# The NanDeSyn Database for *Nannochloropsis* systems and synthetic biology

Yanhai Gong[1,8,+], Nam Kyu Kang[2,17,+], Young Uk Kim[2], Zengbin Wang[1,8], Li Wei[1,8], Yi Xin[1,8], Chen Shen[1,8], Qintao Wang[1,8], Wuxin You[1,3], Jong-Min Lim[4], Seok Won Jeong[5], Youn-Il Park[5], Hee-Mock Oh[6], Yong Keun Chang[2], Kehou Pan[8,9], Eric Poliner[10], Guanping Yang[8,11,12], Yonghua Li-Beisson[13], Yantao Li[16], Qiang Hu[15], Ansgar Poetsch[3,8,11], Eva M. Farre[14], Won-Joong Jeong[4], Byeong-ryool Jeong[1,7,*], Jian Xu[1,8,*]

[1]Single-Cell Center, CAS Key Laboratory of Biofuels and Shandong Key Laboratory of Energy Genetics and Shandong Institute of Energy Research, Qingdao Institute of BioEnergy and Bioprocess Technology (QIBEBT), Chinese Academy of Sciences, Qingdao, Shandong 266101, and University of Chinese Academy of Sciences, Beijing 100049, China

[2]Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea

[3]Department of Plant Biochemistry, Ruhr University Bochum, Bochum, Germany

[4]Plant Systems Engineering Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 34141, Korea

[5]Department of Biological Sciences, Chungnam National University, Daejeon 305-764, Korea

[6]Cell Factory Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Korea

[7]School of Energy and Chemical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Korea

21 [8]Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine
22 Science and Technology, Qingdao 266237, China
23 [9]Laboratory of Applied Microalgae, College of Fisheries, Ocean University of China, Qingdao,
24 266003, China
25 [10]MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan,
26 USA
27 [11]College of Marine Life Sciences, Ocean University of China, Qingdao, Shandong, 266003,
28 China
29 [12]Institutes of Evolution and Marine Biodiversity, Ocean University of China, Qingdao,
30 Shandong, 266003, China
31 [13]Aix Marseille Univ, CEA, CNRS, Institut de Biosciences et Biotechnologies Aix-Marseille,
32 CEA Cadarache, 13108, Saint Paul-Lez-Durance, France
33 [14]Department of Plant Biology, Michigan State University, East Lansing, MI, 48824, USA
34 [15]Center for Microalgal Biotechnology and Biofuels, Institute of Hydrobiology, Chinese
35 Academy of Sciences, Wuhan, 430072 China
36 [16]Institute of Marine and Environmental Technology, University of Maryland Center for
37 Environmental Science and University of Maryland, Baltimore County, Baltimore, MD 21202,
38 USA
39 [17]Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, IL 61801, USA
40 [+]Yanhai Gong and Nam Kyu Kang contributed equally to this work

41 *Correspondence: Jian Xu (xujian@qibebt.ac.cn) and Byeong-ryool Jeong

42 (bjeong@unist.ac.kr)

43 **Keywords**: Genomic database, functional genomics, industrial oleaginous microalgae,

44 *Nannochloropsis*, Tripal

45 **Running title**: Comprehensive genomic database for *Nannochloropsis*

46

**ABSTRACT**

The unicellular industrial oleaginous microalgae *Nannochloropsis* spp. have emerged as model organisms for microalgal systems and synthetic biology. To facilitate community-based annotation and mining of the rapidly accumulated functional genomics resources, we have initiated an international consortium and present a comprehensive, multi-omics resources named Nannochloropsis Design and Synthesis (NanDeSyn; http://nandesyn.single-cell.cn). Via the Tripal toolkit, it featured user friendly interfaces hosting genomic resources with gene annotations, and transcriptomic and proteomic data for six *Nannochloropsis* species. Toolboxes for search, Blast, synteny view, enrichment analysis, metabolic pathway analysis, genome browser, etc. were also included. In addition, functional annotations of genes were also provided based on phenotypes of mutants and relevant bibliography. Furthermore, epigenomic resources were also incorporated, especially for small RNA-Seq including miRNAs and circRNAs. Such comprehensive and integrated landscapes of *Nannochloropsis* genomics and epigenomics will promote and accelerate community efforts in systems and synthetic biology of the industrially important microalgae.

63 **INTRODUCTION**

64 *Nannochloropsis* spp. are a group of unicellular photosynthetic microalgae in the
65 heterokont group (Fawley and Fawley, 2007). Distributed widely in the marine environment as
66 well as in fresh and brackish waters, they are of industrial interest due to their robust growth
67 tolerating a broad range of environmental and culture conditions, and high lipid contents
68 including value-added polyunsaturated fatty acids (PUFAs) such as eicosapentaenoic acid (EPA)
69 (Wang *et al.*, 2014a). Therefore *Nannochloropsis* spp. have already been used as feeds for
70 aquaculture and other purposes at the industrial scale (Freire *et al.*, 2016, Letsiou *et al.*, 2017).
71 Moreover, due to advances in biochemical analyses and molecular tool development these
72 species have been proposed to be suitable model organisms for studying microalgal lipid
73 metabolism and as a chassis for synthetic biology (Liu *et al.*, 2017, Poliner *et al.*, 2018a).

74 *Nannochloropsis* spp. have small haploid genomes (~30Mb, distributed among 30
75 chromosomes), as reported for *N. oceanica* (LAMB0001 (Pan *et al.*, 2011), CCMP1779 (Vieler
76 *et al.*, 2012), IMET1 (Wang *et al.*, 2014a), LAMB2011 (Guo *et al.*, 2019), KB1 and BR2
77 (Brown *et al.*, 2019)), *N. gaditana* (B-31 (Corteggiani Carpinelli *et al.*, 2014), CCMP526
78 (Radakovits *et al.*, 2012) and CCMP1894 (Schwartz *et al.*, 2018)), *N. salina* (CCMP1776
79 (Wang *et al.*, 2014a, Ohan *et al.*, 2019) and CCMP537 (Wang *et al.*, 2014a)), *N. granulate*
80 CCMP529 (Wang *et al.*, 2014a), *N. limnetica* CCMP505 (Wang *et al.*, 2014a) and *N. oculata*
81 CCMP525 (Wang *et al.*, 2014a). The plastid and mitochondrial genomes of five
82 *Nannochloropsis* species were used to produce a pangenome of each organelle (Wei *et al.*,
83 2013). Additional functional resources have also been recently developed, such as

84 transcriptomic data for *N. oceanica* (including IMET1 (Li *et al.*, 2014, Wei *et al.*, 2019) and
85 CCMP1779 (Vieler *et al.*, 2012, Poliner *et al.*, 2015, Mühlroth *et al.*, 2017)), *N. gaditana* (B-31
86 (Corteggiani Carpinelli *et al.*, 2014) and CCMP526 (Radakovits *et al.*, 2012)), *N. salina*
87 (CCMP1776 (Jeong *et al.*, 2017)), and proteomic data for *N. oceanica* IMET1 (Chen *et al.*,
88 2019, Wei *et al.*, 2019) and *N. gaditana* (Fernández-Acero *et al.*, 2019). These resources are
89 currently being exploited to dissect complex metabolic pathways for identification of targets for
90 genetic engineering.

