

Doctoral Dissertation

博士論文

**Studies on parallel evolution of multicellular traits
by comparative embryology and genomics of
the volvocine green algae**

(ボルボックス系列緑藻の比較発生学的解析と比較ゲノム解析による
多細胞形質の平行進化に関する研究)

A Dissertation Submitted for the Degree of Doctor of Science

December 2020

令和 2 年 12 月 博士（理学）申請

Department of Biological Sciences, Graduate School of Science,

The University of Tokyo

東京大学 大学院理学系研究科

生物科学専攻

Shota Yamashita

山下 翔大

Table of Contents

Abstract	1
Chapter 1. General Introduction	6
Figures	16
Chapter 2. Analysis of Cellular Mechanism of Spheroidal Colony Formation during Embryogenesis in <i>Astrephomene</i>	20
2.1. Introduction	21
2.2. Materials and Methods	22
2.2.1. Strains and culture conditions	22
2.2.2. Phylogenetic analysis of the new strain based on coding region of <i>rbcL</i> gene	23
2.2.3. Light microscopy time-lapse imaging	25
2.2.4. Western blot analysis	25
2.2.5. Indirect immunofluorescence microscopy	26
2.3. Results	28
2.3.1. Establishment of a new strain of <i>Astrephomene gubernaculifera</i>	28
2.3.2. Time-lapse imaging of embryogenesis in <i>Astrephomene gubernaculifera</i> and <i>Eudorina</i> sp.	28
2.3.3. Indirect immunofluorescence microscopy of basal bodies and nuclei	31
2.4. Discussion	33
2.5. Tables and Figures	35
Chapter 3. Analysis of Flattened Colony Formation during Embryogenesis in <i>Gonium</i> and <i>Tetrabaena</i>	48
3.1. Introduction	49
3.2. Materials and Methods	51
3.2.1. Establishment of a new strain of <i>Gonium pectorale</i>	51

3.2.2. Strains and culture conditions	52
3.2.3. Light microscopy time-lapse imaging	52
3.2.4. Indirect immunofluorescence microscopy	53
3.3. Results	55
3.3.1. Establishment of a new strain of <i>Gonium pectorale</i>	55
3.3.2. Time-lapse imaging of embryogenesis in <i>Gonium pectorale</i> and <i>Tetrabaena socialis</i>	55
3.3.3. Indirect immunofluorescence microscopy of basal bodies, nuclei, and microtubules	57
3.4. Discussion	60
3.5. Tables and Figures	62
Chapter 4. Analysis of Nuclear Genome of <i>Astrephomene</i>	75
4.1. Introduction	76
4.2. Materials and Methods	77
4.2.1. Strain and culture conditions	77
4.2.2. De novo genome sequencing and assembly	77
4.2.3. Prediction of gene models of <i>Astrephomene gubernaculifera</i> using RNA-seq data	79
4.2.4. Prediction of gene models of <i>Yamagishiella unicocca</i> and <i>Eudorina</i> sp.	79
4.2.5. Identification and analysis of orthologs of genes involved in embryogenesis in <i>Volvox carteri</i>	80
4.2.6. Identification and analysis of VARL genes	81
4.2.7. Identification and analysis of MMP and pherophorin genes	82
4.2.8. Separation of somatic cells and reproductive cells	84
4.2.9. Cell-type RNA-seq analysis	84

4.2.10. Comparison of expression pattern between somatic and reproductive cells	86
4.2.11. Identification and analysis of somatic- and reproductive-specific genes	87
4.2.12. Availability of data	88
4.3. Results and Discussion	89
4.3.1. <i>De novo</i> genome sequence of <i>Astrephomene gubernaculifera</i>	89
4.3.2. Orthologous genes involved in embryogenesis in <i>Volvox carteri</i>	89
4.3.3. <i>RLS1/rlsD</i> homolog and other VARL genes	90
4.3.4. ECM genes	91
4.3.5. Identification of differentially expressed genes based on cell-type RNA-seq analysis	92
4.3.6. Germ–soma differentiation in gene expression patterns	94
4.3.7. Candidate genes for regulators of germ–soma differentiation	97
4.4. Tables and Figures	99
Chapter 5. General Discussion	173
Acknowledgements	180
References	183

Abstract

The evolution from unicellular to multicellular organisms has occurred many times in the history of life and many eukaryotic lineages have acquired their multicellular complexity, such as germ–soma differentiation or three-dimensional body plans, independently. However, molecular bases for initial steps of such evolution have remained unresolved. Volvocine green algae, which range from the unicellular *Chlamydomonas* to the multicellular *Volvox* with a germ–soma division of labor, include various intermediate forms of multicellularity and provide a suitable model for investigation into the evolution of multicellular organisms. Within this group, two independent lineages, Volvocaceae and goniacean *Astrephomene*, exhibit parallel evolution of spheroidal colonies from flattened colonies, germ–soma differentiation and expansion of extracellular matrix (ECM). *Astrephomene* produces spheroidal colonies with posterior somatic cells. The most outstanding feature that distinguishes *Astrephomene* from Volvocaceae is lack of inversion during embryogenesis; the volvocacean embryo undergoes inversion after successive divisions to form a spheroidal colony with flagella outside. Though the mechanisms of inversion at the molecular and cellular levels in volvocacean algae have been assessed in detail, particularly in the model species, *Volvox carteri*, embryogenesis in *Astrephomene* has not been subjected to such investigations. Moreover, the molecular bases involved in the evolution of germ–soma differentiation and expansion of ECM also have been studied mainly in Volvocaceae, without taking *Astrephomene* into consideration. The present thesis aimed to analyze the spheroidal colony formation, germ–soma differentiation and expansion of ECM in *Astrephomene* and to unveil the parallel evolution of these multicellular traits within volvocine green algae in molecular, cellular and genomic levels.

In Chapter 2, I conducted a developmental analysis of *Astrephomene gubernaculifera* by light microscopy time-lapse imaging using an actively growing culture of a newly established strain and compared it with volvocacean *Eudorina* sp. During the successive divisions in *A. gubernaculifera* embryogenesis, gradual rotation of daughter protoplasts resulted in movement of their apical ends (position of basal bodies) toward the embryonic posterior, forming a

convex-to-spheroidal cell sheet with the apical ends of protoplasts on the outside. Differentiation of the posterior somatic cells from the embryo's periphery was traced based on the analysis of cell lineages during embryogenesis. In contrast, in *Eudorina* sp., rotation of daughter protoplasts did not occur during successive cell divisions. Instead, inversion with the formation of acute chloroplast ends occurred after the successive cell divisions to form a spheroidal embryo. Indirect immunofluorescence microscopy of basal bodies and nuclei verified this difference between *A. gubernaculifera* and *Eudorina* sp. in the movement of embryonic protoplasts. These results suggest different tactics for spheroidal colony formation between the two lineages: rotation of daughter protoplasts during successive cell divisions in *Astrephomene*, and inversion after cell divisions in Volvocaceae. However, the evolutionary pathways of the two tactics for formation of spheroidal colony are unclear, since the embryogenesis of extant volvocine genera with ancestral flattened colonies, such as *Gonium* and *Tetrabaena*, also has not previously been investigated in detail.

In Chapter 3, I conducted time-lapse imaging by light microscopy and indirect immunofluorescence microscopy with staining of basal bodies, nuclei, and microtubules to observe embryogenesis in *Gonium pectorale* and *Tetrabaena socialis*, which form 16-celled or 4-celled flattened colonies, respectively. In *G. pectorale*, a cup-shaped cell layer of the 16-celled embryo underwent gradual expansion after successive cell divisions, with the apical ends of protoplasts at the square embryo's periphery separated from each other. In *T. socialis*, on the other hand, there was no apparent expansion of the daughter protoplasts in 4-celled embryos after successive cell divisions, however the two pairs of diagonally opposed daughter protoplasts shifted slightly and flattened after hatching. Neither of these two species exhibited rotation of daughter protoplasts during successive cell divisions as in *Astrephomene* or the formation of acute chloroplast ends of daughter protoplasts as in volvocacean inversion. These results indicate that the ancestor of *Astrephomene* might have newly acquired the rotation of daughter protoplasts after it diverged from the ancestor of *Gonium*, while the ancestor of Volvocaceae might have newly acquired the formation of acute chloroplast

ends to complete inversion after divergence from the ancestor of Goniaceae (*Gonium* and *Astrephomene*).

In Chapter 4, I investigated *de novo* whole genome sequence of *A. gubernaculifera* to resolve molecular bases of parallel evolution of multicellular traits in the volvocine algae. In *V. carteri*, tandem duplication and neofunctionalization of the ancestral transcription factor gene (*RLS1/rlsD*) might have occurred to evolve a master regulator of somatic cell differentiation, *regA*. However, the present genome data demonstrated that *A. gubernaculifera* might not have undergone tandem duplication of *RLS1/rlsD* homolog or acquisition of *regA*-like gene. Pherophorin and matrix metalloprotease (MMP) genes constituting ECM gene families are expanded in *V. carteri* and are thought to be related to the expansion of ECM. However, these genes are not expanded in *A. gubernaculifera*. Thus, the two multicellular traits might have evolved by the modifications of the different molecular bases between *V. carteri* and *A. gubernaculifera*. On the other hand, RNA-seq analysis of isolated somatic and reproductive cells of *A. gubernaculifera* revealed the growth repression of somatic cells via downregulation of photosynthetic pathways like *V. carteri*. Three somatic-specific genes encoding putative MYB and RWP-RK transcription factors were also identified, which might regulate somatic cell differentiation in *A. gubernaculifera*. Thus, germ–soma differentiation in *Volvox* and *Astrephomene* might have originated from different genetic evolution that resulted in similar gene expression patterns in differentiated cell-types which have evolved independently.

In the present thesis, I resolved the parallel evolution of two multicellular traits within the volvocine green algae: the evolution of morphologically similar spheroidal colonies by the acquisition of different cellular mechanisms during embryogenesis, and the evolution of germ–soma differentiation with similar gene expression patterns by the acquisition of different regulatory genes. The present results demonstrated the diversity of cellular and molecular bases involved in the initial steps for evolution of multicellular traits even among closely related groups within a single

multicellular lineage. Further studies on the multicellular traits in *Astrephomene* based on my genome data and the evolution of multicellular traits within volvocine algae would provide the understanding for the diversity or general principles of the evolution of multicellularity in molecular, genetic and genomic levels.

Chapter 1

General Introduction

In the history of life on earth, there have been major transitions in evolution, where the individual lower-level units are integrated into a higher-level unit (Smith and Száthmary, 1995). This major transitions are thought to include following events: the compartmentalization of replicating molecules (birth of the first cell), the coalescence of replicating molecule to chromosomes, the use of DNA as genes and protein as enzymes instead of RNA solely, the union of prokaryotic cells to eukaryotic cells containing mitochondria and chloroplasts, the evolution of sex and sexual reproduction with exchange of genetic information among a sexual population, the union of individual cells to multicellular organisms, and the establishment of social groups composed of multicellular individuals (Smith and Száthmary, 1995; Queller, 2000; Grosberg and Strathmann, 2007). These transition have brought the complexity and diversity of the life on earth up to the present. The evolution of multicellular organisms from unicellular ancestors is one of such major transitions, which occurred at least 25 times independently in a broad range of prokaryotic and eukaryotic lineages (Grosberg and Strathmann, 2007) (Fig. 1-1).

Multicellular organisms have evolved three-dimensional body plans and various differentiation of cells or tissues many times, depending on their lifestyles. Among Metazoa (animals), Bilateria, Cnidaria and Ctenophora have acquired bodies with anterior–posterior axis and bilateral/radial symmetry with muscles, nervous system for prominent motility (Philippe et al., 2009). Though the other two phyla in Metazoa, Porifera and Placozoa, have simpler multicellular bodies, both have several differentiated cell types and body plans with polarity (Martindale, 2005; Philippe et al., 2009; Smith et al., 2014). The spheroidal multicellular forms with cilia or flagella of cells toward outside of the body, which resemble larvae in Porifera (Maldonado, 2005), are also seen in other multicellular protists, such as choanoflagellates (Fairclough et al., 2010; Dayel et al., 2011; Brunet et al., 2019) and golden algae (Chrysophyceae) (Pusztai and Škaloud, 2019; Škaloud et al., 2020). On the other hand, green, red and brown algae, as well as land plants, have evolved similar filamentous body plans and leaf-like body plans composed of parenchymatous tissues independently

(Niklas, 2000; Niklas and Kutschera, 2009). These multicellular lineages exhibit clonal multicellularity, in which the multicellular bodies are developed from a single cell by embryogenesis, which enables the complicated regulation of organismal polarity and cellular differentiation by asymmetric cell divisions (Grosberg and Strathmann, 2007; Rensing, 2016). On the other hand, some eukaryotic lineages also evolved aggregative multicellularity, in which individual cells join and aggregate to form multicellular fruiting bodies or sorocarps for production of spores or cysts (Grosberg and Strathmann, 2007; Du et al., 2015b). The multicellularity of fungi is a modified type of clonal multicellularity; the fruiting bodies with three-dimensional structures for protection and dispersion of spores develop from hyphae (Nagy et al., 2018).

The evolution of such body plans and cellular differentiation has been studied mainly by the comparison of multicellular lineages and their closest unicellular lineages, like Metazoa and choanoflagellates, or land plants and green algae. Recent genetic and genomic studies have revealed that some of genetic toolkits for multicellularity, such as cell adhesion proteins and transcription factors, already existed in the genome of unicellular ancestors (King et al., 2008; Mukherjee et al., 2009; Fairclough et al., 2013; Rensing, 2016). For evolution of animal cell differentiation, two major hypotheses have been proposed: temporal switching of different cell phenotypes in the life cycle of unicellular ancestors co-opted to differentiation between cell-types (temporal-to-spatial transition hypothesis) (Mikhailov et al., 2009), or loss of function from multifunctional ancestral cell-type occurred differentially between cell-types after the evolution of multicellularity (division of labor hypothesis) (Arendt, 2008). Both hypotheses have been supported from mainly morphological and partly genetic data (Brunet and King, 2017). However, the processes or the initial steps of such evolution of cellular differentiation or complex body plans are difficult to resolve in detail, as many multicellular lineages have lost the intermediates between complicated multicellular organisms and their closest unicellular ancestors.

One of the best model groups for investigation into such initial evolution of multicellularity

is volvocine green algae (Fig. 1-2) (Kirk, 2005). The volvocine green algae, which consist of *Volvox* and its closely related genera, include organisms of various intermediate stages of multicellular complexity, ranging from unicellular *Chlamydomonas* to multicellular *Volvox* with germ–soma division of labor. The multicellular group are estimated to have diverged from the unicellular ancestor approximately 200 million years ago (Herron et al., 2009), which is more recent than that of animal or land plants, and *Chlamydomonas*-like cell structure is well conserved within this group. Genome studies on the volvocine algae have been intensively promoted in recent years and whole genome sequence data of *Chlamydomonas reinhardtii* (Merchant et al., 2007), *Volvox carteri* (Prochnik et al., 2010), *Gonium pectorale* (Hanschen et al., 2016), *Tetrabaena socialis* (Featherston et al., 2018), *Yamagishiella unicocca* and *Eudorina* sp. (Hamaji et al., 2018) are currently available. The diverse tools for molecular biology such as transformation method are also available in *C. reinhardtii* and *V. carteri* (Kindle et al., 1989; Schiedlmeier et al., 1994) and possibly applicable to the other species (Lerche and Hallmann, 2009, 2013, 2014). Thus, this group is a suitable model for stepwise evolution of multicellular complexity in genetic and genomic levels.

Within volvocine green algae, there is an evolutionary tendency of multicellular complexity: the larger number of cells composing a vegetative colony (an individual), the more derivative features the organism has (Kirk, 2005). In the early-branched genera with fewer cell numbers, such as *Tetrabaena* with 4-celled colony or *Gonium* with 8-, 16- or 32-celled colony, cells are connected by a thin layer of extracellular matrix (ECM) and form a flattened colony, and every cell in a colony undergoes embryogenesis to form a daughter colony. On the other hand, in phylogenetically more derived genera, the colony has three-dimensional, spheroidal shape and composed of 16, 32 or more cells embedded in ECM that is usually expanded and/or morphologically complicated. Moreover, three of these genera, *Volvox*, *Pleodorina* and *Astrephomene*, have a colony composed of 64 or more cells with germ–soma differentiation, in which a part of cells (reproductive cells) undergo embryogenesis and the other cells (somatic cells) do not perform reproduction (reviewed in Kirk,

2005). Recent studies suggested that these traits might have evolved several times within the volvocine algae; the acquisition of spheroidal colony with expanded ECM and somatic cells might have occurred at least twice in ancestors of Volvocaceae and *Astrephomene* (Fig. 1-2) (Nozaki and Ito, 1994; Herron and Michod, 2008; Herron et al., 2009).

The evolution of these traits associated with increase of cell number can be explained theoretically. In fluid dynamics, a “puller” squirmer, which generate a thrust in front of the organism like unicellular *Chlamydomonas* or a flattened colony with anterior flagella in *Gonium*, has lower swimming efficiency as the Reynolds number (proportional to the organismal size) increases, in contrast to a squirmer with an equally distributed thrust like *Volvox* (Wang and Ardekani, 2012). The spheroidal colonies of volvocine green algae also evolved advantageous features for their swimming, such as asymmetrical flagellar apparatus aligned almost parallel to anterior–posterior axis (Hoops and Floyd, 1983; Hoops, 1993; Hoops et al., 1994; Hoops, 1997) and anterior–posterior division of labor for suitable swimming with phototaxis (Ueki et al., 2010). On the other hand, the evolution of germ–soma differentiation has been explained by the fitness models based on the trade-offs between viability (survival) and fecundity (reproduction), in which specialization in viability (soma) and specialization in fecundity (germ) allow the maximum fitness of the cell group (the multicellular individual) rather than the unspecialized state, when the number of cells in a group increases (Michod, 2006; Michod et al., 2006). The specialization in viability by somatic cells is realized by the maintenance of locomotion of a colony in volvocine green algae; as the number of cells and a time required for embryogenesis increases, the colony cannot prevent sinking away from the euphotic zone during embryogenesis, without differentiation of undivided somatic cells (Koufopanou, 1994; Solari et al., 2006a). The maintenance of locomotion of a colony by somatic cells is also important for the molecular transport of nutrients for growth and reproduction of germ cells (Solari et al., 2006b).

Among volvocine green algae, *Volvox carteri* has been studied as a model species for its

relatively simple multicellularity (Kirk, 1998). A vegetative colony (or a spheroid) of *V. carteri* consists of ~2,000 somatic cells and ~16 reproductive cells (gonidia). The somatic cells are small and biflagellate and distributed throughout the surface of the spheroidal colony, whereas the gonidia are much larger than somatic cells and have no flagella. These cells are embedded in ECM. The ECM is so expanded in *Volvox* that more than 99% of the volume of a mature vegetative colony of *Volvox* consists of ECM. The ECM is not only expanded, but also structurally complex compared to cell wall of *C. reinhardtii* (Kirk et al., 1986). In asexual life cycle (Fig. 1-3a), each reproductive cell grows and undergoes embryogenesis to form a daughter colony. On the other hand, somatic cells do not proliferate and undergo programmed cell death (Kirk, 1998).

The formation of the spheroidal colony during embryogenesis has been investigated extensively in *V. carteri* at the cellular and molecular levels (Fig. 1-3b). Each reproductive cell or gonidium undergoes successive cell divisions (“palintomy” or rapid multiple fissions without cell growth, Desnitski, 1995; Herron et al., 2010) to form a spheroidal-shaped embryo or plakea composed of a single cell layer. Immediately after the successive divisions, the cell layer is inside out compared with the adult configuration—the apical ends of daughter protoplasts of the embryo, in which basal bodies are localized and flagella are formed, are oriented toward the interior of the plakea. Therefore, the embryo undergoes inversion, during which the cell layer is inverted to form a spheroidal daughter colony with the apical ends and flagella of daughter protoplasts positioned outside. This process enables appropriate locomotion of spheroidal colonies of *V. carteri*. The change in shape of daughter protoplasts, from pear-shaped to flask-shaped with stalks in their chloroplast ends (opposite to apical ends) and movement of protoplasts relative to cytoplasmic bridges, which interconnect protoplasts, produce the mechanical force for folding of the cell sheet during inversion (Viamontes and Kirk, 1977; Viamontes et al., 1979; Green et al., 1981). Several genes necessary for progress of inversion, *InvA*, *InvB* and *InvC*, and their orthologs in *C. reinhardtii*, *IAR1*, *IBR1* and *ICR1* have been identified so far (Nishii et al., 2003; Ueki and Nishii, 2008, 2009;

Nishii and Miller, 2010).

Inversion is common among volvocacean species (Nozaki and Ito, 1994; Nozaki et al., 2000; Herron and Michod, 2008). In *Volvox* species, there are two types of inversion, type A and type B, in which the patterns of changes in cell shapes, patterns of cell sheet bending and timing of the opening of the phialopore are different (Hallmann, 2006; Höhn and Hallmann, 2011). Compared with that in *Volvox*, inversion in other volvocacean genera, such as *Pandorina*, *Eudorina* and *Pleodorina*, is relatively simple: the cup-shaped cell layer of an embryo inverts to form a spheroidal colony (Gottlieb and Goldstein, 1977; Marchant, 1977; Fulton, 1978; Hallmann, 2006; Höhn and Hallmann, 2016). In addition, *Platydorina* perform intercalation after inversion to form a perpendicularly flattened colony (Iida et al., 2011). The change in cell shape with the development of stalks or acute chloroplast ends and the location of cytoplasmic bridges at the chloroplast ends during inversion have also been reported in some species of *Volvox* (Pickett-Heaps, 1970; Kelland, 1977; Ireland and Hawkins, 1981; Höhn and Hallmann, 2011) and other volvocacean genera (Marchant, 1977; Fulton, 1978; Iida et al., 2011; Höhn and Hallmann, 2016), which suggests that the volvocacean genera use common cellular mechanisms for inversion.

As well as spheroidal colony formation, germ–soma differentiation in *V. carteri* has also been investigated in molecular level, which resulted in the discovery of several key genes. In early embryogenesis in *V. carteri*, a gonidium undergoes successive cell divisions to form an embryo. After the fifth cell division, 16 cells in the anterior hemisphere of the embryo (the posterior region of a spheroidal colony after inversion) divide asymmetrically, which results in the formation of 16 large cells and 16 small cells (Green and Kirk, 1981) (Fig. 1-3b). This asymmetrical division is regulated by two interacting chaperones, GlcA and Hsp70A (Miller and Kirk, 1999; Cheng et al., 2005). Then, the small cells express *regA*, a master regulator for the differentiation of somatic cells which encodes a putative transcription factor with a DNA-binding SAND-like domain (called VARL domain, volvocine algae *regA*-like), and differentiate into somatic cells, while the large cells do not express

regA and differentiate into gonidia (Kirk et al., 1999). The two cell-types are differentiated in gene expression level (Tam and Kirk, 1991; Choi et al., 1996; Meissner et al., 1999; Klein et al., 2017; Matt and Umen, 2018). The evolutionary analyses of *regA* and its homologs among volvocine algae have revealed that *regA* is derived from tandem duplication of another VARL gene *RLS1/rlsD* at the common ancestor of Volvocaceae (Hanschen et al., 2014, 2016; Grochau-Wright et al., 2017).

Overall, spheroidal colony formation and germ–soma differentiation in *V. carteri* have been investigated in developmental and molecular levels, and the homologous genes involved in these multicellular traits have been revealed to exist also in *C. reinhardtii*. In addition, comparative genomics also revealed the expansion of several ECM gene families in *V. carteri* (Prochnik et al., 2010; Hanschen et al., 2016). However, the information about the intermediate states widely seen in volvocine algae is currently not enough to identify the developmental or molecular modifications which are critical for the evolution of these traits. Moreover, there is another lineage which have evolved the multicellular traits independently of *Volvox* and Volvocaceae, a genus *Astrephomene*, which has not been studied in detail.

Astrephomene was first found by Pocock (1937) as a new species of volvocacean genus *Volvulina*, and later described in detail as a distinct genus and species (Pocock, 1954). Two species, *A. gubernaculifera* and *A. perforata*, have been described so far (Pocock, 1954; Nozaki, 1983). *Astrephomene* has 32- or 64- (or rarely 128-) celled spheroidal vegetative colonies. There are only two or four somatic cells at the posterior pole of a vegetative colony in *Astrephomene* (Pocock, 1954; Nozaki, 1983). While the number of somatic cells in both 32- and 64-celled colony in *A. perforata* and 32-celled colony in *A. gubernaculifera* is two, the number of somatic cells in 64-celled colony in *A. gubernaculifera* is four and these somatic cells are considered as “rudder” cells according to their flagella direction toward the posterior of the colony (Pocock, 1954; Nozaki, 1983). Moreover, in contrast to Volvocaceae but in the same way as *Gonium*, the tripartite boundary of ECM surrounding the entire colony (colonial boundary) is absent and each cell is enclosed by the expanded “cellular

sheaths” surrounded by the tripartite boundary (Hoops and Floyd, 1982; Nozaki, 1990). Unlike many of other volvocine species, heterotrophy of *Astrephomene* is prominent and requires external carbon sources such as acetate for its growth and reproduction (Brooks, 1972). Based on phylogenetic analyses using five protein-coding genes in chloroplast genome, *Astrephomene* was positioned as a sister group of *Gonium* in Goniaceae (Nozaki et al., 1995, 1999, 2000). The evolutionary analysis based on these phylogenetic analyses suggested that *Astrephomene* has evolved a spheroidal colony, germ–soma differentiation and expanded ECM independently of Volvocaceae (Nozaki and Ito, 1994; Herron and Michod, 2008; Herron et al., 2009).

The most outstanding feature by which *Astrephomene* has been distinguished from other volvocacean species is a lack of inversion (Pocock, 1954; Stein, 1958; Brooks, 1966; Nozaki, 1983). However, the cellular or molecular mechanisms to form spheroidal colonies have remained unveiled. Moreover, the molecular data of *Astrephomene* is absent except for some chloroplast genes used in phylogenetic analyses, thus the molecular bases involved in multicellular traits are completely unknown. In addition, there are problems in the strains of *Astrephomene*: the strains used in morphological observation (Nozaki, 1983) and phylogenetic analyses (Nozaki et al., 1995, 1999, 2000) lack their original morphology, probably due to long-term maintenance of the strains via asexual reproduction. Therefore, investigations into the multicellular traits of *Astrephomene* in morphological, developmental, cell biological or molecular levels need a new strain of *Astrephomene* and its various tools for experimental studies.

In the present thesis, I focused on *Astrephomene* for the investigation into the independently evolved multicellular traits within volvocine green algae. In Chapter 2, I established a new strain of *Astrephomene*, analyzed the cellular mechanism of spheroidal colony formation during embryogenesis in *Astrephomene* and compared it with that in Volvocaceae. In Chapter 3, I analyzed the embryogenesis of *Gonium* and *Tetrabaena*, which form more ancestral flattened colonies, and deduced the developmental mechanisms involved in the evolution of spheroidal colonies in cell

biological level. In Chapter 4, I investigated *de novo* whole nuclear genome of *Astrephomene*, conducted cell-type RNA-seq analysis and verified whether the molecular evolution involved in multicellular traits observed in *Volvox* is also found in *Astrephomene* or not. The results from these studies provide insights into the parallel evolution of multicellular traits in the volvocine green algae and would lead to further understanding of the evolution of multicellularity in molecular, genetic and genomic levels.

Figures

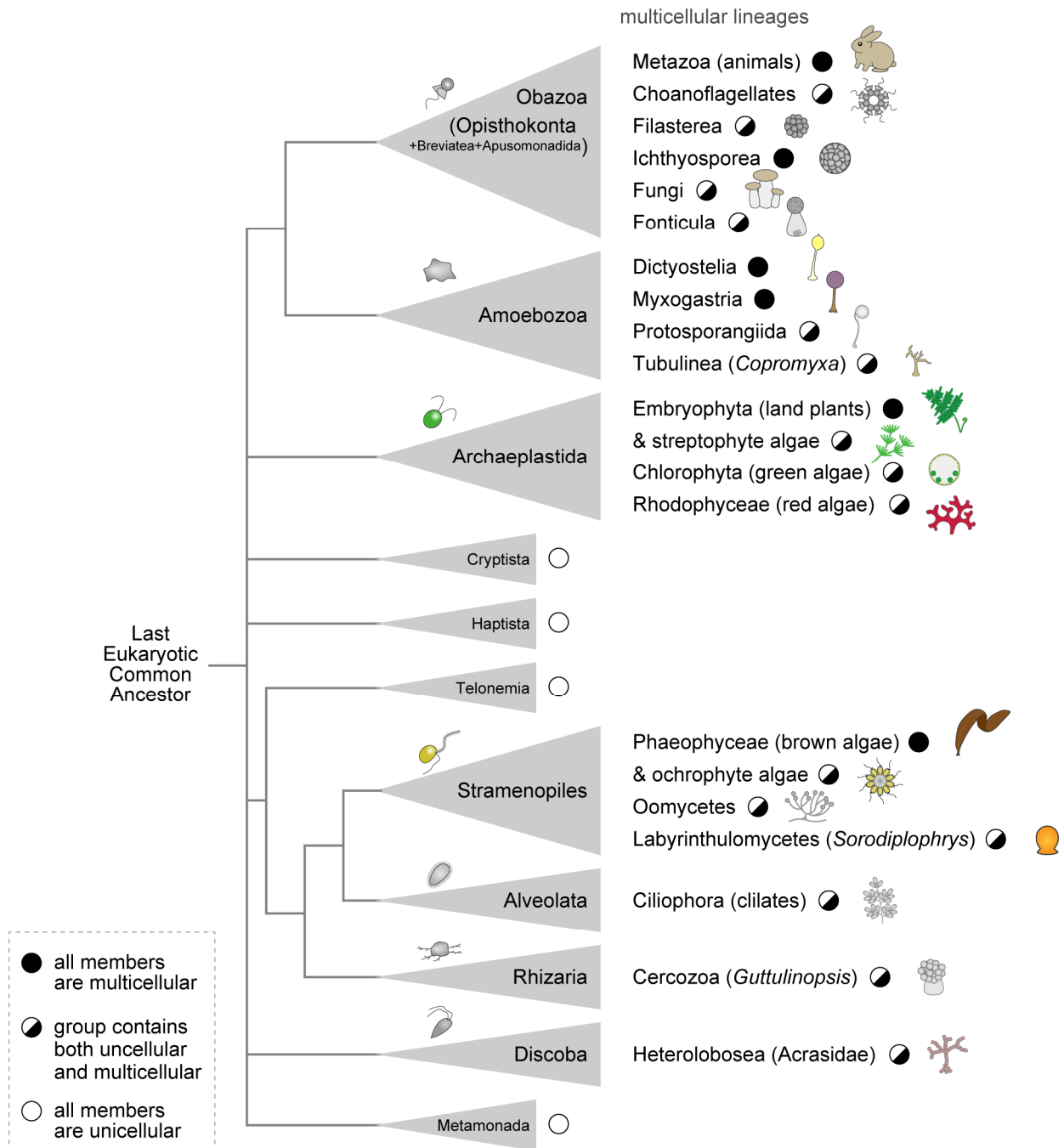


Fig. 1-1: The multiple origins of multicellular organisms in eukaryotes. Phylogenetic relationships and classifications are based on previous studies (Keeling et al., 2014; Adl et al., 2019; Strasser et al., 2019). The distribution of multicellular organisms are indicated by marks according to the key at the lower left, based on previous studies (Grosberg and Strathmann, 2007; Rensing, 2016; Brunet and King, 2017).

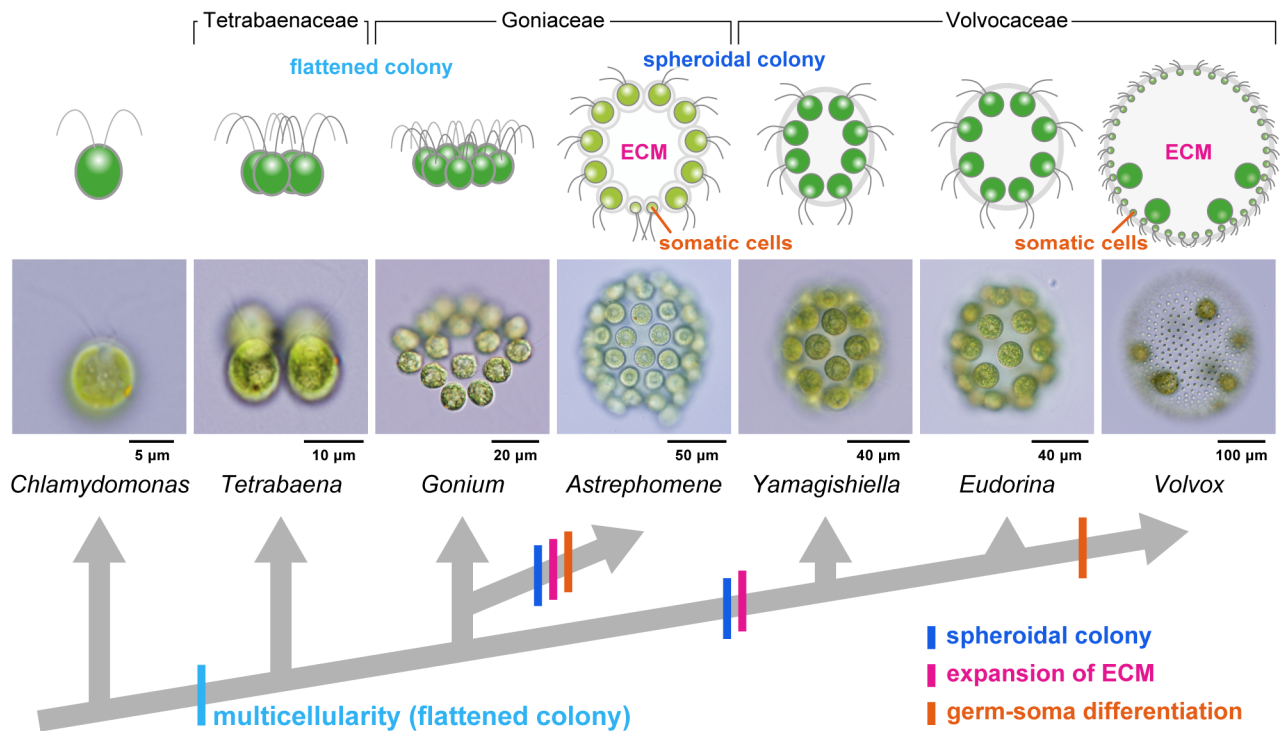


Fig. 1-2: Schematic representation of the phylogenetic relationships of the volvocine algae and the parallel evolution of the multicellular traits. Volvocine algae include unicellular *Chlamydomonas*, multicellular *Volvox* and various intermediate forms and represent a suitable model for research into the evolution of multicellularity (Kirk, 2005). The spheroidal colony, expanded ECM, and germ–soma differentiation are thought to have evolved twice or more independently within this group (Nozaki and Ito, 1994; Nozaki et al., 2000; Herron and Michod, 2008). *Astrephomene* has acquired these traits independently of Volvocaceae. The genomes of exhibited genera except for *Astrephomene* have been sequenced in previous studies (Merchant et al., 2007; Prochnik et al., 2010; Hanschen et al., 2016; Featherston et al., 2018; Hamaji et al., 2018). The phylogenetic relationships and acquisition of traits are based on previous reports (Nozaki et al., 2000; Herron and Michod, 2008; Herron et al., 2009). All drawings and photographs represent side views of individuals with anterior part of individuals orienting toward the top of this figure.

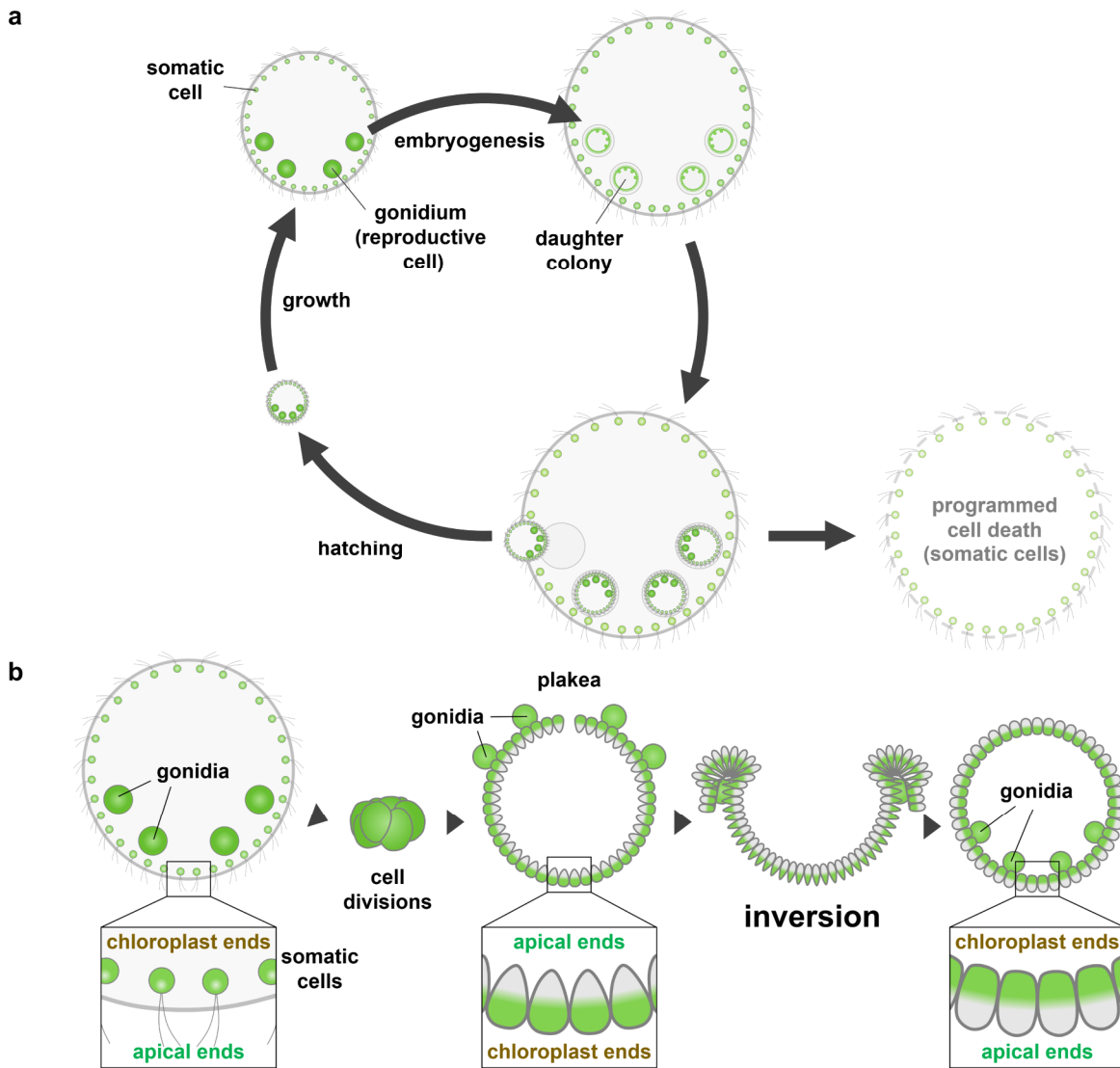


Fig. 1-3: The asexual reproduction of *Volvox carteri*. (a) Asexual life cycle and germ–soma differentiation in *V. carteri*. Based on Kirk (1998). The mature vegetative colony of *V. carteri* consists of a single cell layer of somatic cells with flagella outside and gonidia (reproductive cells) inside it. While gonidia grow largely and perform embryogenesis to form daughter colonies, somatic cells do not undergo much growth or embryogenesis but undergo programmed cell death. (b) The embryogenesis in *V. carteri*. Based on previous studies (Viamontes and Kirk, 1977; Green and Kirk, 1981; Kirk and Nishii, 2001). Each gonidium undergoes successive cell divisions to form a spheroidal plakea (embryo) composed of a single cell layer. However, immediately after the successive divisions, the cell layer is inside out compared with the adult configuration; the gonidia of daughter colony are on the outside and apical ends of the daughter protoplasts, in which basal bodies are localized and flagella are formed, are oriented toward the interior of the plakea. Therefore, the plakea undergoes inversion, during which the cell layer is turned inside out. As a result of inversion, a spheroidal daughter colony with the configuration similar to adult one is formed.

Chapter 2

Analysis of Cellular Mechanism of Spheroidal Colony Formation during Embryogenesis in *Astrephomene*

2.1. Introduction

The most prominent feature of *Astrephomene* that distinguishes it from Volvocaceae is the lack of inversion during embryogenesis; each reproductive cell in a colony divides successively to form a spheroidal daughter colony directly (Pocock, 1954; Stein, 1958). Although the embryogenesis of *Astrephomene* has been visualized by light microscopy in previous studies (Pocock, 1954; Stein, 1958; Nozaki, 1983), spheroidal colony formation in *Astrephomene* has not been observed successively in cellular and subcellular levels unlike previous studies for other volvocine algae using light microscopy time-lapse imaging (Hallmann, 2006; Arakaki et al., 2013; Höhn and Hallmann, 2016). Therefore, the mechanism underlying the formation of a spheroidal colony without inversion is unclear in cell biological level. Moreover, the cell lineage of the somatic cells in the posterior pole of the spheroidal colony in *Astrephomene* has not been investigated, in comparison to volvocacean species with somatic cells, such as *Pleodorina californica* (Gerisch, 1959) or *Volvox carteri* (Green and Kirk, 1981, 1982).

In this Chapter, I analyzed the cellular mechanism of spheroidal colony formation in *Astrephomene* based on observation of embryogenesis. Firstly, I newly established a strain of *A. gubernaculifera*, which has active growth sufficiently for developmental studies. Using the new strain, I conducted light microscopy time-lapse imaging of *A. gubernaculifera* and compared it with that of volvocacean *Eudorina* sp., which has a similar cell number and colony size. The developmental events observed by time-lapse imaging were verified at the subcellular level by tracing the behavior of the basal bodies and nuclei of daughter protoplasts by indirect immunofluorescence microscopy.

2.2. Materials and Methods

2.2.1. Strains and culture conditions

The *Astrephomene* strains established in previous research (Nozaki, 1983) often showed different morphology from previously described ones, possibly due to their long-term maintenance. Thus, a new strain of *Astrephomene* suitable for developmental observation was established in the present study. *Astrephomene gubernaculifera* strain 2014-1002-YkAs8 was isolated from a soil sample collected from a rice paddy field (35° 23'23.1" N, 140° 01'36.2" E) in Yokota, Sodegaura, Chiba Prefecture, Japan on July 21, 2014. The dried soil was rewetted with a boiled pea (*Pisum sativum*) and its extract in Petri dishes (90 × 20 mm) and incubated at 25°C on a 12-h light/12-h dark schedule under cool-white fluorescent lamps at an intensity of 50–90 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. After 3–4 days, vegetative colonies appeared in the Petri dishes. Clonal cultures were established using the pipette-washing method (Pringsheim, 1946) and grown in 10 mL of VTAC medium (Kawachi et al., 2013; the growth medium in Nozaki et al., 1989) in screw-capped tubes (18 × 150 mm). The culture was maintained at 25°C on a 12-h light/12-h dark schedule as described above. The strain 2014-1002-YkAs8 is currently available from Microbial Culture Collection at the National Institute for Environmental Studies (NIES, Kawachi et al., 2013, <http://mcc.nies.go.jp/>) as strain NIES-4017.

To identify the species of strain 2014-1002-YkAs8, morphological observation was conducted. Mature vegetative colonies in one-day-old culture were attached to coverslips coated with polyethyleneimine and the coverslips were put on slides and sealed with Vaseline. Light microscopy was carried out with a BX53 microscope (Olympus, Tokyo, Japan) equipped with Nomarski interference optics.

For comparison with volvocacean species with similar colonial organization, *Eudorina* sp. strain 2010-623-F1-E8 was used. This is a F1 strain of *Eudorina* sp. NIES-2734 and NIES-2735 (Yamada et al., 2008) and a sibling of the strains used in a previous genome study (Hamaji et al.,

2018). This strain was also used in a previous study (Hiraide et al., 2013). The strain 2010-623-F1-E8 is currently available from Microbial Culture Collection at the Institute for National Environmental Studies as strain NIES-4018.

For light microscopy time-lapse imaging, both species were grown in screw-capped tubes (18 × 150 mm) containing 10 mL VTAC medium under the culture conditions described above. For indirect immunofluorescent microscopy, to improve the synchrony, cultures were grown in silicon-capped 500 mL Erlenmeyer flasks containing 250 mL VTAC medium with aeration at 32°C on a 16-h light/8-h dark schedule under cool-white fluorescent lamps at an intensity of 140–180 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Under these conditions, the asexual life cycle of *A. gubernaculifera* was completed in approximately 24 h, and the culture was highly synchronized with the light–dark cycle. Almost all colonies of *A. gubernaculifera* seemed to initiate embryogenesis approximately 10 h after the onset of the light period. In *Eudorina* sp., the asexual life cycle was completed in approximately 48 h, and the culture was not highly synchronized with the light–dark cycle. Embryogenesis was initiated approximately 2 h after the onset of the dark period in 30% of the colonies in *Eudorina* sp.

2.2.2. Phylogenetic analysis of the new strain based on coding region of *rbcL* gene

In order to identify the species of the *Astrephomene* strain 2014-1002-YkAs8 in molecular level, the phylogenetic analysis of the new strain was conducted. The total DNA was prepared based on the previous studies (Nozaki et al., 2000; Fawley and Fawley, 2004; Nakada and Nozaki, 2007).

Concentrated vegetative colonies (ca. 20 μl) were mixed with 300 μl of extraction buffer (1 M NaCl, 70 mM Tris-HCl, 30 mM Na₂EDTA, pH 8.0), 25 μl of 10% hexadecyltrimethylammonium bromide (Sigma-Aldrich, St. Louis, MO, USA), ceramic beads and 300 μl of chloroform in a conical-bottom 2-mL microcentrifuge tube. The mixture in tube was then vibrated in 25 Hz for 8 min using a Mixer Mill MM300 (Retsch, Haan, Germany) and centrifuged at 15,000 rpm for 3 min. The total DNA in

the aqueous phase was purified using illustra blood genomicPrep Mini Spin Kit (GE healthcare, Little Chalfont, UK) according to the manufacture's protocol.

The amplification of coding region of *rbcL*, the chloroplast gene for the large subunit of RuBisCO, was conducted by PCR using TaKaRa Taq (Takara Bio Inc., Shiga, Japan), primers in previous studies (Nozaki et al., 1995, 1997) (Table 2-1), and GeneAmp PCR System 9700 (Thermo Fisher Scientific Inc., Waltham, MA, USA) with following PCR schedule: 31 cycles of 95°C for 2 min, 46°C for 2 min, and 66°C for 3 min, followed by 72°C for 15 min. The amplification of PCR products were checked by agarose gel electrophoresis with standard methods (Sambrook et al., 1989). The PCR products were purified using illustra GFX PCR DNA and Gel Band Purification Kit (GE healthcare) according to the manufacture's protocol. Sequencing of the purified PCR products was carried out using an ABI PRISM 3100 Genetic Analyser (Applied Biosystems, Foster City, CA, USA) with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). The *rbcL* sequence has been deposited to DDBJ/EMBL/GenBank (accession number: LC188997).

The coding region of *rbcL* gene (1,128 bp, corresponding to positions 31–1,158 of *Chlorella ellipsoidea* *rbcL* gene from Yoshinaga et al., 1988) from *Astrephomene* and *Gonium* (as outgroup) (Table 2-2) were subjected to maximum-likelihood analysis based on the GTR + G models and neighbor-joining method (Saitou and Nei, 1987) with maximum composite likelihood method, with 1,000 replicates of bootstrap analyses (Felsenstein, 1985), respectively, performed by MEGA 6.06 (Tamura et al., 2013). As the sequence from *A. gubernaculifera* strain 2014-1002-YkAs8 was completely the same as the sequence from *A. gubernaculifera* strain NIES-418, they were treated as a single operational taxonomic unit. The sequence alignment has been deposited in TreeBASE (<https://treebase.org/treebase-web/home>; study ID: 19979).

2.2.3. Light microscopy time-lapse imaging

Embryogenesis of *A. gubernaculifera* 2014-1002-YkAs8 and *Eudorina* sp. 2010-623-F1-E8 was observed by time-lapse light microscopy based on a method reported previously (Arakaki et al., 2013) with some modifications (Fig. 2-1). To examine the embryos from anterior–lateral, lateral, posterior–lateral and posterior angles, fully mature vegetative colonies of *A. gubernaculifera* were fragmented into several parts using a Dounce tissue grinder (Wheaton Industries Inc., Millville, NJ, USA) and were attached to coverslips coated with polyethyleneimine. Then, the coverslips were placed on slides and sealed with Vaseline. Preparations were observed using a BX-53 microscope (Olympus) equipped with Nomarski interference optics. Plural photomicrographs with different optical sections were obtained using DP Controller 1.2.1108 (Olympus) at 1-min intervals with manual successive changes in focus. In *Eudorina* sp., since fragmentation of the colonies was not possible, fully mature colonies were directly attached to coverslips and observed as described above. Only anterior–lateral view images were obtained from *Eudorina* sp. The images of both species were analyzed and processed using ImageJ 1.50b (National Institutes of Health, Bethesda, MD, USA).

2.2.4. Western blot analysis

To evaluate the specificity of the anti-*Chlamydomonas reinhardtii* SAS-6 (CrSAS-6) antibody (Nakazawa et al., 2007, kindly provided by Dr. Akira Noga at Paul Scherrer Institut and Dr. Masafumi Hirono at Hosei University) for *A. gubernaculifera* 2014-1002-YkAs8 and *Eudorina* sp. 2010-623-F1-E8, western blot analysis was carried out based on the previous study (Nakazawa et al., 2007). The total protein from culture was prepared by mixing concentrated cells (ca. 10 µl) with SDS-PAGE sample buffer (62.5 mM Tris-HCl, 0.1 M DTT, 2% SDS, 10% glycerol, 0.005% Bromophenol Blue) and heating the mixture in 100°C for 3 min. The total protein solution was then subjected to SDS-PAGE in 200 V for 30 min using Mini-PROTEAN Tetra Cell (Bio-Rad Laboratories, Inc., Hercules, CA, USA) and Mini-PROTEAN TGX Precast Gels (10%, Bio-Rad

Laboratories, Inc.) according to the manufacture's protocol. The protein in gel was transferred to membrane Amersham Hybond-P (GE healthcare) in $2 \text{ mA} \cdot \text{cm}^{-2}$ for 60 min using blotting buffer (0.025 M Tris base, 0.192 M glycine, 0.02% SDS, 20% methanol), Whatman 3MM paper (Sigma-Aldrich) and a transferring device NA-1512 (NIHON EIDO Co. Ltd., Tokyo, Japan). The membrane was incubated in blocking buffer (3% skim milk [Morinaga & Co. Ltd., Tokyo, Japan], 0.1% Tween 20 [Sigma-Aldrich] in phosphate-buffered saline [PBS]) at 4°C for 18 h. The membrane was then subjected to a primary antibody and a secondary antibody, respectively for 1 h at room temperature. The primary antibody was a rabbit anti-CrSAS-6 antibody diluted in 1:1,000 with blocking buffer. The secondary antibody was a Horseradish peroxidase-conjugated goat anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc., West Grove, PA, USA) diluted in 1:1,000 with blocking buffer. The antigen–antibody reaction was detected by Amersham ECL Prime Western Blotting Detection Reagent (GE healthcare) and ChemiDoc XRS with Quantity One software (Bio-Rad Laboratories, Inc.). Level correction of the image was conducted using Adobe Photoshop CC (Adobe Systems Inc., San Jose, CA, USA).

2.2.5. Indirect immunofluorescence microscopy

To verify the behavior of daughter protoplasts during the embryogenesis of *A. gubernaculifera* 2014-1002-YkAs8 and *Eudorina* sp. 2010-623-F1-E8, immunostaining of basal bodies and nuclei was performed based on the method described previously (Arakaki et al., 2013). The colonies and fragmented colonies containing developing embryos in culture were attached to coverslips coated with polyethyleneimine. The samples were fixed in fixing solution-A (3.7% formaldehyde [Sigma-Aldrich], 0.1% TritonX-100 [Sigma-Aldrich], 1 mM DTT [Nacalai Tesque Inc., Kyoto, Japan] in PBS) for 15 min. After fixation, chlorophyll and other pigments in samples were extracted in extracting solution (1% IGEPAL CA-630 [Sigma-Aldrich], 1% BSA [bovine serum albumin, Sigma-Aldrich], 1 mM DTT in PBS) for 60 min. The fixed samples were incubated in blocking

buffer-A (2.2% gelatin [Sigma-Aldrich], 0.05% NaN₃, 5% BSA in PBS) for 30 min and then incubated in block buffer-B (10% Goat Serum [Sigma-Aldrich] in blocking buffer-A). The samples were then subjected to two primary antibodies and two secondary antibodies, respectively for 1 h at 37°C. The primary antibodies were a mouse monoclonal anti-histone H1 antibody (clone F152.C25.WJJ, Merck Millipore Corp., Darmstadt, Germany) and a rabbit anti-CrSAS-6 antibody diluted in 1:500 with blocking buffer-C (0.44% gelatin, 0.05% NaN₃, 1% BSA, 0.1% Tween 20 [Sigma-Aldrich] in PBS). The secondary antibodies were an Alexa Fluor 568-conjugated goat anti-mouse IgG (Invitrogen, Carlsbad, CA, USA) and an Alexa Fluor 488-conjugated goat anti-rabbit IgG (Invitrogen) diluted in 1:500 with blocking buffer-C. As the two primary antibodies were derived from different organisms, the double staining of basal bodies and nuclei was feasible. The samples were put on a drop of SlowFade Gold Antifade Mountant (Thermo Fisher Scientific Inc.) Preparations were observed using a BX-60 microscope (Olympus) equipped with Nomarski interference optics, a mercury lamp, and filter sets with DP Controller 1.2.1108 (Olympus). Differential interference contrast images and fluorescence images were merged using Adobe Photoshop CC (Adobe Systems Inc.).

2.3. Results

2.3.1. Establishment of a new strain of *Astrephomene gubernaculifera*

As the *Astrephomene* strains established in previous research (Nozaki, 1983) lack their original morphology probably due to their long-term maintenance, a new strain of *Astrephomene*, 2014-1002-YkAs8, was established by pipette-washing method (Pringsheim, 1946) from a soil sample of rice paddy field in Chiba Prefecture. Light microscopy of mature vegetative colonies was carried out to identify the species of this strain (Fig. 2-2). A 64-celled vegetative colony has four somatic cells at the posterior pole (Fig. 2-2b, c) and cellular sheaths attached to each other without fenestrations (Fig. 2-2d). Based on the morphological characteristics described in Nozaki 1983, strain 2014-1002-YkAs8 was identified as *Astrephomene gubernaculifera*. The phylogenetic analysis using coding region (1,128 bp) of *rbcL* gene also supported the morphological identification (Fig. 2-3): the sequence from strain 2014-1002-YkAs8 was 100% identical with that of *A. gubernaculifera* strain NIES-418.

2.3.2. Time-lapse imaging of embryogenesis in *Astrephomene gubernaculifera* and *Eudorina* sp.

Embryogenesis of the newly established *Astrephomene* strain, *A. gubernaculifera* strain 2014-1002-YkAs8, was observed by light microscopy time-lapse imaging. Time-lapse images of the anterior–lateral view, lateral view, posterior–lateral view, and posterior view were obtained and analyzed as successive images and movies. During embryogenesis, each reproductive cell performed multiple divisions to form a spheroidal colony, as reported previously (Pocock, 1954; Stein, 1958; Nozaki, 1983). Moreover, cell divisions of all daughter protoplasts were synchronized, and six synchronous cell divisions resulted in the formation of a 64-celled daughter colony (Fig. 2-4). The intervals between cell divisions were approximately 20–30 min. Cleavage patterns were essentially identical in all reproductive cells. The cleavage patterns and cell lineages of embryogenesis in *A.*

gubernaculifera were traced by comparing images of the embryos obtained using different optical sections and angles (see Light microscopy time-lapse imaging in Materials and Methods) (Fig. 2-4).

Based on the present time-lapse imaging, gradual rotation of daughter protoplasts was observed during successive divisions. The first two cleavages did not involve rotation of protoplasts and were similar to those in other volvocine algae (Hallmann, 2006); these resulted in the generation of a four-celled embryo (Fig. 2-4a, f), which is typical of the colonies in volvocine algae (Nozaki and Ito, 1994), with two diagonal protoplasts attached to each other in the center (Fig. 2-4f). The rotation of daughter protoplasts was observed after the second division and occurred in conjunction with movement of the outer surface of the cell layer or the apical ends of daughter protoplasts toward the posterior of the embryo. In an optical section from the lateral side of the embryo, daughter protoplasts were positioned at the right side of the embryo rotated clockwise, whereas protoplasts at the left side were rotated counterclockwise. Immediately prior to the third cleavage, each daughter protoplast underwent slight rotation, which resulted in the formation of a convex 8-celled embryo with outside apical ends after the third cleavage (Fig. 2-4b, g). The rotation of daughter protoplasts was most remarkable immediately before the fourth cleavage (Fig. 2-5a–c); the apical ends of protoplasts, which frequently exhibited acute apices opposite to pale-green chloroplasts, moved from the anterior to the posterior direction of the embryo (Fig. 2-5a, b). Moreover, the angles between the fourth cleavage planes and the longitudinal axis of the embryo were larger than those for the planes of the third cleavage. A hemispherical 16-celled embryo was formed after the fourth cleavage (Figs. 2-4c, h and 2-5c). Protoplasts underwent smaller rotation prior to the fifth cleavage than before the fourth cleavage, which resulted in an almost spheroidal 32-celled embryo with a gap at the posterior pole (Fig. 2-4d, i). A very slight rotation of the protoplasts occurred before the sixth cleavage, resulting in the formation of a 64-celled embryo with a smaller gap at the posterior pole (Fig. 2-4e, j).

Immediately after the sixth cleavage, daughter protoplasts narrowed slightly and

commenced to produce flagella, and the posterior gap of a cell sheet of the embryo closed (Figs. 2-4j and 2-5d, e). Daughter protoplasts subsequently became flattened, and the hollow spheroidal daughter colony expanded (Fig. 2-5e, f). Simultaneously, the orientation of the longitudinal axes of the four cells at the posterior pole of the embryo changed from lateral to posterior, and their apical ends became protruded (Fig. 2-5e, f). These four cells were elongate-ovoid or ellipsoidal in shape and slightly larger than the other daughter cells. The most posterior position and orientation of the longitudinal axes of the four cells corresponded to those in posterior somatic cells of a mature vegetative colony (Pocock, 1954; Nozaki, 1983) (Fig. 2-2). Therefore, these four cells are supposed to become somatic cells at maturity.

Based on images of the developing embryos obtained using different optical sections and angles, the cell lineage of somatic cells was traced (Fig. 2-4). The pattern of cell divisions was 180° rotationally symmetrical about the longitudinal axis of the embryo, and the origins of four posterior somatic cells (Fig. 2-4e, j; 1 pp, 2 pp, 1 pp', and 2 pp' cells) were determined; the somatic cells were derived from two diagonally opposite daughter protoplasts in four-celled embryos (Fig. 2-4a, f; A and A' cells) and two opposite pairs of adjoining protoplasts at the periphery of 16-celled embryos (Fig. 2-4c, h; A1, A2, A1' and A2' cells).

To directly compare the embryogenesis in *A. gubernaculifera* with that in Volvocaceae, light microscopy time-lapse images of embryogenesis in *Eudorina* sp. were analyzed (Fig. 2-6). During formation of daughter colonies in *Eudorina* sp., 32 or 16 vegetative cells in a colony divided five times successively to produce 32-celled concave plakeas, which became spheroidal daughter colonies by means of inversion, as reported previously (Gottlieb and Goldstein, 1977; Marchant, 1977; Hallmann, 2006). The present time-lapse analyses of *Eudorina* sp. demonstrated that rotation of daughter protoplasts did not occur during successive cell divisions (Fig. 2-6a, b), and the angles between the cleavage planes and the longitudinal axis of the mother protoplast did not change markedly (Fig. 2-6c). After five successive divisions, a concave plakea, a single cell layer with apical

ends of the daughter protoplasts inside, was formed (Fig. 2-6d). Then, inversion occurred to bend the cell layer to form a spheroidal daughter colony (Fig. 2-6e, f). Each daughter protoplast formed a stalk at the chloroplast end during the initial stage of inversion (Fig. 2-6e).

2.3.3. Indirect immunofluorescence microscopy of basal bodies and nuclei

The cellular behavior observed by light microscopy time-lapse imaging was verified at the subcellular level by indirect immunofluorescence microscopy of basal bodies and nuclei. For immunostaining of basal bodies, an antibody against *Chlamydomonas reinhardtii* SAS-6 (CrSAS-6), which is localized in mature and immature basal bodies (Nakazawa et al., 2007), was reacted with fixed embryos; thus, two mature basal bodies and two immature basal bodies in each cell were observed as four dots. The western blot analysis showed that the anti-CrSAS-6 antibody used in previous study (Nakazawa et al., 2007) cross-reacted with SAS-6 from *A. gubernaculifera* and *Eudorina* sp. (Fig. 2-7). An anti-histone antibody was also reacted with the samples at the same time, allowing simultaneous visualization of basal bodies and nuclei. The localization of basal bodies and nuclei was used to identify the longitudinal axis of daughter protoplasts during embryogenesis.

The position and movement of basal bodies and nuclei in daughter protoplasts during the embryogenesis of *A. gubernaculifera* confirmed the rotation of daughter protoplasts during successive divisions (Fig. 2-8). The developmental stages of fixed embryos were identified by correspondence to the time-lapse images. As the cell divisions progressed, the basal bodies of the daughter protoplasts at the outer surface of the convex embryo moved from anterior to posterior region of the embryo (Fig. 2-8a, b). The position of the basal bodies immediately after the rotation of protoplasts corresponded to the subsequent cleavage furrows (Fig. 2-8b, c). Additionally, the drastic change in the longitudinal axes of the four posterior somatic cells in the 64-celled stages observed by the time-lapse imaging (Fig. 2-5e, f) was verified by the positions of the basal bodies of the four most posterior protoplasts of the embryos (Fig. 2-8d, e).

In contrast to the gradual change of basal body positions between cell divisions in *A. gubernaculifera*, basal bodies moved only during inversion in the embryogenesis of *Eudorina* sp. (Fig. 2-9). The basal bodies of the daughter protoplasts did not move markedly within the embryo of *Eudorina* sp. and remained at the concave face of the plakea during successive cell divisions (Fig. 2-9a–c). This situation is consistent with the absence of rotation of daughter protoplasts between cell divisions in *Eudorina* sp. (Fig. 2-6). Basal bodies inside the concave cell layer after successive cell divisions (Fig. 2-9d) came to be oriented toward the outside of the spheroidal cell layer by means of inversion (Fig. 2-9e, f).

2.4. Discussion

Using light microscopy time-lapse imaging of the embryogenesis of *Astrephomene gubernaculifera* with a high temporal resolution (1 min) and several views, I demonstrated the dynamic cellular events, rotation of daughter protoplasts during successive cell divisions, and drastic change in the orientation of the posterior somatic cells after successive cell divisions. The time-lapse observation was required to figure out the rotation, since the intervals between cell divisions were more than 20 min and the rotational movement was slow. These results were also verified by the indirect immunofluorescence microscopy of basal bodies and nuclei.

Comparison of the cellular events during the embryogenesis of *A. gubernaculifera* with those in volvocacean *Eudorina* sp. revealed two distinct mechanisms of spheroidal colony formation (Fig. 2-10). For the appropriate locomotion of the spheroidal colony, the basal bodies of daughter protoplasts in embryos of both lineages must be oriented toward the outside. In *A. gubernaculifera*, the rotation of daughter protoplasts with the movement of basal bodies occurred during successive cell divisions, resulting in the formation of a convex-to-spheroidal cell layer with basal bodies outside (Figs. 2-5a–c and 2-8a–c). In contrast, I did not observe rotation of daughter protoplasts or movement of basal bodies during successive cell divisions in *Eudorina* sp. (Figs. 2-6a–c and 2-9a–c). Inversion then occurred to form a spheroidal daughter colony with basal bodies outside (Figs. 2-6d–f and 2-9d–f), as in other members of Volvocaceae (Hallmann, 2006). Thus, the following mechanisms of spheroidal colony formation are supposed to have evolved independently in the two lineages: gradual rotation of protoplasts during successive divisions in *Astrephomene*, and inversion after successive divisions in Volvocaceae. The verification of this hypothesis requires more information about the ancestral states of embryogenesis. In Chapter 3, I investigated the embryogenesis of *Gonium* and *Tetrabaena*, which have the ancestral flattened colonies.

This study also demonstrated the cell lineages of posterior somatic cells (Fig. 2-4) and their change of direction toward the posterior pole after successive cell divisions (Fig. 2-6d-f). The volvocacean species *Pleodorina californica* has somatic cells at the anterior region in a spheroidal colony. These anterior somatic cells are derived from the daughter protoplasts in the center of a plakea during embryogenesis (Gerisch, 1959). In *A. gubernaculifera*, the posterior somatic cells were derived from the periphery of the embryo. These phenomena show that the correspondence of the central-peripheral polarity in embryos to the anterior-posterior axis in mature spheroidal colonies is conserved between *Astrephomene* and Volvocaceae. Moreover, no apparent asymmetric cell divisions in cell size was observed to produce somatic cells in *A. gubernaculifera* (Fig. 2-4) in contrast to *Volvox carteri* (Green and Kirk, 1981).

Although the present study revealed the cellular mechanism of spheroidal colony formation in *Astrephomene* (i.e., rotation of daughter protoplasts during successive cell divisions), the underlying molecular mechanism is unclear. In Chapter 4, I identified embryogenesis-related genes in *A. gubernaculifera*, whose orthologs are involved in inversion or asymmetric cell divisions in *V. carteri*, and compared them with those in other volvocine algae.

2.5. Tables and Figures

Table 2-1: List of primers used for amplification and sequencing of *rbcL* gene.

Designation	Positions ^a	Sequence (5'–3')
rbcL-F1 ^b	1–20	ATGGTTCACAAACAGAAAC
rbcL-F4 ^b	320–341	TATTCGAAGAAGGTTTCAGTAAC
rbcL-R2 ^b	395–376 ^d	GCACGTAAAGCTTTGAAACC
rbcL-F9 ^c	475–497	CGTGACAAACTAAACAAATATGG
rbcL-F7 ^b	650–671	GTTTCCTTTTCGTAGCTGAAGC
rbcL-R10 ^c	685–665 ^d	TGTGCTTTGTAAATAGCTTCAG
rbcL-R6 ^b	830–809 ^d	TTAGCTGTGAAACCACCTGTTA
rbcL-R8 ^b	1181–1160 ^d	AAGATTTCAACTAAAGCTGGCA
rbcL-R3 ^b	1421–1402 ^d	TTGTCAATAGTATCAAATTC

^a Coordinate number from the *Chlorella ellipsoidea rbcL* gene (Yoshinaga et al., 1988).

^b Designed in Nozaki et al. (1995).

^c Designed in Nozaki et al. (1997).

^d Reverse primer.

Table 2-2: List of *rbcL* genes used in the phylogenetic analysis (Fig. 2-3).

Taxon	Strain designation	Accession number
<i>Astrephomene gubernaculifera</i>	2014-1002-YkAs8 (= NIES ^a -4017)	LC188997 ^c
	NIES-418	D63428
	UTEX ^b 1394	AB044170
<i>Astrephomene perforata</i>	NIES-564	D63429
<i>Gonium multicoccum</i>	UTEX 2580	D63435
<i>Gonium octonarium</i>	NIES-851	D63436
<i>Gonium pectorale</i>	NIES-569	D63437
<i>Gonium quadratum</i>	NIES-653	D63438
<i>Gonium viridistellatum</i>	UTEX 2519	D86831

^a Microbial Culture Collection at the National Institute for Environmental Studies (Kawachi et al., 2013).

