View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by OIST Institutional Repository

OIST OKINAWA INSTITUTE OF SCIENCE AND TECHNOLOGY GRADUATE UNIVERSITY
沖縄科学技術大学院大学

Differential gene expression in fronds and stolons of the siphonous macroalga, Caulerpa lentillifera

Author	Asuka Arimoto, Koki Nishitsuji, Haruhi Narisoko, Eiichi Shoguchi, Noriyuki Satoh
journal or publication title	Development, Growth & Differentiation
volume	61
number	9
page range	475-484
year	2019-11-11
Publisher	John Wiley and Sons
Rights	(C) 2019 Japanese Society of Developmental Biologists This is the peer reviewed version of the following article: Arimoto, A, Nishitsuji, K, Narisoko, H, Shoguchi, E, Satoh, N. Differential gene expression in fronds and stolons of the siphonous macroalga, Caulerpa lentillifera. Develop Growth Differ. 2019; 61: 475 484., which has been published in final form at https://doi.org/10.1111/dgd.12634. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.
Author's flag	author
URL	http://id.nii.ac.jp/1394/00001238/

doi: info:doi/10.1111/dgd.12634

1 **[DGD-00019-2019 revised form]**
2
3 **Differential gene expression in fronds and stolons of the siphonous macroalga,**
4 ***Caulerpa lentillifera***
5
6
7 **Asuka Arimoto[*], Koki Nishitsuji, Haruhi Narisoko, Eiichi Shoguchi, Noriyuki**
8 **Satoh**
9
10 Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate
11 University, Onna, Okinawa 904-0495, Japan
12
13 [*]Correspondence: Asuka Arimoto
14 Email: asuka.arimoto@oist.jp; Tel: +81-98-966-8653, FAX: +81-98-966-8622
15
16 Running title: Gene expression profile of a siphonous alga
17

Abstract:

The green alga, *Caulerpa lentillifera*, is composed of a single cell with multiple nuclei, but it possesses structures analogous to leaves or fronds, stems or stolons, and roots or rhizoids. To understand molecular mechanisms involved in formation and function of these structures, we carried out RNA-seq analysis of fronds and stolons (including rhizoids). Taking advantage of the decoded genome of *C. lentillifera*, the present RNA-seq analysis addressed transcripts corresponding to 9,311 genes identified in the genome. RNA-seq data suggested that 8,734 genes are expressed in sporophytes. In spite of the siphonous body of the alga, differential gene expression was evident in the two structures. 1,027 (11.8%) and 1,129 (12.9%) genes were preferentially expressed in fronds and stolons, respectively, while the remaining 6,578 (75.3%) genes were expressed at the same level in both. Most genes preferentially expressed in fronds are associated with photosynthesis and plant hormone pathways, including abscisic acid signaling. In contrast, those preferentially expressed in stolons are associated with translation and DNA replication. These results indicate that gene expression is regulated differently between fronds and stolons, which probably governs the function of each structure. Together with genomic information, the present transcriptomic data provide genic information about development and physiology of this unique, siphonous organism.

KEYWORDS

a siphonous green alga, *Caulerpa lentillifera*, differential gene expression, fronds and stolons, plant hormone pathways

44 **1. INTRODUCTION**

45

46 The green alga, *Caulerpa lentillifera*, belongs to the division Chlorophyta, the class
47 Ulvophyceae, the order Bryopsidales, and the family Caulerpaceae (Adl et al. 2005;
48 Draisma et al. 2014). The *Caulerpa* body is composed of a single cell with multiple
49 nuclei (Coneva & Chitwood 2015). Some siphonous algae reach meters in length, likely
50 the largest single-celled organisms on Earth. However, these algae possess structures
51 analogous to leaves or fronds, stems or stolons, and roots or rhizoids (Figure 1). About
52 200 species of *Caulerpa* have been reported, with diverse frond morphologies
53 (AlgaeBase, http://www.algaebase.org). *C. lentillifera* morphology resembles clusters
54 of green grapes (Figure 1); hence, it is commonly known as sea grapes, or "umi-budo"
55 in Japanese. Using rhizoids, *C. lentillifera* anchors itself to substrates such as rocks or
56 ropes. Fronds are formed with new stolons, in which new vesicle-rudiments are added
57 apically. This alga is one of the major edible seaweeds in the subtropical/tropical
58 Asia-Pacific region, especially cultivated for market in Okinawa, Japan (Mary et al.
59 2009).

60

61 In a previous study, Ranjan et al. (2015) carried out RNA-seq analysis using *C. taxifolia*,
62 a well-known toxic, invasive species (Mozzachiodi et al. 2001; Galil 2007). Ranjan et al.
63 (2015) showed that a set of genes was highly expressed in stolons, rachis (frond stem),
64 and apex (frond tips). These genes are associated with DNA replication and chromatin,
65 translation, and coat-protein complex (COP) vesicles and kinases, respectively. They
66 suggested that a global, basal-apical pattern of transcript distribution from the holdfast
67 to the frond apex roughly follows the flow of genetic information in the cell, that is,
68 from transcription to translation. This is an interesting result, given that the alga is a
69 unicellular, multinucleated organism. That is, the plant body comprises a single cell.
70 The stolon corresponds to the nucleus in the middle of the body, the rachis to the central
71 cytoplasm, and the apex to the peripheral cytoplasm, respectively (Ranjan et al. 2015).
72 Although the distribution of nuclei in the entire body has not been shown yet, this result
73 raises several basic questions about how morphology and function of these structures
74 are differentiated and/or maintained in *Caulerpa*.