91 Various genetic tools have been developed for *Nannochloropsis* (Poliner *et al.*, 2018a).
92 These tools include overexpression (Kang *et al.*, 2015, Chen *et al.*, 2017, Zienkiewicz *et al.*,
93 2017, Poliner *et al.*, 2018b, Kang *et al.*, 2019), RNAi (Ma *et al.*, 2017, Wei *et al.*, 2017),
94 multigene expression enabled by bidirectional promoters and ribosome skipping 2A sequences
95 (Kilian *et al.*, 2011, Jinkerson *et al.*, 2013, Moog *et al.*, 2015, Nobusawa *et al.*, 2017, Poliner *et*
96 *al.*, 2018b), unrestrained markerless trait stacking through combined genome editing and
97 marker recycling (Verruto *et al.*, 2018) or the use of episomes (Poliner *et al.*, 2018c), as well as
98 gene targeting via homologous recombination (Kilian *et al.*, 2011, Dolch *et al.*, 2017, Gee and
99 Niyogi, 2017, Nobusawa *et al.*, 2017). Clustered regularly interspaced short palindromic
100 repeats (CRISPR)-based gene editing techniques have also been successfully employed in
101 targeted gene disruption (Wang *et al.*, 2016, Poliner *et al.*, 2018c, Verruto *et al.*, 2018,
102 Naduthodi *et al.*, 2019), leading to improved lipid production in *Nannochloropsis* spp. (Ajjawi
103 *et al.*, 2017). Notably, the development of Single-cell Raman Spectrum based approaches that
104 characterize the profiles of energy-storage molecules in *N. oceanica*, the "ramanome", has

105 accelerated the time-consuming microalgal phenotyping processes (Ji *et al.*, 2014, Wang *et al.*,
106 2014b, He *et al.*, 2017, Llansola-Portoles *et al.*, 2017), by providing single-cell-resolution
107 phenomes in a label-free and non-invasive manner (He *et al.*, 2019).

108 These extensive metabolic and genetic engineering efforts have entitled *Nannochloropsis*
109 as an excellent model system intended for industrial applications in aquaculture and
110 nutraceutical fields as well as biofuel industries (Poliner *et al.*, 2018a). In fact, such efforts have
111 been employed to enhance triacylglycerol (TAG) synthesis (Li *et al.*, 2014, Wang *et al.*, 2014a,
112 Ajjawi *et al.*, 2017, Xin *et al.*, 2017, Xin *et al.*, 2019), and characterize carbon partitioning (Jia
113 *et al.*, 2015, Poliner *et al.*, 2015), sterol metabolism (Lu *et al.*, 2014b), phytohormone function
114 (Lu *et al.*, 2014a, Lu and Xu, 2015), photosynthetic pigments (Koh *et al.*, 2019) and
115 transcriptional-factor regulation (Hu *et al.*, 2014, Kang *et al.*, 2015, Kang *et al.*, 2017, Kang *et*
116 *al.*, 2019).

117 In addition to genomic resources, comprehensive understanding of biological systems
118 requires epigenetic components that are intertwined with genetic regulation. Epigenetic
119 regulation is accomplished mainly by small RNAs and chromatin modifications, which have
120 been characterized in animals and plants (Singh *et al.*, 2019). For example, sirtuins are histone
121 deacetylases, and act as metabolic sensor regulating carbon metabolism via AMPK, which are
122 involved in circadian regulation and aging in animals (Griswold *et al.*, 2008, Furlan *et al.*,
123 2019). Lipid metabolism in animals is also regulated by epigenetic modifications, including
124 histone acetylation/deacetylation (Ferrari *et al.*, 2012, Shimazu *et al.*, 2013). The
125 RNA-mediated epigenetic regulation, for example using microRNAs (miRNAs), can also

126 modulate stress responses and metabolic processes including lipids metabolism (Navarro *et al.*,
127 2006, van Rooij *et al.*, 2007, Thomou *et al.*, 2017). Such epigenetic regulation has been
128 reported in green algae, including histone and DNA modifications and post-transcriptional gene
129 silencing in *Chlamydomonas reinhardtii* (Wu-Scharf *et al.*, 2000, Jeong *et al.*, 2002,
130 Casas-Mollano *et al.*, 2008, Cerutti *et al.*, 2011, Xue *et al.*, 2019). Epigenetic regulation is
131 essential for metabolic regulation including lipid accumulation under stress conditions in
132 *Chlamydomonas*, in which histone modifications are intricately involved in transcriptional
133 regulation by phosphorus stress response 1 (PSR1) (Xue *et al.*, 2019). These epigenetic
134 components have also been described in *Nannochloropsis* (Wei and Xu, 2018), but have not
135 been tested for their functions.

136 In light of the rapidly expanding genomic and epigenomic resources for *Nannochloropsis*,
137 a comprehensive database is needed to store, mine, analyze, and disseminate large-scale
138 multi-omics datasets and to provide a portal for the relevant research community. There have
139 been a few databases, e.g., *Nannochloropsis gaditana* genome portal
140 (http://www.nannochloropsis.org), and JGI MycoCosm
141 (https://genome.jgi.doe.gov/Nanoce1779_2/Nanoce1779_2.home.html), which are however
142 limited to a specific *Nannochloropsis* species or a specific data type, and lack functions for
143 comparative and functional genomics. In addition, they do not offer comprehensive integration
144 and visualization of different data types. Here, we developed a genus wide *Nannochloropsis*
145 genomics database named Nannochloropsis Design and Synthesis (NanDeSyn;
146 http://nandesyn.single-cell.cn), mainly including data for *N. oceanica*, *N. gaditana* and *N.*

147 *salina*. We also incorporated new genomic data of *N. oceanica* and (epi)genomic data of *N.*
148 *salina* with newly annotated gene models and expression data of mRNAs and small RNAs
149 including microRNAs (miRNAs) and circular RNAs (circRNAs). NanDeSyn was built using
150 Tripal (Cheng *et al.*, 2011), a toolkit for construction of online genomic and genetic databases
151 by integrating the GMOD Chado database schema (Jung *et al.*, 2016). Tripal (Cheng *et al.*,
152 2011) has been used to implement a number of widely used databases such as Genome
153 Database for Rosaceae (Jung *et al.*, 2007, Jung *et al.*, 2014), CottonGen (Yu *et al.*, 2014),
154 Coffee Genome Hub (Dereeper *et al.*, 2015), etc. Furthermore, we have generated
155 comprehensive functional annotations for *Nannochloropsis* gene models, identified synteny
156 blocks and homologous genes among different *Nannochloropsis* spp., incorporated expression
157 profiles based on public and new RNA-Seq and liquid chromatography-mass spectrometry
158 (LC-MS/MS) data, and employed or developed modules in NanDeSyn to analyze and visualize
159 comparative genomics and expression datasets of different *Nannochloropsis* species (**Figure 1**).

160 **DATABASE CONTENTS AND FEATURES**

161 **Genome sequences and gene annotations**

162 Currently NanDeSyn contains reference genome sequences across all *Nannochloropsis*
163 species, including seven *N. oceanica* strains (IMET1, CCMP1779, CCMP531, LAMB2011,
164 LAMB0001, KB1 and BR2), four *N. gaditana* strains (CCMP1894, B-31, CCMP526 and
165 CCMP527), two *N. salina* strains (CCMP1776 and CCMP537), one *N. granulate* strain
166 (CCMP529), one *N. oculata* strain (CCMP525) and one *N. limnetica* strain (CCMP505) (**Table**
167 **S1**). *N. oceanica* IMET1 (version 2) and *N. salina* CCMP1776 (version 2) were first released in

168 NanDeSyn, while the other genome sequences were downloaded from GenBank or JGI
169 Genome Portal. The previous version of *N. oceanica* IMET1 genome (Wang *et al.*, 2014a) was
170 fragmented and included ambiguous bases, while gene structure prediction was based on a
171 limited amount of transcriptomic data (mostly Sanger ESTs). To address these shortfalls, both
172 the assembly and annotation of *N. oceanica* IMET1 genome were improved by exploiting a
173 large amount of PacBio whole genome sequencing (WGS) data and transcriptomic data. Thirty
174 chromosomes were resolved using ultra-high depth of reads from Single-molecule real-time
175 (SMRT) sequencing and chromosome conformation capture-based Hi-C. In addition, a
176 complete re-annotation of the genome was carried out based on extensive RNA-Seq data (from
177 PRJNA182180 and PRJNA241382; **Table S2**).

