1 **ALGAEFUN with MARACAS, microALGAE FUNctional enrichment tool for MicroAlgae**

2 **RnA-seq and Chip-seq Analysis**

3

4 Ana B. Romero-Losada^{1,2}, Christina Arvanitidou¹, Pedro de los Reyes¹, Mercedes García-González¹,
5 Francisco J. Romero-Campero^{1,2,*}

6

7 ¹ Institute for Plant Biochemistry and Photosynthesis, Universidad de Sevilla – Consejo Superior de
8 Investigaciones Científicas. Centro de Investigaciones Científicas Isla de la Cartuja. Avenida
9 Américo Vespucio 49, 41092, Seville, Spain

10

11 ² Department of Computer Science and Artificial Intelligence, University of Sevilla, Escuela
12 Técnica Superior en Ingeniería Informática, Avenida Reina Mercedes s/n, 41012, Seville, Spain

13

14 *Corresponding author: Francisco J. Romero-Campero, fran@us.es

15

16

17

18

19

20

21

22

23

24

25

26 **Abstract**

27 **Background:** Microalgae are emerging as promising sustainable sources for biofuels, biostimulants
 28 in agriculture, soil bioremediation, feed and human nutrients. Nonetheless, the molecular
 29 mechanisms underpinning microalgae physiology and the biosynthesis of compounds of
 30 biotechnological interest are largely uncharacterized. This hinders the development of microalgae
 31 full potential as cell-factories. The recent application of omics technologies into microalgae
 32 research aims at unraveling these systems. Nevertheless, the lack of specific tools for analysing
 33 omics raw data generated from microalgae to provide biological meaningful information are
 34 hampering the impact of these technologies. The purpose of ALGAEFUN with MARACAS
 35 consists in providing researchers in microalgae with an enabling tool that will allow them to exploit
 36 transcriptomic and cistromic high-throughput sequencing data.

37 **Results:** ALGAEFUN with MARACAS consists of two different tools. First, MARACAS
 38 (MicroAlgae RnA-seq and Chip-seq AnalysiS) implements a fully automatic computational pipeline
 39 receiving as input RNA-seq (RNA sequencing) or ChIP-seq (chromatin immunoprecipitation
 40 sequencing) raw data from microalgae studies. MARACAS generates sets of differentially
 41 expressed genes or lists of genomic loci for RNA-seq and ChIP-seq analysis respectively. Second,
 42 ALGAEFUN (microALGAE FUNctional enrichment tool) is a web-based application where gene
 43 sets generated from RNA-seq analysis as well as lists of genomic loci from ChIP-seq analysis can
 44 be used as input. On the one hand, it can be used to perform Gene Ontology and biological
 45 pathways enrichment analysis over gene sets. On the other hand, using the results of ChIP-seq data
 46 analysis, it identifies a set of potential target genes and analyses the distribution of the loci over
 47 gene features. Graphical representation of the results as well as tables with gene annotations are
 48 generated and can be downloaded for further analysis.

49 **Conclusions:** ALGAEFUN with MARACAS provides an integrated environment for the
 50 microalgae research community that facilitates the process of obtaining relevant biological

information from raw RNA-seq and ChIP-seq data. These applications are designed to assist researchers in the interpretation of gene lists and genomic loci based on functional enrichment analysis. ALGAEFUN with MARACAS is publicly available on <https://greennetwork.us.es/AlgaeFUN/>.

Keywords

Microalgae, RNA-seq, ChIP-seq, Functional enrichment, GO enrichment, KEGG pathway enrichment

Background

Microalgae are a very diverse non-monophyletic group of photosynthetic microorganisms of special interest due to their physiological plasticity and wide range of biotechnological applications. Microalgae can be found in a wide variety of different habitats, from freshwater to oceans growing under a broad range of temperature, salinity, pH and light intensity values. More than 5000 species have been identified in the oceans accounting for the production of 50% of the oxygen necessary to sustain life on Earth. Microalgae also play a central ecological role as primary producers of biomass establishing the base of aquatic food chains [1]. In the last decades, microalgae have also been of great interest for the scientific community due to the large and yet increasing number of biotechnological applications they present. Specifically, they have been described as a high yield source of carbon compounds and good candidates for the mitigation of CO₂ emissions. In microalgae, fixation of CO₂ is coupled with growth and biosynthesis of compounds of biotechnological interests such as polysaccharides, lipids, vitamins and antioxidants. Their industrial cultivation for the production of biofuels as well as bioproducts used as biostimulants in agriculture, health supplements, pharmaceuticals and cosmetics is also thoroughly explored nowadays [2,3]. Finally, microalgae are successfully applied in wastewater treatment coupled with fixation of atmospheric CO₂ [4].

Due to these promising features, physiological characterization of multiple microalgae species under different cultivation regimes have been thoroughly carried out [5,6]. Nevertheless, the molecular mechanisms underpinning the physiology of microalgae are yet poorly understood. In order to facilitate the progress in the characterization of the molecular systems regulating microalgae physiology, high throughput sequencing technologies have been recently applied to obtain the genome of a wide range of microalgae [7-19]. This has promoted the use of different omics, particularly transcriptomics based on RNA-seq data [20-22] and cistromics based on ChIP-seq data [23,24], to initiate molecular systems biology studies in microalgae. Nonetheless, the impact of this type of studies on microalgae are hampered by the lack of freely available and easy to use online tools to analyze, extract relevant information and integrate omics data. Typically, RNA-seq and ChIP-seq analysis received as input high-throughput sequencing raw data and produce as output, respectively, sets of differentially expressed genes and genomic loci significantly bound by the protein of interest. Processing of the massive amount of high-throughput sequencing data and analysis of the resulting sets of genes and genomic loci obtained from molecular systems biology studies requires computational power, time, effort and expertise that research groups on microalgae may lack. In addition, researchers must explore different data bases separately, which makes the integration of the results and the generation of biological meaningful information more difficult. Therefore, it is imperative the development of frameworks integrating microalgae genome sequences and annotations with tools for high-throughput sequencing data analysis and functional enrichment of gene and genomic loci sets. In order to cover these microalgae research community needs and promote studies in molecular systems biology we have developed the web portal ALGAEFUN with MARACAS using the R package Shiny [25] and other Bioconductor packages. Our web portal consists of two different tools. MARACAS (MicroAlgae RnA-seq and Chip-seq Analysis) implements an automatic computational workflow that receives as input RNA-seq or ChIP-seq raw sequencing data from microalgae studies and produces, respectively, sets of

