A genomic study of the invasive red alga *Bangia atropurpurea* (Mertens ex Roth) C. Agardh

by

Vanessa Poletto Borges

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biology

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Suzanne Fredericq Professor University of Louisiana at Lafayette
Supervisor	Kirsten M. Müller Professor Biology
Supervisor	Andrew C. Doxey Associate Professor Biology
Internal Member	Jonathan D. S. Witt Associate Professor Biology
Internal Member	Jozef Nissimov Assistant Professor Biology
Internal-external Member	Dan Brown Professor Computer Science

Author’s declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis might be made electronically available to the public.

Abstract

The phylum Rhodophyta comprises the red algae, which are morphologically diverse organisms present in marine and freshwater environments. In addition to the economic importance of these organisms, fossil records of red algal species dating 1.2 and 1.4 billion years confirm the evolutionary importance of this group. Studies of plastid evolution have shown that Rhodophyta have their plastid derived from the first endosymbiotic event, having a common ancestor with green algae and glaucophyte. Red algal plastids have also played a crucial role in secondary endosymbiotic events, being responsible for the emergence of other algal plastid lineages. The red algae order, Bangiales, is an ancient lineage and molecular studies have focused on elucidating the phylogenetic relationships and differentiating species within this group, which is complicated by their known phenotypic plasticity.

The genus *Bangia* (Bangiales) has been observed in both marine coastal environments and in freshwater, and for many years it was thought that the organisms in these habitats were conspecific. Studies showed that the species could acclimate well when transitioning from one habitat to another, with preservation of morphology. Molecular research, however, has shown that the species are not conspecific, and that the freshwater species, *Bangia atropurpurea* (Mertens ex Roth) C. Agardh, is more closely related to species from the genus *Porphyra* than to the marine *Bangia*. This discovery points out other curiosities about *B. atropurpurea*: it is the only known freshwater Bangiales, it is considered an invasive species of the Great Lakes, is strictly asexual, and has a filamentous thallus while its closest relative has a foliose thallus. To understand the similarities and differences of *B. atropurpurea* when compared to other Rhodophyta genomes, this thesis investigates the species at a genomic level using high-throughput sequencing and bioinformatic approaches.

Metagenomic sequencing allows high-throughput analysis of DNA from environmental samples, which is relevant when analysing organisms that are difficult to grow in lab conditions, such as *B. atropurpurea*. In this thesis, I used metagenomic techniques to sequence and assemble fragments of a mixed pool of short DNA sequences. Due to the availability of existing reference Bangiales genomes, and the presence of large conserved orthologous regions across the genomes of red algae, it was possible to identify *B. atropurpurea* contigs from the metagenome and assemble a high quality draft chloroplast genome with 97% coverage (Chapter 2). Given the

evolutionary importance the red algae chloroplast, the chloroplast genome of *B. atropurpurea* was then used to calculate the rates of substitution in 74 chloroplast genes, allowing the identification of evolutionary patterns that provide insights into selective forces on genes that may have contributed to the invasion of the Great Lakes (Chapter 3). Chloroplast gene sequences were compared with genes from 29 Rhodophyta species originally isolated from marine, freshwater, salt marshes and hot springs. This extensive analysis revealed trends in substitution rates according to taxonomic groups but also by habitat, where the Bangiales exhibited the most diverse rates for the *atp*, *psa*, *rps* and *rpl* gene families. Moreover, the substitution rates indicated that within the Bangiales, *B. atropurpurea* is the most distinct species, with rates that contradict the trend observed in the other Bangiales species.

Continued analysis of metagenomic data allowed conserved nuclear genes from *B. atropurpurea* to be identified and retrieved. Curiously, several genes were identified that are related to meiosis (Chapter 4). *B. atropurpurea* is known for being an asexual organism, but its reproduction is not well studied as other Bangiales species, which are known for having 22 - 29 meiotic genes. With the use of the Eukaryotic Meiotic Toolkit as a reference, 13 genes required for meiosis machinery were detected in the *B. atropurpurea* draft genome, and a phylogenetic analysis was performed, providing new insights about the evolution of *B. atropurpurea*. Although it is still unclear if these genes are functional, it is interesting that the species has sexual genes and is still suspected of only undergoing asexual reproduction.

The research conducted in this thesis demonstrates the utility of metagenomic sequencing and assembly approaches in the analysis of red algae species, and the extent of biological and evolutionary insights that can be gained from this information. The chloroplast genome of *B. atropurpurea* represents the first chloroplast genome of a freshwater Rhodophyta retrieved using this methodology. The chloroplast genome is consistent with other Bangiales genomes regarding gene order and content, serving as a basis for future research of other *Bangia* and Bangiales species. This study also provides the most detailed information to date about the rates of substitution among chloroplast genes in Rhodophyta species, emphasizing the importance of the use of individual gene analyses when performing such investigations. Lastly, this thesis identified meiotic related genes in *B. atropurpurea*, amplifying the overall understanding of the organism and opening doors for future investigations.

Acknowledgments

I would like to acknowledge the funding for this research which was supported by an NSERC Discovery Grant to Kirsten M. Müller, and by the Science Without Borders Program, a grant provided by the Brazilian Government.

I would like to express my sincere gratitude to my patient and supportive supervisor, Prof. Kirsten M. Müller. I am extremely grateful for our friendly chats, meetings, and your support in my academic and personal endeavours. Thank you for opening the lab door for me and for believing in me when I was ready to give up. I also thank my co-supervisor, Dr. Andrew C. Doxey, for his support and guidance during this project, giving me a chance to learn about bioinformatics from scratch.

I extend my appreciation to the many members of the Müller lab. I was able to see our research group grow tremendously and am proud of it. I am truly grateful for getting to know all of you and being part of this group. Dr. Ellen S. Cameron, thank you for being an amazing colleague, giving me emotional support on top of coding tips and road trips. Gabriela Britto Lopez, we went from TA and student, to researcher and co-op, to Peer Success colleagues, to friends; Thank you for being the great person that you are.

Throughout my years at the University of Waterloo, many people had a great impact on the person I am today. I cannot thank you all enough for the various opportunities I had working on campus and all the great people I've met. Thank you, Science Undergraduate Office, for creating a role for me and making the Science Transition Coach a reality. Thank you, Student Success Office, for the countless opportunities and for contributing immensely to my professional growth.

I would also like to thank my parents. The most supportive people you will ever know, the kindest and most loving. It's been almost 7 years since I'm in Canada, and the 8,470 km of distance was not enough to create any barrier between us. I do miss you though, every day. Thank you for teaching me how to use the microscope when I could barely stand up, for taking me to the beach and for teaching me that every organism is beautiful and has an impact on our lives. I am forever grateful. Thank you for believing in me and being happy with my decisions.

An especial acknowledgement to Mr. Yiu Pong To, your unconditional support and partnership were key for me to be here today.

Dedication

This thesis work is dedicated to my beloved family and parents, Edison and Sandra, for nursing me with love and affection and for their dedication to supporting me for the success of my life. This work is also dedicated to my grandparents, I so wish you all could see this and celebrate this accomplishment with me. Grandma Solange, you would have loved Canada. Grandma Hilda, any of your unforgettable desserts would be much appreciated. Lastly, I dedicate this thesis to all my fellow grad students that had to face grad school while a global pandemic happens.

Table of Content

List of Figures

List of Tables

Chapter 1

General Introduction

1.1 Phylum Rhodophyta

The red algae (Phylum Rhodophyta) are an important evolutionary lineage observed in both freshwater and marine environments and are important primary producers in these habitats. This phylum includes about 7000 species, that range in morphology from single-celled organisms to large and multicellular organisms (Guiry & Guiry, 2017). The morphological characteristics that define this phylum are the presence of phycobilins, including phycoerythrin, which gives the characteristic red colour to this group (Wray, 1977), the absence of flagellated cells at any time during the life cycle, and triphasic isomorphic life history with a tetrasporophyte (most species in Class Florideophyceae) or biphasic heteromorphic life history (other classes) (Gabrielson et al., 1990; Graham et al, 2000; Yoon et al., 2006).

Members of this phylum are important food sources and have considerable economic importance [e.g., *Chondrus crispus* (Irish moss), *Porphyra* & *Pyropia* (nori)]. The substantial economic value of these highlights the great importance of phycological studies on this group. For example, species of *Porphyra* and *Pyropia yezoensis* (Class Bangiophyceae) are of considerable economic importance amounting to US$1.3 billion per year in cultivation fields in China, Japan and Korea (Sutherland et al., 2011; Blouin et al., 2011). This contributes greatly to the economies of these countries (Yang et al., 2017). Another example is the seaweed *Kappaphycus* (Class Florideophyceae), which is farmed extensively in the Philippines, and used around the globe by the food industry, and for its pharmaceutical and nutraceutical properties (Hayashi & Reiss 2012; Hurtado et al., 2014). Moreover, *Chondrus crispus* (Class Florideophyceae), farmed and processed in the United States, Philippines, Indonesia, China, Ireland and Canada (Chopin et al., 1999; Gao et al., 2016), has great economic and medicinal importance, due to the high levels of carrageenan that are studied for their tumour and virus inhibition, hypertension prevention, control of hyperglycemia, immunity enhancement, antibacterial and anti-inflammatory activities (Güven et al., 1991; Zhou et al., 2004; Kalitnik et al., 2015; Gao et al., 2016). In addition, the Rhodophyta are extremely diverse in secondary

metabolites and are considered the most important source of biologically active metabolites when compared to other algal groups (Kasanah et al., 2015).

The classification of red algae has changed extensively over the past few decades due to advances in various techniques, particularly gene sequence analyses and phylogenetics (Fig. 1.1). Before molecular studies, other characteristics were studied to classify members of the group, such as the absence of flagellar cells, the presence of a two-membrane plastid with unstacked thylakoids without the presence of chlorophyll *b* or *c*, and the presence of pit plugs (Pueschel & Cole, 1982; Garbary and Gabrielson, 1990; Bhattacharya & Medlin, 1995; Müller et al., 2010). The first cladistic analysis of the red algae was performed by Gabrielson et al. (1985), which made use of morphologic traits to compare the species and determine the ancestor-descendant relationships to an ordinal rank, such as ultrastructure and life history. As a result, the red algae species were classified as part of one class, the Rhodophyceae, with two subclasses: the Bangiophycidae and the Florideophycidae. In addition, it was suggested that the Porphyridiales were polyphyletic within the Bangiophycidae due to differences in the species life cycle and ultrastructure (Gabrielson et al., 1985). Later, Bidoux and Magne (1989) proposed a classification based on characteristics of the thallus and reproductive cells, as result, Rhodophyceae was divided into three subclasses: Archaeorhodophycidae, with a single order Porphyridiales; Metarhodophycidae, including the orders Erythropeltidales, Rhodochaetales, and Compsopogonales; and Eurhodophycidae, including the order Bangiales and all the orders of the former Florideophycidae.

In 1994, Ragan et al. was the first study to sequence the nuclear small subunit ribosomal rRNA (SSU rRNA) to clarify the relationships among the red algae using molecular phylogenetics and suggested that the Bangiophycidae were paraphyletic and the basal rhodophyte line consisted of at least three distinct lineages. Following this, Oliveira and Bhattacharya (2000) compared Bangiophycidae to Florideophycidae plastid SSU rDNA coding regions; they demonstrated that the Porphyridiales formed three different lineages; they also confirmed that Bangiales and Florideophycidae are a sister-group as suggested by Bidoux and Magne (1989), and, that Cyanidiales is likely an independent order. Müller et al. (2001b) were the first researchers to study a considerable number of samples from the Bangiophycidae using the nuclear SSU rDNA gene and confirmed that the subclass Bangiophycidae is a paraphyletic group. In 2004, Saunders and Hommersand proposed four classes: the Bangiophyceae,

Composopogonophyceae, Florideophyceae, and Rhodellophyceae. Also, they recognize the subphylum Rhodophytina, comprising all classes with exception of the Cyanidiophyceae which is included on the new subphyla Cyanidiophytina. Analyzing the existent literature, it is evident the need for modern genomic research on red algae to elucidate taxonomic relationships and mechanisms of evolution, especially regarding Bangiophyceae. As a result of these studies, the taxonomy of this group has greatly changed since the work of Gabrielson et al. (1985).

The class Florideophyceae comprise morphologically complex species, being a monophyletic group, and a sister group of Bangiophyceae (Garbary and Gabrielson, 1990; Ragan et al., 1994; Saunders and Kraft 1997; Müller et al., 2001b and 2010; Saunders and Hommersand, 2004; Yoon et al., 2006). The Bangiophyceae is commonly represented by morphologically simple organisms when compared to the algal group Florideophyceae, and phylogenetic studies show that the group is ancestral of all red algae (Ragan et al. 1994; Freshwater et al., 1994; Yoon et al., 2006; Müller et al., 2001b, 2010; Lynch et al., 2008; Sutherland et al., 2011). The species within this class are in majority attached to substrata, in intertidal and upper subtidal zones in both temperate and tropical oceans (Sheath & Cole, 1984; Müller et al., 1998, 2010), comprising species with varied morphology (unicellular to multicellular filaments - uniseriate or multiseriate, branched or unbranched - or sheet-like thalli), and habitats (Garbary et al., 1980).

Currently, the most accepted classification is according to Yoon et al. (2006), who examined sequences of the Photosystem I (PSI) P700 chloroplast α protein A1 (*psa*A), RuBisCo large subunit (*rbc*L) coding regions and 18S gene. As result, the Phylum Rhodophyta includes six classes: Bangiophyceae, Florideophyceae, Rhodellophyceae, Porphyridiophyceae, Compsopogonophyceae and Stylonematophyceae (Figure 1). The class Cyanidiophyceae, after intensive investigations and changes regarding taxonomy (Merola et al., 1981; Seckbach, 1991; Ott & Seckbach, 1994), was classified as phylum by Saunders and Hommersand (2004) and is now considered to be part of the subphylum Cyanidiophytina, since it appears to have diverged first in red algal evolution and is clearly separated from the other red algal lineages (Yoon et al., 2006) (Fig. 1.1).