178 The genome sequence of *N. salina* CCMP1776 was improved from an online version
179 listed in the Greenhouse genome database (https://greenhouse.lanl.gov/greenhouse/organisms/),
180 and then merged with new PacBio sequencing results (details in the website). Gene models
181 were performed using combinations of *de novo* transcriptome assembly, *ab initio* prediction
182 and evidence-based alignment methods. (*i*) We proceeded with *de novo* transcriptome assembly
183 using Cufflink v2.1.1 (Trapnell *et al.*, 2010), Velvet v1.2.10 (Zerbino and Birney, 2008, Schulz
184 *et al.*, 2012) and Trinity v2012-06-08 (Grabherr *et al.*, 2011) to obtain the coding sequence of
185 each gene. (*ii*) Augustus v3.2.2 (Stanke and Morgenstern, 2005) and Genewise v2.4.1 (Birney
186 *et al.*, 2004) were applied to predict draft model of each gene by training coding sequences
187 from *de novo* transcriptome assembly for the *ab initio* prediction. (*iii*) The predicted transcripts
188 sequence resulting from *de novo* transcriptome assembly and the protein sequences

189 downloaded from RefSeq non-redundant proteins sequence (NR) database were aligned on the
190 genome using Tophat2 v2.0.9 (Kim *et al.*, 2013). (*iv*) Gene regions were determined by
191 merging aligned information of predicted transcripts, NR proteins and results of *ab initio*
192 prediction. Frameshift correction was performed to make the identical model compared to NR
193 proteins by in-house scripts. We finally identified 10,952 gene models that included both start
194 and stop codons by selecting representative transcripts form each gene region and retraining
195 them with Augustus. BLASTX (e-value ≤ 1e-10, best hit) was performed on amino acid
196 sequences of NR, UniProtKB, GO, KOG, and KEGG to determine the function of the
197 confirmed gene model.

198 A total of 10333~10952 protein-coding genes were exhaustively predicted from four
199 genomes (*N. oceanica* IMET1 (version 2) and CCMP1779, *N. gaditana* B-31, and *N. salina*
200 CCMP1776). For *N. oceanica* IMET1 (version 2; for the other genomes, the published
201 annotation was adopted), protein sequences of the predicted genes were compared against NR,
202 SwissProt, and Arabidopsis protein databases using the BLAST program with an E-value cutoff
203 of 1e-6. The protein sequences were further compared against the InterPro database using
204 InterProScan (Mitchell *et al.*, 2014) to identify functional protein domains. BLAST results
205 against the NR database and the identified InterPro domains were fed to Blast2GO (Conesa and
206 Götz, 2008) for assigning gene ontology (GO) terms to protein-coding genes. The top BLAST
207 hits (homologs), GO terms and InterPro domains assigned to each of the protein-coding genes
208 were imported into NanDeSyn using Tripal Analysis Extension Modules. Each gene in the

209 database contains a detailed feature page for sequence, annotation, homologs, expression
210 analyses, etc. (**Figure S1**).

211 **Synteny blocks and homologous genes**

212 We have identified synteny blocks and homologous gene pairs within the synteny blocks
213 for pairwise comparisons of the *Nannochloropsis* genomes of *N. oceanica* IMET1, *N. gaditana*
214 B-31, *N. oceanica* CCMP1779 and *N. salina* CCMP1776, for which rich omics resources are
215 present. Protein/mRNA sequences were first searched against each other (pairwise comparisons)
216 using BLASTP/BLASTX with an E-value cutoff of 1e-10 and a maximum of five alignments.
217 Based on the BLAST results, synteny blocks were determined using MCScanX (Wang *et al.*,
218 2012) with default parameters. On average, ~7,264 homologous gene pairs (accounting for ~70%
219 genes) were identified and stored in NanDeSyn, which can be visualized using tools from the
220 'SyntenyViewer' module. To share knowledge between species/strains, a section is also
221 included in the "Relationships" panel of each gene-feature page to display all synteny blocks
222 and homologous gene pairs related to a specific gene (**Figure S1**).

223 **Transcriptome and proteome expression profiles**

224 We collected published *Nannochloropsis* mRNA-Seq datasets from NCBI Sequence Read
225 Archive (SRA) and proteomics datasets from the PRIDE database (**Table S2**). For public
226 mRNA-Seq datasets, a unified pipeline (nf-core/rnaseq v1.2; with modification to limit intron
227 length to 20-3000 bp) was applied to quantify gene expression. Briefly, raw reads were
228 processed to remove adaptor and low-quality sequences using Trim Galore v0.5.0
229 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The remaining high-quality

230 reads were aligned to the corresponding genome using STAR v2.6.1b (Dobin *et al.*, 2013). Raw
231 counts were then derived for each predicted gene model and normalized to TPM (Transcripts
232 Per Kilobase Million) using featureCounts (subread v1.6.2) (Liao *et al.*, 2014) and StringTie
233 v1.3.3 (Pertea *et al.*, 2015), followed by cross-sample TMM (Trimmed Mean of M-values)
234 normalization using scripts from Trinity v2.4.0 (Grabherr *et al.*, 2011).

235 Transcriptomic data for *N. salina* CCMP 1776 that had not been previously published,
236 were based on four independent repeats of RNA-seq for normal, nitrogen (N) limitation and
237 high salt stress conditions (details on the website). The quality of raw reads was evaluated by
238 SolexaQA software v3.1.5 (http://solexaqa.sourceforge.net/) (Cox *et al.*, 2010) and cleaned
239 reads were extracted by removing low-quality (phred < 20) reads from the dataset. The cleaned
240 reads were mapped to the transcripts with the bowtie2 (v2.1.0) software (Langmead *et al.*,
241 2009). The number of mapped reads for each transcript was calculated and normalized with
242 DESeq package (Anders and Huber, 2010) in R. The read counts of all genes were published
243 through the expression menu and genome browse menu on "Nannochloropsis WebDB"
244 (http://web.seeders.co.kr/NSK2019/), and now incorporated into NanDeSyn
245 (http://nandesyn.single-cell.cn/organism/5).

246 For proteomic MS/MS data, proteins were identified by Andromeda search engine
247 embedded in MaxQuant v1.6.5.0 (Tyanova *et al.*, 2016). All datasets were searched against
248 respective protein databases with decoy sequences (revert mode) to control false discovery rate
249 (FDR ≤ 0.01 for both peptide spectra match (PSM) and identified protein). Only tryptic
250 peptides with at least seven amino acids were used for the identification of proteins. No fixed

251 modifications were used during the searches, while oxidation of methionine and N-terminal
252 acetylation were set as variable modifications. Proteins were quantified by label-free
253 quantification (LFQ) method integrated in MaxQuant (Tyanova *et al.*, 2016), and used for
254 quantification of proteins with at least two unique peptides. Then, the LFQ-normalized
255 intensities were log2-transformed with Perseus (Tyanova and Cox, 2018) software. Proteins not
256 quantified in at least half of the samples for each experiment were removed. Finally, missing
257 values were imputed from normal distribution in Perseus (Tyanova and Cox, 2018). Gene
258 expression profiles for a specific gene can be accessed under the 'Expression' section of the
259 gene feature page, where after selecting an mRNA-Seq/proteome project, a histogram showing
260 expression profiles across different experimental conditions is displayed (**Fig. 2**). The 'Analysis:
261 Expression' page (chado/analysis-expression) lists all collected projects with descriptions.

262 **Epigenomic resources including small RNAs**

263 As an initial step for establishing epigenomic resources, we isolated small and large RNAs
264 from the same cultures of *N. salina* CCMP1776 used for gene expression analyses (in
265 triplicates). Small RNA libraries were generated with the Illumina protocols, and reads were
266 trimmed using cutAdapt v2.6 (Martin, 2011) to remove remnants of the 3′-adapter sequence.
267 The trimmed sequences were plotted and validated with FastQC
268 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc), and mapped to the reference
269 genome of CCMP1776 using bowtie2 (Langmead and Salzberg, 2012). We identified miRNAs
270 from mapped reads using miRA (Evers *et al.*, 2015), and their expression was normalized to
271 RPM (Reads Per Million Reads) using featureCounts (subread v1.6.2) (Liao *et al.*, 2014).