^b Culture Collection of Algae at the University of Texas at Austin (Starr and Zeikus, 1993).

^c Sequenced in this study.

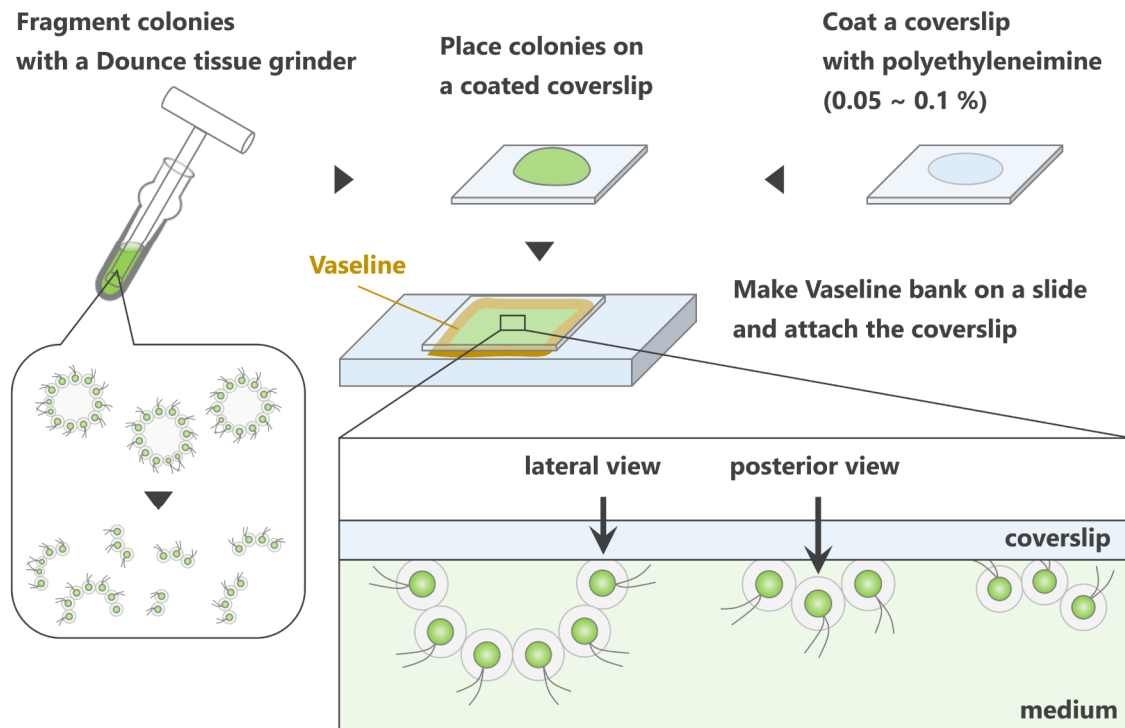


Fig. 2-1: Diagram of making preparations for light microscopy time-lapse imaging of *Astrephomene gubernaculifera* embryogenesis. In order to examine the embryos from anterior–lateral, lateral, posterior–lateral and posterior views, fully mature vegetative colonies of *A. gubernaculifera* were fragmented into several parts with a Dounce tissue grinder and attached to coverslips coated with polyethyleneimine. Then the coverslips were put on slides and sealed with Vaseline.

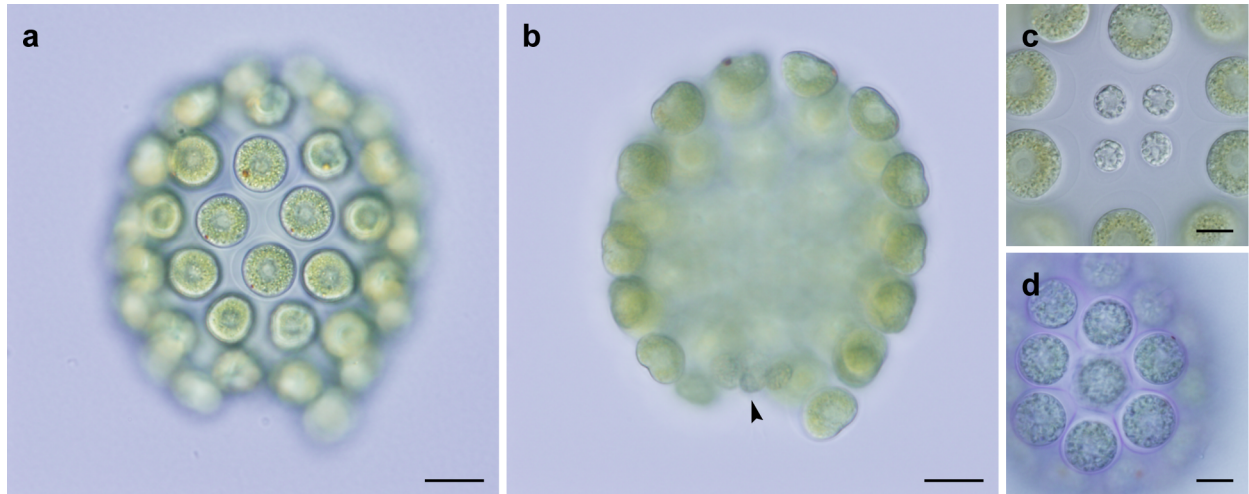


Fig. 2-2: Morphological observation of vegetative colonies in *Astrephomene gubernaculifera* strain 2014-1002-YkAs8. (a) Surface view of a 64-celled colony. Scale bar: 20 μm . (b) Median optical section of a 64-celled colony. Note small somatic cells at the posterior pole of colony (arrowhead). Scale bar: 20 μm . (c) Four somatic cells at the posterior pole of a 64-celled colony. Note the somatic cells are smaller and paler than reproductive cells. Scale bar: 10 μm . (d) Cellular sheaths stained with diluted methylene blue. Note cellular sheaths attached to each other without fenestrations. Scale bar: 10 μm .

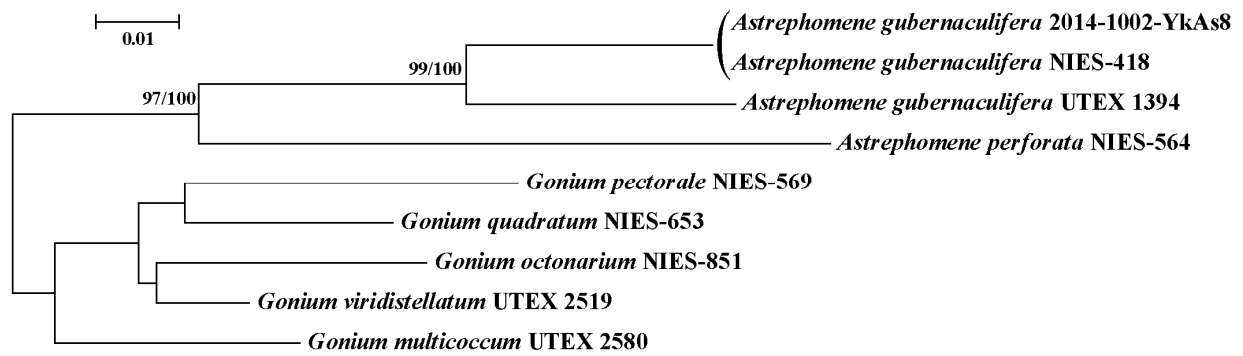


Fig. 2-3: Maximum-likelihood tree of *Astrephomene* and *Gonium* based on *rbcL* genes. The coding region (1,128 bp) of *rbcL* gene from *Astrephomene* and *Gonium* (as outgroup) (Table 2-2) were subjected to maximum-likelihood analysis based on the GTR + G models with 1,000 replicates of bootstrap analyses (Felsenstein, 1985) performed by MEGA 6.06 (Tamura et al., 2013). Bootstrap values ($\geq 50\%$) from maximum-likelihood (left), and neighbor-joining (right) are shown at branches.

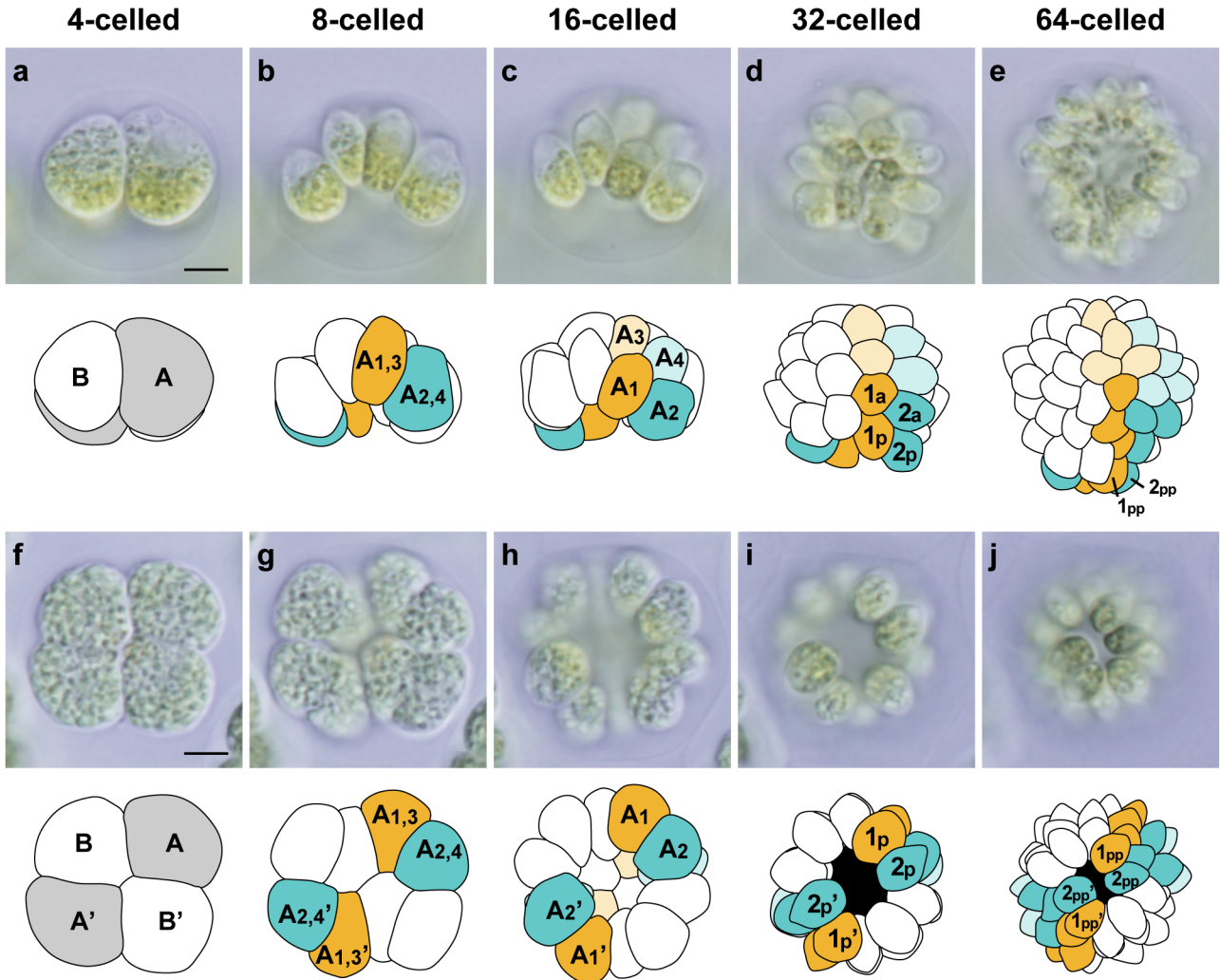


Fig. 2-4: Cleavage patterns and cell lineages involved in embryogenesis of *A. gubernaculifera*. Based on time-lapse analyses of two views of embryos. Images in each row are shown at the same magnification. Scale bars: 5 μ m. Successive lateral view (a–e) and posterior view (f–j) images are shown, together with diagrams. Outlines of daughter protoplasts were traced from images in different optical sections taken during a 1-min period. The cleavage pattern was 180° rotationally symmetrical about the longitudinal axis of the embryo. Four posterior somatic cells in the newly formed colony (1 pp, 2 pp, 1 pp', and 2 pp' cells in e, j) were derived from two diagonally opposite cells in four-celled embryos (A and A' cells in a, f) and four peripheral cells in 16-celled embryos (A1, A2, A1', and A2' cells in c, h).

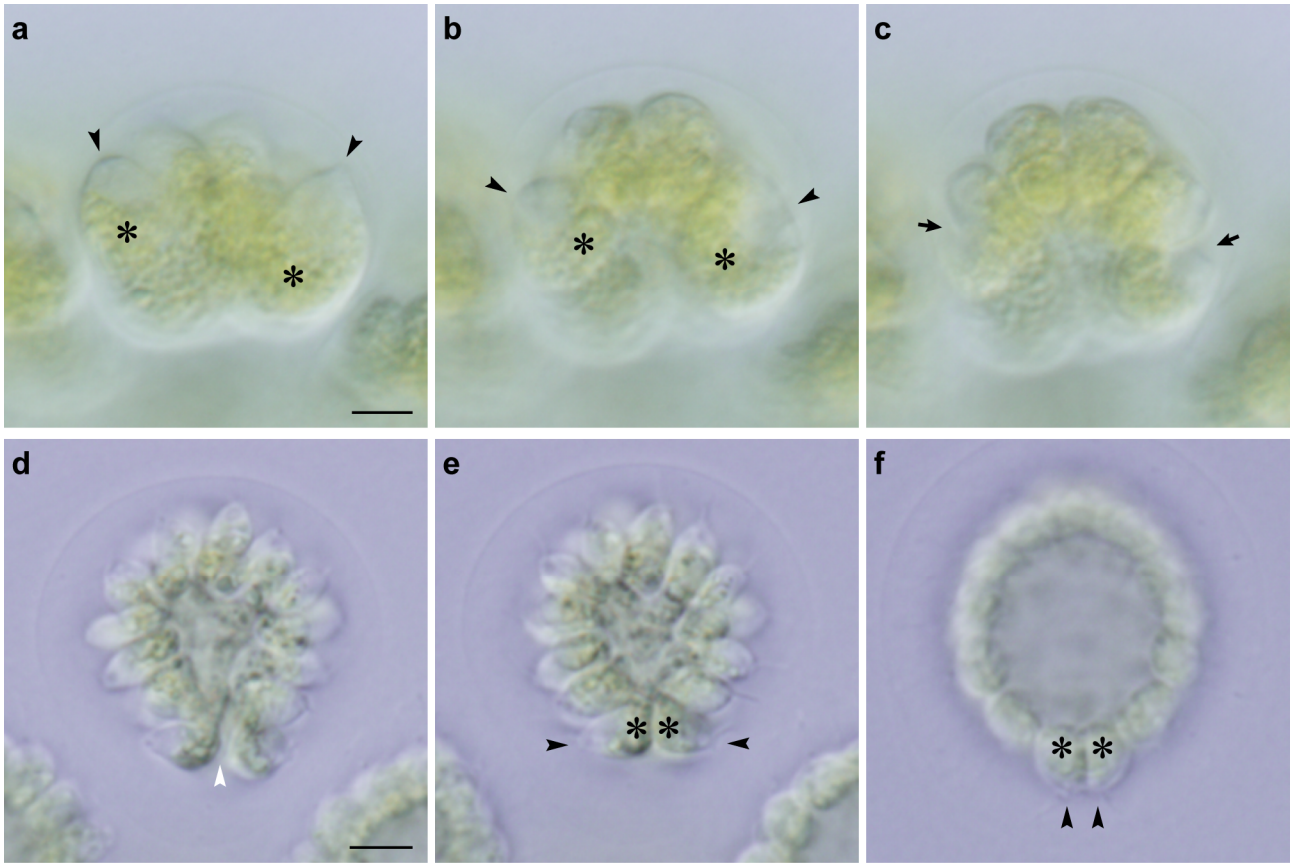


Fig. 2-5: Behavior of daughter protoplasts during embryogenesis in *Astrephomene gubernaculifera*. Time-lapse analyses of lateral views of embryos. Images in the same row are shown at the same magnification. Scale bars: 5 μm . (**a–c**) Successive anterior–lateral view images of an embryo at the 8-celled stage (**a**), late 8-celled stage (**b**), and early 16-celled stage (**c**). Note that the rotation of daughter protoplasts is indicated by the positions of the apical ends (black arrowheads) and chloroplasts (asterisks) (**a, b**). The positions of cleavage furrows (arrows) corresponded approximately to the positions of the apical ends in the preceding stage (**b, c**). (**d–f**) Successive posterior–lateral view images of an early 64-celled embryo immediately after the final cleavage (**d**) and 5 min after the final cleavage (**e**), and an expanded 64-celled embryo or daughter colony (**f**). The posterior gap of the cell layer (white arrowhead, **d**) was closed soon after the last cleavage, and the shape of the daughter protoplasts became narrower slightly. Then the orientation of the apical ends of posterior somatic cells (black arrowheads, **e, f**), opposite to chloroplasts (asterisks), changed from lateral (**e**) to posterior (**f**).

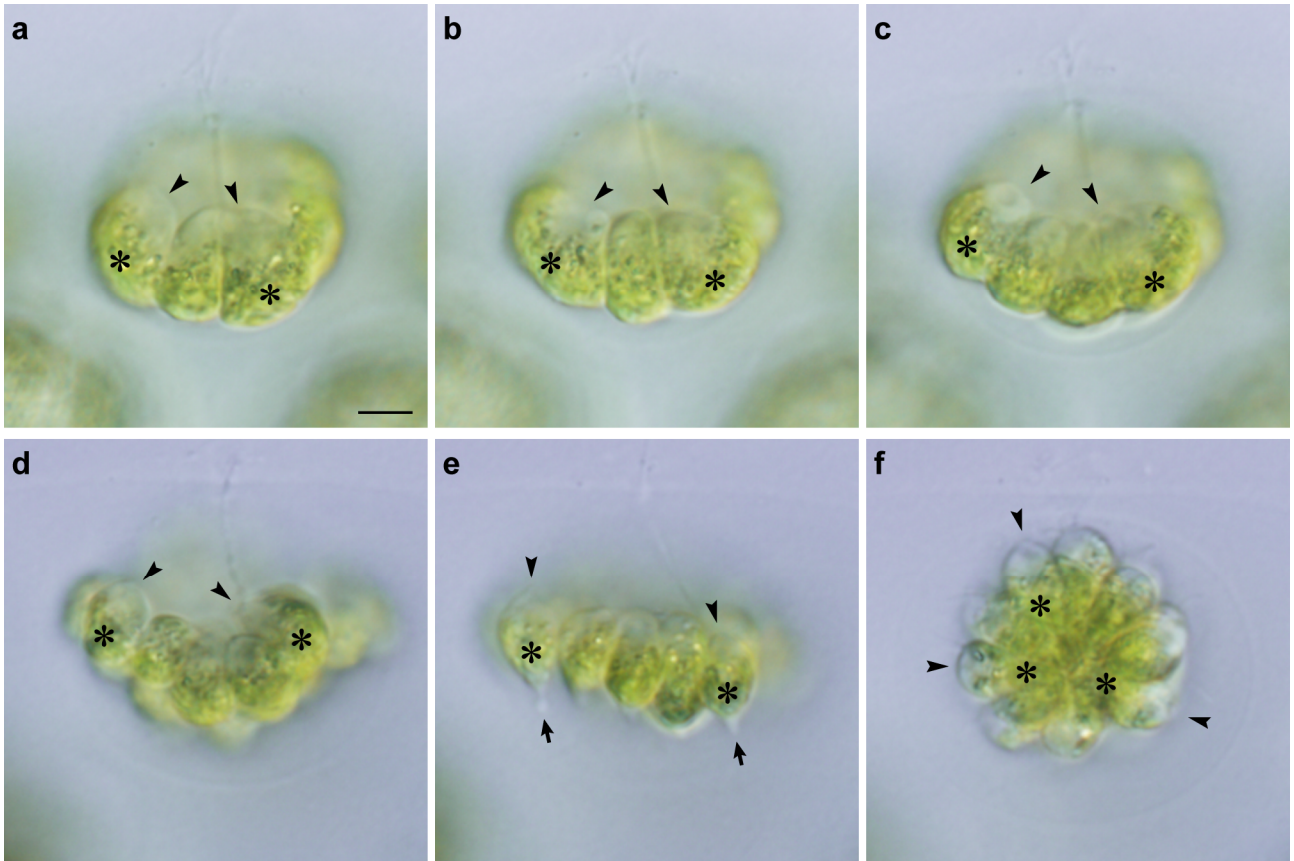


Fig. 2-6: Cell divisions and inversion during embryogenesis in *Eudorina* sp. Successive stages of an embryo observed by time-lapse analysis from anterior–lateral view. All at the same magnification throughout. Scale bar: 5 μ m. Note the longitudinal axis of each daughter protoplasts indicated by positions of apical ends (arrowheads) and chloroplasts (asterisks). Rotation of daughter protoplasts was not observed during cell divisions (**a–c**). The concave surface of plakea or apical ends of the daughter protoplasts (**d**) became outer surface of the spheroid by means of inversion (**e, f**). (**a**) Early 8-celled stage. (**b**) Late 8-celled stage. (**c**) Early 16-celled stage. (**d**) 32-celled stage before inversion. (**e**) Inverting plakea. Note formation of stalks (arrows) at the chloroplast ends of daughter protoplasts. (**f**) Spheroidal embryo or daughter colony just after inversion.

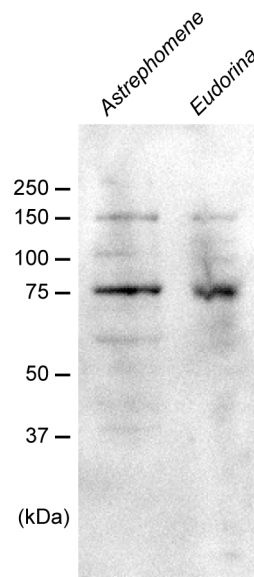


Fig. 2-7: Western blot of two species with anti-CrSAS-6 antibody. The protein bands were detected with anti-CrSAS-6 antibody in *Astrephomene gubernaculifera* strain 2014-1002-YkAs8 and *Eudorina* sp. strain 2010-623-F1-E8. Western blotting showed that the anti-CrSAS-6 antibody cross-reacted with SAS-6 from the two species.

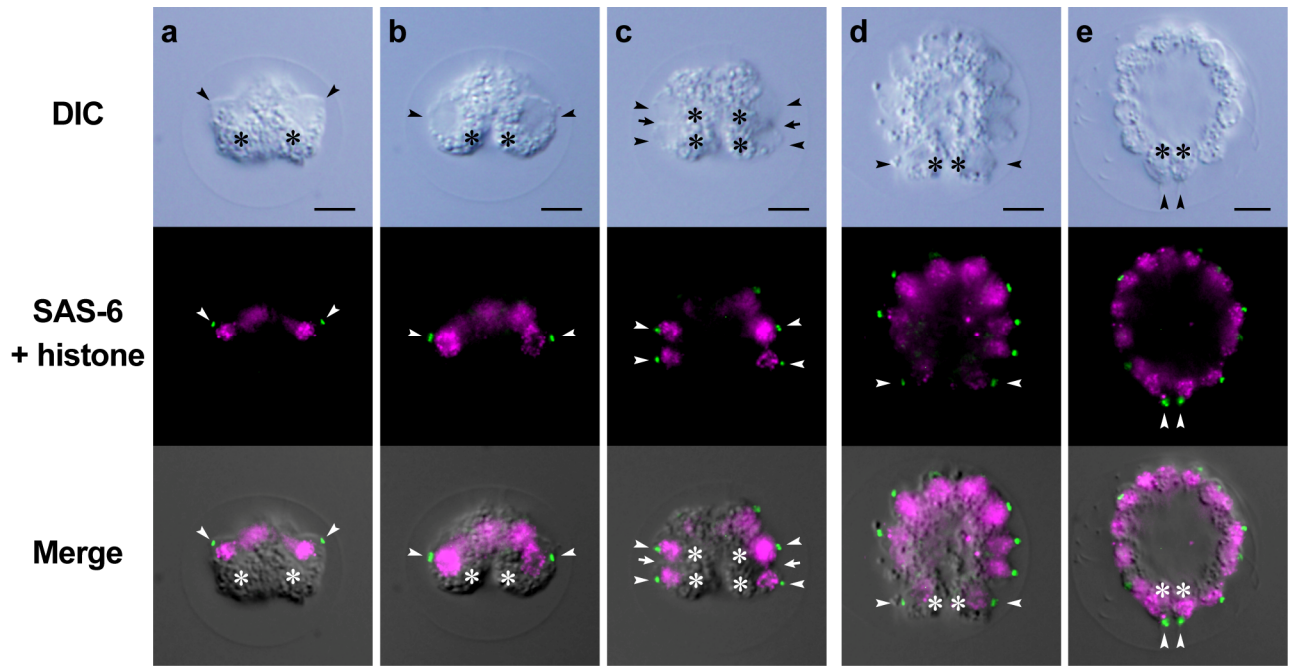


Fig. 2-8: Indirect immunofluorescence microscopy of developing *Astrephomene gubernaculifera* embryos. Differential interference contrast (DIC) images (top row), fluorescence images (middle row) labeled with anti-SAS-6 (green) and anti-histone (magenta) antibodies, and merged images (bottom row) of the same embryos are shown lengthwise. Scale bars: 5 μ m. Positions of basal bodies labeled with an anti-SAS-6 antibody (arrowheads) and chloroplasts (asterisks) are shown. **(a)** Early 8-celled embryo. **(b)** Late 8-celled embryo. Note that the positions of the basal bodies of daughter protoplasts (arrowheads) are changed from the anterior to the posterior region of the embryo during the 8-celled stage. **(c)** Early 16-celled embryo. Note that the positions of the fourth cleavage furrows (arrows) correspond approximately to the positions of basal bodies at the 8-celled stage **(b)**. **(d)** Early 64-celled embryo. **(e)** Late 64-celled embryo or a daughter colony before hatching. Note that the posterior somatic cells are oriented toward the posterior direction of the daughter colony.

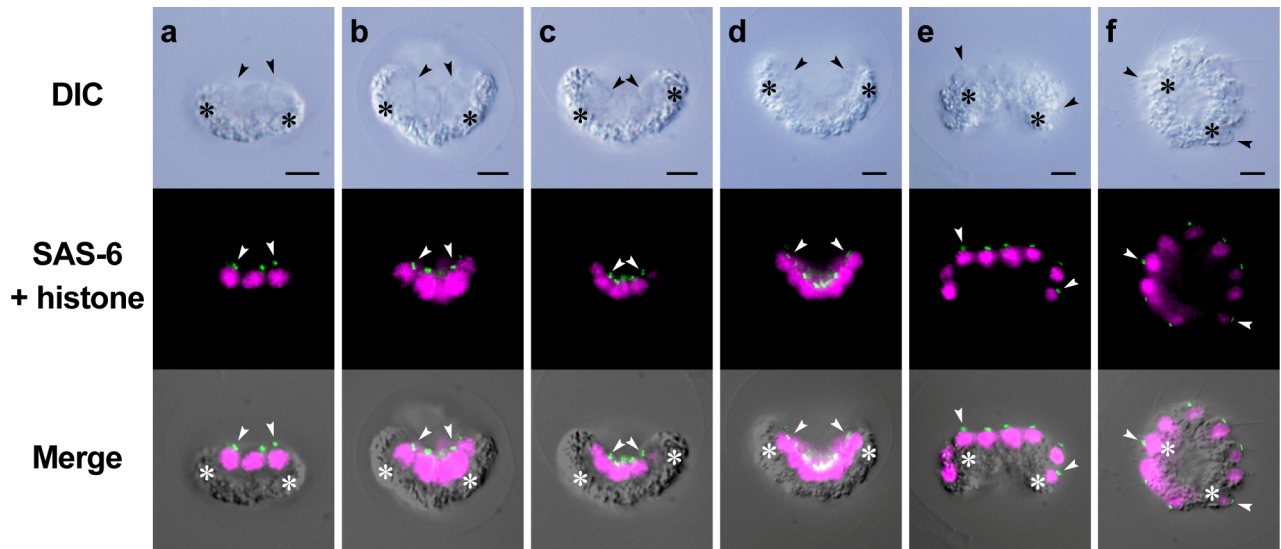


Fig. 2-9: Indirect immunofluorescence microscopy of daughter colony formation in *Eudorina* sp. Differential interference contrast (DIC) images (top row), fluorescence images (middle row) labeled with anti-SAS-6 (green) and anti-histone (magenta) antibodies, and merged images (bottom row) of the same embryos are shown lengthwise. Scale bars: 5 μ m. Positions of basal bodies labeled with anti-SAS-6 antibody (arrowheads) and chloroplasts (asterisks) are shown. Note that basal bodies of daughter protoplasts are positioned in the center of concave surface of plakea before inversion (**a–d**). Basal bodies are moved from interior or concave surface of the embryo to the outer surface of the spheroidal daughter colony by inversion (**d–f**). (**a**) Early 8-celled embryo. (**b**) Late 8-celled embryo. (**c**) Early 16-celled embryo. (**d**) Early 32-celled embryo before inversion. (**e**) Inverting 32-celled embryo. Note that the periphery of the plakea commences bending toward outside. (**f**) 32-celled embryo after inversion.

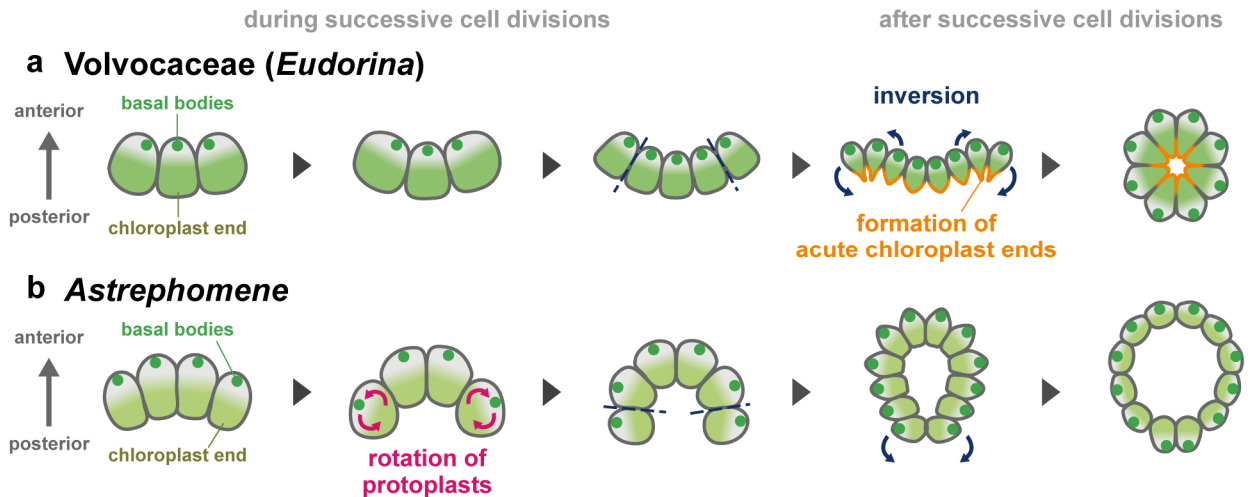


Fig. 2-10. Schematic diagrams of embryogenesis in Volvocaceae and *Astrephomene*. Diagrams represent lateral views of stages of embryogenesis with anterior sides of embryos oriented toward the top of the figure. (a) Embryogenesis in *Eudorina* as a representative of Volvocaceae. A mature reproductive cell undergoes successive cell divisions and forms a cup-shaped embryo composed of a single layer of daughter protoplasts. Following the successive divisions, the embryo inverts its cell layer to form a spheroidal daughter colony, with the formation of acute chloroplast ends (opposite to basal bodies) and the relocation of cytoplasmic bridges at the chloroplast ends. (b) Embryogenesis in *Astrephomene*. In contrast to Volvocaceae, an embryo of *Astrephomene* undergoes rotation of daughter protoplasts during successive cell divisions and forms a nearly spheroidal cell layer just after successive cell divisions.

Chapter 3

Analysis of Flattened Colony Formation during Embryogenesis in *Gonium* and *Tetrabaena*

3.1. Introduction

In Chapter 2, I revealed the major cellular mechanisms of spheroid colony formation in *Astrephomene* and Volvocaceae: the rotation of the daughter protoplast during successive cell divisions and inversion with the formation of acute chloroplast ends, respectively (Chapter 2, Fig. 2-10). How the two different mechanisms for making spheroidal colonies evolved in Volvocaceae and *Astrephomene* is currently unclear, since the embryogenesis of extant genera with ancestral flattened colonies, such as *Gonium* and *Tetrabaena* (Chapter 1, Fig. 1-2), has not been investigated in detail from this point of view. In contrast to inversion in Volvocaceae, *Gonium* performs “partial inversion” or “incomplete inversion”, in which a cup-shaped cell sheet of an embryo expands to form a flattened colony (Harper, 1912; Hartmann, 1924; Hallmann, 2006; Iida et al., 2013). On the other hand, embryogenesis of *Tetrabaena* is much simpler; a four-celled daughter colony is formed after two successive cell divisions (Arakaki et al., 2013). Though embryogenesis in *Gonium* and *Tetrabaena* has been observed by light microscopy time-lapse imaging (Hallmann, 2006; Arakaki et al., 2013; Iida et al., 2013) and transmission electron microscopy (Arakaki et al., 2013; Iida et al., 2013), it is currently unclear whether the rotation of daughter protoplasts seen in *Astrephomene* or the cell shape change of Volvocaceae occurs in these genera, as their embryogenesis has been observed at certain stages and from limited angles.

In this Chapter, I focused on the formation of flattened colonies in *G. pectorale* and *T. socialis* to evaluate whether they employ cellular mechanisms that may have been co-opted to make spheroidal colonies in Volvocaceae and *Astrephomene*. I conducted light microscopy time-lapse imaging showing all successive stages of embryogenesis, as well as indirect immunofluorescence microscopy staining basal bodies, nuclei, and microtubules of both species to investigate their embryogenesis at cellular and subcellular levels. The present results define the similarities and

differences in the developmental events that lead to the formation of flattened and spheroidal colonies, thus providing clues to the evolutionary pathway from flattened to spheroidal colonies.

3.2. Materials and Methods

3.2.1. Establishment of a new strain of *Gonium pectorale*

Gonium pectorale strain 2017-0423-IsgGo1 was isolated from a soil sample collected from a rice paddy field (24°24'54.9" N, 123°46'45.4" E) in Yaeyama, Okinawa Prefecture, Japan on March 20, 2017. The dried soil was rewetted with distilled water in Petri dishes (90 × 20 mm) and incubated at 25°C on a 12-h light/12-h dark schedule under cool-white fluorescent lamps at an intensity of 50–90 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. After 7–8 days, vegetative colonies appeared in the Petri dishes. Clones were established using the pipette-washing method (Pringsheim, 1946). *G. pectorale* strain 2017-0423-IsgGo1 is currently available from Microbial Culture Collection at the Institute for National Environmental Studies (NIES, Kawachi et al., 2013, <http://mcc.nies.go.jp/localeAction.do?lang=en>) as NIES-4327.

For the morphological identification of species, light microscopy of mature vegetative colonies in two-days-old culture was carried out with a BX53 microscope (Olympus, Tokyo, Japan) equipped with Nomarski interference optics. The culture was grown in 10 mL of SVM medium (Kirk and Kirk, 1983) in screw-capped tubes (18 × 150 mm) and maintained at 25°C on a 12-h light/12-h dark schedule under cool-white fluorescent lamps at an intensity of 50–90 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

In order to identify the species of the strain 2017-0423-IsgGo1 in molecular level, the phylogenetic analysis was conducted. The coding region of *rbcL*, the chloroplast gene for the large subunit of RuBisCO, of *G. pectorale* strain 2017-0423-IsgGo1 were sequenced with essentially the same method in previous research (Nozaki et al., 1995, 1997; Fawley and Fawley, 2004; Nakada and Nozaki, 2007). The 1,128 bp corresponded to position 31-1,158 of *rbcL* from *G. pectorale* strain 2017-0423-IsgGo1 and 14 OTUs of *Gonium* as well as 3 OTUs of *Astrephomene* (as outgroup) (Table 3-1) were subjected to Bayesian interference and maximum-likelihood analysis. For the substitution model, GTR + I + G model were selected using MEGA 7.0.21 (Kumar et al., 2016).

Bayesian interference was performed using MrBayes 3.2.6 (Ronquist et al., 2012) with 1,000,000 generations of Markov chain Monte Carlo iterations; the first 25% of the generations were discarded as burn-in. Maximum-likelihood analysis was performed with 1,000 replicates of bootstrap analyses (Felsenstein, 1985) performed by MEGA 7.0.21 (Kumar et al., 2016). The new *rbcL* sequence from *G. pectorale* strain 2017-0423-IsgGo1 has been deposited to DDBJ/EMBL/GenBank (accession number: LC459976) and the sequence alignment has been deposited in TreeBASE (<https://treebase.org/treebase-web/home.html>; study ID: 23894).

3.2.2. Strains and culture conditions

For developmental observation, *G. pectorale* strain 2017-0423-IsgGo1 was grown synchronously in silicon-capped 500-mL Erlenmeyer flasks containing 250 mL SVM medium with aeration at 32°C on a 16-h light/8-h dark schedule under cool-white fluorescent lamps at an intensity of 140–180 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$, and *Tetrabaena socialis* strain NIES-571 was grown synchronously in silicon-capped 500 mL Erlenmeyer flasks containing 250 mL SVM medium with aeration at 20°C on a 12-h light/12-h dark schedule under cool-white fluorescent lamps at an intensity of 240–280 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

3.2.3. Light microscopy time-lapse imaging

Embryogenesis of *G. pectorale* strain 2017-0423-IsgGo1 and *T. socialis* strain NIES-571 was observed by time-lapse light microscopy based on methods in Chapter 2 with some modifications. As the polyethyleneimine coating on coverslips hampered the normal development of *G. pectorale* and *T. socialis* embryos, I used coverslips without coating. Culture containing fully mature vegetative colonies of *G. pectorale* or *T. socialis* was placed on slides, and then the coverslips, surrounded by thin banks of petroleum jelly, were placed on top of the culture. In this method, cells and embryos are attached to the coverslip with weak adhesion of the flagella or cell walls. Preparations were

observed under a BX-53 microscope (Olympus) equipped with Nomarski interference optics. Photomicrographs were obtained using DP Controller 1.2.1108 (Olympus) at 1-min (*G. pectorale*) or 2-min (*T. socialis*) intervals. This difference in timing was implemented in response to the observation that *T. socialis* embryos ceased embryogenesis when photomicrographs were taken at 1-min intervals, likely because they are sensitive to light and/or heat, so that frequent exposure to light under light microscopy prevented *T. socialis* embryos from proceeding with their development. As the embryos of *G. pectorale* and *T. socialis* moved and rotated during observation, positions and angles of images were adjusted manually using Adobe Photoshop CC (Adobe Systems Inc., San Jose, CA, USA). The sequence of adjusted images was analyzed and processed using ImageJ 1.50b (National Institutes of Health, Bethesda, MD, USA). Embryogenesis was observed at least three times in each species and showed almost identical patterns in each observation.

3.2.4. Indirect immunofluorescence microscopy

To verify the behavior of daughter protoplasts during embryogenesis in *G. pectorale* and *T. socialis*, immunostaining of basal bodies and microtubules was initially performed as previously described (Arakaki et al., 2013). However, I found that the cortical microtubules of *G. pectorale* and *T. socialis* were fragile and broke when subjected to that protocol, and thus the following modifications were conducted. The samples were placed in fixing solution-A (3.7% formaldehyde [Sigma-Aldrich, St. Louis, MO, USA], 0.1% Triton X-100 [Sigma-Aldrich], and 1 mM DTT [Nacalai Tesque Inc., Kyoto, Japan] in phosphate-buffered saline [PBS]) for 1 min and then into fixing solution-B (3.7% formaldehyde and 1 mM DTT in PBS; i.e., fixing solution-A without Triton X-100) for 14 min, instead of the original protocol of 15 min in fixing solution-A. Chlorophyll and other pigments in the samples were extracted in -20°C methanol for 1 min, instead of extracting solution (Arakaki et al., 2013) for 60 min. The samples were then subjected to two primary antibodies and two secondary antibodies, each for 1 h at 37°C. The primary antibodies were a rat anti-tubulin α antibody (clone

YL1/2, Bio-Rad Laboratories, Inc., Hercules, CA, USA) and a rabbit anti-*Chlamydomonas reinhardtii* SAS-6 antibody (Nakazawa et al., 2007, kindly provided by Dr. Akira Noga at Paul Scherrer Institut and Dr. Masafumi Hirono at Hosei University) diluted 1:500 with blocking buffer (0.44% gelatin [Sigma-Aldrich], 0.05% NaN₃, and 1% BSA [bovine serum albumin, Sigma-Aldrich] in PBS). The secondary antibodies were an Alexa Fluor 568-conjugated goat anti-rat IgG (Invitrogen, Carlsbad, CA, USA) and an Alexa Fluor 488-conjugated goat anti-rabbit IgG (Invitrogen) diluted 1:500 with the blocking buffer described above. The samples were placed on a drop of SlowFade Gold Antifade Mountant (Thermo Fisher Scientific Inc., Waltham, MA, USA) mixed with 0.5 µL of 0.1 mg/mL DAPI (Sigma-Aldrich). Preparations were observed under a BX-60 microscope (Olympus) equipped with Nomarski interference optics, a mercury lamp, and filter sets with DP Controller 1.2.1108 (Olympus). Differential interference contrast and fluorescence images were merged using Adobe Photoshop CC (Adobe Systems Inc.). For quantitative analysis of the shapes of embryos in successive developmental stages, I measured the diameter of the embryo/daughter colony in both species, as well as the distance between basal bodies of a diagonally opposed pair of peripheral protoplasts/cells in *G. pectorale* or a diagonally opposed pair of protoplasts/cells in *T. socialis*, using ImageJ 1.50b (National Institutes of Health).

3.3. Results

3.3.1. Establishment of a new strain of *Gonium pectorale*

As the *Gonium* strains used in recent genome (Hanschen et al., 2016) and morphological studies (Arakaki et al., 2013, 2017) frequently formed concave colonies and did not form completely flattened 16-celled colonies, a new strains of *Gonium*, 2017-0423-IsgGo1 was established by pipette-washing method (Pringsheim, 1946) from a soil sample of rice paddy field in Okinawa Prefecture. The morphological features of vegetative colonies in this strain (Fig. 3-1) corresponded to those in *G. pectorale*: the maximum number of cells in a vegetative colony is 16, each cell has only one pyrenoid and all 8-celled colonies consist of four zigzag rows of two cells each (Hayama et al., 2010). This strain also showed completely flattened 16-celled colonies. The phylogenetic analysis using coding region (1,128 bp) of *rbcL* gene also positioned this strain within species *Gonium pectorale* (Fig. 3-2).

3.3.2. Time-lapse imaging of embryogenesis in *Gonium pectorale* and *Tetrabaena socialis*

Successive images of embryogenesis in *G. pectorale* from the anterior view and the anterior–lateral view were obtained and analyzed as movies. In *G. pectorale*, each cell in vegetative colonies performed four successive cell divisions to form 16-celled embryos (Fig. 3-3a–e). The cleavage patterns were as described in previous studies on *G. pectorale* (Harper, 1912; Gerisch, 1959; Hayama et al., 2010): the third cell division occurred nearly parallel to the first division to form an 8-celled embryo with a pattern of four zigzag rows of two protoplasts each (Fig. 3-3d), then the fourth cell division occurred nearly parallel to the second division to form a 16-celled embryo with four central protoplasts (A₃, A₃', B₃, and B₃' in Fig. 3-3e) and 12 peripheral protoplasts (Fig. 3-3e).

A time-lapse analysis of the anterior–lateral view of embryogenesis in *G. pectorale* (Fig. 3-4) showed that the outward rotation of daughter protoplasts, as observed in *Astrephomene* (Chapter 2),

did not occur during successive cell divisions, and the apical ends of the daughter protoplasts (transparent region opposite to the green chloroplasts) did not separate from one another (Fig. 3-4a, b). As a result, a cup-shaped cell layer with the apical ends of daughter protoplasts positioned inside the concave surface of the cup-shaped 16-celled embryo was formed just after the fourth cell division (Fig. 3-4c). At this point, each daughter protoplast appeared to be pear-shaped, with an acute apical end and a distended chloroplast end. The concave surface of the cup-shaped embryo then expanded, with the longitudinal axes of the 12 peripheral daughter protoplasts inclining toward the outside of the embryo (Fig. 3-4d). During this expansion, each daughter protoplast began to extend flagella from the apical end and changed its shape from pear-shaped to ovoid (Fig. 3-4d). The formation of stalks or acute chloroplast ends, seen during inversion in volvocacean species, was not observed in daughter protoplasts during this stage. Just prior to hatching, the cell layer of daughter colonies was still concave or distorted and did not fully expand within the mother cell wall (Fig. 3-4e). This distortion of the cell layer dissolved and the daughter colony fully expanded after it hatched out of the mother cell wall, as described in previous reports (Harper, 1912; Hartmann, 1924).

Successive images of embryogenesis in *T. socialis* from the anterior view and the anterior–lateral view were also obtained and analyzed as movies. The cleavage patterns in *T. socialis* (Fig. 3-3f–h) were identical to those previously observed (Arakaki et al., 2013). In the 4-celled embryo just after successive cell divisions, one pair of diagonally opposed daughter protoplasts (B and B' in Fig. 3-3h) were slightly attached to each other at the center of the embryo, while the other pair were separated from each other (A and A' in Fig. 3-3h). A similar daughter protoplast arrangement was also observed in the 4-celled embryos of *G. pectorale* (Fig. 3-3c) and has been observed in other species of volvocine algae (Hallmann, 2006).

The time-lapse analysis from the anterior–lateral view of embryogenesis in *T. socialis* provided detailed insight into the behavior of daughter protoplasts (Fig. 3-5). The daughter protoplasts did not rotate during successive cell divisions (Fig. 3-5a–c), which resulted in a 4-celled embryo with the

longitudinal axes of the daughter protoplasts almost parallel to the anterior–posterior axis of the embryo (or the mother cell before embryogenesis; Fig. 3-5c). After successive cell divisions, one pair of diagonally opposed daughter protoplasts (corresponding to A and A' in Fig. 3-3h) shifted slightly toward the anterior of the embryo relative to the other pair (corresponding to B and B' in Fig. 3-3h; Fig. 3-5c–e). During this process, each daughter protoplast gradually emitted flagella from its apical end and changed its shape from a quarter of an ellipsoid to ovoid, without forming acute chloroplast ends (Fig. 3-5c–e).

3.3.3 Indirect immunofluorescence microscopy of basal bodies, nuclei, and microtubules

Cellular behaviors during embryogenesis in *G. pectorale* and *T. socialis* were examined at the subcellular level using indirect immunofluorescence microscopy of basal bodies, nuclei, and microtubules. The two basal bodies and two pro-basal bodies in each cell were indicated as four dots by localization of SAS-6 (see Materials and Methods). The locations of basal bodies and nuclei (stained with DAPI), as well as chloroplasts (observed in DIC images), clearly showed the longitudinal axis of each daughter protoplast during embryogenesis (Figs. 3-6 and 3-7). For identification of the stages of embryogenesis, an antibody against tubulin α was also used to stain cortical microtubules and the flagella that developed after successive cell divisions.

The staining of basal bodies allowed visualization of the expansion of the apical surface of the concave or cup-shaped 16-celled embryo in *G. pectorale* (Fig. 3-6). During successive cell divisions, the basal bodies of daughter protoplasts were located within the central region of the cup-shaped embryos and did not change their positions much (Fig. 3-6a–c). This situation of the basal bodies during successive divisions is consistent with the absence of rotation of daughter protoplasts observed in time-lapse imaging (Fig. 3-4a–c). After successive cell divisions, the basal bodies of the 12 peripheral protoplasts positioned in the center of the cup-shaped embryo moved from the center to the periphery in accordance with expansion of the cell layer of embryos (Fig. 3-6d). The basal bodies

of each protoplast were positioned more peripherally than the longitudinal axis of each daughter protoplast (Fig. 3-6d, e). After daughter colonies hatched from the mother cell wall, the basal bodies of their peripheral cells were further separated from one another and pointed outward (Fig. 3-6e), but they did not move toward the posterior region as in successive cell divisions of *Astrephomene* (Chapter 2).

In contrast to the apparent movement of the basal bodies in *G. pectorale*, the basal bodies of daughter protoplasts did not show obvious movement during embryogenesis in *T. socialis* (Fig. 3-7). During successive cell divisions, the basal bodies of daughter protoplasts were positioned in the anterior face of the embryo (Fig. 3-7a–c). After successive cell divisions, the basal bodies of a pair of diagonally opposed daughter protoplasts (corresponding to A and A' in Fig. 3-3h) shifted slightly toward the anterior relative to those in the other pair (corresponding to B and B' in Fig. 3-3h; Fig. 3-7d). No significant movement of the basal bodies was observed in developing embryos within the cell wall, except for a slight shifting parallel to the anterior–posterior axis of the embryo. After daughter colonies hatched out from the mother cell wall, I observed no shifting of their cells, and the basal bodies of four cells were arranged to form a square shape in the same plane (Fig. 3-7e), which suggests that flattening and extension of the cell layer occurred after hatching.

The morphological changes in the cell layer after successive cell divisions during embryogenesis in *G. pectorale* and *T. socialis* described above were verified quantitatively by measuring the distance between basal bodies relative to the diameter of the embryo/daughter colony (Figs. 3-8 and 3-9). In *G. pectorale*, the ratio of the distance between basal bodies of a diagonally opposed pair of peripheral protoplasts to the diameter of the 16-celled embryo/daughter colony increased significantly after the formation of flagella (Fig. 3-8a, b). It further increased after the daughter colony hatched from the mother cell wall (Fig. 3-8c, d). In *T. socialis*, on the other hand, the ratio of the distance between the basal bodies of two pairs of diagonally opposed protoplasts/cells to the diameter of the 4-celled embryo/daughter colony did not increase following the formation of

flagella before hatching (Fig. 3-9a, b, d, e), but increased after hatching in both pairs of diagonally opposed cells (Fig. 3-9c, f, g).

3.4. Discussion

The present results clearly showed a fundamental difference in the cellular mechanisms of embryogenesis between ancestral flattened colonies, *Gonium pectorale* and *Tetrabaena socialis* (Fig. 3-10), and spheroidal colonies in the volvocine lineage (Chapter 2, Fig. 2-10). Neither *G. pectorale* nor *T. socialis* showed the rotation of daughter protoplasts during successive cell divisions like *Astrephomene* (Figs. 3-4a–c, 3-5a–c, 3-6a–c and 3-7a–c) (Chapter 2). Moreover, though the daughter protoplasts of *G. pectorale* and *T. socialis* underwent slight changes in shape, they did not display the formation of acute chloroplast ends after cell division that is seen during inversion in volvocacean species (Figs. 3-4c–e, 3-5c–e, 3-6c–e and 3-7c–e) (Pickett-Heaps, 1970; Gottlieb and Goldstein, 1977; Kelland, 1977; Marchant, 1977; Viamontes and Kirk, 1977; Fulton, 1978; Viamontes et al., 1979; Ireland and Hawkins, 1981; Höhn and Hallmann, 2011; Iida et al., 2013; Höhn and Hallmann, 2016) (Chapter 2). Considering these developmental characteristics along with the phylogenetic relationships of these volvocine lineages (Chapter 1, Fig. 1-2) (Nozaki et al., 2000), the most likely scenario for the evolution of these cellular mechanisms in association with the emergence of spheroidal colonies is that the ancestor of *Astrephomene* developed the rotation of daughter protoplasts after it diverged from the ancestors of *Gonium*, while the ancestor of Volvocaceae acquired the formation of acute chloroplast ends after it diverged from the ancestors of Goniaceae (Fig. 3-11).

Moreover, this study clearly demonstrated the details of the formation of flattened colonies in *G. pectorale* and *T. socialis* (Fig. 3-10). A recent study on embryogenesis in *G. pectorale* reported that a “partial inversion” occurred within the mother cell wall (Iida et al., 2013), while earlier studies reported that the flattening of the cell layer coincided with the breaking of the mother cell wall or hatching of the daughter colony (Harper, 1912; Hartmann, 1924). My time-lapse analyses were consistent with both: the cup-shaped embryos formed after successive cell divisions expanded

gradually within the mother cell wall (Fig. 3-4c–e) and the further flattening of the cell layer occurred simultaneously with hatching. Immunofluorescence microscopy staining the basal bodies during these stages also allowed clear visualization of these developmental processes (Figs. 3-6c–e, 3-8). This stepwise formation of flattened colonies was also observed in the embryogenesis of *T. socialis*: the slight shifting between the two pairs of diagonally opposed protoplasts occurred within the mother cell wall (Figs. 3-5c–e, 3-7c, d) and the cell layer was flattened after hatching (Figs. 3-7e, 3-9).

3.5. Tables and Figures

Table 3-1: List of *rbcL* genes used in the phylogenetic analysis of *Gonium* (Fig. 3-2).

Taxon	Strain designation	Accession number
<i>Gonium pectorale</i>	2017-0423-IsgGo1	LC459976 ^c
	NIES ^a -569	D63437
	NIES-1713	AB246189
	NIES-1711	AB246190
<i>Gonium maiaprilis</i>	NIES-2455	AB520743
<i>Gonium multicoccum</i>	UTEX ^b 2580	D63435
	UTEX 783	AB076103
	NIES-1038	AB246187
	NIES-1708	AB246188
<i>Gonium octonarium</i>	NIES-851	D63436
<i>Gonium quadratum</i>	NIES-653	D63438
<i>Gonium viridistellatum</i>	NIES-654	D86831
	NIES-289	AB076091
	NIES-1122	AB076093
<i>Astrephomene gubernaculifera</i>	NIES-418	D63428
	UTEX 1394	AB044170
<i>Astrephomene perforata</i>	NIES-564	D63429

^a Microbial Culture Collection at the National Institute for Environmental Studies (Kawachi et al., 2013).

^b Culture Collection of Algae at the University of Texas at Austin (Starr and Zeikus, 1993).

^c Sequenced in this study.

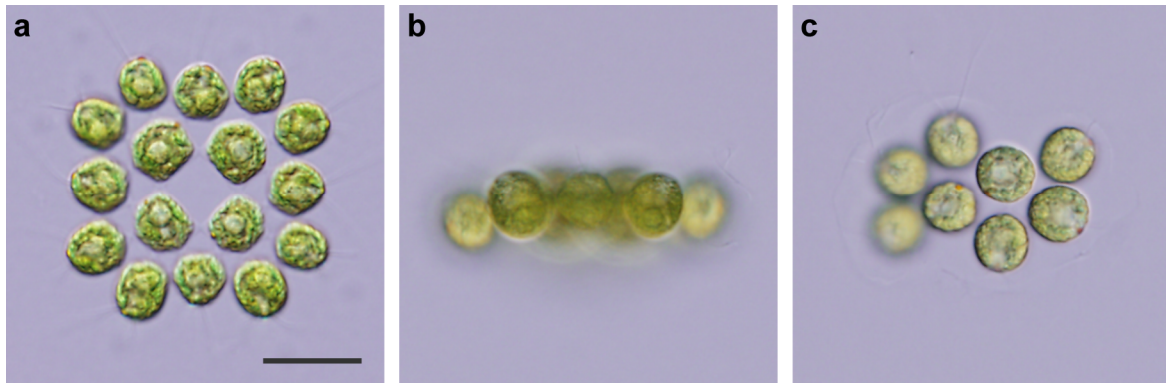


Fig. 3-1: Vegetative colonies in *Gonium pectorale* strain 2017-0423-IsgGo1. (a) 16-celled colony from the anterior view. Each cell has only one pyrenoid. (b) 16-celled colony from the lateral view. (c) 8-celled colony from the anterior view. All 8-celled colonies in this strain consist of four zigzag rows of two cells each, which is a characteristic of *G. pectorale* (Hayama et al., 2010). All images are at the same magnification. Scale bar: 20 μm .

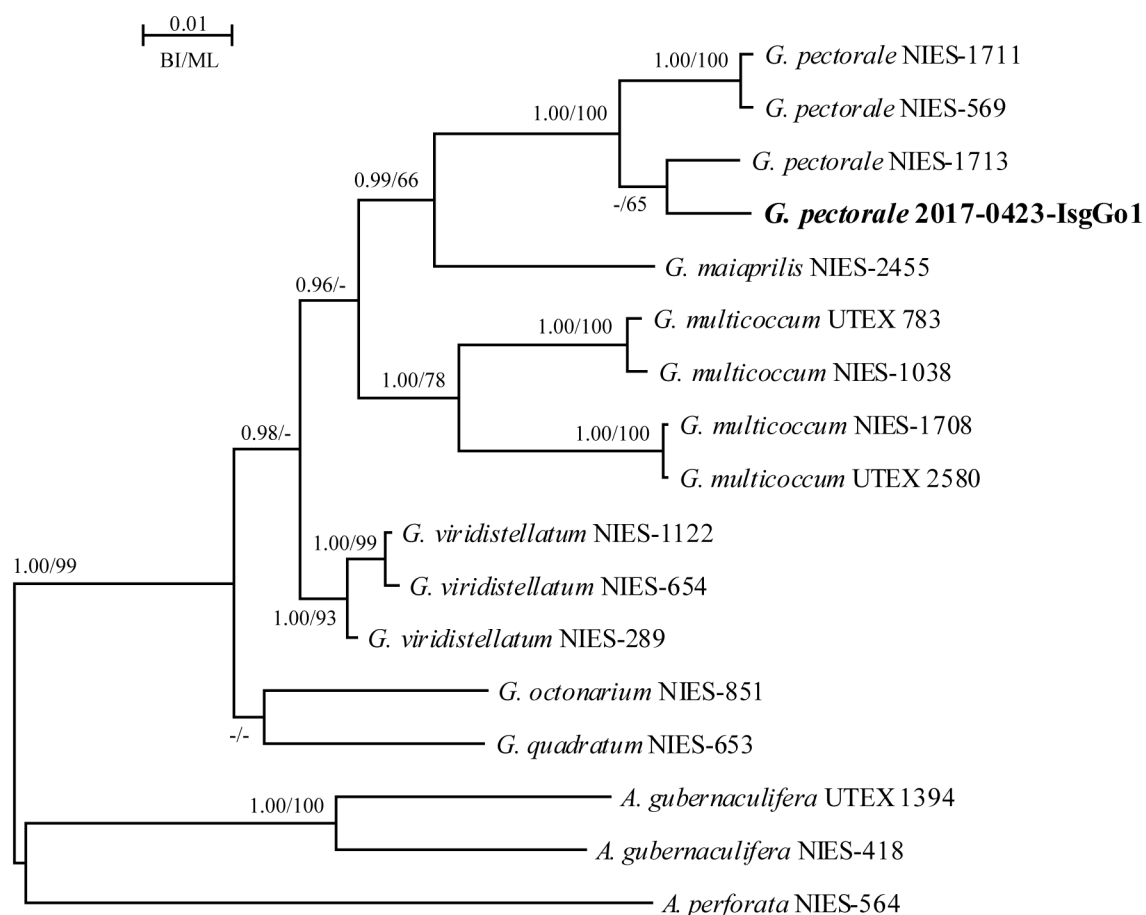
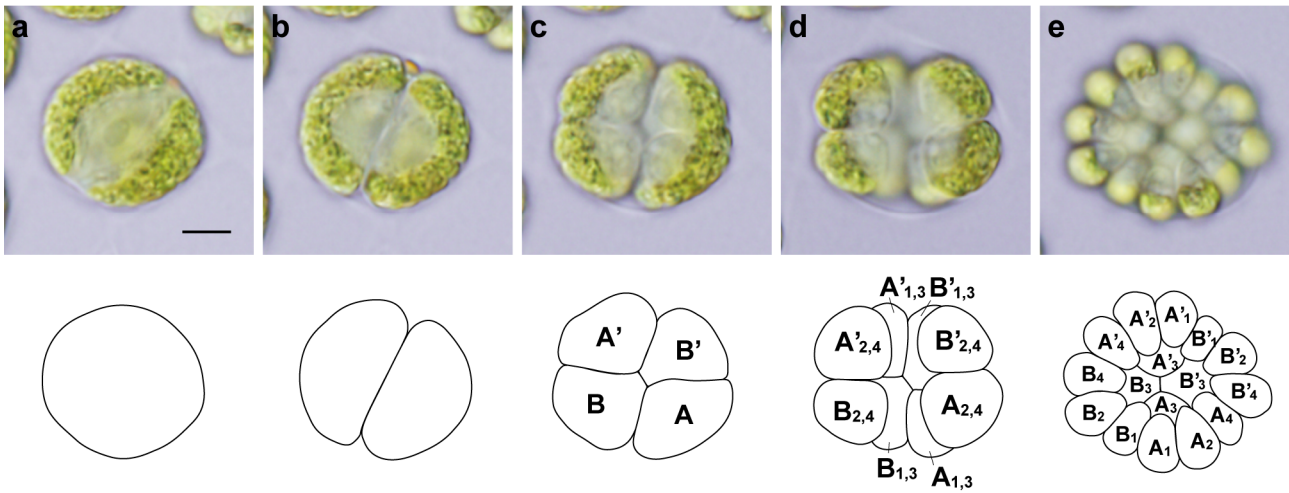


Fig. 3-2: Bayesian phylogenetic tree of *Gonium* based on *rbcL* genes. The 1,128 bp corresponded to position 31-1,158 of *rbcL* from 14 OTUs of *Gonium* as well as 3 OTUs of *Astrephomene* (as outgroup) (Table 3-1) were subjected to Bayesian inference and maximum-likelihood analysis (GTR + I + G model). The posterior probabilities (≥ 0.95) from Bayesian inference (left) and bootstrap values ($\geq 50\%$) from maximum-likelihood (right) are shown at branches.

Gonium



Tetrabaena

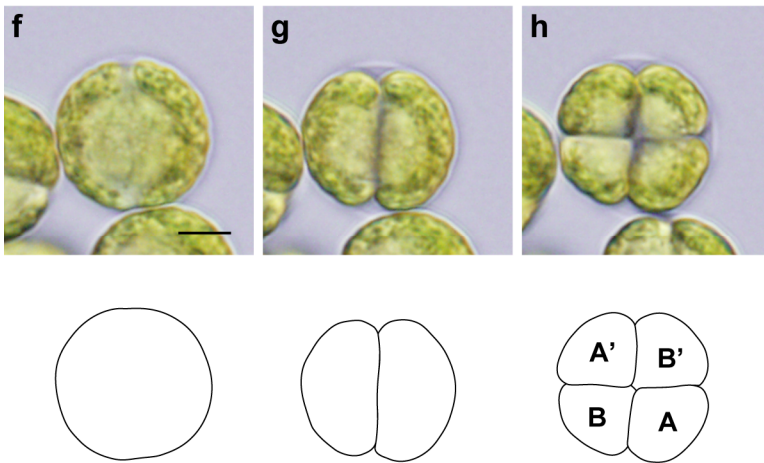


Fig. 3-3: Cleavage patterns and cell lineages in embryogenesis of *Gonium pectorale* and *Tetrabaena socialis*. Based on the time-lapse imaging of each species. Outlines of daughter protoplasts were traced from each image. All images in each row are at the same magnification. Scale bars: 5 μ m. (a–e) Successive images of embryogenesis in *G. pectorale* from the anterior view. (a) Prior to embryogenesis. (b) Two-celled stage. (c) 4-celled stage. (d) 8-celled stage. (e) 16-celled stage. Cleavage patterns were 180° rotationally symmetrical about the longitudinal axis of the embryo and essentially the same as described in previous studies (Harper, 1912; Gerisch, 1959; Iida et al., 2013). The naming of each daughter protoplast followed that in *A. gubernaculifera* (Chapter 2, Fig. 2-4), since the cleavage patterns up to 16-celled stage of *G. pectorale* and *A. gubernaculifera* are almost identical except for the direction of daughter protoplasts (Nozaki and Ito, 1994). (f–h) Successive images of embryogenesis in *T. socialis* from the anterior view. (f) Prior to embryogenesis. (g) Two-celled stage. (h) 4-celled stage. The cleavage patterns were essentially the same as described in a previous study (Arakaki et al., 2013). Note that one pair of diagonally opposed daughter protoplasts were separated from each other (A and A' in h) while the other pair were slightly attached to each other at the center of the embryo (B and B' in h).

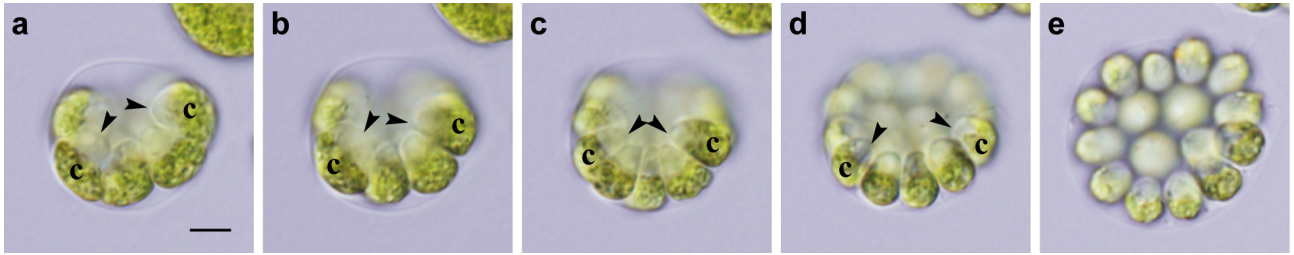


Fig. 3-4: Successive cell divisions and expansion of the cell layer in embryogenesis of *Gonium pectorale*. Successive stages of an embryo observed by time-lapse analysis from the anterior–lateral view. All images are at the same magnification. Scale bar: 5 μ m. Note the longitudinal axis of each daughter protoplast, indicated by the positions of apical ends (arrowheads) and chloroplasts (letter c). (a) Early 8-celled stage. (b) Late 8-celled stage. (c) Early 16-celled stage. Rotation of daughter protoplasts was not observed during cell divisions (a–c). (d) Mid 16-celled stage during partial inversion. The cup-shaped cell layer of the embryo expanded and the apical ends of outer daughter protoplasts (arrowheads) were separated from one another and moved outwards. (e) Late 16-celled stage, just prior to hatching. The cell layer expanded increasingly and rotated slightly to show an almost anterior view.

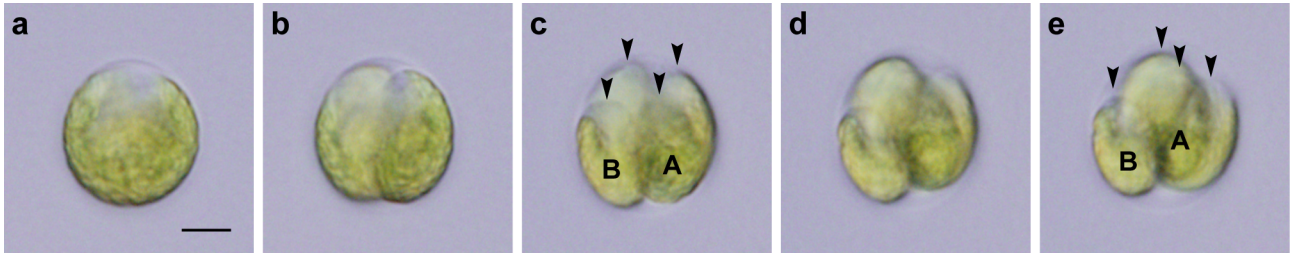


Fig. 3-5: Successive cell divisions and slight shifting of daughter protoplasts during embryogenesis of *Tetrabaena socialis*. Successive stages of an embryo observed by time-lapse analysis from the anterior–lateral view. All images are at the same magnification. Scale bar: 5 μm . Note the longitudinal axis of each daughter protoplast, indicated by the positions of apical ends (arrowheads) and chloroplasts (uppercase letters). Uppercase letters (A and B) correspond with those in Fig. 3-3h. (a) Prior to embryogenesis. (b) Two-celled stage. (c) Early 4-celled stage. Daughter protoplasts do not rotate during cell divisions (a–c). (d) Mid 4-celled stage. (e) Late 4-celled stage. One pair of diagonally opposed daughter protoplasts (A) shifted slightly toward the anterior of the embryo relative to the other pair (B) (c–e).