75

76 We focused on *C. lentillifera*, the size of which is similar to *C. taxifolia*, approximately
77 10 to 30 cm from stolon to frond apex (Figure 1) (Paul et al. 2014; Ranjan et al. 2015).
78 In order to understand the biology of the siphonous body plan, we decoded the
79 approximately 26-Mb genome of *C. lentillifera* (Arimoto et al. 2019). This genome was

80 estimated to contain 9,311 protein-coding genes. Unexpectedly, homologous genes for
81 some components of plant hormone pathways were conserved and/or highly expanded
82 in the *Caulerpa* genome. It has been reported that outgrowth of the plant body is
83 controlled by plant hormone pathways (Bar & Ori 2014; Abad et al. 2017). These
84 hormones are also essential to maintain plant physiological homeostasis (Verma et al.
85 2016). However, it is thought that the green siphonous algal lineage evolved body plans
86 and sessile life styles independently of the land plant lineage (Umen 2014). It is
87 interesting that while these organisms have different body plans and cellularities, they
88 employ similar mechanisms to regulate development (Kaplan & Hagemann 1991).
89
90 The present study carried out RNA-seq analysis of *C. lentillifera* fronds and stolons,
91 because these two regions are the major structural components of this alga.
92
93 **2. MATERIALS AND METHODS**
94
95 Our gene expression analysis was based on sequencing data obtained in the *Caulerpa*
96 *lentillifera* genome decoding project (Arimoto et al. 2019). In that project, genomic
97 DNA and mRNA of *C. lentillifera* were sequenced to reconstruct the genome and to
98 predict gene models. Procedures for RNA isolation and sequencing are described briefly
99 below.
100
101 **2.1. Algal specimens**
102 *Caulerpa lentillifera* was cultivated in aquaria under natural light and harvested during
103 the daytime at the Onna Village Fisheries Cooperative, Okinawa, Japan. Large, intact
104 (undamaged) sporophytes were used for RNA extraction. Debris attached to the
105 sporophytes was removed by flushing with 0.22-μm-filtered seawater. Fronds and
106 stolons were separated using sterilized scalpels and frozen in liquid nitrogen
107 immediately. Frozen samples were stored at -80°C until RNA extraction.
108
109 **2.2. RNA extraction**
110 Frozen samples were ground in a mortar and pestle, and put into Plant RNA reagent (cat.
111 no. 12322012; ThermoFisher, MA, USA). Contaminating DNA was removed by DNase
112 treatment using columns from a Qiagen RNeasy Plant Mini Kit (cat. no. 74904; Qiagen,
113 Germany). RNA extraction and purification procedures followed manufacturer
114 instructions. RNA purity and quantity were verified with a NanoDrop 2000
115 Spectrophotometer (ThermoFisher, MA, USA). The amount of extracted RNA was

116 quantified using a Qubit RNA HS Assay Kit (cat. no. Q32852; ThermoFisher, MA,
117 USA). RNA integrity was confirmed using electropherograms generated by the Agilent
118 2100 Bioanalyzer (Agilent Technologies, CA, USA).
119

120 **2.3. Library preparation and sequencing**

121 Six RNA-seq libraries were prepared from fronds and stolons of three specimens.
122 Extracted RNA was converted into RNA-seq libraries with a TruSeq Stranded mRNA
123 Library Prep Kit (cat. no. RS-122-2101; Illumina, CA, USA). All sequencing libraries
124 were prepared with protocols provided by the manufacturers. RNA libraries were
125 sequenced using the Illumina HiSeq 4000 platform (Illumina, CA, USA).
126

127 **2.4. Expression Analysis**

128 Sequencing adapters and low-quality (< Q20) regions of Illumina reads were removed
129 with Trimmomatic 0.33 (Bolger et al. 2014) and Sickle 1.33
130 (https://github.com/najoshi/sickle), respectively. Trimmed, paired-end reads were
131 mapped onto genome sequences using STAR 2.5.2a (Dobin et al. 2013). Index files for
132 STAR were generated with gene model information using the '--sjdbGTFfile' option.
133 Expression levels of genes were calculated using edgeR 3.20.1 (McCarthy et al. 2012),
134 based on mapping results from which chimeric mapped reads were excluded.
135 Expression levels were converted into counts-per-million (CPM) from RNA-seq data.
136 Data were normalized using the Trimmed mean M-value (TMM) and Relative Log
137 Expression (RLE) methods in edgeR 3.20.1 and DESeq2 version 1.18.1 (Love et al.
138 2014), respectively. After normalization, generalized linear model methods (GLMs)
139 (McCarthy et al. 2012) were applied to calculate gene expression levels in the edgeR
140 analysis. Differences in gene expression level were examined using a likelihood-ratio
141 test.
142

143 **2.5. Quantitative real-time PCR**

144 RNA extraction and purification methods for quantitative real-time PCR (qRT-PCR)
145 were the same as for RNA-seq library preparation. mRNAs in purified total RNA were
146 enriched using a TruSeq Stranded mRNA Library Prep Kit. cDNA was synthesized
147 from enriched mRNA equivalent to 1 μg of total RNA using SuperScript VILO Master
148 Mix (cat. no. 11755050; ThermoFisher, MA, USA). cDNA equivalent to 1 ng of total
149 RNA was used as a template for qRT-PCR. Quantification of gene expression was
150 performed on the StepOnePlus Real-Time PCR System (ThermoFisher, MA, USA)
151 using Probe qPCR Mix (cat. no. RR391A; Takara Bio, Japan) and