272 Differential expression of miRNAs between untreated and N starvation samples were analyzed
273 using DESeq2 (Love *et al.*, 2014). The large RNA counterpart, prepared from the same culture
274 and representing mRNA, were prepared, sequenced, and analyzed as described above.
275 CircRNAs were identified using combinations of tools including find_circ (Memczak *et al.*,
276 2013) and CIRI (Gao *et al.*, 2015). In total, 279 miRNAs and 89 circRNAs were predicted,
277 providing rich targets for genetic manipulations.

278 **Biochemical pathways and metabolic models**

279 We used PathwayTools (Karp *et al.*, 2001) to predict biochemical pathways for
280 *Nannochloropsis* species (*N. oceanica* IMET1, *N. oceanica* CCMP1779 and *N. gaditana* B-31).
281 For each *Nannochloropsis* genome, functional annotations, GO terms, and enzyme commission
282 (EC) numbers were integrated into a file in the PathoLogic format, which was used by
283 PathwayTools (Karp *et al.*, 2001) for pathway prediction. A total of 119-212 biochemical
284 pathways were predicted from each genome (**Table S3**). A *Nannochloropsis* biochemical
285 pathway database (NannoCyc) has been implemented in NanDeSyn using the PathwayTools
286 (Karp *et al.*, 2001) web server. Users can search and browse the predicted pathways/metabolites,
287 as well as perform comparative and omics data analyses through the NannoCyc database.

288 **Mutant collection in NanDeSyn**

289 Extensive literature mining was performed to gather mutant information for
290 *Nannochloropsis* spp., which were then inserted into a custom table with standardized formats
291 (**Table S4**). A dedicated Tripal extension ("Tripal Mutation Viewer") was developed for the
292 loading and displaying genetic information of mutants, together with underlying genes linked

293 to gene features in NanDeSyn. Users can find detailed information of every entry in a
294 searchable list (chado/mutinfos) or add new entries after granting of permissions.

295 **DATABASE FUNCTIONS AND WEB INTERFACE**

296 In NanDeSyn, biologists can locate genes of interest according to functional descriptions or
297 homologous genes. To facilitate the query process of functional annotation data stored in
298 NanDeSyn, we used "Gene Search" interface from "Mainlab Chado Search" module which was
299 based on materialized views to build the search index for different types of annotation data
300 including feature name, GO terms, BLAST and InterPro descriptions, etc. In addition, a global
301 search function, which queries against all the "nodes" stored in the database, was provided
302 under the main menu of NanDeSyn. We also implemented an instance of NCBI's BLAST tool
303 in NanDeSyn to locate similar genes/fragments using the BLAST UI extension module in
304 Tripal (Cheng *et al.*, 2011), which provides downloadable output files in different formats. To
305 view genome synteny and homologous gene pairs between different *Nannochloropsis* species,
306 we used the 'Synteny Viewer' extension module of Tripal (**Figure S2A**), which can not only
307 display synteny blocks for multiple genomes in an intuitive manner, but also connect
308 homologous gene pairs among different *Nannochloropsis* species. Besides, to retrieve
309 sequences, annotations or other features for a list of user-provided genes, the batch query
310 function modified from the 'Sequence Retrieval' page of Tripal has been established in
311 NanDeSyn.

312 For classes of genes (gene sets), enrichment analysis is a powerful method to identify
313 functional categories that are overrepresented in a list of genes, which represent highly affected

314 biological processes under different experimental conditions. In NanDeSyn, we used two
315 extension modules to identify significantly overrepresented GO terms and pathways,
316 respectively. The 'GO tool' module (https://github.com/tripal/tripal_go) determines enriched
317 GO terms, and the 'Pathway tool' module (https://github.com/tripal/tripal_pathway) calculates
318 the significance of enrichment based on the biochemical pathways predicted by PathwayTools
319 (Karp *et al.*, 2001). In addition to viewing the expression profiles of individual genes on the
320 gene feature page (**Figure S1**), the 'Analysis: Expression' module provided the heatmap
321 function to display expression profiles of multiple genes (**Figure S2B**).

322 In NanDeSyn, we also implemented a genome browser using JBrowse (Buels *et al.*, 2016)
323 to display genome sequences, gene models, and expression profile data under different
324 conditions (**Figure S2C**). Each gene was shown as "CanvasFeatures" in specific track of
325 JBrowse (Buels *et al.*, 2016) and linked back to the corresponding feature page, which enabled
326 consecutive examination of genes along chromosomes. The genome browser contained several
327 tracks including expression coverages and read alignments derived from mRNA-Seq and WGS
328 datasets. Other interesting data, such as histone modifications, single-base resolution
329 methylation or variants that are being generated by the community, can be easily added to the
330 genome browser for view in future updates.

331 **CONCLUSIONS AND FUTURE DIRECTIONS**

332 We have developed a comprehensive *Nannochloropsis* database (NanDeSyn) that can serve
333 as a central portal to genomics, transcriptomics, and genetics of all available *Nannochloropsis*
334 species. NanDeSyn contains sequences of genomes, mRNAs, proteins, and comprehensive

335 functional annotations of genes and mRNAs, as well as genome synteny blocks, homologous
336 gene pairs, gene expression profiles, biochemical pathways, and genetic datasets of
337 *Nannochloropsis* species. The database also provides various query, visualization and analysis
338 tools including BLAST, batch query, genome browser, pathway database (NannoCyc), GO
339 terms and pathway enrichment analysis, genome synteny viewer, etc.

340 NanDeSyn will be continuously updated when new genome, mRNA-Seq and genetic
341 datasets of *Nannochloropsis* species become available. We will continue to improve the
342 functionality of the 'Tripal Mutation Viewer' module, and develop novel related data mining
343 and analysis tools which can be used by the Tripal community. We are currently in the process
344 of developing a module for analyses of small RNA (sRNA) datasets, which will be made
345 available in both NanDeSyn and Tripal in the near future. Currently, basic small RNA-seq data
346 are already incorporated in NanDeSyn, and ongoing efforts from the global research
347 community will be incorporated. In addition, genome-wide search for potential off-target sites
348 of all 20-bp Cas9 RNA-guided endonucleases (PAM sequence 5'-NGG-3') have been enabled
349 as a web-based track for *N. oceanica* IMET1 (details in
350 http://nandesyn.single-cell.cn/sgrna_ngg). This serves as a starting for further implementation
351 of web tools to search, design, rank and review CRISPR-Cas9 single guide RNAs (sgRNAs).

352 **ACKNOWLEDGEMENTS**

353 The authors appreciate the feedbacks from the wider Nannochloropsis community for
354 sharing their insights into creation and curation of these web-based resources. We thank the
355 Tripal community for developing, maintaining and updating the Tripal modules. We also thank

CoreIT and Seeders in Korea for allowing Young Uk Kim and Junkyoung Choe to perform bioinformatic analyses, and description and data transfer work, respectively.

**FUNDING**

This work was financially supported by 31900071 from Natural Science Foundation of China, ZR2019QC012 from Natural Science Foundation of Shandong Province, and the Advanced Biomass R&D Center (ABC) of the Global Frontier Project funded by the Ministry of Science and ICT, Republic of Korea (Grant No. ABC-2010-0029728 for YKC; ABC-2011-0031350 for B-rJ; ABC-2011-0031343 for WJJ).