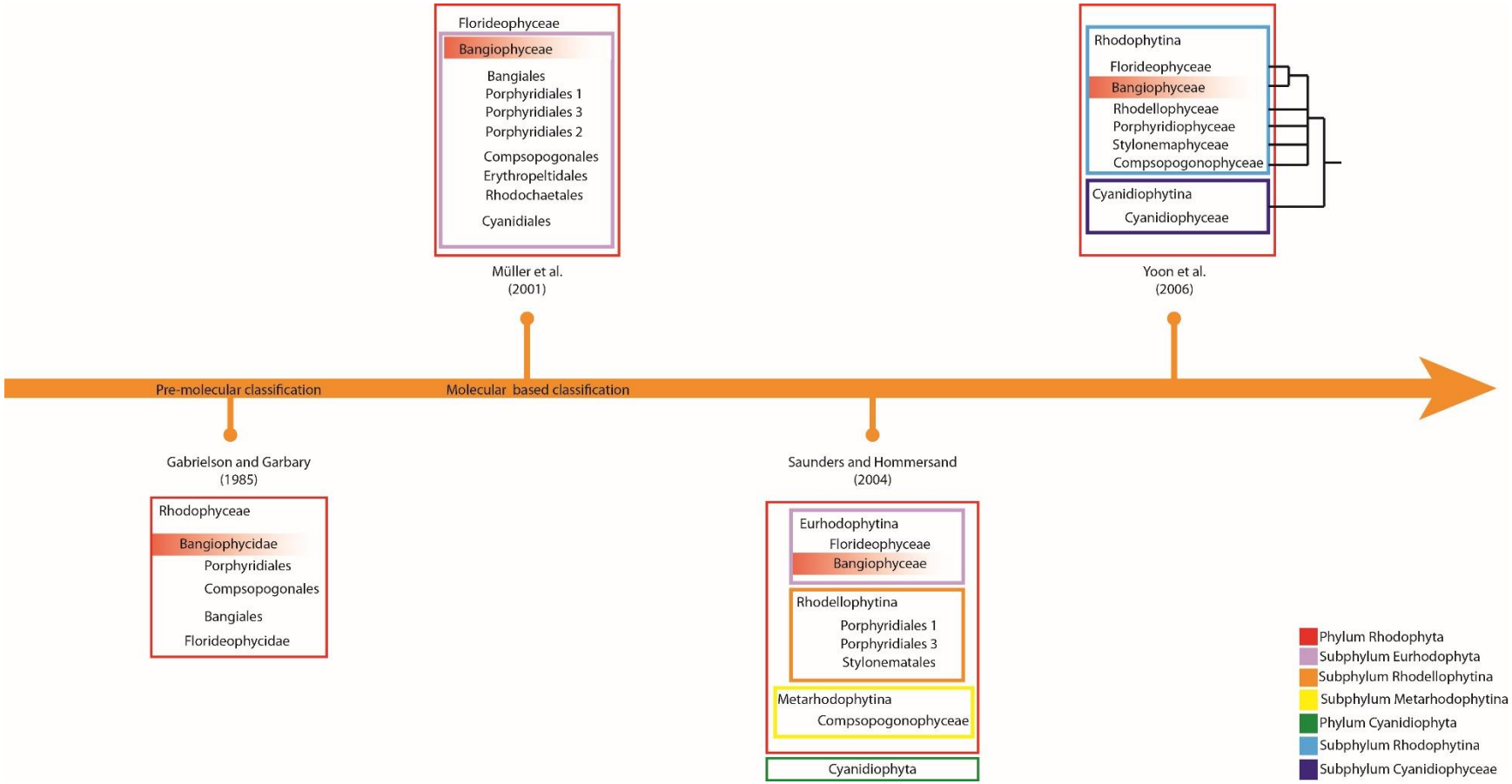

Figure 1. 1. Timeline of the taxonomic changes within the Phylum Rhodophyta. Adaptation from Yoon et al., 2006.

1.2 Red algal fossils

Fossil record data from the Proterozoic Era, around 2.5 billion years ago (Ba), are crucial to understanding the evolution of eukaryotes, especially through the analysis of various microfossils (Butterfield, 2005). Microfossils are microscopic fossils grouped according to their size, with bacteria and protists being the most numerous contributors, along with plants (Signor & Lipps, 1992). Despite the extensive distribution of organisms during the Precambrian Era (from 4.6 Ba to 544 million years ago), microfossils are sparsely distributed, limiting taxonomical and environmental findings (Butterfield, 1988). In addition, unicellular microfossils are challenging to characterize due to the lack of distinctive attributes that can be linked to an extant taxon, and are, most of the time, considered to be part of extant protist groups or cyanobacteria, even when they may have had a eukaryotic derivation (Butterfield, 2005; Yoon 2010).

Although the geological record indicates that eukaryotes evolved by 1.9–1.4 Ba their early evolution is poorly resolved taxonomically and chronologically. The 1.2 Ba microfossil of the red alga *Bangiomorpha pubescens*, a multicellular filamentous fossil, discovered in Canada, has provided a key time point in the evolution of red algae (Butterfield, 1990; Müller, 1998, 2001; Butterfield, 2000). The microfossil is the only recognized eukaryote older than 0.8 Ba and marks the earliest known expression of extant forms of multicellularity and eukaryotic photosynthesis (Butterfield, 1990; Gibson et al., 2018). The good preservation of *B. pubescens* microfossils and the pattern of cell division and morphology allowed the investigation of characteristics that lead to the identification of the organism, which can be compared to the gametophytic generation of the modern red algal genus *Bangia* (Butterfield, 1990; 2000).

Red algal microfossils from the Neoproterozoic Era (1 Ba to 540 Ma) were also discovered in China, with a simple pseudoparenchymatous thalli, implying a multiphasic life cycle; and a complex pseudoparenchymatous thalli, where *Thallophyca* and *Paramecia* show similarities to the coralline algae, part of the Florideophyceae (Xiao et al., 2004). In a more recent discovery, well-preserved fossils part of the phylum Rhodophyta were described by Bengtson et al. (2017). One of these found in Vindhyan, central India, has uniserial rows of large cells, rhomboidal disks recognized as a pyrenoid, septa between cells which may represent pit connections and pit plugs. Assembling these and other characteristics, the authors identified the

fossil to be the filamentous form of *Rafatazmia chitrakootensis,* a possible bangiophycean. In addition, the lobate sessile *Ramathallus lobatus,* which has a pseudoparenchymatous thallus, and apical growth, was identified as part of the Florideophycean group. Although these two microfossils are still under examination for possible ambiguity, the findings possibly represent the origin of multicellular rhodophytes, being 400 Ba older than *Bangiomorpha,* dating 1.6 Ba (Bengtson et al., 2017).

Examples of red algae microfossils that are unambiguous include the fossilized conchocelis stage of *Porphyra,* found in Poland, recorded as dating to 425 Ma (Campbell, 1980; Butterfield et al., 1990). On the other hand, the unicellular microfossils *Huronispora* and *Eosphaera tyleri,* found in Canada, and dating 1.9 million years old (Ma), were designated as blue-green algae (Barghoorn & Tyler, 1965); however, after analyses of reproduction, morphology, habitat and life cycles; it was suggested that these fossils were likely *Porphyridium purpurea,* a eukaryotic red alga (Tappan, 1976). Another example is the microfossil *Tappania plana,* found in Australia, dating 1.450 Ma (Javaux et al., 2001), and considered to be fungi due to the branching pattern and cell fusion, yet, both characteristics are found in red alga species (Butterfield, 2005; Porter, 2006). These are only two examples of many inaccurately classified organisms, revealing the importance of continuous studies of fossils in the field to disclose the potential identity of the samples.

1.3 Plastid evolution in the red algae and related groups

Resolving the tree of life using nuclear, mitochondrial and plastid single-gene phylogenies and phylogenomic investigations is essential to understand the evolution of photosynthetic phyla, including the red algae (Adl et al., 2005; Reyes-Pietro et al., 2007; Burki, 2014; Spang et al., 2015). This is particularly important to comprehend the diversity of photosynthetic life on this planet. Despite the challenge of understanding how and when plastids entered eukaryotic cells, endosymbiosis and serial endosymbiosis has played a major role in the diversity of species cohabiting the planet (Palmer et al, 1993; Bhattacharya & Medlin, 1995; Reyes et al., 2007; McFadden, 2014; Burki, 2014; Rockwell et al., 2014). A putative member of the Bangiophyceae played a central role in the eukaryote tree of life as a donor through a single

or more likely multiple secondary and tertiary plastid endosymbiosis that gave rise to chlorophyll-*c* containing groups, such as diatoms, dinoflagellates, haptophytes, and cryptophytes (Ragan et al.,1994; Yoon et al. 2004; Chan et al. 2011; Bhattacharya et al. 2013; Yang et al., 2017). With still many questions to investigate, understand and attempt to resolve, the phylogenomic of the chloroplast of red algae can provide insights into the early eukaryotic evolution (Adl et al., 2005; Reyes-Pietro, 2007; Burki, 2014; Spang et al., 2015).

Hypothetically, eukaryotes are considered as having originated from a universal ancestor, from the Archaea group, the *Lokiarchaeota* (Spang et al., 2015), and branched into six supergroups: Opisthokonta, Amoebozoa, Archaeplastida (Plantae), Chromalveolata, Rhizaria, and Excavata (Adl et al., 2005, 2012; Keeling et al., 2005; Reyes-Pietro et al., 2007). The study by Spang et al. (2015), proposed that *Lokiarchaeota* is the identity of the putative archaeal ancestor for eukaryotes. According to phylogenomic analyses, the group is monophyletic with eukaryotes and has proteins that are associated with membrane deformation and cell shape formation, including phagocytosis. These findings indicate that the origin of complex cells happened before the mitochondrial endosymbiosis. Nevertheless, the possible ancestor of eukaryotes had phagocyte ability, which could have enabled the invagination of the mitochondrial progenitor (Spang et al., 2015). From the supergroup Archaeplastida, or Plantae, emerged green algae and plants, the red algae and the glaucophyta algae (Adl et al., 2005; Reyes-Pietro et al., 2007; Bhattacharya et al., 2004, 2013; Burki, 2014) (Fig. 1.2).

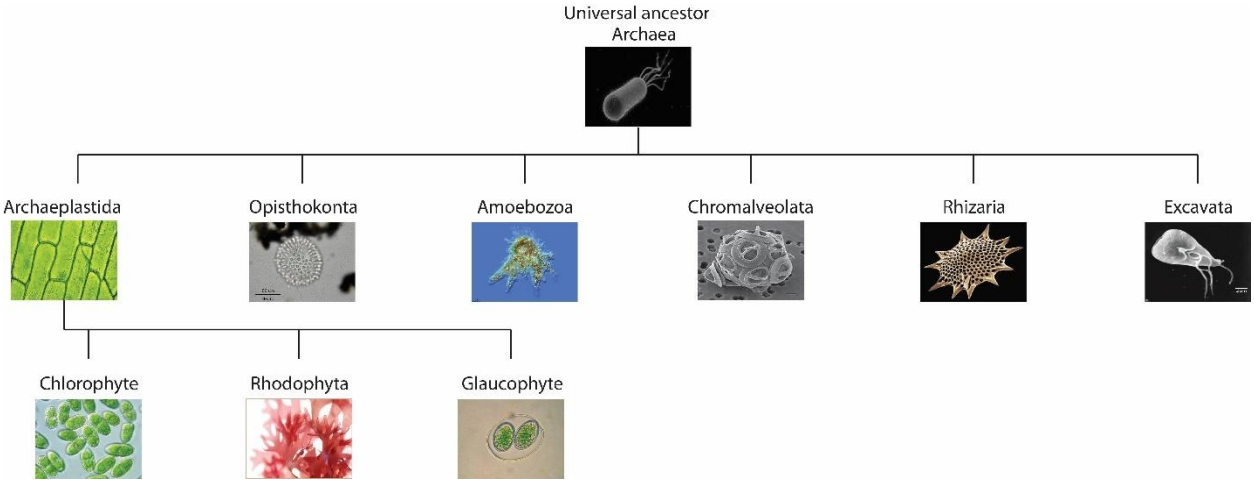

Figure 1. 2 Simplified map of chloroplast evolution from the common ancestor, an Archea organism, to the three main algae branching: Chlorophyta, Rhodophyta and Glaucophyta.

Along with the endosymbiosis that gave rise to mitochondria, the origin of plastids is one of the most important phenomena in the evolution of the eukaryotic cell, culminating in all phototrophic eukaryote lives (Archibald & Keeling, 2002). The most accepted theory about the origin of plastids through endosymbiosis is the single-cell origin of all plastids, where plastids are derived from an engulfed photosynthetic cyanobacterium by a non-photosynthetic host (protist) (Archibald and Keeling, 2002; Ponce-Toledo et al., 2017; Martin et al., 2017) (Fig. 1.3). This event happened in the common ancestor of the supergroup Archaeplastida, giving rise to the green, red and Glaucophyte algae (Gray, 1999; Adl et al., 2005, 2012; Bhattacharya & Medlin, 1995; Bhattacharya et al., 2013; Rockwell et al. 2014; Gould, 2015; Ponce-Toledo et al., 2017) (Figure 1.3). In a secondary endosymbiosis event, a eukaryote obtained a plastid by engulfing a phototrophic eukaryote with a primary plastid, obtained in the primary endosymbiosis (green, red or Glaucophyte); The digestion of the cyanobacteria did not happen, instead, it became endosymbiont (Archibald & Keeling, 2002).

The secondary endosymbiosis led to the origin of plastids in other important algal groups, such as dinoflagellates, heterokonts, cryptomonads, and parasites such as the Apicomplexa (Douglas & Penny, 1999; Müller et al., 2001b; van Dooren et al., 2001; Oliveira & Bhattacharya, 2000; Gould et al., 2015). In addition, the secondary endosymbiosis is responsible for endosymbiotic gene transfer (EGT), in which red algal plastid genes are transferred to the nucleus of other photosynthetic organisms (Douglas et al., 2001; van Dooren et al., 2001; Bhattacharya et al., 2013; Gould, 2015). With time, the endosymbiotic organism was reduced, by gene transfer to the nucleus, to a double membrane-bound plastid, and vertically transmitted to following generations (Archibald & Keeling, 2002; Bhattacharya et al., 2004; Reyes-Pietro et al., 2007; Gould, 2015). (Figure 1.3). To understand the magnitude of these events for the eukaryotic life, it is fundamental to discover which algal lineages arose from the same endosymbiotic association and which arose independently (Delwiche & Palmer, 1997; Cavalier-Smith 1999; Oliveira & Bhattacharya, 2000; Archibald & Keeling, 2002). Since it is definitive that red algae are an early branching within eukaryotes, and, therefore, play a very important role in algal evolution, they are excellent models for new investigations regarding plastid evolution (Gray, 1999; Bhattacharya & Medlin, 1995; Palmer et al., 2003; Bhattacharya et al., 2004; Reyes-Pietro et al., 2007).