differentially expressed genes or a list of genomic loci. These results can be further analyzed using our second tool ALGAEFUN (microAlgae FUNctional enrichment tool). On the one hand, when receiving the results from an RNA-seq analysis, sets of genes are functionally annotated by performing GO (Gene Ontology) [26] and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways [27] enrichment analysis. On the other hand, when genomic loci from a ChIP-seq analysis are inputted, a set of potential target genes is generated together with the analysis of the distribution of the loci over gene features as well as metagene plots representing the average mapping signal. This set of potential target genes can be further studied using the features for functional enrichment analysis in ALGAEFUN as described above. ALGAEFUN with MARACAS supports a wide range of 14 different microalgae species including *Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Nannochloropsis gaditana* and *Phaeodactylum tricornutum* among others. Additionally, our tools can be easily extended to include new microalgae as their genome sequence and annotation are made available. The code for ALGAEFUN with MARACAS is publicly available at their respective GitHub repositories from the following links: <https://github.com/fran-romero-campero/ALGAEFUN> and <https://github.com/fran-romero-campero/MARACAS>.

Implementation

ALGAEFUN with MARACAS supports the analysis of the following microalgae covering an ample spectrum of their phylogeny: *Chlamydomonas reinhardtii* [7], *Volvox carteri* [8], *Chromochloris zofingiensis* [9], *Dunaliella salina* [10], *Haematococcus lacustris* [11] (Chlorophyceae), *Coccomyxa subellipsoidea* [12] (Trebouxiophyceae), *Ostreococcus tauri* [13], *Bathycoccus prasinos* [14], *Micromonas pusilla* CCMP1545 [15] (Mamiellophyceae), *Phaeodactylum tricornutum* [16], *Nannochloropsis gaditana* [17] (Stramenopiles), *Klebsormidium nitens* [18], *Mesotaenium endlicherianum* and *Spirogloea muscicola* [19] (Charophyceae), Fig. 1.

126

127 **Figure 1. Phylogenetic relationship between the different microalgae species supported in**
128 **ALGAEFUN with MARACAS.** The Mamiellophyceae *Ostreococcus tauri*, *Micromonas pusilla*
129 *CCMP1545* and *Bathycoccus prasinos* (olive green rectangle); the Trebouxiophyceae *Coccomyxa*
130 *subellipsoidea* (blue rectangle); the Chlorophyceae *Chlamydomonas reinhardtii*, *Volvox carteri*,
131 *Dunaliella salina*, *Haematococcus lacustris* and *Chromochloris zofingiensis* (red rectangle); the
132 Charophyceae *Klebsormidium nitens*, *Mesotaenium endlicherianum* and *Spirogloea muscicola*
133 (green rectangle); and the Stramenopiles *Phaeodactylum tricornutum*, *Nannochloropsis gaditana*
134 (brown rectangle).

135

136 One of the limiting factors for the development of molecular systems biology studies in microalgae
137 is the fragmentation of the available genomic sequences and functional annotations into different
138 databases. To overcome this issue and generate easily accessible resources, genome sequences,
139 functional annotation and genomic feature annotation files for the previous microalgae species were
140 systematically collected from different freely available databases: Ensembl Protists [28],

PhycoCosm [29], Phytozome [30], Genomes - NCBI Datasets [31] and a Figshare repository [32], see Table 1.

Table 1. Resources used to collect genome sequences, functional and gene feature annotations for each supported microalga.

Ensembl Protists	PhycoCosm	Phytozome	Genomes – NCBI Datasets	Figshare associated to publication
<i>Nannochloropsis</i>	<i>Ostreococcus tauri</i>	<i>Chlamydomonas</i>	<i>Haematococcus</i>	<i>Mesotaenium</i>
<i>gaditana</i>	<i>Micromonas pusilla</i>	<i>reinhardtii</i>	<i>lacustris</i>	<i>endlicherianum</i>
<i>Phaeodactylum</i>	CCMP1545	<i>Volvox carteri</i>		<i>Spirogloea</i>
<i>tricornutum</i>	<i>Bathycoccus prasinos</i>	Chromochloris		<i>musciicola</i>
	<i>Klebsormidium nitens</i>	<i>zofingiensis</i>		
		<i>Dunaliella salina</i>		
		<i>Coccomyxa</i>		
		<i>subellipsoidea</i>		

The MARACAS pipeline is based on the short read mappers HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts 2) [33] and Bowtie2 [34]. Pre-computed genome indexes for each microalga are included in MARACAS to alleviate the high-cost computational tasks consisting of read mapping to reference genomes. Gene features annotation files in GTF (gene transfer format) for each microalga were also collected. Transcript assembly and estimation of gene expression is performed using StringTie [33]. Identification of differentially expressed genes and graphical representation of the results are carried out using the R Bioconductor packages Ballgown [33] and LIMMA (Linear Models for Microarray Analysis) [35]. For ChIP-seq data analysis, the peak caller