152 FAM/TAMRA-labeled probes synthesized by Takara Bio Inc. Composition of reaction
153 mixtures and thermocycler settings followed manufacturer instructions. Primer and
154 probe sequences used for qRT-PCR assays are shown in Table S2. Relative gene
155 expression was calculated using the ΔΔCt method normalized to gene expression of the
156 housekeeping gene, *tubulin alpha*.
157
158 **2.6. Gene annotation**
159 Sequence similarities of *Caulerpa lentillifera* proteins to NCBI RefSeq entries were
160 detected with BLAST searches (E-value cutoff of 10^{-5}; Camacho et al. 2009).
161 Functional domains of *C. lentillifera* proteins were searched using HMMER 3.1b2
162 (http://hmmer.org) and the Pfam-A 29.0 database (Finn et al. 2016). HMMER was
163 employed with default parameters, except for the e-value threshold '-E 1e-5' option.
164 Gene Ontology (GO) IDs were assigned using InterProScan 5.22-61.0 (Jones et al.
165 2014) with a corresponding lookup service. Assigned GO terms were projected onto
166 higher-level GO terms to plot a GO landscape using WEGO 2.0 (Ye et al. 2006) with
167 GO database version 2017-04-01. GO enrichment analysis was performed with goatools
168 under default parameters combined with the '--no_propagate_counts' option
169 (https://github.com/tanghaibao/goatools).
170
171 **3. RESULTS**
172
173 **3.1. Mapping of RNA-seq reads and annotation**
174 mRNAs were extracted from fronds and stolons, respectively, and were paired-end
175 sequenced on the Illumina HiSeq 4000 platform with the 2x150 bp sequencing protocol.
176 We obtained approximately 18 gigabases (Gb) per library from the biological triplicates
177 of fronds and stolons (Arimoto et al. 2019; Table S1). Approximately 110 Gb (2x377
178 million reads) of RNA information were obtained. This read number was much larger
179 than that of the previous study of *Caulerpa taxifolia*, in which a total of 32.8 Gb (2x178
180 million reads) were used for transcriptome assembly (Ranjan et al. 2015).
181
182 We recently decoded the approximately 26-Mb genome of *C. lentillifera* and found
183 9,311 protein-coding genes (Arimoto et al. 2019). RNA-seq reads from sporophytes
184 were assigned to genomic sequences, and 8,734 (93.8%) of the 9,311 genes were
185 supported by corresponding mRNAs (Table 1). We expect that the remaining 577 genes
186 are specifically or primarily expressed in gametes or sporophytes during the
187 reproductive stage. The following analysis focused on the 8,734 mRNA-assigned genes.

Of these, 7,173 (82.1%) were BLAST-matched, 6,498 (74.4%) Pfam protein-domain-matched, and 5,372 (61.5%) gene ontology (GO)-assigned, respectively (Table 1).

3.2. Differential expression of mRNAs in fronds and stolons

Expression profiles of the six specimens were compared using Z-scores from edgeR software. Profiles were categorized into two groups (Figure 2a). The first group included three specimens of fronds and second comprised three specimens of stolons (Figure 2a). In this analysis, globally two major expression profiles were evident. One includes genes with highly expressed mRNAs in fronds and fewer expressed in stolons. The other includes genes with highly expressed mRNAs in stolons and fewer expressed in fronds (Figure 2a).

The magnitude of differences in gene expression between fronds and stolons was determined from analysis using edgeR. The threshold of expression level change was the false discovery rate (FDR) < 0.05 and p-value < 0.05 (Figure 2b). There were 1,027 (11.8%) genes with significantly expressed mRNAs in fronds and 1,129 (12.9%) genes with significant expression in stolons. 6,578 (75.3%) genes were expressed at comparable levels in both regions (Figure 2c). Analysis using DESeq2 also detected expression changes in 421 of the 1,027 genes and 487 of the 1,129 genes, respectively. 1,027 genes preferentially expressed in fronds were characterized using BLAST searches (787 genes were annotated), Pfam domain searches (715), and GO annotation (574), respectively (Table 1). On the other hand, 1,129 genes preferentially expressed in stolons were characterized using BLAST searches (878), Pfam domain searches (793), and GO annotation (700), respectively (Table 1). Further GO annotation analysis of those obtained by edgeR and DESeq2 combined showed that 204 and 276 GO terms were assigned to genes expressed preferentially in fronds and stolons, respectively. Comparison of assigned GO terms in stolon-preferring and frond-preferring genes showed that 104 and 176 were unique to fronds and stolons, respectively, while 100 GO terms were common to both structures (Figure 3a).

Assigned GO terms were categorized as ‘cellular components (CC),’ ‘molecular function (MF),’ or ‘biological process (BP)’ (Figure 3b). Comparison of the count of assigned GO terms showed that two GO terms in CC (membrane part and intrinsic component of membrane), one GO term in MF (oxidoreductase activity), and one GO term in BP (metabolic process) were more frequently observed in fronds than in stolons

224 (Figure 3b). On the other hand, seven (cell part, intracellular, intracellular part,
225 intracellular organelle, non-membrane-bounded organelle, ribonucleoprotein complex,
226 and membrane-bounded organelle), six (small molecule binding, heterocyclic
227 compound binding, organic cyclic compound binding, carbohydrate derivative binding,
228 ion binding and structural constituent of ribosome), and three (nitrogen compound
229 metabolic process, biosynthetic process, and primary metabolic process) GO terms in
230 the CC, MF and BP categories, respectively, were found at higher frequencies in stolons
231 than in fronds (Figure 3b).
232

233 **3.3. Enriched functions based on gene ontology analysis in fronds**

234 To explore biological functions, GO terms assigned to genes preferentially expressed in
235 fronds were estimated by enrichment analysis. A GO term in the CC category assigned
236 to genes that are preferentially expressed in fronds was an "integral component of
237 membrane." Eight GO terms in the MF category assigned to fronds were cytochrome-c
238 oxidase activity, carbohydrate binding, oxidoreductase activity, and others. Six GO
239 terms in BP category also assigned to fronds were aerobic respiration, protein
240 ubiquitination, lipid metabolic process, and others (Figure 4a).
241