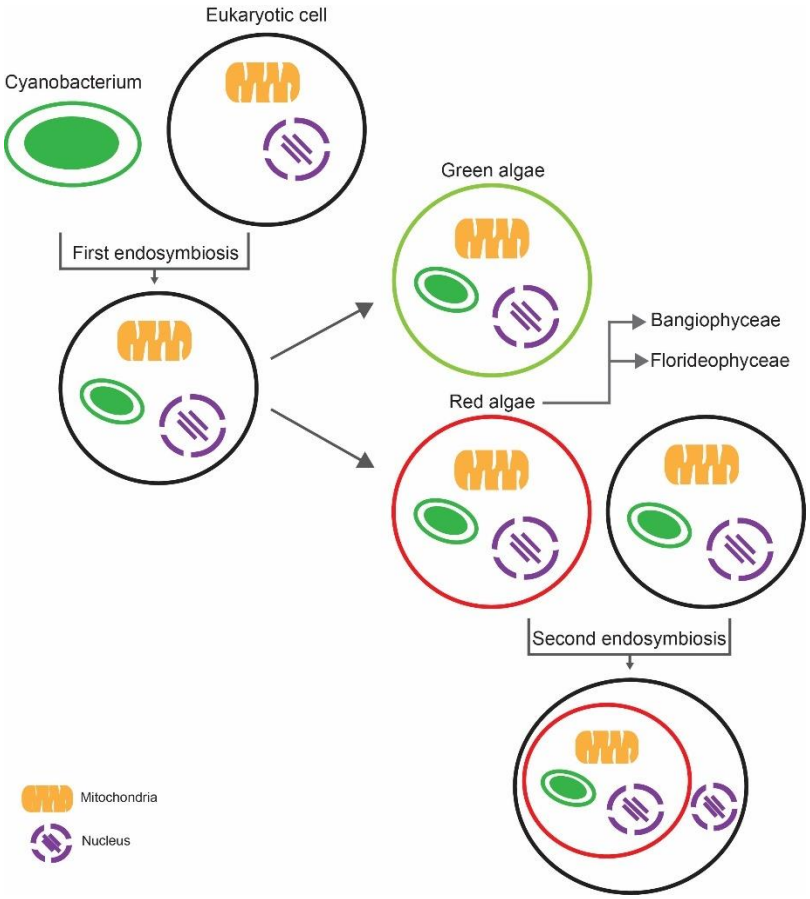

Figure 1.3 Diagram depicting the evolution of plastids deriving from primary and secondary endosymbiosis. The first endosymbiosis event involved the incorporation of a cyanobacterium into a non-photosynthetic eukaryote cell. After this single primary endosymbiotic event, the now phototrophic eukaryote diverged and evolved to be red, green and Glaucophyta (not shown in the figure) algal lines. In the secondary endosymbiosis event, a phototrophic eukaryote was incorporated into a non-photosynthetic eukaryote. Adapted from Douglas et al. (2001) and van Dooren et al. (2001). This diagram does depict tertiary endosymbiosis.

1.4 Class Bangiophyceae

The class Bangiophyceae is one of the seven classes within the phylum Rhodophyta (Yoon et al. 2006; Sutherland et al., 2011) (Fig. 1.1). The phenotypic plasticity of the Bangiales species has led many researchers to conduct studies in the interest of resolving the classification of the class. The thallus morphology can change significantly during different stages of reproduction, characteristic that resulted in misleading taxonomy in pre-molecular studies (Sutherland et al., 2011). The life history of species within the Bangiophyceae is diverse and

biphasic, which includes a macroscopic gametophyte and a microscopic conchocelis phase (Sheath & Cole, 1984; Garbary & Gabrielson, 1990; Müller et al., 2005 and 2010; Yoon et al., 2010). (Fig. 1.4). Sexual reproduction is often visualized by marine members of the Bangiophyceae and is uncommon in other classes (Müller et al., 2010). One example of a Bangiophyceae member that presents both reproductive pathways is within the genus *Porphyra*, which exhibits a heteromorphic life history (Drew, 1949; Blouin et al., 2011). From the sheet-like haploid gametothallus, fertilization happens between spermatium and carpogonium, generating a carposporangium (diploid) which will rise to a microscopic shell-boring conchocelis phase. In some species, the spores developed during the conchocelis phase can undergo meiosis, forming haploid spores (Mitman & Meer, 1994; Brodie & Irvine, 2003). Sexual reproduction, however, has not yet been described for some members of this class, and follow an asexual route where the gametothallus releases a spore (haploid) and the spore will grow into a new gametothallus (Garbary et al., 1980; Gabrielson et al., 1990, Müller et al., 1998, 2005).

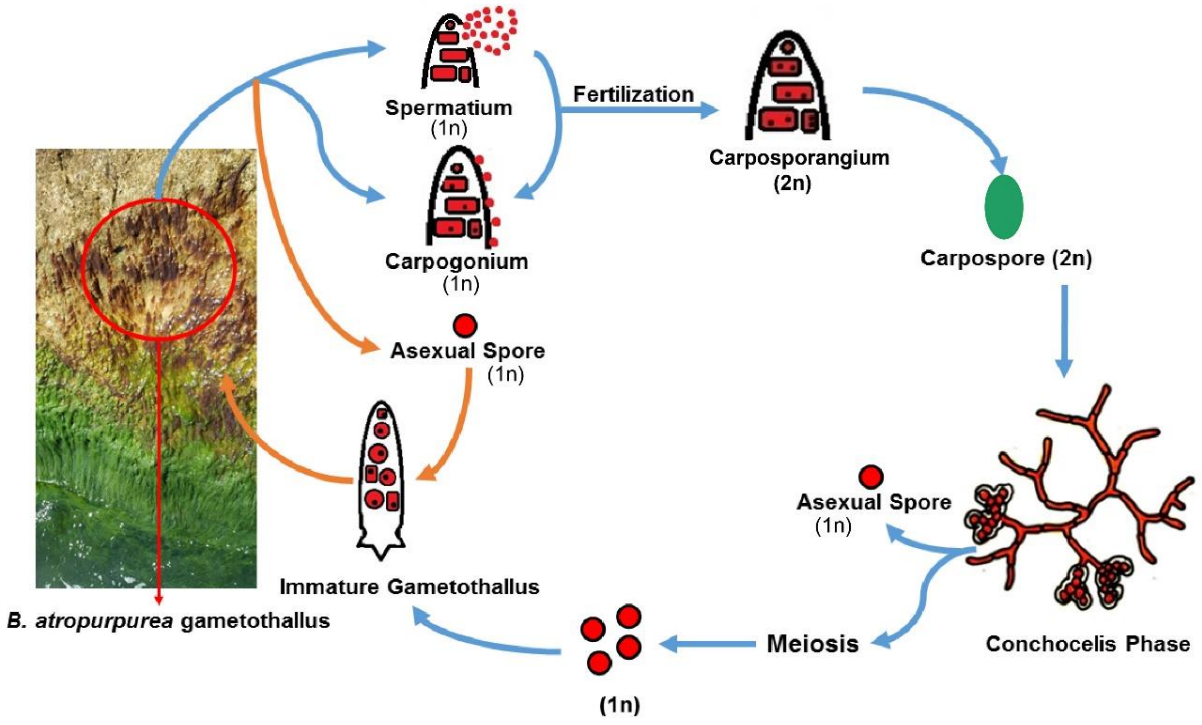

Figure 1. 4 Possible reproduction routes for Bangiophyceae members. Asexual reproduction – orange arrows - is only reported for species of *B. atropurpurea*. Species of the genus *Porphyra*, however, have been reported to exhibit both asexual and sexual reproduction.

1.5 Order Bangiales

Before molecular phylogenetic analyses, the class Bangiophyceae was divided into four distinctive orders: Bangiales, Porphyridiales, Compsogonales and Rhodochaetales (Gabrielson et al., 1990; Garbary & Gabrielson, 1990, Ragan et al. 1994). Currently, the class is comprised of only the order Bangiales (Müller et al., 2001b; 2010; Sutherland et al., 2011). The order Bangiales is also recognized as the starting point of divergence for the other algal orders, comprising morphologically simple red algae, including also the florideophytes, (Ragan et al., 1994; Saunders & Kraft, 1997; Butterfield et al., 2000). Currently, the order comprises 15 genera, including the genera *Porphyra* and *Pyropia*, which are the most highly valued seaweed crops in the world (Yoon et al., 2006; Müller et al., 2005 and 2010; Sutherland et al., 2011; Guiry & Guiry, 2017), and the ancient genera *Bangiomorpha* and *Bangia* (Butterfield, 2000). However, the taxonomy of the Bangiales is not yet fully understood with genera still missing molecular information and phylogenetic analysis.

The most in-depth study of the order and its genera to date was conducted by Sutherland et al. (2011). The authors made use of a vast number of samples from collections distributed worldwide and performed an analysis of both the nuclear small subunit rRNA (SSU rRNA) gene and the plastid *rbc*L gene. This study has extensively modified the number of genera in the order, with several well-supported clades being resolved within the order, and 6 of the15 genera within the order being monotypic. This information could clarify the selection of species, breeding and cultivation to the aquaculture industry, who were possibly comparing members of different genera until then. Although four of the seven filamentous genera have established names and taxonomic position (*Bangia, Dione*, *Minerva* and *Pseudobangia*), new studies are necessary to clarify the filamentous members of the Bangiales that are still not investigated (Müller et al., 2005; Nelson et al., 2006; Sutherland et al., 2011).

1.5.1 Genus *Bangia*

The genus *Bangia* was first described in 1819 by Lyngbye and is commonly observed in freshwater within the Laurentian Great Lakes and in marine coasts of North America (Sheath &

Cole, 1984; Müller et al., 2003). In an algal vegetation study in a lake in the Netherlands, den Hartog (1972) concluded that *Bangia* species in the Ysselmeer may be derived from the marine species. This prediction led to studies that analyzed the gradual adaptation of *B. atropurpurea* in saltwater, and *B. fuscopurpurea* in freshwater (Geesink, 1973). The study demonstrated the ability of acclimation of both species, concluding that they are conspecific. The acclimation was confirmed by Sheath and Cole (1980), by experiments that analyzed the growth and photosynthetic rates of samples of *B. atropurpurea* from the Laurentian Great Lakes, corresponding with Geesink's proposal. A study conducted by Sheath and Cole (1984) aimed to identify different species of *Bangia* using morphological characteristics such as filament diameter, length and pigmentation. The authors noted that there was a north-south trend in increasing filament diameter which suggests the presence of more than one species in the Pacific; moreover, the Atlantic populations could represent a single species, *B. atropurpurea* (Roth) C. Agardh; lastly, the freshwater populations were a variety of *B. atropurpurea*.

In 1998, Müller et al. made use of the SSU rRNA and *rbc*L gene sequences to establish the difference between the two species, which correspond to the study by Seath and Cole (1984), where the freshwater collections of *Bangia* were distinct from marine collections. Moreover, Muller et al. (1998) noted that the sequence divergence among inland samples from North America and Europe is very low, suggesting that the freshwater *Bangia* may be a separate lineage. The authors also observed that the marine *Bangia* populations were heterogeneous. Later, Müller et al. (2003) performed a karyological study with *Bangia* species collected in the freshwater and marine environment. Based on the DNA sequence and phylogenetic analyses, the freshwater collections of *Bangia* were positioned on a separate and well-supported branch, thus the name *B. atropurpurea* should represent this lineage. All marine collections of *Bangia* did not cluster together in gene trees, indicating a mix of different species, and until further identification, should be recognized as *B. fuscopurpurea* (Müller et al., 2003).

1.5.2. Distribution, invasion patterns and adaptation of *Bangia atropurpurea*

The first record of freshwater populations of *Bangia* was described by Roth (1806) as *Conferva atropurpurea*. In 1824, C. Agardh included the species in the genus *Bangia* as *B.*

atropurpurea. Later, due to the ability to acclimate from marine to freshwater environments, species of *Bangia* were conspecific (Geesink 1973; Sheath & Cole, 1980). More recently, molecular data contributed to a better understanding of the species, since freshwaters lineages cluster in a different group of marine species (Müller et al., 1998 and 2003). In Canada, *B. atropurpurea* was observed for the first time in Lake Erie, part of the Laurentian Great Lakes, in 1964 (Lin & Blum, 1977). Damann (1979) stated that *Bangia* expanded through Lake Ontario to Hamlin Beach Park. Jackson (1985) and Sheath (1987) noted that *Bangia* had spread throughout all of the Great Lakes, except for Lake Superior. In the Great Lakes, *Bangia* can be observed 5-15 cm above species of *Cladophora glomerata* (L.) Kütz, or above the waterline (Sheath & Cole, 1980, 1984). These studies acknowledge that the *Bangia* population from the Great Lakes is likely derived from freshwater in Europe (Müller et al., 1998; Shea et al., 2014). The sudden appearance of *B. atropurpurea* in Lake Erie and subsequent spread to the other lakes, except for Lake Superior, demonstrates how effective the adaptation mechanisms and invasion patterns can be, being able to spread quickly and permanently throughout the Great Lakes.

Considering the fast growth and high reproductive yield of this species (Sheath, 1987; Sheath & Cole, 1980; Sheath et al., 1985), *B. atropurpurea* is treated as opportunistic (Müller et al., 2003). Müller et al. (1998) concluded that the unexpected appearance of *B. atropurpurea* in the Laurentian Great Lakes is probably due to vector-assisted transport, indicating a single invasion of the Great Lakes (Sheath, 1984). This hypothesis is supported by the disjunctive distribution of *Bangia* in the Great Lakes and molecular data (Müller et al., 1998). The ability to acclimate to both fresh and marine water provides *Bangia* with even more uniqueness since species tend to be specific to one habitat and only survive when all the optimal conditions are available. The use of seaweed models to test salinity acclimation is extensive. Several authors are intrigued by the ability to survive in such contrasting environments, with emphasis on intertidal species, such as *Bangia*. The successful salinity acclimation of *Bangia* (Geesink, 1973; Sheath and Cole, 1980) could have contributed to the establishment of the species in different habitats, however, there are no studies to clarify and understand how the acclimation takes place on a molecular basis. The fundamental mechanism of salinity tolerance is osmotic acclimation, which will preserve intracellular homeostasis (Kirst, 1990). Many red algal seaweeds have a two-step process when exposed to salinity changes: first, there is a rapid change in turgor due to the water

flux in or out of the cells, and secondly, there is an osmotic acclimation with the regulation of the intracellular concentrations (Karsten et al., 2012).