MACS2 (Model-based Analysis of ChIP-seq 2) [36] is used. MARACAS can be executed in sequential mode or in distributed/parallel mode on a computational cluster managed by the job scheduling system SLURM (Simple Linux Utility for Resource Management). In the distributed/parallel mode, synchronization among the jobs processing different samples is implemented using a blackboard system. A common blackboard file is used where parallel jobs can read and write to ensure that all samples have been processed before proceeding to the next steps. This effectively implements synchronization points during the workflow in MARACAS. For the implementation of ALGAEFUN, functional annotation files were also downloaded for each microalga from the previously mentioned databases, Table 1. Specifically, Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genome) Orthology (KO) terms were collected. For microalgae species lacking these annotation systems HMMER (biological sequence analysis using profile hidden Markov models) [37] was used to identify protein domains according to the PFAM (Protein Family) nomenclature. PFAM terms were subsequently converted into GO terms using pfam2go. KO terms were associated to genes applying KAAS (KEGG Automatic Annotation Server) [38]. Additionally, whenever available other systematic functional annotation formats were also included such as Protein Analysis Through Evolutionary Relationships (PANTHER) terms [39], Enzyme Commission numbers (EC numbers) and Eukaryotic Orthologous Groups (KOG) terms [40]. In order to use all these functional annotation systems in ALGAEFUN two different types of R annotation packages were developed and were made freely available from our Github repository [41]. On the one hand, using the function makeOrgPackage from the Bioconductor R package AnnotationForge [42] we generated annotation packages for each microalga integrating the systematic sources of functional annotation discussed previously. These packages are instrumental when performing functional enrichment analysis over gene sets obtained, for instance, from a differential expression analysis based on RNA-seq data using MARACAS. On the other hand, applying the function makeTxDbFromGFF from the Bioconductor R package GenomicFeatures

[43] we developed a package for each microalga storing gene features annotation available from the previously downloaded and processed GTF files. These packages are central to carry out analysis over genomic loci obtained, for example, using MARACAS and ChIP-seq data to study the genome wide occupation of specific transcription factors or histone modifications. These genomic and functional annotation packages will enable the microalgae research community to perform omics analysis independently from the tools available in ALGAEFUN with MARACAS. The interactive interface of ALGAEFUN was developed using the R package Shiny [25]. In turn, ALGAEFUN functionalities were implemented based on our annotation packages described above. The R Bioconductor packages clusterProfiler [44] and pathview [45] are used for functional enrichment analysis over gene sets. Whereas the annotation and visualization of genomic loci is performed in ALGAEFUN with the R Bioconductor packages ChIPseeker [46] and ChIPpeakAnno [47].

Results and Discussion

Our web-based applications ALGAEFUN (MicroALGAE FUNctional annotation tool) with MARACAS (MicroAlgae RnA-seq and Chip-seq Analysis) seek to become enabling tools that would promote molecular systems biology studies in microalgae. A wide range of microalgae species relevant both in basic research and biotechnological applications are supported. Moreover, our tools are easily extendable to include new microalgae species as their genome sequences and functional annotations become available. Analysis combining the different functionalities implemented in our tools will allow researchers to extract relevant information in the form of functional annotation enrichment analysis over gene sets and genomic loci starting from raw high-throughput sequencing data. ALGAEFUN with MARACAS has been applied recently to determine the molecular systems underpinning astaxanthin accumulation in the microalgae of industrial interest *Haematococcus lacustris* [21].

205 Next, we use two case studies starting from RNA-seq and ChIP-seq raw sequencing data
206 respectively to describe the user interface and discuss the intended uses and benefits of applying
207 ALGAEFUN with MARACAS in microalgae research.

208

209 Case Study 1. From RNA-seq raw sequencing data to biological processes and pathways.

210 A detailed description of the steps to install and execute MARACAS is provided at the
211 corresponding Github repository [48] and the video tutorials available on our webpage [49]. Using
212 the automatic pipeline maracas-rna-seq, users can process raw high-throughput sequencing data in
213 fastq format available either locally in their computers or from a data base such as Gene Expression
214 Omnibus [50] or Sequence Read Archive [51]. This pipeline can be executed either in sequential
215 mode or in distributed/parallel mode on a computational cluster managed by the job scheduling
216 system SLURM. The parameter settings need to be provided in a single text file that constitutes the
217 only input to this pipeline. Besides specifying the microalga of interest, the execution mode and
218 fastq files locations or accession numbers, these parameters describe the experimental design,
219 control and experimental samples, as well as the fold-change and significance level thresholds for
220 the identification of differentially expressed genes. Reports in html and pdf format are generated
221 containing information regarding sequence quality analysis, mapping process and differential gene
222 expression.

223 In order to illustrate this pipeline, we re-analysed RNA-seq data studying the response of the
224 mamiellophyceae microalgae *Ostreococcus tauri* to iron starvation [20]. The parameter file to
225 reproduce this analysis is provided within the MARACAS distribution bundle. The reports
226 produced by MARACAS described all samples as of high quality and notified no problem during
227 read mapping to the reference genome with mapping rates greater than 94%. Scatter plots
228 comparing gene expression between samples are also produced in the MARACAS report. In this
229 case, high Pearson correlations greater than 98% were identified between replicates of the same

230 condition. Accordingly, the automatically performed Principal Components Analysis in MARACAS
 231 identified two clearly separated clusters constituted by the control and iron starvation samples, Fig
 232 2. A volcano plot is used in the report to represent the 45 repressed genes and 554 activated genes
 233 identified with a fold-change threshold of two and a q-value threshold of 0.01, Fig 2. These lists of
 234 genes can be then inputted into ALGAEFUN to determine significantly overrepresented biological
 235 processes or pathways affected during iron starvation in *Ostreococcus*.
 236 The graphical interface in ALGAEFUN allows users to explore the different functionalities from a
 237 navigation side panel, Fig. 2(C). For this case study we selected “Gene Set Functional Analysis”.
 238 Next, the microalga of interest needs to be chosen from a dropdown menu. Here, we chose
 239 *Ostreococcus tauri*. Users need to specify whether GO term and/or KEGG pathway enrichment
 240 analysis are to be performed at a selected significance level. The set of genes to analyse can be
 241 specified through a text box or by uploading the corresponding file. As background gene set, users
 242 can also choose between the entire microalga genome selected by default or provide a custom gene
 243 set. ALGAEFUN provides gene set examples for each microalga so that users can explore our tool
 244 and check the required gene id format. These examples can be accessed and inputted in the
 245 corresponding text box by clicking on the Example button. These examples were generated during
 246 the testing of MARACAS using previously published RNA-seq data sets and have in turn been used
 247 in the testing and validation of ALGAEFUN. In this case study we decided to carry out a GO term
 248 and KEGG pathway enrichment analysis over the set of 554 activated genes under iron starvation
 249 using the default background gene set for *Ostreococcus*.
 250