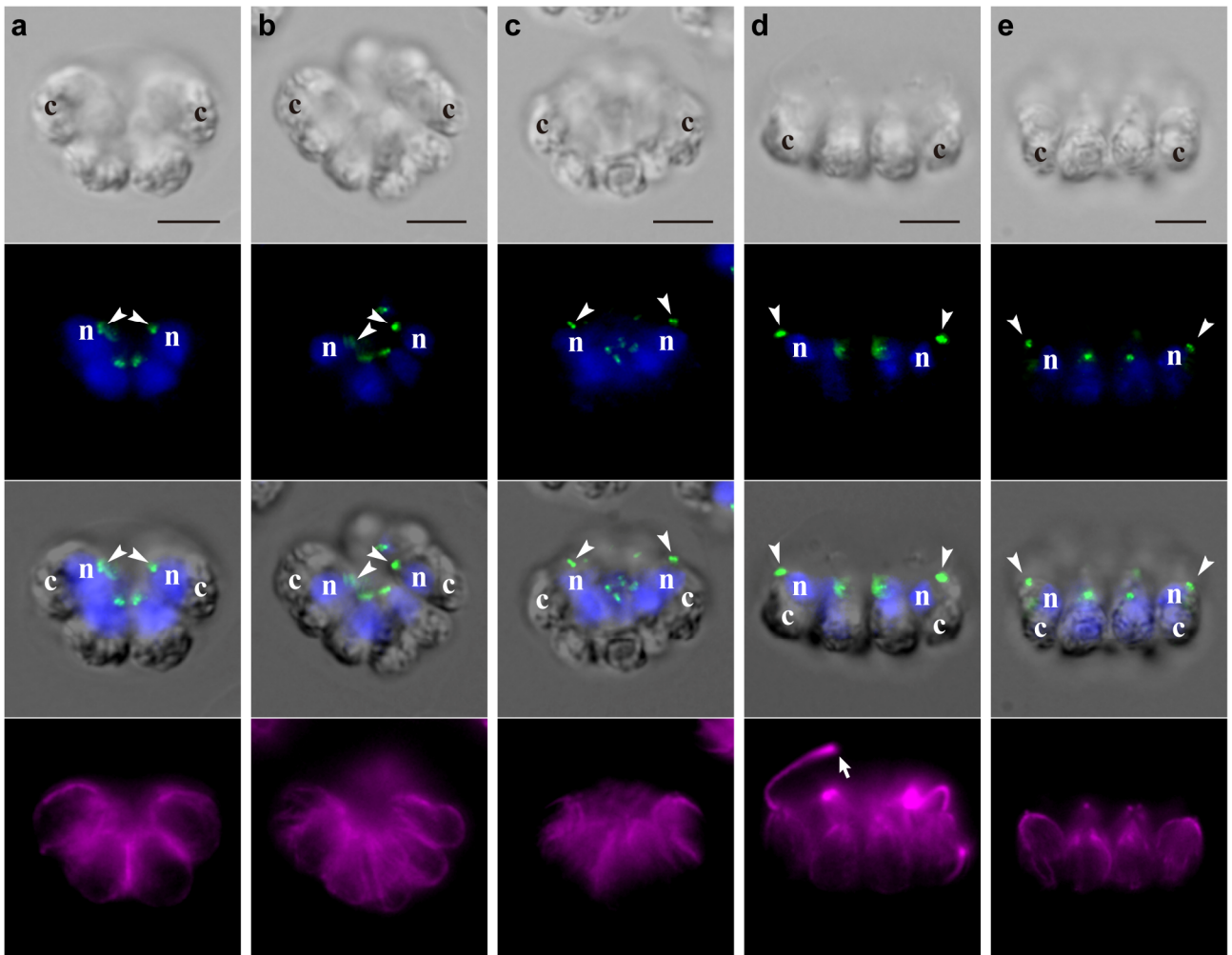


Fig. 3-6: Indirect immunofluorescence microscopy showing successive stages of embryogenesis in *Gonium pectorale*. Each column shows a differential interference contrast (DIC) image (top row), a fluorescence image labeled with anti-SAS-6 antibody (green) and DAPI (blue) (second row), a merged image (third row), and a fluorescence image labeled with anti-tubulin α antibody (magenta) of the same embryo. Positions of nuclei (letter n), chloroplasts (letter c), and basal bodies labeled with the anti-SAS-6 antibody (arrowheads) are shown. Scale bars: 5 μ m. **(a)** Early 8-celled stage. **(b)** Late 8-celled stage. **(c)** Early 16-celled stage, just after the successive cell divisions. Basal bodies of daughter protoplasts are positioned in the center of the concave surface of the cell layer during successive cell divisions **(a–c)**. **(d)** Mid 16-celled stage during partial inversion showing emitted flagella (arrow). The basal bodies of peripheral daughter protoplasts (arrowheads) are separated from each other and are located slightly outside of the position of the nuclei. **(e)** After hatching. The basal bodies of peripheral cells (arrowheads) point toward the outside of the daughter colony.

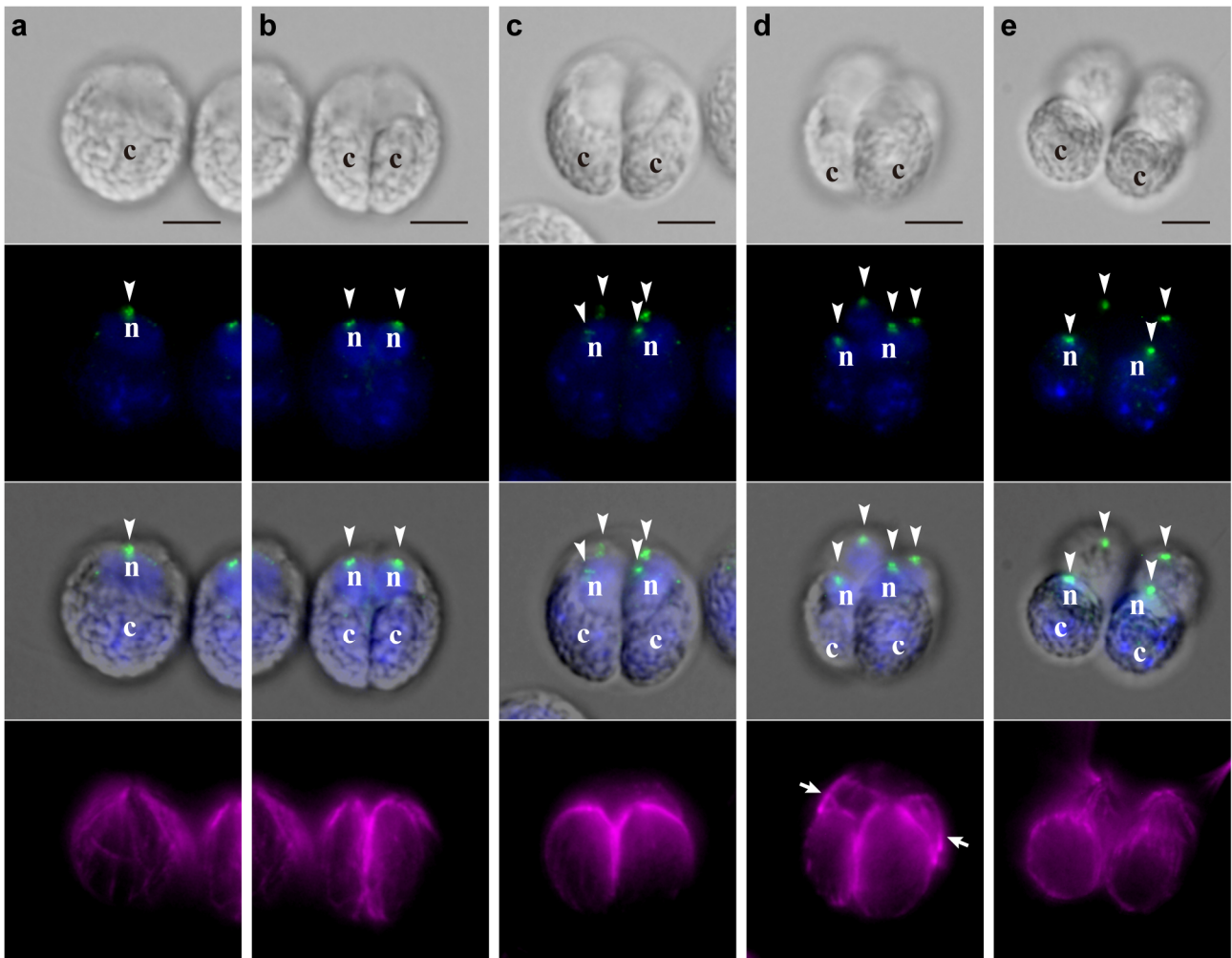


Fig. 3-7: Indirect immunofluorescence microscopy showing successive stages of embryogenesis in *Tetraabaena socialis*. Each column shows a DIC image (top row), a fluorescence image labeled with anti-SAS-6 antibody (green) and DAPI (blue) (second row), a merged image (third row), and a fluorescence image labeled with anti-tubulin α antibody (magenta) of the same embryo. Positions of nuclei (letter n), chloroplasts (letter c), and basal bodies labeled with anti-SAS-6 antibody (arrowheads) are shown. Scale bars: 5 μ m. **(a)** Prior to embryogenesis. **(b)** Two-celled stage. **(c)** Early 4-celled stage. The angles of the longitudinal axes of daughter protoplasts, which are indicated by basal bodies and chloroplasts, did not change during successive cell divisions **(a–c)**. **(d)** Late 4-celled stage showing emitted flagella (arrows). The angles of longitudinal axes of daughter protoplasts did not change after successive cell divisions, though a pair of diagonally opposed daughter protoplasts shifted slightly toward the anterior of the embryo (“A” in Figs. 3-3h, 3-5c, e). **(e)** After hatching. The basal bodies of four cells are arranged in a square shape in the same plane.

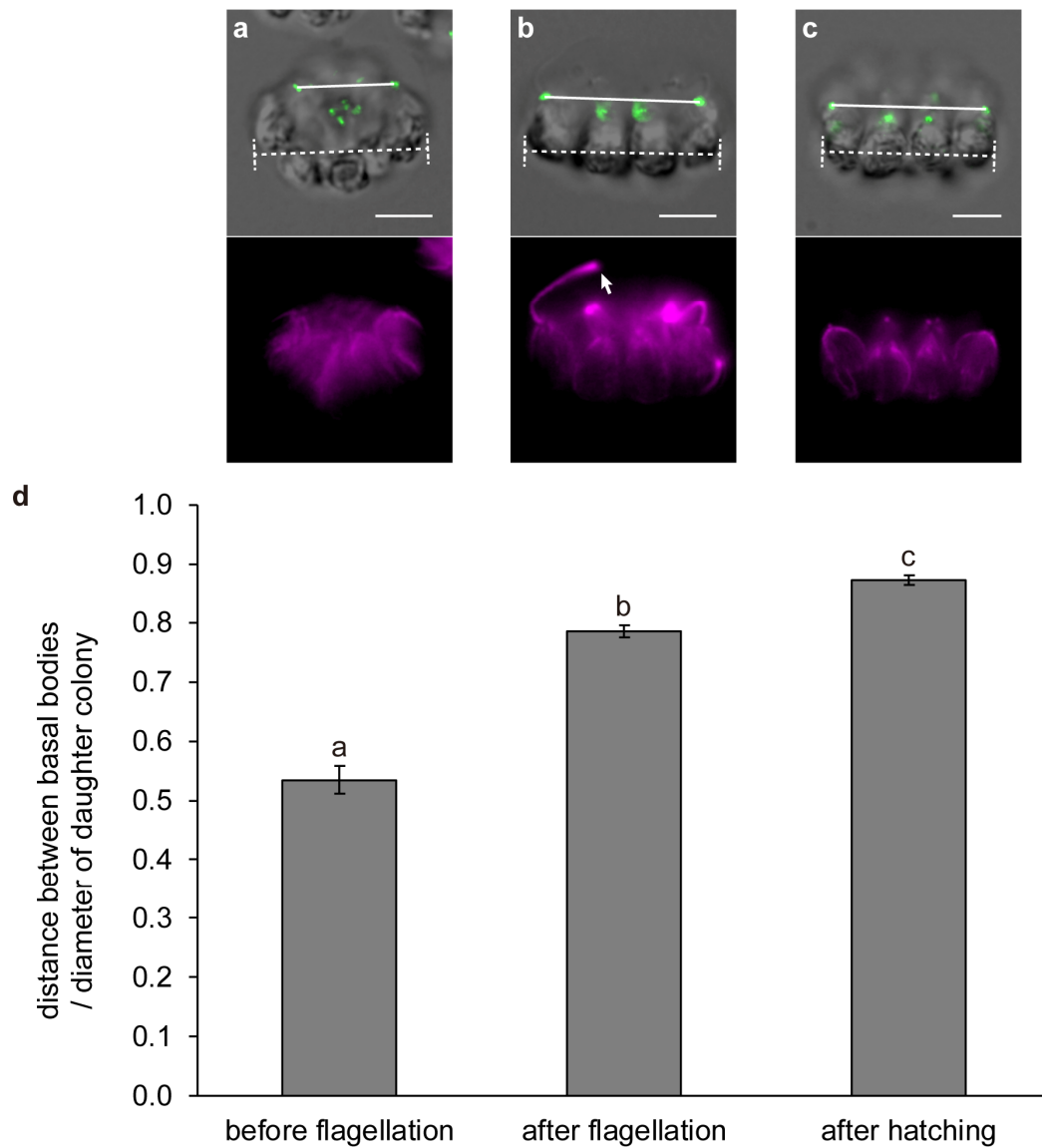


Fig. 3-8: Measurement of morphological changes of the cell layer in 16-celled embryos/daughter colonies during colonial development in *Gonium pectorale*. (a–c) Merged fluorescence images labeled with anti-SAS-6 antibody (green) and DIC images (top row), and fluorescence images labeled with anti-tubulin α antibody (magenta, second row) of examples of 16-celled embryos/daughter colonies before flagellation (a), after flagellation (b; arrow indicates flagellum), and after hatching (c). Scale bars: 5 μm . The distance between basal bodies of diagonally opposed pairs of peripheral protoplasts/cells (solid lines, without distinction of the location of the pair; see Fig. 3-3e) and the diameter of 16-celled embryos/daughter colonies (dotted lines) were measured at each of these stages. (d) The ratio of the distance between the basal bodies of a diagonally opposed pair of peripheral protoplasts/cells to the diameter of embryos/daughter colonies in three stages during development. The stages of development were distinguished as follows: before flagellation (with no flagella or flagella < 1 μm , $n = 22$), after flagellation (with flagella > 5 μm and a mother cell wall, $n = 42$), and after hatching (without a mother cell wall, $n = 20$). Bars represent means \pm standard error (SE). Different letters above bars indicate significant differences of means based on a Tukey–Kramer post-hoc test at $P < 0.05$.

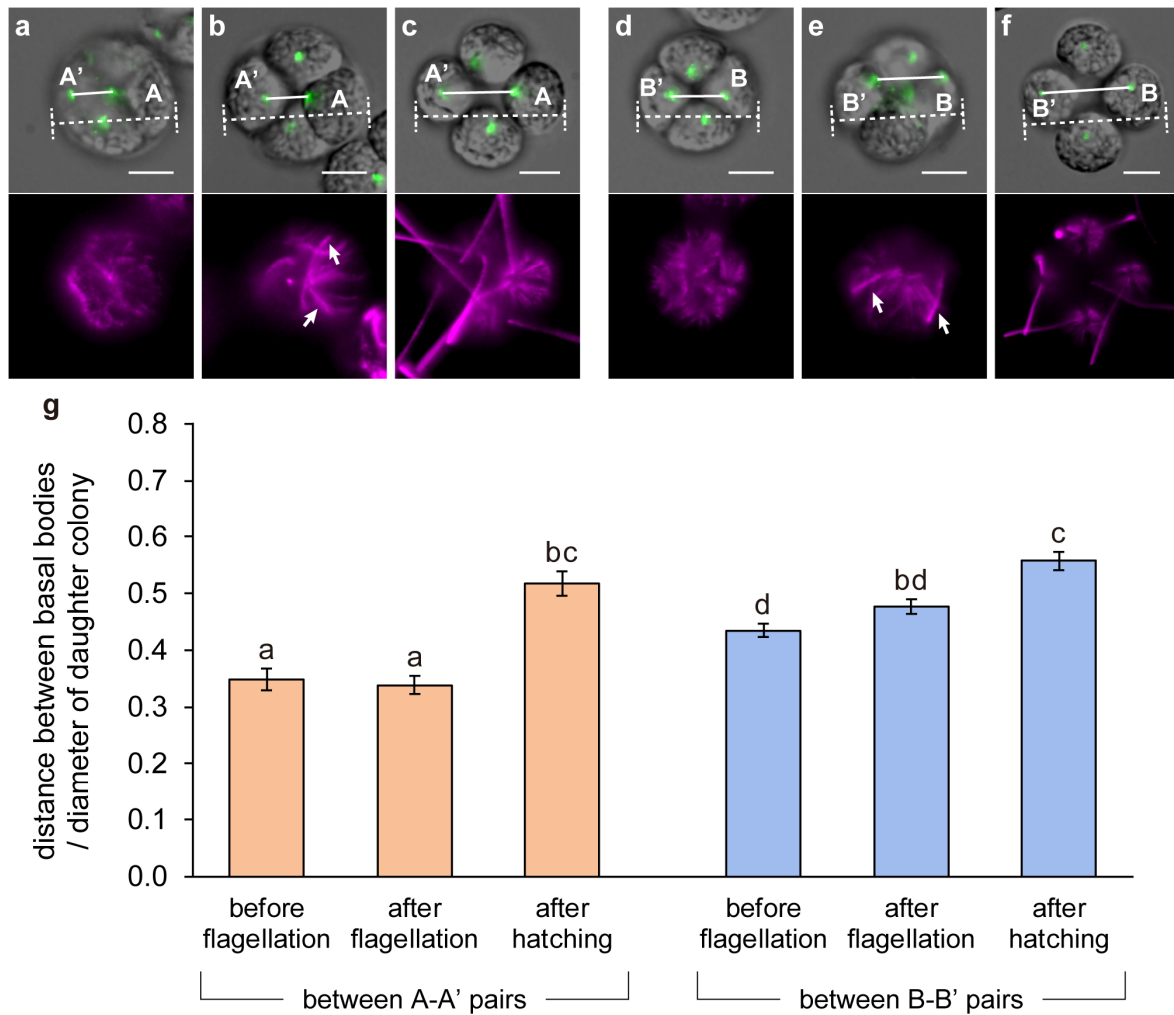


Fig. 3-9: Measurement of morphological changes of the cell layer in 4-celled embryos/daughter colonies during development in *Tetraena socialis*. (a–f) Merged fluorescence images labeled with anti-SAS-6 antibody (green) and DIC images (top row), and fluorescence images labeled with anti-tubulin α antibody (magenta, second row) of examples of 4-celled embryos/daughter colonies before flagellation (a, d), after flagellation (b, e, arrows indicate flagella), and after hatching (c, f). Scale bars: 5 μ m. The distance between basal bodies of a pair of diagonally opposed protoplasts/cells (solid lines) between A-A' pairs (a–c) or B-B' pairs (d–f; see Fig. 3-3h) and the diameter of embryos/daughter colonies (dotted lines) were measured at each of these stages. Uppercase letters (A and B) correspond with those in Fig. 3-3h. (g) The ratio of the distance between basal bodies of a pair of diagonally opposed protoplasts/cells to the diameter of 4-celled embryos/daughter colonies in three developmental stages, with distinction between the two pairs. The stages of development were distinguished as follows: before flagellation (with no flagella or flagella < 1 μ m), after flagellation (with flagella > 5 μ m and a mother cell wall), and after hatching (without a mother cell wall). Bars represent means \pm SE (n = 15–21). As the distance of both pairs could not be measured in some daughter colonies, the distances between basal bodies of A-A' and B-B' cells were treated as independent data. Different letters above bars indicate significant differences of means based on a Tukey–Kramer post-hoc test at $P < 0.05$.

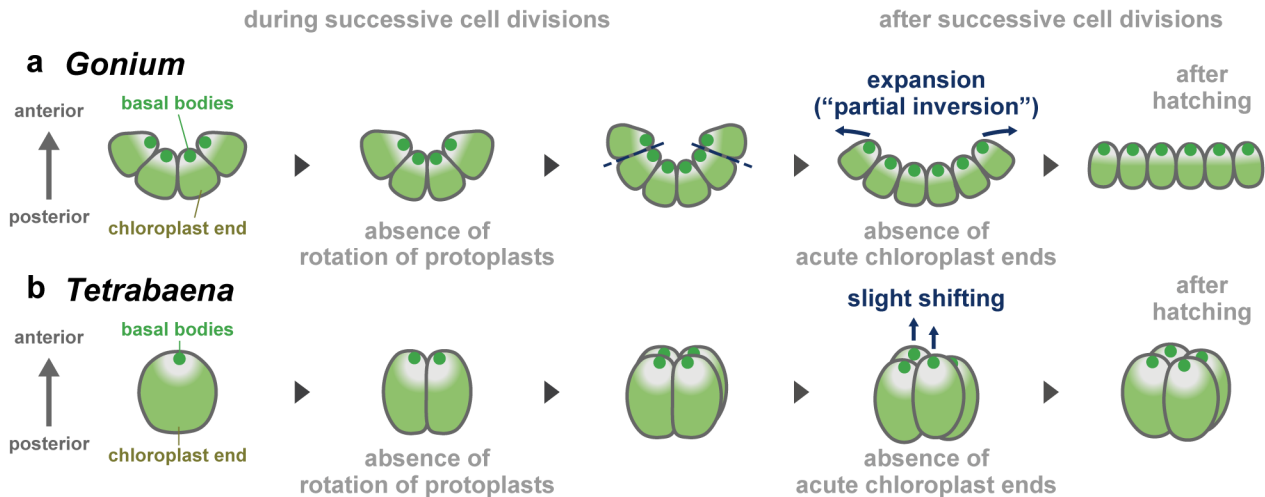


Fig. 3-10: Schematic diagrams of embryogenesis in *Gonium* and *Tetrabaena*. Diagrams are based on the present time-lapse analysis (Figs. 3-4, 3-5) and indirect immunofluorescence microscopy (Figs. 3-6, 3-7) and represent lateral views of stages of embryogenesis with anterior sides of embryos oriented toward the top of the figure. Note that the rotation of daughter protoplasts seen in *Astrephomene* and the formation of acute chloroplast ends of daughter protoplasts seen during inversion in the Volvocaceae were not observed during embryogenesis in *G. pectorale* and *T. socialis*. **(a)** Embryogenesis in *G. pectorale*. Each cell of a vegetative colony performs four successive cell divisions to form a cup-shaped 16-celled embryo with basal bodies positioned in the center of the concave surface of the cell layer. The cell layer of the embryo expands gradually after successive cell divisions and instantly upon hatching of the daughter colony, with the basal bodies moving from the center to the periphery of the cell layer. **(b)** Embryogenesis in *T. socialis*. Each cell of a vegetative colony performs two successive cell divisions to form a 4-celled embryo with the basal bodies positioned in the anterior face of the embryo. The two pairs of diagonally opposed daughter protoplasts shift slightly after successive cell divisions. The cell layer become flattened and expanded after the daughter colony hatches.

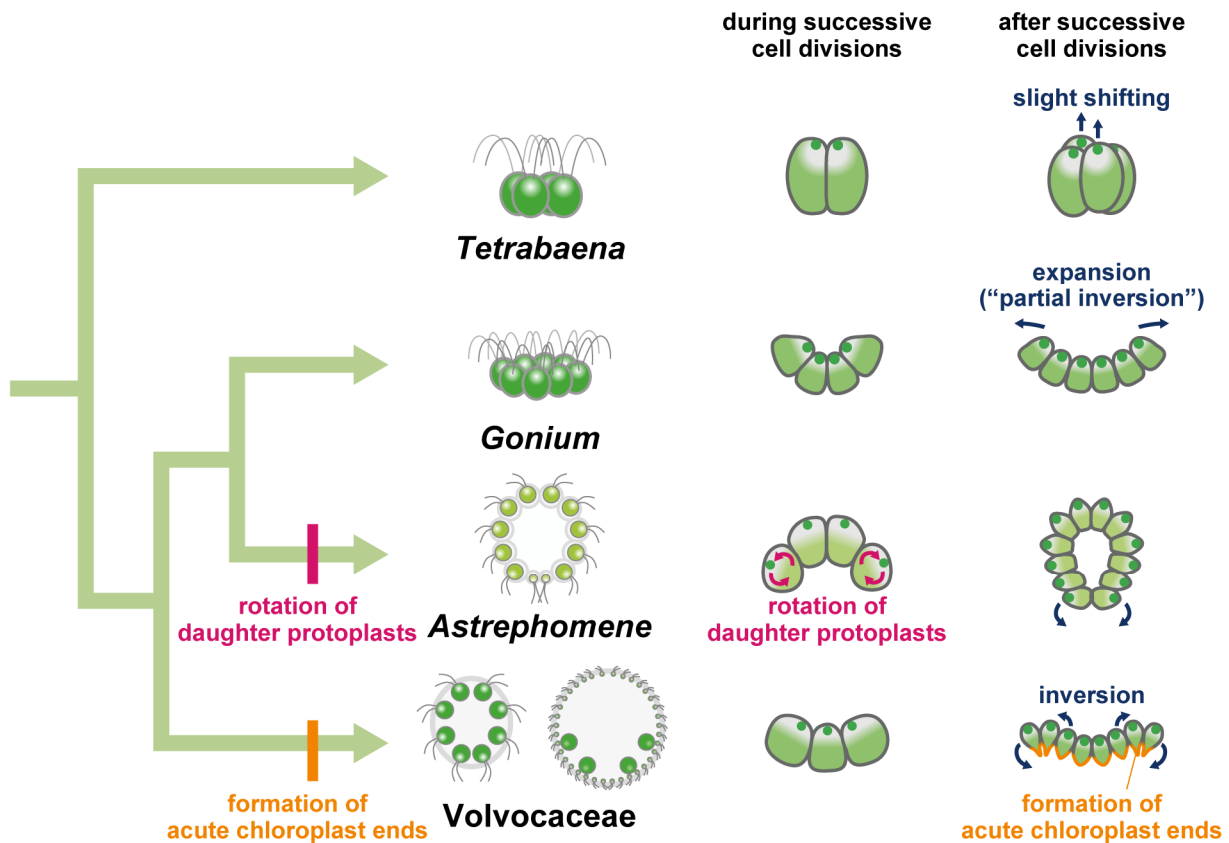


Fig. 3-11: Schematic diagrams of the most likely evolutionary pathways of embryogenesis in volvocine green algae. Diagrams of stages of embryogenesis in *Astrephomene* and *Volvocaceae* (*Eudorina*) shown are based on the previous chapter (summarized in Chapter 2, Fig. 2-10), while those in *Tetrabaena* and *Gonium* are based on the present study (summarized in Fig. 3-10).

Volvocaceae species undergo the formation of acute chloroplast ends of daughter protoplasts, which is one of the principal factors producing the force for folding the cell layer during inversion after successive cell divisions. On the other hand, *Astrephomene* undergoes rotation of daughter protoplasts during successive cell divisions to form a spheroidal shape of the cell layer. Neither *Gonium* nor *Tetrabaena* show the formation of acute chloroplast ends of daughter protoplasts after successive cell divisions or the rotation of daughter protoplasts during successive cell divisions. These results suggest that the ancestor of *Astrephomene* developed the rotation of daughter protoplasts after it diverged from the ancestors of *Gonium*, while the ancestor of *Volvocaceae* acquired the formation of acute chloroplast ends after it diverged from the ancestors of *Goniaceae*.

Chapter 4

Analysis of Nuclear Genome of *Astrephomene*

4.1. Introduction

Astrephomene has evolved germ–soma differentiation and expanded ECM as well as a spheroidal colony independently of Volvocaceae (Nozaki and Ito, 1994; Herron and Michod, 2008; Herron et al., 2009). Though spheroidal colony formation, germ–soma differentiation and expanded ECM and their evolution in *Volvox carteri* have been studied in detail in genetic and genomic levels (see Chapter 1), the genetic bases involved in these multicellular traits in *Astrephomene* and their evolution have not been investigated so far. Thus, whether the genetic and genomic evolution found in *Volvox* also occurred in *Astrephomene* is completely unknown.

In this Chapter, I investigated *de novo* whole nuclear genome of *Astrephomene gubernaculifera* and performed RNA-seq analysis of isolated somatic and reproductive cells of this species. The results demonstrated that the two multicellular traits, germ–soma differentiation and expansion of ECM, might have evolved based on the different molecular genetic bases between *V. carteri* and *A. gubernaculifera*.

4.2. Materials and Methods

4.2.1. Strain and culture conditions

Astrephomene gubernaculifera strain 2014-1002-YkAs8 (established in Chapter 2) was used through this study. In addition, *Eudorina* sp. strain 2010-623-F1-E8 (see Chapter 2), *Gonium pectorale* strain NIES-2863 and *Tetrabaena socialis* strain NIES-571 (from Microbial Culture Collection at the National Institute for Environmental Studies, NIES, Kawachi et al., 2013, <http://mcc.nies.go.jp/>) were used for sequencing of InvA orthologs. All cultures were maintained in 10 mL of VTAC medium (Kawachi et al., 2013; the growth medium in Nozaki et al., 1989) in screw-capped tubes (18 × 150 mm) at 25°C on a 12-h light/12-h dark schedule under cool-white fluorescent lamps at an intensity of 50–90 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$.

For experimental analyses, the synchronous culture of *A. gubernaculifera* was established. *A. gubernaculifera* strain NIES-4017 were grown in silicon-capped 500 mL Erlenmeyer flasks containing 250 mL VTAC medium with aeration, at 32°C on a 16-h light/8-h dark schedule under cool-white fluorescent lamps at an intensity of 140–180 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. The inoculation of culture was conducted every 24 h, just before the onset of the dark period, by transferring 4 mL of culture into new 250 mL VTAC medium in silicon-capped Erlenmeyer flasks. In this condition, the asexual life cycle was completed in approximately 24 h, and the culture was highly synchronized with the light–dark cycle. Almost all reproductive cells (> 95%) of *A. gubernaculifera* initiated embryogenesis approximately 9–10 h after the onset of the light period (Fig. 4-1).

4.2.2. De novo genome sequencing and assembly

For extraction of genomic DNA, young colonies just after embryogenesis in synchronous culture (Fig. 4-1) were used. Genomic DNA from cells was prepared based on the method in a previous study (Miller et al., 1993) with some modifications. The colonies were collected by centrifugation

and mixed with 65 μ l of resuspension buffer (50 mM Tris-HCl, 10 mM EDTA, pH 8.0), 130 μ l of 20% sarcosyl and 500 μ l of glass beads in 15 mL centrifuge tube. The colonies in tube was then vortexed for 1 min six times, diluted by resuspension buffer and 20% sarcosyl and centrifuged at 3,500 rpm for 25 min. The supernatant was divided into approximately 2 ml in new 15 ml centrifuge tubes. An equal volume of 2 \times CTAB buffer (2% CTAB, 100 mM Tris-HCl, 20 mM EDTA, 1.4 M NaCl, 1% polyvinyl pyrrolidone, pH 8.0) prewarmed to 65°C was added to the tubes and the tubes were vortexed for 2 min and incubated at 65°C for 15 min. An equal volume of chloroform was added to the mixture, the tubes were vortexed for 30 s and centrifuged at 3,500 rpm for 25 min. The aqueous layer was transferred to new tubes, one-tenth volume of 10% CTAB buffer (10% CTAB, 0.7 M NaCl) prewarmed to 65°C was added, the tubes were vortexed for 2 min and incubated at 65°C for 20 min, and the samples were extracted with chloroform as described above. The aqueous layer was mixed with an equal volume of CTAB precipitation buffer (1% CTAB, 50 mM Tris-HCl, 10 mM EDTA, pH 8.0) and mixed by inverting and incubated at 65°C for 30 min. The tubes were vortexed for 30 s and centrifuged at 3,500 rpm for 25 min to precipitate DNA. The supernatant was discarded and the pellet was resuspended in high-salt TE buffer (10 mM Tris-HCl, 1 mM EDTA, 1 M NaCl, pH 8.0). DNA was precipitated with ethanol, dissolved and precipitated again, washed with 70% ethanol, dried, and dissolved in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). DNA was then also precipitated with one-tenth volume of 200 mM spermine-4HCl, dissolved to 100 mM EDTA and precipitated with isopropyl alcohol again, washed with 70% ethanol, dried, and dissolved in TE buffer. RNaseA (NIPPON GENE Co. Ltd., Tokyo, Japan) was added to DNA solution and the mixture was incubated at 37°C for 35 min. An equal volume of phenol/chloroform was added to the mixture, the tubes were vortexed for 10 s and centrifuged at 3,500 rpm for 25 min. The aqueous layer was extracted with chloroform as described above and precipitated and washed with ethanol as described above.

The purified DNA was shipped to National Institute of Genetics and sequenced by PacBio

Sequel (Pacific Biosciences, Menlo Park, CA, USA) and Illumina HiSeq 2500 sequencers (Illumina, San Diego, CA, USA). The generated PacBio subreads and Illumina reads were then assembled using Falcon v0.7, Falcon-unzip v0.4 (Chin et al., 2016), SMRT Link v6.0.0.47841 (Pacific Biosciences) and Pilon v1.22 (Walker et al., 2014) for *de novo* nuclear genome assembly of *A. gubernaculifera*. Both sequencing and assembly were conducted by the collaboration with Dr. Yohei Minakuchi and Dr. Atsushi Toyoda at National Institute of Genetics.

4.2.3. Prediction of gene models of *Astrephomene gubernaculifera* using RNA-seq data

For extraction of total RNA for annotation, young colonies and embryos in synchronous culture (Fig. 4-1) were used. RNA was extracted using TRIzol Reagent (Thermo Fisher Scientific Inc., Waltham, MA, USA) according to the manufacture's protocols. The purified RNA was shipped to National Institute of Genetics, Mishima and sequenced by Illumina HiSeq 2500 sequencer (Illumina). The sequenced reads from two samples were merged and subjected to alignment (to *A. gubernaculifera* genome assembly described above), assembly and prediction of CDS using HISAT2 version 2.1.0 (Kim et al., 2019), StringTie v2.0.4 (Kovaka et al., 2019) and TransDecoder v5.5.0 (<https://github.com/TransDecoder/TransDecoder>) for predicted gene models, by the collaboration with Dr. Yohei Minakuchi and Dr. Atsushi Toyoda at National Institute of Genetics.

4.2.4. Prediction of gene models of *Yamagishiella unicocca* and *Eudorina* sp.

The gene prediction against whole genome data of *Yamagishiella unicocca* and *Eudorina* sp. (Hamaji et al., 2018) were conducted by AUGUSTUS v3.3.3 (Stanke et al., 2008) using the default parameter of *V. carteri*. 21,231 gene models were predicted in *Y. unicocca* and 28,661 gene models were predicted in *Eudorina* sp., respectively.

4.2.5. Identification and analysis of orthologs of genes involved in embryogenesis in *Volvox carteri*

The amino acid sequences of genes involved in inversion in *Volvox carteri*, InvA (Acc. No. BAC77722.1) (Nishii et al., 2003), InvB (Acc. No. BAH28849.1) and InvC (Acc. No. BAH03159.1) (Ueki and Nishii, 2008, 2009), and genes involved in asymmetric divisions during embryogenesis, GlsA (Acc. No. AAD26632.1) (Miller and Kirk, 1999) and Hsp70A (Acc. No. AAZ04921.1) (Cheng et al., 2006), were used as queries in BLASTP search (E-value < 1e-10) against gene models of *A. gubernaculifera*, *Y. unicocca* and *Eudorina* sp. described above as well as *Chlamydomonas reinhardtii*, *T. socialis*, *G. pectorale* and *V. carteri* (Table 4-1). A single putative ortholog for each gene in each species was found with identity of sequences higher than other hit genes, except for InvC ortholog in *G. pectorale* and GlsA ortholog in *T. socialis*.

As gene models of *InvA* orthologs in *T. socialis*, *G. pectorale*, *A. gubernaculifera* and *Eudorina* sp. were thought to be incomplete, cDNA sequences were sequenced. The polyadenylated mRNAs from concentrated cells of each species (see Strain and culture conditions above) were isolated using Dynabeads mRNA Purification Kit (Invitrogen, Carlsbad, CA, USA) and then reverse transcribed with Superscript III reverse transcriptase (Invitrogen) according to the manufacture's protocols. For *InvA* ortholog in *G. pectorale* and *Eudorina* sp., The amplification of cDNA sequences was conducted using KOD FX Neo (TOYOBO Co. Ltd., Osaka, Japan), specific primers designed based on genome data (Table 4-2), with following PCR schedule: 94°C for 2 min, following 35 cycles of 98°C for 10 s and 68°C for 40 s. PCR products were purified using illustra GFX PCR DNA and Gel Band Purification Kit (GE healthcare, Little Chalfont, UK) according to the manufacture's protocol. Sequencing of the purified PCR products was carried out using an ABI PRISM 3100 Genetic Analyser (Applied Biosystems, Foster City, CA, USA) with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). For *InvA* orthologs in *T. socialis* and *A. gubernaculifera*, partial cDNA sequences were firstly amplified by nested PCR using TaKaRa Taq (Takara Bio Inc.,

Shiga, Japan) and degenerate primers (Table 4-2) with following PCR schedule: 35 cycles of 95°C for 2 min, 46°C for 2 min, and 66°C for 3 min, followed by 72°C for 15 min. Purification and sequencing the PCR products were conducted as described above. Then, 5' and 3' RACE were conducted for full length cDNA sequences using polyadenylated mRNAs isolated as described above and GeneRacer Kit (Invitrogen) and Superscript III reverse transcriptase (Invitrogen) according to the manufacture's protocols. Nested PCR was conducted using GeneRacer primers in the kit and specific primers (Table 4-2), KOD FX Neo (TOYOBO Co. Ltd.) with 'step-down cycle' on the manufacture's protocol of KOD FX Neo. Purification and sequencing the PCR products were conducted as described above.

The amino acid sequences of InvA, InvB, InvC, GlcA and Hsp70A orthologs were aligned by MUSCLE v3.8.31 (Edgar, 2004) and alignments were corrected manually. The aligned amino acid sequences were subjected to maximum-likelihood analysis. Evolutionary models for maximum-likelihood analysis were selected using ModelTest-NG (Darriba et al., 2020). The maximum-likelihood analysis was conducted with 1,000 replicates of bootstrap analyses (Felsenstein, 1985) using RAxML-NG (Kozlov et al., 2019). Orthologs in *C. reinhardtii* were treated as outgroup based on the phylogenetic relationships of the volvocine algae (Nozaki et al., 2000).

4.2.6. Identification and analysis of VARL genes

The master regulator of somatic cells in *V. carteri*, *regA*, encodes a transcription factor with a DNA-binding SAND-like domain (called VARL domain, volvocine algae *regA*-like) (Kirk et al., 1999). *regA* and other previously reported VARL genes (genes encoding putative transcription factors with VARL domain) in *V. carteri*, *Y. unicocca*, *G. pectorale* and *C. reinhardtii* (Duncan et al., 2007; Hanschen et al., 2016; Grochau-Wright et al., 2017) were used as queries in BLASTP search (E-value < 1e-10) against gene models of *A. gubernaculifera* described above. The VARL domain of each gene, the N-terminal extension and core VARL domain structure (Duncan et al., 2007), was

identified based on alignment with VARL genes in volvocine algae (Grochau-Wright et al., 2017). The 8 VARL genes in *T. socialis* corresponding to those reported in a previous study (Featherston et al., 2018) and their VARL domains were also identified by the same method. The amino acid sequences of VARL domains from *A. gubernaculifera*, *T. socialis* and other volvocine algae (Grochau-Wright et al., 2017) were aligned by MUSCLE (Edgar, 2004) and by manually on MEGA 7.0.21 (Kumar et al., 2016). The alignment of 87 amino acid positions was subjected to maximum-likelihood analysis and Bayesian interference. For evolutionary model for Bayesian interference and maximum-likelihood analysis, LG + I + G model was selected by ModelTest-NG (Darriba et al., 2020). The maximum-likelihood analysis was conducted with 1,000 replicates of bootstrap analyses (Felsenstein, 1985) using RAxML-NG (Kozlov et al., 2019). Bayesian interference was performed using MrBayes 3.2.6 (Ronquist et al., 2012) with 11,000,000 generations of Markov chain Monte Carlo iterations; the first 25% of the generations were discarded as burn-in.

The gene synteny near *RLS1/rlsD* ortholog in volvocine algae were defined based on BLASTP search (E-value < 1e-10) using amino acid sequences of genes near *RLS1/rlsD* ortholog in genome sequences of *V. carteri*, *Y. unicocca*, *G. pectorale* and *C. reinhardtii* and *reg* cluster genes in *V. carteri*, *Y. unicocca* (Hanschen et al., 2014, 2016; Grochau-Wright et al., 2017) against gene models of *A. gubernaculifera*. For gene model of *Y. unicocca*, gene models near *RLS1/rlsD* ortholog and *reg* cluster genes predicted in a previous study (Grochau-Wright et al., 2017) (Acc. No. KU257988.1) was used. According to phylogenetic analysis and gene synteny, Agub_g12775 was identified as *RLS1/rlsD* ortholog. BLASTP search (E-value < 1e-10) using genes near Agub_g12775 in *A. gubernaculifera* genome as queries was also conducted against gene models of *V. carteri* v2.1, *C. reinhardtii* v5.6, *G. pectorale* and *Y. unicocca* to confirm the synteny.

4.2.7. Identification and analysis of MMP and pherophorin genes

The identification of genes in MMP (matrix metalloprotease) family, which is one of expanded ECM

gene families in *V. carteri*, was conducted based on methods in a previous study (Hanschen et al., 2016). The MMP domains, coordinate with Pfam metalloprotease domain (PF05548, Peptidase M11), of all MMP genes listed in previous studies (Prochnik et al., 2010; Hanschen et al., 2016) with available protein model (44 genes in *C. reinhardtii*, 35 genes in *G. pectorale* and 98 genes in *V. carteri*) were aligned by MUSCLE (Edgar, 2004) and by manually. The hidden Markov model for MMP domains was built using the aligned sequences by “hmmbuild” in HMMER 3.3 (<http://hmmer.org/>) and used for protein search against gene models of *C. reinhardtii* v5.6, *T. socialis*, *G. pectorale*, *A. gubernaculifera*, *Y. unicocca*, *Eudorina* sp. *V. carteri* (v2.1, v2.0 and v1.0) (Table 4-1) by “hmmsearch” (E-value < 1e-5) in HMMER. For *V. carteri* v2.0 and v1.0 genome, gene models which are not included in v2.1 genome were counted. Some MMP domains were thought to be separated into two gene models in *Y. unicocca* and *Eudorina* sp. Such gene models were connected and treated as one gene.

The identification of pherophorin genes, which are also expanded ECM genes in *V. carteri*, were conducted by the essentially the same methods with those for MMP genes. However, typical pherophorin proteins have two pherophorin domains, A-type domain and B-type (pheromone-like) domain (Godl et al., 1997; Hallmann, 2003), which are not distinguished in Pfam domain model (PF12499, DUF3707). Therefore, pherophorin domains of all pherophorin genes listed in previous studies (Prochnik et al., 2010; Hanschen et al., 2016) with available protein model (35 genes in *C. reinhardtii*, 20 genes in *G. pectorale* and 75 genes in *V. carteri*, total 225 domains) were first classified into A-type and B-type based on the similarity of amino acid sequences and then used to build the protein models. In addition to A-type and B-type domain models, Pfam domain model (PF12499, DUF3707) were also used for protein search against genome data of volvocine species. As 11 pherophorin genes in *G. pectorale* listed in the previous study (Hanschen et al., 2016) were not found in gene models from NCBI or Phytozome, amino acid sequences of pherophorin domains of these genes were deduced manually based on TBLASTN search (E-value < 1e-10) using amino acid

sequences of previously reported perophorin genes.

4.2.8. Separation of somatic cells and reproductive cells

The overview of separation of somatic and reproductive cells is shown in Fig. 4-2a. Mature vegetative colonies before embryogenesis in synchronous culture (Fig. 4-1) were collected by centrifugation at 2,500 rpm for 3 min and homogenized to single cells by strong pipetting. For isolation of somatic cells, the homogenized culture were diluted by 12% Percoll (Sigma-Aldrich, St. Louis, MO, USA) in VTAC medium, put in 15 mL centrifuge tubes and centrifuged at 3,500 rpm for 10 min. The cells at the upper layer were collected, diluted by 10% Percoll in VTAC medium and centrifuged at 3,500 rpm for 5 min. The sedimented cells after second centrifugation were collected, washed by VTAC medium and used as somatic cell samples. For isolation of reproductive cells, the homogenized culture were diluted by 16% Percoll in VTAC medium, put in 15 mL centrifuge tubes and centrifuged at 3,500 rpm for 10 min. The sedimented cells were collected, washed by VTAC medium and used as reproductive cell samples. For both somatic and reproductive cells, three samples were prepared from different flasks and immediately used for RNA isolation. The percentage of somatic and reproductive cells of each sample was measured by light microscopy (Table 4-3).

4.2.9. Cell-type RNA-seq analysis

The somatic and reproductive cell samples were mixed with RLT buffer in RNeasy Plant Mini Kit (Qiagen, Venlo, Limburg, Netherlands) and ceramic beads in a conical-bottom 2-mL microcentrifuge tube and then vibrated in 25 Hz for 8 min using a Mixer Mill MM300 (Retsch, Haan, Germany). Total RNA in mixture were isolated using RNeasy Plant Mini Kit according to the manufacture's protocol. DNA contamination was removed using TURBO DNA-free Kit (Applied Biosystems). Illumina paired-end libraries (library sizes with approx. 500 bp) were constructed using NEBNext

Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) according to the manufacture's protocol. These libraries were sequenced using Illumina MiSeq sequencer (Illumina). 1.7–2.2 M paired-end reads (250 bp × 2) were generated from each sample (Table 4-3). These reads were processed by Trimmomatic 0.39 (Bolger et al., 2014) and PRINSEQ lite 0.20.4 (Schmieder and Edwards, 2011) for removing adaptor sequences, poly A/T tails, low quality bases and very short reads (less than 30 bases). The processed reads were aligned to *A. gubernaculifera* genome assembly described above using HISAT2 version 2.2.0 (Kim et al., 2019). Mapped reads were assigned to gene models described above using featureCounts (part of Subread 2.0.0) (Liao et al., 2014). More than 80% of high-quality reads were uniquely mapped to 13,198 gene models (Table 4-3) and 515 gene models with no expression in any samples were excluded from further analysis.

Raw read counts were then subjected to differential expression analysis using DESeq2 (Love et al., 2014). The genes with differential expression were detected by Wald test between somatic and reproductive cells with a false discovery rate (FDR) < 0.05. The differentially expressed genes, which were defined as “somatic genes” or “reproductive genes”, were further subclassified based on threshold-based Wald tests with \log_2 fold change threshold = 1 ($\log_2 2$) or 2.322 ($\log_2 5$) in DESeq2 options with a FDR < 0.05. Genes having greater than five-fold significant expression ratio were classified as “somatic-specific” or “reproductive-specific”, genes having greater than two-fold significant expression ratio (but less than five-fold) classified as “somatic-biased” or “reproductive-biased”, and genes having less than two-fold expression ratio were classified as “somatic-upregulated” or “reproductive-upregulated” (Figs. 4-2b, 4-3). The remaining genes without significant difference in expression (FDR > 0.05) were classified as “constitutive” with expression ratio < 2 or “low confidence” with expression ratio > 2 (Figs. 4-2b, 4-3). This classification corresponds to those in a previous study on *V. carteri* (Matt and Umen, 2018) except for “upregulated”. TPM (transcripts per million, Wagner et al., 2012) was also calculated to estimate the

abundance of transcripts. The expression ratios (without genes having no expression in one of the cell-types) derived from DESeq2 and calculated by TPM were highly correlated ($r^2 = 0.999$) (Fig. 4-4).

4.2.10. Comparison of expression pattern between somatic and reproductive cells

GO (gene ontology) enrichment analysis was conducted using Blast2GO (Conesa et al., 2005). GO terms were assigned with BLASTP search results (E-value < $1e^{-3}$) by DIAMOND v2.0.2.140 (Buchfink et al., 2015) using all expressed genes in *A. gubernaculifera* as queries against NCBI nonredundant database and InterProScan (Quevillon et al., 2005) output on Blast2GO. Enrichment of GO terms in “somatic”, “reproductive” (including specific, biased and upregulated) and “constitutive” genes was detected by two-sided Fisher’s exact test with an odds ratio > 1 and a FDR < 0.05 as criteria. GO terms in generic GO subset were selected.

The *V. carteri* genes involved in motility, photosynthesis, central carbon metabolism, nucleotide-sugar metabolism and glycosyltransferases, whose expressions in somatic cells and reproductive cells (gonidia) were reported in the previous study (Matt and Umen, 2018), and their orthologs in *C. reinhardtii* were used as queries in BLASTP search (E-value < $1e^{-10}$) against gene models of *A. gubernaculifera* described above. Some genes which were not included in the previous datasets (Matt and Umen, 2018) were added to queries based on previous studies (Allen et al., 2011; Barker et al., 2014; Zones et al., 2015). The expression in somatic and reproductive cells and classification of hit putative orthologs were listed in Tables 4-8, 4-9, 4-10, 4-11. Metabolic pathways were deduced based on previous studies (Johnson and Alric, 2013; Zones et al., 2015; Matt and Umen, 2018). The expression levels of genes involved in multicellular traits, orthologous genes involved in embryogenesis in *V. carteri* (InvA, InvB, InvC, GlsA and Hsp70A), VARL genes, MMP and pherophorin genes, were also listed in Table 4-12.

As LHCII and LHCI genes were similar to each other and difficult to distinguish solely by

BLASTP search, the LHCII and LHCI genes were identified based on phylogenetic analysis of LHC genes in *A. gubernaculifera*, *C. reinhardtii* and *V. carteri* (Fig. 4-5). The 271 amino acid positions from LHCII and LHCI genes in *V. carteri*, *C. reinhardtii* and *A. gubernaculifera* (Table 4-9) were subjected to maximum-likelihood analysis using LG + I + G + F model selected by ModelTest-NG (Darriba et al., 2020) with 1,000 replicates of bootstrap analyses (Felsenstein, 1985) by RAxML-NG (Kozlov et al., 2019). Putative orthologs of LHCB5 and LHCB7 were not found in gene models of *A. gubernaculifera* genome.

4.2.11. Identification and analysis of somatic- and reproductive-specific genes

The characteristics of proteins encoded in all somatic- and reproductive-specific genes described above were deduced based on BLASTP search against NCBI nonredundant database (E-value < 1e-10) by DIAMOND v2.0.2.140 (Buchfink et al., 2015) and the domain search against all Pfam-A models (El-Gebali et al., 2019) using “hmmscan” (E-value < 1e-5) in HMMER. The putative hydroxyproline-rich domains, which contain prolines at a high ratio, were difficult to detect by these searches and found by checking sequences manually.

Among somatic-specific genes, putative MYB transcription factors, Agub_g2945 and Agub_g5284, and a putative RWP-RK transcription factor, Agub_g8265, were subjected to further analysis. The amino acid sequences of these genes were used as queries for BLASTP search (E-value < 1e-10) against genome data of 15 algal/plant species (Table 4-1). Based on the BLASTP results and phylogeny of MYB transcription factors and RWP-RK transcription factors in previous studies (Chardin et al., 2014; Du et al., 2015a; Bowman et al., 2017), 75, 43 and 127 closely related genes in these species were selected for the phylogenetic analyses of Agub_g2945, Agub_g5284 and Agub_g8265, respectively. The DNA binding domain of these genes were aligned by MUSCLE (Edgar, 2004). The alignments of 105, 87, 49 amino acid positions, respectively, were subjected to maximum-likelihood analysis. For evolutionary model for maximum-likelihood analysis, LG + I + G

model or LG + I + G +F model were selected for the datasets of MYB transcription factors including Agub_g2945 or Agub_g5284, respectively, and LG + G model was selected for the dataset of RWP-RK transcription factors by ModelTest-NG (Darriba et al., 2020). The maximum-likelihood analysis was conducted with 1,000 replicates of bootstrap analyses (Felsenstein, 1985) using RAxML-NG (Kozlov et al., 2019).

4.2.12. Availability of data

The new genome assembly with annotated gene models of *A. gubernaculifera* strain 2014-1002-YkAs8 has been deposited to DDBJ/EMBL/GenBank (BioProject accession number: PRJDB10253; BioSample accession number: SAMD00236769; WGS accession numbers: BMAR01000001–BMAR01000206). The raw read sequence data from cell-type RNA-seq analysis also has been deposited to DDBJ/EMBL/GenBank (BioProject accession number: PRJDB10862; BioSample accession numbers: SAMD00260684–SAMD00260689; DRA accession number: DRA011195) All the sequence alignments used for phylogenetic analyses have been deposited in TreeBASE (<https://treebase.org/treebase-web/home>; study ID: 27376).

4.3. Results and Discussion

4.3.1. *De novo* genome sequence of *Astrephomene gubernaculifera*

The whole nuclear genome of *A. gubernaculifera* strain NIES-4017 was *de novo* sequenced. The genome assembly with 103,829,650 bp from 206 contigs and N50 of 1,506,991 bp was obtained. Moreover, gene prediction based on RNA-seq data identified 13,713 protein-coding gene models on genome assembly of *A. gubernaculifera*. The genome size and number of genes of *A. gubernaculifera* were slightly smaller than those of other volvocine algae (Table 4-4). GC content of *A. gubernaculifera* was 59.9% and intermediate between *Chlamydomonas reinhardtii* (64.1%) and *V. carteri* (56.1%). Overall, the characteristics of nuclear genome in *A. gubernaculifera* are not so much different from those in other volvocine algae.

4.3.2. Orthologous genes involved in embryogenesis in *Volvox carteri*

In early embryogenesis in *V. carteri*, a part of cells of the embryo divide asymmetrically, which results in the formation of large cells (gonidial initials) (Green and Kirk, 1981). As the cell sheet of an embryo after successive cell divisions is inside-out with respect to the adult configuration, the embryo then performs inversion to form the adult configuration (Viamontes and Kirk, 1977; Kirk and Nishii, 2001). Three genes involved in inversion have been identified so far: *InvA* encodes a kinesin and is necessary for movement of cells during inversion (Nishii et al., 2003), while *InvB* and *InvC* encode GDP-mannose transporter and glycosyltransferase, respectively, and are involved in the enlargement of gonidial vesicle (cell wall) and providing sufficient space for inversion (Ueki and Nishii, 2008, 2009). The asymmetrical division has been also revealed to be regulated by two interacting chaperones, GlcA and Hsp70A (Miller and Kirk, 1999; Cheng et al., 2005). On the other hand, the embryo in *Astrephomene* forms the proper configuration directly by the outward rotation of daughter protoplasts during successive cell divisions and does not perform inversion (Chapter 2).

Asymmetric cell division also does not occur to form somatic and reproductive cells in *Astrephomene* (Chapter 2).

Orthologs of *V. carteri* genes which are involved in inversion (*InvA*, *InvB* and *InvC*) and asymmetric cell divisions (*GlsA* and *Hsp70A*) were searched in the present genome data of *A. gubernaculifera* as well as other volvocine algae (Table 4-1). *A. gubernaculifera* had an ortholog of each of the five genes. Orthologs of these genes were also found in all other species, except for *InvC* ortholog in *Gonium pectorale* and *GlsA* ortholog in *Tetrabaena socialis*. Orthologs of each gene were highly conserved in amino acid sequences (Fig. 4-6) and exon–intron structure (Fig. 4-7). As *Astrephomene* embryo does not undergo inversion or asymmetric cell divisions, these orthologous genes in *Astrephomene* might have different functions in cellular level from *Volvox*.

4.3.3. *RLS1/rlsD* homolog and other VARL genes

The differentiation of somatic cells in *V. carteri* is controlled by *regA*, which encodes a putative transcription factor with a DNA-binding SAND-like domain (called VARL domain), and is expressed only in somatic cells (Kirk et al., 1999). *regA* is arranged tandemly with closely related genes with VARL domains (*reg* cluster) in *Volvox* genome (Duncan et al., 2007). This *reg* cluster is derived from tandem duplication of another VARL domain-containing gene (*rlsD*) at the common ancestor of Volvocaceae (Hanschen et al., 2014; Grochau-Wright et al., 2017). *C. reinhardtii* has the most related gene *RLS1* and it is hypothesized that *RLS1/rlsD* ortholog in the volvocine ancestor might have been co-opted to differentiation of somatic cells in the extant species of *Volvox* (Nedelcu and Michod, 2006; Nedelcu, 2009; Hanschen et al., 2014; Grochau-Wright et al., 2017). Some genes with VARL domain (VARL genes) other than *reg* cluster and *RLS1/rlsD* ortholog have been identified in *C. reinhardtii*, *T. socialis*, *G. pectorale*, *Yamagishiella unicocca* and *V. carteri* (Duncan et al., 2007; Hanschen et al., 2016; Grochau-Wright et al., 2017; Featherston et al., 2018).

The homologs of *regA* and other genes with VARL domain were searched in genome of *A. gubernaculifera*. There were six VARL genes in *A. gubernaculifera*. The phylogenetic analysis of VARL genes within volvocine algae assigned Agub_g12775 to *RLS1/rlsD* clade, Agub_g12169 to *RLS5, 6, 9/rlsH* clade, Agub_g8757 to *RLS10/rlsL* clade, Agub_g12954 to *RLS2/rlsI* clade, Agub_g6775 and Agub_g12876 to *rlsF* clade, respectively (Fig. 4-8). Moreover, gene synteny including Agub_g12775 was similar to those including *RLS1/rlsD* genes in other volvocine algae (Fig. 4-9). There was no gene in *A. gubernaculifera* included in *reg* cluster genes (Fig. 4-8) or no tandem duplication of *RLS1/rlsD* ortholog on *A. gubernaculifera* genome (Fig. 4-9) in contrast to species in Volvocaceae (Hanschen et al., 2014; Grochau-Wright et al., 2017). These results suggest that *Astrephomene* might have evolved another molecular basis to differentiate somatic cells.

4.3.4. ECM genes

The components of expanded ECM in *V. carteri* have been identified so far and most of them are revealed to be glycoproteins with hydroxyproline-rich domains (hydroxyproline-rich glycoproteins, HRGPs) (reviewed in Hallmann, 2003). Among them, numbers of MMPs and pherophorins are remarkably increased in *V. carteri* genome compared with those in *C. reinhardtii*, *T. socialis* or *G. pectorale* genome (Prochnik et al., 2010; Hanschen et al., 2016; Featherston et al., 2018). *Astrephomene* also has evolved expanded ECM independently of *Volvox* and Volvocaceae (Herron and Michod, 2008; Herron et al., 2009), though the expanded regions of ECM are different in structure from Volvocaceae (Hoops and Floyd, 1982; Nozaki, 1990).

The number of genes encoding ECM proteins, MMPs and pherophorins, were investigated among genome data of *A. gubernaculifera* as well as other volvocine algae (Fig. 4-10, Tables 4-5, 4-6). In addition to most of genes previously reported in *C. reinhardtii*, *V. carteri*, *G. pectorale* and *T. socialis* (Prochnik et al., 2010; Hanschen et al., 2016; Featherston et al., 2018; Heyde and Hallmann, 2020), I newly found many genes by increasing the number of protein models (see Materials and

Methods). The number of MMP and pherophorin genes in *C. reinhardtii* were 50 and 75, while those in *V. carteri* were 103 and 157, respectively. The ratio of number of genes in *C. reinhardtii* to that in *V. carteri* was approximately 1:2 in both MMPs and pherophorins, as in previous studies (Prochnik et al., 2010; Hanschen et al., 2016). The numbers of MMP and pherophorin genes in *T. socialis* were 32 and 71, and those in *G. pectorale* were 46 and 55, which were similar to, or even slightly smaller than, those in *C. reinhardtii* as in previous studies (Hanschen et al., 2016; Featherston et al., 2018).

In this study, I newly identified MMP and pherophorin genes in *A. gubernaculifera* as well as *Y. unicocca* and *Eudorina* sp. *A. gubernaculifera* had only 20 MMP genes and 54 pherophorin genes, which are both the smallest numbers among volvocine algae (Fig. 4-10). On the other hand, the numbers of both MMP and pherophorin genes in *Y. unicocca* were 61 and 101, and those in *Eudorina* sp. were 173 and 184, which indicate the expansion of the two gene families in the whole Volvocaceae. These results suggest that the expansion of ECM in *Astrephomene* was not accompanied by the expansion of MMPs or pherophorins in contrast to Volvocaceae. Some of pherophorin genes are involved in the signal amplification of sex inducer (Sumper et al., 1993; Godl et al., 1995). As the number of these genes in anisogamous *Eudorina* sp. and oogamous *V. carteri* were greater than those in isogamous *Y. unicocca*, the expansion of these ECM gene families might be involved in also the evolution of sexual reproduction, the transition from isogamy to anisogamy and oogamy (Kirk, 2005).

4.3.5. Identification of differentially expressed genes based on cell-type RNA-seq analysis

In *V. carteri*, somatic cells and gonidia show different patterns of gene expression. Many of genes related to photosynthesis and anabolism are upregulated in gonidia (Tam and Kirk, 1991; Choi et al., 1996; Meissner et al., 1999; Klein et al., 2017; Matt and Umen, 2018), while genes related to motility and ECM synthesis are upregulated in somatic cells (Klein et al., 2017; Matt and Umen, 2018). *Astrephomene* is also thought to have germ–soma differentiation in gene expression level,

though the characteristics of the somatic cells in *Astrephomene* are little known except for difference in cellular size. To identify the gene expression patterns in somatic and reproductive cells in *A. gubernaculifera*, I conducted the cell-type RNA-seq analysis.

By homogenization and density centrifugation of mature colonies in synchronous culture of *A. gubernaculifera*, I obtained “somatic cell samples” and “reproductive cell samples”, which contained > 70% somatic cells and > 95% reproductive cells, respectively, in total cell number (Table 4-3, Fig. 4-2a). The cell-type RNA-seq was conducted using RNA from three biological replicates of somatic and reproductive cell samples. 1.6–2.1 million paired-end reads were generated from each RNA-seq library and 81–82% reads were uniquely mapped to gene models on genome assembly of *A. gubernaculifera* (Table 4-3).

Based on differential expression analysis, I identified 2,529 “somatic genes” and 2,328 “reproductive genes” among 13,198 expressed genes, which showed the significant differential expression between somatic and reproductive cell samples (Fig. 4-2b, 4-3). Then the somatic and reproductive genes were classified as follows; genes with greater than five-fold expression ratio were classified as “somatic-specific” or “reproductive-specific”, genes with between two-fold and five-fold expression ratio were classified as “somatic-biased” or “reproductive-biased” and genes with less than two-fold expression ratio but with significant different expression were classified as “somatic-upregulated” or “reproductive-upregulated”, respectively (Fig. 4-2b, 4-3). The criteria for “specific” and “biased” correspond to those in the previous RNA-seq study on *V. carteri* (Matt and Umen, 2018), which showed that nearly half of the expressed genes in *V. carteri* have greater than two-fold expression ratio (“specific” or “biased”) and 40% genes have greater than five-fold expression ratio (“specific”). Another RNA-seq study also showed that nearly 55% genes have greater than two-fold expression ratio in *V. carteri* (Klein et al., 2017). By contrast, only 893 genes (7%) showed greater than two-fold expression ratio (“specific” or “biased”) and 148 genes (1%) showed greater than five-fold expression ratio (“specific”) in *A. gubernaculifera*. This situation may

have resulted from the difference in isolation of somatic and reproductive cells between *V. carteri* and *A. gubernaculifera*; though the isolation of somatic cells and gonidia were able to be separated completely (Klein et al., 2017; Matt and Umen, 2018), somatic and reproductive cells were difficult to separate and both cell samples contained another type cells to some extent in *A. gubernaculifera*.

4.3.6. Germ–soma differentiation in gene expression patterns

To identify the overview of germ–soma differentiation, GO (gene ontology) enrichment analysis was conducted for somatic, reproductive and constitutive genes of *A. gubernaculifera* (Table 4-7) and the results were compared with those of the previous GO enrichment analysis in *V. carteri*. (Matt and Umen, 2018). The enriched GO terms in *A. gubernaculifera* somatic cells included “signal transduction”, “cytoskeleton-dependent intracellular transport” and “cilium”, which were also common with somatic enriched terms in *V. carteri* (Matt and Umen, 2018). On the other hand, the enriched GO terms in *A. gubernaculifera* reproductive cells included “generation of precursor metabolites and energy” and “photosynthesis”, which were also common with gonidial enriched terms in *V. carteri* (Matt and Umen, 2018). This result suggests that the germ–soma differentiation in *Astrephomene* and *Volvox* have similar features in gene expression. Moreover, *A. gubernaculifera* reproductive enriched terms included “translation” and “ribosome”, which implies that the whole gene expression is higher in reproductive cells than somatic cells. I further investigated the expression patterns of *A. gubernaculifera* genes involved in respective biological functions.

Among the genes involved in motility (Table 4-8, Fig. 4-11), many of the genes encoding flagella components belonged to somatic-biased and somatic-upregulated genes. Especially, most of intraflagellar transport (IFT) genes, Bardet–Biedel Syndrome (BBS) genes were somatic genes. Though not only somatic cells but reproductive cells in *Astrephomene* have flagella unlike *Volvox* (Fig. 4-16) (Pocock, 1954), this result implies that the somatic cells in *Astrephomene* are also more specialized for swimming than reproductive cells. Moreover, the movement of flagella in somatic

cells might be specialized, as the direction of flagella is different between somatic and reproductive cells (Pocock, 1954; Nozaki, 1983). On the other hand, genes involved in transition zone and basal bodies included some reproductive-upregulated genes as well as somatic genes.

Most obvious difference in gene expression in *A. gubernaculifera* between somatic and reproductive cells was observed in a part of photosynthetic genes (Table 4-9, Fig. 4-12). Almost all of the core subunits of photosynthetic complexes on the thylakoid membranes belonged to reproductive genes (Fig. 4-12b). In contrast, gene expression of assembly factors for photosynthetic complexes and enzymes for chlorophyll biosynthesis varied among respective genes (Fig. 4-12a). This is in contrast with *V. carteri*, in which many of genes encoding assembly factors and enzymes for chlorophyll biosynthesis belonged to gonidia-specific or gonidial-biased, while few of core photosynthetic genes were gonidial-specific or gonidial-biased (Matt and Umen, 2018). Thus, both *Astrephomene* and *Volvox* are thought to have downregulation of photosynthesis in somatic cells, though the targets of downregulation might be different between the two species (Fig. 4-16).