In addition to the acclimation to different salinities, species of *B. atropurpurea* can survive in low temperatures (-22°C) and have great tolerance to many environmental factors, a possible consequence of their high metabolic flexibility (Bischoff-Bäsmann & Wiencke 1996; Karsten & West 2000). However, in comparison to green algal studies, there is little information regarding the ecophysiology of the *B. atropurpurea*, and more study is needed to fully understand the survival mechanisms of the species (Kirst, 1990; Karsten et al., 1991; Wiencke 2007). In addition, species of *B. atropurpurea* is observed in the upper intertidal zone, being out of water for long periods, therefore experiencing a broad variation in salinity and temperature, which leads to desiccation and exposure to different ultraviolet ranges (Karsten & West, 2000). Russel (1985) explained that since intertidal species regularly experience wide fluctuations in salinity, they can be preadapted to higher salinity levels because of their genetic plasticity. Moreover, Pearson et al. (2000), noted that species in intertidal zones may have special mechanisms to tolerate simultaneously freezing, desiccation, and precipitation stresses. Taking into consideration the well-established knowledge about the adaptation mechanisms which green and brown algae express when facing abiotic stresses, it is important to build the same understanding in species of red algae. In addition, make use of modern molecular techniques to compare gene regulation and further our understanding of how *B. atropurpurea* can acclimated and adapt to such harsh environmental conditions and compare these mechanisms with other seaweeds.

1.6 Molecular studies of Red Alga

With the advance of technology and techniques that evolved over the 20th century, molecular studies became essential in phycological investigations since they provide insights into characteristics never studied before, from morphology differentiation to timing of the origin of species. Following the course of many animal molecular studies, algal systematics also relies on molecular tools to resolve and identify species (Saunders et al., 2005). Most genetic information is contained in nuclear DNA, however, plastids and mitochondria also contain DNA,

and each has special characteristics, and provides unique insights into different aspects of algal evolution and biology. For the purpose of this study, I will be focusing on the nuclear and chloroplast genomes.

Nuclear genes

The nuclear DNA of red algae will vary between one or more copies of genes due to gene duplication, in the copies, numerous tandem repeats of the same sequence can be observed; in the same region is the sequence that codes for the ribosomal RNA (rRNA) (Coleman & Goff, 1991). The rRNA consist of a small subunit (SSU, or 18S rRNA gene) and a large subunit (LSU or 28S rRNA gene), the gene 5.8S rRNA, which has internal transcribed spacers (ITS1 and ITS2) and intergenic spacers (IGS), that is composed by a non-transcribed spacer and an external transcribed spacer (ETS), which separate continuous transcription units such as the SSU, 5.8S rDNA, and LSU (Coleman & Goff 1991; Harper & Saunders, 2001; Xu et al., 2016). Within Bangiales, the ITS-1 and ITS-2 have served as genetic markers for population genetic analysis of *Porphyra* (Kunimoto et al., 1999) and species identification (Broom et al. 2002). In addition, a group I introns have also been identified in the SSU rDNA of collections of *Bangia* and *Porphyra* (Stiller & Waaland 1993; Oliveira & Ragan, 1994; Müller et al. 1998, 2001a; Kunimoto et al. 1999), in the helix 51 at the 3' end of the gene (Klein et al., 2003). These groups of introns have also been used as molecular markers for genus recognition (Oliveira & Ragan 1994; Kunimoto et al. 1999), for phylogenetic analysis (Müller et al. 2001) and species differentiation (Broom et al. 2002). However, the lengths of these introns can vary significantly within species; in Bangiales for example, the sequence and structural elements of the introns differ from each other and other intron sequences available (Müller et al., 2001), therefore, the phylogenetic properties and effectiveness as a marker are unclear (Müller et at. 1998, 2001).

Studies have confirmed the relevance of the SSU rDNA gene in phylogenetic studies, since it is present in all eukaryotes and has a conserved structure and function, evolving at a stable rate and, therefore, can be used as a molecular clock to obtain divergence times between species (Van de Peer, 1993; Oliveira et al., 1995; Müller et al., 1998 and 2001; Broom et al., 1999; Milstein & Oliveira, 2005; Yu et al., 2010; Xu et al., 2017). Among Bangiales, the SSU rDNA was sequenced for *Porphyra* phylogeny (Oliveira, 1993; Oliveira et al., 1995; Broom et al. 1999, Milstein & Oliveira, 2005), for the distinction among *Porphyra* species (Broom et al.

1999; Kunimoto et al. 1999; Nelson et al. 2001; Klein et al. 2003), and biogeographic and systematic of *Bangia* (Müller et al., 1998 and 2003). The SSU rDNA of red algae has unexpected high levels of sequence variation (15%) when compared with higher plants and green algae (Bird et al., 1992). The gene regions that are used for phylogenetic studies are the S' end, from helix 6 to 19, and the middle of the gene, from helix 19-27 (Bird et al., 1992; Klein et al., 2003), however, has relevant variability between species.

Chloroplast genes

Chloroplasts are organelles present in plants that are responsible for the process of photosynthesis, participating in the biosynthesis of amino acids, nucleotides, lipids and starch (Sugiura, 1992; Alberts et al., 1994). As mentioned previously in section 1.3, red algal plastids have characteristics in common with cyanobacteria, such as the presence of phycobiliproteins and the thylakoid arrangement (Kostrzewa et al., 1990; Müller et al., 2001). The application of molecular markers has provided insights into the taxonomy of Rhodophyta, and several chloroplast genes are currently used as marker genes for phylogenetic reconstruction (Lim et al., 2017). The Rubisco enzyme for example, which is responsible for the CO_2 fixation step in photosynthesis, has a small (*rbc*S) and a large subunit (*rbc*L), and both are encoded by plastid genes (Kostrzewa et al., 1990; Freshwater et al, 1994). The *rbc*L gene is the most studied chloroplast marker gene and has been demonstrated to be reliable when used to assess relationships between red algal taxa at species, familial and ordinal levels (Hommersand et al., 1994; Freshwater et al., 1994). On the other hand, studies indicate that the gene is not as useful for species differentiation when the species have low levels of divergence (Freshwater et al., 1994). The *rbc*L gene and intergenic spacers were also studied to discriminate *Porphyra* and *Pyropia* species (Brodie et al. 1996, 1998; Neefus et al., 2000; Teasdale et al., 2002; Klein et al., 2003; Milstein & Oliveira, 2005; Choi et al., 2019; Meynard et al., 2019) as well as for phylogeny, biogeography and systematic analysis of *Bangia* (Müller et al. 1998, 2003) and *Porphyra* (Teasdale et al. 2002; Klein et al. 2003; Lindstrom & Fredericq 2003; Milstein & Oliveira, 2005).

1.7 Genomic studies in the Red Alga

With the advance of molecular tools, genomic studies of red algae involving nuclear, mitochondrial and plastid complete sequences can be accomplished faster and with more details than sequencing just part of the organism's genetic material. Genome sequencing provides insights into adaptability, stress resistance, invasion mechanisms and metabolism. With still many questions to investigate regarding red algae evolution and their unique biological traits, continuous research is extremely important to understand the relationships among species molecular changes that can result in adaptability in different habitats (Adl et al., 2005; Reyes-Pietro, 2007; Burki, 2014; Archibald, 2015; Spang et al., 2015, Hu et al., 2016). Completed genomes are available for a growing number of algal species, offering a new understanding regarding comparative evolution, molecular biology, biochemistry, physiology, and developmental biology (Waaland et al., 2004).

Because of the small size, prokaryotic genomes have been more extensively sequenced and compared, resulting in more than forty thousand bacteria genomes, and more than 2000 Archaea (https://www.ncbi.nlm.nih.gov/genome/?term=Bacteria and https://www.ncbi.nlm.nih.gov/genome/?term=Archaea). In the last few decades, although genome sequencing studies have become more popular, complete algal genomes are still limited and are largely restricted to unicellular organisms, such as the green microalgae *Chlamydomonas* (Shrager et al., 2003) and the unicellular red algae *Cyanidioschyzon merolae* (Matsuzaki et al., 2004), *Galdieria sulphuraria* (Schönknecht et al., 2013) and *Porphyridium purpureum* (Bhattacharya et al., 2013). Partial sequences (nuclear, plastid or mitochondrial DNAs) of algal samples are much more abundant than the list of complete genome sequences. Since whole genome sequencing is still a relatively recent technology, the total number of genomics studies of macroalgae is still insufficient to answer fundamental questions regarding these important organisms. To date, only *Ectocarpus siliculosus* (Cock et al., 2010), *Chondrus crispus* (Collén et al., 2013), *Pyropia yezoensis* (Nakamura et al., 2013) and *Porphyra umbilicalis* (Brawley et al., 2017) have been completely sequenced (Table 1.1).

Table 1. 1 List of algae species with complete genome available at NCBI*

Subdivision	Species	Habitat	Genome lengh (Mb)	Morphology	Assembly Number	Reference
Rhodophyta	*Chondrus crispus*	Marine Intertidal	104.981	Macroalga bush talli	GCA_000350225.2	Collén et al., 2013
Rhodophyta	*Cyanidioschyzon merolae*	Hot springs	16.547	Unicellular	ASM9120v1	Matsuzaki et al., 2004
Rhodophyta	*Galdieria sulphuraria*	Hot springs	13.712	Unicellular	ASM34128v1	Barbier et al., 2005
Rhodophyta	*Porphyra umbilicalis*	Marine Intertidal	87.767	Macroalga sheet thalli	GCA_002049455.2	Brawley et al.,2017
Rhodophyta	*Porphyridium purpureum*	Marine and soil	22.193	Unicellular	GCA_008690995.1	Bhattacharya, et al., 2013
Chlorophyta	*Auxenochlorella protothecoides*	Soil	22.925	Unicellular	ASM73321v1	Darienko & Pröschold, 2015
Chlorophyta	*Bathycoccus prasinos*	Marine	15.075	Non-motile picoplanktonic	GCF_002220235.1	Moreau et al., 2012
Chlorophyta	*Chlorella variabilis*	Freshwater and soil	46.160	Unicellular	GCF_000147415.1	Junej et al., 2016
Chlorophyta	*Coccomyxa subellipsoidea*	Marine	48.827	Unicellular	GCF_000258705.1	Blanc et al., 2012
Chlorophyta	*Micromonas commoda*	Marine	21.110	Unicellular	GCF_000090985.2	Worden et al., 2009
Chlorophyta	*Micromonas pusilla*	Marine	21.959	Picoplanktonic	GCF_000151265.2	Cottrell & Suttle, 1991
Chlorophyta	*Ostreococcus lucimarinus*	Marine	13.205	Unicellular	GCF_000092065.1	Palenik et al., 2007
Chlorophyta	*Ostreococcus tauri*	Marine	13.668	Unicellular	GCA_000214015.2	Derelle et al., 2006
Chlorophyta	*Volvox carteri f. nagariensis*	Freshwater	137.685	Spherical multicellular	GCF_000143455.1	Smith & Lee, 2010
Phaeophyta	*Ectocarpus siliculosus*	Rocky shorelines	195.812	Macroalga filamentous	GCA_000310025.1	Delaroque et al., 2001

*Genomes available to date.

Chloroplast genomes

Red algal chloroplast genomes are considered ancestral and evolutionarily stable, due to the gene content and highly compact genomic organization (Kim et al., 2009; Lang & Nedelcu, 2012; Janouškovec et al., 2013), carrying characteristics that resemble the common ancestral plastid genome of all photosynthetic eukaryotes (Lang & Nedelcu, 2012; Janouškovec et al., 2013; Muñoz-Gómez et al., 2017). According to Lee et al. (2017), Rhodophyta chloroplast genomes contain the largest collection of plastid genes ever described, including genes specific for red algae, such as regulatory proteins, membrane transporters and biosynthetic enzymes (Kim et al., 2009; Muñoz- Gómez et al., 2017). The map of a complete plastid circular chromosome was first characterized in 1979 (Bedbrook & Kolodner), since then the organization and expression of chloroplast genomes have been extensively studied in plant molecular biology, especially in higher plants and green algae (Sugiura, 1992). The chloroplast genomes have distinguished inverted repeats (IR), separated by one large and one small single-copy region, and can be organized into three groups: chloroplast genomes without IRs, chloroplast genomes containing IRs, and chloroplast genomes with tandem repeats (Sugiura, 1992). Since red algae are the earliest divergence point of eukaryotes, phylogenetic studies with chloroplast genomes are very common to understand the evolution, relationships and changes that have occurred throughout history.

As previously mentioned, after SSU RNA and *rbc*L gene sequencing analyses, Bangiales was restructured, to include seven filamentous and eight foliose genera (Sutherland et al., 2011).

One unexpected finding within the analyses was the reassignment of species to the genera *Wildemania* de Toni, and to confirm, Hughey (2016) did a complete plastid genome study with the algae. As result, the chloroplast genomes display similar organization and content to other Bangiaceae, however, the commonly present 16S and 23S rRNA gene repeats are lacking. Moreover, a mitogenome of the *Wildemania schizophylla* was made (Silva & Hughey, 2016), resulting in a smaller genome due to the lack of large intronic Open Reading Frames (ORFs), that are present in other species of Bangiales. Due to these differences, a phylogenetic analysis strongly supports the resurgence of the genera *Wildemania*, positioned between *Porphyra* C. Agardh and *Pyropia* J. Agardh; in addition, *Porphyra* and *Wildemania* species are viewed as distinct (Hughey, 2016; Silva & Hughey, 2016). These unpredictable observations about *Wildemania* species emphasize the importance of continuous analyses within species of the family Bangiales.