***Conflict of interest statement***. None declared.

367 **REFERENCES**


368 **Ajjawi, I., Verruto, J., Aqui, M., Soriaga, L.B., Coppersmith, J., Kwok, K., Peach, L.,**
369 **Orchard, E., Kalb, R., Xu, W., Carlson, T.J., Francis, K., Konigsfeld, K., Bartalis,**
370 **J., Schultz, A., Lambert, W., Schwartz, A.S., Brown, R. and Moellering, E.R.** (2017)
371 Lipid production in *Nannochloropsis gaditana* is doubled by decreasing expression of a
372 single transcriptional regulator. *Nature Biotechnology*, **35**, 647-652.
373 **Anders, S. and Huber, W.** (2010) Differential expression analysis for sequence count data.
374 *Genome Biol*, **11**, R106.
375 **Birney, E., Clamp, M. and Durbin, R.** (2004) GeneWise and Genomewise. *Genome Research*,
376 **14**, 988-995.
377 **Brown, R.B., Wass, T.J., Thomas-Hall, S.R. and Schenk, P.M.** (2019) Chromosome-scale
378 genome assembly of two australian *Nannochloropsis oceanica* isolates exhibiting
379 superior lipid characteristics. *Microbiology Resource Announcements*, **8**, e01288-01219.
380 **Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M.,**
381 **Elsik, C.G., Lewis, S.E. and Stein, L.** (2016) JBrowse: a dynamic web platform for
382 genome visualization and analysis. *Genome Biology*, **17**, 66.
383 **Casas-Mollano, J.A., Jeong, B.-r., Xu, J., Moriyama, H. and Cerutti, H.** (2008) The
384 MUT9p kinase phosphorylates histone H3 threonine 3 and is necessary for heritable
385 epigenetic silencing in *Chlamydomonas*. *Proc. Natl. Acad. Sci. U. S. A.*, **105**,
386 6486-6491.
387 **Cerutti, H., Ma, X., Msanne, J. and Repas, T.** (2011) RNA-mediated silencing in algae:
388 biological roles and tools for analysis of gene function. *Eukaryot Cell*, **10**, 1164-1172.
389 **Chen, C., Harst, A., You, W., Xu, J., Ning, K. and Poetsch, A.** (2019) Proteomic study
390 uncovers molecular principles of single-cell-level phenotypic heterogeneity in lipid
391 storage of *Nannochloropsis oceanica*. *Biotechnology for Biofuels*, **12**, 21.
392 **Chen, J.-W., Liu, W.-J., Hu, D.-X., Wang, X., Balamurugan, S., Alimujiang, A., Yang,**
393 **W.-D., Liu, J.-S. and Li, H.-Y.** (2017) Identification of a malonyl CoA-acyl carrier
394 protein transacylase and its regulatory role in fatty acid biosynthesis in oleaginous
395 microalga *Nannochloropsis oceanica*. *Biotechnology and Applied Biochemistry*, **64**,
396 620-626.
397 **Cheng, C.-H., Cho, I.-H., Bett, K.E., Sanderson, L.-A., Staton, M.E., Jung, S., Ficklin, S.P.,**
398 **Lee, T. and Main, D.** (2011) Tripal: a construction toolkit for online genome databases.
399 *Database*, **2011**.
400 **Conesa, A. and Götz, S.** (2008) Blast2GO: A comprehensive suite for functional analysis in
401 plant genomics. *International Journal of Plant Genomics*, **2008**, 619832.
402 **Corteggiani Carpinelli, E., Telatin, A., Vitulo, N., Forcato, C., D'Angelo, M., Schiavon, R.,**
403 **Vezzi, A., Giacometti, G.M., Morosinotto, T. and Valle, G.** (2014) Chromosome scale
404 genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen
405 depletion. *Molecular Plant*, **7**, 323-335.
406 **Cox, M.P., Peterson, D.A. and Biggs, P.J.** (2010) SolexaQA: At-a-glance quality assessment
407 of Illumina second-generation sequencing data. *BMC Bioinformatics*, **11**, 485.

**Dereeper, A., Bocs, S., Rouard, M., Guignon, V., Ravel, S., Tranchant-Dubreuil, C., Poncet, V., Garsmeur, O., Lashermes, P. and Droc, G.** (2015) The coffee genome hub: a resource for coffee genomes. **43**, D1028-D1035.

**Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R.** (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15-21.

**Dolch, L.-J., Rak, C., Perin, G., Tourcier, G., Broughton, R., Leterrier, M., Morosinotto, T., Tellier, F., Faure, J.-D., Falconet, D., Jouhet, J., Sayanova, O., Beaudoin, F. and Maréchal, E.** (2017) A palmitic acid elongase affects eicosapentaenoic acid and plastidial monogalactosyldiacylglycerol levels in *Nannochloropsis*. *Plant Physiology*, **173**, 742-759.

**Evers, M., Huttner, M., Dueck, A., Meister, G. and Engelmann, J.C.** (2015) miRA: adaptable novel miRNA identification in plants using small RNA sequencing data. *BMC Bioinformatics*, **16**, 370.

**Fawley, K.P. and Fawley, M.W.** (2007) Observations on the diversity and ecology of freshwater *Nannochloropsis* (Eustigmatophyceae), with descriptions of new taxa. *Protist*, **158**, 325-336.

**Fernández-Acero, F.J., Amil-Ruiz, F., Durán-Peña, M.J., Carrasco, R., Fajardo, C., Guarnizo, P., Fuentes-Almagro, C. and Vallejo, R.A.** (2019) Valorisation of the microalgae *Nannochloropsis gaditana* biomass by proteomic approach in the context of circular economy. *Journal of Proteomics*, **193**, 239-242.

**Ferrari, A., Fiorino, E., Giudici, M., Gilardi, F., Galmozzi, A., Mitro, N., Cermenati, G., Godio, C., Caruso, D., De Fabiani, E. and Crestani, M.** (2012) Linking epigenetics to lipid metabolism: Focus on histone deacetylases. *Molecular Membrane Biology*, **29**, 257-266.

**Freire, I., Cortina-Burgueño, A., Grille, P., Arizcun Arizcun, M., Abellán, E., Segura, M., Witt Sousa, F. and Otero, A.** (2016) *Nannochloropsis limnetica*: A freshwater microalga for marine aquaculture. *Aquaculture*, **459**, 124-130.

**Furlan, A., Jacquier, M., Woller, A., Heliot, L., Duez, H., Staels, B. and Lefranc, M.** (2019) Mathematical models converge on PGC1alpha as the key metabolic integrator of SIRT1 and AMPK regulation of the circadian clock. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 13171–13172.

**Gao, Y., Wang, J. and Zhao, F.** (2015) CIRI: an efficient and unbiased algorithm for *de novo* circular RNA identification. *Genome Biol*, **16**, 4.

**Gee, C.W. and Niyogi, K.K.** (2017) The carbonic anhydrase CAH1 is an essential component of the carbon-concentrating mechanism in *Nannochloropsis oceanica*. *Proceedings of the National Academy of Sciences*, 201700139.

**Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A.** (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644-652.