In carbon metabolism genes, *A. gubernaculifera* showed both similarities and differences with *V. carteri* (Table 4-10 Fig. 4-13). Almost all genes in Calvin–Benson–Bassham (CBB) cycle were reproductive-biased or reproductive-upregulated. Moreover, many of fatty acid biosynthesis genes and a half of amino acid biosynthesis genes were reproductive genes. These anabolic pathways were also upregulated in gonidia of *V. carteri* (Matt and Umen, 2018). In addition, genes involved in acetate metabolism were reproductive-upregulated in *A. gubernaculifera*. In heterotrophic algae, acetate is mainly metabolized in TCA cycle and glyoxylate cycle (Perez-Garcia et al., 2011). In *A. gubernaculifera*, genes involved in both metabolic pathways, including enzymes for acetyl-CoA synthesis from acetate (ACK2, PAT1 and ACS1 for TCA cycle, ACS2 and ACS3 for glyoxylate cycle) (Yang et al., 2014; Zones et al., 2015) and downstream TCA and glyoxylate cycles were mainly reproductive-upregulated and highly expressed (with high TPM). This is in contrast with *V. carteri*, in which glyoxylate cycle genes, along with upstream fatty acid degradation, were

upregulated in somatic cells (Fig. 4-16) (Matt and Umen, 2018). The different expression patterns in glyoxylate cycle might reflect the different trophic states; unlike other volvocine algae, *Astrephomene* has strong heterotrophy and requires acetate or some carboxylic acid as carbon sources for its growth (Brooks, 1972). Thus, the growth in somatic cells of *A. gubernaculifera* is thought to be downregulated via both phototrophic and heterotrophic pathways (Fig. 4-16).

ECM-related genes were mainly upregulated in somatic cells (Tables 4-11, 4-12, Fig. 4-13). The majority of genes involved in biosynthesis of nucleotide sugars and glycosyltransferase genes, which are used in glycosylation of ECM glycoprotein in volvocine algae (Hallmann, 2003), belonged to somatic genes. The half of MMP genes also belonged to somatic genes, while the expression of phosphorin genes varied. In *V. carteri*, the most part of ECM is produced by somatic cells as the sugar sink (Matt and Umen, 2018). As the somatic cells of *Astrephomene* do not seem to contribute to production of whole ECM of a colony, this upregulation of ECM genes in somatic cells might reflect the larger proportion of ECM volume to cell volume (see Chapter 2, Fig. 2-2c).

Moreover, expression of the other genes related in multicellular traits was also examined (Table 4-12). Among orthologous genes involved in embryogenesis in *V. carteri*, Agub_g14864 (*InvB* ortholog) was reproductive-upregulated and Agub_g8128 (*GlsA* ortholog) was reproductive-biased. As the reproductive cells are going to undergo embryogenesis while somatic cells are not, these genes might be involved in embryogenesis also in *Astrephomene*. In terms of VARL genes, on the other hand, Agub_g12775 (*rlsD/RLS1* ortholog) and Agub_g8757 (*RLS10/rlsL* related) were somatic-upregulated, Agub_g12876 (*rlsF* related) was reproductive-upregulated and Agub_g12169 (*RLS5, 6, 9/rlsH* related) was somatic-biased. In *V. carteri*, some of VARL genes were upregulated in somatic cells and *rlsH* expression was somatic-specific (Matt and Umen, 2018). Thus, the *rlsH* homologs in *A. gubernaculifera* and *V. carteri* might have similar function, which is currently unknown. However, there is no VARL genes which had high somatic/gonidial expression

ratio like *regA* in *V. carteri*, which was reported to be 150-fold (Nematollahi et al., 2006) or 1,000-fold (Matt and Umen, 2018).

4.3.7. Candidate genes for regulators of germ–soma differentiation

To find candidates for a master regulator of germ–soma differentiation, I searched the transcripts of 147 somatic- and one reproductive-specific genes (Table 4-13) and found three genes encoding putative transcription factors: Agub_g2945, Agub_g8265 and Agub_g5284 (Figs. 4-14, 4-15). Agub_g2945 had 2 MYB repeats in a DNA binding domain (Fig. 4-14a) and belonged to R2R3-MYB transcription factors, which are expanded in green plants and especially in land plants (Du et al., 2015a). The DNA binding domain of Agub_g2945 was similar to that in *Arabidopsis thaliana* MYB64, MYB98, MYB119, *Marchantia polymorpha* FGMYB (involved in female gametophyte development) (Kasahara et al., 2005; Rabiger and Drews, 2013; Hisanaga et al., 2019), *A. thaliana* MYB115, MYB118 (embryonic development) (Wang et al., 2009), FOUR LIPS (FLP, MYB124) and MYB88 (stomata development) (Yang, 2016), though it did not belong to the land plant clades including them (Fig. 4-14b). Agub_g5284 had 1 MYB repeat and was thought to be 1R-MYB transcription factor (Fig. 4-14a) and belonged to a group expanded within volvocine green algae (Fig. 4-14c). On the other hand, Agub_g8265 encoded an RWP-RK transcription factor (Fig. 4-15a) which was included in RKD(B) subfamily expanded in volvocine algae (Fig. 4-15b) (Chardin et al., 2014) and was separated from MID (sex determination) (Ferris and Goodenough, 1997) and NIT2 (nitrate signaling) (Camargo et al., 2007). Though the functions of these genes and their closely related homologs in other volvocine algae are unknown so far, these genes would be worth investigating further as candidates for a master regulator of germ–soma differentiation in *A. gubernaculifera*.

In addition, many of somatic-specific genes had proline-rich regions which are similar to those of hydroxyproline-rich glycoproteins of ECM (Table 4-13). These genes include fasciclin-like

domain containing protein or cysteine proteases as well as MMPs and pherophorins. Along with ECM genes discussed above, the differential expression of these genes between somatic and reproductive cells might indicate the difference in ECM structure or component, which have not been investigated so far.

4.4. Tables and Figures

Table 4-1: List of genome data used in this study.

Species, version	Reference	Source
<i>Astrephomene gubernaculifera</i>	this study	NCBI (BioProject accession number: PRJDB10253)
<i>Chlamydomonas reinhardtii</i> v5.6	(Merchant et al., 2007)	Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Creinhardtii)
<i>Volvox carteri</i> v1.0, v2.0, v2.1	(Prochnik et al., 2010)	Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Vcarteri)
<i>Gonium pectorale</i>	(Hanschen et al., 2016)	NCBI (https://www.ncbi.nlm.nih.gov/genome/16856?genome_assembly_id=266986)
<i>Tetrabaena socialis</i>	(Featherston et al., 2018)	NCBI (https://www.ncbi.nlm.nih.gov/genome/66769?genome_assembly_id=360948)
<i>Yamagishiella unicocca</i>	(Hamaji et al., 2018)	NCBI (https://www.ncbi.nlm.nih.gov/genome/53185?genome_assembly_id=374518)
<i>Arabidopsis thaliana</i> TAIR10	(Lamesch et al., 2012)	Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Athaliana)
<i>Amborella trichopoda</i>	(Albert et al., 2013)	Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Atrichopoda)
<i>Picea abies</i>	(Nystedt et al., 2013)	ConGenIE (ftp://plantgenie.org/Data/ConGenIE/Picea_abies/v1.0)
<i>Selaginella moellendorffii</i>	(Banks et al., 2011)	Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Smoellendorffii)
<i>Physcomitrella patens</i> v3.0	(Rensing et al., 2008)	Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Ppatens)

Table 4-1: Continued.

<i>Marchantia polymorpha</i> v3.1	(Bowman et al., 2017)	Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mpolymorpha)
<i>Klebsormidium nitens</i>	(Hori et al., 2014)	NCBI (https://www.ncbi.nlm.nih.gov/genome/54879?genome_assembly_id=319335)
<i>Ostreococcus tauri</i>	(Derelle et al., 2006)	https://bioinformatics.psb.ugent.be/gdb/ostreococcus/
<i>Cyanidioscyzon merolae</i>	(Matsuzaki et al., 2004)	http://czon.jp/

Table 4-2: List of specific primers used for amplification and sequencing of *InvA* orthologs.

Designation	Positions ^a	Sequence (5'–3')
<i>Gonium pectorale</i>		
GpInvA_F1	(5' UTR)	CCTCTGACTTCTTTTCTTGCCGATC
GpInvA_F2	396–418	ACGTGTTATTCGCCACATCTTCG
GpInvA_F3	843–865	CGAGTTCACCAGCAAGCTCCACT
GpInvA_F4	1412–1435	AGCAGATGAAAGCAAAGGGATTCC
GpInvA_F5	1873–1896	GGAGATGTCGGCACTGAGGACTAC
GpInvA_F6	2410–2432	GACAAGATGACGTACCTGCTCGC
GpInvA_R1	692–668 ^b	GTAAAGAAGGACAGAAGCTCCTCCG
GpInvA_R2	1182–1158 ^b	GTTGGCATACTTCAGGGTGTGTGTTG
GpInvA_R3	1589–1568 ^b	CTTTCCACAGAGTTCGCCACCT
GpInvA_R4	2352–2331 ^b	CTTTGTGCAGGCCTTGAAAACG
GpInvA_R5	2756–2733 ^b	CAAGTGATTGTGTTGGCCTTTTCG
GpInvA_R6	(3' UTR) ^b	AAGTTCGAGCCAGCCATGAAAAG
<i>Eudorina</i> sp.		
EuInvA_F1	(5' UTR)	TTAAGCTGATCCCTTTGCCTGTTG
EuInvA_F2	390–412	ACGGGTTATTCGGCACATCTTTG
EuInvA_F3	917–939	TCAAGGAGGCCACGCACATTAAC
EuInvA_F4	1359–1381	CAAGGAGAAGGAGGCCAAACGAG
EuInvA_F5	1975–1999	GGCGACTCAATGTTTACGGAAGAAG
EuInvA_F6	2582–2604	CCTCGTGCATCTTCGACTTTGAC
EuInvA_R1	501–478 ^b	CAAGTCCTCGTTGTACAGCTCCAG
EuInvA_R2	1174–1150 ^b	CGTACTTCAGGGTGTGTTCGTCTC
EuInvA_R3	1686–1664 ^b	GGTGTTTCGGCAGGACTTTCGTAC
EuInvA_R4	2222–2200 ^b	CTATCGACCGGCTCCACTGTTTC
EuInvA_R5	2646–2622 ^b	AATTCGCTCGTCCACACCGTAGTAC
EuInvA_R6	(3' UTR) ^b	TGATACCAAGCGCACACTTCAAG

Table 4-2: Continued.

degenerated primers		
InvA-F11	1141–1163	GAYTTYGARGARACIAAYAAYAC
InvA-F12	1183–1205	GCITGYAARATHAARAAYCARCC
InvA-F13	1405–1427	GARGARCARATGAARGCIAARGG
InvA-F0	2311–2330	GCIGTITTYAARGCITGYAC
InvA-F1	2332–2351	AARGARGAYTTYAARGAYGT
InvA-F2	2521–2540	ATGGCIACITTYGGIACIAA
InvA-R1	2801–2782 ^b	CCRTTYTCDATIGCRTGRTT
InvA-R2	2867–2848 ^b	TTIACICCI GCYTGCATCCA
InvA-R3	2972–2953 ^b	ACYTCIGCIACCATRAACCA
<i>Astrephomene gubernaculifera</i>		
InvA Primer SR1	2383–2359 ^b	GTAGGTAGGTCATCTTGTCTGAGGC
InvA Primer SR2	2556–2532 ^b	CTCCTCTGGGTCCTCAAAGTCGAAG
InvA Primer SR3	2587–2563 ^b	CAATCCGCTCGTCCACACCGTAATA
AgInvA_R1	1522–1498 ^b	CAATGACCTGATCCGTCGGGATGCT
AgInvA_R2	1554–1532 ^b	CCGCGTACCGAACCCCTTCAACA
AgInvA_R3	709–685 ^b	TGGATCGGTTGTCCATGCACAGGTT
AgInvA_R4	782–758 ^b	ACGGTGCGCTGGATGGTGATAGTGA
AgInvA_F1	2564–2588	ATTACGGTGTGGACGAGCGGATTGA
AgInvA_F2	2675–2698	AGGCGTCGAAGGACAACACGATCA
AgInvA_F5	(3' UTR)	GGGTACACTCTGTTTCTGCACGCACA
AgInvA_F6	(3' UTR)	CCACAAGGTCTCCGCTAGTATGATAG
<i>Tetrabaena socialis</i>		
TsInvA_F1	95–119	TGAACTTTGATGACGAAACCAGACA
TsInvA_R1	405–384 ^b	CGCGAAAATGTGCTGGATGACC
TsInvA_F2	259–282	GAGGCGGTTTTCAAGGGTTACAAC
TsInvA_R2	2462–2439 ^b	AGCAGGTAGGTCATCTTGTCGGAG
TsInvA_R3	164–141 ^b	CGTAGCTGGAGAAGGGTGTTCCTTG

Table 4-2: Continued.

TsInvA_R4	287–263 ^b	GTTGCGTTGTAAACCCTTGAAAACCG
TsInvA_F3	2764–2786	TCCAAGGACAGCACCATCACCTG
TsInvA_F4	2806–2829	AACATGTCCATCGGCCTGTTGTAC
TsInvA_R5	2599–2578 ^b	CGATCATGTTGGTGCCGAACGT
TsInvA_F5	(3' UTR)	CCGGGTCTGCAATGGTTCGTGG
TsInvA_F6	(5' UTR)	GCCTGGGAAAGAAACTGCATCATC
TsInvA_R6	(3' UTR) ^b	ACGAACCATTGCAGACCCGGTG
TsInvA_F7	909–930	CAAGGAGGCGACCCACATCAAC
TsInvA_F8	1482–1502	CTCCATCCCCAGCGACCAGGT
TsInvA_R7	1838–1817 ^b	CCCACAAAGAAGTCGTCCGTCC

^a Coordinate number from CDS of cDNA sequence determined in this study.

^b Reverse primer.

Table 4-3: Characteristics of the samples for cell-type RNA-seq in *Astrephomene gubernaculifera* and their reads mapping statistics.

Sample	somatic1	somatic2	somatic3	reproductive1	reproductive2	reproductive3
Percentage of somatic cells	74.0%	70.7%	70.7%	3.9%	3.5%	2.9%
Percentage of reproductive cells	25.5%	28.7%	28.9%	96.1%	96.5%	97.1%
Number of total reads	1,800,628	1,932,566	2,081,114	2,179,646	1,663,646	1,765,606
Number of high-quality reads	1,766,185 (98.09%)	1,892,324 (97.92%)	2,032,021 (97.64%)	2,135,021 (97.95%)	1,629,527 (97.95%)	1,725,977 (97.76%)
Number of reads uniquely mapped to genome	1,552,330 (86.21%)	1,642,394 (84.99%)	1,768,757 (84.99%)	1,891,595 (86.78%)	1,435,092 (86.26%)	1,522,967 (86.26%)
Number of reads uniquely mapped to gene models	1,491,965 (82.86%)	1,579,356 (81.72%)	1,700,723 (81.72%)	1,801,205 (82.64%)	1,364,596 (82.02%)	1,454,112 (82.36%)

Table 4-4: Summary statics for genomes of volvocine green algae used in this study.

Species ^a	<i>Chlamydomonas reinhardtii</i> v5.6	<i>Tetrabaena socialis</i>	<i>Gonium pectorale</i>	<i>Astrephomene gubernaculifera</i>	<i>Yamagishiella unicocca</i> (plus)	<i>Eudorina</i> sp. (female)	<i>Volvox carteri</i> v2.1
Genome size (Mb)	111.1	135.8	148.8	103.8	134.2	184.0	131.2
Scaffold N50 (Mb)	7.78	0.14	1.27	1.50	0.66	0.56	2.6
Number of contigs/scaffolds	54	5,858	2,373	206	1,461	3,180	434
% G and C	64.1	66.0	64.5	59.9	61.0	61.0	56.1
Protein coding loci	17,741	15,513	17,984	13,713	21,231 ^b	28,661 ^b	14,247
References	(Merchant et al., 2007)	(Featherston et al., 2018)	(Hanschen et al., 2016)	this study	(Hamaji et al., 2018)	(Hamaji et al., 2018)	(Prochnik et al., 2010)

^a The sources of genome data are listed in Table 4-1.

^b Predicted by AUGUSTUS in this study.

Table 4-5: List of MMP genes searched from genome data of volvocine algae.

Species ^a	Protein Model ^b	Scaffold ^c	Start Location ^c	End Location ^c	Strand ^c	No. of MMP Domain	Evalue	Common Name ^d
Ag	Agub_g1239.1.p1	Scf001	6252002	6253063	-	1	4.00E-57	
Ag	Agub_g2744.1.p1	Scf002	3390675	3396457	-	1	3.30E-52	
Ag	Agub_g2953.1.p1	Scf002	4409425	4417200	-	1	2.80E-78	
Ag	Agub_g2954.1.p1	Scf002	4417913	4425183	-	1	2.40E-90	
Ag	Agub_g6282.1.p1	Scf009	126662	145795	-	1	9.90E-89	
Ag	Agub_g6672.1.p1	Scf009	2196046	2200703	-	1	2.60E-85	
Ag	Agub_g8972.1.p1	Scf017	1180977	1186195	+	1	1.10E-74	
Ag	Agub_g10447.1.p1	Scf024	544839	550628	+	1	9.50E-83	
Ag	Agub_g10832.1.p1	Scf026	906132	914360	+	1	5.70E-78	
Ag	Agub_g10833.1.p1	Scf026	915138	920646	+	1	3.90E-68	
Ag	Agub_g10834.1.p1	Scf026	925572	932244	+	1	1.50E-74	
Ag	Agub_g11444.1.p1	Scf030	70795	76832	-	1	2.60E-64	
Ag	Agub_g11544.1.p1	Scf030	743501	753310	-	1	1.80E-65	
Ag	Agub_g12060.1.p1	Scf033	885844	891675	-	1	5.70E-78	
Ag	Agub_g12871.1.p1	Scf038	729966	737953	-	1	8.20E-74	
Ag	Agub_g13771.1.p1	Scf049	123679	124280	+	1	1.10E-23	
Ag	Agub_g14056.1.p1	Scf052	431928	438008	-	1	2.60E-46	
Ag	Agub_g14212.1.p1	Scf054	183899	189624	+	1	1.30E-77	
Ag	Agub_g14455.1.p1	Scf056	34904	39409	+	1	1.00E-79	
Ag	Agub_g14986.1.p1	Scf064	54642	60261	+	1	5.40E-74	
Cr	Cre01.g006900.t1.1	chromosome_1	1369582	1389407	-	1	1.40E-50	
Cr	Cre01.g025200.t1.2	chromosome_1	3877297	3882212	-	1	7.30E-99	MMP4
Cr	Cre02.g078950.t1.1	chromosome_2	838688	845320	+	1	5.50E-69	
Cr	Cre02.g142366.t1.1	chromosome_2	8752133	8759548	-	1	6.00E-92	
Cr	Cre03.g144564.t1.1	chromosome_3	290354	295831	-	1	2.70E-92	MMP13
Cr	Cre03.g170850.t1.1	chromosome_3	3852659	3856163	-	1	1.50E-55	
Cr	Cre03.g201950.t1.2	chromosome_3	7968186	7972617	+	1	5.00E-101	
Cr	Cre03.g201550.t1.2	chromosome_3	8008220	8014495	-	1	4.20E-92	
Cr	Cre03.g201552.t1.1	chromosome_3	8528853	8535028	-	1	1.90E-87	MMP8
Cr	Cre05.g232600.t1.1	chromosome_5	437193	445137	+	1	5.50E-49	
Cr	Cre06.g278152.t1.1	chromosome_6	3736492	3743409	+	1	9.60E-90	
Cr	Cre06.g278153.t1.1	chromosome_6	3743512	3749073	+	1	1.40E-72	
Cr	Cre06.g278161.t1.1	chromosome_6	3777576	3783461	-	1	3.90E-90	
Cr	Cre06.g278164.t1.1	chromosome_6	3795586	3799951	-	1	7.10E-85	
Cr	Cre06.g278165.t1.1	chromosome_6	3800006	3806821	-	1	5.00E-85	
Cr	Cre06.g292050.t1.2	chromosome_6	6351938	6359316	-	1	1.40E-77	
Cr	Cre06.g302950.t1.1	chromosome_6	7842367	7848741	+	1	3.60E-76	
Cr	Cre07.g324500.t1.2	chromosome_7	1562754	1568281	+	1	2.30E-80	
Cr	Cre07.g333400.t1.1	chromosome_7	3015216	3024274	+	1	7.20E-84	
Cr	Cre07.g353600.t1.2	chromosome_7	5851540	5856919	-	1	2.00E-84	MMP7
Cr	Cre07.g353750.t1.2	chromosome_7	5873700	5877081	+	1	4.60E-53	MMP10
Cr	Cre08.g376950.t1.1	chromosome_8	3717857	3723352	-	1	1.10E-79	MMP14
Cr	Cre09.g393700.t1.1	chromosome_9	2105436	2110682	+	1	1.70E-87	MMP3
Cr	Cre09.g391150.t1.1	chromosome_9	2478042	2482696	+	1	3.40E-85	
Cr	Cre09.g388350.t1.1	chromosome_9	2888111	2893452	+	1	7.00E-89	MMP11
Cr	Cre09.g396772.t1.1	chromosome_9	4667357	4674289	-	1	1.40E-89	
Cr	Cre09.g407050.t1.2	chromosome_9	6417251	6427088	-	1	4.00E-74	
Cr	Cre10.g419100.t1.1	chromosome_10	196272	203723	-	1	1.40E-11	
Cr	Cre10.g421350.t1.2	chromosome_10	536562	542782	+	1	1.20E-82	
Cr	Cre11.g467608.t1.1	chromosome_11	562486	570827	+	1	6.00E-89	MMP12
Cr	Cre11.g467683.t1.1	chromosome_11	1140060	1146445	-	1	6.70E-63	
Cr	Cre11.g467684.t1.1	chromosome_11	1146613	1152195	-	1	3.30E-66	

Table 4-5: Continued.

Cr	Cre11.g467686.t1.1	chromosome_11	1154620	1159945	-	1	1.10E-72	
Cr	Cre11.g467756.t1.1	chromosome_11	1552481	1561531	+	1	8.50E-86	MMP5
Cr	Cre11.g467600.t1.1	chromosome_11	1813212	1819113	-	1	1.30E-76	
Cr	Cre13.g605200.t1.1	chromosome_13	4800766	4806621	-	1	6.20E-88	
Cr	Cre14.g609202.t1.1	chromosome_14	218628	232029	+	1	4.90E-06	
Cr	Cre14.g621700.t1.1	chromosome_14	2018756	2024095	-	1	1.60E-94	
Cr	Cre14.g625850.t1.1	chromosome_14	2595078	2604975	-	1	2.70E-92	
Cr	Cre16.g694200.t1.1	chromosome_16	277646	284021	-	1	4.90E-48	
Cr	Cre16.g692200.t1.1	chromosome_16	601758	609759	+	1	2.60E-86	MMP6
Cr	Cre16.g652200.t1.2	chromosome_16	1401332	1408423	+	1	1.90E-91	MMP9
Cr	Cre16.g680902.t1.1	chromosome_16	7150116	7157137	+	1	6.70E-82	
Cr	Cre17.g705200.t1.1	chromosome_17	1234959	1242965	+	1	4.60E-76	
Cr	Cre17.g708450.t1.1	chromosome_17	1675745	1683067	+	1	9.60E-71	
Cr	Cre17.g715800.t1.1	chromosome_17	2443966	2449838	+	1	2.50E-76	
Cr	Cre17.g718468.t1.2	chromosome_17	2755532	2760353	+	1	3.40E-75	MMP2
Cr	Cre17.g718500.t1.1	chromosome_17	2760552	2766223	+	1	1.80E-90	MMP1
Cr	Cre17.g720150.t1.1	chromosome_17	2965182	2972213	+	1	4.90E-67	
Cr	Cre17.g728750.t1.2	chromosome_17	3974807	3979694	-	1	4.80E-60	
Eu	Eu_g8928	scaffold2	1	5510	-	1	1.60E-22	
Eu	Eu_g5891	scaffold5	1102746	1112347	-	1	1.90E-112	
Eu	Eu_g9648	scaffold7	222763	230187	+	1	1.20E-103	
Eu	Eu_g14203	scaffold12	384438	399229	-	2	7.30E-75	
Eu	Eu_g12700	scaffold13	539859	547110	+	1	9.60E-46	
Eu	Eu_g11837	scaffold18	72554	83128	-	1	3.40E-70	
Eu	Eu_g11867	scaffold18	212823	222577	-	2	5.00E-40	
Eu	Eu_g11870	scaffold18	230566	234756	-	1	4.80E-54	
Eu	Eu_g11873	scaffold18	241245	246163	-	1	1.10E-28	
Eu	Eu_g11875	scaffold18	251559	257270	-	1	7.30E-61	
Eu	Eu_g11878	scaffold18	279756	286060	-	1	1.80E-47	
Eu	Eu_g3275	scaffold20	94624	99130	-	1	9.80E-108	
Eu	Eu_g3276	scaffold20	104217	106490	+	1	9.00E-108	
Eu	Eu_g3278	scaffold20	113310	115217	+	1	4.80E-113	
Eu	Eu_g3280	scaffold20	120780	123113	+	1	1.90E-116	
Eu	Eu_g3284	scaffold20	129471	132415	-	1	4.40E-116	
Eu	Eu_g3287	scaffold20	138718	141980	-	1	2.10E-108	
Eu	Eu_g3379	scaffold20	752173	758451	+	1	2.20E-78	
Eu	Eu_g819	scaffold30	235181	241841	+	1	1.40E-88	
Eu	Eu_g18682_g18683	scaffold32	503	6134	+	1	2.50E-17	
Eu	Eu_g18598	scaffold35	1503	5471	-	1	7.20E-26	
Eu	Eu_g18600	scaffold35	13412	21677	-	1	6.50E-26	
Eu	Eu_g18602	scaffold35	28198	33931	-	1	4.90E-16	
Eu	Eu_g18632_g18633	scaffold35	235407	248513	+	1	9.60E-39	
Eu	Eu_g18717	scaffold37	12708	15942	-	1	7.90E-68	
Eu	Eu_g18720	scaffold37	26278	30287	-	1	3.60E-70	
Eu	Eu_g11018	scaffold40	429729	434042	+	1	8.30E-85	
Eu	Eu_g18404	scaffold50	18292	24261	+	1	2.10E-75	
Eu	Eu_g18405	scaffold50	28386	34212	+	1	2.30E-83	
Eu	Eu_g18406	scaffold50	36903	42998	+	1	2.80E-59	
Eu	Eu_g18409	scaffold50	50255	58097	+	1	3.50E-45	
Eu	Eu_g18410	scaffold50	60035	68300	+	1	1.20E-80	
Eu	Eu_g18411	scaffold50	70409	81134	+	2	9.90E-46	
Eu	Eu_g18413	scaffold50	89158	97305	+	1	1.80E-86	
Eu	Eu_g18415	scaffold50	109927	111964	+	1	3.50E-42	
Eu	Eu_g20463	scaffold61	91898	96270	+	1	5.20E-93	

Table 4-5: Continued.

Eu	Eu_g10552	scaffold67	268272	273224	+	1	4.00E-115	
Eu	Eu_g10553	scaffold67	279554	285201	+	1	3.60E-112	
Eu	Eu_g10556	scaffold67	292121	297850	+	1	4.00E-114	
Eu	Eu_g20782_g20781	scaffold72	2041	4720	-	1	4.10E-40	
Eu	Eu_g20784	scaffold72	12951	15149	-	1	2.10E-72	
Eu	Eu_g20786	scaffold72	19989	22401	-	1	4.50E-83	
Eu	Eu_g20788	scaffold72	43910	47866	-	1	3.40E-86	
Eu	Eu_g20789	scaffold72	53398	56623	-	1	3.20E-87	
Eu	Eu_g20790	scaffold72	58262	58840	-	1	4.40E-22	
Eu	Eu_g20791	scaffold72	61361	64432	-	1	4.20E-76	
Eu	Eu_g20792	scaffold72	67193	70300	-	1	2.00E-57	
Eu	Eu_g20794	scaffold72	74261	75880	-	1	4.30E-18	
Eu	Eu_g5014_g5015	scaffold89	1124664	1129386	+	1	9.90E-48	
Eu	Eu_g6125	scaffold97	51760	56570	+	1	9.40E-08	
Eu	Eu_g8074	scaffold106	520786	525302	+	1	5.80E-115	
Eu	Eu_g8076	scaffold106	529636	531236	+	1	4.30E-109	
Eu	Eu_g1367	scaffold120	867653	874893	-	1	8.90E-83	
Eu	Eu_g22588	scaffold172	1032	7764	-	1	2.20E-36	
Eu	Eu_g22590	scaffold172	13013	15139	-	1	1.40E-27	
Eu	Eu_g22464	scaffold216	275	1967	+	1	1.70E-29	
Eu	Eu_g22466	scaffold216	4808	7719	+	1	1.80E-70	
Eu	Eu_g22468	scaffold216	11371	14411	+	1	1.30E-67	
Eu	Eu_g22253	scaffold225	11851	13997	+	1	2.20E-62	
Eu	Eu_g22255	scaffold225	17680	19792	+	1	2.50E-15	
Eu	Eu_g22963	scaffold230	1	4082	+	1	4.30E-23	
Eu	Eu_g22157	scaffold254	7164	9529	+	1	5.50E-62	
Eu	Eu_g22159	scaffold254	12973	15200	+	1	6.70E-29	
Eu	Eu_g22161_g22162	scaffold254	18724	21096	+	1	1.90E-33	
Eu	Eu_g28271	scaffold319	4579	10192	+	1	2.30E-64	
Eu	Eu_g26471	scaffold382	8738	12648	-	2	8.90E-16	
Eu	Eu_g26793	scaffold428	5320	10837	-	1	4.60E-19	
Eu	Eu_g28151	scaffold548	3726	9398	-	1	1.90E-25	
Eu	Eu_g26587	scaffold671	3606	6771	-	1	1.40E-86	
Eu	Eu_g8189	scaffold1012	240579	244029	+	1	1.80E-116	
Eu	Eu_g8191	scaffold1012	248491	251177	+	1	1.60E-113	
Eu	Eu_g8192	scaffold1012	255443	259795	+	1	3.10E-112	
Eu	Eu_g7042	scaffold1014	735817	742633	+	1	3.20E-56	
Eu	Eu_g7043	scaffold1014	746426	752411	-	1	4.90E-77	
Eu	Eu_g7044	scaffold1014	754664	763356	-	1	1.20E-71	
Eu	Eu_g7045	scaffold1014	765582	770012	-	1	1.80E-46	
Eu	Eu_g7046	scaffold1014	770101	773953	-	1	1.90E-45	
Eu	Eu_g7047	scaffold1014	776662	791359	-	1	4.60E-91	
Eu	Eu_g7048	scaffold1014	791662	798199	-	1	1.80E-81	
Eu	Eu_g12075	scaffold1017	106358	111904	+	1	1.40E-58	
Eu	Eu_g8444	scaffold1021	942842	945908	-	1	4.30E-90	
Eu	Eu_g8447	scaffold1021	952312	957446	-	1	3.10E-86	
Eu	Eu_g8449	scaffold1021	960395	963148	-	1	3.20E-93	
Eu	Eu_g8450	scaffold1021	966357	971326	-	1	1.90E-34	
Eu	Eu_g11477	scaffold1022	106175	110065	-	1	4.00E-71	
Eu	Eu_g11483	scaffold1022	130070	134146	-	1	1.30E-50	
Eu	Eu_g11485	scaffold1022	144061	148778	-	1	1.50E-69	
Eu	Eu_g11486	scaffold1022	153790	157149	-	1	4.20E-80	
Eu	Eu_g11488	scaffold1022	163359	167911	-	1	9.80E-77	
Eu	Eu_g11490	scaffold1022	174276	178307	-	1	1.50E-76	

Table 4-5: Continued.

Eu	Eu_g9261	scaffold1024	1	7378	-	1	3.10E-103	
Eu	Eu_g9277	scaffold1024	92327	100190	-	1	1.70E-102	
Eu	Eu_g15541	scaffold1039	51397	56960	+	1	3.90E-88	
Eu	Eu_g15543	scaffold1039	61378	70251	-	1	3.10E-78	
Eu	Eu_g15544	scaffold1039	74418	79978	-	1	1.50E-89	
Eu	Eu_g15545	scaffold1039	84017	91178	+	1	1.20E-80	
Eu	Eu_g15548	scaffold1039	101866	104986	+	1	1.70E-74	
Eu	Eu_g15550	scaffold1039	112174	119743	+	1	2.20E-72	
Eu	Eu_g15553	scaffold1039	127752	133185	+	1	1.10E-90	
Eu	Eu_g15557	scaffold1039	143921	150719	+	1	7.90E-95	
Eu	Eu_g15562	scaffold1039	160144	166754	+	1	1.90E-48	
Eu	Eu_g15571	scaffold1039	232061	235867	+	1	4.00E-74	
Eu	Eu_g20067	scaffold1043	24298	30145	-	1	4.80E-56	
Eu	Eu_g20071	scaffold1043	41807	46772	+	1	7.50E-60	
Eu	Eu_g20172_g20171	scaffold1047	69498	74382	-	1	6.90E-40	
Eu	Eu_g15308	scaffold1048	81690	87298	+	1	5.10E-54	
Eu	Eu_g15343	scaffold1048	356613	359233	-	1	5.30E-76	
Eu	Eu_g20630	scaffold1053	21637	44813	-	1	8.00E-70	
Eu	Eu_g18142	scaffold1062	5916	12679	+	1	2.70E-86	
Eu	Eu_g21797	scaffold1114	16837	19520	-	1	8.50E-65	
Eu	Eu_g22622	scaffold1330	1	2558	-	1	1.70E-07	
Eu	Eu_g2483	scaffold1985	969339	971564	+	1	1.50E-51	
Eu	Eu_g2484	scaffold1985	972439	975122	+	1	7.60E-68	
Eu	Eu_g2486	scaffold1985	989537	993702	+	1	1.30E-90	
Eu	Eu_g2493	scaffold1985	1016955	1023898	+	1	6.20E-88	
Eu	Eu_g2663	scaffold1986	152679	160413	+	1	2.50E-77	
Eu	Eu_g2666	scaffold1986	166727	171802	-	1	8.60E-86	
Eu	Eu_g2667	scaffold1986	176613	180904	-	1	1.20E-91	
Eu	Eu_g2668	scaffold1986	183060	187156	+	1	3.40E-91	
Eu	Eu_g2753	scaffold1986	743271	750140	+	1	1.90E-70	
Eu	Eu_g2858	scaffold1986	1456566	1459815	+	1	2.20E-83	
Eu	Eu_g19250	scaffold2003	99372	104645	+	1	4.00E-71	
Eu	Eu_g19138	scaffold2012	4802	10917	+	1	1.00E-83	
Eu	Eu_g19169	scaffold2012	197037	205726	-	1	3.00E-65	
Eu	Eu_g19171	scaffold2012	211312	216709	-	1	2.50E-67	
Eu	Eu_g20984	scaffold2026	10654	17198	-	1	2.70E-52	
Eu	Eu_g20986	scaffold2026	47155	48136	+	1	3.50E-44	
Eu	Eu_g21827	scaffold2076	1	3769	-	1	9.60E-37	
Eu	Eu_g21828	scaffold2076	5365	8489	-	1	2.80E-37	
Eu	Eu_g21829	scaffold2076	11314	14465	-	1	3.10E-67	
Eu	Eu_g21831_g21830	scaffold2076	17837	20961	-	1	3.00E-40	
Eu	Eu_g21832	scaffold2076	23835	25493	-	1	3.40E-28	
Eu	Eu_g26517	scaffold2601	1498	4138	-	2	1.80E-38	
Eu	Eu_g26519	scaffold2601	7809	11261	-	1	2.30E-42	
Eu	Eu_g5514	scaffold2949	1162622	1169864	+	1	3.10E-40	
Eu	Eu_g14358	scaffold2950	247480	251010	+	1	7.10E-106	
Eu	Eu_g16685	scaffold2952	384135	389487	+	1	2.20E-85	
Eu	Eu_g16691	scaffold2952	416393	418321	+	1	2.90E-16	
Eu	Eu_g16379	scaffold2958	183105	187402	+	1	4.10E-71	
Eu	Eu_g22729	scaffold3100	12143	15982	+	1	7.90E-21	
Eu	Eu_g22730	scaffold3100	18575	19213	+	1	1.40E-16	
Eu	Eu_g25015_g25014	scaffold3185	1913	5029	-	1	5.40E-29	
Eu	Eu_g25016	scaffold3185	7863	11017	-	1	4.80E-46	
Eu	Eu_g25017	scaffold3185	13850	15554	-	1	2.90E-22	

Table 4-5: Continued.

Eu	Eu_g23975	scaffold3207	15761	16917	-	1	3.90E-17	
Eu	Eu_g4593	scaffold3898	1528821	1531881	+	1	3.90E-83	
Eu	Eu_g4594	scaffold3898	1533466	1552495	+	3	9.20E-85	
Eu	Eu_g4595	scaffold3898	1557178	1560639	+	1	2.80E-81	
Eu	Eu_g4597	scaffold3898	1569958	1573094	+	1	2.80E-86	
Eu	Eu_g4598	scaffold3898	1589280	1592428	+	1	7.70E-87	
Eu	Eu_g4599	scaffold3898	1595245	1598389	+	1	1.30E-86	
Eu	Eu_g4600	scaffold3898	1601242	1603339	+	1	1.80E-17	
Eu	Eu_g12224	scaffold3901	335832	342142	-	1	2.50E-82	
Eu	Eu_g18121	scaffold3911	207418	215225	+	1	2.00E-75	
Eu	Eu_g18123	scaffold3911	220434	223350	+	1	3.10E-84	
Eu	Eu_g18124	scaffold3911	229996	232470	+	1	1.60E-83	
Eu	Eu_g18126	scaffold3911	237909	241245	+	1	2.20E-83	
Eu	Eu_g18132	scaffold3911	265948	268461	+	1	2.40E-83	
Eu	Eu_g18134	scaffold3911	275294	276983	+	1	9.40E-53	
Eu	Eu_g18135	scaffold3911	279291	282486	+	1	3.60E-80	
Eu	Eu_g18136	scaffold3911	287177	291994	-	1	3.80E-87	
Eu	Eu_g18138	scaffold3911	301220	301674	-	1	1.00E-16	
Eu	Eu_g18140	scaffold3911	306313	308121	-	1	4.80E-42	
Eu	Eu_g20605	scaffold3922	57101	62368	-	1	1.90E-82	
Eu	Eu_g20606	scaffold3922	69680	72079	-	1	1.30E-25	
Eu	Eu_g20610	scaffold3922	94978	96940	-	1	3.20E-25	
Eu	Eu_g20340	scaffold3923	3120	10686	-	1	6.30E-71	
Eu	Eu_g20543	scaffold3924	22178	36006	+	1	3.60E-103	
Eu	Eu_g23052_g23053	scaffold4219	2351	14228	+	1	3.00E-29	
Eu	Eu_g28545	scaffold4497	4807	6005	-	1	3.40E-21	
Eu	Eu_g23516	scaffold4507	1	2096	-	1	2.20E-49	
Eu	Eu_g23517	scaffold4507	5468	8603	-	1	2.00E-57	
Eu	Eu_g23518	scaffold4507	11438	14589	-	1	7.00E-47	
Gp	GPECTOR_1g282	scaffold00001	1961968	1962348	+	1	3.80E-24	
Gp	GPECTOR_1g283	scaffold00001	1963231	1963958	+	1	1.30E-11	
Gp	GPECTOR_4g645	scaffold00004	907160	911009	-	1	1.50E-62	
Gp	GPECTOR_4g665	scaffold00004	1115162	1120277	-	1	4.30E-72	
Gp	GPECTOR_6g583	scaffold00006	861830	870345	-	1	4.80E-79	
Gp	GPECTOR_6g740	scaffold00006	2007982	2011073	-	1	5.30E-76	
Gp	GPECTOR_10g909	scaffold00010	1195787	1200501	+	1	3.00E-92	
Gp	GPECTOR_11g263	scaffold00011	2070608	2079461	-	1	6.20E-06	
Gp	GPECTOR_12g378	scaffold00012	393789	395116	+	2	1.80E-23	
Gp	GPECTOR_13g774	scaffold00013	1525468	1529659	-	1	3.30E-74	
Gp	GPECTOR_14g57	scaffold00014	504017	510354	-	2	6.40E-81	
Gp	GPECTOR_14g58	scaffold00014	510903	511970	-	1	2.80E-32	
Gp	GPECTOR_14g159	scaffold00014	1227678	1228681	+	1	5.90E-24	
Gp	GPECTOR_21g660	scaffold00021	634680	638214	+	1	7.10E-73	
Gp	GPECTOR_21g723	scaffold00021	1147735	1151116	+	1	6.20E-108	
Gp	GPECTOR_23g167	scaffold00023	1453156	1456379	-	1	1.00E-86	
Gp	GPECTOR_24g222	scaffold00024	886211	893853	-	1	1.20E-16	
Gp	GPECTOR_24g298	scaffold00024	1793265	1793575	+	1	7.40E-11	
Gp	GPECTOR_27g673	scaffold00027	668298	669200	-	1	1.00E-10	
Gp	GPECTOR_27g675	scaffold00027	708986	712405	+	1	1.30E-31	
Gp	GPECTOR_28g796	scaffold00028	661011	663584	+	1	8.90E-87	
Gp	GPECTOR_28g827	scaffold00028	975299	981479	-	1	1.50E-91	
Gp	GPECTOR_28g830	scaffold00028	1001353	1011832	-	1	3.50E-65	
Gp	GPECTOR_28g855	scaffold00028	1265316	1270733	+	1	1.90E-90	
Gp	GPECTOR_40g574	scaffold00040	666844	673684	-	1	7.40E-91	

Table 4-5: Continued.

Gp	GPECTOR_40g608	scaffold000040	895772	900769	+	1	8.40E-97	
Gp	GPECTOR_42g765	scaffold000042	149412	153781	-	1	1.60E-37	
Gp	GPECTOR_45g182	scaffold000045	824998	833206	+	1	1.10E-21	
Gp	GPECTOR_48g409	scaffold000048	194745	200779	+	1	7.60E-59	
Gp	GPECTOR_48g434	scaffold000048	1086599	1090399	-	1	6.10E-31	
Gp	GPECTOR_55g263	scaffold000055	129271	134113	+	1	7.90E-88	
Gp	GPECTOR_60g721	scaffold000060	182655	185316	-	1	5.80E-60	
Gp	GPECTOR_60g722	scaffold000060	188500	194374	-	1	1.10E-61	
Gp	GPECTOR_60g723	scaffold000060	197026	201701	-	1	1.40E-65	
Gp	GPECTOR_60g774	scaffold000060	674969	679324	+	1	5.20E-26	
Gp	GPECTOR_63g68	scaffold000063	494348	499897	-	1	6.30E-89	
Gp	GPECTOR_68g385	scaffold000068	374981	382252	-	1	4.20E-73	
Gp	GPECTOR_76g794	scaffold000076	196410	201152	-	1	1.70E-87	
Gp	GPECTOR_76g795	scaffold000076	202309	208497	-	1	2.20E-87	
Gp	GPECTOR_76g801	scaffold000076	249745	252125	+	1	3.80E-14	
Gp	GPECTOR_78g75	scaffold000078	210160	216359	+	1	1.20E-89	
Gp	GPECTOR_97g754	scaffold000097	82806	85711	+	1	4.50E-81	
Gp	GPECTOR_99g816	scaffold000099	192983	194705	+	1	3.80E-06	
Gp	GPECTOR_104g85	scaffold00104	122985	133682	-	1	1.60E-73	
Gp	GPECTOR_183g266	scaffold00183	19436	19870	+	1	7.50E-29	
Gp	GPECTOR_1481g664	scaffold01481	938	1864	+	1	1.70E-27	
Ts	TSOC_001165	scaffold_18	216227	220647	-	1	1.60E-83	
Ts	TSOC_002689	scaffold_52	192824	196014	+	1	3.40E-31	
Ts	TSOC_002690	scaffold_52	196222	197901	+	1	8.60E-41	
Ts	TSOC_002755	scaffold_54	80431	113258	-	5	9.40E-56	
Ts	TSOC_002954	scaffold_60	31244	37791	-	1	1.50E-38	
Ts	TSOC_003163	scaffold_65	179265	182824	-	1	5.40E-76	
Ts	TSOC_003195	scaffold_66	297543	303589	-	1	2.10E-19	
Ts	TSOC_003228	scaffold_68	402	1917	-	1	4.40E-38	
Ts	TSOC_003441	scaffold_74	129461	132110	+	1	5.90E-102	
Ts	TSOC_003589	scaffold_79	125998	129316	-	1	3.20E-93	
Ts	TSOC_003598	scaffold_79	173908	181250	+	1	2.20E-86	
Ts	TSOC_005473	scaffold_149	184899	191211	-	1	9.70E-13	
Ts	TSOC_005565	scaffold_153	114449	121915	-	1	8.90E-81	
Ts	TSOC_005975	scaffold_174	33713	44900	+	1	5.90E-69	
Ts	TSOC_006833	scaffold_216	6730	11876	-	1	1.20E-57	
Ts	TSOC_006886	scaffold_219	96732	102655	+	1	6.40E-33	
Ts	TSOC_007229	scaffold_239	16417	26773	-	1	3.10E-33	
Ts	TSOC_007440	scaffold_250	119832	124284	+	1	1.40E-72	
Ts	TSOC_008656	scaffold_333	56042	58261	-	1	3.90E-36	
Ts	TSOC_008716	scaffold_338	78194	79367	+	1	2.40E-17	
Ts	TSOC_009489	scaffold_397	79040	88560	-	1	1.80E-16	
Ts	TSOC_010606	scaffold_515	31249	47722	-	1	1.10E-57	
Ts	TSOC_011331	scaffold_613	550	7436	+	1	5.50E-73	
Ts	TSOC_012255	scaffold_785	29786	31261	+	1	2.60E-48	
Ts	TSOC_012758	scaffold_922	1223	4319	+	1	5.70E-62	
Ts	TSOC_013403	scaffold_1187	46	988	+	1	5.60E-28	
Ts	TSOC_014009	scaffold_1633	1	2379	-	1	8.10E-21	
Ts	TSOC_014219	scaffold_1847	6	6917	-	1	5.30E-28	
Ts	TSOC_014403	scaffold_2114	2414	3843	+	1	4.80E-16	
Ts	TSOC_014803	scaffold_2957	1	1942	+	1	2.80E-48	
Ts	TSOC_014938	scaffold_3417	2	1508	+	1	6.00E-42	
Ts	TSOC_015250	scaffold_4646	279	1151	+	1	2.00E-32	
Vc	Vocar.0001s0355.1.p	scaffold_1	2666898	2678643	+	1	6.60E-90	VMP38

Table 4-5: Continued.

Vc	Vocar.0001s1363.1.p	scaffold_1	10493026	10504466	-	1	1.20E-77	
Vc	Vocar.0002s0001.1.p	scaffold_2	48039	73079	+	1	6.10E-105	VMP4
Vc	Vocar.0002s0127.1.p	scaffold_2	830457	843175	-	1	5.00E-83	
Vc	Vocar.0003s0191.1.p	scaffold_3	1985628	1989330	-	1	2.70E-77	
Vc	Vocar.0003s0288.1.p	scaffold_3	2820882	2830841	-	1	6.10E-89	
Vc	Vocar.0003s0314.1.p	scaffold_3	3123208	3133001	+	1	1.30E-86	
Vc	Vocar.0003s0417.1.p	scaffold_3	4126470	4131979	+	1	4.40E-108	VMP39
Vc	Vocar.0003s0448.1.p	scaffold_3	4477596	4488472	-	1	6.60E-55	
Vc	Vocar.0004s0009.1.p	scaffold_4	54312	62329	+	1	1.00E-89	
Vc	Vocar.0004s0166.1.p	scaffold_4	1389463	1393428	-	1	5.10E-89	
Vc	Vocar.0004s0228.1.p	scaffold_4	1855178	1863410	+	1	1.60E-72	
Vc	Vocar.0004s0341.1.p	scaffold_4	2770548	2780079	+	1	1.50E-65	
Vc	g89659	scaffold_7	2933697	2946948	+	1	3.20E-59	
Vc	Vocar.0007s0366.1.p	scaffold_7	2947545	2953534	-	1	1.10E-113	VMP1
Vc	Vocar.0007s0367.1.p	scaffold_7	2965243	2969712	-	1	4.40E-111	VMP37
Vc	Vocar.0008s0181.1.p	scaffold_8	1516532	1524169	+	1	8.10E-84	
Vc	Vocar.0008s0182.1.p	scaffold_8	1525426	1531433	+	1	6.20E-81	
Vc	Vocar.0008s0183.1.p	scaffold_8	1532168	1538622	+	1	1.00E-82	
Vc	Vocar.0008s0184.1.p	scaffold_8	1538951	1547096	+	1	1.40E-69	
Vc	Vocar.0008s0187.1.p	scaffold_8	1553105	1563757	+	1	4.00E-79	
Vc	Vocar.0008s0188.1.p	scaffold_8	1566523	1577476	+	1	2.40E-70	
Vc	Vocar.0011s0345.1.p	scaffold_11	2855655	2863577	+	1	2.60E-96	
Vc	Vocar.0011s0346.1.p	scaffold_11	2865263	2871936	-	1	9.80E-99	
Vc	g42097	scaffold_11	2893443	2895629	+	1	2.80E-107	
Vc	g66631	scaffold_11	2899398	2901340	+	1	4.40E-112	
Vc	g97008	scaffold_11	2908834	2915916	+	2	1.80E-45	
Vc	g66589	scaffold_11	2918542	2922165	+	1	2.60E-102	
Vc	g41285	scaffold_11	2932756	2938960	+	1	4.80E-108	
Vc	Vocar.0011s0347.1.p	scaffold_11	2964294	2974436	+	1	1.60E-85	
Vc	g66557	scaffold_11	2975372	2978859	-	1	1.90E-111	VMP12
Vc	g127216	scaffold_11	2979400	2982673	-	1	2.30E-111	VMP15
Vc	g66578	scaffold_11	2982491	2985573	-	1	1.20E-114	VMP13
Vc	g127215	scaffold_11	2986914	2989827	-	1	2.60E-112	VMP14
Vc	g127218	scaffold_11	2996390	2999112	+	1	1.50E-89	VMP16
Vc	g127219	scaffold_11	3000305	3003114	+	1	7.30E-55	VMP17
Vc	Vocar.0012s0179.1.p	scaffold_12	1709206	1720103	+	1	1.50E-97	LSG2
Vc	Vocar.0012s0180.1.p	scaffold_12	1722202	1728207	+	1	4.60E-98	
Vc	Vocar.0012s0181.1.p	scaffold_12	1734942	1748807	+	1	5.60E-98	
Vc	Vocar.0012s0269.1.p	scaffold_12	2569104	2581881	+	1	8.00E-84	
Vc	Vocar.0012s0270.1.p	scaffold_12	2583023	2591871	-	1	8.40E-89	
Vc	Vocar.0012s0271.1.p	scaffold_12	2598327	2605482	-	1	2.10E-90	VMP42
Vc	Vocar.20007296m	scaffold_13	4915	11911	-	1	4.40E-120	VMP29
Vc	Vocar.0013s0081.1.p	scaffold_13	669665	678207	-	1	6.40E-95	
Vc	Vocar.0014s0029.1.p	scaffold_14	233106	249634	-	1	6.30E-84	
Vc	Vocar.0014s0030.1.p	scaffold_14	254309	263958	-	1	3.90E-100	
Vc	Vocar.0014s0044.1.p	scaffold_14	337429	356802	-	1	1.30E-96	
Vc	Vocar.0014s0045.1.p	scaffold_14	357659	364727	-	1	5.70E-36	
Vc	Vocar.0014s0062.1.p	scaffold_14	494984	503695	-	1	2.00E-83	
Vc	Vocar.0014s0203.1.p	scaffold_14	1655662	1665458	-	1	2.20E-95	
Vc	Vocar.0014s0204.1.p	scaffold_14	1667750	1678752	-	1	2.90E-93	
Vc	Vocar.0015s0120.1.p	scaffold_15	917050	922509	-	1	3.10E-86	
Vc	g56439	scaffold_15	1490502	1492470	+	1	3.40E-98	VMP36
Vc	g101266	scaffold_15	1501958	1508717	+	1	2.40E-84	VMP35/VMP41
Vc	g41348	scaffold_15	1509941	1515697	+	1	2.10E-98	VMP34

Table 4-5: Continued.

Vc	g103856	scaffold_16	202380	226172	-	1	3.00E-76	
Vc	Vocar.0016s0063.1.p	scaffold_16	505005	509850	-	1	2.90E-116	VMP10
Vc	Vocar.0016s0064.1.p	scaffold_16	512712	519655	+	1	3.80E-116	VMP3
Vc	g40166	scaffold_16	525269	527809	+	1	9.30E-114	VMP11
Vc	Vocar.0016s0065.1.p	scaffold_16	533737	541225	-	1	2.60E-108	VMP9
Vc	Vocar.0016s0066.1.p	scaffold_16	546375	552941	-	1	5.60E-117	VMP8
Vc	g105896	scaffold_17	263264	271179	-	1	2.80E-109	VMP26
Vc	Vocar.0017s0039.1.p	scaffold_17	392324	399595	-	1	2.00E-118	VMP7
Vc	Vocar.0017s0047.1.p	scaffold_17	458133	463639	+	1	7.70E-106	VMP28
Vc	Vocar.0017s0049.1.p	scaffold_17	470432	473174	+	1	2.00E-115	VMP27
Vc	Vocar.0017s0175.1.p	scaffold_17	1921198	1925978	-	1	3.10E-87	
Vc	Vocar.0017s0176.1.p	scaffold_17	1934768	1940076	+	1	1.20E-68	
Vc	Vocar.0019s0056.1.p	scaffold_19	494412	503398	+	1	3.30E-93	
Vc	g107831	scaffold_20	2210311	2224914	-	1	6.50E-08	
Vc	Vocar.0021s0204.1.p	scaffold_21	2115807	2124887	-	1	3.30E-92	
Vc	Vocar.0021s0205.1.p	scaffold_21	2132219	2145517	-	1	4.10E-82	
Vc	Vocar.0021s0206.1.p	scaffold_21	2147169	2159486	-	1	2.20E-90	
Vc	Vocar.0021s0208.1.p	scaffold_21	2164624	2173123	-	1	6.80E-87	
Vc	Vocar.20003206m	scaffold_21	2185613	2193173	-	1	3.30E-74	
Vc	Vocar.0021s0209.1.p	scaffold_21	2197958	2210683	-	1	1.30E-89	
Vc	Vocar.0028s0043.1.p	scaffold_28	469012	493605	-	1	5.30E-07	
Vc	Vocar.0028s0044.2.p	scaffold_28	497109	525380	-	1	8.70E-06	
Vc	Vocar.0028s0111.1.p	scaffold_28	1058569	1066253	+	1	1.40E-82	
Vc	Vocar.0028s0114.1.p	scaffold_28	1100516	1109154	-	1	9.90E-73	
Vc	Vocar.0028s0115.1.p	scaffold_28	1114433	1117581	-	1	1.30E-88	
Vc	Vocar.0028s0173.1.p	scaffold_28	1558417	1571919	-	1	5.10E-62	
Vc	Vocar.0028s0174.1.p	scaffold_28	1576357	1586783	-	1	5.40E-56	
Vc	g90284	scaffold_29	1357781	1374747	+	1	6.60E-79	VMP40
Vc	Vocar.0031s0068.1.p	scaffold_31	909659	922624	+	1	8.10E-77	
Vc	Vocar.0031s0069.1.p	scaffold_31	927391	946324	-	1	1.10E-54	
Vc	Vocar.0037s0139.1.p	scaffold_37	806608	813717	+	1	7.60E-86	
Vc	Vocar.0040s0051.1.p	scaffold_40	400478	407147	+	1	1.90E-71	
Vc	Vocar.0040s0065.1.p	scaffold_40	502538	510517	+	1	2.40E-85	
Vc	g127222	scaffold_40	776364	779145	-	1	5.60E-105	VMP32
Vc	g127223	scaffold_40	779959	783054	-	1	3.00E-76	VMP33
Vc	g66400	scaffold_40	789632	792930	-	1	1.00E-108	VMP30
Vc	g61971	scaffold_45	584917	594876	+	1	9.30E-107	VMP6
Vc	g121539	scaffold_50	144275	152652	-	1	3.30E-06	
Vc	g42295	scaffold_54	467309	469321	-	1	1.10E-113	VMP20
Vc	Vocar.0054s0052.1.p	scaffold_54	473466	483143	-	1	4.50E-118	VMP21
Vc	Vocar.0054s0053.1.p	scaffold_54	492602	501966	-	1	2.00E-117	VMP22
Vc	Vocar.0054s0055.1.p	scaffold_54	517009	521871	-	1	3.20E-121	VMP23
Vc	Vocar.0062s0024.1.p	scaffold_62	182569	189712	+	1	3.00E-112	VMP2
Vc	Vocar.0062s0029.1.p	scaffold_62	239653	252330	+	1	8.20E-115	
Vc	Vocar.0062s0031.1.p	scaffold_62	270906	279611	+	1	2.20E-115	VMP19
Vc	Vocar.0062s0033.1.p	scaffold_62	319529	324025	+	1	2.00E-120	VMP25
Vc	Vocar.0062s0036.1.p	scaffold_62	398693	405803	+	1	7.40E-120	VMP24
Vc	Vocar.0094s0001.1.p	scaffold_94	81933	94051	-	1	2.70E-76	VMP5
Yu	Yu_g969	Scaffold0002	2089226	2093927	-	1	3.40E-76	
Yu	Yu_g1263	Scaffold0003	764709	768555	-	1	5.40E-57	
Yu	Yu_g1482	Scaffold0003	2206033	2210053	-	1	1.00E-85	
Yu	Yu_g1494	Scaffold0003	2291293	2294564	-	1	2.20E-66	
Yu	Yu_g1543	Scaffold0003	2671119	2680822	-	1	2.40E-46	
Yu	Yu_g1610	Scaffold0004	333083	342979	-	1	9.10E-37	

Table 4-5: Continued.

Yu	Yu_g1722	Scaffold0004	1007894	1010883	+	1	4.50E-69
Yu	Yu_g1943	Scaffold0004	2449055	2455283	-	1	2.20E-69
Yu	Yu_g1944	Scaffold0004	2457770	2461618	-	1	7.70E-58
Yu	Yu_g1972	Scaffold0004	2616639	2624894	+	1	1.40E-77
Yu	Yu_g2796	Scaffold0007	224076	229416	+	1	7.70E-41
Yu	Yu_g2798	Scaffold0007	230304	232504	+	1	4.10E-45
Yu	Yu_g2800	Scaffold0007	237248	239357	+	1	3.80E-30
Yu	Yu_g2802	Scaffold0007	245721	252537	+	1	2.10E-55
Yu	Yu_g3387	Scaffold0008	1836852	1842466	-	1	4.10E-92
Yu	Yu_g4008	Scaffold0011	329899	333149	-	1	5.50E-104
Yu	Yu_g4819	Scaffold0014	699030	705130	-	1	1.20E-72
Yu	Yu_g5143	Scaffold0015	1373023	1377239	+	1	2.40E-72
Yu	Yu_g6731	Scaffold0023	779927	783206	+	1	1.20E-90
Yu	Yu_g6732	Scaffold0023	788553	793495	+	1	9.80E-94
Yu	Yu_g7397	Scaffold0027	519210	524237	+	1	8.50E-83
Yu	Yu_g7517	Scaffold0028	253460	256552	+	1	4.70E-85
Yu	Yu_g7519	Scaffold0028	262345	263679	+	1	1.50E-76
Yu	Yu_g7521	Scaffold0028	267371	269897	+	1	1.90E-82
Yu	Yu_g7522	Scaffold0028	274734	276658	+	1	3.30E-72
Yu	Yu_g7524	Scaffold0028	285905	288239	+	1	2.40E-82
Yu	Yu_g8093	Scaffold0031	939667	945852	+	1	4.90E-80
Yu	Yu_g9404	Scaffold0041	247877	251437	-	1	2.60E-99
Yu	Yu_g10351	Scaffold0049	240486	244110	+	1	6.70E-112
Yu	Yu_g10352	Scaffold0049	250385	253533	+	1	1.20E-107
Yu	Yu_g10353	Scaffold0049	259154	261031	+	1	2.40E-108
Yu	Yu_g10355	Scaffold0049	266911	269603	+	1	1.90E-107
Yu	Yu_g10678	Scaffold0052	335237	341667	-	1	2.70E-67
Yu	Yu_g10927	Scaffold0054	579457	585401	+	1	3.50E-88
Yu	Yu_g11015	Scaffold0055	458301	460885	-	1	1.10E-106
Yu	Yu_g11395	Scaffold0059	316136	320527	-	1	4.40E-57
Yu	Yu_g11765	Scaffold0063	363943	369223	-	1	9.30E-90
Yu	Yu_g11941	Scaffold0065	238508	244084	-	1	1.70E-60
Yu	Yu_g11942	Scaffold0065	250668	255162	-	1	5.80E-79
Yu	Yu_g12176	Scaffold0068	36266	37897	+	1	9.40E-77
Yu	Yu_g12981	Scaffold0076	516929	523636	-	1	3.50E-48
Yu	Yu_g13119	Scaffold0078	451499	453867	+	1	6.00E-78
Yu	Yu_g13120	Scaffold0078	457023	459382	+	1	8.50E-78
Yu	Yu_g13121	Scaffold0078	461055	463275	+	1	8.30E-79
Yu	Yu_g13182	Scaffold0079	372452	376184	-	1	1.10E-56
Yu	Yu_g14195	Scaffold0093	244159	253977	+	1	9.50E-53
Yu	Yu_g14196	Scaffold0093	259478	267192	-	1	5.30E-50
Yu	Yu_g14255	Scaffold0094	269500	272571	-	1	4.10E-90
Yu	Yu_g14258	Scaffold0094	278697	281771	+	1	3.60E-90
Yu	Yu_g14259	Scaffold0094	284948	288005	+	1	1.70E-90
Yu	Yu_g14758	Scaffold0102	30867	38120	-	1	8.10E-06
Yu	Yu_g16332	Scaffold0134	4380	14216	+	1	2.40E-46
Yu	Yu_g16905	Scaffold0149	142065	143533	-	1	7.30E-60
Yu	Yu_g17013	Scaffold0153	63681	77306	-	1	6.00E-52
Yu	Yu_g17268	Scaffold0167	52343	56031	-	1	1.60E-102
Yu	Yu_g17445	Scaffold0175	86521	91821	+	1	1.30E-83
Yu	Yu_g17660	Scaffold0187	1	10093	+	1	7.80E-44
Yu	Yu_g18334	Scaffold0294	16218	28507	-	1	1.30E-43
Yu	Yu_g18403	Scaffold0313	1	8413	-	1	4.90E-44
Yu	Yu_g20594	Scaffold1116	1	10449	-	1	8.10E-44

Table 4-5: Continued.

Yu	Yu_g20871	Scaffold1249	306	11572	-	1	1.10E-43	
----	-----------	--------------	-----	-------	---	---	----------	--

^a Ag: *A. gubernaculifera*, Cr: *C. reinhardtii*, Eu: *Eudorina* sp., Gp: *G. pectorale*, Ts: *T. socialis*, Vc: *V. carteri*, Yu: *Y. unicocca*. The sources of genome data are listed in Table 4-1.

^b For *V. carteri* : g##### (v1.0); Vocar#####m (v2.0); Vocar.#####s#####.p (v2.1).

^c For *V. carteri*, coordinate locations in v2.1 genome are shown in all protein models.

^d Based on Prochnik et al. (2010) and Hanschen et al. (2016).

Table 4-6: List of pherophorin genes in genome data of volvocine algae.

Species ^a	Protein Model ^b	Scaffold ^c	Start Location ^c	End Location ^c	Strand ^c	No. of Pherophorin Domain	Evalue ^d	Common Name ^e
Ag	Agub_g803.1.p1	Scf001	3465172	3468895	-	2	8.4E-35	
Ag	Agub_g1106.2.p1	Scf001	5588642	5598450	-	1	8.3E-24	
Ag	Agub_g1107.1.p1	Scf001	5598555	5605703	+	1	1.3E-23	
Ag	Agub_g2415.1.p1	Scf002	1409326	1413719	-	2	1E-27	
Ag	Agub_g2416.1.p1	Scf002	1414453	1418816	-	2	2.8E-28	
Ag	Agub_g2613.2.p1	Scf002	2478819	2482333	-	2	3.5E-26	
Ag	Agub_g2619.1.p1	Scf002	2512149	2517367	+	2	1.5E-26	
Ag	Agub_g2911.1.p1	Scf002	4920798	4927965	+	1	1.2E-18	
Ag	Agub_g3149.1.p1	Scf003	540475	542183	+	1	3E-26	
Ag	Agub_g3150.1.p1	Scf003	542358	544424	+	1	4.8E-28	
Ag	Agub_g3615.1.p1	Scf003	3476767	3478466	-	1	1.1E-22	
Ag	Agub_g3653.1.p1	Scf003	3739244	3742918	+	1	2.2E-43	
Ag	Agub_g4456.1.p1	Scf005	2238221	2241846	+	4	1.2E-16	
Ag	Agub_g4922.1.p1	Scf006	1472787	1480416	+	1	4.1E-42	
Ag	Agub_g5596.1.p1	Scf007	2151373	2153407	+	1	9.3E-23	
Ag	Agub_g6523.1.p1	Scf009	1142271	1143512	-	1	3.7E-35	
Ag	Agub_g7134.1.p1	Scf010	1854419	1856334	-	1	2.9E-28	
Ag	Agub_g7135.1.p1	Scf010	1860021	1863894	-	1	2.1E-24	
Ag	Agub_g7136.1.p1	Scf010	1866943	1869465	-	1	1.9E-26	
Ag	Agub_g7137.1.p1	Scf010	1871481	1874288	-	1	5.7E-25	
Ag	Agub_g7272.1.p1	Scf010	2463699	2466879	+	2	3.3E-38	
Ag	Agub_g7275.2.p1	Scf010	2478936	2482240	-	2	9.3E-37	
Ag	Agub_g7482.2.p1	Scf011	1027902	1032637	+	1	3.3E-37	
Ag	Agub_g7884.1.p1	Scf013	1421414	1425220	+	2	2.3E-31	
Ag	Agub_g10451.1.p1	Scf024	568943	570347	+	1	6.7E-35	
Ag	Agub_g10668.1.p1	Scf025	453795	460460	+	2	1.3E-37	
Ag	Agub_g11010.1.p1	Scf027	439549	441289	-	1	9.1E-25	
Ag	Agub_g11011.1.p1	Scf027	442760	444746	-	1	7.4E-28	
Ag	Agub_g11013.1.p1	Scf027	451270	454587	+	1	4.5E-26	
Ag	Agub_g11014.1.p1	Scf027	455031	457889	+	1	2.3E-28	
Ag	Agub_g11083.2.p1	Scf027	722194	725439	+	2	2.5E-16	
Ag	Agub_g10884.1.p1	Scf028	36984	38047	-	1	1.1E-10	
Ag	Agub_g11471.1.p1	Scf030	189972	193036	-	1	7.7E-26	
Ag	Agub_g11472.1.p1	Scf030	195326	196260	-	1	7.4E-19	
Ag	Agub_g11474.1.p1	Scf030	198853	202142	-	1	2.1E-26	
Ag	Agub_g11475.1.p1	Scf030	203143	207142	-	1	9.4E-28	
Ag	Agub_g11589.1.p1	Scf031	64368	69413	-	1	3.1E-10	
Ag	Agub_g11719.1.p1	Scf032	821473	824665	-	1	3.1E-23	
Ag	Agub_g11720.1.p1	Scf032	824748	825692	-	1	6.4E-27	
Ag	Agub_g12144.3.p1	Scf034	453608	504436	-	2	1.9E-42	
Ag	Agub_g12153.1.p1	Scf034	504564	506386	+	1	2.7E-39	
Ag	Agub_g12154.1.p1	Scf034	508425	510771	+	1	4.4E-31	
Ag	Agub_g12504.1.p1	Scf036	748712	752748	+	1	9.6E-43	
Ag	Agub_g12816.1.p1	Scf038	423470	424541	-	1	1.7E-14	
Ag	Agub_g12817.1.p1	Scf038	425026	430975	-	1	3.2E-26	
Ag	Agub_g12688.1.p1	Scf038	790162	792644	-	1	5.1E-57	
Ag	Agub_g12687.1.p1	Scf038	793869	794975	-	1	2.2E-56	
Ag	Agub_g13723.1.p1	Scf050	183722	188208	+	1	1.4E-13	
Ag	Agub_g13916.1.p1	Scf051	130457	134346	+	1	2.2E-43	
Ag	Agub_g14323.1.p1	Scf055	215371	216654	+	1	5.6E-35	
Ag	Agub_g14324.1.p1	Scf055	216874	219400	+	1	7.8E-15	

Table 4-6: Continued.