Figure 2. RNA-seq MARACAS results and Graphical Interface in AlgaeFUN to perform Gene Set Functional Enrichment Analysis. (A) Volcano plot automatically generated by MARACAS marking activated genes in red and repressed genes in blue. (B) Principal Components Analysis automatically generated by MARACAS clearly separating samples into two different clusters depending on the analysed conditions. (C) AlgaeFUN Graphical Interface for Functional Enrichment Analysis, navigation side panel for selecting the different functionalities (blue box); dropdown menu for microalgae selection and sidebar panel with parameter settings for GO and/or KEGG pathways enrichment analysis (red box); text box for gene set input and gene background selections (yellow box).

The outputs of the functional enrichment analysis are presented in the graphical interface in different tabs. Downloadable tables are generated consisting of columns with GO/KEGG term identifiers, human readable descriptions, p-values, q-values, enrichment values and the list of genes associated in the input set with the corresponding GO/KEGG term. Gene names can be clicked to access their annotation from different data bases. Furthermore, ALGAEFUN also generates several graphs that represent the GO/KEGG term enrichment that can be visualized and downloaded from the different tabs such as acyclic graphs, barplots, dotplots, enrichment maps, gene-concept networks and KEGG pathway maps. Our results were in agreement with the published ones [20]

identifying ribosome biogenesis and DNA metabolic process as key biological processes
overrepresented in the set of activated genes.

Case Study 2. From ChIP-seq raw sequencing data to marked genes.

Similar to the previous case study, MARACAS contains an automatic pipeline that can be executed in sequential or parallel mode, maracas-chip-seq. This pipeline starts from raw high-throughput sequencing data in fastq format generated in ChIP-seq experiments and produces lists of genomic loci occupied by the histone modification or bound by the transcription factor under study. For this type of analysis, a parameter file needs to be provided specifying the microalga of interest, execution mode, the fastq files locations or accession numbers, whether an input or mock control sample is included and whether the data has been generated for a histone modification or transcription factor. The results of the pipeline include reports in html and pdf format with information regarding sequence quality analysis, mapping process and identification of the genomic loci occupied by the histone modification or bound by the transcription factor under study. The coordinates of these genomic loci are generated in BED (Browser Extensible Data) format and the genome wide mapping signal is produced in a BigWig (Big Wiggle) file. These formats are supported by ALGAEFUN so that these files can be uploaded directly to proceed with the analysis. To illustrate this pipeline, we re-analysed ChIP-seq data studying the genome wide distribution of the histone modification H3K4me3 in the chlorophyceae microalgae *Chlamydomonas reinhardtii* [24] using the parameter file provided within the MARACAS distribution bundle. The reports produced by MARACAS did not inform of any issue during the processing, leading to the generation of the genomic loci in BED format and the mapping signal in BigWig format.

Figure 3. AlgaeFUN Graphical Interface and Results for Genomic Loci Annotation generated from ChIP-seq Analysis. (A) Navigation side panel for selecting the different functionalities (blue box); dropdown menu for microalgae selection and sidebar panel with parameter settings for identifying genes associated with the analysed genomic loci (red box); text box and upload buttons for genomic loci and signal level (yellow box). (B) Visualization of signal distribution over a specific gene and identified peaks in MARACAS. (C) Average signal level around the TSS (Transcription Start Site) for the identified target genes generated as one of the output graphs in AlgaeFUN. (D) Average signal level around the TES (Transcription End Site) for the identified target genes generated as one of the output graphs in AlgaeFUN.

The graphical interface in ALGAEFUN allows users to input these files to study the genome-wide distribution of specific transcription factors or histone modifications like the H3K4me3 in this case. First, users need to select “Genomic Loci Functional Analysis” from the navigation side panel, Fig.

305 3(A). Next the microalga of interest has to be chosen from a dropdown menu, *Chlamydomonas*
306 *reinhardtii* in this example. Users must also set the distance around the transcription start site (TSS)
307 defining gene promoters as well as the gene features that a genomic locus needs to overlap to
308 consider the corresponding gene a target; promoter, 5',3'- UTR (5',3'- Untranslated Regions), Exon
309 and Intron, Fig 3(A). The set of genomic loci must be specified by either uploading a BED file or
310 by pasting them in the corresponding text box. Optionally, the genome wide mapping signal can be
311 uploaded in BigWig format, Fig 3(A).

312 For some microalgae an example consisting of a list of genomic loci can be accessed and inputted
313 in the text box by clicking on the Example button. This would allow users to explore the type of
314 results generated by ALGAEFUN when analysing the outcomes of a ChIP-seq experiment. These
315 examples were generated during the testing of MARACAS using previously published ChIP-seq
316 data sets and have in turn been used in the testing and validation of ALGAEFUN.

317 In this case study we uploaded to ALGAEFUN the 12,814 genomic loci identified by MARACAS
318 as significantly occupied by H3K4me3 in the *Chlamydomonas* genome under standard growth
319 conditions and the corresponding genome wide mapping signal file in BigWig format. We
320 considered as gene promoter the region two kilobases around the TSS and selected all the gene
321 features to determine the H3K4me3 marked genes. The outputs are presented in the graphical
322 interface in different tabs. A downloadable table with the marked genes and their available
323 annotation is generated. This gene list can in turn be analysed by ALGAEFUN to perform a GO
324 term and/or pathways enrichment analysis. We identified 11,558 H3K4me3 marked genes. Graphs
325 representing the distribution of the genomic loci overlapping different gene features and the
326 distance distribution upstream and downstream from genes TSS are also represented. In agreement
327 with previously published results, we found that 90.75% of the genomic loci occupied by H3K4me3
328 are located at gene promoters in *Chlamydomonas*. As in this case study, when a BigWig file with
329 the genome wide mapping signal is provided, specific marked genes can be selected to visualize the

330 signal profile over their gene bodies and promoters. A gene example presenting two H3K4me3
 331 peaks on its promoter is depicted to illustrate this functionality in Fig 3 (B). Moreover, DNA motifs
 332 recognized by specific transcription factors and regulators in photosynthetic organisms can be
 333 identified in the promoter of the selected gene. Finally, a visualization of the average level of signal
 334 around Transcriptional Start Site (TSS) and Transcriptional End Site (TES) across all marked genes
 335 is generated. For the case of H3K4me3 in *Chlamydomonas* we obtained further evidence showing
 336 that this epigenetic mark specifically and exclusively locates at the TSS of marked genes and not at
 337 the TES, Fig 3(C,D).