242 In fronds, GO terms associated with photosynthesis, including carbohydrate binding and
243 oxidation-reduction process are enriched among preferentially expressed genes. Protein
244 modification genes associated with phosphatase activity and protein ubiquitination are
245 also enriched in fronds. These enriched metabolic processes suggest that primary
246 photosynthetic products such as carbohydrates are produced mainly in fronds. Abundant
247 genes associated with "integral component of membrane" may reflect localization of
248 plastids in fronds. These results are consistent with those of Ranjan et al. (2015).
249

250 **3.4. Enriched functions based on gene ontology analysis in stolons**

251 Overviews of assigned GO terms in preferentially expressed genes in stolons were also
252 estimated by enrichment analysis. Six GO terms in the CC category were assigned to
253 genes that are preferentially expressed in stolons: associated with ribosome, intercellular
254 and nucleus. Ten assigned GO terms in the MF category included structural constituent
255 of ribosome, histone acetyltransferase activity, transcription-related activity, and others.
256 Nine GO terms in the category BP were translation, transcription, DNA metabolic
257 process, and others (Figure 4b).
258

259 Translation-associated GO terms including ribosome and rRNA processing are enriched

260 among genes preferentially expressed in stolons. GO terms that are related to
261 DNA-associated activities, e.g. DNA metabolism, histone acetylation and transcription
262 are also enriched in stolons. These enriched GO terms potentially indicate that basic
263 cellular processes such as DNA replication and translation are more active in stolons.
264 Preferentially expressed genes associated with DNA activity are consistent with a
265 previous study of *Caulerpa taxifolia* (Ranjan et al. 2015). On the other hand, genes with
266 translation activity were different. In *C. taxifolia*, genes with translation activity were
267 previously reported to be more active in fronds.
268

269 **3.5. Characterization of loci involved in plant hormone pathways**

270 Plant hormones affect growth of macroalgae as they do land plants (Bradley 1991;
271 Evans & Trewavas 1991; Lobban & Harrison 1997). In addition, genomic data of
272 *Caulerpa lentillifera* suggest that plant hormone-associated genes are expanded in this
273 alga (Arimoto et al. 2019). Therefore, we further analyzed genes involved in plant
274 hormone pathways. These genes are classified into seven categories involved in plant
275 hormone pathways for auxin, abscisic acid (ABA), jasmonic acid (JA), cytokinin (CK),
276 ethylene, brassinosteroid, and strigolactone (SL) (Table 2 and Table S3). Of two
277 *tryptophan synthase alpha subunit* (*TSA*) genes in the auxin hormone pathway, only one
278 was preferentially expressed in fronds (Table 2). This was also the case with zeaxanthin
279 epoxidase (ZEP) and abscisic acid-responsive elements-binding factor (AREB) in the
280 ABA pathway, acyl-coenzyme A oxidase (ACX) in the JA pathway, LONELY GUY
281 lysine decarboxylase (LOG) in the CK pathway, aminocyclopropane 1-carboxylate
282 synthase (ACS) in the ethylene pathway, and sterol 1/dwarf7 type C-5 sterol desaturase
283 (STE1/DWF7) in the brassinosteroid pathway. ZEP, ACX, LOG, ACS, and
284 STE1/DWF7 are enzymatic signaling molecules in each plant hormone pathway
285 whereas AREB is a transcription factor activated by ABA signaling. These results
286 indicate that a larger variety of genes involved in plant hormone biosynthesis and/or
287 signaling is preferentially expressed in fronds than in stolons.
288

289 *Protein phosphatase 2C* (*PP2C*) and *snf1-related protein kinase 2* (*SnRK2*) in the ABA
290 pathway and *more axillary growth* (*MAX*) in the SL pathway are expanded gene families
291 in the *C. lentillifera* genome, constituting 17, 22, and 32 genes, respectively (Table 2
292 and Table S3) (Arimoto et al. 2019). The number of *MAX* genes preferentially expressed
293 in either fronds or stolons was the same (five genes, Table 2 and Table S3). On the other
294 hand, four genes each in *PP2C* and *SnRK2* were preferentially expressed in fronds,
295 while three *PP2C* and two *SnRK2* were preferentially expressed in stolons, respectively

296 (Table 2). In addition, 9 *PP2C* and 16 *SnRK2* were expressed at the same level in both
297 parts. Preferential expression of one, *PP2C* (g6397), was also supported by quantitative
298 real-time PCR assays (Figure S1).
299
300 **4. DISCUSSION**
301
302 Individual *Caulerpa* plants are composed of single cells with multiple nuclei. These
303 algae, however, possess fronds, stolons, and rhizoids, structures analogous to leaves,
304 stems, and roots, respectively. In order to explore differential gene expression
305 responsible for functional differences of these structures, we performed RNA-seq
306 analysis of *Caulerpa lentillifera* fronds and stolons. Previously, Ranjan et al. (2015)
307 carried out a similar RNA-seq analysis using *Caulerpa taxifolia*. The advantage of their
308 analysis was that six structures including apex, pinnules, the rachis, basal fronds,
309 stolons, and holdfasts were used to make a transcriptomic atlas of *C. taxifolia*. We
310 subjected two structures, stolons (including rhizoids) and fronds, because these two are
311 the major structural components of *C. lentillifera*. On the other hand, an advantage of
312 the present analysis is that the gene expression profile is based on genomic information.
313 Transcriptomes assigned to 8,734 of 9,311 estimated genes were analyzed. In addition,
314 this study yielded a much greater mass of data than the previous study. The number of
315 genes categorized with GO terms in the present study was 5,372, corresponding to
316 61.5% of the 8,734 genes. In contrast, only 25% of 57,118 assembled transcripts were
317 analyzed in the previous study. In addition, bioinformatic tools used in the two studies
318 were different. Therefore, results obtained in the two studies are not entirely
319 comparable.
320
321 Ranjan et al. (2015) suggested that in *C. taxifolia*, a global pattern of transcript
322 distribution from the holdfast to the frond apex corresponds roughly to the flow of
323 genetic information in the cell. That is, DNA replication occurs in stolons, mRNAs are
324 translated in the rachis, and proteins such as kinases accumulate in apex. This is an
325 interesting idea, given that the alga is a unicellular, multinucleate organism. Our results
326 show that approximately 75.3% of 8,734 genes are expressed at a comparable level
327 throughout the individual. On the other hand, 11.8% and 12.9% of genes are
328 preferentially expressed in fronds and stolons, respectively. In addition, the present
329 study confirmed that genes associated with DNA replication are preferentially
330 expressed in stolons of *C. lentillifera* and that genes associated with phosphatase
331 activity are preferentially expressed in the fronds.