450 **Griswold, A.J., Chang, K.T., Runko, A.P., Knight, M.A. and Min, K.T.** (2008) Sir2
451 mediates apoptosis through JNK-dependent pathways in Drosophila. *Proc. Natl. Acad.*
452 *Sci. U. S. A.*, **105**, 8673-8678.
453 **Guo, L., Liang, S., Zhang, Z., Liu, H., Wang, S., Pan, K., Xu, J., Ren, X., Pei, S. and Yang,**
454 **G.** (2019) Genome assembly of *Nannochloropsis oceanica* provides evidence of host
455 nucleus overthrow by the symbiont nucleus during speciation. *Communications Biology*,
456 **2**, 249.
457 **He, Y., Wang, X., Ma, B. and Xu, J.** (2019) Ramanome technology platform for label-free
458 screening and sorting of microbial cell factories at single-cell resolution. *Biotechnology*
459 *Advances*, **37**, 107388.
460 **He, Y., Zhang, P., Huang, S., Wang, T., Ji, Y. and Xu, J.** (2017) Label-free, simultaneous
461 quantification of starch, protein and triacylglycerol in single microalgal cells.
462 *Biotechnology for Biofuels*, **10**, 275.
463 **Hu, J., Wang, D., Li, J., Jing, G., Ning, K. and Xu, J.** (2014) Genome-wide identification of
464 transcription factors and transcription-factor binding sites in oleaginous microalgae
465 *Nannochloropsis*. *Scientific Reports*, **4**, 5454.
466 **Jeong, B.-r., Wu-Scharf, D., Zhang, C. and Cerutti, H.** (2002) Suppressors of transcriptional
467 transgenic silencing in *Chlamydomonas* are sensitive to DNA-damaging agents and
468 reactivate transposable elements. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 1076-1081.
469 **Jeong, S.W., Nam, S.W., HwangBo, K., Jeong, W.J., Jeong, B.-r., Chang, Y.K. and Park,**
470 **Y.-I.** (2017) Transcriptional regulation of cellulose biosynthesis during the early phase
471 of nitrogen deprivation in *Nannochloropsis salina*. *Scientific Reports*, **7**, 5264.
472 **Ji, Y., He, Y., Cui, Y., Wang, T., Wang, Y., Li, Y., Huang, W.E. and Xu, J.** (2014) Raman
473 spectroscopy provides a rapid, non-invasive method for quantitation of starch in live,
474 unicellular microalgae. *Biotechnology Journal*, **9**, 1512-1518.
475 **Jia, J., Han, D., Gerken, H.G., Li, Y., Sommerfeld, M., Hu, Q. and Xu, J.** (2015) Molecular
476 mechanisms for photosynthetic carbon partitioning into storage neutral lipids in
477 *Nannochloropsis oceanica* under nitrogen-depletion conditions. *Algal Research*, **7**,
478 66-77.
479 **Jinkerson, R.E., Radakovits, R. and Posewitz, M.C.** (2013) Genomic insights from the
480 oleaginous model alga *Nannochloropsis gaditana*. *Bioengineered*, **4**, 37-43.
481 **Jung, S., Ficklin, S.P., Lee, T., Cheng, C.-H., Blenda, A., Zheng, P., Yu, J., Bombarely, A.,**
482 **Cho, I., Ru, S., Evans, K., Peace, C., Abbott, A.G., Mueller, L.A., Olmstead, M.A.**
483 **and Main, D.** (2014) The Genome Database for Rosaceae (GDR): year 10 update. **42**,
484 D1237-D1244.
485 **Jung, S., Lee, T., Ficklin, S., Yu, J., Cheng, C.H. and Main, D.** (2016) Chado use case:
486 storing genomic, genetic and breeding data of Rosaceae and Gossypium crops in Chado.
487 *Database the Journal of Biological Databases & Curation*, **2016**, baw058.
488 **Jung, S., Staton, M., Lee, T., Blenda, A., Svancara, R., Abbott, A. and Main, D.** (2007)
489 GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae
490 genomics and genetics data. **36**, D1034-D1040.

**Kang, N.K., Jeon, S., Kwon, S., Koh, H.G., Shin, S.-E., Lee, B., Choi, G.-G., Yang, J.-W., Jeong, B.-r. and Chang, Y.K.** (2015) Effects of overexpression of a bHLH transcription factor on biomass and lipid production in *Nannochloropsis salina*. *Biotechnology for Biofuels*, **8**, 200.

**Kang, N.K., Kim, E.K., Kim, Y.U., Lee, B., Jeong, W.J., Jeong, B.R. and Chang, Y.K.** (2017) Increased lipid production by heterologous expression of AtWRI1 transcription factor in *Nannochloropsis salina*. *Biotechnol Biofuels*, **10**, 231.

**Kang, N.K., Kim, E.K., Sung, M.G., Kim, Y.U., Jeong, B.R. and Chang, Y.K.** (2019) Increased biomass and lipid production by continuous cultivation of *Nannochloropsis salina* transformant overexpressing a bHLH transcription factor. *Biotechnol Bioeng*, **116**, 555-568.

**Karp, P.D., Paley, S. and Romero, P.** (2001) The Pathway Tools software. *Bioinformatics*, **18 Suppl 1**, S225.

**Kilian, O., Benemann, C.S.E., Niyogi, K.K. and Vick, B.** (2011) High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proceedings of the National Academy of Sciences*, **108**, 21265-21269.

**Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L.** (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, **14**, R36.

**Koh, H.G., Kang, N.K., Jeon, S., Shin, S.-E., Jeong, B.-r. and Chang, Y.K.** (2019) Heterologous synthesis of chlorophyll b in *Nannochloropsis salina* enhances growth and lipid production by increasing photosynthetic efficiency. *Biotechnol Biofuels*, **12**, 122.

**Langmead, B. and Salzberg, S.L.** (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.

**Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L.** (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, **10**, R25.

**Letsiou, S., Kalliampakou, K., Gardikis, K., Mantecon, L., Infante, C., Chatzikonstantinou, M., Labrou, N.E. and Flemetakis, E.** (2017) Skin protective effects of *Nannochloropsis gaditana* extract on $H_2O_2$-stressed human dermal fibroblasts. *Front Mar Sci*, **4**, 1-15.

**Li, J., Han, D., Wang, D., Ning, K., Jia, J., Wei, L., Jing, X., Huang, S., Chen, J., Li, Y., Hu, Q. and Xu, J.** (2014) Choreography of transcriptomes and lipidomes of *Nannochloropsis* reveals the mechanisms of oil synthesis in microalgae. *The Plant Cell*, **26**, 1645-1665.

**Liao, Y., Smyth, G.K. and Shi, W.** (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923-930.

**Liu, J., Song, Y. and Qiu, W.** (2017) Oleaginous microalgae Nannochloropsis as a new model for biofuel production: Review & analysis. *Renewable and Sustainable Energy Reviews*, **72**, 154-162.

532 **Llansola-Portoles, M.J., Litvin, R., Ilioaia, C., Pascal, A.A., Bina, D. and Robert, B.** (2017)
533 Pigment structure in the violaxanthin–chlorophyll-a-binding protein VCP.
534 *Photosynthesis Research*, **134**, 51-58.
535 **Love, M.I., Huber, W. and Anders, S.** (2014) Moderated estimation of fold change and
536 dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
537 **Lu, Y., Tarkowská, D., Turečková, V., Luo, T., Xin, Y., Li, J., Wang, Q., Jiao, N., Strnad, M.**
538 **and Xu, J.** (2014a) Antagonistic roles of abscisic acid and cytokinin during response to
539 nitrogen depletion in oleaginous microalga *Nannochloropsis oceanica* expand the
540 evolutionary breadth of phytohormone function. *The Plant Journal*, **80**, 52-68.
541 **Lu, Y. and Xu, J.** (2015) Phytohormones in microalgae: a new opportunity for microalgal
542 biotechnology? *Trends in Plant Science*, **20**, 273-282.
543 **Lu, Y., Zhou, W., Wei, L., Li, J., Jia, J., Li, F., Smith, S.M. and Xu, J.** (2014b) Regulation of
544 the cholesterol biosynthetic pathway and its integration with fatty acid biosynthesis in
545 the oleaginous microalga *Nannochloropsis oceanica*. *Biotechnology for Biofuels*, **7**, 81.
546 **Ma, X., Yao, L., Yang, B., Lee, Y.K., Chen, F. and Liu, J.** (2017) RNAi-mediated silencing of
547 a pyruvate dehydrogenase kinase enhances triacylglycerol biosynthesis in the
548 oleaginous marine alga *Nannochloropsis salina*. *Scientific Reports*, **7**, 11485.
549 **Martin, M.** (2011) Cutadapt removes adapter sequences from high-throughput sequencing
550 reads. *EMBnet.journal*, **17**, 3.
551 **Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L.,**
552 **Mackowiak, S.D., Gregersen, L.H., Munschauer, M., Loewer, A., Ziebold, U.,**
553 **Landthaler, M., Kocks, C., le Noble, F. and Rajewsky, N.** (2013) Circular RNAs are a
554 large class of animal RNAs with regulatory potency. *Nature*, **495**, 333-338.
555 **Mitchell, A., Sangrador-Vegas, A., Quinn, A.F., McAnulla, C., Binns, D., Nuka, G.,**
556 **McWilliam, H., Chang, H.-Y., Maslen, J., Fraser, M., Scheremetjew, M., Jones, P.,**
557 **Lopez, R., Hunter, S., Pesseat, S., Yong, S.-Y. and Li, W.** (2014) InterProScan 5:
558 genome-scale protein function classification. *Bioinformatics*, **30**, 1236-1240.
559 **Moog, D., Stork, S., Reislöhner, S., Grosche, C. and Maier, U.-G.** (2015) In vivo localization
560 studies in the stramenopile alga *Nannochloropsis oceanica*. *Protist*, **166**, 161-171.
561 **Mühlroth, A., Winge, P., El Assimi, A., Jouhet, J., Maréchal, E., Hohmann-Marriott, M.F.,**
562 **Vadstein, O. and Bones, A.M.** (2017) Mechanisms of phosphorus acquisition and lipid
563 class remodeling under P limitation in a marine microalga. *Plant Physiology*, **175**,
564 1543-1559.
565 **Naduthodi, M.I.S., Mohanraju, P., Südfeld, C., D'Adamo, S., Barbosa, M.J. and van der**
566 **Oost, J.** (2019) CRISPR-Cas ribonucleoprotein mediated homology-directed repair for
567 efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1.
568 *Biotechnology for Biofuels*, **12**, 66.
569 **Navarro, L., Dunoyer, P., Jay, F., Arnold, B., Dharmasiri, N., Estelle, M., Voinnet, O. and**
570 **Jones, J.D.G.** (2006) A plant miRNA contributes to antibacterial resistance by
571 repressing auxin signaling. *Science*, **312**, 436-439.