Ag	Agub_g15106.2.p1	Scf066	183991	192297	-	1	1.1E-31	
Ag	Agub_g15292.1.p1	Scf069	758	32215	+	1	5E-32	
Ag	Agub_g16031.1.p1	Scf118	31555	34721	+	2	4.5E-41	
Cr	Cre01.g047265.t1.1	chromosome_1	6598706	6608552	+	1	5.7E-31	
Cr	Cre01.g049826.t1.1	chromosome_1	6879928	6923528	+	1	1.5E-26	
Cr	Cre02.g077550.t1.2	chromosome_2	558207	562518	+	2	3.9E-17	
Cr	Cre02.g078777.t1.1	chromosome_2	771148	777657	+	2	6.4E-30	phC16a
Cr	Cre02.g078804.t1.1	chromosome_2	778398	782412	+	2	2.2E-30	phC16b
Cr	Cre02.g079926.t1.1	chromosome_2	954805	963880	-	1	9.2E-26	
Cr	Cre02.g081700.t1.1	chromosome_2	1153656	1157872	+	1	2.4E-16	
Cr	Cre02.g094450.t1.2	chromosome_2	2808946	2813363	+	2	4.8E-31	phC21
Cr	Cre02.g095095.t1.1	chromosome_2	3345375	3351647	+	2	7E-38	phC12
Cr	Cre02.g118650.t1.1	chromosome_2	6767677	6773962	-	1	1.1E-18	
Cr	Cre03.g155300.t1.1	chromosome_3	1922313	1926910	+	4	2.6E-16	
Cr	Cre03.g155750.t1.1	chromosome_3	1972170	1977210	-	1	5.5E-22	
Cr	Cre03.g161150.t1.1	chromosome_3	2671641	2686777	-	1	9.7E-09	
Cr	Cre03.g192201.t1.1	chromosome_3	6167505	6172700	-	1	7.8E-27	
Cr	Cre03.g192250.t1.1	chromosome_3	6172881	6185562	+	1	4.9E-24	
Cr	Cre04.g219050.t1.1	chromosome_4	2047750	2062645	-	1	3.8E-17	
Cr	Cre04.g221450.t1.2	chromosome_4	2430687	2436859	+	1	2E-26	
Cr	Cre04.g222600.t1.2	chromosome_4	2586817	2590382	-	2	5.5E-16	
Cr	Cre05.g238600.t1.2	chromosome_5	3029411	3032735	-	2	1.3E-42	
Cr	Cre05.g238650.t1.1	chromosome_5	3034210	3039095	-	2	1.7E-43	phC5
Cr	Cre05.g238687.t1.1	chromosome_5	3039142	3044935	-	2	2.8E-41	phC17
Cr	Cre06.g272250.t1.1	chromosome_6	2891774	2896177	-	4	3.8E-17	
Cr	Cre06.g278162.t1.1	chromosome_6	3784323	3787464	-	2	2.8E-28	phC3
Cr	Cre06.g290676.t1.1	chromosome_6	6215092	6233523	+	1	2.8E-26	
Cr	Cre06.g292249.t1.1	chromosome_6	6377397	6410484	+	1	1.7E-27	
Cr	Cre06.g294250.t1.2	chromosome_6	6677839	6683140	-	1	1.3E-28	
Cr	Cre06.g299150.t1.1	chromosome_6	7361752	7368597	-	2	1.7E-56	phC9
Cr	Cre06.g301806.t1.2	chromosome_6	7689350	7691791	+	1	4.7E-10	
Cr	Cre07.g319452.t1.1	chromosome_7	1021604	1027325	-	1	3.4E-16	
Cr	Cre07.g330750.t1.1	chromosome_7	2665869	2670182	-	2	2.4E-33	phC30
Cr	Cre07.g346698.t1.1	chromosome_7	4961036	4969687	+	1	1E-09	
Cr	Cre08.g368800.t1.2	chromosome_8	2102110	2105235	-	1	4.3E-07	
Cr	Cre09.g404201.t1.1	chromosome_9	495000	503449	-	2	1.1E-52	phC10
Cr	Cre09.g396100.t1.2	chromosome_9	1809519	1811468	+	1	5.6E-32	phC15
Cr	Cre09.g388351.t1.1	chromosome_9	2893511	2896174	+	2	7.7E-45	phC20c
Cr	Cre09.g388353.t1.1	chromosome_9	2901591	2904798	-	2	7.9E-45	phC20b
Cr	Cre09.g388355.t1.1	chromosome_9	2905893	2909768	-	2	9E-45	phC20a
Cr	Cre09.g409901.t1.1	chromosome_9	6848947	6852664	-	1	1.7E-37	phC29
Cr	Cre10.g420200.t1.2	chromosome_10	319809	324313	-	2	9E-16	
Cr	Cre10.g451752.t1.1	chromosome_10	4481576	4488961	-	2	9.3E-27	phC14
Cr	Cre11.g468359.t1.1	chromosome_11	1966744	1970333	-	2	1.8E-33	GAS31
Cr	Cre11.g468362.t1.1	chromosome_11	1974304	1979395	+	1	1.8E-13	
Cr	Cre11.g468800.t1.2	chromosome_11	2067870	2070655	-	1	2.5E-12	
Cr	Cre11.g480400.t1.1	chromosome_11	3348540	3360615	-	2	1.5E-27	phC27a
Cr	Cre11.g480451.t1.1	chromosome_11	3363251	3377643	+	2	2.8E-27	phC27b
Cr	Cre11.g481600.t1.1	chromosome_11	3554838	3557651	+	2	2.2E-42	GAS28
Cr	Cre11.g481750.t1.2	chromosome_11	3571847	3575932	-	2	1.3E-41	GAS30
Cr	Cre11.g483250.t1.1	chromosome_11	3779247	3784383	+	2	6.5E-21	
Cr	Cre11.g483351.t1.1	chromosome_11	3788767	3794267	+	2	5.4E-22	
Cr	Cre12.g506750.t1.1	chromosome_12	2463822	2471857	-	2	4.4E-36	
Cr	Cre12.g496500.t1.1	chromosome_12	3428360	3432421	+	1	8.2E-20	

Table 4-6: Continued.

Cr	Cre12.g552300.t1.2	chromosome_12	7944759	7951496	+	1	1.1E-22	
Cr	Cre12.g549000.t1.2	chromosome_12	8385044	8387899	+	2	6.9E-43	phC4
Cr	Cre12.g546800.t1.1	chromosome_12	8642192	8647925	-	1	7.8E-27	
Cr	Cre12.g551552.t1.1	chromosome_12	9625548	9630819	+	1	6.2E-27	phC26
Cr	Cre12.g551977.t1.1	chromosome_12	9631660	9635847	-	1	3.2E-13	
Cr	Cre14.g610700.t1.1	chromosome_14	422816	426421	-	1	4.8E-10	
Cr	Cre14.g612633.t1.1	chromosome_14	690473	728895	-	1	7.9E-29	
Cr	Cre14.g620600.t1.2	chromosome_14	1865969	1869014	-	2	3.8E-56	phC2
Cr	Cre14.g620702.t1.1	chromosome_14	1873369	1877024	-	2	7.7E-56	phC13
Cr	Cre16.g654600.t1.1	chromosome_16	1662676	1678650	+	1	1.5E-27	
Cr	Cre16.g654700.t1.1	chromosome_16	1679930	1691825	-	1	1.5E-26	
Cr	Cre17.g696500.t1.1	chromosome_17	49581	52364	-	2	3.1E-41	phC19
Cr	Cre17.g696700.t1.2	chromosome_17	66834	69000	+	2	7.4E-41	phC22
Cr	Cre17.g699750.t1.1	chromosome_17	558187	563756	+	1	5.3E-26	
Cr	Cre17.g699800.t1.1	chromosome_17	567710	570561	+	1	5.2E-23	
Cr	Cre17.g705300.t1.2	chromosome_17	1252305	1254898	-	1	3.4E-23	
Cr	Cre17.g705500.t1.1	chromosome_17	1275409	1279411	+	3	2.3E-35	phC32
Cr	Cre17.g710300.t1.1	chromosome_17	1905236	1908962	+	2	1.1E-43	phC28
Cr	Cre17.g717850.t1.1	chromosome_17	2685191	2692102	+	2	1.6E-52	phC8
Cr	Cre17.g717900.t1.2	chromosome_17	2693361	2696479	+	2	1.8E-55	phC1
Cr	Cre17.g717950.t1.2	chromosome_17	2698091	2702210	-	2	1.7E-55	phC6
Cr	Cre17.g718000.t1.1	chromosome_17	2702311	2708092	+	2	1.5E-54	phC7/phC23
Cr	Cre17.g729101.t1.1	chromosome_17	4039840	4049045	-	2	7.7E-27	phC31
Cr	Cre24.g755997.t1.1	scaffold_24	118246	122121	+	2	9.5E-36	phC18
Eu	Eu_g8584	scaffold3	828733	832315	-	1	5.2E-35	
Eu	Eu_g8586	scaffold3	840342	844257	-	1	9.7E-46	
Eu	Eu_g8605	scaffold3	1026997	1028199	-	1	4.5E-13	
Eu	Eu_g10876	scaffold8	388586	392062	-	2	6.3E-49	
Eu	Eu_g13506	scaffold9	333055	344069	-	1	1.9E-23	
Eu	Eu_g8663	scaffold11	394472	405382	+	2	1.4E-50	
Eu	Eu_g8664	scaffold11	410459	414596	+	2	2.4E-49	
Eu	Eu_g15282	scaffold17	438523	442928	-	2	6.6E-33	
Eu	Eu_g15283	scaffold17	450579	456018	-	2	7.8E-32	
Eu	Eu_g3332	scaffold20	382537	385250	-	2	7.9E-24	
Eu	Eu_g3334	scaffold20	394861	396568	-	1	3.6E-31	
Eu	Eu_g3336	scaffold20	406060	407842	-	1	1E-29	
Eu	Eu_g3349	scaffold20	485511	489085	+	2	9.9E-34	
Eu	Eu_g3350	scaffold20	493711	496123	+	2	6E-34	
Eu	Eu_g3351	scaffold20	499059	502448	+	2	2.6E-36	
Eu	Eu_g3391	scaffold20	830345	836665	+	2	4E-24	
Eu	Eu_g12945	scaffold25	39890	44577	-	2	1.1E-55	
Eu	Eu_g12970	scaffold25	338220	338603	-	1	2.3E-11	
Eu	Eu_g12855	scaffold26	60378	62278	-	1	1.9E-07	
Eu	Eu_g16712	scaffold29	115030	118337	+	1	3.1E-23	
Eu	Eu_g16714	scaffold29	123211	126512	+	1	2.6E-23	
Eu	Eu_g16717	scaffold29	134218	134884	+	1	1.2E-13	
Eu	Eu_g16745	scaffold29	340013	350426	-	2	2.6E-38	
Eu	Eu_g822	scaffold30	268044	271877	-	2	1.2E-53	
Eu	Eu_g960	scaffold30	926405	927089	+	1	1.9E-27	
Eu	Eu_g962	scaffold30	929468	930173	+	1	5.7E-40	
Eu	Eu_g964	scaffold30	936021	936705	+	1	3E-27	
Eu	Eu_g966	scaffold30	938835	939539	+	1	5.7E-40	
Eu	Eu_g969	scaffold30	948236	948871	-	1	1.6E-26	
Eu	Eu_g971	scaffold30	950714	951873	-	1	6.2E-40	

Table 4-6: Continued.

Eu	Eu_g972	scaffold30	954142	954774	-	1	6.8E-40
Eu	Eu_g974	scaffold30	957135	957868	-	1	2.1E-27
Eu	Eu_g975	scaffold30	960135	960769	-	1	6.8E-40
Eu	Eu_g977	scaffold30	962695	971747	-	1	2.5E-31
Eu	Eu_g1095	scaffold30	1641459	1643369	-	2	5.1E-07
Eu	Eu_g1096	scaffold30	1645346	1647814	-	1	7.4E-18
Eu	Eu_g10982	scaffold40	216103	219061	+	2	5E-43
Eu	Eu_g10984	scaffold40	224910	227649	+	2	4.9E-43
Eu	Eu_g10988	scaffold40	245151	248495	-	2	7.2E-54
Eu	Eu_g10989	scaffold40	256063	261594	-	2	9.4E-53
Eu	Eu_g10991	scaffold40	274322	277539	-	1	2.7E-45
Eu	Eu_g10993	scaffold40	279851	280684	-	1	3.4E-12
Eu	Eu_g10994	scaffold40	284783	289942	-	2	6.4E-48
Eu	Eu_g10995	scaffold40	293485	296415	-	1	1.7E-40
Eu	Eu_g10997	scaffold40	297908	299242	-	1	8E-15
Eu	Eu_g10999	scaffold40	305640	313544	-	2	7.9E-49
Eu	Eu_g11000	scaffold40	322029	326398	-	2	3.4E-49
Eu	Eu_g11001	scaffold40	329507	332890	-	2	5.3E-41
Eu	Eu_g11039	scaffold40	537862	543420	+	1	1.4E-24
Eu	Eu_g11044	scaffold40	552562	553469	+	1	4.8E-18
Eu	Eu_g11082	scaffold40	791786	793479	-	1	9.3E-23
Eu	Eu_g11083	scaffold40	794909	796339	+	1	1.7E-29
Eu	Eu_g19752	scaffold42	83076	85556	-	1	2.3E-43
Eu	Eu_g19753	scaffold42	85607	87110	-	1	9E-32
Eu	Eu_g19757	scaffold42	113895	117230	-	2	9.5E-43
Eu	Eu_g19758	scaffold42	124305	127895	-	2	1.6E-41
Eu	Eu_g19760	scaffold42	133407	143822	-	4	5.6E-44
Eu	Eu_g19761	scaffold42	146992	153188	-	2	5.1E-41
Eu	Eu_g4077	scaffold47	88172	90953	-	1	1.1E-13
Eu	Eu_g4177	scaffold47	593583	598865	-	2	1.1E-46
Eu	Eu_g4178	scaffold47	604727	608288	-	2	1.2E-43
Eu	Eu_g4179	scaffold47	612581	616535	-	2	1.4E-49
Eu	Eu_g4185_g4184	scaffold47	626904	629358	-	1	5.8E-28
Eu	Eu_g4187	scaffold47	631271	634351	-	1	6.5E-49
Eu	Eu_g4274	scaffold47	1188978	1191499	-	3	5.5E-18
Eu	Eu_g17309	scaffold62	254914	260268	+	2	4.2E-43
Eu	Eu_g11290	scaffold66	432063	437147	+	2	1.1E-52
Eu	Eu_g11293	scaffold66	448331	452808	-	2	1E-47
Eu	Eu_g11294_g11295	scaffold66	461057	469504	+	2	6E-38
Eu	Eu_g20508	scaffold71	76022	81463	-	1	2.2E-22
Eu	Eu_g20870	scaffold81	43272	43736	-	1	1.7E-11
Eu	Eu_g184	scaffold96	1190131	1192153	+	1	1.3E-18
Eu	Eu_g253	scaffold96	1651937	1654258	+	3	3.2E-35
Eu	Eu_g297	scaffold96	1961499	1971755	+	1	1.4E-08
Eu	Eu_g325	scaffold96	2157402	2159381	+	2	3.3E-26
Eu	Eu_g6299	scaffold97	1181226	1183784	-	2	1E-20
Eu	Eu_g8071	scaffold106	495651	501042	-	2	8.1E-43
Eu	Eu_g11698	scaffold111	717375	721219	-	1	5.5E-21
Eu	Eu_g11700	scaffold111	725239	729939	-	1	9.2E-22
Eu	Eu_g11701	scaffold111	734073	735455	-	1	2.3E-21
Eu	Eu_g11705	scaffold111	741608	747415	+	1	4.5E-24
Eu	Eu_g1286	scaffold120	437284	438951	+	1	9.2E-14
Eu	Eu_g1287	scaffold120	442650	443784	+	1	2E-15
Eu	Eu_g1543	scaffold120	2064558	2068997	+	1	2.1E-19

Table 4-6: Continued.

Eu	Eu_g20804	scaffold142	43128	59994	-	2	4.7E-30
Eu	Eu_g4746	scaffold1009	1010430	1015532	+	1	3.7E-25
Eu	Eu_g4825	scaffold1009	1529662	1530429	+	1	3.6E-17
Eu	Eu_g4827	scaffold1009	1535700	1537051	+	1	7E-24
Eu	Eu_g4828	scaffold1009	1540670	1545959	+	2	8.5E-25
Eu	Eu_g1971	scaffold1010	193991	195046	-	1	1E-36
Eu	Eu_g1973	scaffold1010	200115	200733	-	1	1.6E-16
Eu	Eu_g1990	scaffold1010	296975	301229	+	2	2.2E-37
Eu	Eu_g1991	scaffold1010	305055	308854	+	2	1.4E-41
Eu	Eu_g737	scaffold1011	2270540	2272315	-	1	4.5E-27
Eu	Eu_g8255	scaffold1012	652600	666728	+	1	3.5E-23
Eu	Eu_g8299	scaffold1012	925291	931197	-	2	2.7E-41
Eu	Eu_g7087	scaffold1014	1021745	1027240	+	2	2.5E-35
Eu	Eu_g7089	scaffold1014	1032336	1033237	+	1	6E-10
Eu	Eu_g7091	scaffold1014	1035565	1037789	+	1	8.1E-38
Eu	Eu_g7098	scaffold1014	1064064	1067751	+	1	5.7E-08
Eu	Eu_g7099	scaffold1014	1070962	1075798	+	2	1.5E-26
Eu	Eu_g7109	scaffold1014	1137697	1141147	+	1	5.1E-31
Eu	Eu_g7112	scaffold1014	1146566	1158984	+	1	1.9E-24
Eu	Eu_g7115	scaffold1014	1166959	1167345	+	1	6.1E-14
Eu	Eu_g7118	scaffold1014	1173783	1174338	+	1	5.6E-17
Eu	Eu_g7120	scaffold1014	1192069	1193704	+	2	3.1E-38
Eu	Eu_g7123	scaffold1014	1199602	1199949	+	1	1.3E-23
Eu	Eu_g7124	scaffold1014	1205105	1206959	+	2	2.3E-31
Eu	Eu_g11521	scaffold1022	390995	403689	+	1	5.7E-22
Eu	Eu_g17954	scaffold1033	153738	159706	-	1	2.5E-42
Eu	Eu_g17956	scaffold1033	165086	168606	-	2	6.6E-42
Eu	Eu_g17957	scaffold1033	171005	173251	-	1	2.5E-37
Eu	Eu_g17959	scaffold1033	174751	175811	-	1	7.3E-15
Eu	Eu_g17960	scaffold1033	181244	184713	-	2	8E-52
Eu	Eu_g17806	scaffold1035	2281	3072	+	2	6.4E-30
Eu	Eu_g17807	scaffold1035	11458	13342	+	2	8.4E-35
Eu	Eu_g17809	scaffold1035	17666	19563	+	2	5.9E-33
Eu	Eu_g17810	scaffold1035	26349	27761	+	2	1.2E-35
Eu	Eu_g17812	scaffold1035	44692	45558	-	1	9.6E-21
Eu	Eu_g17822	scaffold1035	99127	99531	-	1	4.4E-13
Eu	Eu_g9796	scaffold1036	195986	196903	+	1	4.7E-16
Eu	Eu_g9799	scaffold1036	218521	223392	-	2	2.2E-47
Eu	Eu_g9800	scaffold1036	226772	229973	-	2	1.3E-45
Eu	Eu_g9801	scaffold1036	233554	236771	+	2	7.7E-50
Eu	Eu_g9874	scaffold1036	803464	808581	-	2	8E-45
Eu	Eu_g9875	scaffold1036	811617	816882	-	2	4.8E-47
Eu	Eu_g20083	scaffold1043	128159	132173	+	2	4.7E-26
Eu	Eu_g20174	scaffold1047	80750	82386	-	1	1.1E-20
Eu	Eu_g20178	scaffold1047	99942	105605	-	1	5.9E-22
Eu	Eu_g15376	scaffold1123	146662	162544	-	2	6.7E-28
Eu	Eu_g2456	scaffold1985	818218	819741	+	1	1E-23
Eu	Eu_g2458	scaffold1985	821827	823001	+	1	1.2E-20
Eu	Eu_g2851	scaffold1986	1394567	1405706	-	3	3.9E-25
Eu	Eu_g10479	scaffold1987	772810	775265	-	1	5.2E-10
Eu	Eu_g12612	scaffold1989	593588	596870	+	1	7.4E-12
Eu	Eu_g13838	scaffold1990	435248	437555	-	1	6.2E-50
Eu	Eu_g13840	scaffold1990	442185	444649	-	1	9.9E-51
Eu	Eu_g17567	scaffold1993	92649	93080	+	1	1.6E-14

Table 4-6: Continued.

Eu	Eu_g18498	scaffold1996	149039	152849	+	2	1.4E-48	
Eu	Eu_g14290	scaffold1997	388429	389491	-	1	1.2E-16	
Eu	Eu_g14292	scaffold1997	393536	396827	-	1	1.3E-17	
Eu	Eu_g19296	scaffold2002	1	2545	-	1	1.9E-30	
Eu	Eu_g19307	scaffold2002	73012	76424	+	1	1E-26	
Eu	Eu_g19310	scaffold2002	81444	84896	+	1	3.3E-14	
Eu	Eu_g19655	scaffold2015	54816	56210	+	1	8.3E-31	
Eu	Eu_g19658	scaffold2015	58006	58377	+	1	1.4E-28	
Eu	Eu_g19664	scaffold2015	94320	99048	+	1	3.6E-20	
Eu	Eu_g3622	scaffold2050	632891	634202	+	1	4.4E-08	
Eu	Eu_g3624	scaffold2050	636550	638807	+	1	8.9E-34	
Eu	Eu_g3665	scaffold2050	912386	915187	-	1	5.8E-10	
Eu	Eu_g26047_g26046	scaffold2084	1	5752	-	1	2.8E-17	
Eu	Eu_g26048	scaffold2084	12053	14103	-	1	6.6E-25	
Eu	Eu_g3153	scaffold2943	1183172	1184878	+	1	2.3E-18	
Eu	Eu_g5984	scaffold2945	407209	409738	-	2	6.4E-39	
Eu	Eu_g13226	scaffold2948	535649	537582	-	1	3.1E-06	
Eu	Eu_g13227	scaffold2948	539989	541610	-	1	5.9E-07	
Eu	Eu_g15198	scaffold2951	489171	499634	-	2	1.1E-07	
Eu	Eu_g16384	scaffold2958	220541	223268	+	3	5.1E-34	
Eu	Eu_g18824	scaffold2965	174	3378	+	1	1.9E-21	
Eu	Eu_g18830	scaffold2965	42469	46210	+	1	2.6E-08	
Eu	Eu_g18833	scaffold2965	66717	68718	+	1	4.6E-24	
Eu	Eu_g19585	scaffold2978	116607	119749	+	2	6.5E-41	
Eu	Eu_g19587	scaffold2978	130698	131554	+	1	6.2E-17	
Eu	Eu_g19589	scaffold2978	132993	133502	+	1	1.7E-41	
Eu	Eu_g21069	scaffold2984	47801	49379	-	1	3.5E-26	
Eu	Eu_g15693	scaffold3057	286477	287627	+	1	1.2E-36	
Eu	Eu_g15695	scaffold3057	293252	294418	-	1	1.1E-36	
Eu	Eu_g15697	scaffold3057	295768	296935	-	1	1.9E-44	
Eu	Eu_g15698	scaffold3057	299002	300168	-	1	1.1E-36	
Eu	Eu_g15700	scaffold3057	305605	306756	-	1	1.1E-36	
Eu	Eu_g7508	scaffold3899	1152683	1153030	+	1	4.2E-24	
Eu	Eu_g7510	scaffold3899	1165230	1171481	+	2	3.9E-23	
Eu	Eu_g10675	scaffold3900	45796	46565	-	1	2.5E-20	
Eu	Eu_g10693	scaffold3900	110905	111252	-	1	6.6E-11	
Eu	Eu_g10713	scaffold3900	269772	270540	+	1	8E-21	
Eu	Eu_g10813	scaffold3900	920025	921617	-	1	2.5E-39	
Eu	Eu_g10817	scaffold3900	930828	932103	-	1	2.4E-25	
Eu	Eu_g10260	scaffold3902	313271	317384	+	2	4.8E-15	
Eu	Eu_g10268	scaffold3902	405939	411819	+	1	2.3E-08	
Eu	Eu_g14796	scaffold3905	65087	68423	+	2	8.7E-46	
Eu	Eu_g14797	scaffold3905	72423	75811	+	2	9.9E-51	
Eu	Eu_g18199	scaffold3909	82016	86131	-	2	1.3E-44	
Eu	Eu_g18398	scaffold3910	261883	266261	-	2	2.4E-42	
Eu	Eu_g28347	scaffold4144	3087	5633	+	1	3.9E-46	
Gp	phG6(manual)	scaffold00001	25300	30452	+	2	5E-52	phG6
Gp	phG23(manual)	scaffold00001	128470	130082	+	1	6.9E-51	phG23
Gp	GPECTOR_1g48	scaffold00001	462857	464815	+	1	1.6E-09	
Gp	GPECTOR_1g810	scaffold00001	5558137	5558900	+	1	2.8E-09	
Gp	GPECTOR_1g900	scaffold00001	6153365	6154899	-	1	9.8E-10	
Gp	GPECTOR_4g647	scaffold00004	931982	934890	+	2	3.7E-17	
Gp	GPECTOR_5phG20	scaffold00005	2596599	2599412	-	2	9.8E-44	phG20
Gp	GPECTOR_6g468	scaffold00006	6602	15353	+	2	6E-25	

Table 4-6: Continued.

Gp	GPECTOR_6g470	scaffold00006	33498	34877	+	1	6.2E-22	
Gp	GPECTOR_6phG18	scaffold00006	3143265	3146432	-	2	2.1E-34	phG18
Gp	GPECTOR_7g967	scaffold00007	502699	507905	-	1	1.2E-11	
Gp	GPECTOR_7g1021	scaffold00007	933893	936745	-	1	2.1E-21	
Gp	GPECTOR_7phG4	scaffold00007	1747691	1749613	-	2	1E-42	phG4
Gp	GPECTOR_7g1334	scaffold00007	3253502	3256307	+	1	2.7E-16	
Gp	GPECTOR_7phG26	scaffold00007	3363379	3366573	-	1	2.2E-27	phG26
Gp	GPECTOR_9g650	scaffold00009	2311190	2315353	+	1	2.9E-28	
Gp	GPECTOR_10phG3	scaffold00010	1191767	1194208	+	2	4.7E-30	phG3
Gp	GPECTOR_10g1087	scaffold00010	2379863	2382562	-	3	7.9E-14	
Gp	GPECTOR_11g297	scaffold00011	2297814	2299862	-	1	1.9E-21	
Gp	GPECTOR_14phG21	scaffold00014	1258538	1262072	+	3	5.2E-35	phG21
Gp	GPECTOR_14phG2a	scaffold00014	1387241	1391918	+	2	2.1E-52	phG2a
Gp	GPECTOR_14phG2b	scaffold00014	1394959	1399935	-	2	1.1E-55	phG2b
Gp	phG8(manual)	scaffold00014	1813170	1815624	+	2	2.9E-43	phG8
Gp	phG9(manual)	scaffold00014	1823134	1825277	-	2	1E-43	phG9
Gp	GPECTOR_14g266	scaffold00014	2018167	2021795	+	1	3.3E-08	
Gp	GPECTOR_14g267	scaffold00014	2024167	2025556	+	1	5.7E-27	
Gp	GPECTOR_15g322	scaffold00015	383803	388050	-	2	5.8E-23	
Gp	GPECTOR_15g358	scaffold00015	708816	711416	+	1	2.9E-16	
Gp	GPECTOR_15phG17	scaffold00015	1851755	1858839	+	2	2E-32	phG17
Gp	GPECTOR_16phG13	scaffold00016	1392853	1395606	-	2	4.3E-32	phG13
Gp	GPECTOR_18g141	scaffold00018	1181909	1183778	+	1	2.6E-20	
Gp	GPECTOR_18phG15	scaffold00018	1185779	1187130	+	1	5.5E-31	phG15
Gp	GPECTOR_20phG10	scaffold00020	345729	348190	-	2	6.3E-36	phG10
Gp	GPECTOR_21phG11	scaffold00021	1368438	1372827	-	2	1.6E-34	phG11
Gp	GPECTOR_28phG5a	scaffold00028	1101217	1104525	-	2	3.9E-57	phG5a
Gp	phG5b(manual)	scaffold00028	1117505	1120386	-	2	2E-56	phG5b
Gp	GPECTOR_31phG1	scaffold00031	732596	734322	-	1	3.1E-25	phG1
Gp	GPECTOR_31g362	scaffold00031	735525	737922	-	1	3.1E-25	
Gp	GPECTOR_31g405	scaffold00031	1308627	1311718	+	1	2.4E-12	
Gp	GPECTOR_33g554	scaffold00033	166725	170785	-	1	2.9E-23	
Gp	GPECTOR_42phG14	scaffold00042	178546	180779	+	1	8.9E-17	phG14
Gp	GPECTOR_42GAS30	scaffold00042	310005	311868	+	2	1.5E-36	GAS30
Gp	GPECTOR_48phG19	scaffold00048	1312056	1314370	+	2	1.1E-39	phG19
Gp	GPECTOR_57g449	scaffold00057	30827	45528	+	1	4.5E-16	
Gp	GPECTOR_60phG7	scaffold00060	336548	340285	-	1	5.2E-29	phG7
Gp	GPECTOR_64g89	scaffold00064	77422	80632	-	1	2.5E-08	
Gp	GpGAS31(manual)	scaffold00068	485900	488677	+	2	1.3E-31	GAS31
Gp	phG22(manual)	scaffold00090	261	4184	+	1	1.5E-26	phG22
Gp	GPECTOR_230g516	scaffold00230	948	5587	-	1	3.2E-24	
Gp	phG12(manual)	scaffold00280: scaffold00018	sc18:274048	sc280:2554	-	1	6.8E-34	phG12
Gp	GPECTOR_506g472	scaffold00506	6483	11131	+	1	9E-24	
Gp	GPECTOR_590phG16	scaffold00590	7420	11326	+	2	1.1E-33	phG16
Gp	phG25(manual)	scaffold00695	1	1917	+	1	1.3E-47	phG25
Gp	GpGAS28(manual)	scaffold00860: scaffold00042	sc42:331225	sc860:1112	-	1	5.1E-32	GAS28
Gp	phG24(manual)	scaffold01216	5240	6451	+	1	4.5E-48	phG24
Ts	TSOC_000066	scaffold_1	499671	502484	-	2	2.2E-39	
Ts	TSOC_000087	scaffold_1	670058	677097	-	1	8.9E-38	
Ts	TSOC_000577	scaffold_8	233329	237928	-	1	4.1E-12	
Ts	TSOC_000970	scaffold_15	5647	15814	+	2	5.2E-40	
Ts	TSOC_001818	scaffold_30	277755	284171	+	1	6.2E-20	

Table 4-6: Continued.

Ts	TSOC_001878	scaffold_33	86572	94050	+	1	2.5E-30
Ts	TSOC_001881	scaffold_33	105904	116012	+	1	5.6E-23
Ts	TSOC_002146	scaffold_39	289173	292763	-	1	2.8E-30
Ts	TSOC_002196	scaffold_41	78702	79787	-	2	1.1E-18
Ts	TSOC_002944	scaffold_59	256859	258712	-	1	1.5E-21
Ts	TSOC_003208	scaffold_67	91066	92696	+	1	2.8E-30
Ts	TSOC_003594	scaffold_79	150678	153567	-	2	9E-27
Ts	TSOC_004158	scaffold_98	138022	141280	+	2	1.1E-40
Ts	TSOC_004162	scaffold_98	150986	159667	-	2	7.9E-39
Ts	TSOC_004172	scaffold_99	73695	78706	+	1	2.8E-22
Ts	TSOC_004307	scaffold_104	105290	111810	-	1	4.4E-13
Ts	TSOC_004360	scaffold_107	136983	143582	+	1	3.8E-20
Ts	TSOC_004714	scaffold_118	213959	224354	-	1	1.8E-14
Ts	TSOC_004991	scaffold_130	17967	23637	+	1	2.1E-20
Ts	TSOC_005497	scaffold_150	185427	188719	-	1	8.8E-12
Ts	TSOC_005734	scaffold_162	2	1280	-	1	2.6E-40
Ts	TSOC_005871	scaffold_166	156775	159915	-	2	2.4E-42
Ts	TSOC_005912	scaffold_171	39986	40585	-	1	3.7E-21
Ts	TSOC_005914	scaffold_171	47147	52809	+	1	1.9E-22
Ts	TSOC_006039	scaffold_177	36318	37759	-	1	1.3E-07
Ts	TSOC_006150	scaffold_183	9792	11618	+	1	4.9E-20
Ts	TSOC_006151	scaffold_183	14997	16823	+	1	6.9E-20
Ts	TSOC_006152	scaffold_183	17879	20746	+	1	2.7E-22
Ts	TSOC_006153	scaffold_183	31558	33634	+	1	2.8E-22
Ts	TSOC_006170	scaffold_183	158458	163655	-	1	2.5E-21
Ts	TSOC_006171	scaffold_183	165995	166979	-	1	1.9E-23
Ts	TSOC_006172	scaffold_183	168376	170290	-	1	1.6E-09
Ts	TSOC_006496	scaffold_199	93665	97020	-	2	9.8E-41
Ts	TSOC_007379	scaffold_247	2	6878	-	4	6.1E-16
Ts	TSOC_007774	scaffold_271	55344	57446	-	1	1.9E-23
Ts	TSOC_007800	scaffold_272	138738	140155	-	1	1.5E-22
Ts	TSOC_008322	scaffold_307	45243	48095	-	2	7.1E-33
Ts	TSOC_008536	scaffold_323	27294	29125	-	1	9.4E-32
Ts	TSOC_008957	scaffold_350	241	1626	+	1	1.3E-30
Ts	TSOC_009048	scaffold_364	92601	97654	+	2	6.1E-28
Ts	TSOC_009114	scaffold_369	9959	12262	-	1	1.8E-23
Ts	TSOC_009167	scaffold_371	64640	69055	+	1	4.3E-09
Ts	TSOC_009258	scaffold_378	36177	39828	+	2	3E-52
Ts	TSOC_009772	scaffold_424	92901	94151	+	1	5.5E-15
Ts	TSOC_010325	scaffold_484	60381	63293	-	3	5.7E-32
Ts	TSOC_010595	scaffold_512	82	1910	+	1	4E-25
Ts	TSOC_011328	scaffold_611	49912	50794	-	1	7.6E-16
Ts	TSOC_011329	scaffold_611	52120	52473	-	1	8E-08
Ts	TSOC_011393	scaffold_622	1	2530	-	1	9.2E-23
Ts	TSOC_011580	scaffold_656	37500	45825	+	2	1.3E-37
Ts	TSOC_011723	scaffold_679	307	2912	+	1	2.5E-43
Ts	TSOC_011753	scaffold_687	44040	49239	-	1	2.2E-28
Ts	TSOC_011919	scaffold_717	14840	17916	+	2	5.6E-19
Ts	TSOC_012046	scaffold_741	2758	3768	-	1	1.6E-16
Ts	TSOC_012047	scaffold_741	5132	6742	-	1	4.6E-12
Ts	TSOC_012186	scaffold_751	39400	41218	-	1	1.1E-20
Ts	TSOC_012576	scaffold_869	4083	6444	-	1	2.1E-25
Ts	TSOC_012577	scaffold_869	17622	21266	-	1	1.4E-11
Ts	TSOC_012718	scaffold_908	6048	7997	+	1	5.4E-17

Table 4-6: Continued.

Ts	TSOC_013100	scaffold_1042	17662	18981	+	1	4.2E-39	
Ts	TSOC_013386	scaffold_1177	1772	8209	-	1	1.6E-33	
Ts	TSOC_013642	scaffold_1329	8528	12867	+	1	2.4E-11	
Ts	TSOC_013883	scaffold_1522	7234	10229	+	1	1.5E-29	
Ts	TSOC_014438	scaffold_2169	2112	3469	-	1	4.3E-31	
Ts	TSOC_014648	scaffold_2564	1597	2337	+	1	5.3E-44	
Ts	TSOC_014893	scaffold_3206	602	1724	-	1	7.7E-32	
Ts	TSOC_014918	scaffold_3326	31	1655	-	1	6E-24	
Ts	TSOC_014928	scaffold_3384	852	1613	+	1	3E-42	
Ts	TSOC_015044	scaffold_3788	40	411	-	1	5.4E-33	
Ts	TSOC_015053	scaffold_3821	510	1188	-	1	1.7E-41	
Ts	TSOC_015393	scaffold_5270	812	1030	-	1	3.8E-15	
Vc	g89093	scaffold_1	136	4295	-	1	2.3E-16	
Vc	g89111	scaffold_1	191521	195209	+	1	3.3E-15	
Vc	Vocar.0001s0294.1.p	scaffold_1	2151796	2154844	-	1	1.4E-35	
Vc	g101495	scaffold_1	2159334	2159939	-	1	3.8E-11	
Vc	Vocar.0001s0295.1.p	scaffold_1	2183501	2188794	+	2	1.6E-45	phV53
Vc	Vocar.0001s0296.1.p	scaffold_1	2190643	2193280	-	1	9.2E-46	
Vc	Vocar.0001s0297.1.p	scaffold_1	2194869	2197708	-	1	4.7E-44	
Vc	Vocar.0001s0298.1.p	scaffold_1	2225937	2230078	+	2	2E-44	phV54
Vc	g107992	scaffold_1	4677999	4705726	+	2	5.6E-21	
Vc	g102804	scaffold_2	2917567	2918633	+	1	5.1E-47	
Vc	Vocar.0002s0417.1.p	scaffold_2	3174180	3177629	+	3	1.4E-19	
Vc	Vocar.0002s0430.1.p	scaffold_2	3268309	3276191	+	2	9.9E-52	phV38
Vc	Vocar.0002s0564.1.p	scaffold_2	4191410	4198020	+	2	7.1E-30	ssgA/ssg185
Vc	Vocar20012572m	scaffold_3	240421	242399	+	1	1.1E-47	
Vc	Vocar.0003s0019.1.p	scaffold_3	258170	261328	-	1	1E-49	phV40b
Vc	Vocar.0003s0020.1.p	scaffold_3	267953	272478	-	2	1.3E-52	phV40a
Vc	g106282	scaffold_3	275937	285359	-	2	2.9E-51	phV40c
Vc	Vocar.0003s0021.1.p	scaffold_3	288889	291448	-	2	2.6E-51	phV40d
Vc	Vocar.0003s0023.1.p	scaffold_3	297570	303150	+	2	2.8E-53	phV40e
Vc	Vocar.0003s0024.1.p	scaffold_3	305358	311084	+	2	1.6E-53	phV40f
Vc	Vocar.0003s0025.1.p	scaffold_3	330185	332910	-	2	1.7E-54	phV40g
Vc	Vocar.0003s0026.1.p	scaffold_3	335846	339789	-	2	2E-54	phV40h
Vc	g98863	scaffold_3	3208753	3218820	-	1	5.2E-08	
Vc	Vocar.0003s0325.1.p	scaffold_3	3252727	3258505	+	1	4.6E-06	
Vc	Vocar20010554m	scaffold_4	272928	278142	+	1	5.6E-16	
Vc	Vocar.0004s0106.1.p	scaffold_4	851046	858205	-	2	3E-45	phV45
Vc	g80812	scaffold_4	909128	914816	+	2	1.6E-50	phV14
Vc	g90188	scaffold_4	921342	936215	+	2	4.9E-36	phDZ2
Vc	g104453	scaffold_4	944174	951729	+	1	1E-38	phS
Vc	g59555	scaffold_4	954674	963100	+	2	1.1E-50	phV12
Vc	Vocar.0004s0116.1.p	scaffold_4	971082	976053	+	2	8.7E-51	phV11
Vc	Vocar.0004s0117.1.p	scaffold_4	987729	994364	+	2	1.8E-51	phV10
Vc	Vocar20010621m	scaffold_4	1884312	1888361	+	3	2.4E-34	phV32
Vc	Vocar.0004s0253.1.p	scaffold_4	2018925	2023516	+	2	2.2E-53	phV55
Vc	Vocar.0004s0255.1.p	scaffold_4	2026661	2032058	-	2	3.8E-52	phI
Vc	Vocar.0004s0257.1.p	scaffold_4	2037429	2045105	+	2	1E-52	phV15
Vc	g107649	scaffold_4	3085416	3111447	+	2	5.7E-29	phV34
Vc	Vocar.0004s0371.1.p	scaffold_4	3122539	3134540	+	2	2.8E-23	
Vc	Vocar20010671m	scaffold_4	3138401	3143891	+	2	4.1E-25	
Vc	Vocar.0004s0372.1.p	scaffold_4	3147271	3153816	+	2	3.1E-25	
Vc	Vocar.0004s0374.1.p	scaffold_4	3158380	3164279	+	1	3.2E-09	
Vc	Vocar.0004s0376.1.p	scaffold_4	3169158	3173652	+	1	2.8E-10	

Table 4-6: Continued.

Vc	g98453	scaffold_4	3187958	3190741	+	1	2.4E-11	
Vc	Vocar.0005s0504.1.p	scaffold_5	4082284	4086243	+	2	1.7E-41	phV35
Vc	g105540	scaffold_6	975675	981748	+	1	1.1E-15	
Vc	Vocar.0007s0316.1.p	scaffold_7	2563819	2569918	-	1	4.8E-24	
Vc	Vocar.0007s0317.1.p	scaffold_7	2573754	2587564	-	1	1.1E-21	
Vc	Vocar.0008s0161.1.p	scaffold_8	1413441	1417467	-	2	1.4E-34	phV56
Vc	Vocar.20004596m	scaffold_8	1654086	1666382	+	1	6.2E-19	
Vc	g97603	scaffold_9	248770	264280	-	2	2.2E-19	
Vc	Vocar.0009s0095.1.p	scaffold_9	868830	877399	-	1	1.8E-22	
Vc	g100253	scaffold_9	886085	887843	-	1	2.8E-11	
Vc	Vocar.0009s0096.1.p	scaffold_9	893687	896323	-	1	1.5E-09	
Vc	Vocar.0009s0102.1.p	scaffold_9	993374	1000776	-	2	1.9E-27	
Vc	Vocar.0009s0117.1.p	scaffold_9	1139993	1151476	+	1	3.6E-23	
Vc	Vocar.20005232m	scaffold_9	1290523	1294917	+	2	6.2E-24	
Vc	g95705	scaffold_9	1329232	1350974	+	2	1.2E-24	
Vc	Vocar.0009s0136.1.p	scaffold_9	1402394	1415004	-	2	9.9E-26	
Vc	Vocar.0009s0297.1.p	scaffold_9	2885725	2888686	+	1	7.4E-06	
Vc	Vocar.0009s0298.1.p	scaffold_9	2908649	2913619	+	1	2.4E-06	
Vc	g87907	scaffold_9	2913422	2921440	+	1	5.8E-09	
Vc	Vocar.0012s0191.1.p	scaffold_12	1835801	1839551	-	1	6.7E-20	
Vc	g104151	scaffold_12	2002556	2005916	-	2	5E-20	phV31
Vc	Vocar.0012s0256.1.p	scaffold_12	2441709	2447890	-	2	4.9E-32	phV46
Vc	g110102	scaffold_13	976726	1058228	+	1	4.4E-49	phV3
Vc	Vocar.20008133m	scaffold_15	452053	455553	-	1	7.2E-08	
Vc	Vocar.0016s0049.1.p	scaffold_16	371485	374576	-	1	1.1E-23	
Vc	g88561	scaffold_16	1348222	1349697	+	1	1.6E-10	
Vc	g63512	scaffold_17	189721	194498	-	2	5.2E-53	phV8
Vc	g75691	scaffold_17	220440	227838	+	2	1.1E-50	phV9/sef5
Vc	Vocar.0017s0136.1.p	scaffold_17	1512268	1520996	+	1	5.2E-36	
Vc	Vocar.0017s0137.1.p	scaffold_17	1524215	1528618	+	1	3E-28	
Vc	Vocar.0018s0041.1.p	scaffold_18	341189	346978	+	2	2.3E-52	phV36
Vc	g89855	scaffold_18	532485	535306	+	1	6E-07	
Vc	Vocar.20000037m	scaffold_18	541453	541791	+	1	1.4E-10	
Vc	Vocar.0019s0064.1.p	scaffold_19	556389	563049	+	1	1.2E-19	
Vc	g104012	scaffold_20	501611	508392	+	1	1.1E-16	
Vc	Vocar.0021s0055.1.p	scaffold_21	881346	885660	-	2	8.2E-43	phV51
Vc	Vocar.0022s0224.1.p	scaffold_22	1990278	1996110	-	1	3E-25	
Vc	Vocar.0024s0151.1.p	scaffold_24	1115594	1122017	-	2	8.4E-51	phV23a
Vc	Vocar.0024s0152.1.p	scaffold_24	1126287	1131775	-	2	5.2E-51	phV23b
Vc	Vocar.0024s0216.1.p	scaffold_24	1596034	1601391	-	2	5.2E-46	phV22b
Vc	Vocar.0024s0217.1.p	scaffold_24	1606820	1610491	-	2	5.9E-49	phV22a
Vc	g91705	scaffold_25	25038	27606	+	1	1.6E-16	
Vc	Vocar.0026s0098.1.p	scaffold_26	1117390	1119819	-	1	6E-21	
Vc	Vocar.0026s0099.1.p	scaffold_26	1125392	1129134	-	1	4.7E-24	
Vc	Vocar.0027s0017.1.p	scaffold_27	139594	146217	-	1	5.2E-49	
Vc	Vocar.20014546m	scaffold_27	1535364	1538327	+	1	8.4E-21	
Vc	Vocar.0027s0164.1.p	scaffold_27	1542465	1547696	+	1	5.4E-33	sex2
Vc	Vocar.0027s0177.1.p	scaffold_27	1644945	1654145	+	2	1E-33	phV41
Vc	Vocar.0029s0006.1.p	scaffold_29	49517	54230	-	2	3.3E-53	phV25
Vc	g127208	scaffold_29	60615	61493	-	1	3.9E-41	phV39a
Vc	Vocar.20015019m	scaffold_29	66147	68849	-	2	5.2E-30	phV39b
Vc	g90439	scaffold_29	76431	80164	-	1	1.6E-23	phV24
Vc	g90431	scaffold_29	141804	146119	+	2	2.7E-50	phV21
Vc	Vocar.0029s0031.1.p	scaffold_29	320431	326473	-	2	2.6E-48	phV42

Table 4-6: Continued.

Vc	Vocar.0029s0041.1.p	scaffold_29	425310	432942	+	2	7.9E-54	phV20
Vc	g80872	scaffold_29	444655	446511	-	2	1E-45	phV18
Vc	Vocar.0030s0081.1.p	scaffold_30	717804	722763	+	2	4.2E-12	
Vc	Vocar20013339m	scaffold_31	256731	257051	+	1	1.6E-11	
Vc	g91736	scaffold_31	490052	494138	+	1	1.8E-22	
Vc	Vocar20013296m	scaffold_31	792397	795286	-	1	3E-07	
Vc	g91772	scaffold_31	800186	806925	-	1	6.9E-09	
Vc	Vocar20013362m	scaffold_31	1023899	1027504	-	2	8.2E-47	phV43
Vc	Vocar.0031s0086.1.p	scaffold_31	1104183	1109320	+	2	1.1E-49	phV44a
Vc	Vocar.0031s0087.1.p	scaffold_31	1111737	1115725	+	2	3.1E-49	phV44b
Vc	Vocar.0033s0047.1.p	scaffold_33	357473	362532	-	2	1.2E-14	
Vc	Vocar.0033s0108.1.p	scaffold_33	850814	856523	-	2	5.2E-55	phV52
Vc	Vocar.0033s0126.1.p	scaffold_33	969658	980827	-	2	8.2E-48	phV47
Vc	Vocar.0033s0128.1.p	scaffold_33	1001204	1003662	-	2	1.8E-44	phV48
Vc	Vocar.0033s0180.1.p	scaffold_33	1439273	1441872	+	2	1.3E-47	phV33c
Vc	Vocar20003633m	scaffold_33	1444403	1449596	+	2	7.9E-45	phV33b
Vc	Vocar.0033s0185.1.p	scaffold_33	1463693	1467353	-	2	3.8E-44	phV33a
Vc	Vocar.0034s0002.1.p	scaffold_34	8519	11800	+	2	2E-47	
Vc	g91290	scaffold_34	199034	207043	-	2	3.2E-27	
Vc	Vocar20004187m	scaffold_34	212576	217730	+	3	2.1E-16	
Vc	Vocar.0034s0061.1.p	scaffold_34	534023	538072	-	3	5.5E-32	
Vc	g91237	scaffold_34	643842	649756	+	1	3.2E-21	
Vc	g104851	scaffold_34	699532	707375	-	1	2E-09	
Vc	Vocar.0034s0075.1.p	scaffold_34	732748	736023	+	1	1.8E-20	
Vc	g107455	scaffold_34	951649	958042	+	2	2.6E-55	phV6
Vc	g83859	scaffold_34	961775	968594	+	2	5.6E-55	phIII
Vc	g77338	scaffold_34	974279	978515	-	2	2.1E-55	phV5
Vc	g127206	scaffold_34	981439	982774	-	1	5.5E-29	phV13a
Vc	g108791	scaffold_34	993874	995277	+	1	7.7E-56	phV7a
Vc	g85077	scaffold_34	1008329	1010109	+	1	2.6E-55	phII
Vc	g77335	scaffold_34	1015476	1019903	-	2	6.6E-55	phV4
Vc	Vocar.0034s0104.1.p	scaffold_34	1022025	1028496	+	2	1.1E-54	phV1
Vc	g83847	scaffold_34	1050022	1054050	+	1	5E-33	phV19/sef6
Vc	g98305	scaffold_35	837378	841599	-	1	3E-23	
Vc	Vocar.0036s0021.1.p	scaffold_36	130969	135967	+	2	8.5E-44	phV50
Vc	Vocar20003965m	scaffold_39	185552	185966	-	1	3.1E-19	
Vc	g104835	scaffold_41	743031	751739	+	1	8.3E-14	
Vc	Vocar.0041s0078.1.p	scaffold_41	785136	799337	+	2	2.3E-31	
Vc	Vocar.0043s0060.1.p	scaffold_43	632273	638314	+	2	2.7E-30	
Vc	g95614	scaffold_44	8959	16371	+	1	9.7E-44	
Vc	Vocar.0046s0015.1.p	scaffold_46	277817	281459	-	1	8.5E-24	
Vc	g98215	scaffold_48	280752	284732	+	1	3E-27	
Vc	Vocar.0048s0010.1.p	scaffold_48	396869	401014	+	2	9.3E-38	phV29
Vc	g67890	scaffold_48	403339	405270	+	2	2E-35	phV27
Vc	Vocar.0048s0012.1.p	scaffold_48	409032	413326	+	2	1.2E-33	phV26
Vc	Vocar.0048s0013.1.p	scaffold_48	417685	419806	+	2	4.1E-38	phV30
Vc	Vocar.0048s0016.1.p	scaffold_48	436341	439989	-	2	5.1E-34	phV28
Vc	g107539	scaffold_48	489059	497522	+	1	3E-10	
Vc	Vocar.0048s0025.1.p	scaffold_48	497750	505027	+	2	1.4E-24	
Vc	Vocar.0050s0044.1.p	scaffold_50	443615	445835	+	1	1.3E-19	
Vc	Vocar.0061s0002.1.p	scaffold_61	12154	15060	+	1	9.2E-25	
Vc	Vocar.0064s0013.1.p	scaffold_64	166389	174698	+	1	1.1E-14	
Vc	Vocar.0070s0021.1.p	scaffold_70	191077	196438	+	2	5.6E-23	
Vc	Vocar.0073s0006.1.p	scaffold_73	33007	39878	+	2	3.1E-46	phV2

Table 4-6: Continued.

Vc	g100178	scaffold_73	46056	51006	+	2	5.6E-46	phV17
Vc	g78132	scaffold_73	54503	64075	-	3	8.2E-51	phV16
Vc	Vocar.0083s0018.1.p	scaffold_83	147399	158606	-	2	6.6E-30	
Vc	g100727	scaffold_101	14487	18935	-	1	8.4E-15	
Vc	Vocar.0252s0001.1.p	scaffold_252	14	1511	+	1	4.1E-21	
Vc	Vocar20005313m	scaffold_929	1913	2648	-	1	3E-42	
Vc	Vocar20005827m	scaffold_1049	97	797	+	1	1.1E-10	
Yu	Yu_g1226	Scaffold0003	519125	525341	+	1	5E-15	
Yu	Yu_g1248	Scaffold0003	682264	684415	-	3	1.5E-35	
Yu	Yu_g1464	Scaffold0003	2079412	2083920	+	2	1.9E-24	
Yu	Yu_g1559	Scaffold0004	71241	73084	+	2	1.2E-42	
Yu	Yu_g1727	Scaffold0004	1036618	1039352	+	3	2.2E-32	
Yu	Yu_g1823	Scaffold0004	1754673	1758254	-	1	9.2E-23	
Yu	Yu_g1824	Scaffold0004	1763058	1766872	-	1	1.5E-24	
Yu	Yu_g1852	Scaffold0004	1935953	1937725	+	2	1.2E-39	
Yu	Yu_g1857	Scaffold0004	1962167	1963504	+	1	7.1E-17	
Yu	Yu_g1859	Scaffold0004	1964735	1965250	+	1	5.8E-44	
Yu	Yu_g1929	Scaffold0004	2375794	2376809	+	1	4.2E-46	
Yu	Yu_g1931	Scaffold0004	2378389	2380086	+	1	4.7E-53	
Yu	Yu_g1932	Scaffold0004	2391447	2394602	+	2	7.3E-56	
Yu	Yu_g1935	Scaffold0004	2406761	2409646	+	2	1.1E-52	
Yu	Yu_g1936	Scaffold0004	2411914	2416522	-	2	3.8E-54	
Yu	Yu_g2095	Scaffold0005	764983	768976	-	1	3.2E-11	
Yu	Yu_g2517	Scaffold0006	760077	762834	+	2	3.8E-48	
Yu	Yu_g2760	Scaffold0007	2	1589	+	1	3.1E-22	
Yu	Yu_g2761	Scaffold0007	8599	10541	+	1	3.5E-21	
Yu	Yu_g2763	Scaffold0007	16299	19067	+	1	1.9E-23	
Yu	Yu_g2766	Scaffold0007	30788	34727	+	2	1.5E-14	
Yu	Yu_g2768	Scaffold0007	40522	46115	+	1	5.7E-21	
Yu	Yu_g2769	Scaffold0007	51924	58074	+	2	3.5E-24	
Yu	Yu_g2830	Scaffold0007	388939	390525	+	1	1E-14	
Yu	Yu_g2938	Scaffold0007	1100931	1103915	-	2	4.1E-16	
Yu	Yu_g2939	Scaffold0007	1106074	1113053	-	2	3.1E-16	
Yu	Yu_g3354	Scaffold0008	1643094	1645152	-	1	1.1E-51	
Yu	Yu_g3356	Scaffold0008	1650125	1653677	-	1	2.7E-40	
Yu	Yu_g3508	Scaffold0009	567769	569382	-	1	3.3E-28	
Yu	Yu_g3780	Scaffold0010	571058	573916	+	2	5.4E-52	
Yu	Yu_g3781	Scaffold0010	576762	580450	+	2	2.8E-35	
Yu	Yu_g5062	Scaffold0015	805894	809244	+	2	7.7E-49	
Yu	Yu_g5162	Scaffold0016	58450	60275	+	1	9.4E-20	
Yu	Yu_g5164	Scaffold0016	61976	63056	+	1	6.2E-19	
Yu	Yu_g5165	Scaffold0016	66688	68743	+	1	6.1E-20	
Yu	Yu_g5167	Scaffold0016	69627	70592	+	1	1.3E-21	
Yu	Yu_g5169	Scaffold0016	76577	78624	+	1	9.2E-20	
Yu	Yu_g5171	Scaffold0016	79505	80254	+	1	8.8E-15	
Yu	Yu_g5826	Scaffold0018	1282709	1287395	-	1	7.9E-28	
Yu	Yu_g5843	Scaffold0019	99178	102924	-	2	2.6E-50	
Yu	Yu_g5844	Scaffold0019	105299	109304	-	2	2E-52	
Yu	Yu_g5845	Scaffold0019	113104	116956	-	2	8.8E-53	
Yu	Yu_g5912	Scaffold0019	508434	512010	+	2	3.3E-54	
Yu	Yu_g6069	Scaffold0020	104552	106281	-	1	1.7E-33	
Yu	Yu_g6071	Scaffold0020	110423	114390	-	1	2.8E-32	
Yu	Yu_g6399	Scaffold0021	1082736	1089977	+	1	4.5E-21	
Yu	Yu_g6542	Scaffold0022	712473	714414	-	1	1E-25	

Table 4-6: Continued.

Yu	Yu_g6543	Scaffold0022	722434	724012	-	1	3.6E-33
Yu	Yu_g6545	Scaffold0022	728942	730248	+	1	4.2E-30
Yu	Yu_g6546	Scaffold0022	731868	733500	+	1	2.1E-32
Yu	Yu_g6602	Scaffold0022	1116033	1122087	-	1	2.1E-29
Yu	Yu_g6889	Scaffold0024	621451	624567	-	2	1.3E-27
Yu	Yu_g6890	Scaffold0024	627744	628933	-	1	1.2E-17
Yu	Yu_g6891	Scaffold0024	629528	632013	-	2	6.5E-23
Yu	Yu_g6894	Scaffold0024	648182	651949	-	2	1.9E-29
Yu	Yu_g7416	Scaffold0027	651535	654239	-	2	5.1E-25
Yu	Yu_g7494	Scaffold0028	130863	132015	+	1	9.7E-22
Yu	Yu_g7497	Scaffold0028	139664	140443	+	1	4.7E-24
Yu	Yu_g7499	Scaffold0028	145332	147787	+	1	1.8E-15
Yu	Yu_g7500	Scaffold0028	148911	149690	+	1	4E-24
Yu	Yu_g7573	Scaffold0028	569910	579317	+	1	6.6E-22
Yu	Yu_g7615	Scaffold0028	826029	828030	-	1	1.4E-53
Yu	Yu_g7877	Scaffold0030	577335	578719	-	1	3E-25
Yu	Yu_g7882	Scaffold0030	585211	593718	+	1	6.8E-23
Yu	Yu_g8081	Scaffold0031	858188	863588	+	2	2.2E-51
Yu	Yu_g8084	Scaffold0031	871851	874794	+	2	5.8E-54
Yu	Yu_g8366	Scaffold0033	764860	767821	-	2	2.7E-45
Yu	Yu_g8368	Scaffold0033	773462	774009	-	1	3.1E-28
Yu	Yu_g8460	Scaffold0034	352007	354466	+	3	2.8E-15
Yu	Yu_g8669	Scaffold0035	688921	694054	+	2	8.6E-30
Yu	Yu_g9278	Scaffold0040	276318	279473	+	2	7.5E-28
Yu	Yu_g9683	Scaffold0043	496479	501612	-	2	2.4E-35
Yu	Yu_g10467	Scaffold0050	341882	344153	-	1	1.4E-28
Yu	Yu_g10472	Scaffold0050	370569	376481	-	1	1.6E-13
Yu	Yu_g10730	Scaffold0052	624443	627399	-	1	5.3E-27
Yu	Yu_g11177	Scaffold0057	243579	246865	+	2	1.4E-40
Yu	Yu_g11179	Scaffold0057	251975	254499	+	2	1.9E-40
Yu	Yu_g11181	Scaffold0057	259401	262534	+	2	6E-45
Yu	Yu_g11403	Scaffold0059	383675	386129	+	2	2.3E-25
Yu	Yu_g11785	Scaffold0063	486582	489560	+	2	4E-52
Yu	Yu_g11824	Scaffold0064	134728	135383	-	1	5.6E-12
Yu	Yu_g11827	Scaffold0064	144664	145441	-	1	3.2E-22
Yu	Yu_g11984	Scaffold0065	559614	563254	+	2	1.4E-49
Yu	Yu_g12320	Scaffold0069	345219	347198	+	1	6.9E-28
Yu	Yu_g12435	Scaffold0070	390814	394299	+	2	3.6E-44
Yu	Yu_g12718	Scaffold0073	521050	525148	+	2	1.1E-52
Yu	Yu_g12898	Scaffold0075	477028	479404	+	2	1.5E-20
Yu	Yu_g13816	Scaffold0088	85417	86871	+	2	1E-39
Yu	Yu_g13820	Scaffold0088	105513	107210	-	2	3.8E-38
Yu	Yu_g13829	Scaffold0088	153512	154121	-	1	8E-25
Yu	Yu_g13840	Scaffold0088	202256	207062	-	2	1.3E-26
Yu	Yu_g13841	Scaffold0088	210047	213034	-	1	2.8E-14
Yu	Yu_g13954	Scaffold0090	3438	7089	-	1	1.7E-24
Yu	Yu_g14110	Scaffold0092	63858	66738	+	1	2.4E-38
Yu	Yu_g14798	Scaffold0102	222670	225230	-	1	9.1E-20
Yu	Yu_g14991	Scaffold0105	218381	224548	-	1	1.1E-42
Yu	Yu_g16221	Scaffold0131	1548	3014	-	1	7.9E-19
Yu	Yu_g16313	Scaffold0133	132493	135890	-	1	3.3E-18
Yu	Yu_g17114	Scaffold0158	76269	79124	-	2	1.8E-46
Yu	Yu_g17115	Scaffold0158	82292	85258	-	2	4.1E-42
Yu	Yu_g18546	Scaffold0354	19654	23742	+	2	5.6E-50

^a Ag: *A. gubernaculifera*, Cr: *C. reinhardtii*, Eu: *Eudorina* sp., Gp: *G. pectorale*, Ts: *T. socialis*, Vc: *V. carteri*, Yu: *Y. unicocca*. The sources of genome data are listed in Table 4-1.

^b For *V. carteri* : g##### (v1.0); Vocar#####m (v2.0); Vocar.#####s#####.p (v2.1).

^c For *V. carteri*, coordinate locations in v2.1 genome are shown in all protein models.

^d The lowest E-value of all domains from HMMER search by A domain model, B domain model or PF12499 (DUF3707) is shown.

^e Based on Prochnik et al. (2010) and Hanschen et al. (2016).

Table 4-7: GO terms enriched in somatic, reproductive and constitutive genes in *Astrephomene gubernaculifera*.

GO ID	GO Name	GO Category	FDR
Somatic enriched terms			
GO:0007165	signal transduction	BIOLOGICAL_PROCESS	5.38E-09
GO:0005794	Golgi apparatus	CELLULAR_COMPONENT	1.14E-07
GO:0016301	kinase activity	MOLECULAR_FUNCTION	2.26E-07
GO:0030705	cytoskeleton-dependent intracellular transport	BIOLOGICAL_PROCESS	2.00E-04
GO:0003700	DNA-binding transcription factor activity	MOLECULAR_FUNCTION	3.42E-04
GO:0006464	cellular protein modification process	BIOLOGICAL_PROCESS	1.04E-03
GO:0031410	cytoplasmic vesicle	CELLULAR_COMPONENT	4.28E-03
GO:0005856	cytoskeleton	CELLULAR_COMPONENT	4.68E-03
GO:0005576	extracellular region	CELLULAR_COMPONENT	5.26E-03
GO:0005929	cilium	CELLULAR_COMPONENT	1.91E-02
GO:0016192	vesicle-mediated transport	BIOLOGICAL_PROCESS	2.12E-02
Reproductive enriched terms			
GO:0006091	generation of precursor metabolites and energy	BIOLOGICAL_PROCESS	5.25E-26
GO:0044281	small molecule metabolic process	BIOLOGICAL_PROCESS	1.22E-23
GO:0015979	photosynthesis	BIOLOGICAL_PROCESS	7.34E-22
GO:0016491	oxidoreductase activity	MOLECULAR_FUNCTION	1.49E-20
GO:0009058	biosynthetic process	BIOLOGICAL_PROCESS	1.49E-20
GO:0006520	cellular amino acid metabolic process	BIOLOGICAL_PROCESS	7.08E-17
GO:0042254	ribosome biogenesis	BIOLOGICAL_PROCESS	3.47E-14
GO:0006412	translation	BIOLOGICAL_PROCESS	8.95E-14
GO:0005622	intracellular	CELLULAR_COMPONENT	3.88E-13
GO:0005730	nucleolus	CELLULAR_COMPONENT	9.50E-10
GO:0034641	cellular nitrogen compound metabolic process	BIOLOGICAL_PROCESS	1.05E-09
GO:0016874	ligase activity	MOLECULAR_FUNCTION	3.07E-09
GO:0003735	structural constituent of ribosome	MOLECULAR_FUNCTION	1.05E-08
GO:0006457	protein folding	BIOLOGICAL_PROCESS	2.28E-08
GO:0005739	mitochondrion	CELLULAR_COMPONENT	2.48E-08
GO:0005840	ribosome	CELLULAR_COMPONENT	7.33E-08
GO:0032991	protein-containing complex	CELLULAR_COMPONENT	1.74E-07
GO:0043226	organelle	CELLULAR_COMPONENT	3.66E-07
GO:0009579	thylakoid	CELLULAR_COMPONENT	7.46E-07
GO:0005198	structural molecule activity	MOLECULAR_FUNCTION	1.03E-06
GO:0051082	unfolded protein binding	MOLECULAR_FUNCTION	4.00E-06

Table 4-7: Continued.