332
333 The repertory of plant hormone-associated genes preferentially expressed in fronds is
334 broader than that in stolons. However, relationships among preferentially expressed
335 plant hormone-associated genes within the two structures seem complicated, especially
336 for abscisic acid (ABA) signaling. The ABA signaling pathway is associated with
337 environmental stress responses (Vishwakarma et al. 2017). Preferential expression of
338 genes for two components of ABA signaling, *PP2C* and *SnRK2*, were observed in both
339 of fronds and stolons. However, the combinations of these genes were different in each
340 structure. PP2Cs negatively regulate SnRK2 activity in land plants (Umezawa et al.
341 2009). However, these two gene families are highly expanded in the *Caulerpa* genome
342 (Arimoto et al. 2019). Therefore, it is likely that each PP2C may suppress specific
343 SnRK2s expressed in each structure, so that genes that are expressed in a given part
344 might maintain structure-specific signaling in the other part. Further functional analyses
345 are needed to support or reject this hypothesis.
346
347 Ranjan et al. (2015) reported a basal-apical pattern of transcript distribution that
348 corresponds roughly to the flow of genetic information in the cell,
349 transcription-to-translation. In order to interpret the results, information regarding the
350 distribution of nuclei throughout the entire cell body is essential and should be
351 examined in future studies.
352
353 First, if nuclei are distributed unevenly, preferentially, or solely in stolons, this might
354 indicate that differential distribution of mRNAs, especially those for nuclear activity,
355 translation, and real functions may be interpreted as follows. Most mRNAs are
356 transcribed in the stolons and then flow through fronds toward the apexes. Because
357 different sets of mRNAs are present in stolon, rachis, and apex, there may be a special
358 transport and accumulation system in which sets of mRNAs might be localized and
359 moved to the three regions. If so, future studies should explore cellular and molecular
360 mechanisms that achieve specific mRNA flows and accumulations.
361
362 Second, if nuclei are distributed through the entire body, nuclei have to recognize their
363 locations or respond to positional signals emanating from a certain region. Then, the
364 nuclei would regulate a set of genes specific to that region. According to the previous
365 study, the pattern of gene expression in the body resembles that of the cell itself.
366 However, the present study confirmed preferential expression of mRNAs in the two
367 structures, but not a flow of genetic information corresponding to transcription to

368 translation.
369
370 *Caulerpa* nuclei are very small, approximately 2 µm in diameter (Varela-Álvarez et al.
371 2012). In addition, nuclei of various epiphytic and endophytic microbes are embedded
372 in the cell wall or cytoplasm of *Caulerpa* (Singh & Reddy 2014). Normal methods such
373 as DAPI staining cannot distinguish algal nuclei from microbial nuclei. Improved
374 methods must be developed to answer this fundamental question. However, we have
375 some information about nuclear distribution in *C. lentillifera*. In decoding the genome
376 of this alga, we isolated DNA from fronds, indicating that a considerable number of
377 nuclei are present in fronds. Fronds can regenerate the entire body when they are
378 isolated, and regenerated stolons also contain many nuclei (Guo et al. 2015). Therefore,
379 during this regeneration process, nuclei of fronds must divide and move to the newly
380 forming stolons. These circumstantial data support an even distribution of nuclei in
381 fronds and stolons.
382
383 1,027 and 1,129 genes were preferentially expressed in fronds and stolons, respectively.
384 We tried to confirm these results by other methods, including quantitative PCR. There
385 were some difficulties regarding specific amplification of target sites, because the *C.*
386 *lentillifera* genome contains paralogs of target genes with high sequence similarity.
387 Although further technical improvement may be required, we finally found that
388 sequence-specific probes such as TaqMan mitigated the problems. Nevertheless,
389 molecular functions associated with photosynthesis and environmental stress responses
390 are preferentially activated in fronds, which are the edible parts of *Caulerpa*. DNA
391 replication seems to occur more frequently in stolons. These results may help to develop
392 methods for efficiently inducing frond growth for more productive aquaculture.
393 Furthermore, genomic and transcriptomic information in this study shed additional light
394 on the biology of siphonous algae.
395
396 **5. DATA AVAILABILITY**
397 All sequence data obtained in this study are accessible in the DDBJ/EMBL/NCBI
398 database at the BioProject ID, PRJDB5734. Gene annotation data are available at
399 http://marinegenomics.oist.jp/algae/gallery/.
400
401 **ACKNOWLEDGEMENTS**
402 We thank the DNA Sequencing Section of Okinawa Institute of Science and
403 Technology Graduate University for preparing and sequencing RNA-seq libraries. We

404 thank Mr. Munekazu Mekaru and Mr. Makoto Toma (Onna Village Fisheries
405 Cooperative in Okinawa) for providing algal specimens. Dr. Steven D. Aird is
406 acknowledged for editing the manuscript. This study was supported by funding from the
407 Okinawa Institute of Science and Technology Graduate University to the Marine
408 Genomics Unit.
409

410 **References**

411

412 Abad, U., Sassi, M. & Traas, J. 2017. Flower development: from morphodynamics to
413 morphomechanics. *Philos*. *Trans*. *R*. *Soc*. *Lond*. *B Biol*. *Sci*. **372**, 20150545. doi:
414 10.1098/rstb.2015.0545.