More recent investigations have performed complete chloroplast genomes sequencing of the red alga *Pyropia endiviifolia,* revealing a compact genome organization in comparison to other *Pyropia* species, probably due to the different copy number of rRNA operons in the chloroplast genomes (Xu et al., 2017). Moreover, the complete genome of *Porphyra umbilicalis* was published (Brawley et al., 2017), facilitating comparison with the genomes of other red algae (Table 1). The 87.7-Mbp haploid *Porphyra* genome was sequenced using a combination of PacBio and Illumina whole-genome shotgun sequencing (Sturm et al., 2013; Brawley et al., 2017). The authors agree that genomic reduction has happened in the red algal ancestor since some pathways and genes from *Porphyra* are absent, such as flagellar and autophagy proteins. Since the majority of bangiophytes survive severe intertidal conditions, one of the most important discoveries was the 11 genes that are essential for photoacclimation and cell viability when the plant undergoes stress.

Moreover, *Porphyra* exhibits different pathways to protect themselves from the intense UV radiation, being one of them the biosynthesis of mycosporine-like amino acids, which act as photo protectors in living organisms (Gröninger et al., 1999; Conde et al., 2000; Brawley et al., 2017). Analyzing samples, Brawley et al. (2017) concluded that *Porphyra* has genes that encode the MAAs catalysts proteins MysA and MysB, which are also found in dinoflagellates. The existence of these proteins in an ancestral lineage, such as *Porphyra*, agrees that dinoflagellates possibly acquired these genes from red algae through secondary endosymbiosis (Waller et al.,

2006; Brawley et al., 2017). Due to the similarities in lineage evolution, morphological differences and many other questions that remain between species of *Porphyra* and *Bangia*, and the possibility of discoveries to understand the relationship between them, a genomic study of *Bangia* is necessary.

Although the origin and evolution of the red algae have been studied for decades, the relationship among species and the molecular pathways that red algae possess, which might have facilitated their survival throughout time and in different habitats, are examples of unanswered questions which potentially can be explained by new molecular studies. Studies that had focused on the pathways of evolution and adaptation to survival, such as Keeling and Palmer (2008), conclude that all photosynthetic eukaryotes had an extensive foreign gene transfer, with emphasis on the plastid donor through endosymbiotic gene transfer (EGT). Moreover, the genes of these organisms also underwent a horizontal gene transfer (HGT) when receiving genes from non-cyanobacterial prokaryotes (Qiu et al., 2013). When analyzing the genome of *Porphyridium purpureum*, the authors were able to identify that ancient red algae likely had mediated transfers of ~300 prokaryotic genes into chromalveolates, recognizing red algae as a relevant source of a genetic novelty among photosynthetic eukaryotes. Furthermore, Qiu et al. (2013 and 2015) suggest that Rhodophyta have an important role in eukaryote genome evolution due to the ability to assemble and to be the source of foreign genes through HGT and endosymbiosis. These observations have led to new investigations regarding the gene reduction in red algae (Qiu et al., 2015).

It is hypothesized that the red algal ancestor likely had an intron-rich genome, and was able to invade regions with extreme conditions, such as high temperature and acidic environments; and following this invasion, the genome became greatly reduced (Collén, 2013; Qiu et al., 2013, 2015; Csuros et al., 2011). This genome reduction likely happened during the first stages of red algae evolution, because environmental pressures (high temperature and low pH) induced a genome reduction, with loss of 25% of genes and flagella (Yoon et al., 2004; Csuros et al., 2011; Collén et al., 2013; Qiu et al., 2013 and 2015). Observations made by Collén et al. (2013) when analyzing the *Chondrus crispus* genome indicate the presence of the gene mannosylglycerate synthase (MGS), which is very similar to the gene found in marine bacteria and archaea. The gene encodes an enzyme responsible for thermal adaptation in thermophilic microorganisms is found in other red algae, but absent in green, glaucophytes and land plants

(Martins et al., 1999; Collén et al., 2013). This unusual metabolic feature in addition to the evolutionary characteristics of *C. crispus*, such as the compact structure, is evidence that the red alga experienced an evolutionary bottleneck (Collén et al., 2013). Only then did red algae conquer freshwaters and marine environments, having their genome expanded by the activity of transposable elements and evolving to what we know as Bangiophyceae and Florideophyceae (Collén, 2013).

1.8 Rates of substitution

Natural genetic variation among species is the raw material for evolutionary processes (Gillespie, 1994; Thompson, 1999; Whitehead & Crawford, 2006; Bódi et al., 2017). These variations, or mutations, can be caused by genetic or environmental variations, or a mix of the two. They are usually linked with the adaptation of a trait or characteristic that increases the fitness of an organism, as a result of natural selection. Early studies suggested that variation in regulatory non-coding DNA, rather than in protein-coding genes, was likely an important source of adaptive variation (King & Wilson 1975). Following, studies focused on determining the relative roles of nucleotide mutation, drift, and natural selection affecting protein variation (Gillespie, 1994). The number of mutations is expected to be minimal between genetically identical organisms and to increase among more distantly related species (Whitehead & Crawford, 2006). Tests for selection pressures are applied to better understand how species evolve and adapt to their environment. For these tests, the accepted theory is that parts of the genome that are responsible for adaptive phenotypic changes evolve faster than other parts (Nowick et al., 2019). Therefore, the chloroplast genes can help elucidate questions about species evolution and adaptation, since changes or mutations affecting chloroplast gene expression can quickly alter the plant phenotype.

The observations and calculations on sequence mutations began in the 60s, with the rise of the field of molecular evolution (Zuckerkandl & Pauling,1965; Kimura,1968). These studies suggested that there are three types of mutation/selection acting on sequences: neutral, positive, and purifying. Neutral selection can be explained according to the Neutral Theory of Molecular Evolution (Kimura, 1968 and 1983; King & Jukes, 1969; Ohta, 1973 and 1992; Yang et al.,

2000), which states that most evolutionary changes at the molecular level are caused by random genetic drift of selectively neutral or nearly neutral mutations rather than by natural selection. In the absence of selection, the rates of nucleotide substitutions that change the corresponding amino acid should be equal, not affecting the fitness of the organism (Kimura, 1968; Ohta, 1973; Bush, 2001). The neutral theory has been widely accepted and is the guiding principle for studying evolutionary changes, where it is usually used as a null hypothesis (Nei et al., 2010). The use of statistical methods to detect deviations from this neutral model is, therefore, able to detect other phenomena including positive and purifying selection (Yang et al., 2000; Nei et al., 2010).

Positive selection perturbs patterns of genetic variation relative to what is expected under the standard neutral theory (Biswas & Akey, 2006). The latest review on statistical methods for selection detection states that positive mutations are often lost to drift while still in low frequency, but otherwise tend to rise to fixation by natural selection, increasing the fitness in a population (Hejase et al., 2020). Identifying signatures of positive selection is important since they indicate regions of the genome that are functionally important (Biswas & Akey, 2006). Positive selection is a crucial source of evolutionary innovation, and after many decades of debates, it is now accepted that both neutral drift and positive selection play major roles in evolutionary change (Kosiol, et al., 2008). Lastly, purifying or deleterious mutations, which are present among all populations, are responsible for genomic sequence conservation across long evolutionary timescales (Cvijović et al., 2018; Brunet et al., 2021). They tend to be eliminated by natural selection, having fewer contributions to molecular evolution, and there is evidence for reducing genetic diversity at sites under direct selection and at linked neutral sites (Comeron, 2014; Elyashiv et al., 2016; Cvijović et al., 2018). It is important to note, however, that purifying mutations appear less likely to lead to strong effects on diversity in natural populations having a small absolute effect on fitness but can be substantial if a large portion of individuals in the population acquire the mutations in every generation (Cvijović et al., 2018). Despite the large number of DNA sequences available, positive, and purifying selection has been understudied and is poorly understood (Yan & Bielawski, 2000; Cvijović et al., 2018). To comprehend genomic evolution, it is important to study the evolution of protein-coding genes and gene-regulatory elements in the context of their structure and molecular interactions (Adams et al., 2017), as well as their developmental and physiological processes (Nei et al., 2010). Nevertheless, much

remains to be learned about the types of selection acting on genes and genomes, even within protein-coding regions.

1.9 The Eukaryotic Meiosis Toolbox

The origins of meiosis in early eukaryotic history remain elusive and it is considered one of the most difficult evolutionary problems yet to be resolved (Smith & Maynard-Smith, 1978; Hamilton & Hamilton,1999; Wilkins & Holliday, 2009). However, the study of the evolution of meiosis based on cytology and genetics shows evidence for meiosis to have evolved from mitosis (Simchen & Hugerat, 1993; Cavalier-Smith, 2002). The similarity between the events in mitosis and meiosis (Table 1.2) argue for a close evolutionary relationship between them, and the greater complexity of meiosis indicates that it is the derived process (Wilkins & Holliday, 2009). Comparative evidence suggests that meiosis arose early in eukaryotic cell history (Ramesh et al., 2005; Schurko & Logsdon, 2008), and its high degree of similarity in different taxonomic groups suggests that it happened once in the eukaryotic history (Hamilton & Hamilton,1999; Ramesh et al., 2005). Four main events differentiate mitosis and meiosis: the acquisition of homolog pairing, the occurrence of efficient intergenic recombination between homologs during pairing, the suppression of sister-chromatid separation in the first division, and the absence of S phase at the start of the second division (Wilkins & Holliday, 2009) (Table 1).

Meiotic cell division consists of a single DNA replication followed by two rounds of chromosome segregation (meiosis I and meiosis II), which splits the chromosome number to ultimately produce haploid gametes (Haversat et al., 2021; Takemoto et al., 2020). Several highly conserved genes are active during each stage of meiosis, and together they are referred to as the Meiotic Toolkit (Malik et al., 2008; Schurko & Logsdon, 2008). According to a review done by Hofstatter et al. (2020), there are a total of 12 meiosis-specific proteins identified in eukaryote model organisms that will work alongside other proteins that although not specific to meiosis, have an essential role in meiosis success (*hap*2, *spo*11, *rec*8, *hop*1, *pch*2, *dmc*1, *mnd*1, *hop*2, *mer*3, *msh*4, *msh*5 and *zip*4). A previous review was done by Malik et al. (2008), which mentions that the eukaryotic meiosis toolkit has 29 genes in total, being 9 of them meiosis-specific (*dmc*1, *hop*1, *hop*2, *mer*3, *mdn*1, *msh*4, *msh*5, *rec*8 and *spo*11). Much of the discussion

about the maintenance of sex and meiosis has been done using model organisms, particularly in animals, and does not analyze the phylogeny of these genes. Since animals are a group of organisms that arose long after meiosis originated (Wilkins & Holliday, 2009), it is interesting to consider and study organisms that are early eukaryotes evolution such as red algae.

Unfortunately, there is a lack of genetic information regarding the meiotic genes of Rhodophyta species. Only recently, transcriptome studies of red algae species such as *Thorea hispida* (Nan et al., 2020) and *Bostrychia moritziana* (Shim et al., 2021) include more details about possible meiotic genes in action. Previous studies do identify the presence of some meiotic-related genes, but not an in-depth understanding of how the red algal meiotic toolbox looks like if the meiotic machinery among these organisms is conserved, or about the phylogenetic relationship of the meiotic genes among the different taxonomic groups. In chapter 4 of this thesis, I will analyze the meiotic genes present in Rhodophyta species, using the meiosis toolbox of *Arabidopsis thaliana* as a reference (Malik et al., 2008). From this toolbox, I analyzed a total of 6 gene groups were analyzed phylogenetically: *spo*11; *rad*51 and *dmc*1; *mnd*1 and *hop*2; *smc* gene family; *mut*L, *mlh* gene family and *pms*1; *mut*S and *msh* gene family. These results, in addition to the new sequencing of the chloroplast genome, and the detailed analysis of the rates of substitution of 74 chloroplast genes among 30 red algae species, will deepen the understanding of the molecular structure, and the relationship of *B. atropurpurea* to other red algae species.

Table 1. 2 Comparison of mitotic and meiosis stages*.

Mitotic stage	Result	Meiotic stage	Result
S phase	Chromatid duplication	S phase, I	Chromatid duplication; DNA breaks introduced
Prophase	Chromosome condensation	Prophase, I	Chromosome condensation; *homolog pairing , recombination*
Metaphase	Chromosome alignment in center of spindle body	Metaphase, I	Alignment of homologs in center of spindle body
Anaphase	Centromere splitting; chromatids separated	Anaphase, I	Separation of homologs with independent assortment; *centromere splitting suppressed*
Telophase	Chromatid decondensation; two daughter nuclei with mother-cell ploidy, single-chromatid chromosomes	Telophase, I	Partial or complete chromatid decondensation; two haploid nuclei with replicated chromatids
		Prophase, II	*No S phase* ; chromosome condensation
		Metaphase, II	Alignment of replicated chromatids
		Anaphase, II	Centromere splitting; separation of chromatids
		Telophase, II	Chromatid decondensation; four haploid nuclei, single-chromatid chromosomes

*Table based on Wilkins & Holliday, 2009. The four novel events of meiosis are indicated by italics and red.

1.10 Thesis objectives

This thesis aims to address the following objectives in order to understand the similarities and differences of *Bangia atropurpurea* when compared to other Rhodophyta genomes:

Chapter 2:

- To sequence, assemble and analyze a draft genome of the *B. atropurpurea* chloroplast.

- To provide a detailed comparison of the first freshwater chloroplast genome of the Bangiales order to pertinent species.

Chapter 3:

- To estimate the rates of substitutions of the chloroplast genes from Rhodophyta species, pointing to similarities and differences that could have contributed to the survival of *B. atropurpurea* in the Great Lakes.

Chapter 4:

- To investigate the presence or absence of specific genes related to meiosis in *Bangia atropurpurea*, which may provide information on the evolution of this species.