GO:0003723	RNA binding	MOLECULAR_FUNCTION	9.89E-05
GO:0006790	sulfur compound metabolic process	BIOLOGICAL_PROCESS	3.77E-04
GO:0005634	nucleus	CELLULAR_COMPONENT	4.40E-04
GO:0006399	tRNA metabolic process	BIOLOGICAL_PROCESS	6.34E-04
GO:0007005	mitochondrion organization	BIOLOGICAL_PROCESS	1.70E-03
GO:0019843	rRNA binding	MOLECULAR_FUNCTION	5.83E-03
GO:0005694	chromosome	CELLULAR_COMPONENT	6.70E-03
GO:0043167	ion binding	MOLECULAR_FUNCTION	6.79E-03
GO:0016829	lyase activity	MOLECULAR_FUNCTION	8.98E-03
GO:0008135	translation factor activity, RNA binding	MOLECULAR_FUNCTION	1.47E-02
GO:0016853	isomerase activity	MOLECULAR_FUNCTION	1.55E-02
GO:0005737	cytoplasm	CELLULAR_COMPONENT	1.78E-02
GO:0006913	nucleocytoplasmic transport	BIOLOGICAL_PROCESS	1.92E-02
Constitutive enriched terms			
GO:0016192	vesicle-mediated transport	BIOLOGICAL_PROCESS	2.54E-02
GO:0005654	nucleoplasm	CELLULAR_COMPONENT	4.87E-02

Table 4-8: List of expression of flagella, transition zone and basal body genes in *Astrephomene gubernaculifera*.

gene name	locus	soma TPM	repr TPM	log2 fold change	classification	<i>V. carteri</i> homolog ^a	<i>C. reinhardtii</i> homolog ^b	reference
Flagella, Axoneme								
CNK2	Agub_g578	46.20	6.46	2.634	somatic- biased	Vocar.0001s1051	Cre12.g560350	(Zones et al., 2015)
MBO2	Agub_g3858	21.68	17.67	0.109	constitutive	Vocar.0011s0271	Cre09.g416550	(Zones et al., 2015)
PP2A	Agub_g7013	362.31	282.25	0.168	constitutive	Vocar.0027s0044	Cre11.g477300	(Zones et al., 2015)
RIB43a	Agub_g1749	102.20	155.58	-0.798	reproductive- upregulated	Vocar.0015s0345	Cre03.g178500	(Zones et al., 2015)
RIB72	Agub_g797	410.36	531.84	-0.567	reproductive- upregulated	Vocar.0001s1159	Cre02.g091700	(Zones et al., 2015)
PF16	Agub_g6364	326.89	217.28	0.397	somatic- upregulated	Vocar.0009s0310	Cre09.g394251	(Zones et al., 2015)
PP1c	Agub_g12031	240.72	225.63	-0.099	constitutive	Vocar.0024s0253	Cre06.g292550	(Zones et al., 2015)
PF20	Agub_g2605	182.84	136.21	0.233	constitutive	Vocar.0001s1782	Cre04.g227900	(Zones et al., 2015)
C1a-32	Agub_g9432	36.37	49.88	-0.649	reproductive- upregulated	Vocar.0026s0152	Cre09.g389282	(Zones et al., 2015)
C1a-34	Agub_g4154	38.45	30.95	0.117	constitutive	Vocar.0025s0019	Cre06.g256450	(Zones et al., 2015)
C1a-86	Agub_g10055	16.50	16.26	-0.175	constitutive	Vocar.0001s1652	Cre02.g112100	(Zones et al., 2015)
PF6	Agub_g7062	27.05	23.86	-0.010	constitutive	Vocar.0014s0117	Cre10.g434400	(Zones et al., 2015)
C1b-135	Agub_g1947	20.83	15.81	0.199	constitutive	Vocar.0040s0092	Cre03.g168200	(Zones et al., 2015)
C1b-350	Agub_g4841	55.63	44.58	0.128	constitutive	Vocar.0037s0136	Cre12.g519950	(Zones et al., 2015)
CPC1	Agub_g1600	82.49	64.26	0.170	constitutive	Vocar.0015s0258	Cre03.g183200	(Zones et al., 2015)
PGH1	Agub_g4701	1744.11	2173.66	-0.510	reproductive- upregulated	Vocar.0006s0398	Cre12.g513200	(Zones et al., 2015)
HY3/HYD3	Agub_g13654	65.98	36.95	0.645	somatic- upregulated	Vocar.0007s0373	Cre01.g025400	(Zones et al., 2015)
KLP1	Agub_g9392	84.02	54.14	0.436	somatic- upregulated	Vocar.0056s0026	Cre02.g073750	(Zones et al., 2015)
FAP250/DRC2	Agub_g3566	49.84	44.12	-0.017	constitutive	Vocar.0024s0131	Cre13.g607750	(Zones et al., 2015)
FAP134/DRC3	Agub_g5202	164.31	117.41	0.291	constitutive	Vocar.0028s0074	Cre12.g531500	(Zones et al., 2015)
PF2/DRC4	Agub_g7225	43.30	20.28	0.909	somatic- upregulated	Vocar.0027s0036	Cre11.g476850	(Zones et al., 2015)
FAP155/DRC5	Agub_g10098	55.96	50.71	-0.044	constitutive	Vocar.0001s0692	Cre01.g045350	(Zones et al., 2015)
FAP169/DRC6	Agub_g4745	260.33	309.48	-0.439	reproductive- upregulated	Vocar.0006s0317	Cre12.g515200	(Zones et al., 2015)
FAP50/DRC7	Agub_g3646	87.79	53.31	0.525	somatic- upregulated	Vocar.0059s0027	Cre14.g612700	(Zones et al., 2015)
FAP122/DRC9	Agub_g432	221.63	206.94	-0.093	constitutive	Vocar.0001s0833	Cre01.g025450	(Zones et al., 2015)
FAP84/DRC10	Agub_g10256	127.12	139.40	-0.326	constitutive	Vocar.0008s0009	Cre02.g141350	(Zones et al., 2015)
FAP82/DRC11	Agub_g2188	97.06	74.23	0.195	constitutive	Vocar.0021s0107	Cre07.g325762	(Zones et al., 2015)
DHC1/IDA1/PF 9/ROC34	Agub_g14818	140.97	49.47	1.320	somatic- upregulated	Vocar.0031s0108	Cre12.g484250	(Zones et al., 2015)
DHC10/DYH1B /IDA2	Agub_g5687	90.33	50.26	0.654	somatic- upregulated	Vocar.0022s0111	Cre14.g624950	(Zones et al., 2015)
DIC4/BOP5/IC1 38	Agub_g4867	38.76	30.65	0.150	constitutive	Vocar.0037s0110	Cre12.g520950	(Zones et al., 2015)
IDA7	Agub_g9598	125.21	99.16	0.146	constitutive	Vocar.0018s0215	Cre16.g674515	(Zones et al., 2015)
DII6/IC97	Agub_g5505	13.67	8.66	0.468	somatic- upregulated	Vocar.20001728m	Cre14.g631200	(Zones et al., 2015)
LC7a ^c	Agub_g14920	323.41	194.15	0.546	somatic- upregulated	Vocar.0006s0122	Cre08.g376550	(Zones et al., 2015)

Table 4-8: Continued.

LC7b	Agub_g9717	373.01	237.07	0.466	somatic-upregulated	Vocar.0007s0162	Cre12.g546400	(Zones et al., 2015)
ODA7/DAU1	Agub_g406	33.43	52.32	-0.834	reproductive-upregulated	Vocar.0001s0903	Cre01.g029150	(Zones et al., 2015)
DLT3/TCTEX1/TCT1	Agub_g6140	139.25	90.20	0.433	constitutive	Vocar.0013s0042	Cre01.g004250	(Zones et al., 2015)
Tctex2b	Agub_g6363	119.54	101.26	0.048	constitutive	Vocar.0009s0309	Cre09.g394213	(Zones et al., 2015)
VFL2/CEN1	Agub_g6512	257.57	346.11	-0.621	reproductive-upregulated	Vocar.0012s0203	Cre11.g468450	(Zones et al., 2015)
	Agub_g3600	40.73	80.60	-1.182	reproductive-upregulated	Vocar.0050s0015		
DHC2	Agub_g12697	95.21	46.69	0.836	somatic-upregulated	Vocar.0044s0047	Cre09.g392282	(Zones et al., 2015)
DHC5	Agub_g15379	15.93	12.40	0.168	constitutive	Vocar.0022s0148	Cre02.g107050	(Zones et al., 2015)
DHC6	Agub_g5814	30.11	13.77	0.938	somatic-upregulated	Vocar.0026s0029	Cre05.g244250	(Zones et al., 2015)
DHC8	Agub_g10752	62.72	37.20	0.561	somatic-upregulated	Vocar.0018s0113	Cre16.g685450	(Zones et al., 2015)
DHC9	Agub_g7181	69.85	38.71	0.659	somatic-upregulated	Vocar.0027s0019	Cre02.g141606	(Zones et al., 2015)
DHC11	Agub_g593	16.45	10.17	0.502	somatic-upregulated	Vocar.0001s1061	Cre12.g555950	(Zones et al., 2015)
DHC12	Agub_g12938	6.70	5.17	0.185	constitutive	Vocar.0020s0100	Cre06.g297850	(Zones et al., 2015)
IDA4	Agub_g7958	280.50	157.01	0.642	somatic-upregulated	Vocar.0031s0023	Cre12.g494800	(Zones et al., 2015)
FAP146	Agub_g995	49.58	32.60	0.416	constitutive	Vocar.0015s0204	Cre03.g186300	(Zones et al., 2015)
TPR5	Agub_g12290	63.63	69.84	-0.321	constitutive	Vocar.0036s0122	Cre08.g374650	(Zones et al., 2015)
DHC13	Agub_g2868	144.84	107.89	0.232	somatic-upregulated	Vocar.0014s0194	Cre03.g145127	(Zones et al., 2015)
ODA4	Agub_g15612	150.75	92.12	0.519	somatic-upregulated	Vocar.0008s0338	Cre09.g403800	(Zones et al., 2015)
DC1/ODA3	Agub_g10479	66.19	63.45	-0.132	constitutive	Vocar.0004s0202	Cre17.g703850	(Zones et al., 2015)
DC2/ODA1	Agub_g12488	141.36	110.46	0.161	constitutive	Vocar.0066s0015	Cre16.g666150	(Zones et al., 2015)
DHC15/ODA2	Agub_g3000	120.56	82.35	0.358	somatic-upregulated	Vocar.0027s0068	Cre11.g476050	(Zones et al., 2015)
IC1/ODA9	Agub_g3995	91.99	86.32	-0.099	constitutive	Vocar.0043s0052	Cre12.g536550	(Zones et al., 2015)
IC2/ODA6	Agub_g8560	175.70	178.08	-0.211	constitutive	Vocar.0016s0285	Cre12.g506000	(Zones et al., 2015)
LC1	Agub_g3256	104.76	95.16	-0.054	constitutive	Vocar.0001s1196	Cre02.g092850	(Zones et al., 2015)
LC2	Agub_g5104	99.04	80.21	0.112	constitutive	Vocar.0037s0060	Cre12.g527750	(Zones et al., 2015)
LC3	Agub_g4886	117.95	88.88	0.217	constitutive	Vocar.0037s0033	Cre12.g528850	(Zones et al., 2015)
LC4	Agub_g15715	203.76	87.64	1.021	somatic-upregulated	Vocar.0001s0404	Cre01.g051250	(Zones et al., 2015)
LC5	Agub_g12834	129.35	98.41	0.200	constitutive	Vocar.0026s0127	Cre17.g714250	(Zones et al., 2015)
LC6	Agub_g1576	116.91	104.94	-0.032	constitutive	Vocar.0011s0170	Cre03.g187200	(Zones et al., 2015)
LC10	Agub_g5105	100.28	118.68	-0.429	constitutive	Vocar.0037s0059	Cre12.g527800	(Zones et al., 2015)
ODA8	Agub_g205	30.89	18.26	0.563	somatic-upregulated	Vocar.0001s0651	Cre01.g043650	(Zones et al., 2015)
RSP1	Agub_g1241	247.35	123.52	0.809	somatic-upregulated	Vocar.0047s0016	Cre03.g201900	(Zones et al., 2015)
RSP2	Agub_g2141	74.83	43.32	0.599	somatic-upregulated	Vocar.0005s0309	Cre10.g427300	(Zones et al., 2015)
RSP3	Agub_g12046	107.20	66.64	0.493	somatic-upregulated	Vocar.0024s0129	Cre06.g291700	(Zones et al., 2015)
RSP4	Agub_g5954	179.32	129.50	0.278	constitutive	Vocar.0073s0018	Cre05.g242500	(Zones et al., 2015)

Table 4-8: Continued.

RSP5	Agub_g11696	199.62	124.54	0.487	somatic-upregulated	Vocar.0007s0124	Cre12.g544000	(Zones et al., 2015)
RSP6	Agub_g5955	241.46	113.35	0.899	somatic-upregulated	Vocar.0073s0020	Cre05.g242550	(Zones et al., 2015)
RSP7	Agub_g13721	37.68	15.59	1.080	somatic-upregulated	Vocar.0005s0074	Cre07.g347050	(Zones et al., 2015)
RSP8	Agub_g8992	48.23	43.29	-0.025	constitutive	Vocar.0003s0051	Cre02.g078600	(Zones et al., 2015)
RSP9	Agub_g14347	245.70	142.90	0.592	somatic-upregulated	Vocar.0031s0003	Cre07.g330200	(Zones et al., 2015)
RSP10	Agub_g6155	52.55	29.01	0.667	somatic-upregulated	Vocar.0024s0236	Cre01.g005450	(Zones et al., 2015)
RSP11	Agub_g12890	133.87	91.15	0.361	constitutive	Vocar.0029s0100	Cre10.g465250	(Zones et al., 2015)
RSP12	Agub_g8420	60.40	40.93	0.368	constitutive	Vocar.0017s0097	Cre02.g114700	(Zones et al., 2015)
RSP14	Agub_g3588	145.36	103.59	0.297	constitutive	Vocar.0024s0056	Cre14.g617500	(Zones et al., 2015)
RSP16	Agub_g6232	136.42	102.01	0.227	constitutive	Vocar.0016s0197	Cre12.g511050	(Zones et al., 2015)
RSP17	Agub_g12650	130.78	82.04	0.480	somatic-upregulated	Vocar.0005s0012	Cre07.g340400	(Zones et al., 2015)
CAM1/RSP20	Agub_g1742	2606.06	2521.24	-0.146	constitutive	Vocar.0015s0353	Cre03.g178150	(Zones et al., 2015)
RSP23	Agub_g15296	125.66	107.22	0.030	constitutive	Vocar.0041s0085	Cre16.g654300	(Zones et al., 2015)
Flagella, IFT								
D1bLIC	Agub_g15875	143.72	88.81	0.505	somatic-upregulated	Vocar.0032s0018	Cre09.g398882	(Zones et al., 2015)
DHC1b/FLA24	Agub_g12346	154.19	44.49	1.602	somatic-biased	Vocar.0038s0044	Cre06.g250300	(Zones et al., 2015)
FAP133/DIC5	Agub_g11972	98.38	29.23	1.567	somatic-upregulated	Vocar.0001s1610	Cre02.g110950	(Zones et al., 2015)
FLA10/KHP1	Agub_g8363	135.79	52.23	1.187	somatic-upregulated	Vocar.0009s0040	Cre17.g730950	(Zones et al., 2015)
FLA8	Agub_g5013	180.50	98.94	0.674	somatic-upregulated	Vocar.0006s0438	Cre12.g522550	(Zones et al., 2015)
FLA3/KAP1	Agub_g5895	233.38	108.63	0.912	somatic-upregulated	Vocar.0070s0018	Cre10.g449250	(Zones et al., 2015)
IFT43	Agub_g4209	178.58	96.45	0.694	somatic-upregulated	Vocar.0086s0003	Cre06.g251200	(Zones et al., 2015)
IFT121	Agub_g10706	87.12	34.72	1.131	somatic-upregulated	Vocar.0027s0106	Cre11.g475000	(Zones et al., 2015)
IFT122	Agub_g9946	250.41	92.62	1.241	somatic-upregulated	Vocar.0001s0466	Cre01.g065822	(Zones et al., 2015)
IFT139	Agub_g4382	155.66	50.91	1.421	somatic-biased	Vocar.0002s0151	Cre06.g268800	(Zones et al., 2015)
IFT140	Agub_g10531	222.22	72.39	1.425	somatic-biased	Vocar.0004s0184	Cre08.g362650	(Zones et al., 2015)
IFT144	Agub_g9212	219.60	69.83	1.462	somatic-biased	Vocar.0001s0188	Cre13.g572700	(Zones et al., 2015)
IFT20	Agub_g710	66.83	27.35	1.099	somatic-upregulated	Vocar.0001s1095	Cre02.g089950	(Zones et al., 2015)
IFT25	Agub_g6942	153.25	83.28	0.687	somatic-upregulated	Vocar.0008s0425	Cre10.g450350	(Zones et al., 2015)
IFT27	Agub_g10117	210.23	85.73	1.102	somatic-upregulated	Vocar.0001s0561	Cre01.g047950	(Zones et al., 2015)
IFT52	Agub_g2329	234.37	133.34	0.623	somatic-upregulated	Vocar.0025s0090	Cre04.g219250	(Zones et al., 2015)
IFT57/IFT55	Agub_g12901	45.89	27.69	0.537	somatic-upregulated	Vocar.0029s0088	Cre10.g467000	(Zones et al., 2015)
IFT70	Agub_g12659	75.90	54.03	0.298	somatic-upregulated	Vocar.0067s0035	Cre07.g342200	(Zones et al., 2015)
IFT74/72	Agub_g400	72.11	32.33	0.967	somatic-upregulated	Vocar.0001s0895	Cre01.g027950	(Zones et al., 2015)
IFT80	Agub_g1523	940.64	873.06	-0.085	constitutive	Vocar.0047s0029	Cre03.g204150	(Zones et al., 2015)
IFT81	Agub_g3529	67.07	38.48	0.610	somatic-upregulated	Vocar.0024s0088	Cre17.g723600	(Zones et al., 2015)
IFT88	Agub_g12596	65.98	27.03	1.097	somatic-upregulated	Vocar.0011s0115	Cre07.g335750	(Zones et al., 2015)

Table 4-8: Continued.

IFT172	Agub_g10478	100.10	37.74	1.216	somatic-upregulated	Vocar.0004s0203	Cre17.g703900	(Zones et al., 2015)
IFT54	Agub_g6564	185.26	101.67	0.675	somatic-upregulated	Vocar.0030s0182	Cre11.g467739	(Zones et al., 2015)
IFT46	Agub_g6918	48.21	16.41	1.362	somatic-upregulated	Vocar.0019s0210	Cre05.g241637	(Zones et al., 2015)
Flagella, BBS								
ARL6/BBS3	Agub_g8028	1782.13	1084.61	0.525	somatic-upregulated	Vocar.0016s0153	Cre12.g486250	(Zones et al., 2015)
BBS1	Agub_g9536	77.91	27.86	1.289	somatic-upregulated	Vocar.0003s0317	Cre17.g741950	(Zones et al., 2015)
BBS2	Agub_g12034	117.72	54.61	0.914	somatic-upregulated	Vocar.0002s0661	Cre06.g257250	(Zones et al., 2015)
BBS3A	Agub_g8539	10.54	2.47	1.860	somatic-upregulated	Vocar.0016s0307	Cre12.g506851	(Zones et al., 2015)
BBS3B	Agub_g7438	31.56	23.43	0.228	constitutive	Vocar.0041s0017	Cre16.g664500	(Zones et al., 2015)
BBS4	Agub_g9807	67.96	27.88	1.086	somatic-upregulated	Vocar.0007s0218	Cre12.g548650	(Zones et al., 2015)
BBS5	Agub_g4401	128.03	53.19	1.073	somatic-upregulated	Vocar.0002s0180	Cre06.g267550	(Zones et al., 2015)
BBS7	Agub_g202	116.21	56.30	0.854	somatic-upregulated	Vocar.0001s0649	Cre01.g043750	(Zones et al., 2015)
BBS8	Agub_g12497	24.83	10.18	1.091	somatic-upregulated	Vocar.0066s0006	Cre16.g666500	(Zones et al., 2015)
BBS9	Agub_g2332	28.97	14.11	0.848	somatic-upregulated	Vocar.0025s0100	Cre04.g219700	(Zones et al., 2015)
Transition zone								
MKS1	Agub_g7952	5.24	1.77	1.341	low expression	Vocar.0031s0027	Cre12.g495350	(Barker et al., 2014)
TMEM216	Agub_g11353	34.74	27.48	0.135	constitutive	Vocar.0042s0022	Cre08.g379750	(Barker et al., 2014)
TMEM67	Agub_g15327	2.79	2.92	-0.257	constitutive	Vocar.0001s0050	Cre13.g584901	(Barker et al., 2014)
CEP290	Agub_g1955	15.36	8.07	0.737	somatic-upregulated	Vocar.0040s0104	Cre03.g167550	(Barker et al., 2014)
RPGR1PL1	Agub_g5456	68.31	8.63	2.792	somatic-biased	Vocar.0022s0202	Cre14.g632050	(Barker et al., 2014)
CC2D2A	Agub_g7350	11.05	3.05	1.667	somatic-upregulated	Vocar.0004s0137	Cre16.g660750	(Barker et al., 2014)
B9D1	Agub_g9599	37.21	32.91	-0.021	constitutive	Vocar.0018s0214	Cre16.g674627	(Barker et al., 2014)
B9D2	Agub_g1411	20.76	11.39	0.657	constitutive	Vocar.0035s0113	Cre03.g196050	(Barker et al., 2014)
AHI1	Agub_g7770	3.22	0.58	2.294	somatic-biased	Vocar.0012s0258	Cre16.g692902	(Barker et al., 2014)
TMEM17	Agub_g3281	22.01	20.14	-0.062	constitutive	Vocar.0024s0023	Cre10.g435600	(Barker et al., 2014)
TMEM231	Agub_g651	15.63	40.09	-1.547	reproductive-upregulated	Vocar.0001s1268	Cre10.g441250	(Barker et al., 2014)
NPHP4	Agub_g5201	31.72	12.18	1.187	somatic-upregulated	Vocar.0028s0072	Cre12.g531400	(Barker et al., 2014)
TCTN ^c	Agub_g1677	24.96	31.61	-0.536	reproductive-upregulated		Cre12.g181450	(Barker et al., 2014)
Basal bodies, core (genes with validated basal body/centriolar function)								
BLD10/CEP135	Agub_g2559	4.94	2.82	0.620	constitutive	Vocar.0016s0003	Cre10.g418250	(Zones et al., 2015)
BLD2/TUE1	Agub_g1835	16.41	10.16	0.502	constitutive	Vocar.0015s0161	Cre03.g172650	(Zones et al., 2015)
FA2	Agub_g5259	10.31	4.70	0.950	somatic-upregulated	Vocar.0011s0057	Cre07.g351150	(Zones et al., 2015)
SAS6	Agub_g4784	16.32	12.44	0.190	constitutive	Vocar.0006s0339	Cre12.g516950	(Zones et al., 2015)
SFA1	Agub_g14186	86.62	20.04	1.922	somatic-biased	Vocar.0009s0093	Cre07.g332950	(Zones et al., 2015)
SFI1 ^c	Agub_g15635	0.84	0.49	0.564	constitutive	Vocar.0008s0350	Cre09.g404500	(Zones et al., 2015)
TUG1	Agub_g14649	27.30	58.66	-1.305	reproductive-upregulated	Vocar.0001s0264	Cre06.g299300	(Zones et al., 2015)
UNI2	Agub_g6376	1.38	1.15	0.081	constitutive	Vocar.0030s0154	Cre09.g394695	(Zones et al., 2015)

Table 4-8: Continued.

UNI3/TUD1	Agub_g1581	8.99	8.77	-0.154	constitutive	Vocar.0011s0165	Cre03.g187350	(Zones et al., 2015)
VFL1	Agub_g4909	1.93	4.22	-1.339	low confidence	Vocar.0065s0013	Cre08.g372900	(Zones et al., 2015)
VFL3	Agub_g13324	19.77	9.47	0.866	constitutive	Vocar.0002s0080	Cre06.g279900	(Zones et al., 2015)
FA1	Agub_g2568	3.74	4.25	-0.381	constitutive	Vocar.0002s0679	Cre06.g257600	(Zones et al., 2015)
SAS4 ^c	Agub_g4541	3.88	2.78	0.267	constitutive	Vocar.0002s0400	Cre06.g275400	(Zones et al., 2015)
VFL2/CEN1	Agub_g6512	257.57	346.11	-0.621	reproductive-upregulated	Vocar.0012s0203	Cre11.g468450	(Zones et al., 2015)
	Agub_g3600	40.73	80.60	-1.182	reproductive-upregulated	Vocar.0050s0015		
UNI1	Agub_g10454	8.30	8.74	-0.273	constitutive	Vocar.0004s0236	Cre17.g705650	(Zones et al., 2015)
FAP134	Agub_g5202	164.31	117.41	0.291	constitutive	Vocar.0028s0074	Cre12.g531500	(Zones et al., 2015)
SSA17	Agub_g5912	40.56	33.16	0.099	constitutive	Vocar.0026s0085	Cre10.g444450	(Hodges et al., 2010)
CEP76	Agub_g11197	2.71	5.63	-1.271	low confidence	Vocar.0020s0014	Cre13.g588250	(Hodges et al., 2010)
DIP13	Agub_g3502	267.28	233.99	-0.001	constitutive	Vocar.0024s0043	Cre17.g724550	(Hodges et al., 2010)
NPHP1	Agub_g6810	8.04	0.89	3.025	somatic-upregulated	Vocar.0073s0036 Vocar.0008s0075	Cre08.g385250	(Hodges et al., 2010)
NPHP3	Agub_g14870	40.13	59.19	-0.752	reproductive-upregulated	Vocar.0031s0051	Cre12.g495250	(Hodges et al., 2010)
NPHP5	Agub_g6029	26.26	17.40	0.408	constitutive	Vocar.0013s0057	Cre06.g284300	(Hodges et al., 2010)
POC1	Agub_g8414	80.89	31.82	1.166	somatic-upregulated	Vocar.0009s0060	Cre02.g115250	(Hodges et al., 2010)
POC5	Agub_g2075	10.12	17.67	-1.013	low confidence	Vocar.0005s0174	Cre03.g163000	(Hodges et al., 2010)
WDR16	Agub_g11452	67.99	56.24	0.083	constitutive	Vocar.0004s0391	Cre17.g728650	(Hodges et al., 2010)
ASQ2	Agub_g6379	25.37	43.10	-0.958	reproductive-upregulated	Vocar.0009s0148	Cre09.g394880	(Feldman and Marshall, 2009)
Basal bodies, BUG								
TEK1	Agub_g11258	708.42	238.79	1.376	somatic-biased	Vocar.0039s0062	Cre16.g655750	(Keller et al., 2005)
DIP13	Agub_g3502	267.28	233.99	-0.001	constitutive	Vocar.0024s0043	Cre17.g724550	(Keller et al., 2005)
Rib43a	Agub_g1749	102.20	155.58	-0.798	reproductive-upregulated	Vocar.0015s0345	Cre03.g178500	(Keller et al., 2005)
Rib72	Agub_g797	410.36	531.84	-0.567	reproductive-upregulated	Vocar.0001s1159	Cre02.g091700	(Keller et al., 2005)
CCT3	Agub_g5871	205.75	303.29	-0.754	reproductive-upregulated	Vocar.0026s0092	Cre10.g443250	(Keller et al., 2005)
BUG1	Agub_g14785	185.32	249.31	-0.620	reproductive-upregulated	Vocar.0019s0112	Cre10.g432850	(Keller et al., 2005)
BUG2	Agub_g3121	156.59	249.22	-0.864	reproductive-upregulated	Vocar.0002s0489	Cre06.g261900	(Keller et al., 2005)
BUG3	Agub_g4100	56.51	37.75	0.394	constitutive	Vocar.0028s0056	Cre13.g564400	(Keller et al., 2005)
BUG4	Agub_g9365	329.99	371.15	-0.361	reproductive-upregulated	Vocar.0056s0010	Cre09.g390949	(Keller et al., 2005)
BUG5	Agub_g550	226.18	146.22	0.438	somatic-upregulated	Vocar.0001s1138	Cre12.g558700	(Keller et al., 2005)
BUG6	Agub_g6029	26.26	17.40	0.408	constitutive	Vocar.0013s0057	Cre06.g284300	(Keller et al., 2005)
BUG7	Agub_g10434	186.77	295.62	-0.852	reproductive-upregulated	Vocar.0004s0214	Cre17.g701500	(Keller et al., 2005)
BUG8	Agub_g1184	61.52	21.73	1.303	somatic-upregulated	Vocar.0011s0349	Cre03.g199500	(Keller et al., 2005)
BUG9	Agub_g10639	112.80	123.53	-0.329	constitutive	Vocar.0027s0128	Cre02.g095149	(Keller et al., 2005)
BUG10	Agub_g12906	630.72	453.47	0.284	somatic-upregulated	Vocar.0029s0081	Cre10.g466650	(Keller et al., 2005)

Table 4-8: Continued.

BUG11	Agub_g10522	4.69	7.29	-0.829	reproductive-upregulated	Vocar.0004s0175	Cre17.g703600	(Keller et al., 2005)
BUG12	Agub_g9238	127.95	115.28	-0.037	constitutive	Vocar.0001s0359	Cre13.g573500	(Keller et al., 2005)
BUG13	Agub_g2219	391.97	610.69	-0.831	reproductive-upregulated	Vocar.0005s0328	Cre10.g424250	(Keller et al., 2005)
BUG14	Agub_g8088	156.33	162.08	-0.242	constitutive	Vocar.0016s0100	Cre12.g489750	(Keller et al., 2005)
BUG15	Agub_g13522	155.28	78.32	0.798	somatic-upregulated	Vocar.0008s0314	Cre18.g749947	(Keller et al., 2005)
BUG16	Agub_g3999	50.01	131.33	-1.585	reproductive-upregulated	Vocar.0043s0073	Cre12.g536100	(Keller et al., 2005)
BUG17	Agub_g15178	70.71	91.91	-0.572	reproductive-upregulated	Vocar.0007s0423	Cre01.g015850	(Keller et al., 2005)
BUG18	Agub_g10112	5.82	6.28	-0.306	constitutive	Vocar.0001s0563	Cre01.g048250	(Keller et al., 2005)
BUG19	Agub_g14917	11.18	11.71	-0.265	constitutive	Vocar.0006s0030	Cre01.g012800	(Keller et al., 2005)
BUG20	Agub_g15298	8.51	2.33	1.683	somatic-upregulated	Vocar.0041s0089	Cre16.g654150	(Keller et al., 2005)
BUG21	Agub_g10214	258.38	374.45	-0.727	reproductive-upregulated	Vocar.0001s1595	Cre02.g113400	(Keller et al., 2005)
BUG22	Agub_g5270	591.84	445.05	0.219	constitutive	Vocar.0011s0048	Cre07.g351650	(Keller et al., 2005)
BUG23	Agub_g15211	26.03	7.04	1.692	somatic-upregulated	Vocar.0029s0015	Cre10.g461350	(Keller et al., 2005)
BUG24	Agub_g4869	6.41	4.42	0.347	constitutive	Vocar.0037s0107	Cre12.g521100	(Keller et al., 2005)
BUG28	Agub_g5456	68.31	8.63	2.792	somatic-biased	Vocar.0022s0202	Cre14.g632050	(Keller et al., 2009)
BUG31	Agub_g3258	48.03	47.72	-0.180	constitutive	Vocar.0001s1194	Cre02.g092700	(Keller et al., 2009)
BUG32	Agub_g8451	311.05	295.92	-0.120	constitutive	Vocar.0017s0171	Cre17.g737100	(Keller et al., 2009)
Basal bodies, POC								
POC1	Agub_g8414	80.89	31.82	1.166	somatic-upregulated	Vocar.0009s0060	Cre02.g115250	(Keller et al., 2005)
POC2	Agub_g3382	8.10	5.24	0.434	somatic-upregulated	Vocar.0057s0038	Cre10.g436000	(Keller et al., 2005)
POC3	Agub_g1955	15.36	8.07	0.737	somatic-upregulated	Vocar.0040s0104	Cre03.g167550	(Keller et al., 2005)
POC4	Agub_g1895	7.18	1.02	2.575	somatic-biased	Vocar.0040s0044	Cre03.g170550	(Keller et al., 2005)
POC5	Agub_g2075	10.12	17.67	-1.013	low confidence	Vocar.0005s0174	Cre03.g163000	(Keller et al., 2005)
POC6	Agub_g13251	359.19	135.65	1.219	somatic-upregulated	Vocar.0002s0032	Cre06.g283550	(Keller et al., 2005)
POC7	Agub_g5354	467.05	453.37	-0.150	constitutive	Vocar.0051s0011	Cre07.g356400	(Keller et al., 2005)
POC8	Agub_g4217	1305.65	1340.22	-0.231	reproductive-upregulated	Vocar.0002s0678	Cre06.g257500	(Keller et al., 2005)
POC9	Agub_g3617	8.46	14.64	-0.981	reproductive-upregulated	Vocar.0050s0057	Cre14.g611250	(Keller et al., 2005)
POC10	Agub_g5201	31.72	12.18	1.187	somatic-upregulated	Vocar.0028s0072	Cre12.g531400	(Keller et al., 2005)
POC11	Agub_g1396	12.74	3.53	1.680	somatic-upregulated	Vocar.0033s0121	Cre03.g195450	(Keller et al., 2005)
POC12	Agub_g7953	3.95	9.82	-1.478	low confidence	Vocar.0031s0027	Cre12.g495350	(Keller et al., 2005)
POC13	Agub_g2916	157.77	104.45	0.408	constitutive	Vocar.0019s0011	Cre03.g156250	(Keller et al., 2005)
POC14	Agub_g538	2535.03	2288.96	-0.045	constitutive	Vocar.0001s1177	Cre12.g559250	(Keller et al., 2005)
POC15	Agub_g7506	1.42	0.79	0.660	somatic-upregulated	Vocar.0014s0277	Cre13.g607650	(Keller et al., 2005)
POC17	Agub_g6054	107.34	220.00	-1.227	reproductive-upregulated	Vocar.0004s0447	Cre02.g088000	(Keller et al., 2005)
POC19	Agub_g8083	4.10	0.00	7.309	somatic-specific	Vocar.0016s0103	Cre12.g492850	(Keller et al., 2009)
POC20	Agub_g13484	121.67	100.36	0.085	constitutive	Vocar.0008s0298	Cre09.g386400	(Keller et al., 2009)

Table 4-8: Continued.

POC21	Agub_g11441	85.26	60.65	0.294	constitutive	Vocar.0028s0169	Cre10.g456100	(Keller et al., 2009)
-------	-------------	-------	-------	-------	--------------	-----------------	---------------	--------------------------

^a Vocar#####m (v2.0); Vocar.#####s##### (v2.1).

^b Loci in v5.6 genome of *C. reinhardtii*

^c Not included in Matt & Umen (2018) but added in this study.

Table 4-9: List of expression of photosynthesis genes in *Astrephomene gubernaculifera*.

gene name	locus	soma TPM	repr TPM	log2 fold change	classification	<i>V. carteri</i> homolog ^a	<i>C. reinhardtii</i> homolog ^b	reference	
Core photosynthesis, LHCII									
LHCBM1	Agub_g9939	1157.13	2759.72	-1.449	reproductive- biased	Vocar.0001s0479	Cre01.g066917	(Elrad and Grossman, 2004; Matt and Umen, 2018; Zones et al., 2015; this study)	
LHCBM5	Agub_g7108	356.87	839.84	-1.430	reproductive- biased	Vocar.0014s0016	Cre03.g156900		
Major LHCII Clade 1	Agub_g2631 (LHCBM3.1)	3406.21	4944.92	-0.731	reproductive- upregulated	Vocar.0002s0046 (LHCBM3.1); Vocar.0002s0051 (LHCBM3.2); Vocar.0002s0043 (LHCBM3.3); Vocar.0002s0047 (LHCBM3.4); Vocar.0002s0041 (LHCBM3.5); Vocar.0002s0042 (LHCBM3.6); Vocar.0709s0001 (LHCBM3.7); Vocar.0002s0050 (LHCBM3.8); Vocar.0002s0044 (LHCBM3.9); Vocar.0002s0048 (LHCBM3.10); Vocar.0002s0055 (LHCBM3.11); Vocar.0002s0054 (LHCBM3.12); Vocar.0006s0236 (LHCBM3.13); Vocar.0002s0052 (LHCBM3.14); Vocar.0050s0033 (LHCBM3.15); Vocar.0025s0135 (LHCBM3.16); Vocar.0001s1775 (LHCBM3.17)	Cre04.g232104 (LHCBM3); Cre06.g283950 (LHCBM4); Cre06.g285250 (LHCBM6); Cre06.g284250 (LHCBM8); Cre06.g284200 (LHCBM9)		
	Agub_g13253 (LHCBM3.2)	377.29	1034.54	-1.650	reproductive- upregulated				
	Agub_g13256 (LHCBM3.3)	20.23	115.83	-2.713	reproductive- biased				
	Agub_g13257 (LHCBM3.4)	46.59	299.75	-2.897	reproductive- biased				
Major LHCII Clade 4	Agub_g9801 (LHCBM2.1)	700.77	1720.56	-1.490	reproductive- biased	Vocar.0012s0221 (LHCBM2.1); Vocar.0007s0216 (LHCBM2.2); Vocar.0007s0215 (LHCBM2.3); Vocar.0012s0144 (LHCBM2.4); Vocar.0012s0172 (LHCBM2.5); Vocar.0012s0145 (LHCBM2.6); Vocar.0030s0156 (LHCBM2.7)	Cre12.g548400 (LHCBM2); Cre12.g548950 (LHCBM7)		
	Agub_g9805 (LHCBM2.2)	1613.31	4844.59	-1.780	reproductive- biased				
LHCB4	Agub_g12868	1001.73	2591.74	-1.566	reproductive- biased	Vocar.0021s0124	Cre17.g720250		
Core photosynthesis, PSII									
PSBY	Agub_g11432	261.44	803.52	-1.815	reproductive- biased	Vocar.0028s0156	Cre10.g452100	(Zones et al., 2015)	
PSBX	Agub_g11517	1070.21	2841.52	-1.604	reproductive- biased	Vocar.0004s0497	Cre02.g082750	(Zones et al., 2015)	
PSBPQ	Agub_g4054	962.33	2093.64	-1.316	reproductive- upregulated	Vocar.0065s0007	Cre08.g372450	(Zones et al., 2015)	
PSBP1	Agub_g9918	969.90	2095.47	-1.306	reproductive- upregulated		Cre12.g550850	(Zones et al., 2015)	
PSBW	Agub_g6979	797.70	1922.60	-1.463	reproductive- biased	Vocar.0027s0094		(Zones et al., 2015)	
PSBR	Agub_g3093	1304.05	2625.23	-1.204	reproductive- upregulated	Vocar.0002s0459	Cre06.g261000	(Zones et al., 2015)	
PSBPO	Agub_g6653	1731.67	2920.66	-0.948	reproductive- upregulated	Vocar.0009s0198	Cre09.g396213	(Zones et al., 2015)	
PSBS ^c	Agub_g15171	3.48	6.99	-1.184	low confidence	Vocar.0007s0069	Cre01.g016600 (PSBS1); Cre01.g016750 (PSBS2)	(Allen et al., 2011)	

Table 4-9: Continued.

Core photosynthesis, b6f								
PETC	Agub_g6336	347.96	850.80	-1.485	reproductive-biased	Vocar.0009s0349	Cre11.g467689	(Zones et al., 2015)
PETM	Agub_g9725	143.33	551.53	-2.141	reproductive-biased	Vocar.0007s0157	Cre12.g546150	(Zones et al., 2015)
PETOc	Agub_g547	164.00	571.19	-1.996	reproductive-biased	Vocar.0001s1168	Cre12.g558900	(Zones et al., 2015)
Core photosynthesis, LHCI								
LHCA1	Agub_g13241	1258.02	3119.16	-1.505	reproductive-biased	Vocar.0002s0071	Cre06.g283050	(Zones et al., 2015)
LHCA2	Agub_g5995	454.20	1050.83	-1.405	reproductive-biased	Vocar.0016s0243	Cre12.g508750	(Zones et al., 2015)
	Agub_g7947	9.82	20.35	-1.250	reproductive-upregulated			
LHCA3	Agub_g6442	779.34	1539.56	-1.177	reproductive-upregulated	Vocar.0009s0370 (LHCA3.1); Vocar.0015s0072 (LHCA3.2)	Cre11.g467573	(Zones et al., 2015)
LHCA4	Agub_g11433	430.65	1044.13	-1.472	reproductive-biased	Vocar.0028s0157	Cre10.g452050	(Zones et al., 2015)
LHCA5	Agub_g2119	542.74	1658.58	-1.808	reproductive-biased	Vocar.0005s0361	Cre10.g425900	(Zones et al., 2015)
LHCA6	Agub_g3243	363.08	1058.98	-1.740	reproductive-biased	Vocar.0002s0281	Cre06.g278213	(Zones et al., 2015)
LHCA7	Agub_g15062	436.52	1327.21	-1.799	reproductive-biased	Vocar.0023s0180	Cre16.g687900	(Zones et al., 2015)
LHCA8	Agub_g4431	653.63	1249.88	-1.131	reproductive-upregulated	Vocar.0002s0333	Cre06.g272650	(Zones et al., 2015)
LHCA9	Agub_g5310	449.32	1303.55	-1.732	reproductive-biased	Vocar.0011s0017	Cre07.g344950	(Zones et al., 2015)
Core photosynthesis, PSI								
PSAN	Agub_g11513	425.38	1129.75	-1.604	reproductive-biased	Vocar.0004s0500	Cre02.g082500	(Zones et al., 2015)
PSAI	Agub_g2096	529.19	1174.60	-1.344	reproductive-biased	Vocar.0005s0152	Cre03.g165100	(Zones et al., 2015)
PSAG	Agub_g686	560.46	1345.69	-1.459	reproductive-biased	Vocar.0001s1032	Cre12.g560950	(Zones et al., 2015)
PSAK	Agub_g3499	510.70	1110.33	-1.314	reproductive-upregulated	Vocar.0024s0047	Cre17.g724300	(Zones et al., 2015)
PSAH	Agub_g14344	224.83	610.18	-1.634	reproductive-biased	Vocar.0031s0002	Cre07.g330250	(Zones et al., 2015)
PSAL	Agub_g8033	607.92	1511.06	-1.509	reproductive-biased	Vocar.0016s0158	Cre12.g486300	(Zones et al., 2015)
PSAO	Agub_g3749	288.49	1037.28	-2.044	reproductive-biased	Vocar.0013s0247	Cre07.g334550	(Zones et al., 2015)
PSAD	Agub_g12158	1112.32	1470.18	-0.595	reproductive-upregulated	Vocar.0017s0129	Cre05.g238332	(Zones et al., 2015)
PSAF	Agub_g14488	1165.25	1823.15	-0.840	reproductive-upregulated	Vocar.0006s0098	Cre09.g412100	(Zones et al., 2015)
Core photosynthesis, electron transport								
PETE ^c	Agub_g1623	509.08	1667.13	-1.908	reproductive-biased	Vocar.0015s0197 Vocar.0015s0273	Cre03.g182551	(Allen et al., 2011)
PETF ^c	Agub_g15387	825.06	2530.53	-1.812	reproductive-biased	Vocar.0022s0106	Cre14.g626700	(Allen et al., 2011)
PETH ^c	Agub_g7224	183.29	380.49	-1.247	reproductive-upregulated	Vocar.0027s0034	Cre11.g476750	(Allen et al., 2011)
Core photosynthesis, ATP synthase								
ATPC	Agub_g4150	1183.35	1513.50	-0.547	reproductive-upregulated	Vocar.0002s0456	Cre06.g259900	(Zones et al., 2015)
ATPG	Agub_g7279	417.97	634.12	-0.795	reproductive-upregulated	Vocar.0048s0019	Cre11.g481450	(Zones et al., 2015)
Assembly factors of PSII								
Pitt	Agub_g11584	4.01	0.56	2.523	low confidence	Vocar.0007s0029	Cre12.g554103	(Nickelsen and Rengstl, 2013)
Psb27/LPA19	Agub_g9391	30.82	45.23	-0.742	reproductive-upregulated	Vocar.0056s0023	Cre02.g073850	(Nickelsen and Rengstl, 2013)
	Agub_g6042	113.29	203.53	-1.036	reproductive-upregulated	Vocar.0055s0028	Cre05.g243800	(Nickelsen and Rengstl, 2013)
Psb29/ThfI	Agub_g4079	195.83	306.61	-0.840	reproductive-upregulated	Vocar.0028s0018	Cre13.g562850	(Nickelsen and Rengstl, 2013)

Table 4-9: Continued.

Alb3/Alb3.1	Agub_g4118	72.77	46.57	0.451	constitutive	Vocar.0038s0039	Cre06.g251900	(Nickelsen and Rengstl, 2013)
PAM68/SII0933	Agub_g14292	6.78	1.21	2.298	somatic-upregulated	Vocar.0017s0026	Cre07.g329000	(Nickelsen and Rengstl, 2013)
Slr1768	Agub_g4837	115.20	149.50	-0.568	reproductive-upregulated	Vocar.0006s0255	Cre12.g519350	(Nickelsen and Rengstl, 2013)
Alb3.2/Slr1471	Agub_g11564	44.96	35.15	0.164	constitutive	Vocar.0004s0298	Cre17.g729800	(Nickelsen and Rengstl, 2013)
FKBP20-2	Agub_g9198	2.32	3.33	-0.680	constitutive	Vocar.0001s0168	Cre13.g577850	(Nickelsen and Rengstl, 2013)
TLP18.3/Psb32	Agub_g1662	50.39	157.49	-1.840	reproductive-biased	Vocar.0015s0285	Cre03.g182150	(Nickelsen and Rengstl, 2013)
LTO1	Agub_g7944	10.39	5.48	0.704	constitutive	Vocar.0031s0029	Cre12.g493150	(Nickelsen and Rengstl, 2013)
Deg1	Agub_g4648	34.12	6.88	2.133	somatic-biased	Vocar20014026m	Cre12.g498500	(Nickelsen and Rengstl, 2013)
CYP38	Agub_g1336	30.67	87.97	-1.711	reproductive-biased	Vocar.0035s0034	Cre03.g189800	(Nickelsen and Rengstl, 2013)
CtpA	Agub_g3186	15.30	13.31	0.010	constitutive	Vocar.0016s0059	Cre10.g420550	(Nickelsen and Rengstl, 2013)
	Agub_g10947	10.34	0.62	3.838	somatic-biased			
Psb28	Agub_g3306	61.03	38.32	0.470	constitutive	Vocar.0001s1283	Cre10.g440450	(Nickelsen and Rengstl, 2013)
LPA3	Agub_g1465	10.88	28.55	-1.584	reproductive-upregulated	Vocar.0015s0236	Cre03.g184550	(Nickelsen and Rengstl, 2013)
Slr1768	Agub_g4837	115.20	149.50	-0.568	reproductive-upregulated	Vocar.0006s0255	Cre12.g519350	(Nickelsen and Rengstl, 2013)
Assembly factors of b6f								
CCB4	Agub_g11329	5.22	0.86	2.261	low confidence	Vocar.0042s0052	Cre08.g382300	(Schöttler et al., 2015)
CCDA	Agub_g5385	37.41	13.32	1.297	somatic-upregulated	Vocar.0051s0044	Cre07.g357800	(Schöttler et al., 2015)
TCA1	Agub_g5830	0.14	0.21	-0.749	constitutive	Vocar.0011s0287	Cre09.g415500	(Schöttler et al., 2015)
CCS5	Agub_g10435	20.36	9.60	0.901	constitutive	Vocar.0004s0216	Cre17.g702150	(Schöttler et al., 2015)
CCB2	Agub_g3979	5.19	1.72	1.371	low confidence	Vocar.0034s0004	Cre12.g537850	(Schöttler et al., 2015)
CCS1	Agub_g9236	49.41	5.36	3.030	somatic-biased	Vocar.0001s0213	Cre13.g575000	(Schöttler et al., 2015)
Q	Agub_g6065	8.82	3.44	1.138	low confidence	Vocar.0013s0082	Cre01.g000850	(Schöttler et al., 2015)
MCA1	Agub_g11939	65.95	57.04	0.021	constitutive	Vocar.0020s0117	Cre08.g358250	(Schöttler et al., 2015)
Assembly factors of PSI								
PPD1	Agub_g14696	4.46	3.38	0.186	constitutive	Vocar.0030s0036	Cre08.g362900	(Yang et al., 2015)
Psa2	Agub_g3006	67.23	61.71	-0.069	constitutive	Vocar.0027s0086	Cre11.g475850	(Yang et al., 2015)
Hcf101	Agub_g10094	4.88	7.39	-0.794	constitutive	Vocar.0001s0627	Cre01.g045902	(Yang et al., 2015)
Ycf3	Agub_g13698	1.12	1.18	-0.317	low expression	Vocar.0007s0325	Cre12.g552851	(Yang et al., 2015)
Ycf37/Pyg7	Agub_g5063	23.94	41.08	-0.963	constitutive	Vocar.0037s0004	Cre12.g524300	(Yang et al., 2015)
Y3IP1	Agub_g4415	4.75	5.64	-0.424	constitutive	Vocar.0002s0069	Cre06.g280650	(Yang et al., 2015)
Alb3	Agub_g4118	72.77	46.57	0.451	constitutive	Vocar.0038s0039	Cre06.g251900	(Yang et al., 2015)
CnfU	Agub_g13525	14.84	21.79	-0.760	constitutive	Vocar.0008s0312	Cre18.g748447	(Yang et al., 2015)
VIPP1	Agub_g9065	123.30	113.88	-0.070	constitutive	Vocar.0001s0066	Cre13.g583550	(Yang et al., 2015)
RubA	Agub_g8819	56.86	55.30	-0.143	constitutive	Vocar.0003s0094 (RubA); Vocar.0072s0015 (RubA.2)	Cre07.g315150	(Yang et al., 2015)
Assembly factors of ATP synthase								
TDA1	Agub_g11711	1.00	0.44	0.992	low expression	Vocar20002206m	Cre08.g358350	(Schöttler et al., 2015)

Table 4-9: Continued.

Tetrapyrrole biosynthesis								
GTS1	Agub_g4005	85.43	128.75	-0.788	reproductive-upregulated	Vocar.0014s0250	Cre11.g467547	(Zones et al., 2015)
GSA1	Agub_g3377	12.50	11.85	-0.117	constitutive	Vocar.0019s0198	Cre03.g158000	(Zones et al., 2015)
PBGD1	Agub_g7418	23.83	22.94	-0.140	constitutive	Vocar.0041s0034	Cre16.g663900	(Zones et al., 2015)
UROD3	Agub_g9393	7.22	5.60	0.162	constitutive	Vocar.0056s0027	Cre02.g073700	(Zones et al., 2015)
GTR	Agub_g12658	64.89	116.97	-1.039	reproductive-upregulated	Vocar.0067s0036	Cre07.g342150	(Zones et al., 2015)
ALAD	Agub_g742	202.95	79.91	1.159	somatic-upregulated	Vocar.0001s1205	Cre02.g091050	(Zones et al., 2015)
CPX1	Agub_g5720	46.85	108.75	-1.396	reproductive-upregulated	Vocar.0028s0183	Cre02.g085450	(Zones et al., 2015)
CPX2	Agub_g3260	11.44	15.50	-0.631	constitutive	Vocar.0001s1192	Cre02.g092600	(Zones et al., 2015)
PBGD2	Agub_g8392	16.65	12.93	0.157	constitutive	Vocar.0009s0001	Cre02.g113850	(Zones et al., 2015)
PPX1	Agub_g15114	10.22	5.41	0.729	constitutive	Vocar.0008s0226	Cre09.g396300	(Zones et al., 2015)
UROS	Agub_g817	13.71	8.15	0.583	constitutive	Vocar.0036s0010	Cre09.g409100	(Zones et al., 2015)
GTS2	Agub_g9037	24.23	25.09	-0.244	constitutive	Vocar.0003s0007	Cre07.g313700	(Zones et al., 2015)
UROD2	Agub_g10210	37.55	20.45	0.683	constitutive	Vocar.0001s1599	Cre02.g076300	(Zones et al., 2015)
Chlorophyll biosynthesis								
CHLH2	Agub_g3011	18.66	5.95	1.461	somatic-upregulated	Vocar.0027s0037	Cre11.g477625	(Zones et al., 2015)
GUN4	Agub_g5539	9.71	5.97	0.506	constitutive	Vocar.0022s0158	Cre05.g246800	(Zones et al., 2015)
POR1	Agub_g12615	20.30	67.90	-1.928	reproductive-biased	Vocar.0011s0096	Cre01.g015350	(Zones et al., 2015)
CHLH1	Agub_g8978	27.20	43.27	-0.860	reproductive-upregulated	Vocar.0003s0179	Cre07.g325500	(Zones et al., 2015)
CRD1	Agub_g13724	155.76	295.71	-1.118	reproductive-upregulated	Vocar.0005s0072	Cre07.g346050	(Zones et al., 2015)
HCAR	Agub_g6509	18.93	33.22	-1.006	reproductive-upregulated	Vocar.0012s0189	Cre11.g468700	(Zones et al., 2015)
CHLD	Agub_g5979	7.23	9.09	-0.542	constitutive	Vocar.0055s0003	Cre05.g242000	(Zones et al., 2015)
DVR1	Agub_g186	11.74	11.25	-0.137	constitutive	Vocar.0001s0674	Cre01.g042800	(Zones et al., 2015)
CHL11	Agub_g11113	13.12	18.63	-0.702	constitutive	Vocar.0015s0053	Cre06.g306300	(Zones et al., 2015)
CHL12	Agub_g6227	15.12	8.09	0.722	constitutive	Vocar.0016s0202	Cre12.g510800	(Zones et al., 2015)
CHLM	Agub_g4647	18.54	13.43	0.282	constitutive	Vocar.0006s0270	Cre12.g498550	(Zones et al., 2015)
CAO1	Agub_g218	38.46	31.29	0.107	constitutive	Vocar.0001s0705	Cre01.g043350	(Zones et al., 2015)

^a Vocar#####m (v2.0); Vocar.#####s#### (v2.1).

^b Loci in v5.6 genome of *C. reinhardtii*

^c Not included in Matt & Umen (2018) but added in this study.

Table 4-10: List of expression of carbon metabolism genes in *Astrephomene gubernaculifera*.

gene name	locus	soma TPM	repr TPM	log2 fold change	classification	<i>V. carteri</i> homolog ^a	<i>C. reinhardtii</i> homolog ^b	reference
CBB cycle								
RBCS	Agub_g4502	256.81	706.23	-1.651	reproductive- biased	Vocar.0033s0070 (RBCS1); Vocar.0033s0068 (RBCS2); Vocar.0033s0067 (RBCS3); Vocar.0033s0069 (RBCS4); Vocar.0033s0072 (RBCS5); Vocar.0033s0065 (RBCS6); Vocar.0033s0066 (RBCS7)	Cre02.g120100 (RBCS1); Cre02.g120150 (RBCS2)	(Zones et al., 2015)
	Agub_g8284	320.07	579.04	-1.049	reproductive- upregulated			
	Agub_g8285	524.07	1215.42	-1.405	reproductive- biased			
PGK1	Agub_g6547	999.30	1317.19	-0.591	reproductive- upregulated	Vocar.0030s0188	Cre11.g467770	(Zones et al., 2015)
GAP3	Agub_g8638	446.46	996.89	-1.354	reproductive- biased	Vocar.0013s0021	Cre01.g010900	(Zones et al., 2015)
FBA1	Agub_g12196	12.75	22.68	-1.026	reproductive- upregulated	Vocar.0024s0211	Cre01.g006950	(Zones et al., 2015)
FBA3	Agub_g6901	2776.18	3950.25	-0.702	reproductive- upregulated	Vocar.0008s0418	Cre05.g234550	(Zones et al., 2015)
FBP2	Agub_g6225	22.78	50.05	-1.336	reproductive- upregulated	Vocar.0016s0204	Cre12.g510650	(Zones et al., 2015)
TRK1 ^c	Agub_g11422	107.73	373.14	-1.988	reproductive- biased	Vocar.0004s0460	Cre02.g080200	(Zones et al., 2015)
RPE1	Agub_g6235	204.86	407.85	-1.191	reproductive- upregulated	Vocar.0016s0191	Cre12.g511900	(Zones et al., 2015)
PRK1	Agub_g13663	299.04	959.77	-1.876	reproductive- biased	Vocar.0007s0361	Cre12.g554800	(Zones et al., 2015)
SBP1	Agub_g1004	184.39	313.85	-0.961	reproductive- upregulated	Vocar.0015s0215	Cre03.g185550	(Zones et al., 2015)
RPI1	Agub_g1580	51.59	227.41	-2.334	reproductive- biased	Vocar.0011s0167	Cre03.g187450	(Zones et al., 2015)
CP12	Agub_g11293	852.28	1746.25	-1.228	reproductive- upregulated	Vocar.0042s0039	Cre08.g380250	(Zones et al., 2015)
RAF1	Agub_g11075	18.98	17.49	-0.076	constitutive	Vocar.0030s0064	Cre06.g308450	(Zones et al., 2015)
RCA1	Agub_g2594	71.15	270.74	-2.126	reproductive- biased	Vocar.0001s1749	Cre04.g229300	(Zones et al., 2015)
Starch biosynthesis								
PGI1	Agub_g1809	384.15	532.84	-0.666	reproductive- upregulated	Vocar.0015s0194	Cre03.g175400	(Gargouri et al., 2015)
GPM1	Agub_g3226	462.46	133.62	1.599	somatic- biased	Vocar.0002s0273	Cre06.g278210	(Gargouri et al., 2015)
GPM2	Agub_g13175	48.98	27.79	0.619	somatic- upregulated	Vocar.0045s0067	Cre01.g012600	(Gargouri et al., 2015)
STA1	Agub_g10311	346.69	456.76	-0.591	reproductive- upregulated	Vocar.0001s1701	Cre13.g567950	(Ball and Deschamps, 2009)
AGP2	Agub_g10757	109.97	185.00	-0.945	reproductive- upregulated	Vocar.0018s0105	Cre16.g683450	(Ball and Deschamps, 2009)
STA6	Agub_g1609	117.09	198.48	-0.955	reproductive- upregulated	Vocar.0011s0326	Cre03.g188250	(Goodenough et al., 2014)
SSI1	Agub_g4851	0.99	0.00	1.797	low expression	Vocar.0037s0127	Cre12.g521700	(Gargouri et al., 2015)
SSS2	Agub_g1447	22.96	29.48	-0.552	reproductive- upregulated	Vocar.0015s0224	Cre03.g185250	(Gargouri et al., 2015)
SSS4	Agub_g7450	1.46	0.00	0.932	low expression	Vocar.0041s0007	Cre16.g665800	(Gargouri et al., 2015)
SSS5	Agub_g7421	7.91	13.43	-0.970	constitutive	Vocar.0007s0013	Cre16.g663850	(Gargouri et al., 2015)
STA2	Agub_g3568	285.17	104.91	1.250	somatic- upregulated	Vocar.0001s1467	Cre17.g721500	(Goodenough et al., 2014)
STA3	Agub_g13199	42.90	19.19	0.961	somatic- upregulated	Vocar.0002s0103	Cre06.g282000	(Goodenough et al., 2014)
SBE2	Agub_g4269	30.38	32.08	-0.267	constitutive	Vocar.0002s0315	Cre06.g270100	(Gargouri et al., 2015)

Table 4-10: Continued.

SBE3	Agub_g5907	358.86	264.23	0.250	somatic-upregulated	Vocar.0026s0071	Cre10.g444700	(Gargouri et al., 2015)
ISA1/STA7	Agub_g7114	46.45	30.61	0.411	constitutive	Vocar.0014s0023	Cre03.g155001	(Hemschemeier et al., 2013)
ISA2	Agub_g12806	19.54	23.27	-0.449	constitutive	Vocar.0004s0314	Cre17.g698850	(Hemschemeier et al., 2013)
ISA3	Agub_g977	62.68	52.57	0.061	constitutive	Vocar.0013s0145	Cre03.g207713	(Hemschemeier et al., 2013)
PHOA	Agub_g12544	182.93	172.85	-0.112	constitutive	Vocar.0023s0216	Cre07.g336950	(Blaby et al., 2015)
PHOB	Agub_g9829	593.35	623.67	-0.264	reproductive-upregulated	Vocar.0007s0315	Cre12.g552200	(Blaby et al., 2015)
Starch degradation								
PWD1	Agub_g12852	23.12	21.73	-0.102	constitutive	Vocar.0004s0171	Cre17.g719900	(Ball and Deschamps, 2009)
GWD1	Agub_g8979	23.49	10.37	0.994	somatic-upregulated	Vocar.0003s0059	Cre07.g319300	(Ball and Deschamps, 2009)
GWD2	Agub_g2521	30.84	23.30	0.215	constitutive	Vocar.0009s0098	Cre07.g332300	(Ball and Deschamps, 2009)
GWD3	Agub_g14708	2.26	0.00	3.648	low expression	Vocar20013571m	Cre08.g363874	(Ball and Deschamps, 2009)
AMA1	Agub_g6794	1.13	0.99	0.007	constitutive	Vocar.0008s0085	Cre08.g385500	(Gargouri et al., 2015)
AMA2	Agub_g11888	53.83	89.96	-0.934	reproductive-upregulated	Vocar.0021s0194	Cre08.g362450	(Gargouri et al., 2015)
AMB1	Agub_g15406	48.43	30.70	0.469	constitutive	Vocar.0015s0109	Cre06.g307150	(Gargouri et al., 2015)
AMB2	Agub_g14684	9.41	21.49	-1.377	reproductive-upregulated	Vocar.0003s0276	Cre06.g270350	(Gargouri et al., 2015)
AMB3	Agub_g197	109.89	59.82	0.686	somatic-upregulated	Vocar.0001s0659	Cre01.g044100	(Gargouri et al., 2015)
DPE2	Agub_g10651	17.24	18.18	-0.268	constitutive	Vocar.0027s0145	Cre02.g095126	(Ball and Deschamps, 2009)
HXK1	Agub_g8351	62.92	48.14	0.192	constitutive	Vocar.0009s0046	Cre02.g117500	(Gargouri et al., 2015)
Glycolysis/gluconeogenesis								
PFK1	Agub_g4578	62.48	77.35	-0.504	reproductive-upregulated	Vocar.0002s0496	Cre06.g262900	(Zones et al., 2015)
PFK2	Agub_g13689	80.68	89.80	-0.346	constitutive	Vocar.0007s0331	Cre12.g553250	(Zones et al., 2015)
FBP1	Agub_g12627	2899.64	2179.99	0.218	somatic-upregulated	Vocar.0005s0037	Cre07.g338451	(Zones et al., 2015)
FBA4	Agub_g8300	12.00	13.06	-0.318	constitutive	Vocar.0033s0026	Cre02.g115650	(Zones et al., 2015)
GAP1	Agub_g7971	37.81	69.49	-1.069	reproductive-upregulated	Vocar.0031s0009	Cre12.g485150	(Zones et al., 2015)
	Agub_g7985	14.20	17.75	-0.516	constitutive			
GAP2	Agub_g915	204.36	211.15	-0.240	constitutive	Vocar.0017s0057	Cre07.g354200	(Zones et al., 2015)
GAPN1 ^c	Agub_g3067	110.67	108.61	-0.165	constitutive	Vocar.0001s1126	Cre12.g556600	(Zones et al., 2015)
PGK2	Agub_g914	5.76	7.79	-0.620	constitutive	Vocar.0017s0058	Cre07.g354250	(Zones et al., 2015)
PGM4	Agub_g8815	8.96	3.82	1.066	somatic-upregulated	Vocar.0003s0446	Cre05.g232550	(Zones et al., 2015)
PGH1	Agub_g4701	1744.11	2173.66	-0.510	reproductive-upregulated	Vocar.0006s0398	Cre12.g513200	(Zones et al., 2015)
PYK1	Agub_g5224	51.90	44.46	0.024	constitutive	Vocar.0028s0006	Cre12.g533550	(Zones et al., 2015)
PYK2	Agub_g13351	30.25	12.47	1.094	somatic-upregulated	Vocar.0002s0096	Cre06.g280950	(Zones et al., 2015)
PYK3	Agub_g6874	79.68	67.95	0.037	constitutive	Vocar.0008s0405	Cre05.g234700	(Zones et al., 2015)
PYK4	Agub_g2966	55.09	27.26	0.825	somatic-upregulated	Vocar.0014s0005	Cre03.g144847	(Zones et al., 2015)
PYK5	Agub_g9331	20.03	37.67	-1.108	reproductive-upregulated	Vocar.0044s0014	Cre02.g147900	(Zones et al., 2015)
MDH3	Agub_g10057	281.46	144.83	0.765	somatic-upregulated	Vocar.0054s0004	Cre02.g145800	(Zones et al., 2015)

Table 4-10: Continued.

PCK1,2,3	Agub_g10241	2225.20	2236.02	-0.200	reproductive-upregulated	Vocar.0021s0021 (PCK1); Vocar.0008s0023 (PCK2); Vocar.0021s0022 (PCK3)	Cre02.g141400	(Zones et al., 2015)
	Agub_g12788	59.31	136.47	-1.397	reproductive-biased			
	Agub_g12789	11.45	15.12	-0.589	constitutive			
PEPC2	Agub_g1873	19.53	14.33	0.254	constitutive	Vocar.0015s0128	Cre03.g171950	(Zones et al., 2015)
TCA cycle								
PDC1	Agub_g12551	165.46	233.92	-0.693	reproductive-upregulated	Vocar.0023s0211	Cre07.g337650	(Burgess et al., 2016)
PDH1a	Agub_g5406	193.42	190.75	-0.171	constitutive	Vocar.0023s0019	Cre16.g677026	(Burgess et al., 2016)
DLA1	Agub_g13504	114.08	118.36	-0.246	constitutive	Vocar.0008s0302	Cre09.g386735	(Burgess et al., 2016)
DLD1	Agub_g7169	340.23	570.17	-0.939	reproductive-upregulated	Vocar.0027s0029	Cre18.g749847	(Burgess et al., 2016)
CIS1	Agub_g4741	677.48	890.74	-0.589	reproductive-upregulated	Vocar.0006s0313	Cre12.g514750	(Zones et al., 2015)
ACH1	Agub_g190	3932.60	4894.31	-0.509	reproductive-upregulated	Vocar.0001s0673	Cre01.g042750	(Zones et al., 2015)
IDH1	Agub_g11448	36.51	49.57	-0.629	constitutive	Vocar.0004s0404	Cre17.g728800	(Zones et al., 2015)
IDH2	Agub_g10303	691.11	1161.68	-0.942	reproductive-upregulated	Vocar.0054s0043	Cre02.g143250	(Zones et al., 2015)
IDH3	Agub_g2587	1771.99	1767.69	-0.189	constitutive	Vocar.0025s0166	Cre04.g214500	(Zones et al., 2015)
OGD1	Agub_g3987	892.65	1042.84	-0.417	reproductive-upregulated	Vocar.0043s0062	Cre12.g537200	(Knoop et al., 2013)
OGD2	Agub_g5333	284.36	397.72	-0.676	reproductive-upregulated	Vocar.0067s0006	Cre07.g343700	(Knoop et al., 2013)
DLD2	Agub_g15175	53.33	83.37	-0.835	reproductive-upregulated	Vocar.0007s0425	Cre01.g016514	(Knoop et al., 2013)
DLD1	Agub_g7169	340.23	570.17	-0.939	reproductive-upregulated	Vocar.0027s0029	Cre18.g749847	(Knoop et al., 2013)
SCLA1	Agub_g1273	1214.08	1757.81	-0.726	reproductive-upregulated	Vocar.0035s0094	Cre03.g193850	(Zones et al., 2015)
SCLB1	Agub_g10481	1283.06	1871.64	-0.737	reproductive-upregulated	Vocar.0004s0200	Cre17.g703700	(Zones et al., 2015)
SDH1	Agub_g12858	415.70	541.24	-0.574	reproductive-upregulated	Vocar.0004s0240	Cre14.g619133	(Zones et al., 2015)
SDH2	Agub_g3180	786.86	1117.54	-0.698	reproductive-upregulated	Vocar.0002s0549	Cre06.g264200	(Zones et al., 2015)
SDH3	Agub_g13151	974.05	1925.74	-1.177	reproductive-upregulated	Vocar.0045s0012	Cre01.g020350	(Zones et al., 2015)
FUM1	Agub_g4228	287.26	226.11	0.154	constitutive	Vocar.0007s0256	Cre06.g254400	(Zones et al., 2015)
MDH4	Agub_g14825	2367.07	2543.01	-0.297	reproductive-upregulated	Vocar.0031s0114	Cre12.g483950	(Zones et al., 2015)
ACK2	Agub_g12774	1255.92	2023.14	-0.882	reproductive-upregulated	Vocar.0026s0122	Cre17.g709850	(Zones et al., 2015)
PAT1 ^c	Agub_g11086	2026.59	2515.28	-0.505	reproductive-upregulated	Vocar.0004s0311	Cre17.g699000	(Zones et al., 2015)
ACS1	Agub_g9933	270.05	361.90	-0.615	reproductive-upregulated	Vocar.0001s0493	Cre01.g071662	(Zones et al., 2015)
ACLB1	Agub_g780	39.55	34.94	-0.009	constitutive	Vocar.0001s1237	Cre02.g088600	(Zones et al., 2015)
Glyoxylate cycle								
ACS2	Agub_g10030	24.58	54.08	-1.333	reproductive-upregulated	Vocar.0001s0439	Cre01.g055408	(Zones et al., 2015)
ACS3	Agub_g13891	1117.19	1705.20	-0.803	reproductive-upregulated	Vocar.0005s0104	Cre07.g353450	(Zones et al., 2015)
ACH1	Agub_g190	3932.60	4894.31	-0.509	reproductive-upregulated	Vocar.0001s0673	Cre01.g042750	(Zones et al., 2015)
CIS2	Agub_g2818	294.09	538.23	-1.066	reproductive-upregulated	Vocar.0014s0142	Cre03.g149100	(Zones et al., 2015)
ICL1	Agub_g13246	5432.45	4517.07	0.073	constitutive	Vocar.0002s0037	Cre06.g282800	(Zones et al., 2015)

Table 4-10: Continued.