415

416 Adl, S. M., Simpson, A. G., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J.
417 R., Bowser, S. S., Brugerolle, G., Fensome, R. A., Fredericq, S., James, T. Y., Karpov, S.,
418 Kugrens, P., Krug, J., Lane, C. E., Lewis, L. A., Lodge, J., Lynn, D. H., Mann, D. G.,
419 McCourt, R. M., Mendoza, L., Moestrup, O., Mozley-Standridge, S. E., Nerad, T. A.,
420 Shearer, C. A., Smirnov, A. V., Spiegel, F. W. & Taylor, M. F. 2005. The new higher
421 level classification of eukaryotes with emphasis on the taxonomy of protists. *J*.
422 *Eukaryot*. *Microbiol*. **52**, 399-451.

423

424 Arimoto, A., Nishitsuji, K., Higa, Y., Arakaki, N., Hisata, K., Shinzato, C., Satoh, N. &
425 Shoguchi, E. 2019. A siphonous macroalgal genome suggests convergent functions of
426 homeobox genes in algae and land plants. *DNA Res*. **26**, 183-192.

427

428 Bar, M. & Ori, N. 2014. Leaf development and morphogenesis. *Development* **141**,
429 4219-4230.

430

431 Bolger, A. M., Lohse, M., Usadel, B. 2014. Trimmomatic: a flexible trimmer for
432 Illumina sequence data. *Bioinformatics* **30**, 2114-2120.

433

434 Bradley, P. M. 1991. Plant hormones do have a role in controlling growth and
435 development of algae. *J. Phycol*. **27**, 317-321.

436

437 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. &
438 Madden, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*
439 10:421.

440

441 Coneva, V. & Chitwood, D. H. 2015. Plant architecture without multicellularity:
442 quandaries over patterning and the soma-germline divide in siphonous algae. *Front*.
443 *Plant Sci*. 6:287.

444

445 Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P.,
446 Chaisson, M. & Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner.
447 *Bioinformatics* **29**, 15-21.
448
449 Draisma, S. G., van Reine, W. F., Sauvage, T., Belton, G. S., Gurgel, C. F., Lim, P. E. &
450 Phang, S. M. 2014. A re-assessment of the infra-generic classification of the genus
451 *Caulerpa* (Caulerpaceae, Chlorophyta) inferred from a time-calibrated molecular
452 phylogeny. *J. Phycol*. **50**, 1020-1034.
453
454 Evans, L. V. & Trewavas, A. J. 1991. Is algal development controlled by plant-growth
455 substances? *J. Phycol*. **27**, 322-326.
456
457 Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter,
458 S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J. & Bateman,
459 A. 2016. The Pfam protein families database: towards a more sustainable future.
460 *Nucleic Acids Res*. **44**, D279-D285.
461
462 Galil, B. S. 2007. Loss or gain? Invasive aliens and biodiversity in the Mediterranean
463 Sea. *Mar. Pollut. Bull*. **55**, 314-322.
464
465 Guo, H., Yao, J., Sun, Z. & Duan, D. 2015. Effects of salinity and nutrients on the
466 growth and chlorophyll fluorescence of *Caulerpa lentillifera*. *Chin. J. Oceanol. Limnol*.
467 **33**, 410-418.
468
469 Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H.,
470 Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A.,
471 Scheremetjew, M., Yong, S. Y., Lopez, R. & Hunter, S. 2014. InterProScan 5:
472 genome-scale protein function classification. *Bioinformatics* **30**, 1236-1240.
473
474 Kaplan, D.R. & Hagemann, W. 1991. The relationship of cell and organism in vascular
475 plants. *Bioscience* **41**, 693-703.
476
477 Lobban, C. S. & Harrison, P. S. 1997. Seaweed Ecology and Physiology. Cambridge
478 University Press, Cambridge.
479

480 Love, M. I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and
481 dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550.
482
483 Mary, A., Mary, V., Lorella, A. & Matias J. R. 2009. Rediscovery of naturally occurring
484 seagrape Caulerpa lentillifera from the Gulf of Mannar and its mariculture. *Curr*. *Sci*. **97**,
485 1418-1420.
486
487 Mashiguchi, K., Tanaka, K., Sakai, T., Sugawara, S., Kawaide, H., Natsume, M.,
488 Hanada, A., Yaeno, T., Shirasu, K., Yao, H., McSteen, P., Zhao, Y., Hayashi, K., Kamiya,
489 Y. & Kasahara, H. 2011. The main auxin biosynthesis pathway in *Arabidopsis*. *Proc*.
490 *Natl*. *Acad*. *Sci*. *USA*. **108**, 18512-18517.
491
492 McCarthy, D. J., Chen, Y. & Smyth, G. K. 2012. Differential expression analysis of
493 multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids*
494 *Res*. **40**, 4288-4297.
495
496 Mozzachiodi, R., Scuri, R., Roberto, M. & Brunelli, M. 2001. Caulerpenyne, a toxin
497 from the seaweed *Caulerpa taxifolia*, depresses afterhyperpolarization in invertebrate
498 neurons. *Neuroscience* **107**, 519-526.
499
500 Paul, N. A., Neveux, N., Magnusson, M & de Nys, R. 2014. Comparative production
501 and nutritional value of “sea grapes” — the tropical green seaweeds *Caulerpa*
502 *lentillifera* and *C*. *racemosa*. *J*. *Appl*. *Phycol*. **26**, 1833-1844.
503
504 Ranjan, A, Townsley, B. T., Ichihashi, Y., Sinha, N. R. & Chitwood, D. H. 2015. An
505 intracellular transcriptomic atlas of the giant coenocyte *Caulerpa taxifolia*. *PLoS Genet*.
506 **11,** e1004900.
507
508 Singh, R. P. & Reddy, C. R. 2014. Seaweed-microbial interactions: key functions of
509 seaweed-associated bacteria. *FEMS Microbiol*. *Ecol*. **88**, 213-230.
510
511 Umen, J. G. 2014. Green algae and the origins of multicellularity in the plant kingdom.
512 *Cold Spring Harb*. *Perspect*. *Biol*. **6**, a016170.
513
514 Umezawa, T., Sugiyama, N., Mizoguchi, M., Hayashi, S., Myouga, F.,
515 Yamaguchi-Shinozaki, K., Ishihama, Y., Hirayama, T. & Shinozaki, K. 2009. Type 2C