Chapter 2

The chloroplast genome of the sole freshwater species, *Bangia atropurpurea*, in the Bangiales (Rhodophyta), using metagenomic sequencing

2.1 Introduction

Studies revolving around chloroplast genes and genomes in red algae are biologically relevant since these organelles retain unique genes that tell us the history of the evolution of photosynthetic organisms on Earth. Because of the number of molecular evidence produced in the last decade, the hypothesis of a monophyletic origin of all chloroplasts is broadly accepted (Bhattacharya and Medlin, 1995; Turner et al., 1999; Simpson and Stern, 2002; Palmer, 2000; Qiu et al., 2013; Kim et al., 2014; Rockwell et al., 2014). The universal ancestor for the Eukaryotes belongs to the Archaea group, the *Lokiarchaeota* (Spang et al., 2015), which branched into six supergroups: Opisthokonta, Amoebozoa, Archaeplastida (Plantae), Chromalveolata, Rhizaria, and Excavata (Adl et al., 2005, 2012; Keeling et al., 2005; Reyes-Prieto et al., 2007). From the supergroup Archaeplastida emerged green algae and plants, the red algae and the glaucophyte algae (Adl et al., 2005; Reyes-Prieto et al., 2007; Bhattacharya et al., 2004, 2013; Burki, 2014). The endosymbiotic events that are supported by recent research provide evidence that the primary endosymbiosis evolved from a cyanobacterium that can be compared with the modern cyanobacterium *Gloeomargarita lithophora*, which is present in freshwater and microbial mats (Ponce-Toledo, 2017; de Vries and Archibald, 2017).

The secondary endosymbiosis led to the origin of plastids in important algal groups, such as dinoflagellates, heterokonts, cryptomonads, and parasites such as the apicomplexa (Douglas and Penny, 1999; Müller et al., 2001b; van Dooren et al., 2001; Oliveira and Bhattacharya, 2000; Simpson and Stern, 2002; Gould et al., 2015). Since it is settled that red algae are an early branching within eukaryotes, and, therefore, play a very important role in algal evolution, they continue to be excellent models for new investigations regarding plastid evolution (Gray, 1993; Bhattacharya & Medlin, 1995; Palmer et al., 2003; Bhattacharya et al., 2004; Reyes-Prieto et al., 2007). During the process of endosymbiosis, part of the gene content of the original organism was transferred to the nucleus of the host, and part was maintained in the chloroplast to keep the organelle machinery active, mostly involved in photosynthesis and proteins needed for building ribosomes (Glöckner et al., 2000; Zhang et al., 2020). In addition, the chloroplast genome has a complete independent set of tRNA, rRNA and consequently, exclusive gene expression (Martin et al., 1998; Zhang et al., 2020).

The order Bangiales (Bangiophyceae) includes both foliose and filamentous forms and the latter have been reported in the fossil record dating 1.2 MA and 1.6 BYA (Butterfield, 2000; Bengtson et al. 2017). Some foliose members within the genus *Pyropia* and *Porphyra* have considerable economic importance as nori and are generally well studied, (Broom et al., 2002; Chan et al., 2011; Sutherland et al., 2011; Xie et al., 2013; Brawley et al., 2017; Yang et al., 2017); however, less is known about the filamentous forms except several phylogenetic studies using marker genes (Müller et al. 1998, 2003, 2005). Members of this order primarily occupy marine environments (Nelson, 2006; Moenne et al., 2016; Kogame et al., 2017) except *Bangia atropurpurea*, initially described from the Weser River in Germany (Roth, 1806), and subsequently reported in other European lakes and rivers and also in Asia (Den Hartog, 1972; Reed, 1980; Kumano, 1980; Müller et al., 2003 and 2005). In addition, this species is a known invader to the Laurentian Great Lakes, likely the result of ballast water discharge (Müller et al., 2003; Shea et al., 2014). This species was first observed in Lake Erie in 1964 (Lin and Blum, 1977) and has spread to all the Laurentian Great Lakes except Lake Superior (Sheath, 1987; Shea et al., 2014). Phylogenetic analyses of chloroplast and nuclear gene markers have noted that the populations in the Great Lakes are all genetically identical to each other and closely related to freshwater populations from Europe (Müller et al., 1998; Müller et al., 2003; Shea et al., 2014). More recently, there are reports of this species in the Finger Lakes in New York State (Müller et al. unpublished).

The most extended taxonomic revision of the order Bangiales using molecular markers (nuclear SSU rRNA and *rbc*L genes) was done by Sutherland et al. (2011) where 157 Bangiales species were analyzed. This study resulted in 15 genera of Bangiales being recognized, of which 7 are filamentous - four previously described and three newly recognized - and 8 are foliose. According to their findings and those noted in Müller et al. (1998, 2003, 2005), the species, *B. atropurpurea,* is phylogenetically affiliated with two foliose genera, *Clymene* and *Porphyra*. However, this relationship was only moderately supported and the position of *B. atropurpurea* is still unresolved. Despite the extensive work on the phylogeny of members of the Bangiales using various marker genes, there are still few nuclear and organellar genomes available for members of this group. There is only one chloroplast genome from a *Bangia* species, collected from an estuary in China (Cao et al., 2018), that groups with marine collections in the *"Bangia* 2" clade noted in Sutherland et al. (2011) and is phylogenetically distant from *Bangia atropurpurea*.

Even though several genomes (nuclear and chloroplast) were sequenced for foliose members of the Bangiales, knowledge is lacking for filamentous species. Given the long evolutionary history of filamentous Bangiales in the fossil record (Butterfield, 1990; 2000), and that the species *Bangia atropurpurea* is the sole freshwater species within the order, in addition, to being a known invader into North American lakes, this species is ideal for genomic sequencing. Isolating and growing algal species for genomic analyses is often challenging due to an inability to grow in culture, slow growth, or contamination; since obtaining an axenic culture of algae tends to be more difficult than for higher plants (Polne-Fuller, 1988; Fries, 1980 and 1985; Baweja et al., 2009). Although *B. atropurpurea* is fairly easy to isolate, growth is very slow (~9 mm/month) and hence obtaining enough biomass for DNA extraction for genomic sequencing is difficult. Hence, the present study mapped a draft of the first chloroplast genome (97% coverage) from environmental metagenomic data in the only freshwater species in the order Bangiales. Not only does this study provide clarity regarding the relationship of this species within the order Bangiales, but it also highlights a method by which chloroplast genomes can be obtained from environmental samples of small macroalgae.

2.2 Materials and Methods

To overcome the slow-growing *in vitro* challenges, we employed a strategy involving the following steps: 1) collection of filaments from a location in the Laurentian Great Lakes; 2) metagenomic sequencing of the environmental sample; and 3) isolation and assembly of the red algal chloroplast genome through a targeted bioinformatics workflow (Fig. 2.1).

2.2.1 Sample preparation and sequencing:

Samples of the multicellular filamentous red alga *B. atropurpurea* were collected from rocks on the Lake Ontario shoreline, in the city of Burlington, Ontario, Canada (43°19'12"N 79°47'59"W), on July 17th, 2017. The filaments were manually cleaned of macroscopic epiphytes, micro-algae, *Cladophora* sp. filaments and sediment under a dissecting microscope and rinsed twice with sterile autoclaved lake water and frozen at -20°C until DNA extraction (Fig. 2.1-1). To extract DNA, the filaments were ground using a clean mortar and pestle and

liquid nitrogen until the material was a fine powder. The Sox DNA isolation kit (Metagenom Bio Inc., Waterloo, Canada, Appendix A) was then used to extract DNA with the following modification: After adding 100 mg of ground cells and buffer Sox1 to a bead tube, the mixture was incubated at 70^{o}C for 10 min. Cells were then homogenized using FastPrep-24 (MP Biomedicals) at 6M/second for 40 seconds. The quality of genomic DNA was evaluated with 0.8% TAE agarose gel. The final DNA prep was quantified using the dsDNA HS Assay Kit (ThermoFisher) and diluted to 0.25 ng/ml. The genomic sequencing library was constructed using the Nextera XT kit (Illumina) based on the supplier's guide. Index sequences (i5, GTACTGAC; i7, TCATGAGC) were incorporated into PCR primers. Library DNA was evaluated with 2% TAE agarose gel and quantified with the dsDNA HS Assay Kit (Fig.1-2a). The DNA of 0.4 kb to 1 kb was denatured and pooled with other sequencing amplicons. Sequencing was performed for 2×250 cycles using the MiSeq platform (Illumina). After trimming off adapter sequences, DNA sequences were assembled (Fig. 2.1-2b).

2.2.2 Assembly, alignment, and annotation:

The mixed pool of short DNA sequences, paired-end reads (250 bp) from three runs of the same sample, were combined and were the input for the ATLAS framework (v1.0.31; White et al., 2017) with default parameters using SPAdes (v3.11.0; Bankevich et al., 2012) as the assembly tool. Reads were trimmed, filtered, error-corrected, and normalized prior to assembly with SPAdes. To separate algal chloroplast-derived contigs from non-algal (e.g., bacterial, Fig. 2.1-3) contigs, all contigs were aligned to reference chloroplast genomes (*Porphyra purpurea*, accession number NC_000925, and *Porphyra umbilicalis*, accession number NC_035573.1) using BWA-MEM (Li and Durbin, 2010) with default parameters. Reads were aligned back onto the reordered contigs to visualize read coverage using Bowtie2 (Langmead and Salzberg, 2012) (with default parameters). The resulting BAM file and contig FASTA file were manipulated with in-house scripts and the shinyCircos R package (Yu et al., 2018) to produce a Circos plot (Fig. 2.1-5b and Fig. 2.2 in details) having the *Porphyra purpurea* chloroplast genome as reference. The contigs were then reordered to produce a scaffold (Fig. 1-4) using MeDuSa (Bosi et al., 2015). Gene annotation was then performed using GeSeq (Tillich et al., 2017), refined manually using Geneious Prime, and visualized using OGDRAW (http://ogdraw.mpimp-golm.mpg.de). Finally, several assembly artifacts (insertions of likely

bacterial origin) were removed from the scaffold and replaced with gaps. The final *B. atropurpurea* assembly was deposited into the NCBI (GenBank accession number MN484624).

2.2.3 Synteny and phylogeny

The *B. atropurpurea* chloroplast genome organization was analyzed through syntenic comparison to related red algal chloroplasts using Progressive Mauve (Darling et al., 2010), which generated a synteny map and identified conserved orthologous blocks (Fig. 2.1-5a and Fig. 2.5 in details). The concatenated alignment of orthologous chloroplast genes was then analyzed, and the orthologous genes were extracted from the output file with an in-house script that was uploaded to GitHub (https://github.com/doxeylab/labScripts/tree/master/mauveExtract). The data was then used for the phylogenetic analyses with other 20 chloroplast genomes (Table 2.1), performed using MEGA 10 (Kumar et al., 2016) with the evolutionary history inferred by using the Maximum Likelihood Method using the General Time Reversible + G model (Posada and Crandall, 1998) (Fig.2.1-6 and Fig. 2.6 in details).

Table 2. 1. Chloroplast genomes of Rhodophyta species available at NCBI database

Species	Phylum	Class	Order	Genome size (bp)	GC%	Coding genes	rRNA	tRNA	Accession number	References
Bangia atropurpurea *	Rhodophyta	Bangiophyceae	Bangiales	186.840	32.9	207	3	35	MN484624	This study
Bangia fuscopurpurea *	Rhodophyta	Bangiophyceae	Bangiales	196.913	33.5	205	6	37	KP714733	Cao et al., 2018
Porphyra purpurea *	Rhodophyta	Bangiophyceae	Bangiales	191.028	33	250	2	35	NC_000925	Reith and Munholland, 1995
Porphyra umbilicalis *	Rhodophyta	Bangiophyceae	Bangiales	189.933	32.9	209	6	37	JQ408795	Smith et al., 2012
Pyropia dentata	Rhodophyta	Bangiophyceae	Bangiales	192.266	32.6	92	12	44	LC521919.1	Choi et al., 2020
Pyropia perforata *	Rhodophyta	Bangiophyceae	Bangiales	189.788	32.9	209	3	35	KF515973	Hughey et al., 2014
Pyropia haitanensis *	Rhodophyta	Bangiophyceae	Bangiales	195.597	32.98	213	6	37	NC_021189	Wang L, 2013
Pyropia kanakaensis *	Rhodophyta	Bangiophyceae	Bangiales	194.631	32.0	ND	ND	ND	KJ776836	Hughey et al., 2014
Pyropia endiviifolia *	Rhodophyta	Bangiophyceae	Bangiales	195.784	33.3	210	6	37	KT716756	Xu et al., 2018
Pyropia fucicola *	Rhodophyta	Bangiophyceae	Bangiales	191.982	32.7	ND	ND	ND	KJ776837	Hughey et al., 2014
Pyropia yezoensis *	Rhodophyta	Bangiophyceae	Bangiales	191.975	32.09	213	6	37	KC517072	Wang L, 2013
Wildemania schizophylla *	Rhodophyta	Bangiophyceae	Bangiales	193.008	34.4	211	4	34	NC_029576	Hughey et al., 2016
Cyanidioschyzon merolae *	Rhodophyta	Cyanidiophyceae	Cyanidiales	149.987	37.6	243	3	31	AB002583	Otha et al., 2003
Cumathamnion serrulatum	Rhodophyta	Florideophyceae	Ceramiales	174.192	27.2	193	3	29	MW292565.1	Kim et al., 2021
Corallina officinalis	Rhodophyta	Florideophyceae	Corallinales	178.183	30.2	205	3	31	MT211887.1	Yesson et al., 2020
Chondrus crispus *	Rhodophyta	Florideophyceae	Gigartinales	180.09	28.7	204	3	24	NC_020795	Collen et al., 2013
Betaphycus gelatinus	Rhodophyta	Florideophyceae	Gigartinales	178.394	28.93	240	3	30	MN240356	Zhang et al., 2020
Eucheuma denticulatum	Rhodophyta	Florideophyceae	Gigartinales	177.003	29.61	240	3	30	MN240357	Zhang et al., 2020
Kappaphycus striatus	Rhodophyta	Florideophyceae	Gigartinales	176.763	29.35	238	3	30	MN240358	Zhang et al., 2020
Kappaphycus alvarezii	Rhodophyta	Florideophyceae	Gigartinales	178,205	29.6	238	3	30	KU892652	Liu et al., 2019
Gracilaria bailiniae	Rhodophyta	Florideophyceae	Gracilariales	185.129	27.9	193	3	29	NC_038100.1	Liu et al., 2019
Palmaria palmata *	Rhodophyta	Florideophyceae	Palmariales	192.960	33.9	204	2	22	NC_031147	Cho et al., 2018
Porphyridium purpureum *	Rhodophyta	Porphyridiaceae	Porphyridiales	217.69	30.3	224	6	29	NC_023133.1	Tajima et al., 2014
Gloeochaete wittrockiana *	Glaucophyta	Gloeochaetaceae	Gloeochaetales	143.343	29.6	129	3	31	MF167426	Reyes-Pietro et al., 2018
Glaucocystis sp. *	Glaucophyta	Gloeochaetaceae	Gloeochaetales	130.276	33.4	137	3	32	MF167424	Reyes-Pietro et al., 2018
Glaucocystis incassata *	Glaucophyta	Gloeochaetaceae	Gloeochaetales	137.017	33.6	137	3	31	MF167425	Reyes-Pietro et al., 2018
Rhodomonas salina *	Cryptophyta	Pyrenomonadace	Pyrenomonadal	135.854	34	183	ND	31	EF508371.1	Khan et al., 2016
Guillardia theta *	Cryptophyta	Geminigeraceae	Pyrenomonadal	121.524	33	ND	8	30	NC_000926.1	Douglas and Penny, 1999

Available chloroplast genomes until the moment this study was conducted. ND = not determined/not on NCBI.