MAS1	Agub_g6115	997.40	918.32	-0.073	constitutive	Vocar.0014s0003	Cre03.g144807	(Zones et al., 2015)
	Agub_g2965	17.85	44.24	-1.504	reproductive-upregulated			
MDH2	Agub_g14225	61.83	99.14	-0.871	reproductive-upregulated	Vocar.0009s0119	Cre10.g423250	(Zones et al., 2015)
Fatty acid biosynthesis								
PDC2	Agub_g15673	270.71	400.44	-0.756	reproductive-upregulated	Vocar.0012s0101	Cre02.g099850	(Burgess et al., 2016)
PDH2	Agub_g1280	118.21	264.59	-1.357	reproductive-upregulated	Vocar.0035s0072	Cre03.g194200	(Burgess et al., 2016)
DLA2	Agub_g14790	208.00	316.31	-0.797	reproductive-upregulated	Vocar.0019s0109	Cre03.g158900	(Burgess et al., 2016)
ACX1	Agub_g4820	45.55	118.29	-1.568	reproductive-biased	Vocar.0031s0113	Cre12.g484000	(Li-Beisson et al., 2015)
BCX1	Agub_g14823	34.74	185.55	-2.613	reproductive-biased	Vocar.0026s0109	Cre17.g715250	(Li-Beisson et al., 2015)
BCC2	Agub_g99	51.55	157.57	-1.807	reproductive-biased	Vocar.0001s0827	Cre01.g037850	(Li-Beisson et al., 2015)
BCR1	Agub_g11876	114.87	318.07	-1.665	reproductive-biased	Vocar.0003s0267	Cre08.g359350	(Li-Beisson et al., 2015)
ACC1	Agub_g5120	12.34	12.49	-0.205	constitutive	Vocar.0043s0027	Cre08.g373050	(Li-Beisson et al., 2015)
ACP2	Agub_g14473	787.60	2806.40	-2.028	reproductive-biased	Vocar.0006s0130	Cre13.g577100	(Li-Beisson et al., 2015)
MCT1	Agub_g12761	49.16	70.78	-0.723	reproductive-upregulated	Vocar.0004s0113	Cre14.g621650	(Li-Beisson et al., 2015)
KAS1	Agub_g10637	202.18	540.43	-1.613	reproductive-biased	Vocar.0027s0125	Cre11.g467723	(Li-Beisson et al., 2015)
KAS2	Agub_g3761	17.34	54.56	-1.843	reproductive-biased	Vocar.0013s0258	Cre07.g335300	(Li-Beisson et al., 2015)
KAS3	Agub_g7678	15.38	15.66	-0.227	constitutive	Vocar.0013s0277	Cre04.g216950	(Li-Beisson et al., 2015)
HAD1	Agub_g1346	45.95	209.67	-2.389	reproductive-biased	Vocar.0011s0181	Cre03.g208050	(Li-Beisson et al., 2015)
KAR1	Agub_g1874	40.86	134.08	-1.908	reproductive-biased	Vocar.0015s0127	Cre03.g172000	(Li-Beisson et al., 2015)
ENR1	Agub_g11997	79.37	192.95	-1.476	reproductive-upregulated	Vocar.0001s1644	Cre06.g294950	(Li-Beisson et al., 2015)
Fatty acid beta-oxidation								
CTS	Agub_g14150	29.89	40.03	-0.616	reproductive-upregulated	Vocar.0071s0022	Cre15.g637761	(Li-Beisson et al., 2015)
LCS3	Agub_g8568	25.05	45.13	-1.041	reproductive-upregulated	Vocar.0016s0276	Cre12.g507400	(Li-Beisson et al., 2015)
ACX1	Agub_g5369	75.88	96.18	-0.534	reproductive-upregulated	Vocar.0023s0122	Cre16.g689050	(Li-Beisson et al., 2015)
ACX2	Agub_g14059	94.87	160.81	-0.957	reproductive-upregulated	Vocar.0003s0447	Cre05.g232002	(Li-Beisson et al., 2015)
ACX3	Agub_g12620	17.18	24.84	-0.723	reproductive-upregulated	Vocar.0011s0091	Cre16.g687350	(Li-Beisson et al., 2015)
ACX4	Agub_g7786	23.19	26.91	-0.405	constitutive	Vocar.0012s0263	Cre16.g695100	(Li-Beisson et al., 2015)
ECH1	Agub_g7785	141.47	290.17	-1.231	reproductive-upregulated	Vocar.0012s0264	Cre16.g695050	(Li-Beisson et al., 2015)
ATO1	Agub_g3530	276.83	358.87	-0.568	reproductive-upregulated	Vocar.0024s0089	Cre17.g723650	(Li-Beisson et al., 2015)
Amino acid biosynthesis								
AST1	Agub_g8246	149.95	139.00	-0.083	constitutive	Vocar.0033s0131	Cre09.g387726	(Vallon and Spalding, 2009)
AST2	Agub_g15711	22.19	10.32	0.918	somatic-upregulated	Vocar.0001s0411	Cre01.g051800	(Vallon and Spalding, 2009)
AST3	Agub_g15515	147.22	660.52	-2.360	reproductive-biased	Vocar.0012s0085	Cre02.g097900	(Vallon and Spalding, 2009)
AST5	Agub_g3053	2.55	2.64	-0.245	constitutive	Vocar.0002s0199	Cre06.g278249	(Vallon and Spalding, 2009)
GSF1	Agub_g4716	16.94	18.90	-0.353	constitutive	Vocar.0006s0290	Cre12.g514050	(Fernandez et al., 2009)
GSN1	Agub_g11798	82.44	170.64	-1.243	reproductive-upregulated	Vocar.0064s0005	Cre13.g592200	(Fernandez et al., 2009)
GS2	Agub_g5148	42.77	119.27	-1.675	reproductive-biased	Vocar.0028s0089	Cre12.g530650	(Fernandez et al., 2009)

Table 4-10: Continued.

GS1	Agub_g11963	1261.72	1408.82	-0.352	reproductive-upregulated	Vocar.0001s1591	Cre02.g113200	(Fernandez et al., 2009)
GS3	Agub_g1497	22.78	13.59	0.552	somatic-upregulated	Vocar.0011s0254 (GS3); Vocar.0028s0090 (GS4)	Cre03.g207250	(Fernandez et al., 2009)
ASNS	Agub_g6137	114.95	184.21	-0.873	reproductive-upregulated	Vocar.0013s0043	Cre01.g004300	(Vallon and Spalding, 2009)
Amino acid degradation								
GDH1	Agub_g9528	276.52	456.04	-0.914	reproductive-upregulated	Vocar.0003s0460	Cre05.g232150	(Fontaine et al., 2012)
GDH2	Agub_g9517	158.61	107.52	0.375	constitutive	Vocar.0033s0159	Cre09.g388800	(Fontaine et al., 2012)

^a Vocar#####m (v2.0); Vocar.####s#### (v2.1).

^b Loci in v5.6 genome of *C. reinhardtii*

^c Not included in Matt & Umen (2018) but added in this study.

Table 4-11: List of expression of nucleotide sugar metabolism genes and glycosyltransferases in *Astrephomene gubernaculifera*.

gene name	locus	soma TPM	repr TPM	log2 fold change	classification	<i>V. carteri</i> homolog ^a
nucleotide sugar metabolism						
PGI1	Agub_g1809	384.15	532.84	-0.666	reproductive- upregulated	Vocar.0015s0194
GPM1	Agub_g3226	462.46	133.62	1.599	somatic- biased	Vocar.0002s0273
GPM2	Agub_g13175	48.98	27.79	0.619	somatic- upregulated	Vocar.0045s0067
UGP1	Agub_g2461	193.67	7.85	4.446	somatic- specific	Vocar.0001s1758
UGE1	Agub_g2581	378.13	216.83	0.609	somatic- upregulated	Vocar.0025s0149
UGD1	Agub_g5373	1486.00	539.38	1.270	somatic- upregulated	Vocar.0051s0034 (UGD1); Vocar.0002s0240 (UGD2)
UGD3	Agub_g5178	23.87	16.41	0.347	constitutive	Vocar.0017s0014
UXS1	Agub_g1912	516.66	193.06	1.227	somatic- upregulated	Vocar.0040s0069
RGP1	Agub_g4062	3885.43	1797.29	0.920	somatic- upregulated	Vocar.0028s0033 (RGP1); Vocar.0001s0378 (RGP2); Vocar.0006s0025 (RGP3)
PMI1	Agub_g9330	79.77	36.03	0.958	somatic- upregulated	Vocar.0044s0015
GMP1	Agub_g9603	421.29	314.43	0.228	constitutive	Vocar.0018s0209
GME1	Agub_g13100	963.97	400.80	1.074	somatic- upregulated	Vocar.0007s0276
	Agub_g13494	78.46	139.25	-1.025	reproductive- upregulated	
UGM1	Agub_g4434	594.37	254.21	1.031	somatic- upregulated	Vocar.0002s0336 (UGM1); Vocar.0030s0093 (UGM2); Vocar.0030s0102 (UGM3)
Glycosyltransferase						
HPAT1	Agub_g311	423.62	192.64	0.944	somatic- upregulated	Vocar.0001s0970
HPAT4	Agub_g9430	154.54	69.55	0.958	somatic- upregulated	Vocar.0001s0554
RRA1	Agub_g8736	75.89	31.74	1.064	somatic- upregulated	Vocar.0001s1406
RRA2	Agub_g11974	222.43	121.83	0.676	somatic- upregulated	Vocar.0001s1609
RRA3	Agub_g7126	78.66	37.70	0.866	somatic- upregulated	Vocar.0048s0029
XEG113-1	Agub_g5475	74.13	29.02	1.157	somatic- upregulated	Vocar.0022s0025
XEG113-2	Agub_g14403	96.83	50.30	0.750	somatic- upregulated	Vocar.0034s0083
XEG113-3	Agub_g6062	19.69	18.92	-0.134	constitutive	Vocar.0024s0120
XEG113-5	Agub_g9167	9.31	14.36	-0.821	constitutive	Vocar.0001s0206
	Agub_g9230	12.29	15.30	-0.517	constitutive	
XEG113-6	Agub_g7646	101.63	78.50	0.177	constitutive	Vocar.0013s0264
XEG113-8	Agub_g3705	118.22	45.64	1.186	somatic- upregulated	Vocar.0013s0214
	Agub_g14386	23.19	35.15	-0.793	reproductive- upregulated	
XEG113-like	Agub_g11989	28.72	15.03	0.746	somatic- upregulated	

Table 4-11: Continued.

XEG113-like	Agub_g12685	30.05	18.20	0.532	constitutive	
NDST1	Agub_g9823	23.72	9.37	1.141	somatic-upregulated	Vocar.0007s0318
GALT1	Agub_g4697	8.52	15.02	-1.008	reproductive-upregulated	Vocar.0006s0401
PGT1	Agub_g3854	41.17	26.02	0.477	constitutive	Vocar.0011s0294
SGT1	Agub_g4863	114.39	52.92	0.918	somatic-upregulated	Vocar.0037s0116
INVC	Agub_g3178	8.10	7.09	0.007	constitutive	Vocar.0002s0555
INVCL1	Agub_g12026	9.97	2.97	1.541	somatic-upregulated	Vocar.0063s0023
INVCL2	Agub_g13537	62.95	17.10	1.693	somatic-upregulated	Vocar.0079s0012
INVCL3	Agub_g12509	9.33	15.96	-0.967	somatic-upregulated	Vocar.0004s0167
INVCL4	Agub_g135	18.72	20.50	-0.309	constitutive	Vocar.0001s0747

^a Loci in v2.1 genome of *V. carteri* based on Matt and Umen (2018)

Table 4-12: List of expression of genes involved in multicellular traits in *Astrephomene gubernaculifera*.

locus	soma TPM	repr TPM	log2 fold change	classification
Embryogenesis-related genes				
Agub_g2578 (InvA/IAR1)	117.63	97.12	0.088	constitutive
Agub_g14864 (InvB/IBR1)	149.42	236.66	-0.860	reproductive-upregulated
Agub_g3178 (InvC/ICR1)	8.10	7.09	0.007	constitutive
Agub_g8128 (GlsA/GAR1)	2.42	9.25	-2.140	reproductive-biased
Agub_g14480 (Hsp70A)	2847.76	2321.56	0.102	constitutive
VARL genes				
Agub_g6775 (rlsF1)	3.40	2.25	0.403	constitutive
Agub_g8757 (RLS10/rlsL)	8.39	4.81	0.616	somatic-upregulated
Agub_g12169 (rlsH)	5.96	0.61	3.105	somatic-biased
Agub_g12775 (RLS1/rlsD)	11.38	5.76	0.795	somatic-upregulated
Agub_g12876 (rlsF2)	1.85	3.10	-0.938	reproductive-upregulated
Agub_g12954 (RLS2/rlsI)	2.83	0.97	1.328	low expression
MMP genes				
Agub_g1239	9.91	1.07	3.007	low confidence
Agub_g2744	10.30	9.56	-0.081	constitutive
Agub_g2953	8.11	8.01	-0.166	constitutive
Agub_g2954	637.13	127.31	2.131	somatic-biased
Agub_g6282	404.95	110.24	1.685	somatic-biased
Agub_g6672	46.80	95.19	-1.221	reproductive-upregulated
Agub_g8972	34.40	18.21	0.730	somatic-upregulated
Agub_g10447	58.47	54.32	-0.085	constitutive
Agub_g10832	31.50	16.93	0.709	somatic-upregulated
Agub_g10833	55.42	1.95	4.640	somatic-specific
Agub_g10834	18.50	3.38	2.260	somatic-biased
Agub_g11444	27.00	4.52	2.365	somatic-biased
Agub_g11544	123.12	29.56	1.870	somatic-biased
Agub_g12060	29.24	13.22	0.957	somatic-upregulated
Agub_g12871	110.68	26.11	1.896	somatic-biased
Agub_g13771	1.39	0.00	0.932	low expression
Agub_g14056	13.18	15.24	-0.419	constitutive
Agub_g14212	5.63	0.14	5.145	somatic-biased
Agub_g14455	10.61	14.07	-0.597	constitutive
Agub_g14986	10.85	10.97	-0.221	constitutive
Pherophorin genes				
Agub_g803	12.25	12.44	-0.219	constitutive
Agub_g1106	0.45	0.15	1.421	somatic-upregulated
Agub_g1107	0.82	0.12	2.607	somatic-upregulated
Agub_g2415	136.83	19.38	2.636	somatic-biased
Agub_g2416	0.72	0.28	1.101	low expression
Agub_g2613	276.05	595.66	-1.303	reproductive-biased
Agub_g2619	1.27	0.45	1.266	low confidence
Agub_g2911	51.59	12.94	1.796	somatic-biased

Table 4-12: Continued.

Agub_g3149	224.76	48.91	2.019	somatic-biased
Agub_g3150	391.27	97.63	1.814	somatic-biased
Agub_g3615	11.15	24.95	-1.360	reproductive-upregulated
Agub_g3653	3.06	1.85	0.537	constitutive
Agub_g4456	57.41	87.74	-0.810	reproductive-upregulated
Agub_g4922	6.74	0.02	8.286	somatic-specific
Agub_g5596	1.02	0.99	-0.116	low expression
Agub_g6523	116.14	385.37	-1.926	reproductive-biased
Agub_g7134	39.55	2.14	4.060	somatic-biased
Agub_g7135	5.93	0.40	3.738	somatic-biased
Agub_g7136	3.27	2.32	0.276	constitutive
Agub_g7137	1.39	2.38	-0.964	constitutive
Agub_g7272	38.75	0.00	9.498	somatic-specific
Agub_g7275	15.29	9.56	0.498	constitutive
Agub_g7482	2.22	1.43	0.429	constitutive
Agub_g7884	487.34	326.60	0.385	somatic-upregulated
Agub_g10451	90.52	190.91	-1.278	reproductive-upregulated
Agub_g10668	33.73	4.96	2.565	somatic-biased
Agub_g11010	15.27	5.07	1.362	low confidence
Agub_g11011	17.33	8.52	0.800	constitutive
Agub_g11013	0.44	0.00	2.374	low expression
Agub_g11014	1.17	0.00	2.838	low expression
Agub_g11083	29.65	6.65	1.969	somatic-biased
Agub_g10884	4.37	6.31	-0.698	constitutive
Agub_g11471	3.01	0.76	1.820	low confidence
Agub_g11472	4.81	2.82	0.579	low expression
Agub_g11474	533.42	2.14	7.752	somatic-specific
Agub_g11475	168.26	1.10	7.018	somatic-specific
Agub_g11589	3.49	0.00	7.229	somatic-specific
Agub_g11719	1.05	0.48	0.955	constitutive
Agub_g12144	277.49	86.22	1.495	somatic-biased
Agub_g12153	177.45	52.22	1.565	somatic-upregulated
Agub_g12154	508.00	186.95	1.249	somatic-upregulated
Agub_g12504	13.11	31.12	-1.439	reproductive-upregulated
Agub_g12816	39.40	0.64	5.398	somatic-biased
Agub_g12817	9.90	2.95	1.565	somatic-upregulated
Agub_g12688	630.77	723.99	-0.393	reproductive-upregulated
Agub_g12687	452.42	531.46	-0.425	reproductive-upregulated
Agub_g13723	35.12	11.01	1.485	somatic-upregulated
Agub_g13916	242.41	190.86	0.153	constitutive
Agub_g14323	112.19	160.89	-0.715	reproductive-upregulated
Agub_g14324	81.14	109.24	-0.624	reproductive-upregulated
Agub_g15106	22.48	14.23	0.475	constitutive
Agub_g15292	0.07	0.03	0.922	constitutive
Agub_g16031	23.98	0.00	7.640	somatic-specific

Table 4-13: List of reproductive- and somatic-specific genes.

locus	soma TPM	repr TPM	log2 fold change	description ^a
reproductive-specific genes				
Agub_g9352	119.86	837.98	-3.006	
somatic-specific genes				
Agub_g6446	92.04	0.04	11.117	
Agub_g8	36.57	0.00	10.301	HRGP like
Agub_g3194	528.60	0.46	10.161	deoxyhypusine synthase 1
Agub_g4645	147.75	0.11	9.875	aurora-like kinase
Agub_g7272	38.75	0.00	9.498	extracellular matrix glycoprotein ptherophorin
Agub_g11211	39.25	0.00	9.363	glutamate receptor
Agub_g9	12.66	0.02	9.358	HRGP like
Agub_g7	104.55	0.13	9.349	HRGP like
Agub_g14165	68.61	0.11	9.050	
Agub_g3504	45.11	0.00	8.762	
Agub_g13507	1160.17	2.44	8.709	
Agub_g7514	401.57	0.94	8.554	
Agub_g4918	78.79	0.15	8.543	HRGP like
Agub_g8314	26.80	0.00	8.530	beta-galactosidase like
Agub_g7041	25.46	0.00	8.516	HRGP like
Agub_g6563	36.58	0.09	8.473	
Agub_g5339	681.47	1.87	8.447	
Agub_g4917	51.41	0.00	8.353	
Agub_g4922	6.74	0.02	8.286	extracellular matrix glycoprotein ptherophorin
Agub_g7325	65.90	0.15	8.253	
Agub_g1045	44.50	0.15	8.007	fibrocystin-like protein
Agub_g13519	41.81	0.00	7.961	WD domain
Agub_g8186	78.42	0.28	7.905	HRGP like
Agub_g15414	31.33	0.00	7.777	
Agub_g8375	130.96	0.54	7.766	heat shock protein 22A
Agub_g11474	533.42	2.14	7.752	extracellular matrix glycoprotein ptherophorin
Agub_g12818	17.60	0.07	7.719	
Agub_g14952	33.82	0.15	7.691	
Agub_g16031	23.98	0.00	7.640	extracellular matrix protein ptherophorin
Agub_g10	15.68	0.06	7.529	HRGP like
Agub_g5474	14.79	0.00	7.420	
Agub_g5910	2.61	0.00	7.412	
Agub_g1811	289.82	1.50	7.410	actin-like protein NAP1
Agub_g5887	87.29	0.00	7.376	
Agub_g4314	60.46	0.36	7.331	
Agub_g8083	4.10	0.00	7.309	basal boby protein POC19
Agub_g3465	230.75	1.34	7.278	
Agub_g3974	11.00	0.00	7.275	
Agub_g2945	7.35	0.00	7.241	R2R3 Myb-like transcription factor
Agub_g11589	3.49	0.00	7.229	extracellular matrix glycoprotein ptherophorin
Agub_g10004	15.82	0.00	7.135	Ring finger domain

Table 4-13: Continued.

Agub_g11475	168.26	1.10	7.018	extracellular matrix glycoprotein ptherophorin
Agub_g8601	7.01	0.05	7.014	
Agub_g5340	807.85	5.67	6.930	
Agub_g6793	32.42	0.25	6.880	HRGP like
Agub_g11410	192.17	1.16	6.842	luminal binding protein Bip1
Agub_g15447	140.03	1.29	6.701	
Agub_g10050	18.60	0.17	6.606	
Agub_g6615	9.89	0.10	6.572	
Agub_g7351	20.20	0.21	6.402	HRGP like
Agub_g14044	111.68	1.21	6.331	zinc-dependent metalloprotease
Agub_g15458	8.79	0.08	6.290	
Agub_g15392	138.99	1.67	6.241	ricin-type beta-trefoil lectin domain-like
Agub_g4828	107.81	1.29	6.228	collagen-related protein
Agub_g1046	41.60	0.50	6.120	fibrocystin-like protein
Agub_g8895	35.75	0.43	6.119	DUF4497 like
Agub_g15505	17.66	0.23	6.085	
Agub_g14	7.22	0.09	6.031	HRGP like
Agub_g13520	24.83	0.34	6.024	WD domain
Agub_g8130	6.87	0.10	6.002	
Agub_g14951	50.57	0.73	5.852	
Agub_g13632	49.30	0.77	5.851	Ca2+/H+ antiporter, cation antiporter, membrane protein
Agub_g7577	87.59	1.46	5.782	HRGP like
Agub_g9322	25.69	0.39	5.782	putative serine/threonine-protein kinase
Agub_g8399	386.50	6.34	5.726	glycoside hydrolase family 16 protein
Agub_g15446	309.11	5.36	5.693	ricin-type beta-trefoil lectin domain-like
Agub_g12459	101.86	1.73	5.650	Kinesin
Agub_g3193	133.92	2.41	5.648	SET domain-containing protein
Agub_g14639	154.16	2.86	5.550	long-chain-fatty-acid CoA ligase
Agub_g3740	11.73	0.23	5.456	
Agub_g8453	51.08	0.94	5.455	
Agub_g10829	41.20	0.85	5.438	aurora like protein kinase
Agub_g8995	11.92	0.26	5.276	HRGP like
Agub_g12434	40.36	1.02	5.206	
Agub_g14724	4.77	0.12	5.200	HRGP like
Agub_g11411	205.21	5.19	5.176	luminal binding protein Bip1
Agub_g7170	57.66	1.50	5.153	methyltransferase like
Agub_g9035	91.95	2.43	5.148	
Agub_g14725	134.56	3.32	5.135	HRGP like
Agub_g12550	44.26	1.31	4.890	
Agub_g5920	13.80	0.42	4.863	
Agub_g591	730.36	22.28	4.855	
Agub_g7576	25.34	0.78	4.849	
Agub_g15314	27.41	0.85	4.843	
Agub_g6169	168.63	5.03	4.820	
Agub_g12442	152.56	4.99	4.759	glyoxal or galactose oxidase
Agub_g15609	84.30	2.80	4.729	aurora like protein kinase

Table 4-13: Continued.

Agub_g192	73.76	2.53	4.681	right handed beta helix region containing
Agub_g8804	71.07	2.45	4.676	HRGP like
Agub_g10833	55.42	1.95	4.640	matrix metalloprotease, MMP
Agub_g2900	121.61	4.37	4.607	ankyrin repeats containing
Agub_g12700	33.67	1.22	4.588	
Agub_g6530	741.37	27.47	4.576	small Arf-related GTPase
Agub_g13315	20.53	0.81	4.505	glyoxal or galactose oxidase
Agub_g8735	43.05	1.75	4.497	
Agub_g15692	549.15	21.35	4.493	F-box like domain containing
Agub_g8994	92.51	3.56	4.490	HRGP like
Agub_g7794	97.59	3.78	4.489	PLAC8 family
Agub_g11132	46.10	1.80	4.484	
Agub_g12421	10.39	0.42	4.453	HRGP like
Agub_g2461	193.67	7.85	4.446	UGP1, UTP--glucose-1-phosphate uridylyltransferase
Agub_g14326	45.88	1.84	4.446	
Agub_g10377	715.54	28.75	4.445	HRGP like
Agub_g15666	8.50	0.37	4.409	
Agub_g7874	68.36	2.81	4.399	
Agub_g8400	212.24	9.52	4.318	
Agub_g14442	15.91	0.82	4.126	fasciclin domain containing
Agub_g13940	9.36	0.47	4.125	flagella associated protein ANK17
Agub_g13601	46.87	2.42	4.098	protein kinase like
Agub_g9968	522.66	26.66	4.080	Bax inhibitor-1 like
Agub_g3422	304.91	16.01	4.022	
Agub_g5072	38.26	2.13	3.995	adenylate/guanylate cyclase
Agub_g7725	67.07	3.91	3.914	HRGP like
Agub_g14043	73.19	4.21	3.913	zinc-dependent metalloprotease
Agub_g15894	18.59	1.11	3.903	protein kinase like
Agub_g11228	45.22	2.72	3.859	laccase like
Agub_g12262	21.43	1.27	3.855	
Agub_g8265	31.07	1.88	3.845	RWP-RK domain-containing transcription factor
Agub_g8803	68.19	4.31	3.794	
Agub_g7065	237.48	15.28	3.768	flagella associated protein FAP148
Agub_g14640	63.19	4.23	3.720	
Agub_g2110	105.43	7.01	3.711	pyruvate phosphate dikinase
Agub_g9239	355.65	24.70	3.649	uridine kinase like
Agub_g10269	48.62	3.51	3.581	
Agub_g10623	10.99	0.81	3.561	pyruvate-ferredoxin oxidoreductase
Agub_g10627	977.66	74.98	3.509	pyruvate-ferredoxin oxidoreductase
Agub_g2161	40.65	3.24	3.476	
Agub_g7712	37.91	3.07	3.441	
Agub_g5332	36.40	2.99	3.402	cyclic nucleotide-binding domain containing
Agub_g10680	3670.07	306.74	3.388	papain family cysteine protease
Agub_g14741	127.34	10.69	3.377	
Agub_g9628	49.74	4.24	3.358	HRGP like
Agub_g14426	48.60	4.25	3.327	bacterial Alpha-2-macroglobulin MG10 domain like

Table 4-13: Continued.

Agub_g2259	121.85	11.31	3.259	
Agub_g5499	38.59	3.54	3.232	
Agub_g7105	71.50	6.72	3.224	tetratricopeptide repeat
Agub_g14780	37.23	3.59	3.197	Sulfatase
Agub_g12101	92.78	8.93	3.164	
Agub_g8444	257.30	25.21	3.162	RabGAP/TBC protein
Agub_g15310	184.49	18.42	3.140	sodium/calcium exchanger protein domain containing
Agub_g9235	378.95	38.13	3.122	flagellar associated protein FAP228, callose synthase-like protein
Agub_g11004	99.03	10.36	3.053	presenilin protease
Agub_g5284	155.24	16.41	3.051	Myb-like DNA-binding domain containing
Agub_g14990	99.07	10.92	2.999	glycosyl transferase
Agub_g10650	327.59	36.69	2.972	
Agub_g4501	1146.74	128.69	2.964	cysteine-rich secretory protein family
Agub_g10679	3757.28	518.96	2.664	papain family cysteine protease

^a Green background color: proline-rich domain like HRGPs, red background color: putative transcription factors.

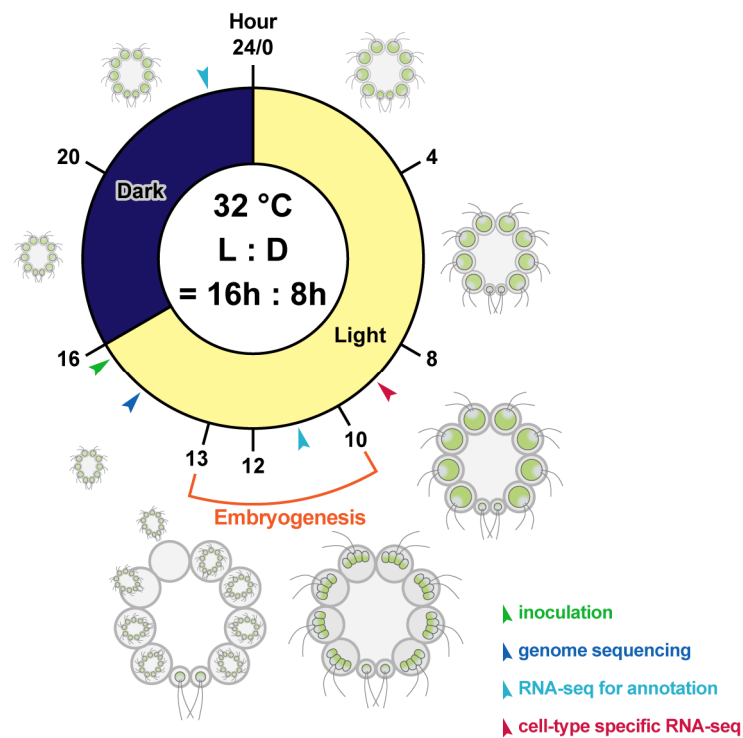


Fig. 4-1: Schematic diagram of the asexual life cycle of *Astrephomene gubernaculifera* in the synchronous culture. The asexual life cycle is completed in approximately 24 h, and the culture is highly synchronized with the light–dark cycle. Almost all reproductive cells undergo embryogenesis between approximately 10–13 h after the onset of the light period. The timing of inoculation (green arrowhead) and sampling for genome sequencing (blue arrowhead) RNA-seq for annotation (cyan arrowhead) and cell-type RNA-seq (red arrowhead) are shown.

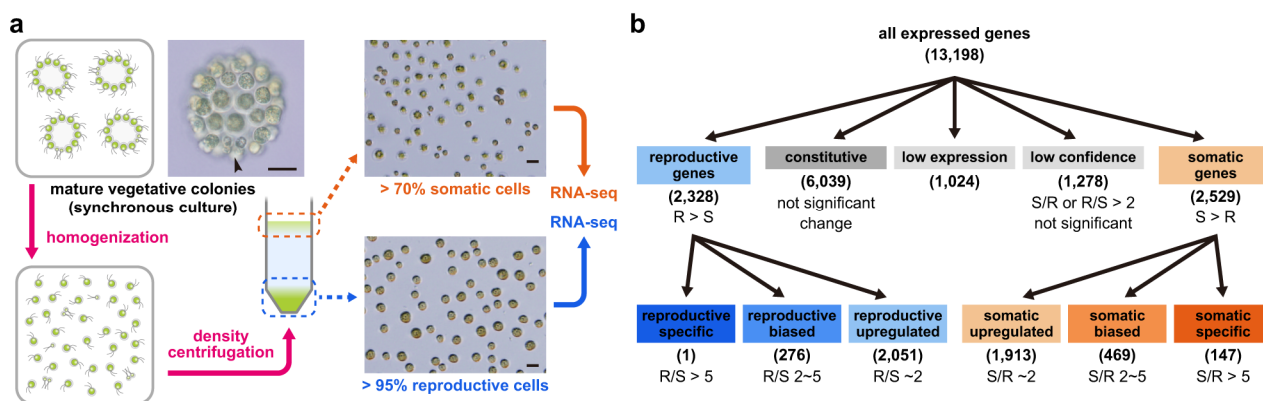


Fig. 4-2: Schematic diagrams of the workflow of cell-type RNA-seq analysis of *Astrephomene gubernaculifera*. (a) The workflow of isolation of somatic cells and reproductive cells. Mature vegetative colonies in synchronous culture have large reproductive cells and small somatic cells (arrowhead), and these two cell-types are distinguishable in size. Colonies are homogenized by strong pipetting and separated by density centrifugation. The top of the gradient is subjected to another centrifugation and collected as somatic cell samples, which contain > 70% of somatic cells (often connected to other somatic cells), while the bottom of the gradient is collected as reproductive cell samples, which contain > 95% of reproductive cells. Both samples are then subjected to RNA extraction for RNA-seq. Scale bars = 20 μ m. (b) The classification of cell-type gene expression. All expressed genes are first classified by differential expression analysis, and then the differential expressed genes are further classified by fold changes (see Materials and Methods for detail). R: reproductive expression, S: somatic expression.

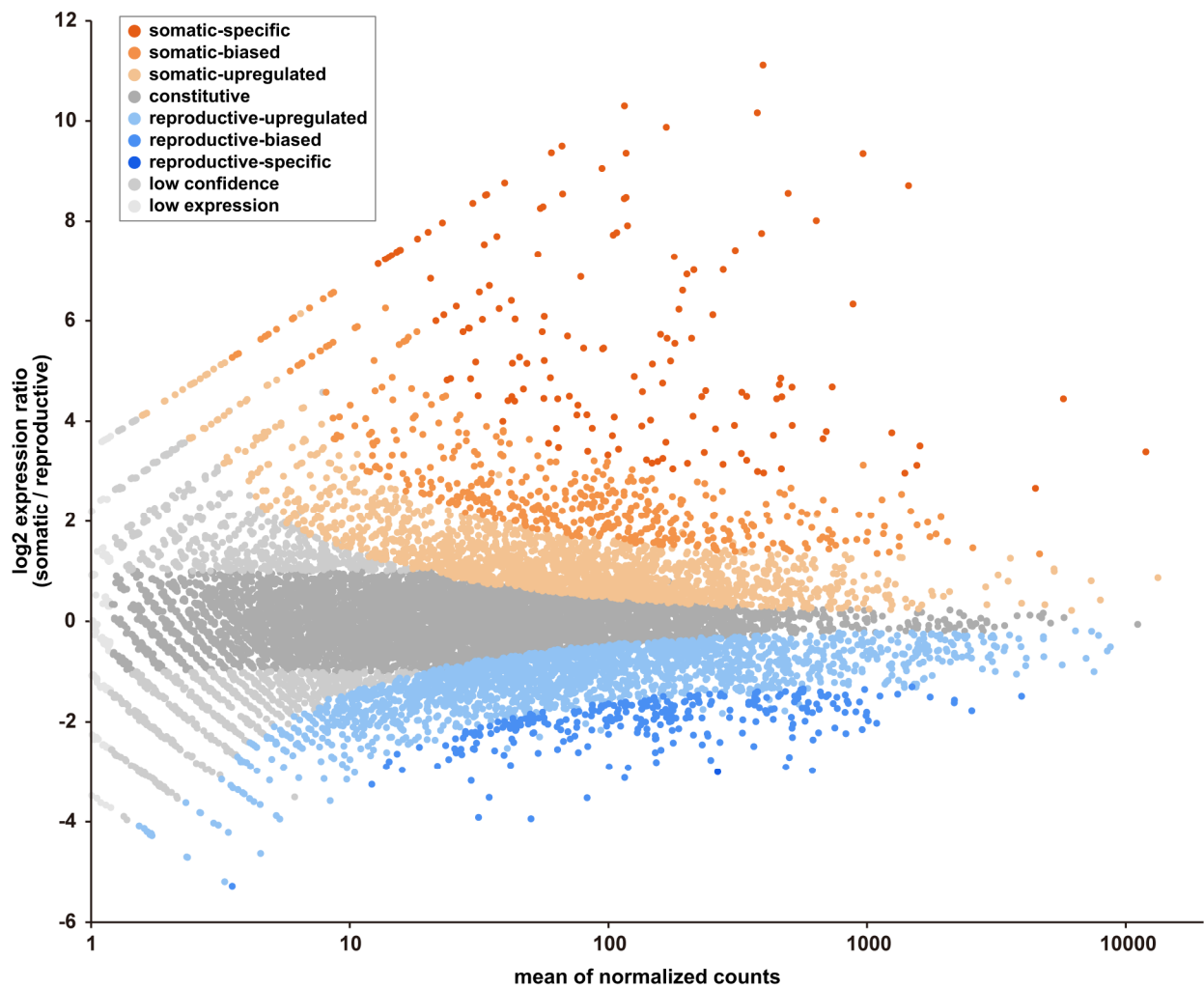


Fig. 4-3: MA plot of the cell-type RNA-seq analysis of *Astrephomene gubernaculifera*. Each dot represents a single gene with coloring by classification shown on upper-left. Mean of normalized read counts is plotted on x-axis, log scale. Ratio of somatic expression to reproductive expression calculated by DESeq2 (Love et al., 2014) is plotted on y-axis, log2 scale.

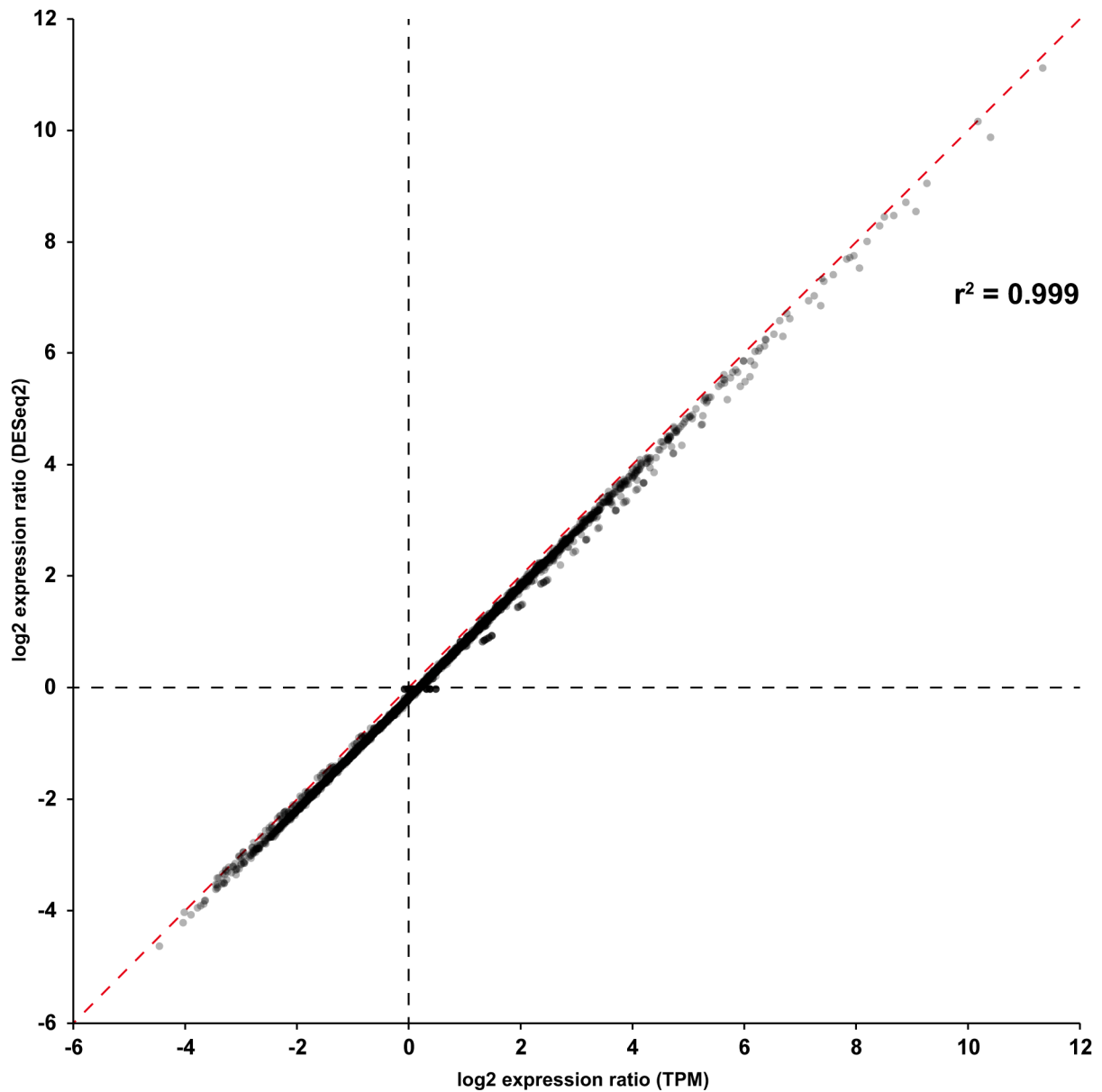


Fig. 4-4: Correlation of expression ratio normalized by DESeq2 and TPM from cell-type RNA-seq analysis in *Astrephomene gubernaculifera*. Each dot represents a single gene. Ratio of somatic expression to reproductive expression normalized by TPM (transcripts per million, Wagner et al., 2012) is plotted on x-axis and ratio by DESeq2 (Love et al., 2014) are plotted on y-axis, both in log2 scale. The red dashed line indicates $y = x$.

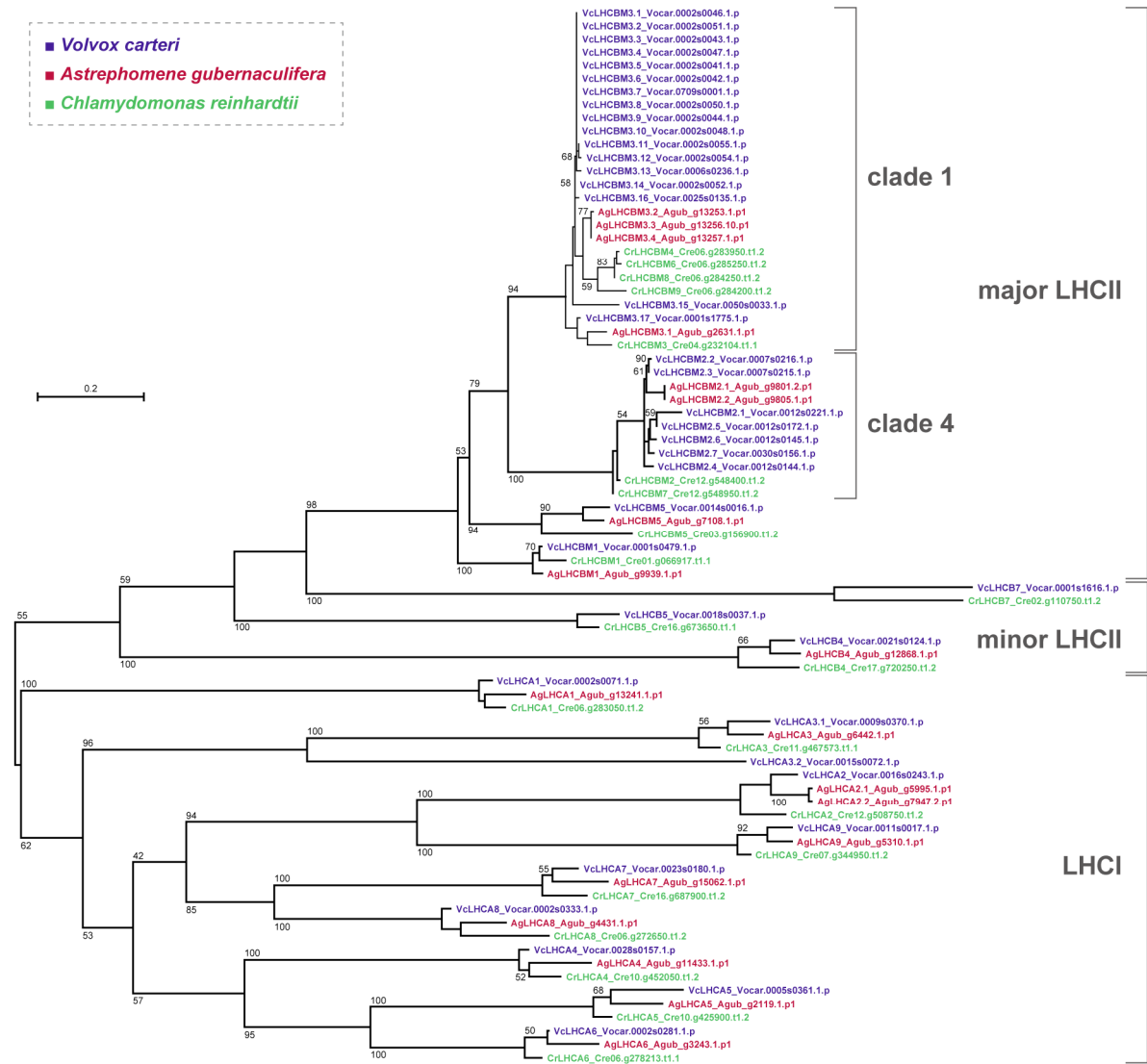


Fig. 4-5: Identification of LHC gene family in *Astrephomene gubernaculifera*. The 271 amino acid positions from LHCII and LHCI genes in *Volvox carteri*, *Chlamydomonas reinhardtii* and *A. gubernaculifera* (Table 4-9) were subjected to maximum-likelihood analysis. The bootstrap values ($\geq 50\%$) are shown at branches. The colors indicate species according to the color key at the upper left. The classification of LHCII genes are based on a previous study (Matt and Umen, 2018). LHCBS5 and LHCBS7 are absent in gene models of *A. gubernaculifera* genome.

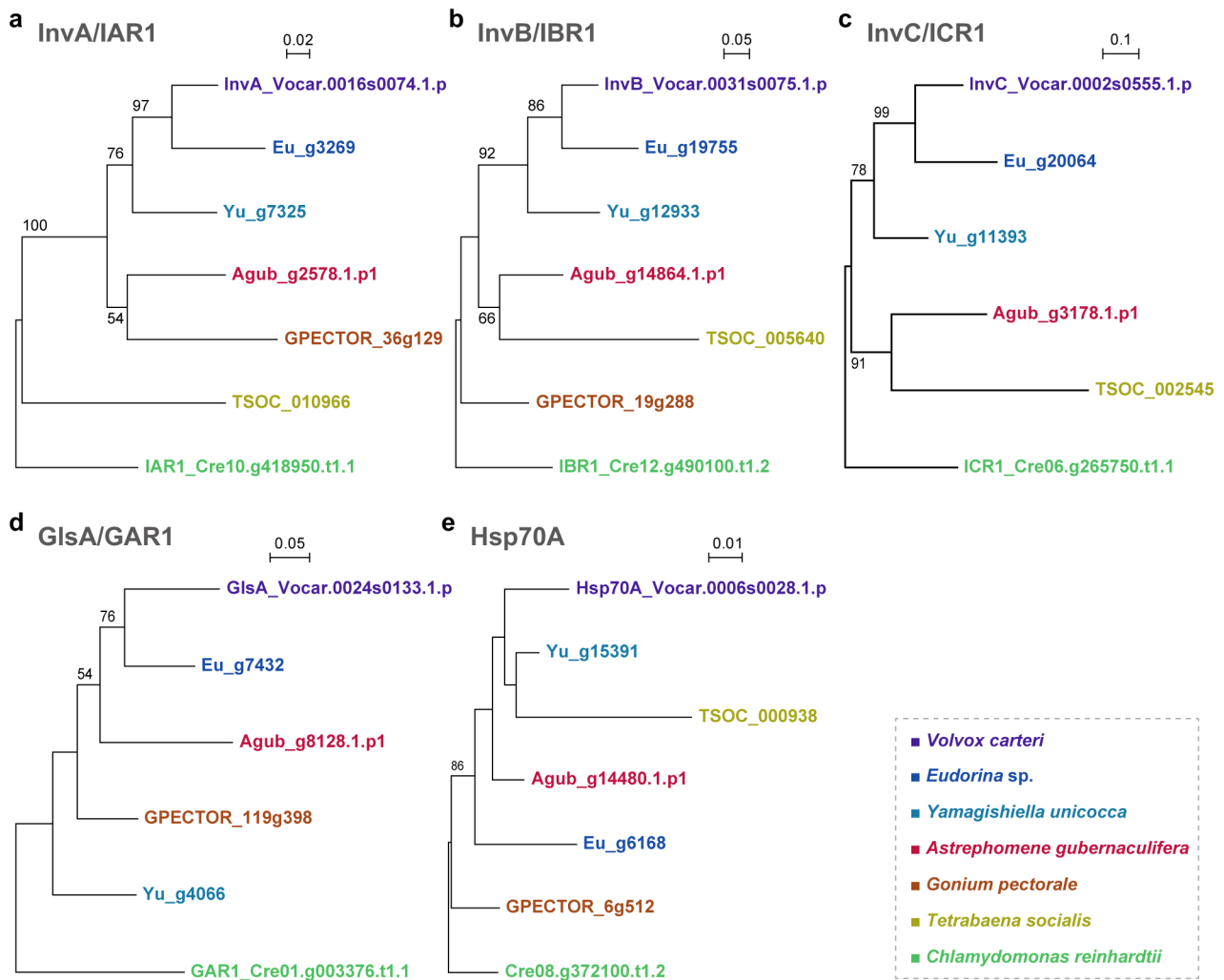


Fig. 4-6: Maximum-likelihood phylogenetic trees of orthologs of five genes involved in embryogenesis in *Volvox carteri*. Six or seven volvocine orthologs were analyzed in each tree. The bootstrap values ($\geq 50\%$) are shown at branches. The colors indicate species according to the color key at the lower right. **(a)** InvA/IAR1 orthologs (1,007 amino acid positions, DAYHOFF + I + G + F model). **(b)** InvB/IBR1 orthologs (347 amino acid positions, JTT + I + G + F model). **(c)** InvC/ICR1 orthologs (391 amino acid positions, LG + G + F model). **(d)** GlisA/GAR1 orthologs (634 amino acid positions, DAYHOFF + G + F model). **(e)** Hsp70A orthologs (653 amino acid positions, LG + I + G + F model).

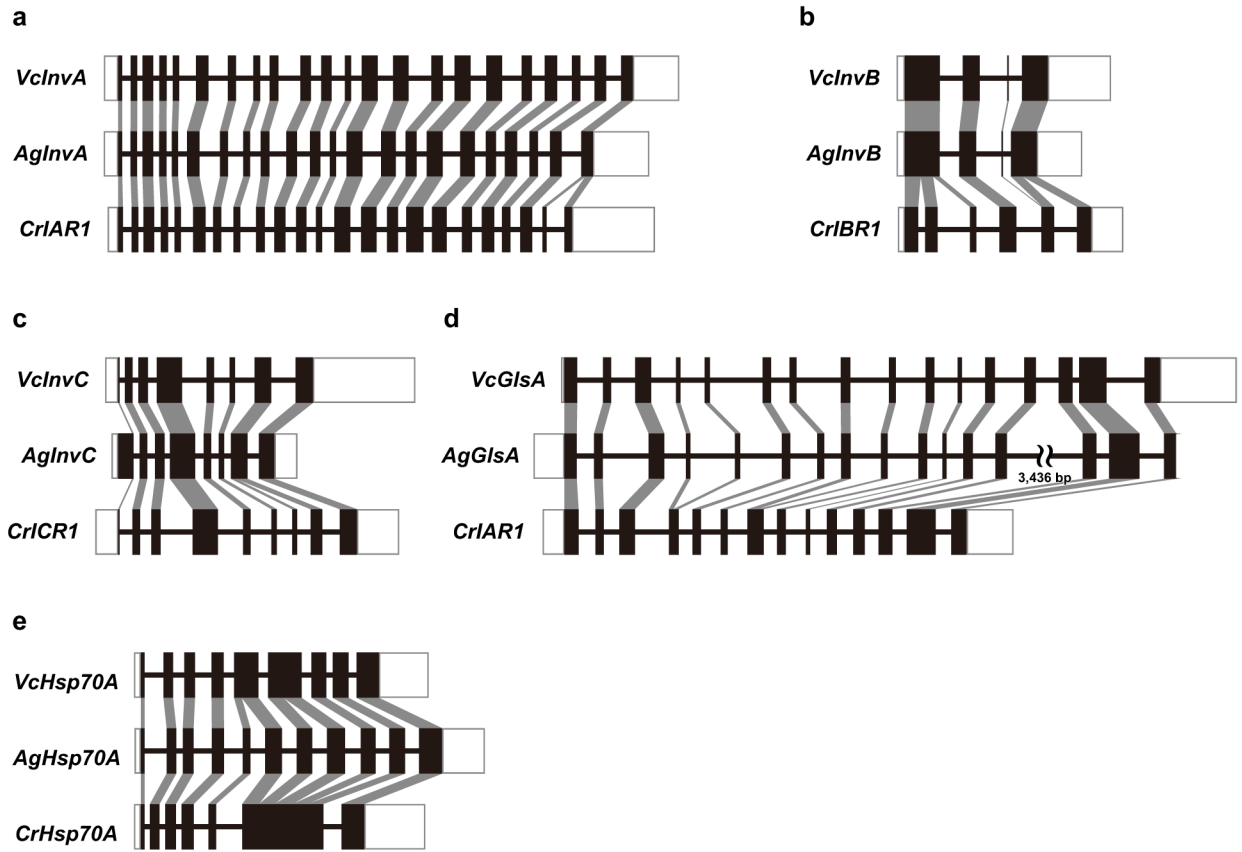


Fig. 4-7: Comparison of the exon–intron structure of orthologs of five genes involved in embryogenesis in *Volvox carteri*. The exon–intron structures from gene model in *V. carteri* (Vc), *Astrephomene gubernaculifera* (Ag) and *Chlamydomonas reinhardtii* (Cr) are shown in line. The thick and thin lines correspond to exons and introns respectively, black thick lines indicate CDS and white thick line indicate 5' and 3' UTRs. **(a)** InvA/IAR1 orthologs. **(b)** InvB/IBR1 orthologs. **(c)** InvC/ICR1 orthologs. **(d)** GlsA/GAR1 orthologs. **(e)** Hsp70A orthologs.

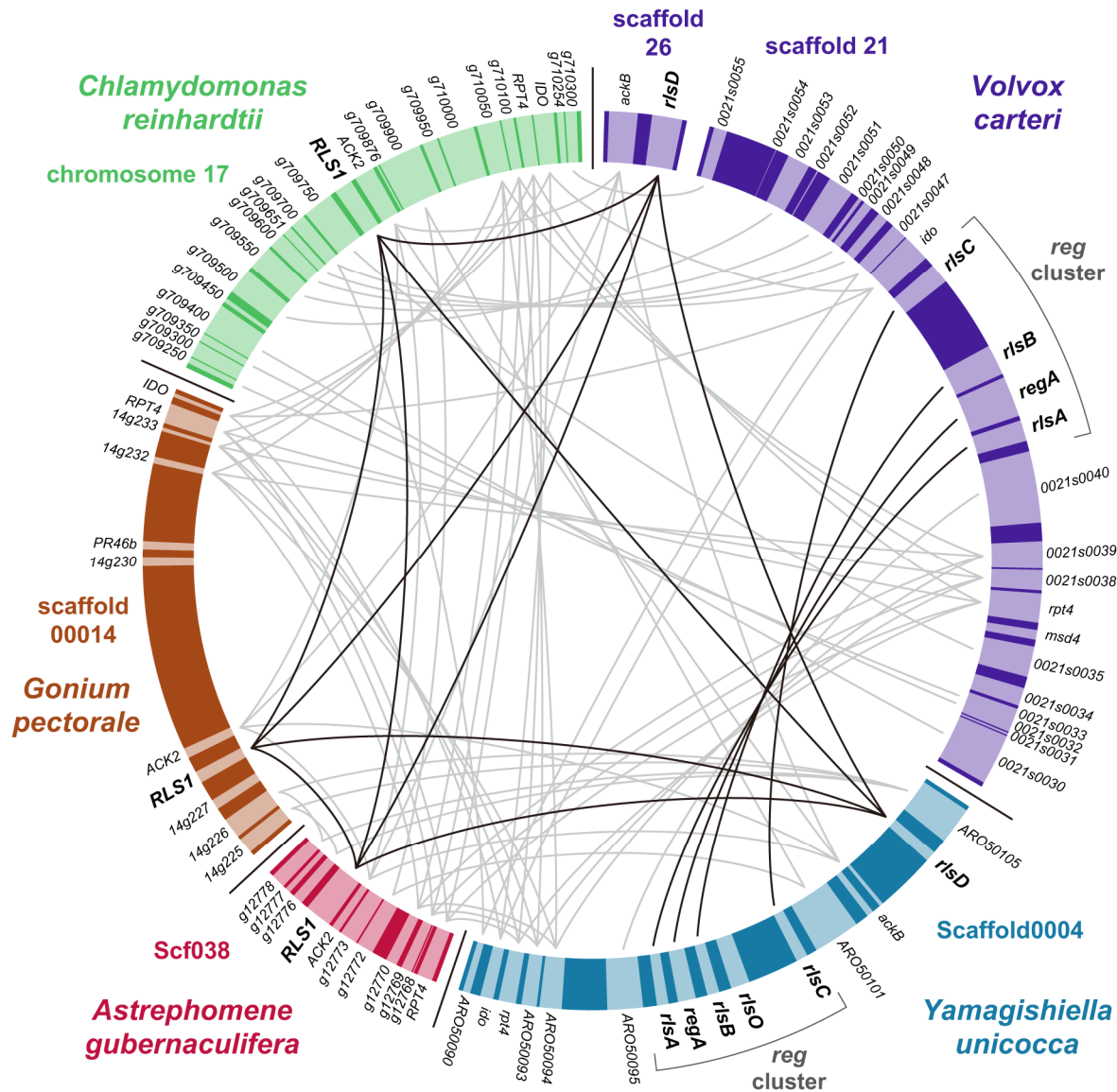


Fig. 4-9: Gene synteny near the *RLS1/rlsD* orthologs and *reg* cluster among volvocine algae. Synteny of *Chlamydomonas reinhardtii*, *Gonium pectorale*, *Astrephomene gubernaculifera*, *Yamagishiella unicocca* and *Volvox carteri* is shown. The synteny of *RLS1/rlsD* orthologs and *reg* cluster genes (Fig 4-8) is indicated with black lines, and the synteny of other genes are indicated with grey lines based on BLASP search (see Materials and Methods).

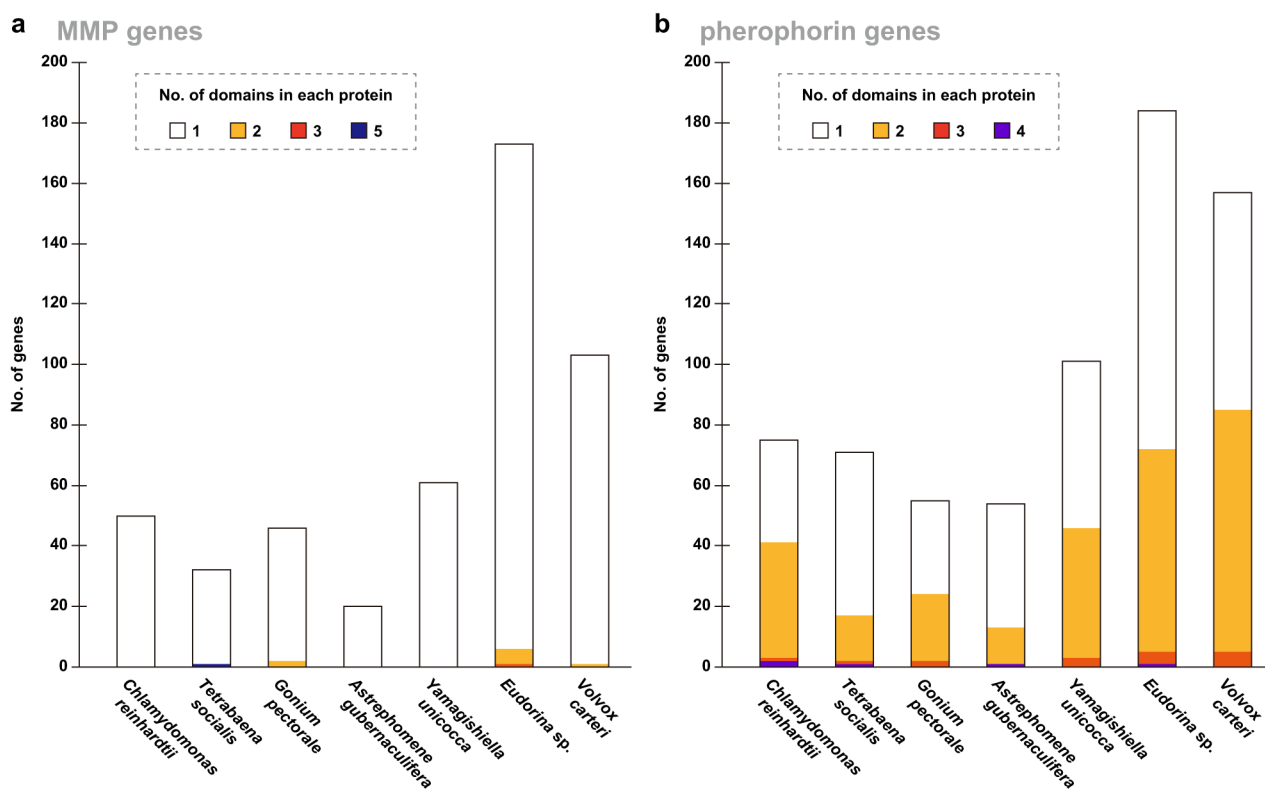


Fig. 4-10: The number of ECM genes in volvocine green algae. (a) Stacked bar graph of the number of MMP genes (Table 4-5). **(b)** Stacked bar graph of the number of pherophorin genes (Table 4-6). The colors indicate the number of MMP or pherophorin domains in each transcript, according to the color key at the upper left.

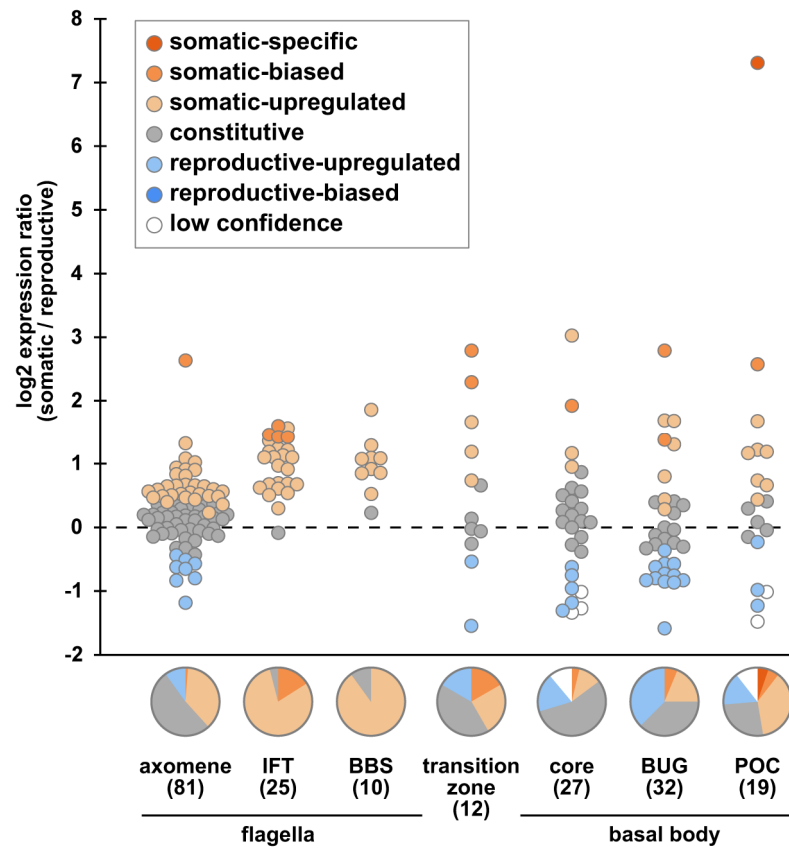


Fig. 4-11: Cell-type expression of flagella, transition zone and basal body genes in *Astrephomene gubernaculifera*. Expression ratio of axoneme genes, intraflagellar transport (IFT) genes, Bardet–Biedel Syndrome (BBS) genes, transition zone genes and basal bodies genes (core genes with validated function and localization, BUG genes and POC genes) (Table 4-8) is shown. The number in brackets indicates the number of genes in each category.

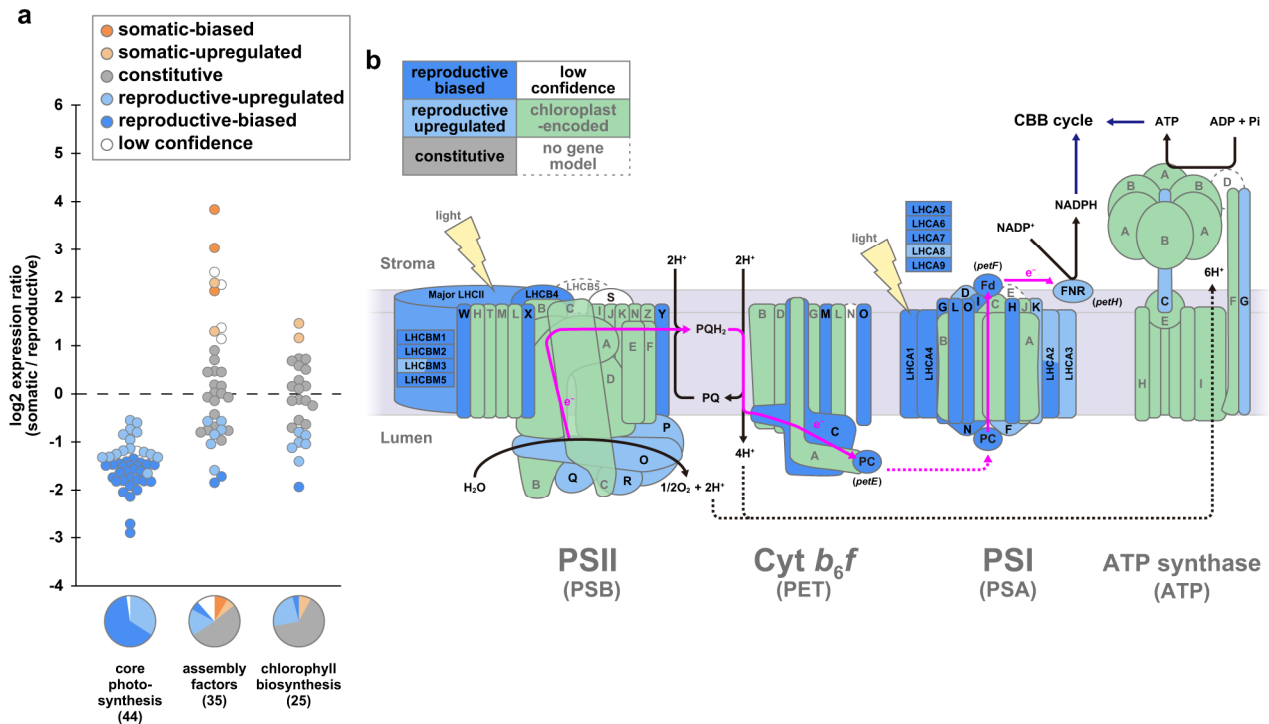


Fig. 4-12: Cell-type expression of photosynthesis-related genes in *Astrephomene gubernaculifera*. (a) Expression ratio of genes for photosynthetic light reactions (core photosynthesis), assembly factors for photosynthetic complexes and chlorophyll biosynthesis (Table 4-9). The number in brackets indicates the number of genes in each category. (b) Schematic diagram of complexes in photosynthetic light reactions and expression pattern of genes encoding their subunits and electron transport proteins. The illustrated structure of the complexes is based on a previous study (Allen et al., 2011). Each component is colored by classification shown on upper-left. Note the green-colored components are chloroplast-encoded in *Chlamydomonas reinhardtii* (Merchant et al., 2007) and also not found in nuclear genome of *A. gubernaculifera*.

