriptome, it is possible to perform searches for genes, aiming the heterologous expression of proteins with biotechnological potential, such as antifreeze proteins, which act to inhibit freezing of intracellular fluids (Nath et al., 2013), heat-shock proteins that play an important role in maintain biological activities in algae present in these acclimatization process (Li and Brawley, 2004) and mycosporine-like amino acids responsive to high incidence of ultraviolet radiation (Kováčik and Pereira, 2001). Proteomics approaches may also be employed, aiming at the confirmation of gene expression at the translational level.

AUTHOR CONTRIBUTIONS

EC, LM, GW, PP: Conducted the experiment; EC, LM, PM, FD, MA, GW: Performed analysis on the data; FV, AP, JB, PP:

Conceived the project and acquired funding; EC, LM, PM, GW, PP: Wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by the National Council for Scientific and Technological Development (CNPq-Brazil), the Coordination for the Improvement of Higher Education Personnel (CAPES-Brazil), the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS-Brazil) and National Institute of Science and Technology—Antarctic Environmental Research (INCT-APA). We thank the Bioinformatic Core at the Aggeu Magalhães Institute for the support with the bioinformatic analysis. This is for Sofia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2017.00089/full#supplementary-material>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search toll. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Carniel, F. B., Gerdol, M., Montagner, A., Banchi, E., De Moro, G., Manfrin, C., et al. (2016). New features of desiccation tolerance in the lichen photobiont *Trebouxia gelatinosa* are revealed by a transcriptomic approach. *Plant Mol. Biol.* 91, 319–339. doi: 10.1007/s11103-016-0468-5
- Carvalho, E. L., Wallau, G. L., Rangel, D. L., Machado, L. C., Pereira, A. B., Victoria, F. C., et al. (2017). Phylogenetic positioning of the Antarctic alga *Prasiola crisper* (Trebouxiophyceae) using organellar genomes and their structural analysis. *J. Phycol.* 53, 908–915. doi: 10.1111/jpy.12541
- Carvalho, E. L., Wallau, G. L., Rangel, D. L., Machado, L. C., da Silva, A. F., da Silva, L. F., et al. (2015). Draft plastid and mitochondrial genome sequences from Antarctic alga *Prasiola crisper*. *Genome Announc.* 3:e01151-15. doi: 10.1128/genomeA.01151-15
- Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832
- Convey, P., Gibson, J. A. E., Hillenbrand, C. D., Hodgson, D. A., Pugh, P. J., Smellie, J. L., et al. (2008). Antarctic terrestrial life – challenging the history of the frozen continent? *Biol. Rev. Camb. Philos. Soc.* 83, 103–117. doi: 10.1111/j.1469-185X.2008.00034.x
- Eklom, R., and Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107, 1–15. doi: 10.1038/hdy.2010.152
- Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150–3152. doi: 10.1093/bioinformatics/bts565
- Gordon, A., and Hannon, G. J. (2010). *Fastx-Toolkit*. FASTQ/A Short-Reads Pre-Processing Tools (unpublished). Available online at: http://hannonlab.Cshl.Edu/fastx_toolkit/
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–512. doi: 10.1038/nprot.2013.084
- Im, S., Choi, S., Hwang, M. S., Park, E. J., Jeong, W. J., and Choi, D. W. (2015). *De novo* assembly of transcriptome from the gametophyte of the marine red algae *Pyropia seriata* and identification of abiotic stress response genes. *J. Appl. Phycol.* 27, 1343–1353. doi: 10.1007/s10811-014-0406-3
- Jackson, A. E., and Seppelt, R. D. (1997). “Physiological adaptations to freezing and UV radiation exposure in *Prasiola crisper*, an Antarctic terrestrial alga,” in *Antarctic Communities: Species, Structure, and Survival*, eds B. Battaglia, J. Valencia, and D. W. H. Walton (Cambridge: University Press), 226–233.
- Jacob, A., Wiencke, C., Lehmann, H., and Kirst, G. O. (1992). Physiology and ultrastructure of desiccation in the green alga *Prasiola crisper* from Antarctica. *Botanica Marina* 35, 297–303. doi: 10.1515/botm.1992.35.4.297
- Koid, A. E., Liu, Z., Terrado, R., Jones, A. C., Caron, D. A., and Heidelberg, K. B. (2014). Comparative transcriptome analysis of four Prymnesiophyte algae. *PLoS ONE* 9:e97801. doi: 10.1371/journal.pone.0097801
- Kováčik, L., and Pereira, A. B. (2001). “Green alga *Prasiola crisper* and its lichenized form *Mastodia tessellata* in Antarctic,” in *Algae and Extreme Environments*, eds J. Elster, J. Seckbach, W. F. Vincent, and O. Lhotský (Czech Republic: Nova Hedwigia 123), 465–478.
- Kuttippurath, J., and Nair, P. J. (2017). The signs of Antarctic ozone hole recovery. *Sci. Rep.* 7:585. doi: 10.1038/s41598-017-00722-7
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Li, L., Zhang, G., and Wang, Q. (2016). *De novo* transcriptomic analysis of *Chlorella sorokiniana* reveals differential genes expression in photosynthetic carbon fixation and lipid production. *BMC Microbiol.* 16:223. doi: 10.1186/s12866-016-0839-8
- Li, R., and Brawley, S. H. (2004). Improved survival under heat stress in intertidal embryos (*Fucus* spp.) simultaneously exposed to hypersalinity and the effect of parental thermal history. *Mar. Biol.* 144, 205–213. doi: 10.1007/s00227-003-1190-9
- Marizcurrena, J. J., Morel, M. A., Braña, V., Morales, D., Martínez-López, W., and Castro-Sowinsk, S. (2017). Searching for novel photolyases in UVC-resistant Antarctic bacteria. *Extremophiles* 21, 409–418. doi: 10.1007/s00792-016-0914-y
- Martínez-Rosales, C., Fullana, N., Musto, H., and Castro-Sowinski, S. (2012). Antarctic DNA moving forward: genomic plasticity and biotechnological potential. *FEMS Microbiol. Lett.* 331, 1–9. doi: 10.1111/j.1574-6968.2012.02531.x