338 As described above, ALGAEFUN with MARACAS constitutes one of the first steps that has been
 339 taken for the development of tools that would enable the microalgae research community to exploit
 340 high throughput next generation sequencing data by applying systems biology techniques. For the
 341 model microalgae *Chlamydomonas reinhardtii*, researchers can find several online tools to
 342 functionally annotate set of genes, such as Algal Functional Annotation Tool [52] and ChlamyNET
 343 [53]. Only the online tool AgriGO [54] offers the possibility of analysing a restrictive number of
 344 different microalgae species beyond *Chlamydomonas*. However, these tools are only applicable to
 345 perform functional annotation of gene sets based exclusively on Gene Ontology (GO) enrichment
 346 analysis. The identification of significantly enriched Kyoto Encyclopedia of Genes and Genomes
 347 (KEGG) pathways in the inputted sets of genes is not supported. Moreover, none of these tools can
 348 be used to process high-throughput sequencing raw data from RNA-seq or ChIP-seq experiments,
 349 or functionally annotate genomic loci obtained from a ChIP-seq analysis. In this respect,
 350 ALGAEFUN with MARACAS improves and implements several novel functionalities of similar
 351 already existing software tools, Table 2.

352

353 **Table 2.** Comparison between ALGAEFUN with MARACAS and other functional enrichment
 354 analysis tools.

	Algal Functional Annotation Tool	AgriGO	ChlamyNET	ALGAEFUN with MARACAS
Gene Sets as Input	YES	YES	YES	YES
Genomic Loci as Input	NO	NO	NO	YES
GO enrichment	YES	YES	YES	YES
KEGG pathways enrichment	YES	NO	NO	YES
Several Microalgae	NO	YES	NO	YES

Conclusion

The main contributions of the tool introduced herein, ALGAEFUN with MARACAS, are enumerated below:

1. ALGAEFUN with MARACAS provides a platform specifically designed for microalgae to analyse raw high-throughput sequencing data from RNA-seq and ChIP-seq studies and functionally annotate the resulting genes sets and/or genomic loci. Our goal consists of developing freely available and easy to use tools that would enable the microalgae research community to perform molecular systems biology analysis.
2. In order to interpret the biological relevance of the gene sets and genomic loci obtained from RNA-seq and ChIP-seq data analysis ALGAEFUN with MARACAS provides simple and informative graphical representations of the GO functional and KEGG pathways enrichment results.
3. Most similar annotation tools only support functional enrichment analysis for gene sets and are restricted to the model microalgae *Chlamydomonas*. In contrast, our annotation tool ALGAEFUN also supports results obtained from ChIP-seq analysis and a wide range of different microalgae species.

372 **Availability and requirements**

373 Project name: AlgaeFUN with MARACAS

374 Project home page: <https://greennetwork.us.es/AlgaeFUN/>

375 Operating system(s): Platform independent

376 Programming language: Bash, R, shiny

377 Other requirements: none

378 License: GPL-3.0 License

379

380 **Availability of data and materials.**

381 The code for ALGAEFUN with MARACAS together with the R scripts used in data processing are
382 publicly available at their respective GitHub repositories from the following links:

383 <https://github.com/fran-romero-campero/ALGAEFUN> and [https://github.com/fran-romero-](https://github.com/fran-romero-campero/MARACAS)

384 [campero/MARACAS](https://github.com/fran-romero-campero/MARACAS). The *Ostreococcus tauri* RNA-seq data set used in the first case study is

385 freely available on the SRA database identified with accession number SRP066656. The

386 *Chlamydomonas reinhardtii* ChIP-seq data set used in the second case study is freely available on

387 the GEO database identified with the accession number GSE59629.

388

389 **Abbreviations**

390 AlgaeFUN (MicroAlgae Functional Annotation Tool)

391 BED (Browser Extensible Data)

392 ChIP-seq (Chromatin ImmunoPrecipitation sequencing)

393 EC (Enzyme Commission)

394 FDR (False Discovery Rate)

395 GO (Gene Ontology)

396 HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts 2)

397 HMMER (Biological sequence analysis using profile Hidden Markov Models)

398 KAAS (KEGG Automatic Annotation Server)

399 KEGG (Kyoto Encyclopedia of Genes and Genomes)

400 KOG (Eukaryotic Orthologous Groups)

401 LIMMA (Linear Models for Microarray Analysis)

402 MACS2 (Model-based Analysis of ChIP-seq 2)

403 MARACAS (MicroAlgae RnA-seq and Chip-seq Analysis)

404 RNA-seq (RNA sequencing)

405 PANTHER (Protein Analysis Through Evolutionary Relationships)

406 PFAM (Protein Family)

407 SLURM (Simple Linux Utility for Resource Management)

408 TES (Transcription End Site)

409 TSS (Transcription Start Site)

410 UTR (Untranslated Region)

411

412 **References**

413 1. Chapman RL. Algae: the world’s most important “plants”—an introduction. Mitig Adapt Strateg

414 Glob Change 2013;18:5-12. <https://doi.org/10.1007/s11027-010-9255-9>

415 2. Chen H, Li T, Wang Q. Ten years of algal biofuel and bioproducts: gains and pains. Planta

416 2019;249:195-219. <https://doi.org/10.1007/s00425-018-3066-8>

417 3. Lee SM, Ryu CM. Algae as New Kids in the Beneficial Plant Microbiome. Front Plant Sci