***** Species used for phylogenetic analysis. The remain of species were not available during the analysis as complete chloroplast genomes.

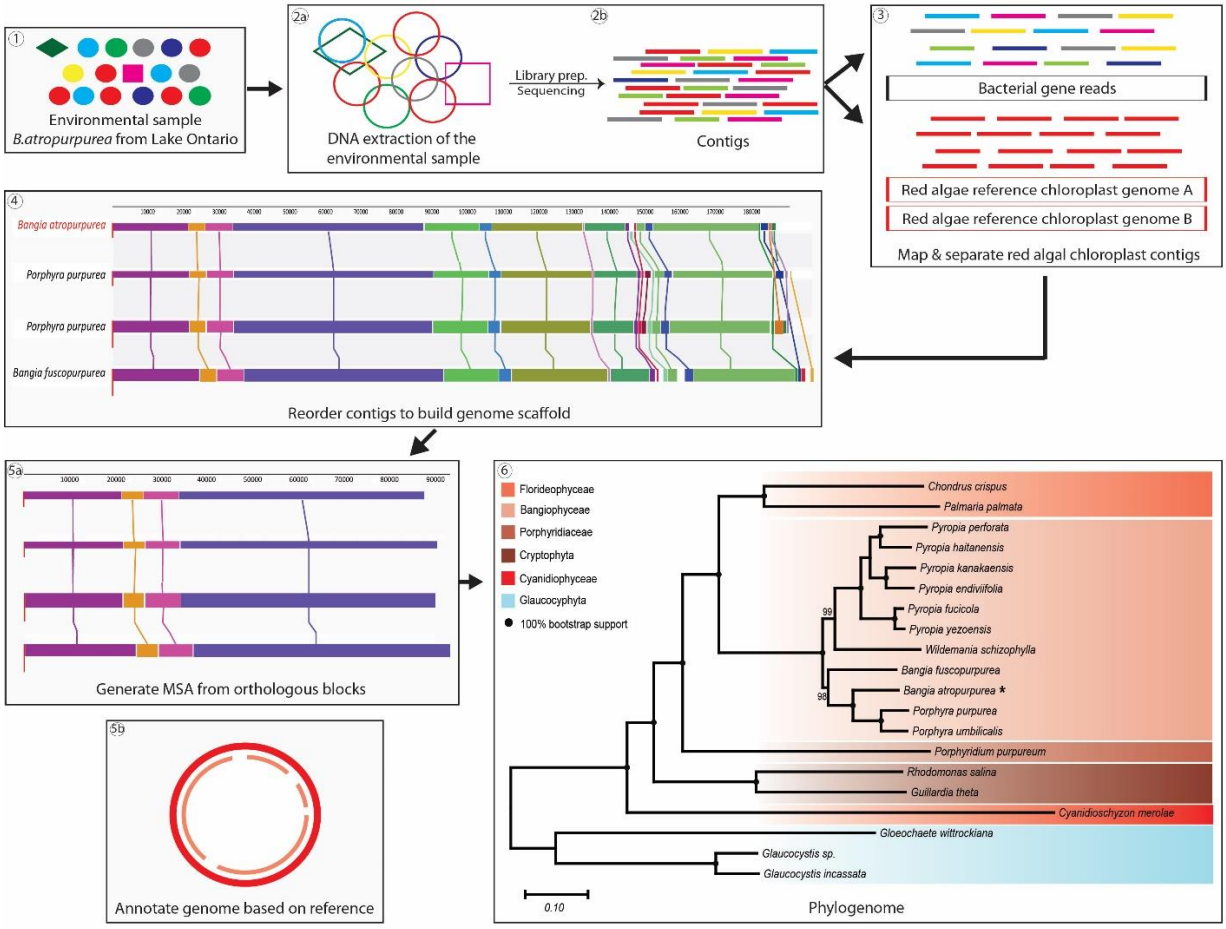

Figure 2.1. Bioinformatic workflow for assembling the chloroplast genome of *B. atropurpurea* from metagenome data. An environmental sample containing mixed organisms was obtained (1), DNA was extracted from all the organisms in the sample (2a) and sequenced using the Illumina MiSeq platform (2b). A mixed pool of assembled contigs (2b) was then separated into contigs from *B. atropurpurea* and those from the other organisms (mainly bacterial) by taxonomic binning (3), using the chloroplast genome of *Porphyra purpurea* (genome A, NC_000925) and *P. umbilicalis* (genome B, JQ408795) as reference. The *Bangia* chloroplast genome was re-aligned and re-ordered by comparison with the synteny of close related red algal genomes (4; Figure 2.5). Conserved, orthologous blocks were concatenated into a whole-chloroplast-genome multiple sequence alignment (5a), and once the coverage was obtained (97%), the chloroplast genome was annotated to produce a gene map (5b; Figure 2.2). The concatenated alignment was used to produce a whole-chloroplast-genome phylogeny (6; Figure 2.6).

2.3 Results

2.3.1 Sequence analysis

Using a metagenomic assembly pipeline, I was able to successfully bin and assemble a high-quality draft chloroplast genome of *B. atropurpurea* (Fig. 2.2). The assembly is 186,840 bp in length, has a G+C content of 32.9%, and has estimated coverage of 97% (Table 2.1; Fig. 2.2). The depth of coverage across the full circle genome was visualized by a Circos plot, which shows variation in coverage distributed across 13 total contigs. We inspected regions of high coverage to rule out the possibility of spurious or non-specific read alignments from other species and found that these regions were genomically conserved within the class Bangiophyceae (Lee et al., 2016), and distinct from green algae and bacterial DNA present in the original environmental material.

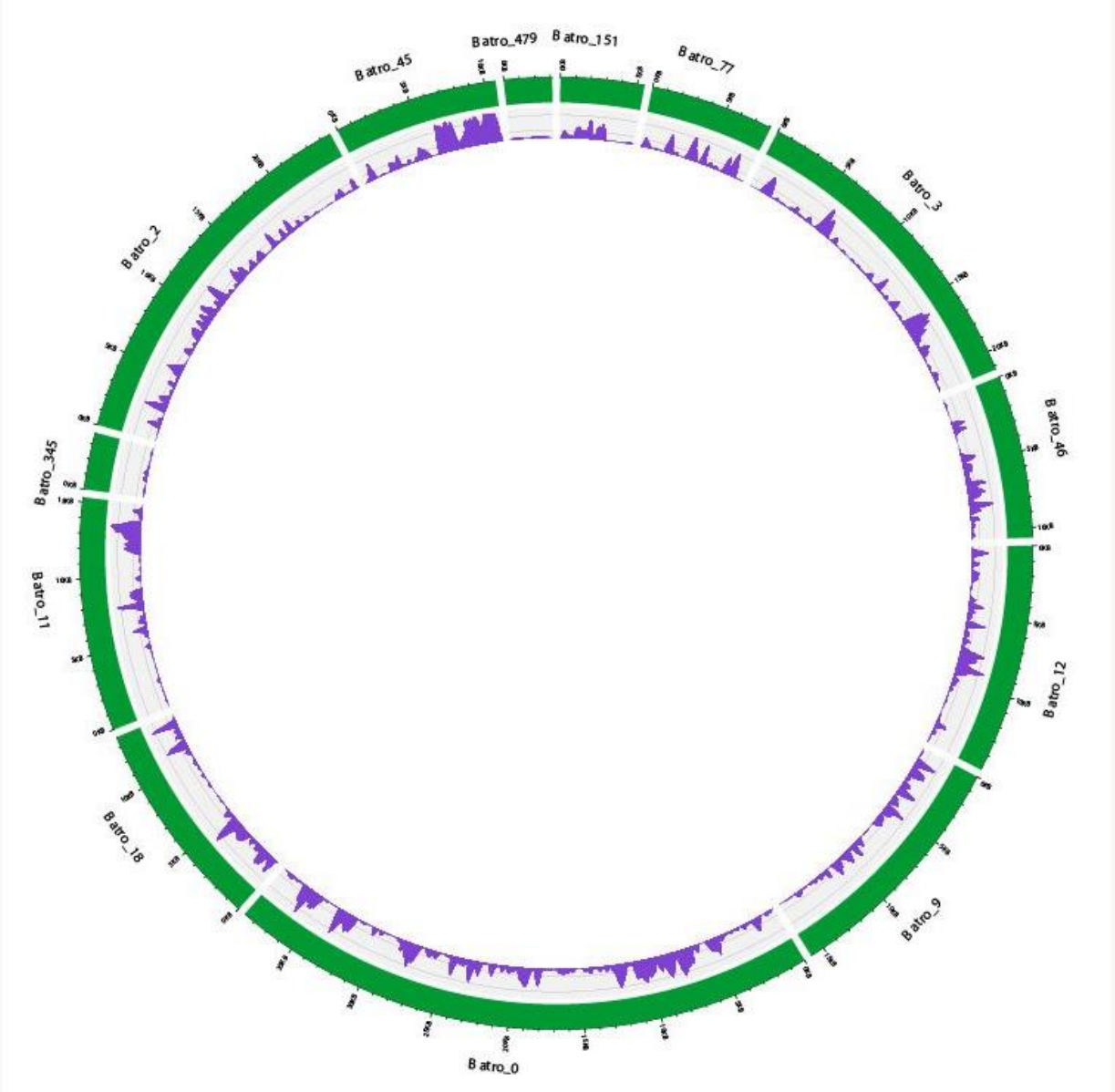

Figure 2.2 Estimated coverage of *B. atropurpurea* chloroplast genome having as reference the *Porphyra purpurea* chloroplast genome (NC_000925). Once the *Bangia* contigs were re-aligned and re-ordered by synteny, the BAM and FASTA files were manipulated using scrips and shinyCircos R package to create this circular plot. The name of the contigs is around the outer-circle. The purple inner-circle indicates the 97% coverage of *B. atropurpurea* in comparison to *P. purpurea* chloroplast genome. The gap between contigs accounts for 3% of the chloroplast genome.

2.3.2 Genome organization

The chloroplast genome of *B. atropurpurea* (Fig. 2.3) encodes a core set of 207 protein-coding sequences (CDS), 35 tRNAs and 3 rRNAs, and, despite the 3% estimated gap content in the chloroplast genome sequence based on alignment to the reference genome, these numbers are similar with what has been observed in chloroplast genomes from related red algal species such as *P. purpurea*, *P. umbilicalis* and *B. fuscopurpurea* (Table 2.1) (Reith and Munholland, 1995; Brawley et al., 2017; Cao et al., 2018). The rRNA and tRNA content are more variable than the protein-coding genes, with the *B. atropurpurea* assembly lacking both copies of the *trn*L tRNA and *trn*A tRNA genes. These are present in both *Porphyra* species, and they are known to occur inside of the repeated rRNA operons (Reith and Munholland, 1995; Brawley et al., 2017; DePriest et al., 2013). The same genes (*trn*L and *trn*A) are present in *B. fuscopurpurea* but not duplicated. The similarity between the gene content of chloroplast genomes from closely related Bangiales species and *B. atropurpurea* is visualized by a Venn Diagram (Fig. 2.4 and Table 2), revealing that there are no exclusive or novel genes in our assembly. Plus, the 207 genes are shared between all species, and only 11 genes are in common with PU and PP but not with BF, and 11 other genes are shared with PP and BF and not with PU. Interestingly, there are 11 genes shared exclusively with PP, and no exclusive genes are shared with BF. The chloroplast map of the same four species presents high similarity in terms of translation direction: PP is the only chloroplast genome with a significant difference in direction. When comparing the gene content of *B. atropurpurea* chloroplast genome with other Bangiales and Rhodophyta species (Appendix B), a few genes are missing from the assembly, being absent in *P. umbilicalis* and *B. atropurpurea*, but present in most of the other Bangiales and Rhodophyta species used for comparison in this study: *pet*N, *ycf*27, *orf*38, *orf*68, *ftr*C, *bas*1, *suf*C, *ccd*A, *psb*Y. Moreover, there are genes absent in *P. umbilicalis* but present in *B. atropurpurea* and other red algae species: *orf*27, *orf*111, *orf*114, *orf*199, *orf*287, *orf*382, *orf*621, *pre*A and *rpl*29 (Table 2.2 and Appendix B).