572 **Nobusawa, T., Hori, K., Mori, H., Kurokawa, K. and Ohta, H.** (2017) Differently localized
573 lysophosphatidic acid acyltransferases crucial for triacylglycerol biosynthesis in the
574 oleaginous alga *Nannochloropsis*. *The Plant Journal*, **90**, 547-559.
575 **Ohan, J.A., Hovde, B.T., Zhang, X.L., Davenport, K.W., Chertkov, O., Han, C., Twary,**
576 **S.N. and Starkenburg, S.R.** (2019) Nuclear genome assembly of the microalga
577 *Nannochloropsis salina* CCMP1776. *Microbiology Resource Announcements*, **8**,
578 e00750-00719.
579 **Pan, K., Qin, J., Li, S., Dai, W., Zhu, B., Jin, Y., Yu, W., Yang, G. and Li, D.** (2011) Nuclear
580 monoploidy and asexual propagation of *Nannochloropsis oceanica* (Eustigmatophyceae)
581 as revealed by its genome sequence(1). *Journal of Phycology*, **47**, 1425-1432.
582 **Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T. and Salzberg, S.L.**
583 (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq
584 reads. *Nature Biotechnology*, **33**, 290.
585 **Poliner, E., Farré, E.M. and Benning, C.** (2018a) Advanced genetic tools enable synthetic
586 biology in the oleaginous microalgae *Nannochloropsis* sp. *Plant Cell Reports*, **37**,
587 1383-1399.
588 **Poliner, E., Panchy, N., Newton, L., Wu, G., Lapinsky, A., Bullard, B., Zienkiewicz, A.,**
589 **Benning, C., Shiu, S.-H. and Farré, E.M.** (2015) Transcriptional coordination of
590 physiological responses in *Nannochloropsis oceanica* CCMP1779 under light/dark
591 cycles. *The Plant Journal*, **83**, 1097-1113.
592 **Poliner, E., Pulman, J.A., Zienkiewicz, K., Childs, K., Benning, C. and Farré, E.M.** (2018b)
593 A toolkit for *Nannochloropsis oceanica* CCMP1779 enables gene stacking and genetic
594 engineering of the eicosapentaenoic acid pathway for enhanced long-chain
595 polyunsaturated fatty acid production. *Plant Biotechnology Journal*, **16**, 298-309.
596 **Poliner, E., Takeuchi, T., Du, Z.-Y., Benning, C. and Farré, E.M.** (2018c) Nontransgenic
597 marker-free gene disruption by an episomal CRISPR system in the oleaginous
598 microalga, *Nannochloropsis oceanica* CCMP1779. *ACS Synthetic Biology*, **7**, 962-968.
599 **Radakovits, R., Jinkerson, R.E., Fuerstenberg, S.I., Tae, H., Settlage, R.E., Boore, J.L.**
600 **and Posewitz, M.C.** (2012) Draft genome sequence and genetic transformation of the
601 oleaginous alga *Nannochloropsis gaditana*. *Nature Communications*, **3**, 686.
602 **Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E.** (2012) Oases: robust de novo
603 RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, **28**,
604 1086-1092.
605 **Schwartz, A.S., Brown, R., Ajjawi, I., McCarren, J., Atilla, S., Bauman, N. and**
606 **Richardson, T.H.** (2018) Complete genome sequence of the model oleaginous alga
607 *Nannochloropsis gaditana* CCMP1894. *Genome Announcements*, **6**, e01448-01417.
608 **Shimazu, T., Hirschey, M.D., Newman, J., He, W., Shirakawa, K., Le Moan, N., Grueter,**
609 **C.A., Lim, H., Saunders, L.R., Stevens, R.D., Newgard, C.B., Farese, R.V., Jr., de**
610 **Cabo, R., Ulrich, S., Akassoglou, K. and Verdin, E.** (2013) Suppression of oxidative
611 stress by beta-hydroxybutyrate, an endogenous histone deacetylase inhibitor. *Science*,
612 **339**, 211-214.

613 **Singh, P.P., Demmitt, B.A., Nath, R.D. and Brunet, A.** (2019) The genetics of aging: A
614 vertebrate perspective. *Cell*, **177**, 200-220.
615 **Stanke, M. and Morgenstern, B.** (2005) AUGUSTUS: a web server for gene prediction in
616 eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, **33**,
617 W465-W467.
618 **Thomou, T., Mori, M.A., Dreyfuss, J.M., Konishi, M., Sakaguchi, M., Wolfrum, C., Rao,**
619 **T.N., Winnay, J.N., Garcia-Martin, R., Grinspoon, S.K., Gorden, P. and Kahn, C.R.**
620 (2017) Adipose-derived circulating miRNAs regulate gene expression in other tissues.
621 *Nature*, **542**, 450.
622 **Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J.,**
623 **Salzberg, S.L., Wold, B.J. and Pachter, L.** (2010) Transcript assembly and
624 quantification by RNA-Seq reveals unannotated transcripts and isoform switching
625 during cell differentiation. *Nature Biotechnology*, **28**, 511-515.
626 **Tyanova, S. and Cox, J.** (2018) Perseus: a bioinformatics platform for integrative analysis of
627 proteomics data in cancer research. In *Cancer Systems Biology: Methods and Protocols*
628 (von Stechow, L. ed. New York, NY: Springer New York, pp. 133-148.
629 **Tyanova, S., Temu, T. and Cox, J.** (2016) The MaxQuant computational platform for mass
630 spectrometry-based shotgun proteomics. *Nature Protocols*, **11**, 2301.
631 **van Rooij, E., Sutherland, L.B., Qi, X., Richardson, J.A., Hill, J. and Olson, E.N.** (2007)
632 Control of stress-dependent cardiac growth and gene expression by a microRNA.
633 *Science*, **316**, 575-579.
634 **Verruto, J., Francis, K., Wang, Y., Low, M.C., Greiner, J., Tacke, S., Kuzminov, F.,**
635 **Lambert, W., McCarren, J., Ajjawi, I., Bauman, N., Kalb, R., Hannum, G. and**
636 **Moellering, E.R.** (2018) Unrestrained markerless trait stacking in *Nannochloropsis*
637 *gaditana* through combined genome editing and marker recycling technologies.
638 *Proceedings of the National Academy of Sciences*, **115**, E7015-E7022.
639 **Vieler, A., Wu, G., Tsai, C.-H., Bullard, B., Cornish, A.J., Harvey, C., Reca, I.-B.,**
640 **Thornburg, C., Achawanantakun, R., Buehl, C.J., Campbell, M.S., Cavalier, D.,**
641 **Childs, K.L., Clark, T.J., Deshpande, R., Erickson, E., Armenia Ferguson, A.,**
642 **Handee, W., Kong, Q., Li, X., Liu, B., Lundback, S., Peng, C., Roston, R.L.,**
643 **Sanjaya, Simpson, J.P., TerBush, A., Warakanont, J., Zäuner, S., Farre, E.M.,**
644 **Hegg, E.L., Jiang, N., Kuo, M.-H., Lu, Y., Niyogi, K.K., Ohlrogge, J., Osteryoung,**
645 **K.W., Shachar-Hill, Y., Sears, B.B., Sun, Y., Takahashi, H., Yandell, M., Shiu, S.-H.**
646 **and Benning, C.** (2012) Genome, functional gene annotation, and nuclear
647 transformation of the heterokont oleaginous alga *Nannochloropsis oceanica*
648 CCMP1779. *PLOS Genetics*, **8**, e1003064.
649 **Wang, D., Ning, K., Li, J., Hu, J., Han, D., Wang, H., Zeng, X., Jing, X., Zhou, Q., Su, X.,**
650 **Chang, X., Wang, A., Wang, W., Jia, J., Wei, L., Xin, Y., Qiao, Y., Huang, R., Chen,**
651 **J., Han, B., Yoon, K., Hill, R.T., Zohar, Y., Chen, F., Hu, Q. and Xu, J.** (2014a)
652 *Nannochloropsis* genomes reveal evolution of microalgal oleaginous traits. *PLOS*
653 *Genetics*, **10**, e1004094.