418 2021;12:91. <https://doi.org/10.3389/fpls.2021.599742>

419 4. Shahid A, Malik S, Zhu H, Xu J, Nawaz MZ, Nawaz S, Alam MA, Mehmood MA. Cultivating

420 microalgae in wastewater for biomass production, pollutant removal, and atmospheric carbon

- 421 mitigation; a review. *Sci Total Environ* 2020;704:135303.
- 422 <https://doi.org/10.1016/j.scitotenv.2019.135303>
- 423 5. Al Jabri H, Taleb A, Touchard R, Saadaoui I, Goetz V, Pruvost J. Cultivating Microalgae in
- 424 Desert Conditions: Evaluation of the Effect of Light-Temperature Summer Conditions on the
- 425 Growth and Metabolism of *Nannochloropsis* QU130. *Appl Sci* 2021;11:3799.
- 426 <https://doi.org/10.3390/app11093799>
- 427 6. Patil PP, Vass I, Kodru S, Szabó M. A multi-parametric screening platform for photosynthetic
- 428 trait characterization of microalgae and cyanobacteria under inorganic carbon limitation. *PLoS One*
- 429 2020;15:e0236188. <https://doi.org/10.1371/journal.pone.0236188>
- 430 7. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The
- 431 *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*
- 432 2007;318:245-50. <https://doi.org/10.1126/science.1143609>
- 433 8. Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, et al. Genomic analysis of
- 434 organismal complexity in the multicellular green alga *Volvox carteri*. *Science* 2010;329:223-6
- 435 <https://doi.org/10.1126/science.1188800>
- 436 9. Roth MS, Cokus SJ, Gallaher SD, Walter A, Lopez D, Erickson E, et al. Chromosome-level
- 437 genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates
- 438 astaxanthin production. *Proc Natl Acad Sci USA* 2017;114:E4296-E4305
- 439 <https://doi.org/10.1073/pnas.1619928114>
- 440 10. Polle JEW, Barry K, Cushman J, Schmutz J, Tran D, Hathwaik LT, et al. Draft Nuclear Genome
- 441 Sequence of the Halophilic and Beta-Carotene-Accumulating Green Alga *Dunaliella salina* Strain
- 442 CCAP19/18. *Genome announc* 2017;5:e01105-17 <https://doi.org/10.1128/genomeA.01105-17>
- 443 11. Morimoto D, Yoshida T, Sawayama S. Draft Genome Sequence of the Astaxanthin-Producing
- 444 Microalga *Haematococcus lacustris* Strain NIES-144. *Microbiol Resour Announc* 2020;9:e00128-
- 445 20 <https://doi.org/10.1128/MRA.00128-20>

- 446 12. Blanc G, Agarkova I, Grimwood J, Kuo A, Brueggeman A, Dunigan DD, et al. The genome of
447 the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome*
448 *Biol* 2012;13:R39 <https://doi.org/10.1186/gb-2012-13-5-r39>
- 449 13. Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, Putnam N, et al. The tiny eukaryote
450 *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci*
451 *USA* 2007;104:7705-10 <https://doi.org/10.1073/pnas.0611046104>
- 452 14. Moreau H, Verhelst B, Couloux A, Derelle E, Rombauts S, Grimsley N, et al. Gene
453 functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the
454 base of the green lineage. *Genome Biol* 2012;13:R74 <https://doi.org/10.1186/gb-2012-13-8-r74>
- 455 15. Worden AZ, Lee JH, Mock T, Rouzé P, Simmons MP, Aerts AL, et al. Green evolution and
456 dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*
457 2009;324:268-72 <https://doi.org/10.1126/science.1167222>
- 458 16. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, et al. The *Phaeodactylum*
459 genome reveals the evolutionary history of diatom genomes. *Nature* 2008;456:239-44
460 <https://doi.org/10.1038/nature07410>
- 461 17. Corteggiani Carpinelli E, Telatin A, Vitulo N, Forcato C, D'Angelo M, Schiavon R, et al.
462 Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in
463 nitrogen depletion. *Mol Plant* 2014;7:323-35 <https://doi.org/10.1093/mp/sst120>
- 464 18. Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, et al. *Klebsormidium*
465 *flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nat Commun*
466 2014;5:3978 <https://doi.org/10.1038/ncomms4978>
- 467 19. Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, et al. Genomes of subaerial
468 *Zygnematophyceae* provide insights into land plant evolution. *Cell* 2019;179:1057-1067
469 <https://doi.org/10.1016/j.cell.2019.10.019>

- 470 20. Lelandais G, Scheiber I, Paz-Yepes J, Lozano JC, Botbol H, Pilatova J, et al. *Ostreococcus*
471 *tauri* is a new model green alga for studying iron metabolism in eukaryotic phytoplankton. *BMC*
472 *Genomics* 2016;17:319 <https://doi.org/10.1186/s12864-016-2666-6>
- 473 21. Hoys C, Romero-Losada AB, Del Río E, Guerrero MG, Romero-Campero FJ, García-González
474 M. Unveiling the underlying molecular basis of astaxanthin accumulation in *Haematococcus*
475 through integrative metabolomic-transcriptomic analysis. *Bioresour Technol* 2021;332:125150
476 <https://doi.org/10.1016/j.biortech.2021.125150>
- 477 22. Monte I, Kneeshaw S, Franco-Zorrilla JM, Chini A, Zamarreño AM, García-Mina JM, Solano
478 R. An Ancient COI1-Independent Function for Reactive Electrophilic Oxylinins in
479 Thermotolerance. *Curr Biol* 2020;30:962-971 <https://doi.org/10.1016/j.cub.2020.01.023>
- 480 23. Zhao X, Rastogi A, Deton-Cabanillas AF, Mohamed OA, Cantrel C, Lombard B, et al. Genome
481 wide natural variation of H3K27me3 selectively marks genes predicted to be important for cell
482 differentiation in *Phaeodactylum tricornutum*. *New Phytol* 2021;229:3208-3220
483 <https://doi.org/10.1111/nph.17129>
- 484 24. Ngan CY, Wong CH, Choi C, Yoshinaga Y, Louie K, Jia J, et al. Lineage-specific chromatin
485 signatures reveal a regulator of lipid metabolism in microalgae. *Nat Plants* 2015;1:15107
486 <https://doi.org/10.1038/nplants.2015.107>
- 487 25. Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A,
488 Borges B. shiny: Web Application Framework for R. R package version 1.6.0. 2021
489 <http://shiny.rstudio.com/>
- 490 26. Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology
491 Resource: 20 years and still GOing strong. *Nucleic Acids Res* 2019;47:D330-8
492 <https://doi.org/10.1093/nar/gky1055>
- 493 27. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource

- 494 for gene and protein annotation. Nucleic Acids Res 2016;44:D457-62
- 495 <https://doi.org/10.1093/nar/gkv1070>
- 496 28. Ensembl Protists release 51. EMBL-EBI. 2021. <https://protists.ensembl.org> Accessed August
- 497 2021.
- 498 29. Grigoriev IV, Hayes RD, Calhoun S, Kamel B, Wang A, Ahrendt S, Dusheyko S, Nikitin R,
- 499 Mondo SJ, Salamov A, Shabalov I, Kuo A. PhycoCosm, a comparative algal genomics resource,
- 500 Nucleic Acids Res 2021;49:D1004–D1011 <https://doi.org/10.1093/nar/gkaa898>
- 501 30. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten
- 502 U, Putnam N, Rokhsar DS. Phytozome: a comparative platform for green plant genomics, Nucleic
- 503 Acids Res 2012;40:D1178-D1186 <https://doi.org/10.1093/nar/gkr944>
- 504 31. Genomes – NCBI Datasets Beta Accession Number GCA_011766145.1. 2021.
- 505 <https://www.ncbi.nlm.nih.gov/datasets/genomes/> Accessed August 2021.
- 506 32. Figshare repository for subaerial Zygnematophyceae. 2019. [https://figshare.com/articles/dataset/](https://figshare.com/articles/dataset/Genomes_of_subaerial_Zygnematophyceae_provide_insights_into_land_plant_evolution/9911876/1)
- 507 [Genomes of subaerial Zygnematophyceae provide insights into land plant evolution/](https://figshare.com/articles/dataset/Genomes_of_subaerial_Zygnematophyceae_provide_insights_into_land_plant_evolution/9911876/1)
- 508 [9911876/1](https://figshare.com/articles/dataset/Genomes_of_subaerial_Zygnematophyceae_provide_insights_into_land_plant_evolution/9911876/1) Accessed August 2021.
- 509 33. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of
- 510 RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 2016;11:1650-1667 [https://](https://doi.org/10.1038/nprot.2016.095)
- 511 doi.org/10.1038/nprot.2016.095
- 512 34. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-
- 513 359 <https://doi.org/10.1038/nmeth.1923>
- 514 35. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential
- 515 expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47
- 516 <https://doi.org/10.1093/nar/gkv007>
- 517 36. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson D, Bernstein BE, Nusbaum C, Myers RM,
- 518 Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137

519 <https://doi.org/10.1186/gb-2008-9-9-r137>

520 37. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in Homology Search: HMMER3
521 and Convergent Evolution of Coiled-Coil Regions. Nucleic Acids Res 2013;41:e121 [https://doi.org/](https://doi.org/10.1093/nar/gkt263)
522 [10.1093/nar/gkt263](https://doi.org/10.1093/nar/gkt263)

523 38. Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. KAAS: an automatic genome
524 annotation and pathway reconstruction server. Nucleic Acids Res 2007;35:W182-5
525 <https://doi.org/10.1093/nar/gkm321>

526 39. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a
527 library of protein families and subfamilies indexed by function. Genome Res 2003;13:2129-41
528 <https://doi.org/10.1101/gr.772403>

529 40. Galperin M, Wolf Y, Makarova KS, Vera-Álvarez R, Landsman D, Koonin EV. COG database
530 update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids
531 Res 2021;49:D274-D281 <https://doi.org/10.1093/nar/gkaa1018>

532 41. ALGAEFUN Github repository 2021. <https://github.com/fran-romero-campero/ALGAEFUN>
533 Accessed August 2021.

534 42. Carlson M, Pagès H. AnnotationForge: Tools for building SQLite-based annotation data
535 packages. R package version 1.34.0. 2021 <https://bioconductor.org/packages/AnnotationForge>

536 43. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ.
537 Software for computing and annotating genomic ranges. PloS Comput Biol 2013;9:e1003118
538 <https://doi.org/10.1371/journal.pcbi.1003118>

539 44. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes
540 among gene clusters. OMICS 2012;16:284-7 <https://doi.org/10.1089/omi.2011.0118>

541 45. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration
542 and visualization. Bioinformatics 2013;15:1830-1 <https://doi.org/10.1093/bioinformatics/btt285>

46. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation,
comparison and visualization. *Bioinformatics* 2015;31:2382-3
<https://doi.org/10.1093/bioinformatics/btv145>

47. Zhu LJ. Integrative analysis of ChIP-chip and ChIP-seq dataset. *Methods Mol Biol*
2013;1067:105-24 https://doi.org/10.1007/978-1-62703-607-8_8

48. MARACAS Github repository 2021. <https://github.com/fran-romero-campero/MARACAS>
Accessed August 2021.

49. ALGAEFUN with MARACAS webpage 2021. <https://greennetwork.us.es/AlgaeFUN/> Accessed
August 2021.

50. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO:
archive for functional genomics data sets – update. *Nucleic Acids Res* 2013;41:D991-5.
<https://doi.org/10.1093/nar/gks1193>

51. Leinonen R, Sugarawa H, Shumway M. The Sequence Read Archive. *Nucleic Acids Res*
2011;39:D19-D21. <https://doi.org/10.1093/nar/gkq1019>