- Nath, A., Chaube, R., and Subbiah, K. (2013). An insight in to the molecular basis for convergent evolution in fish antifreeze proteins. *Comput. Biol. Med.* 43, 817–821. doi: 10.1016/j.combiomed.2013.04.013
- Peng, E., Wei, D., Chen, G., and Chen, F. (2016). Transcriptome analysis reveals global regulation in response to CO₂ supplementation in oleaginous microalga *Coccomyxa subellipsoidea* C-169. *Biotechnol. Biofuels.* 9:151. doi: 10.1186/s13068-016-0571-5
- Riesgo, A., Andrade, S. C., Sharma, P. P., Novo, M., Pérez-Porro, A. R., Vahtera, V., et al. (2012). Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. *Front. Zool.* 9:33. doi: 10.1186/1742-9994-9-33
- Rindi, F., McIvor, L., Sherwood, A. R., Friedl, T., Guiry, M. D., and Sheath, R. G. (2007). Molecular phylogeny of the green algal order Prasiolales (Trebouxiophyceae, Chlorophyta). *J. Phycol.* 43, 811–822. doi: 10.1111/j.1529-8817.2007.00372.x
- Rismani-Yazdi, H., Haznedaroglu, B. Z., Bibby, K., and Peccia, J. (2011). Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: pathway description and gene discovery for production of next-generation biofuels. *BMC Genomics* 12:148. doi: 10.1186/1471-2164-12-148
- Shuangxiu, W., Sun, J., Chi, S., Wang, L., Wang, X., Liu, C., et al. (2014). Transcriptome sequencing of essential marine brown and red algal species in China and its significance in algal biology and phylogeny. *Acta Oceanol. Sin.* 33, 1–12. doi: 10.1007/s13131-014-0435-4
- Talarski, A., Manning, S. R., and La Claire, J. W. II. (2016). Transcriptome analysis of the euryhaline alga, *Prymnesium parvum* (Prymnesiophyceae): effects of salinity on differential gene expression. *Phycologia* 55, 33–44. doi: 10.2216/15-74.1
- Xu, Z., He, J., Qi, S., and Liu, J. (2015). Nitrogen deprivation-induced de novo transcriptomic profiling of the oleaginous green alga *Botryococcus braunii* 779. *Genom. Data* 6, 231–233. doi: 10.1016/j.gdata.2015.09.019
- Yu, M., Yang, S., and Lin, X. (2016). *De-novo* assembly and characterization of *Chlorella minutissima* UTEX2341 transcriptome by paired-end sequencing and the identification of genes related to the biosynthesis of lipids for biodiesel. *Mar. Genomics* 25, 69–74. doi: 10.1016/j.margen.2015.11.005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Carvalho, Maciel, Macedo, Dezordi, Abreu, Victória, Pereira, Boldo, Wallau and Pinto. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.