516 protein phosphatases directly regulate abscisic acid-activated protein kinases in
517 *Arabidopsis*. *Proc*. *Natl*. *Acad*. *Sci*. *USA*. **106**, 17588-17593.
518
519 Varela-Álvarez, E., Garreta, A. G., Lluch, J. R., Soler, N. S., Serrao, E. A. & Siguán, M.
520 A. R. 2012. Mediterranean species of *Caulerpa* are polyploid with smaller genomes in
521 the invasive ones. *PLoS One*. **7**, e47728.
522
523 Verma, V., Ravindran, P. & Kumar, P. P. 2016. Plant hormone-mediated regulation of
524 stress responses. *BMC Plant Biol*. **16**, 86. doi: 10.1186/s12870-016-0771-y.
525
526 Vishwakarma, K., Upadhyay, N., Kumar, N., Yadav, G., Singh, J., Mishra, R. K., Kumar,
527 V., Verma, R., Upadhyay, R. G., Pandey, M. & Sharma, S. 2017. Abscisic acid signaling
528 and abiotic stress tolerance in plants: a review on current knowledge and future
529 prospects. *Front*. *Plant Sci*. **8**, 161. doi: 10.3389/fpls.2017.00161.
530
531 Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R.,
532 Bolund, L. & Wang, J. 2006. WEGO: a web tool for plotting GO annotations. *Nucleic*
533 *Acids Res*. **34**, W293-W297.
534
535

536 **Figure Legends**

537

538 **Figure 1. The siphonous green alga, *Caulerpa lentillifera.*** (a) Cultivated *Caulerpa*
539 *lentillifera*. This alga consists of many grape-like vesicles connected by stolons.
540 Filamentous rhizoids also project from the stolons. The entire alga is composed of a large
541 single cell with many nuclei. Scale bar, 20 mm. A cross section of a *C. lentillifera* stolon
542 (b) and a frond vesicle (c). Organelles are localized peripherally in the body, where many
543 fibrous components are found in the central part of the stolon. Scale bar, 200 µm.

544

545 **Figure 2. Comparison of gene expression profiles in *Caulerpa lentillifera* stolons**
546 **and fronds.** (a) Comparison of gene expression profiles among three individuals A, B,
547 and C. Expression level was categorized as high (red) or low (blue). Sources of RNA
548 are shown on the X axis. The Y axis corresponds to expression levels of each gene. (b)
549 Changes of gene expression between fronds and stolons in *Caulerpa lentillifera*. Genes
550 highly expressed in fronds and stolons have positive and negative y-axis values,
551 respectively. Differentially expressed genes that have a false discovery rate < 0.05 are
552 indicated with red circles. Black circles are genes expressed at comparable levels in
553 both structures, judged by this threshold value. (c) The pie chart shows the number of
554 differentially expressed genes (DEG). The threshold of DEG detection was the false
555 discovery rate.

556

557 **Figure 3. Gene ontology overview of preferentially expressed genes in *Caulerpa***
558 ***lentillifera*.** (a) A Venn diagram showing GO terms assigned to loci preferentially
559 expressed in fronds (cyan) and stolons (magenta). (b) A histogram of GO terms in
560 stolons and fronds. GO terms were assigned to the intersection of edgeR and DESeq2
561 analyses. The X axis shows assigned GO terms and the Y axis shows the percentages of
562 GO-assigned genes in each structure. CC, cellular component. MF, molecular function.
563 BP, biological process.

564

565 **Figure 4. Gene ontology enrichment analysis comparing stolons and fronds in**
566 ***Caulerpa lentillifera*.** Enriched gene ontology (GO) terms pertaining to fronds (a) and
567 stolons (b) are shown in histograms. Each bar corresponds to p-values converted to
568 logarithms. The numbers shown at the tips of each bar equal the number of genes
569 assigned to that GO term. CC, cellular component. MF, molecular function. BP,
570 biological process.

571 **Supplementary Figure Legends**

572
573 **Figure S1. Validation of gene expression levels using quantitative real-time PCR.**
574 Differences of gene expression levels of g6397, which is a component of abscisic acid
575 signaling, between stolons and fronds corresponded to the estimation based on
576 RNA-seq data. Error bars show standard deviation.
577
578

Table 1. Statistical summary of genes expressed in *Caulerpa lentillifera*.

Features	Total	Frond[1]	Stolon[1]
Number of genes[2]	9,311		
% genes have RNA-seq support	8,734/9,311 (93.8%)	1,027/8,734 (11.8%)	1,129/8,734 (12.9%)
Number of BLAST match genes[3]	7,173/8,734 (82.1%)	787/1,027 (76.6%)	878/1,129 (77.8%)
Number of Pfam domain match genes[3]	6,498/8,734 (74.4%)	715/1,027 (69.6%)	793/1,129 (70.2%)
Number of GO assigned genes[3]	5,372/8,734 (61.5%)	574/1,027 (55.9%)	700/1,129 (62.0%)

1) The number of differentially expressed genes

2) Arimoto et al. 2019

3) Percentage based on the RNA-seq supported genes

579

Table 2. Differential expression of plant hormone-associated genes in *Caulerpa lentillifera*.