52. Lopez D, Casero D, Cokus SJ, Merchant SS, Pellegrini M. Algal Functional Annotation Tool: A
web-based analysis suite to functionally interpret large gene lists using integrated annotation and
expression data. *BMC Bioinformatics* 2011;12:282 <https://doi.org/10.1186/1471-2105-12-282>

53. Romero-Campero FJ, Perez-Hurtado I, Lucas-Reina E, Romero JM, Valverde F. ChlamyNET: A
Chlamydomonas gene co-expression network reveals global properties of the transcriptome and the
early setup of key co-expression patterns in the green lineage. *BMC Genomics* 2016;17:227 <https://doi.org/10.1186/s12864-016-2564-y>

54. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, Xu W, Su Z. AgriGO v2.0: A GO analysis toolkit for
the agricultural community, 2017 update. *Nucleic Acids Res* 2017;45:W122–9
<https://doi.org/10.1093/nar/gkx382>

568 **Acknowledgements**

569 This work was supported by the research project MINOTAUR (BIO2017-84066-R) from the
570 Spanish Ministry of Science and Innovation and Intramural Action of the Spanish National
571 Research Council (201420E035).

572

573 **Funding**

574 This work was supported by the research project MINOTAUR (BIO2017-84066-R) from the
575 Spanish Ministry of Science and Innovation and Intramural Action of the Spanish National
576 Research Council (201420E035).

577

578 **Authors' contributions**

579 ABRL, FJRC, CA and PdlR co-developed all the components of MARACAS, ALGAEFUN and the
580 different annotation R packages. FJRC and MGG selected the microalgae species to be included in
581 ALGAEFUN with MARACAS, designed the application interface and wrote the manuscript. All
582 authors read and approved the manuscript.

583 **Declarations**

584 Ethics approval and consent to participate. Not applicable

585 Consent for publication. Not applicable

586

587 **Competing interests**

588 The authors declare that they have no competing interests

(A) **(B)** **(C)**

- ☐ Home
- ☒ Gene Set Functional Analysis
- ☐ Genomic Loci Functional Analysis
- ☐ MARACAS, MicroAlgae RnA-seq and Chip-seq Analysis
- ☐ Tutorials
- ☐ GitHub repository
- ☐ Citation and Contact

ALGAEFUN with MARACAS

microALGAE FUNctional enrichment tool for MicroAlgae RnA-seq and Chip-seq Analysis

AlgaeFUN allows researchers to perform **functional annotation** over gene sets. **Gene Ontology (GO) enrichment** analysis as well as **KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment** analysis are supported. The gene set of interest can be obtained, for example, as the result of a differential expression analysis carried out using **MARACAS**. See our **video tutorial** for details or follow the next steps to perform your analysis:

1. In the left panel choose your **microalgae** of interest, the type of enrichment analysis to perform and the **p-value threshold**.
2. Insert your **gene set** in the text box or load it from a file using the **Browse ...** button. An example can be loaded by clicking on the **Example** button. Click on **Clear** button to remove the loaded gene set.
3. Users can choose between the default **background** gene provided by AlgaeFUN or a custom one that can be specified.
4. Click on the **Have Fun** button to perform the specified functional enrichment analysis. The results will be shown in the different tabs below.

Choose your favourite microalgae

Ostreococcus tauri

Choose your desirable analysis

- ☒ GO terms enrichment
- ☐ KEGG pathways enrichment analysis
- ☐ Both

Choose gene ontology:

- ☒ Biological process
- ☐ Cellular Component
- ☐ Molecular Function

Which will be your chosen p-value?

0.05

Insert a set of genes

Insert set of genes

Example

Clear

Choose File with Gene Set to Upload

Browse...

No file selected

(A)

ALGAEFUN with MARACAS

microALGAE FUNctional enrichment tool for MicroAlgae RnA-seq and Chip-seq Analysis

- Home
- Gene Set Functional Analysis
- Genomic Loci Functional Analysis
- MARACAS, MicroAlgae RnA-seq and Chip-seq Analysis
- Tutorials
- GitHub repository
- Citation and Contact

AlgaeFUN allows researchers to perform annotation analysis of genomic loci or regions. These are typically generated from ChIP-seq studies of the genome-wide distribution of epigenetic marks or transcription factor binding sites. Our tool MARACAS can be used to perform this type of analysis. The set of marked genes can be obtained as well as the distribution of the genomic loci overlapping specific genes parts. Also, individual marked genes and the average signal level around the TSS (Transcription Start Site) and TES (Transcription End Site) over the complete set of marked genes can be visualized. See our [video tutorial](#) for details or follow the next steps to perform your analysis:

- In the left panel choose your **microalgae** of interest, the **gene promoter length** and the **gene parts** that will be considered when determining the marked genes.
- Insert in the text box your **set of genomic regions** as a table consisting of three tab-separated columns representing the chromosome, the start and end position of the regions. An example can be loaded by clicking on the **Example** button. Click on **Clear** button to remove the loaded gene set. Alternatively, using the **Browse...** button, the genomic regions can be uploaded from a file in BED format as described previously containing at least three columns. This file can be obtained using our tool **MARACAS**.
- Optionally, users can upload the genome wide signal level of an epigenetic mark or transcription factor binding in a BigWig file. This file can be obtained using our tool **MARACAS**.
- Click on the **Have Fun** button to perform the specified analysis. The results will be shown in the different tabs below.

Choose your favourite microalgae

Chlamydomonas reinhardtii

Choose the distance in base pairs around the Transcriptional Start Site defining gene promoters

A gene will be associated to an input genomic locus when it overlaps one of the following gene features:

- ☒ Promoter
- ☒ 5' UTR
- ☒ Exon
- ☒ Intron
- ☒ 3' UTR

Insert a set of genomic regions

Insert set of genomic regions

Example

Clear

Choose File with the Genomic Regions to Upload

Browse...

output_peaks.narrowPeak

Upload complete

Choose BigWig File to Upload for Profile Representations: (Optional)

Browse...

chip_average.bw

Upload complete

(B)

(C)

(D)