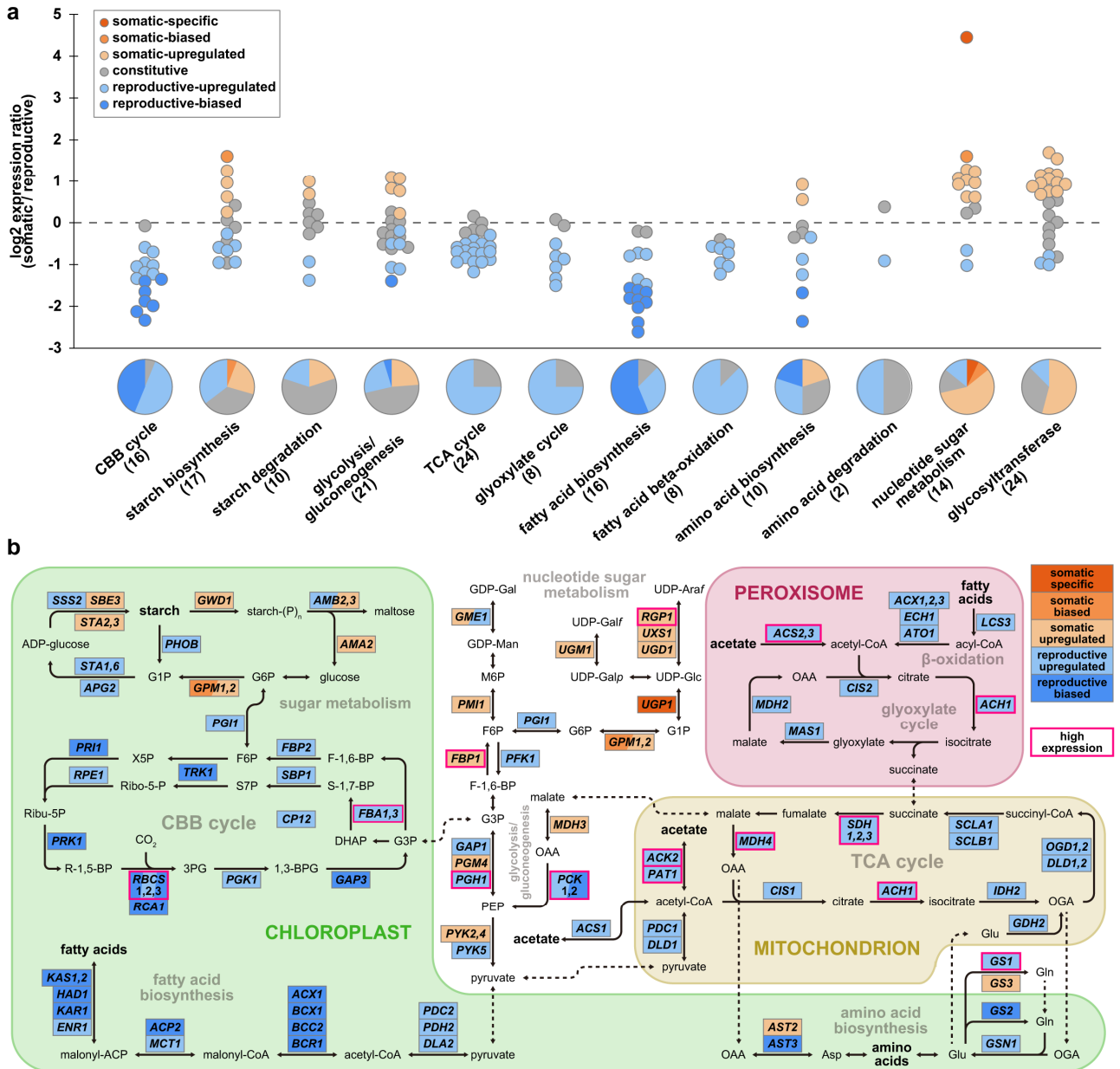


Fig. 4-13: Cell-type expression of carbon metabolism genes in *Astrephomene gubernaculifera*.

(a) Gene expression ratio (Tables 4-10, 4-11). In swarm plot, each dot represents a single gene with coloring by classification shown on upper-left. The number in brackets indicates the number of genes in each category. (b) Schematic diagram of carbon metabolic pathways and differentially expressed genes. Solid arrows indicate chemical reactions, and dashed arrows indicate transportation of substrates between subcellular compartments. The metabolic pathways and the subcellular compartments (chloroplast, mitochondria, peroxisome and cytosol) are mainly based on previous studies (Johnson and Alric, 2013; Zones et al., 2015; Matt and Umen, 2018) (see Materials and Methods for detail). Each enzyme is colored by classification shown on upper-right. High expression genes with TPM (transcripts per million, Wagner et al., 2012) > 1,000 (or total TPM of gene copies > 1,000) in both somatic and reproductive cells are surrounded by pink bold squares.

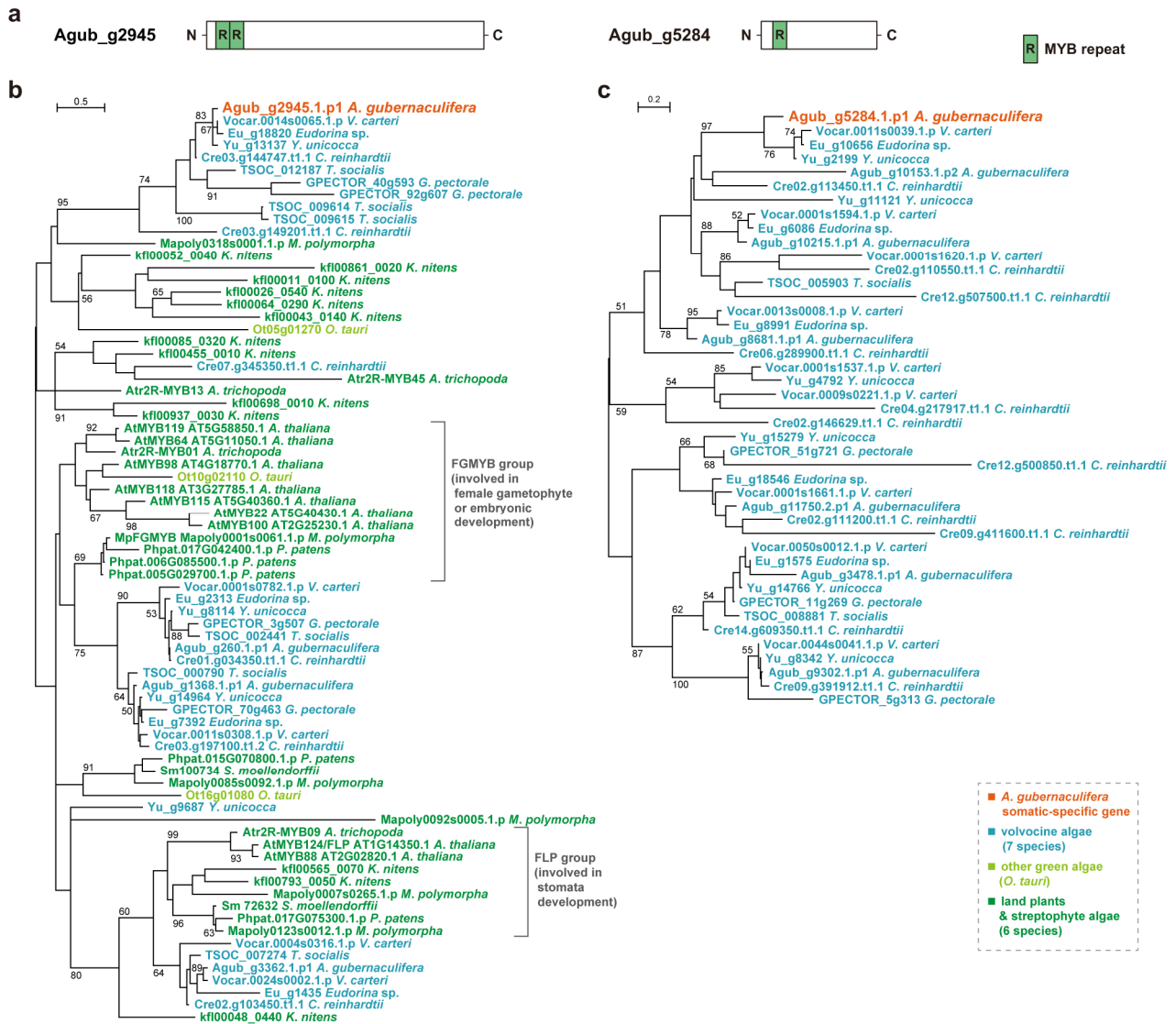


Fig. 4-14: Characteristics of two somatic specific MYB transcription factors in *Astrephomene gubernaculifera*. (a) The domain structure of Agub_g2945 and Agub_g5284. (b) Phylogenetic tree of R2R3-MYB transcription factors including Agub_g2945. The 105 amino acid positions from DNA binding domain of 75 MYBs were subjected to maximum-likelihood analysis (LG + I + G model). The bootstrap values from maximum-likelihood ($\geq 50\%$) are shown at branches. (c) Phylogenetic tree of 1R-MYB-like genes including Agub_g5284. The 87 amino acid positions from DNA binding domain of 43 MYBs were subjected to maximum-likelihood analysis (LG + I + G + F model). The bootstrap values from maximum-likelihood ($\geq 50\%$) are shown at branches.

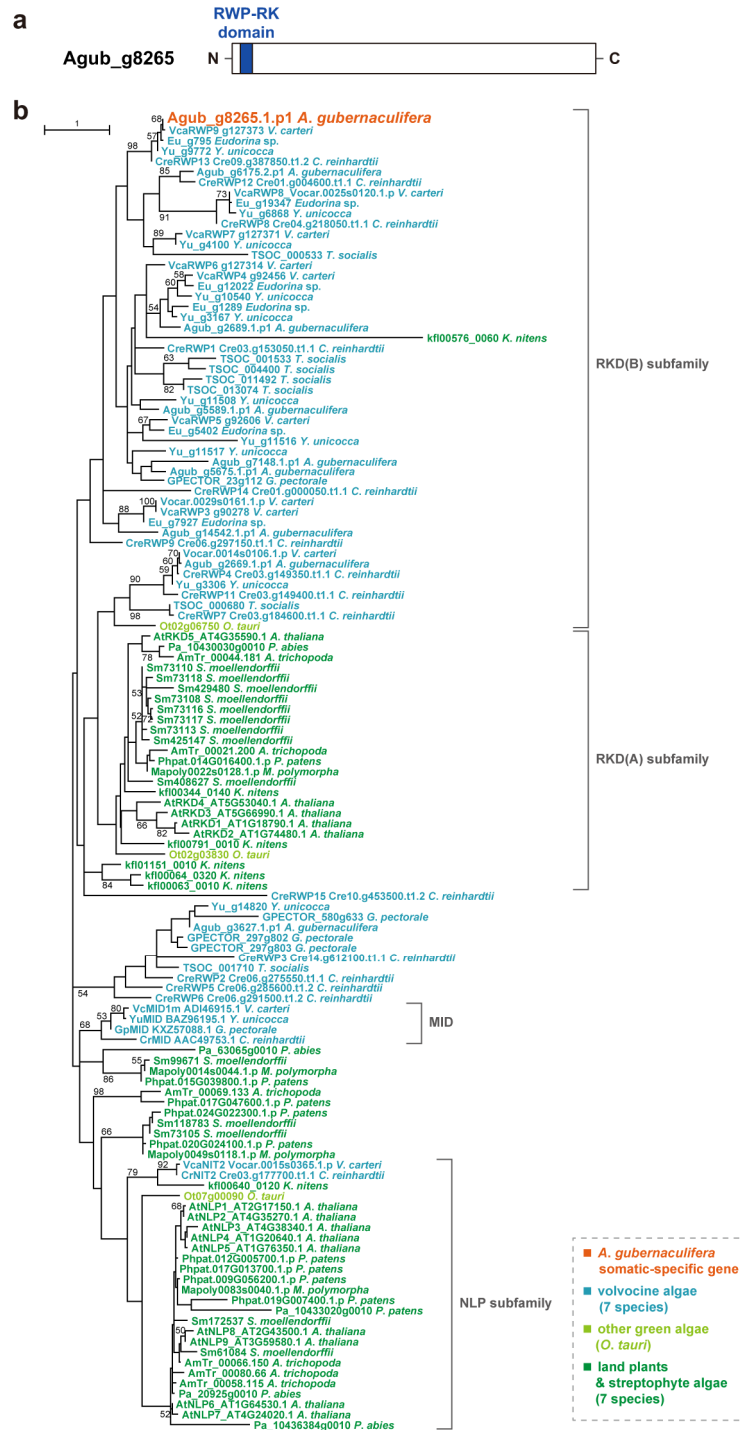


Fig. 4-15: Characteristics of a somatic specific RWP-RK transcription factor in *A. gubernaculifera*. (a) The domain structure of Agub_g8265. (b) Phylogenetic trees of RWP-RK transcription factors including Agub_g8265. The 49 amino acid positions from 127 RWP-RK transcription factors were subjected to maximum-likelihood analysis (LG + I + G model). The bootstrap values from maximum-likelihood ($\geq 50\%$) are shown at branches.

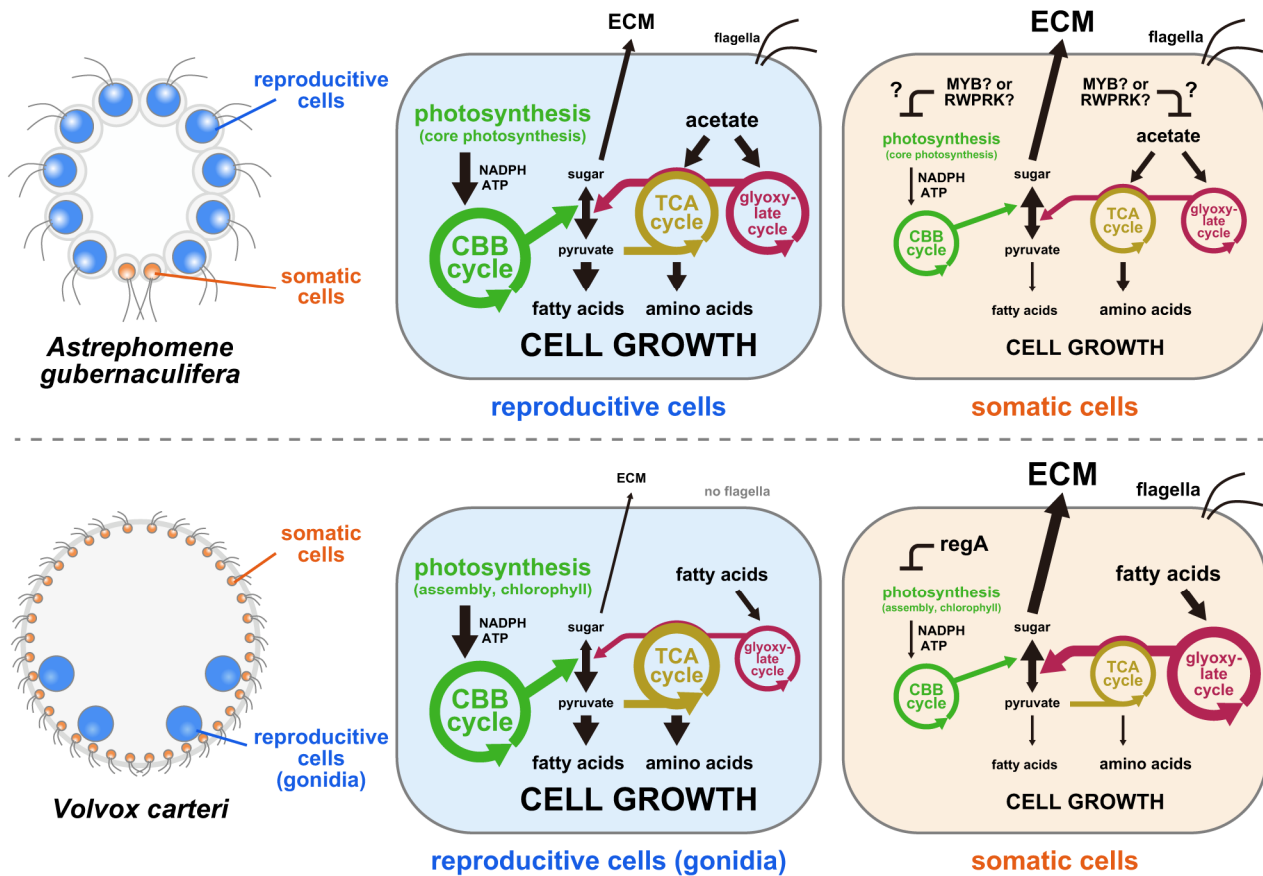


Fig. 4-16: Schematic diagrams of the comparison of germ–soma differentiation in *Astrephomene gubernaculifera* and *Volvox carteri*. The simplified metabolic pathways in *A. gubernaculifera* are based on this study, and those in *V. carteri* are based on a previous study (Matt and Umen, 2018). The size of letters and thickness of arrows indicate gene expression levels. The genes involved in photosynthesis, photosynthetic light reactions and Calvin–Benson–Bassham (CBB) cycle, are more expressed in reproductive cells than in somatic cells in *A. gubernaculifera*. Similar regulation of photosynthesis also occurs in *V. carteri*. Moreover, anabolic pathways are upregulated in reproductive cells in both species, which might affect the difference in cell growth between the two cell-types. While TCA cycle and glyoxylate cycle are upregulated in reproductive cells in *A. gubernaculifera* as downstream pathways of acetate uptake, glyoxylate cycle are upregulated in somatic cells in *V. carteri* for breakdown of fatty acids. Genes involved in ECM biosynthesis are upregulated in somatic cells in both species.

Chapter 5

General Discussion

Within the volvocine green algae, evolution of spheroidal colonies, germ–soma differentiation and expansion of ECM has occurred independently in two different lineages, Volvocaceae and *Astrephomene*. While the developmental mechanisms and molecular bases for these multicellular traits have been investigated in detail in Volvocaceae, *Astrephomene* has not been studied in such levels. In this thesis, I conducted the developmental analyses of embryogenesis in *Astrephomene* and in species with ancestral flattened colonies, *Gonium* and *Tetrabaena*, with comparison to Volvocaceae. I also investigated *de novo* sequenced genome of *Astrephomene* to resolve molecular bases for multicellular traits and their evolution within the volvocine algae.

In Chapter 2 and Chapter 3, I revealed the two independent evolution of the cellular mechanisms of spheroid colony formation during embryogenesis: the rotation of the protoplast during successive cell divisions in *Astrephomene* and inversion with acute chloroplast ends after successive cell divisions in Volvocaceae (Chapter 3, Fig. 3-11). In addition, the analyses of embryogenesis in volvocine species in Chapter 2 and Chapter 3 indicated a constraint of cell divisions in this lineage: as basal bodies are used for both flagellar apparatus and cell divisions and they cannot move freely because of a massive chloroplast, cell division planes should take place along the flagella–chloroplast axis. The similar constraint in cell division is also seen in spherical colony (rosette) formation in choanoflagellates and embryogenesis in some groups of sponges (Brunet and King, 2017). Moreover, the inversion of a cell layer which resembles the inversion in *Volvox* is observed in a multicellular choanoflagellate *Choanoeca flexa* (Brunet et al., 2019) and in a calcareous sponge *Sycon ciliatum* (Franzen, 1988). Thus, this constraint might apply to many multicellular lineages, which form spheroidal colonies composed of flagellated cells. Under this constraint, *Astrephomene* and Volvocaceae have evolved different tactics to form spheroidal colonies with flagella outside (Chapter 2, Fig. 2-10).

In addition, the orthologs of genes involved in inversion in *Volvox carteri*, *InvA*, *InvB* and *InvC*, were revealed to be conserved among volvocine algae in Chapter 4. However, the functions of

these orthologs during embryogenesis or cell divisions other than in *V. carteri* are currently unknown, even in *C. reinhardtii*. Since the gradual expansion of mother cell wall or ECM was observed during embryogenesis in *A. gubernaculifera* (Chapter 2, Figs. 2-4, 2-5), *Eudorina* sp. (Chapter 2, Fig. 2-6) *G. pectorale* (Chapter 3, Figs. 3-3a–e, 3-4) and *T. socialis* (Chapter 3, Figs. 3-3f–h, 3-5), the function of *InvB* and *InvC* orthologs in these species might be expanding of mother cell wall or ECM and providing sufficient space for embryogenesis as in *V. carteri* (Ueki and Nishii, 2008, 2009). On the other hand, function of *InvA* orthologs in *Astrephomene*, *Gonium* and *Tetrabaena* might be different from that in Volvocaceae. During the inversion of *V. carteri*, *InvA*, a kinesin, is localized in the cytoplasmic bridges connecting daughter protoplasts and interacts with cortical microtubules to move daughter protoplasts against the cytoplasmic bridges, which produces the driving force for bending of the cell sheet (Nishii et al., 2003). Since embryos of *A. gubernaculifera*, *G. pectorale* and *T. socialis* also have cytoplasmic bridges between daughter protoplasts (Hoops and Floyd, 1982; Arakaki et al., 2013; Iida et al., 2013), the *InvA* orthologs might be involved in the rotation of daughter protoplasts or the change in direction of the posterior somatic cells after cell divisions in *Astrephomene* (Chapter 2), expansion of the cell layer in *Gonium* and the shifting of protoplasts in *Tetrabaena* (Chapter 3). This speculation would be ascertained by the detailed observation of dynamics of cytoplasmic bridges during embryogenesis in these species as well as functional analyses of *InvA* orthologs. Whether the *InvA* ortholog in *Astrephomene* is involved in the rotation of daughter protoplasts or not, the mechanisms to regulate the timing, directions and angles of the rotation of daughter protoplasts are also thought to be acquired during the evolution of spheroidal colony formation in *Astrephomene*.

One of the molecular bases associated with the formation of acute chloroplast ends in Volvocaceae might be the regulation of cortical microtubules. In flask-shaped cells during inversion in *V. carteri*, the elongated stalks at the chloroplast ends are lined with cortical microtubules (Viamontes and Kirk, 1977). Moreover, colchicine (which prevents the assembly of microtubules)

and cold treatment (which causes rapid disassembly of microtubules) have both been reported to block the formation of stalks, which suggests that the microtubules play an important role in the formation of stalks in flask-shaped cells (Viamontes et al., 1979). Similar patterns of microtubules lining the acute chloroplast ends during inversion have also been observed in *V. tertius* (Pickett-Heaps, 1970) and *Eudorina elegans* (Marchant, 1977). By contrast, the indirect immunofluorescence microscopy used to stain the microtubules of embryos in *G. pectorale* and *T. socialis* in Chapter 3 did not demonstrate the distinctive pattern of cortical microtubules at the chloroplast ends: the cortical microtubules lined the entire cell bodies and concentrated around the apical ends after successive cell divisions (Chapter 3, Figs. 3-6c, d, 3-7c, d). Thus, the ancestral genes involved in the remodeling or the maintenance of cortical microtubules, which have not previously been reported, might have been co-opted or new genes have been acquired to form the acute chloroplast ends in the ancestor of the Volvocaceae.

Chapter 4 also indicated that the different genetic bases have been involved in the evolution of germ–soma differentiation between *Astrephomene* and *Volvox*. While the master regulator genes in *V. carteri* is *regA*, which belongs to VARL gene family with SAND domain, the candidates for the master regulator genes in *A. gubernaculifera* belonged to MYB gene family or RWP-RK gene family (Chapter 4), though it remains unveiled whether one of these candidates is actually the master regulator of differentiation of somatic cells in *A. gubernaculifera*. In *C. reinhardtii*, *RLS1* is expressed under dark or stressful conditions (Nedelcu and Michod, 2006; Nedelcu, 2009) and therefore it is hypothesized that *RLS1/rlsD* ortholog in volvocine ancestors might have responded to environment and regulated growth and reproduction of cells and has been co-opted to downregulation of growth and reproduction of somatic cells in the extant species of *Volvox* (Nedelcu and Michod, 2006; Nedelcu, 2009; Hanschen et al., 2014; Grochau-Wright et al., 2017), though whether *RLS1* contributes to repression of growth via downregulation of photosynthetic pathways is yet to be confirmed. This hypothesis corresponds to the temporal-to-spatial transition hypothesis for

animal multicellularity: temporal switching of different cell phenotypes of unicellular ancestors co-opted to differentiation between cell-types (Mikhailov et al., 2009). The candidates for the master regulator genes in *A. gubernaculifera* have putative homologs in *C. reinhardtii* (Chapter 4, Figs. 4-14b, c, 4-15b) and this hypothesis might be applicable to the evolution of somatic cells in *A. gubernaculifera*. Moreover, while the cell fate after embryogenesis is determined solely by the cell size in *V. carteri* (Kirk et al., 1993), asymmetric cell divisions or a distinct difference in cell size between somatic and reproductive cells were not observed during embryogenesis in *A. gubernaculifera* (Chapter 2). Therefore, the mechanisms for the expression of the master regulator gene in proper cells might be also different between *Astrephomene* and *Volvox* and evolved independently.

The difference in regulator genes between *A. gubernaculifera* and *V. carteri* might affect the differences in metabolic pathways which are downregulated to repress the growth of somatic cells (Chapter 4, Fig. 4-16). In terms of photosynthesis, genes encoding core subunits of photosynthetic complexes were mainly downregulated in somatic cells of *A. gubernaculifera* (Chapter 4, Fig. 4-12), while genes encoding assembly factors for photosynthetic complexes and enzymes for chlorophyll biosynthesis were mainly downregulated in somatic cells of *V. carteri* (Matt and Umen, 2018). In addition, downregulation of glyoxylate cycle, which is thought to be used as the acetate metabolic pathway, is also seen in somatic cells of *A. gubernaculifera* (Chapter 4, Fig. 4-13) in contrast to upregulation of glyoxylate cycle in somatic cells of *V. carteri* (Matt and Umen, 2018). This might reflect the strong dependence on heterotrophy of *Astrephomene* (Brooks, 1972). However, *A. gubernaculifera* and *V. carteri* showed the common overall tendency in terms of the function of two cell-types: reproductive cells are specialized in cell growth and reproduction, while somatic cells are specialized in motility and ECM biosynthesis instead of cell growth (Chapter 4, Fig. 4-16).

Thus, this thesis suggested the differences in developmental and molecular bases involved in the evolution of spheroidal colonies and germ–soma differentiation, respectively, between

Astrephomene and Volvocaceae. The “parallel evolution” of these multicellular traits in volvocine algae might have to be regarded as “convergent evolution” according to their different origins in developmental and molecular levels, though the distinction of “parallel evolution” and “convergent evolution” is sometimes difficult and might be arbitrary (Arendt and Reznick, 2008). The convergent evolution of multicellular traits within a single multicellular lineage is also reported in other multicellular lineages by recent comparative genomic analyses. For example, Ctenophora (comb jellies) has evolved a nervous system without known patterning genes and with expansion of receptor families different from Bilateria and Cnidaria (Moroz et al., 2014). In fungi, on the other hand, two clades which have evolved fruiting bodies independently, Agaricomycotina and Pezizomycotina, exhibited the convergent expansion of developmental gene families (Merényi et al., 2020). The present results indicate that such convergence of molecular bases within a single multicellular lineage also occurs in the initial steps of the evolution of multicellularity, simple three-dimensional body plans and germ–soma differentiation.

This thesis also provided the tools to resolve the parallel evolution of multicellular traits within volvocine algae: the newly established strain, synchronous culture, the time-lapse analysis method, the isolation method of somatic and reproductive cells, and genome data of *A. gubernaculifera*. On the other hand, transformation methods for *Gonium*, *Pandorina* and *Eudorina* have been established in recent years (Lerche and Hallmann, 2009, 2013, 2014), which might be also applicable to *Astrephomene* easily. The establishment of a transformation method for *Astrephomene* would enable functional analyses of orthologs of the embryogenesis-related genes or the somatic-specific transcription factors in *Astrephomene* and a search for molecular bases involved in rotation of daughter protoplasts. Moreover, the genome data of main genera in volvocine green algae have been constructed in recent years (Merchant et al., 2007; Prochnik et al., 2010; Hanschen et al., 2016; Featherston et al., 2018; Hamaji et al., 2018), which would facilitate further evolutionary studies using this group. Further studies on the parallel evolution of multicellular traits within

volvocine algae will provide deeper understanding for the evolution of multicellularity.

Acknowledgements

I would like to express my deep gratitude to my supervisor, Dr. Hisayoshi Nozaki (University of Tokyo), for his suggestion, discussion, proofreading of the manuscript and encouragements through this study. I also deeply appreciate Dr. Hiroko Kawai-Toyooka (Hosei University) and Dr. Yoko Arakaki (former University of Tokyo) for technical guidance, supports, suggestion and a lot of discussion about the developmental analyses of volvocine algae. I would like to thank Dr. Osami Misumi (Yamaguchi University) and Ms. Mizuho Sugawara (former University of Tokyo) for providing soil samples to establish new strains of *Astrephomene*. I would like to thank Dr. Kaoru Kawafune (former Tokyo Institute of Technology) and Dr. Takumi Murakami (National Institute of Genetics) for providing soil samples to establish new strains of *Gonium*. I would like to thank Dr. Masafumi Hirono (Hosei University) and Mr. Akira Noga (Paul Scherrer Institut) for providing anti-CrSAS-6 antibody and helpful suggestion. I would like to thank Dr. Atsushi Toyoda (National Institute of Genetics) and Dr. Yohei Minakuchi (National Institute of Genetics) for the great contribution in the *de novo* sequencing, assembly and annotation of the whole genome data of *A. gubernaculifera*. I would like to thank Dr. Masanobu Kawachi (National Institute for Environmental Studies), Dr. Haruyo Yamaguchi (National Institute for Environmental Studies), Dr. Shigekatsu Suzuki (National Institute for Environmental Studies) and Dr. Ryo Matsuzaki (National Institute for Environmental Studies) for helping the RNA-seq analysis of *A. gubernaculifera*. I would like to thank Dr. Shin-ya Miyagishima (National Institute of Genetics) and Dr. Shunsuke Hirooka (National Institute of Genetics), Ms. Kayoko Yamamoto (University of Tokyo) and Mr. Atsushi Suguchi (former University of Tokyo) for technical guidance, supports, suggestion and a lot of discussion about the analyses for the genome data of *A. gubernaculifera*. I would like to thank Dr. Zachariah Grochau-Wright (The College of New Jersey) for providing the alignment data of VARL genes. And at last I would like thank all the members of Laboratory of Origin of Eukaryote Biodiversity (University of Tokyo) for a lot of suggestion and discussion.

Chapter 2 was reproduced from Yamashita et al. (2016) BMC Evol. Biol. 16: 243 and

Chapter 3 was reproduced from Yamashita and Nozaki (2019) BMC Evol. Biol. 19: 120.

References

- Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., et al. (2019). Revisions to the classification, nomenclature, and diversity of eukaryotes. *J. Eukaryot. Microbiol.* *66*, 4–119.
- Albert, V.A., Barbazuk, W.B., dePamphilis, C.W., Der, J.P., Leebens-Mack, J., Ma, H., Palmer, J.D., Rounsley, S., Sankoff, D., Schuster, S.C., et al. (2013). The *Amborella* genome and the evolution of flowering plants. *Science* *342*, 1241089.
- Allen, J.F., de Paula, W.B.M., Puthiyaveetil, S., and Nield, J. (2011). A structural phylogenetic map for chloroplast photosynthesis. *Trends Plant Sci.* *16*, 645–655.
- Arakaki, Y., Kawai-Toyooka, H., Hamamura, Y., Higashiyama, T., Noga, A., Hirono, M., Olson, B.J.S.C., and Nozaki, H. (2013). The simplest integrated multicellular organism unveiled. *PLOS ONE* *8*, e81641.
- Arakaki, Y., Fujiwara, T., Kawai-Toyooka, H., Kawafune, K., Featherston, J., Durand, P.M., Miyagishima, S., and Nozaki, H. (2017). Evolution of cytokinesis-related protein localization during the emergence of multicellularity in volvocine green algae. *BMC Evol. Biol.* *17*, 243.
- Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* *9*, 868–882.
- Arendt, J., and Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* *23*, 26–32.
- Ball, S.G., and Deschamps, P. (2009). Chapter 1- Starch Metabolism. In *The Chlamydomonas Sourcebook Second Edition Volume 2*, (San Diego, CA: Academic Press), pp. 1–40.
- Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., et al. (2011). The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* *332*, 960–963.
- Barker, A.R., Renzaglia, K.S., Fry, K., and Dawe, H.R. (2014). Bioinformatic analysis of ciliary transition zone proteins reveals insights into the evolution of ciliopathy networks. *BMC*

Genomics 15, 531.

- Blaby, I.K., Blaby-Haas, C.E., Pérez-Pérez, M.E., Schmollinger, S., Fitz-Gibbon, S., Lemaire, S.D., and Merchant, S.S. (2015). Genome-wide analysis on *Chlamydomonas reinhardtii* reveals the impact of hydrogen peroxide on protein stress responses and overlap with other stress transcriptomes. *Plant J.* 84, 974–988.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171, 287–304.e15.
- Brooks, A.E. (1966). The sexual cycle and intercrossing in the genus *Astrephomene*. *J. Protozool.* 13, 367–375.
- Brooks, A.E. (1972). The physiology of *Astrephomene gubernaculifera*. *J. Protozool.* 19, 195–199.
- Brunet, T., and King, N. (2017). The origin of animal multicellularity and cell differentiation. *Dev. Cell* 43, 124–140.
- Brunet, T., Larson, B.T., Linden, T.A., Vermeij, M.J.A., McDonald, K., and King, N. (2019). Light-regulated collective contractility in a multicellular choanoflagellate. *Science* 366, 326–334.
- Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60.
- Burgess, S.J., Taha, H., Yeoman, J.A., Iamshanova, O., Chan, K.X., Boehm, M., Behrends, V., Bundy, J.G., Bialek, W., Murray, J.W., et al. (2016). Identification of the elusive pyruvate reductase of *Chlamydomonas reinhardtii* chloroplasts. *Plant Cell Physiol.* 57, 82–94.
- Camargo, A., Llamas, Á., Schnell, R.A., Higuera, J.J., González-Ballester, D., Lefebvre, P.A., Fernández, E., and Galván, A. (2007). Nitrate signaling by the regulatory gene *NIT2* in

Chlamydomonas. Plant Cell 19, 3491–3503.

- Chardin, C., Girin, T., Roudier, F., Meyer, C., and Krapp, A. (2014). The plant RWP-RK transcription factors: key regulators of nitrogen responses and of gametophyte development. J. Exp. Bot. 65, 5577–5587.
- Cheng, Q., Pappas, V., Hallmann, A., and Miller, S.M. (2005). Hsp70A and GlbA interact as partner chaperones to regulate asymmetric division in *Volvox*. Dev. Biol. 286, 537–548.
- Cheng, Q., Hallmann, A., Edwards, L., and Miller, S.M. (2006). Characterization of a heat-shock-inducible *hsp70* gene of the green alga *Volvox carteri*. Gene 371, 112–120.
- Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods 13, 1050–1054.
- Choi, G., Przybylska, M., and Straus, D. (1996). Three abundant germ line-specific transcripts in *Volvox carteri* encode photosynthetic proteins. Curr. Genet. 30, 347–355.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676.
- Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. 37, 291–294.
- Dayel, M.J., Alegado, R.A., Fairclough, S.R., Levin, T.C., Nichols, S.A., McDonald, K., and King, N. (2011). Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. Dev. Biol. 357, 73–82.
- Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A.Z., Robbens, S., Partensky, F., Degroove, S., Echeynie, S., Cooke, R., et al. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proc. Natl. Acad. Sci. U. S. A.

103, 11647–11652.

- Desnitski, A.G. (1995). A review on the evolution of development in *Volvox* — morphological and physiological aspects. *Eur. J. Protistol.* 31, 241–247.
- Du, H., Liang, Z., Zhao, S., Nan, M.-G., Tran, L.-S.P., Lu, K., Huang, Y.-B., and Li, J.-N. (2015a). The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: new insights into subfamily classification and expansion. *Sci. Rep.* 5, 11037.
- Du, Q., Kawabe, Y., Schilde, C., Chen, Z., and Schaap, P. (2015b). The evolution of aggregative multicellularity and cell–cell communication in the *Dictyostelia*. *J. Mol. Biol.* 427, 3722–3733.
- Duncan, L., Nishii, I., Harryman, A., Buckley, S., Howard, A., Friedman, N.R., and Miller, S.M. (2007). The *VARL* gene family and the evolutionary origins of the master cell-type regulatory gene, *regA*, in *Volvox carteri*. *J. Mol. Evol.* 65, 1–11.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432.
- Elrad, D., and Grossman, A.R. (2004). A genome’s-eye view of the light-harvesting polypeptides of *Chlamydomonas reinhardtii*. *Curr. Genet.* 45, 61–75.
- Fairclough, S.R., Dayel, M.J., and King, N. (2010). Multicellular development in a choanoflagellate. *Curr. Biol.* 20, R875–R876.
- Fairclough, S.R., Chen, Z., Kramer, E., Zeng, Q., Young, S., Robertson, H.M., Begovic, E., Richter, D.J., Russ, C., Westbrook, M.J., et al. (2013). Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* 14, R15.
- Fawley, M.W., and Fawley, K.P. (2004). A simple and rapid technique for the isolation of DNA from

- microalgae. *J. Phycol.* *40*, 223–225.
- Featherston, J., Arakaki, Y., Hanschen, E.R., Ferris, P.J., Michod, R.E., Olson, B.J.S.C., Nozaki, H., and Durand, P.M. (2018). The 4-celled *Tetrabaena socialis* nuclear genome reveals the essential components for genetic control of cell number at the origin of multicellularity in the volvocine lineage. *Mol. Biol. Evol.* *35*, 855–870.
- Feldman, J.L., and Marshall, W.F. (2009). *ASQ2* encodes a TBCC-like protein required for mother-daughter centriole linkage and mitotic spindle orientation. *Curr. Biol.* *19*, 1238–1243.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* *39*, 783–791.
- Fernandez, E., Llamas, A., and Galvan, A. (2009). Chapter 3 - Nitrogen Assimilation and its Regulation. In *The Chlamydomonas Sourcebook Second Edition Volume 2*, (San Diego, CA: Academic Press), pp. 69–113.
- Ferris, P.J., and Goodenough, U.W. (1997). Mating Type in *Chlamydomonas* Is Specified by *mid*, the Minus-Dominance Gene. *Genetics* *146*, 859–869.
- Fontaine, J.-X., Tercé-Laforgue, T., Armengaud, P., Clément, G., Renou, J.-P., Pelletier, S., Catterou, M., Azzopardi, M., Gibon, Y., Lea, P.J., et al. (2012). Characterization of a NADH-dependent glutamate dehydrogenase mutant of *Arabidopsis* demonstrates the key role of this enzyme in root carbon and nitrogen metabolism. *Plant Cell* *24*, 4044–4065.
- Franzen, W. (1988). Oogenesis and larval development of *Scypha ciliata* (Porifera, Calcarea). *Zoomorphology* *107*, 349–357.
- Fulton, A.B. (1978). Colonial development in *Pandorina morum*: II. Colony morphogenesis and formation of the extracellular matrix. *Dev. Biol.* *64*, 236–251.
- Gargouri, M., Park, J.-J., Holguin, F.O., Kim, M.-J., Wang, H., Deshpande, R.R., Shachar-Hill, Y., Hicks, L.M., and Gang, D.R. (2015). Identification of regulatory network hubs that control lipid metabolism in *Chlamydomonas reinhardtii*. *J. Exp. Bot.* *66*, 4551–4566.

- Gerisch, G. (1959). Die Zelldifferenzierung bei *Pleodorina californica* Shaw und die Organisation der Phytomonadenkolonien. Arch. Für Protistenkd. 104, 292–358.
- Godl, K., Hallmann, A., Rappel, A., and Sumper, M. (1995). Pherophorins: a family of extracellular matrix glycoproteins from *Volvox* structurally related to the sex-inducing pheromone. Planta 196, 781–787.
- Godl, K., Hallmann, A., Wenzl, S., and Sumper, M. (1997). Differential targeting of closely related ECM glycoproteins: the pherophorin family from *Volvox*. EMBO J. 16, 25–34.
- Goodenough, U., Blaby, I., Casero, D., Gallaher, S.D., Goodson, C., Johnson, S., Lee, J.-H., Merchant, S.S., Pellegrini, M., Roth, R., et al. (2014). The path to triacylglyceride obesity in the *sta6* strain of *Chlamydomonas reinhardtii*. Eukaryot. Cell 13, 591–613.
- Gottlieb, B., and Goldstein, M.E. (1977). Colony development in *Eudorina elegans* (Chlorophyta, Volvocales). J. Phycol. 13, 358–364.
- Green, K., and Kirk, D.L. (1981). Cleavage patterns, cell lineages, and development of a cytoplasmic bridge system in *Volvox* embryos. J. Cell Biol. 91, 743–755.
- Green, K.J., and Kirk, D.L. (1982). A revision of the cell lineages recently reported for *Volvox carteri* embryos. J. Cell Biol. 94, 741–742.
- Green, K., Viamontes, G.I., and Kirk, D.L. (1981). Mechanism of formation, ultrastructure, and function of the cytoplasmic bridge system during morphogenesis in *Volvox*. J. Cell Biol. 91, 756–769.
- Grochau-Wright, Z.I., Hanschen, E.R., Ferris, P.J., Hamaji, T., Nozaki, H., Olson, B.J.S.C., and Michod, R.E. (2017). Genetic basis for soma is present in undifferentiated volvocine green algae. J. Evol. Biol. 30, 1205–1218.
- Grosberg, R.K., and Strathmann, R.R. (2007). The evolution of multicellularity: a minor major transition? Annu. Rev. Ecol. Evol. Syst. 38, 621–654.
- Hallmann, A. (2003). Extracellular matrix and sex-inducing pheromone in *Volvox*. Int. Rev. Cytol.

227, 131–182.

- Hallmann, A. (2006). Morphogenesis in the family Volvocaceae: different tactics for turning an embryo right-side out. *Protist* 157, 445–461.
- Hamaji, T., Kawai-Toyooka, H., Uchimura, H., Suzuki, M., Noguchi, H., Minakuchi, Y., Toyoda, A., Fujiyama, A., Miyagishima, S., Umen, J.G., et al. (2018). Anisogamy evolved with a reduced sex-determining region in volvocine green algae. *Commun. Biol.* 1, 17.
- Hanschen, E.R., Ferris, P.J., and Michod, R.E. (2014). Early evolution of the genetic basis for soma in the Volvocaceae: evolution of the genetic basis for soma. *Evolution* 68, 2014–2025.
- Hanschen, E.R., Marriage, T.N., Ferris, P.J., Hamaji, T., Toyoda, A., Fujiyama, A., Neme, R., Noguchi, H., Minakuchi, Y., Suzuki, M., et al. (2016). The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat. Commun.* 7, 11370.
- Harper, R.A. (1912). The structure and development of the colony in *Gonium*. *Trans. Am. Microsc. Soc.* 31, 65–82.
- Hartmann, M. (1924). Über die Veränderung der Koloniebildung von *Eudorina elegans* und *Gonium pectorale* unter dem Einfluß äußerer Bedingungen. IV. Mitt. der Untersuchungen über die Morphologie und Physiologie des Formwechsels der Phytomonaden (Volvocales). *Arch. Für Protistenkd.* 49, 375–395.
- Hayama, M., Nakada, T., Hamaji, T., and Nozaki, H. (2010). Morphology, molecular phylogeny and taxonomy of *Gonium maiaprilis* sp. nov. (Goniaceae, Chlorophyta) from Japan. *Phycologia* 49, 221–234.
- Hemschemeier, A., Casero, D., Liu, B., Benning, C., Pellegrini, M., Happe, T., and Merchant, S.S. (2013). COPPER RESPONSE REGULATOR1-dependent and -independent responses of the *Chlamydomonas reinhardtii* transcriptome to dark anoxia. *Plant Cell* 25, 3186–3211.
- Herron, M.D., and Michod, R.E. (2008). Evolution of complexity in the volvocine algae: transitions

- in individuality through Darwin's eye. *Evolution* 62, 436–451.
- Herron, M.D., Hackett, J.D., Aylward, F.O., and Michod, R.E. (2009). Triassic origin and early radiation of multicellular volvocine algae. *Proc. Natl. Acad. Sci. U. S. A.* 106, 3254–3258.
- Herron, M.D., Desnitskiy, A.G., and Michod, R.E. (2010). Evolution of developmental programs in *Volvox* (Chlorophyta). *J. Phycol.* 46, 316–324.
- Heyde, B., and Hallmann, A. (2020). Targeted migration of pteroporphyrin-S indicates extensive extracellular matrix dynamics in *Volvox carteri*. *Plant J.* 103, 2301–2317.
- Hiraide, R., Kawai-Toyooka, H., Hamaji, T., Matsuzaki, R., Kawafune, K., Abe, J., Sekimoto, H., Umen, J., and Nozaki, H. (2013). The evolution of male–female sexual dimorphism predates the gender-based divergence of the mating locus gene *MAT3/RB*. *Mol. Biol. Evol.* 30, 1038–1040.
- Hisanaga, T., Okahashi, K., Yamaoka, S., Kajiwar, T., Nishihama, R., Shimamura, M., Yamato, K.T., Bowman, J.L., Kohchi, T., and Nakajima, K. (2019). A *cis*-acting bidirectional transcription switch controls sexual dimorphism in the liverwort. *EMBO J.* 38.
- Hodges, M.E., Scheumann, N., Wickstead, B., Langdale, J.A., and Gull, K. (2010). Reconstructing the evolutionary history of the centriole from protein components. *J. Cell Sci.* 123, 1407–1413.
- Höhn, S., and Hallmann, A. (2011). There is more than one way to turn a spherical cellular monolayer inside out: type B embryo inversion in *Volvox globator*. *BMC Biol.* 9, 89.
- Höhn, S., and Hallmann, A. (2016). Distinct shape-shifting regimes of bowl-shaped cell sheets – embryonic inversion in the multicellular green alga *Pleodorina*. *BMC Dev. Biol.* 16, 35.
- Hoops, H.J. (1993). Flagellar, cellular and organismal polarity in *Volvox carteri*. *J. Cell Sci.* 104, 105–117.
- Hoops, H.J. (1997). Motility in the colonial and multicellular Volvocales: structure, function, and evolution. *Protoplasma* 199, 99–112.
- Hoops, H.J., and Floyd, G.L. (1982). Mitosis, cytokinesis and colony formation in the colonial green

- alga *Astrephomene gubernaculifera*. Br. Phycol. J. 17, 297–310.
- Hoops, H.J., and Floyd, G.L. (1983). Ultrastructure and development of the flagellar apparatus and flagellar motion in the colonial green-alga *Astrephomene gubernaculifera*. J. Cell Sci. 63, 21–41.
- Hoops, H.J., Long, J.J., and Hile, E.S. (1994). Flagellar apparatus structure is similar but not identical in *Volvox steinii*, *Eudorina elegans* and *Pleodorina illinoisensis* (Chlorophyta): implications for the “volvocine evolutionary lineage.” J. Phycol. 30, 679–689.
- Hori, K., Maruyama, F., Fujisawa, T., Togashi, T., Yamamoto, N., Seo, M., Sato, S., Yamada, T., Mori, H., Tajima, N., et al. (2014). *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. Nat. Commun. 5, 3978.
- Iida, H., Nishii, I., and Inouye, I. (2011). Embryogenesis and cell positioning in *Platydorina caudata* (Volvocaceae, Chlorophyta). Phycologia 50, 530–540.
- Iida, H., Ota, S., and Inouye, I. (2013). Cleavage, incomplete inversion, and cytoplasmic bridges in *Gonium pectorale* (Volvocales, Chlorophyta). J. Plant Res. 126, 699–707.
- Ireland, G.W., and Hawkins, S.E. (1981). Inversion in *Volvox tertius*: the effects of Con A. J. Cell Sci. 48, 355–366.
- Johnson, X., and Alric, J. (2013). Central carbon metabolism and electron transport in *Chlamydomonas reinhardtii*: metabolic constraints for carbon partitioning between oil and starch. Eukaryot. Cell 12, 776–793.
- Kasahara, R.D., Portereiko, M.F., Sandaklie-Nikolova, L., Rabiger, D.S., and Drews, G.N. (2005). *MYB98* is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. Plant Cell 17, 2981–2992.
- Kawachi, M., Ishimoto, M., Mori, F., Yumoto, K., Sato, M., and Noël, M.-H. (2013). MCC-NIES list of strains, 9th edition, microbial culture collection at National Institute for Environmental Studies, Tsukuba, Japan. https://mcc.nies.go.jp/download/list9th_e.pdf

- Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., et al. (2014). The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol.* *12*, e1001889.
- Kelland, J.L. (1977). Inversion in *Volvox* (Chlorophyceae). *J. Phycol.* *13*, 373–378.
- Keller, L.C., Romijn, E.P., Zamora, I., Yates, J.R., and Marshall, W.F. (2005). Proteomic analysis of isolated *Chlamydomonas* centrioles reveals orthologs of ciliary-disease genes. *Curr. Biol.* *15*, 1090–1098.
- Keller, L.C., Geimer, S., Romijn, E., Yates, J., Zamora, I., and Marshall, W.F. (2009). Molecular architecture of the centriole proteome: The conserved WD40 domain protein POC1 is required for centriole duplication and length control. *Mol. Biol. Cell* *20*, 1150–1166.
- Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* *37*, 907–915.
- Kindle, K.L., Schnell, R.A., Fernández, E., and Lefebvre, P.A. (1989). Stable nuclear transformation of *Chlamydomonas* using the *Chlamydomonas* gene for nitrate reductase. *J. Cell Biol.* *109*, 2589–2601.
- King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* *451*, 783–788.
- Kirk, D.L. (1998). *Volvox: Molecular-genetic origins of multicellularity and cellular differentiation* (Cambridge: Cambridge University Press).
- Kirk, D.L. (2005). A twelve-step program for evolving multicellularity and a division of labor. *BioEssays* *27*, 299–310.
- Kirk, D.L., and Kirk, M.M. (1983). Protein synthetic patterns during the asexual life cycle of *Volvox carteri*. *Dev. Biol.* *96*, 493–506.

- Kirk, D.L., and Nishii, I. (2001). *Volvox carteri* as a model for studying the genetic and cytological control of morphogenesis. *Dev. Growth Differ.* *43*, 621–631.
- Kirk, D.L., Birchem, R., and King, N. (1986). The extracellular matrix of *Volvox*: a comparative study and proposed system of nomenclature. *J. Cell Sci.* *80*, 207–231.
- Kirk, M.M., Ransick, A., McRae, S.E., and Kirk, D.L. (1993). The relationship between cell size and cell fate in *Volvox carteri*. *J. Cell Biol.* *123*, 191–208.
- Kirk, M.M., Stark, K., Miller, S.M., Muller, W., Taillon, B.E., Gruber, H., Schmitt, R., and Kirk, D.L. (1999). *regA*, a *Volvox* gene that plays a central role in germ-soma differentiation, encodes a novel regulatory protein. *Development* *126*, 639–647.
- Klein, B., Wibberg, D., and Hallmann, A. (2017). Whole transcriptome RNA-Seq analysis reveals extensive cell type-specific compartmentalization in *Volvox carteri*. *BMC Biol.* *15*.
- Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., and Steuer, R. (2013). Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis* sp. PCC 6803. *PLOS Comput. Biol.* *9*, e1003081.
- Koufopanou, V. (1994). The evolution of soma in the Volvocales. *Am. Nat.* *143*, 907–931.
- Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* *20*, 278.
- Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* *35*, 4453–4455.
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* *33*, 1870–1874.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., et al. (2012). The Arabidopsis Information Resource

- (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* *40*, D1202–D1210.
- Lerche, K., and Hallmann, A. (2009). Stable nuclear transformation of *Gonium pectorale*. *BMC Biotechnol.* *9*, 64.
- Lerche, K., and Hallmann, A. (2013). Stable nuclear transformation of *Eudorina elegans*. *BMC Biotechnol.* *13*, 11.
- Lerche, K., and Hallmann, A. (2014). Stable nuclear transformation of *Pandorina morum*. *BMC Biotechnol.* *14*, 65.
- 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.
- Li-Beisson, Y., Beisson, F., and Riekhof, W. (2015). Metabolism of acyl-lipids in *Chlamydomonas reinhardtii*. *Plant J.* *82*, 504–522.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Maldonado, M. (2005). Choanoflagellates, choanocytes, and animal multicellularity. *Invertebr. Biol.* *123*, 1–22.
- Marchant, H.J. (1977). Colony formation and inversion in the green alga *Eudorina elegans*. *Protoplasma* *93*, 325–339.
- Martindale, M.Q. (2005). The evolution of metazoan axial properties. *Nat. Rev. Genet.* *6*, 917–927.
- Matsuzaki, M., Misumi, O., Shin-i, T., Maruyama, S., Takahara, M., Miyagishima, S., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., et al. (2004). Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* *428*, 653–657.
- Matt, G.Y., and Umen, J.G. (2018). Cell-type transcriptomes of the multicellular green alga *Volvox carteri* yield insights into the evolutionary origins of germ and somatic differentiation programs. *G3 (Bethesda)* *8*, 531–550.
- Meissner, M., Stark, K., Cresnar, B., Kirk, D.L., and Schmitt, R. (1999). *Volvox* germline-specific

- genes that are putative targets of RegA repression encode chloroplast proteins. *Curr. Genet.* 36, 363–370.
- Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L., et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318, 245–250.
- Merényi, Z., Prasanna, A.N., Wang, Z., Kovács, K., Hegedüs, B., Bálint, B., Papp, B., Townsend, J.P., and Nagy, L.G. (2020). Unmatched level of molecular convergence among deeply divergent complex multicellular fungi. *Mol. Biol. Evol.* 37, 2228–2240.
- Michod, R.E. (2006). On the transfer of fitness from the cell to the multicellular organism. *Biol. Philos.* 20, 967–987.
- Michod, R.E., Viossat, Y., Solari, C.A., Hurand, M., and Nedelcu, A.M. (2006). Life-history evolution and the origin of multicellularity. *J. Theor. Biol.* 239, 257–272.
- Mikhailov, K.V., Konstantinova, A.V., Nikitin, M.A., Troshin, P.V., Rusin, L.Yu., Lyubetsky, V.A., Panchin, Y.V., Mylnikov, A.P., Moroz, L.L., Kumar, S., et al. (2009). The origin of Metazoa: a transition from temporal to spatial cell differentiation. *BioEssays* 31, 758–768.
- Miller, S.M., and Kirk, D.L. (1999). *glsA*, a *Volvox* gene required for asymmetric division and germ cell specification, encodes a chaperone-like protein. *Development* 126, 649–658.
- Miller, S.M., Schmitt, R., and Kirk, D.L. (1993). Jordan, an active *Volvox* transposable element similar to higher plant transposons. *Plant Cell* 5, 1125–1138.
- Moroz, L.L., Kocot, K.M., Citarella, M.R., Dosung, S., Norekian, T.P., Povolotskaya, I.S., Grigorenko, A.P., Dailey, C., Berezikov, E., Buckley, K.M., et al. (2014). The ctenophore genome and the evolutionary origins of neural systems. *Nature* 510, 109–114.
- Mukherjee, K., Brocchieri, L., and Burglin, T.R. (2009). A comprehensive classification and evolutionary analysis of plant homeobox genes. *Mol. Biol. Evol.* 26, 2775–2794.
- Nagy, L.G., Kovács, G.M., and Krizsán, K. (2018). Complex multicellularity in fungi: evolutionary

- convergence, single origin, or both? *Biol. Rev.* *93*, 1778–1794.
- Nakada, T., and Nozaki, H. (2007). Re-evaluation of three *Chlorogonium* (Volvocales, Chlorophyceae) species based on 18S ribosomal RNA gene phylogeny. *Eur. J. Phycol.* *42*, 177–182.
- Nakazawa, Y., Hiraki, M., Kamiya, R., and Hirono, M. (2007). SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. *Curr. Biol.* *17*, 2169–2174.
- Nedelcu, A.M. (2009). Environmentally induced responses co-opted for reproductive altruism. *Biol. Lett.* *5*, 805–808.
- Nedelcu, A.M., and Michod, R.E. (2006). The evolutionary origin of an altruistic gene. *Mol. Biol. Evol.* *23*, 1460–1464.
- Nematollahi, G., Kianianmomeni, A., and Hallmann, A. (2006). Quantitative analysis of cell-type specific gene expression in the green alga *Volvox carteri*. *BMC Genomics* *7*, 321.
- Nickelsen, J., and Rengstl, B. (2013). Photosystem II assembly: from cyanobacteria to plants. *Annu. Rev. Plant Biol.* *64*, 609–635.
- Niklas, K. (2000). The evolution of plant body plans—a biomechanical perspective. *Ann. Bot.* *85*, 411–438.
- Niklas, K.J., and Kutschera, U. (2009). The evolutionary development of plant body plans. *Funct. Plant Biol.* *36*, 682.
- Nishii, I., and Miller, S.M. (2010). *Volvox*: Simple steps to developmental complexity? *Curr. Opin. Plant Biol.* *13*, 646–653.
- Nishii, I., Ogiwara, S., and Kirk, D.L. (2003). A kinesin, InvA, plays an essential role in *Volvox* morphogenesis. *Cell* *113*, 743–753.
- Nozaki, H. (1983). Morphology and taxonomy of two species of *Astrephomene* (Chlorophyta, Volvocales) in Japan. *J. Jpn. Bot.* *58*, 345–352.
- Nozaki, H. (1990). Ultrastructure of the extracellular matrix of *Gonium* (Volvocales, Chlorophyta).

Phycologia 29, 1–8.

- Nozaki, H., and Ito, M. (1994). Phylogenetic relationships within the colonial Volvocales (Chlorophyta) inferred from cladistic analysis based on morphological data. *J. Phycol.* 30, 353–365.
- Nozaki, H., Kuroiwa, H., Mita, T., and Kuroiwa, T. (1989). *Pleodorina japonica* sp. nov. (Volvocales, Chlorophyta) with bacteria-like endosymbionts. *Phycologia* 28, 252–267.
- Nozaki, H., Ito, M., Sano, R., Uchida, H., Watanabe, M.M., and Kuroiwa, T. (1995). Phylogenetic relationships within the colonial Volvocales (Chlorophyta) inferred from *rbcL* gene sequence data. *J. Phycol.* 31, 970–979.
- Nozaki, H., Ito, M., Sano, R., Uchida, H., Watanabe, M.M., Takahashi, H., and Kuroiwa, T. (1997). Phylogenetic analysis of *Yamagishiella* and *Platydorina* (Volvocaceae, Chlorophyta) based on *rbcL* gene sequences. *J. Phycol.* 33, 272–278.
- Nozaki, H., Ohta, N., Takano, H., and Watanabe, M.M. (1999). Reexamination of phylogenetic relationships within the colonial Volvocales (Chlorophyta): an analysis of *atpB* and *rbcL* gene sequences. *J. Phycol.* 35, 104–112.
- Nozaki, H., Misawa, K., Kajita, T., Kato, M., Nohara, S., and Watanabe, M.M. (2000). Origin and evolution of the colonial Volvocales (Chlorophyceae) as inferred from multiple, chloroplast gene sequences. *Mol. Phylogenet. Evol.* 17, 256–268.
- Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., et al. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584.
- Perez-Garcia, O., Escalante, F.M.E., de-Bashan, L.E., and Bashan, Y. (2011). Heterotrophic cultures of microalgae: Metabolism and potential products. *Water Res.* 45, 11–36.
- Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Quéinnec, E., et al. (2009). Phylogenomics revives traditional views on deep

- animal relationships. *Curr. Biol.* *19*, 706–712.
- Pickett-Heaps, J.D. (1970). Some ultrastructural features of *Volvox*, with particular reference to the phenomenon of inversion. *Planta* *90*, 174–190.
- Pocock, M.A. (1937). Studies in South African Volvocales. *Proc. Linn. Soc. Lond.* *149*, 55–58.
- Pocock, M.A. (1954). Two multicellular motile green algae, *Volvulina* Playfair and *Astrephomene*, a new genus. *Trans. R. Soc. South Afr.* *34*, 103–127.
- Pringsheim, E.G. (1946). *Pure Cultures of Algae* (Cambridge: Cambridge University Press).
- Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., et al. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* *329*, 223–226.
- Pusztai, M., and Škaloud, P. (2019). Elucidating the evolution and diversity of *Uroglena* -like colonial flagellates (Chrysophyceae): polyphyletic origin of the morphotype. *Eur. J. Phycol.* *54*, 404–416.
- Queller, D.C. (2000). Relatedness and the fraternal major transitions. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* *355*, 1647–1655.
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* *33*, W116–W120.
- Rabiger, D.S., and Drews, G.N. (2013). MYB64 and MYB119 are required for cellularization and differentiation during female gametogenesis in *Arabidopsis thaliana*. *PLOS Genet.* *9*, e1003783.
- Rensing, S.A. (2016). (Why) Does evolution favour embryogenesis? *Trends Plant Sci.* *21*, 562–573.
- Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.-F., Lindquist, E.A., Kamisugi, Y., et al. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* *319*, 64–69.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L.,

- Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* *61*, 539–542.
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* *4*, 406–425.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, second edition (New York: Cold Spring Harbor Laboratory Press).
- Schiedlmeier, B., Schmitt, R., Muller, W., Kirk, M.M., Gruber, H., Mages, W., and Kirk, D.L. (1994). Nuclear transformation of *Volvox carteri*. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 5080–5084.
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* *27*, 863–864.
- Schöttler, M.A., Tóth, S.Z., Boulouis, A., and Kahlau, S. (2015). Photosynthetic complex stoichiometry dynamics in higher plants: biogenesis, function, and turnover of ATP synthase and the cytochrome b6f complex. *J. Exp. Bot.* *66*, 2373–2400.
- Škaloud, P., Škaloudová, M., Jadrná, I., Bestová, H., Pusztai, M., Kapustin, D., and Siver, P.A. (2020). Comparing morphological and molecular estimates of species diversity in the freshwater genus *Synura* (Stramenopiles): A model for understanding diversity of eukaryotic microorganisms. *J. Phycol.* *56*, 574–591.
- Smith, J.M., and Száthmary, E. (1995). *The Major Transitions in Evolution* (W. H. Freeman).
- Smith, C.L., Varoqueaux, F., Kittelmann, M., Azzam, R.N., Cooper, B., Winters, C.A., Eitel, M., Fasshauer, D., and Reese, T.S. (2014). Novel cell types, neurosecretory cells, and body plan of the early-diverging Metazoan *Trichoplax adhaerens*. *Curr. Biol.* *24*, 1565–1572.
- Solari, C.A., Kessler, J.O., and Michod, R.E. (2006a). A hydrodynamics approach to the evolution of multicellularity: flagellar motility and germ-soma differentiation in Volvocalean green algae. *Am. Nat.* *167*, 537–554.
- Solari, C.A., Ganguly, S., Kessler, J.O., Michod, R.E., and Goldstein, R.E. (2006b). Multicellularity

- and the functional interdependence of motility and molecular transport. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 1353–1358.
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* *24*, 637–644.
- Starr, R.C., and Zeikus, J.A. (1993). UTEX—The culture collection of algae at the University of Texas at Austin 1993 List of Cultures. *J. Phycol.* *29*, 1–106.
- Stein, J. (1958). A morphological study of *Astrephomene gubernaculifera* and *Volvulina steinii*. *Am. J. Bot.* *45*, 388–397.
- Strassert, J.F.H., Jamy, M., Mylnikov, A.P., Tikhonenkov, D.V., and Burki, F. (2019). New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. *Mol. Biol. Evol.* *36*, 757–765.
- Sumper, M., Berg, E., Wenzl, S., and Godl, K. (1993). How a sex pheromone might act at a concentration below 10^{−16} M. *EMBO J.* *12*, 831–836.
- Tam, L.-W., and Kirk, D.L. (1991). Identification of cell-type-specific genes of *Volvox carteri* and characterization of their expression during the asexual life cycle. *Dev. Biol.* *145*, 51–66.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* *30*, 2725–2729.
- Ueki, N., and Nishii, I. (2008). *Idaten* is a new cold-inducible transposon of *Volvox carteri* that can be used for tagging developmentally important genes. *Genetics* *180*, 1343–1353.
- Ueki, N., and Nishii, I. (2009). Controlled enlargement of the glycoprotein vesicle surrounding a *Volvox* embryo requires the InvB nucleotide-sugar transporter and is required for normal morphogenesis. *Plant Cell* *21*, 1166–1181.
- Ueki, N., Matsunaga, S., Inouye, I., and Hallmann, A. (2010). How 5000 independent rowers coordinate their strokes in order to row into the sunlight: Phototaxis in the multicellular green alga *Volvox*. *BMC Biol.* *8*, 103.

- Vallon, O., and Spalding, M.H. (2009). Chapter 4 - Amino Acid Metabolism. In *The Chlamydomonas Sourcebook* Second Edition Volume 2, (San Diego, CA: Academic Press), pp. 115–158.
- Viamontes, G.I., and Kirk, D.L. (1977). Cell shape changes and the mechanism of inversion in *Volvox*. *J. Cell Biol.* 75, 719–730.
- Viamontes, G.I., Fochtmann, L.J., and Kirk, D.L. (1979). Morphogenesis in *Volvox*: analysis of critical variables. *Cell* 17, 537–550.
- Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281–285.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* 9, e112963.
- Wang, S., and Ardekani, A. (2012). Inertial squirmer. *Phys. Fluids* 24, 101902.
- Wang, X., Niu, Q.-W., Teng, C., Li, C., Mu, J., Chua, N.-H., and Zuo, J. (2009). Overexpression of *PGA37/MYB118* and *MYB115* promotes vegetative-to-embryonic transition in *Arabidopsis*. *Cell Res.* 19, 224–235.
- Yamada, T.K., Tsuchikane, Y., Wu, J.-T., Sekimoto, H., Miyaji, K., and Nozaki, H. (2008). Morphology and molecular phylogeny of *Eudorina* sp. (Volvocaceae, Chlorophyceae) from Taiwan. *Hikobia* 15, 135–143.
- Yang, M. (2016). The *FOUR LIPS (FLP)* and *MYB88* genes conditionally suppress the production of nonstomatal epidermal cells in *Arabidopsis* cotyledons. *Am. J. Bot.* 103, 1559–1566.
- Yang, H., Liu, J., Wen, X., and Lu, C. (2015). Molecular mechanism of photosystem I assembly in oxygenic organisms. *Biochim. Biophys. Acta BBA - Bioenerg.* 1847, 838–848.
- Yang, W., Catalanotti, C., D'Adamo, S., Wittkopp, T.M., Ingram-Smith, C.J., Mackinder, L., Miller, T.E., Heuberger, A.L., Peers, G., Smith, K.S., et al. (2014). Alternative acetate production pathways in *Chlamydomonas reinhardtii* during dark anoxia and the dominant role of

chloroplasts in fermentative acetate production. *Plant Cell* 26, 4499–4518.

Yoshinaga, K., Ohta, T., Suzuki, Y., and Sugiura, M. (1988). *Chlorella* chloroplast DNA sequence containing a gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and a part of a possible gene for the β' subunit of RNA polymerase. *Plant Mol. Biol.* 10, 245–250.

Zones, J.M., Blaby, I.K., Merchant, S.S., and Umen, J.G. (2015). High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation. *Plant Cell* 27, 2743–2769.