654 **Wang, Q., Lu, Y., Xin, Y., Wei, L., Huang, S. and Xu, J.** (2016) Genome editing of model
655 oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9. *The Plant Journal*, **88**,
656 1071-1081.
657 **Wang, T., Ji, Y., Wang, Y., Jia, J., Li, J., Huang, S., Han, D., Hu, Q., Huang, W.E. and Xu,**
658 **J.** (2014b) Quantitative dynamics of triacylglycerol accumulation in microalgae
659 populations at single-cell resolution revealed by Raman microspectroscopy.
660 *Biotechnology for Biofuels*, **7**, 58.
661 **Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B.**
662 **and Guo, H.** (2012) MCScanX: A toolkit for detection and evolutionary analysis of
663 gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49-e49.
664 **Wei, L., El Hajjami, M., Shen, C., You, W., Lu, Y., Li, J., Jing, X., Hu, Q., Zhou, W.,**
665 **Poetsch, A. and Xu, J.** (2019) Transcriptomic and proteomic responses to very low
666 CO2 suggest multiple carbon concentrating mechanisms in *Nannochloropsis oceanica*.
667 *Biotechnology for Biofuels*, **12**, 168.
668 **Wei, L., Xin, Y., Wang, D., Jing, X., Zhou, Q., Su, X., Jia, J., Ning, K., Chen, F., Hu, Q.**
669 **and Xu, J.** (2013) *Nannochloropsis* plastid and mitochondrial phylogenomes reveal
670 organelle diversification mechanism and intragenus phylotyping strategy in microalgae.
671 *BMC Genomics*, **14**, 534.
672 **Wei, L., Xin, Y., Wang, Q., Yang, J., Hu, H. and Xu, J.** (2017) RNAi-based targeted gene
673 knockdown in the model oleaginous microalgae *Nannochloropsis oceanica*. *The Plant*
674 *Journal*, **89**, 1236-1250.
675 **Wei, L. and Xu, J.** (2018) Optimized methods of chromatin immunoprecipitation (ChIP) for
676 profiling histone modifications in industrial microalgae *Nannochloropsis* spp. *J Phycol*,
677 **54**, 358-367.
678 **Wu-Scharf, D., Jeong, B.-r., Zhang, C. and Cerutti, H.** (2000) Transgene and transposon
679 silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase. *Science*, **290**,
680 1159-1162.
681 **Xin, Y., Lu, Y., Lee, Y.-Y., Wei, L., Jia, J., Wang, Q., Wang, D., Bai, F., Hu, H., Hu, Q., Liu,**
682 **J., Li, Y. and Xu, J.** (2017) Producing designer oils in industrial microalgae by rational
683 modulation of co-evolving type-2 diacylglycerol acyltransferases. *Molecular Plant*, **10**,
684 1523-1539.
685 **Xin, Y., Shen, C., She, Y., Chen, H., Wang, C., Wei, L., Yoon, K., Han, D., Hu, Q. and Xu,**
686 **J.** (2019) Biosynthesis of triacylglycerol molecules with a tailored PUFA profile in
687 industrial microalgae. *Molecular Plant*, **12**, 474-488.
688 **Xue, J.-H., Chen, G.-D., Hao, F., Chen, H., Fang, Z., Chen, F.-F., Pang, B., Yang, Q.-L.,**
689 **Wei, X., Fan, Q.-Q., Xin, C., Zhao, J., Deng, X., Wang, B.-A., Zhang, X.-J., Chu, Y.,**
690 **Tang, H., Yin, H., Ma, W., Chen, L., Ding, J., Weinhold, E., Kohli, R.M., Liu, W.,**
691 **Zhu, Z.-J., Huang, K., Tang, H. and Xu, G.-L.** (2019) A vitamin-C-derived DNA
692 modification catalysed by an algal TET homologue. *Nature*, **569**, 581-585.
693 **Yu, J., Jung, S., Cheng, C.-H., Ficklin, S.P., Lee, T., Zheng, P., Jones, D., Percy, R.G. and**
694 **Main, D.** (2014) CottonGen: a genomics, genetics and breeding database for cotton
695 research. **42**, D1229-D1236.

**Zerbino, D.R. and Birney, E.** (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821-829.

**Zienkiewicz, K., Zienkiewicz, A., Poliner, E., Du, Z.-Y., Vollheyde, K., Herrfurth, C., Marmon, S., Farré, E.M., Feussner, I. and Benning, C.** (2017) *Nannochloropsis*, a rich source of diacylglycerol acyltransferases for engineering of triacylglycerol content in different hosts. *Biotechnology for Biofuels*, **10**, 8.

704 **FIGURE LEGEND**

705 **Figure 1. Design of the NanDeSyn database.**

706

707 **SUPPLEMENTAL TABLES**

708 **Table S1. Genomic resources of *Nannochloropsis* spp. in NanDeSyn at present.**

709

710 **Table S2. Transcriptome and proteome datasets of *Nannochloropsis* spp. in NanDeSyn at**

711 **present.**

712

713 **Table S3. Summary of the Pathway/Genome Databases (PGDBs) in NannoCyc at present.**

714

715 **Table S4. Selected representatives of *Nannochloropsis* genes that have been mutated,**

716 **perturbed or experimentally validated for subcellular localization (as recorded in the**

717 **present version of the NanDeSyn website).**

718

719 **Figure S1. The gene feature page in NanDeSyn.** (**A**) Overview of a specific gene. (**B**)

720 Synteny blocks and homologous genes in the "Relationships" section. (**C**) The histogram

721 shows expression profiles of a specific gene under a selected project in the "Expression"

722 section. (**D**) Links to all the analyses associated with this gene.

723

724 **Figure S2. Web interface for accessing the functions in NanDeSyn.** (**A**) Synteny viewer in
725 NanDeSyn. Left: synteny blocks displayed in the circos plot; Right: selected synteny block.
726 Homologous gene pairs are connected by grey lines. (**B**) Heatmap showing expression profiles
727 of user supplied genes. (**C**) Integrated analysis of multi-omics data using JBrowse in
728 NanDeSyn.