Pathways	Genes	Annotations	# Genes[1]	# Frond DEGs	# Stolon DEGs
Auxin	TSA	Tryptophan synthase alpha subunit	2	1 (1)	0
Abscisic acid	ZEP	Zeaxanthin epoxidase	2	1	0
(ABA)	PP2C	Protein phosphatase 2C	17	4 (2)	3 (1)
	SnRK2	Serine/threonine-protein kinase	22	4 (1)	2
	AREB	ABSCISIC ACID-INSENSITIVE 5-like protein	2	1 (1)	0
Jasmonic acid (JA)	ACX	Acyl-CoA oxidase	3	2	0
Cytokinin (CK)	LOG	Cytokinin nucleoside 5'-monophosphate phophoribohydrolase	2	1 (1)	0
Ethylene	ACS	1-aminocyclopropane-1-carboxylate synthase	2	1 (1)	0
Brassinosteroids	STE1/DWF7	Delta7 sterol C-5 desaturase	1	1	0
Strigolactone (SL)	MAX	Carotenoid cleavage dioxygenase	32	5 (2)	5 (2)

1) Arimoto et al. 2019

DEG, differentially expressed genes.

Numbers in parentheses correspond to numbers of DEGs detected by both of two different algorithms.

580

581

Table S1. Sequence data summary[1].

Library types	Insert size	Read length	Raw data		Filtered data			Source
			Number of reads	Total amount of data	Number of reads	Total amount of data	Uniquely mapped reads	
Paired-end	200 bp	2x150 bp	2x61.4 million reads	18.1 Gb	2x59.1 million reads	17.4 Gb	96.23%	Individual A frond
Paired-end	200 bp	2x150 bp	2x60.9 million reads	18.1 Gb	2x58.9 million reads	17.4 Gb	96.67%	Individual A stolon
Paired-end	200 bp	2x150 bp	2x64.7 million reads	19.2 Gb	2x62.7 million reads	18.6 Gb	93.69%	Individual B frond
Paired-end	200 bp	2x150 bp	2x61.4 million reads	18.2 Gb	2x59.7 million reads	17.6 Gb	95.57%	Individual B stolon
Paired-end	200 bp	2x150 bp	2x70.1 million reads	20.8 Gb	2x67.9 million reads	20.1 Gb	94.67%	Individual C frond
Paired-end	200 bp	2x150 bp	2x58.6 million reads	17.3 Gb	2x56.8 million reads	16.8 Gb	96.33%	Individual C stolon

1) Arimoto et al. 2019

582

Table S2. Primer and probe sequences for quantitative real-time PCR.

Target genes	Gene ID	Primers	Sequences (5' to 3')
tubulin alpha	g4057	Forward	GCCAAGCCGGTATCCAAGTC
		Reverse	GCGTTAAATGCGTCGTCTTCG
		Probe	CCAACCCGACGGCCAACTACCATCTGA
PP2C	g6397	Forward	GCCGTGGATTTTACGTCGTCT
		Reverse	TGCACTGGATGCGTCCTG
		Probe	TCTGCTGCGATTCATGTTACTTCACGATGC

583

584

Table S3. Differential expression levels of plant hormone associated genes.

Pathway	Annotation	Gene ID	Parts	logFC	FDR
Auxin	TSA	g7447*	Frond	0.45006	9.43E-03
Abscisic acid	ZEP	g2372	Frond	0.75646	4.35E-04
	PP2C	g508	Frond	0.38404	1.15E-02
	PP2C	g4317*	Frond	0.37332	4.85E-02
	PP2C	g4936	Frond	0.46760	1.07E-03
	PP2C	g8444*	Frond	0.37847	2.84E-02
	PP2C	g2460	Stolon	-0.31301	4.04E-02
	PP2C	g6397*	Stolon	-0.38438	8.68E-03
	PP2C	g8298	Stolon	-0.58589	1.91E-04
	SnRK2	g3174*	Frond	0.38396	3.59E-02
	SnRK2	g6486	Frond	1.40935	9.82E-20
	SnRK2	g7448	Frond	0.46533	1.32E-03
	SnRK2	g8496	Frond	1.50088	9.94E-12
	SnRK2	g3093	Stolon	-0.40972	2.26E-02
	SnRK2	g8008	Stolon	-0.47855	1.69E-03
	AREB	g6338*	Frond	0.46958	1.64E-02
Jasmonic acid	ACX	g3712	Frond	0.69534	2.11E-06
	ACX	g8317	Frond	0.91244	4.36E-11
Cytokinin	LOG	g7625*	Frond	0.69523	1.91E-06
Ethylene	ACS	g3569*	Frond	0.54005	1.01E-04
Brassinosteroids	STE1/DWF7	g6980	Frond	0.31979	3.55E-02
Strigolactone	MAX	g334	Frond	0.65272	1.53E-03
	MAX	g342*	Frond	0.61833	2.04E-02
	MAX	g499	Frond	0.50666	6.37E-03
	MAX	g4218*	Frond	0.74305	3.61E-07
	MAX	g5131	Frond	0.70679	6.68E-06
	MAX	g335	Stolon	-0.41606	1.18E-02
	MAX	g3820	Stolon	-0.41255	2.41E-02
	MAX	g3827*	Stolon	-0.46190	5.72E-03
	MAX	g3830*	Stolon	-0.57019	1.91E-03
	MAX	g8395	Stolon	-0.47942	7.40E-04

* Gene expression differences were detected by both algorithms.

585