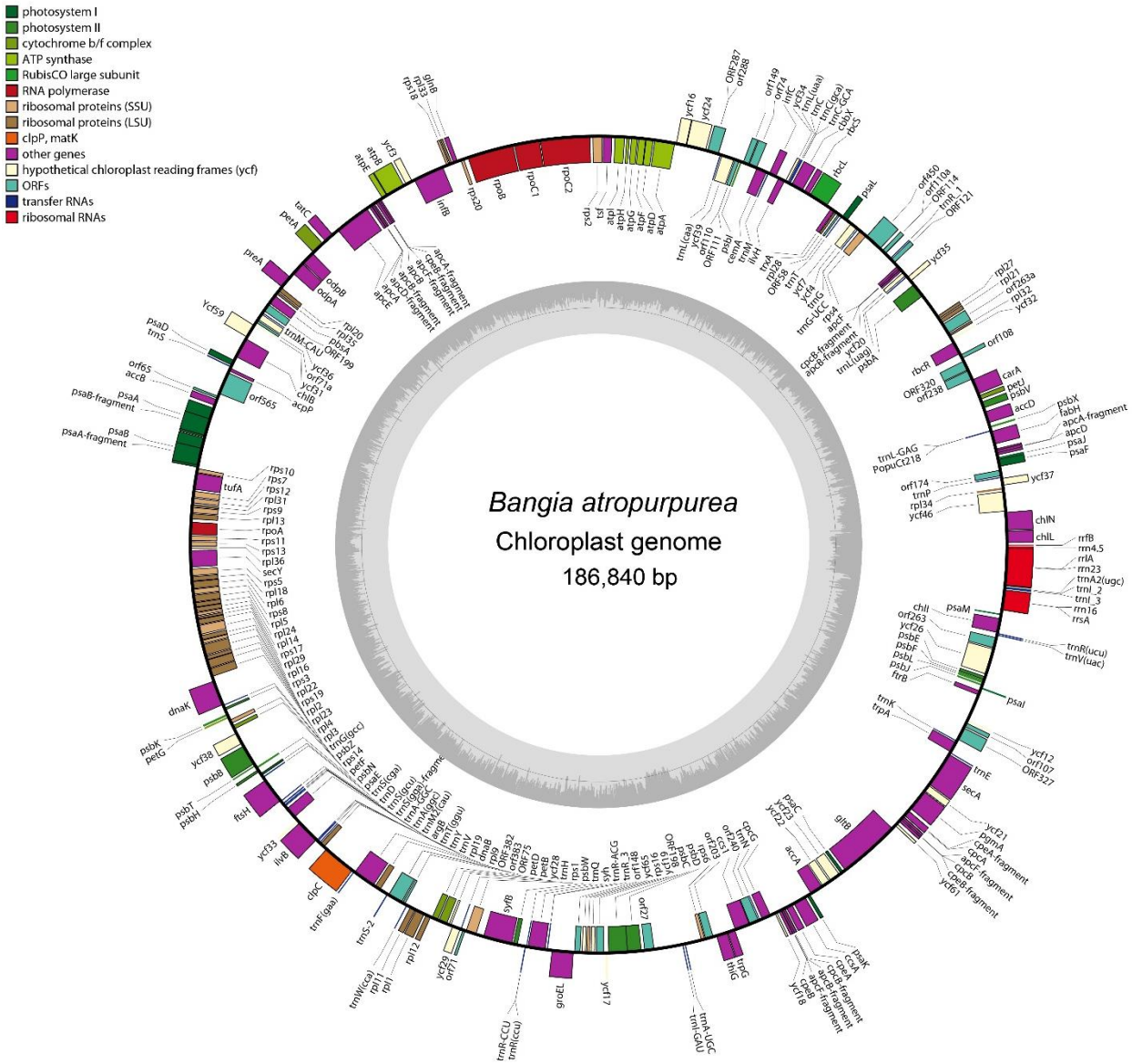

Figure 2.3 Gene map of the *B. atropurpurea* chloroplast genome. Annotated genes are indicated in different colours and correspond to the functional categories listed in the legend. Genes inside the circle are transcribed clockwise whereas those outside the circle are transcribed counterclockwise. The inner circle displays the GC content represented by dark gray bars.

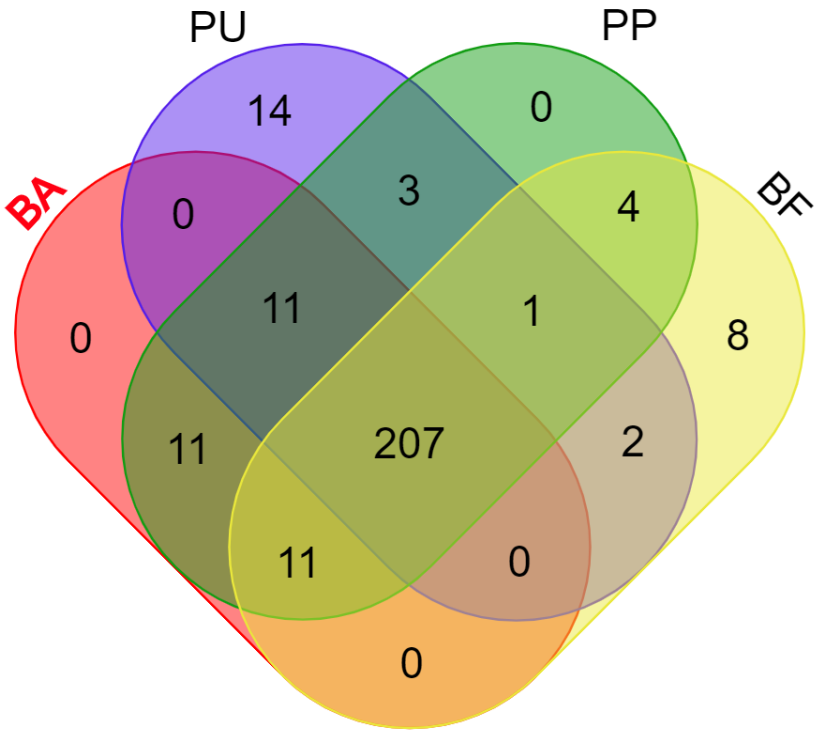

Figure 2.4 Venn diagram of genes shared among four closely related Bangiophyceae chloroplast genomes. BA: *B. atropurpurea* (accession number: MN484624); PU: *Porphyra umbilicalis* (accession number: JQ408795); PP: *Porphyra purpurea* (accession number:NC_000925); BF: *Bangia fuscopurpurea* (accession number:KP71473337). PU and PP chloroplast genomes were used as reference genomes in this study. BF is the only other *Bangia* species with a complete chloroplast genome available to date.

Table 2.2 Gene content of *B. atropurpurea* in comparison to the most common chloroplast encoded genes of other red algae species.

Group of Genes	Gene name								Putative gene role	Reference
Atp synthase	atpA	atpB	atpD	atpE	atpF	atpG	atpH	atpI	Catalyzes the light-driven synthesis of ATP. Is activated in the light and inactivated in the dark by redox-modulation through the thioredoxin system.	Kohzuma et al., 2017
ClpP	clpC								Controls the availability of short-lived regulatory proteins and in removing abnormal or damaged proteins.	Shikanai et al., 2001
Cytochrome b f complex	petA	petB	petD	petF	petG	petN			Part of the chloroplast thylakoids three main complexes involved in electron transfers, along with Photosystem I and Photosystem II.	Hope, 1993
Hypothetical chloroplast reading frames (ycf)	ycf3	ycf4	ycf7	ycf12	ycf16	ycf17	ycf18	ycf19	Conserved hypothetical chloroplast open reading frame.	Hallick and Bairoch, 1994
	ycf20	ycf21	ycf22	ycf23	ycf24	ycf26	ycf27	ycf28		
	ycf29	ycf31	ycf32	ycf33	ycf34	ycf35	ycf36	ycf37		
	ycf38	ycf39	ycf46	ycf59	ycf61	ycf65				
ORFs	orf27	orf38	orf58	orf65	orf68	orf71	orf71a	orf74	Conserved open reading frames among plants, and often plants, algae, and cyanobacteria, functions are unknown.	Hallick and Bairoch, 1994
	orf75	orf107	orf108	orf110	orf110a	orf111	orf114	orf121		
	orf148	orf149	orf174	orf198	orf199	orf203	orf238	orf240		
	orf263	orf263a	orf287	orf288	orf320	orf327	orf382	orf383		
	orf450	orf565	orf621							
Photosystem I	psaA	psaB	psaC	psaD	psaE	psaF	psaJ	psaK	Mediates the light-driven electron transportfrom plastocyanin to ferredoxin.	Scheller et al., 2001
	psaL	psaM	psaI							
Photosystem II	psbA	psbB	psbC	psbD	psbE	psbF	psbI	psbJ	First protein complex in the light-dependent reactions of oxygenic photosynthesis; Primary site of photoinhibition	Baker, 1991
	psbK	psbL	psbH	psbN	psbT	psbV	psbW	psbX		
	psbY	psbZ								
Ribosomal proteins (50S)	rpl1	rpl2	rpl3	rpl4	rpl5	rpl6	rpl9	rpl11	Protein biosynthesis and translational regulation	Stelzl et al., 2001
	rpl12	rpl13	rpl14	rpl16	rpl18	rpl19	rpl20	rpl21		
	rpl22	rpl23	rpl24	rpl27	rpl29	rpl28	rpl31	rpl32		
	rpl33	rpl34	rpl35	rpl36						
Ribosomal proteins (30S)	rps1	rps2	rps3	rps4	rps5	rps6	rps7	rps8	Protein biosynthesis and translational regulation	Stelzl et al., 2001
	rps9	rps10	rps11	rps12	rps13	rps14	rps16	rps17		
	rps18	rps19	rps20							
RNA polymerase	rpoA	rpoB	rpoC1	rpoC2					Transcription initiation	Roeder, 1996
RubisCO	rbcL	rbcS	rbcR						First major step of carbon fixation	Feller at al., 2008
Other genes	accA	accB	accD	acpP	apcA	apcB	apcD	apcE		
	apcF	argB	carA	cbbX	ccs1	ccsA	cemA	chlB		
	chlL	chlN	cpcA	cpcB	cpcG	cpeA	cpeB	dnaB		
	dnaK	fabH	ftrB	ftsH	glnB	gltB	groEL	ilvB		
	ilvH	infB	odpA	odpB	pgmA	pbsA	rne	secA		
	secY	tatC	thiG	trpA	trxA	tsf	tufA	syfB		
	syh	chlI	petJ	trpG	infC	preA	bas1	sufC		

Genes absent from B. atropurpurea and P. umbilicalis, present in most of the other red algae
Genes present in B. atropurpurea and absent in P. umbilicalis

2.3.3 Synteny and Phylogeny

In the synteny analysis of 12 red algal chloroplast genomes (Fig. 2.5), *B. atropurpurea* exhibits extensive conservation with *P. purpurea*, *P. umbilicalis* and other Bangiales species. In general, all the Bangiales species have a highly conserved chloroplast genome, and, therefore, a high degree of identity, which can be verified when comparing the genomes against *Palmaria palmata*, a red alga that belongs to the class Florideophyceae, order Palmariales, and show different degrees of rearrangement. The Maximum Likelihood tree constructed from 17 species of 5 different classes of Rhodophyta organisms was used to estimate the relationships of *B. atropurpurea* within the Bangiales (Fig. 2.6). The tree exhibit 100% support for the placement of *B. atropurpurea*, being closer to the marine species *Porphyra* species than to the marine species with the same genus *B. fuscopurpurea*.

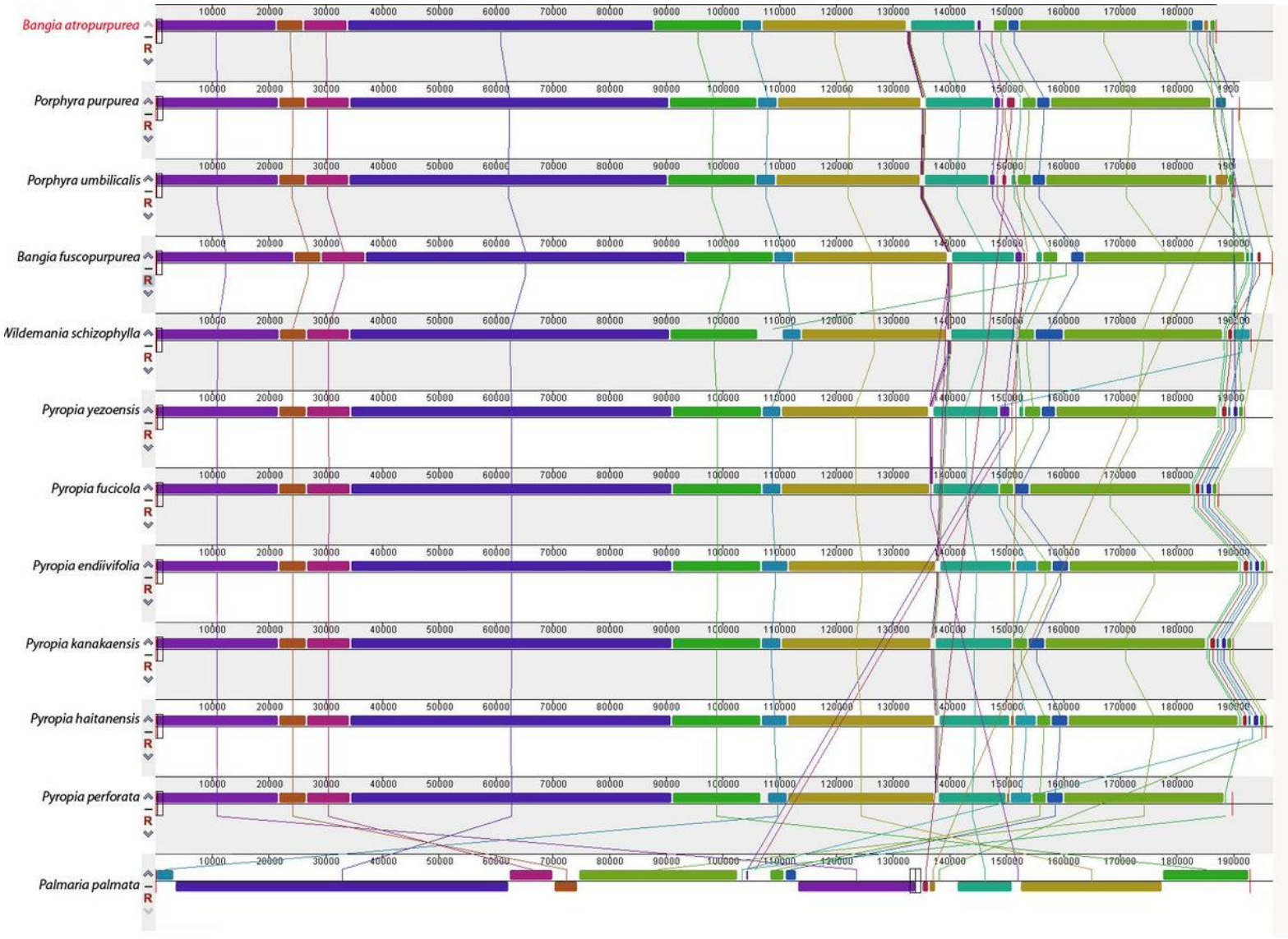

Figure 2.5 Synteny based on the chloroplast genomes of red algae species closely related to *B. atropurpurea*, and one member of the Florideophycidae, *Palmaria palmata*. The orthologous blocks shown here provided the concatenated alignment used for the phylogenetic analysis.

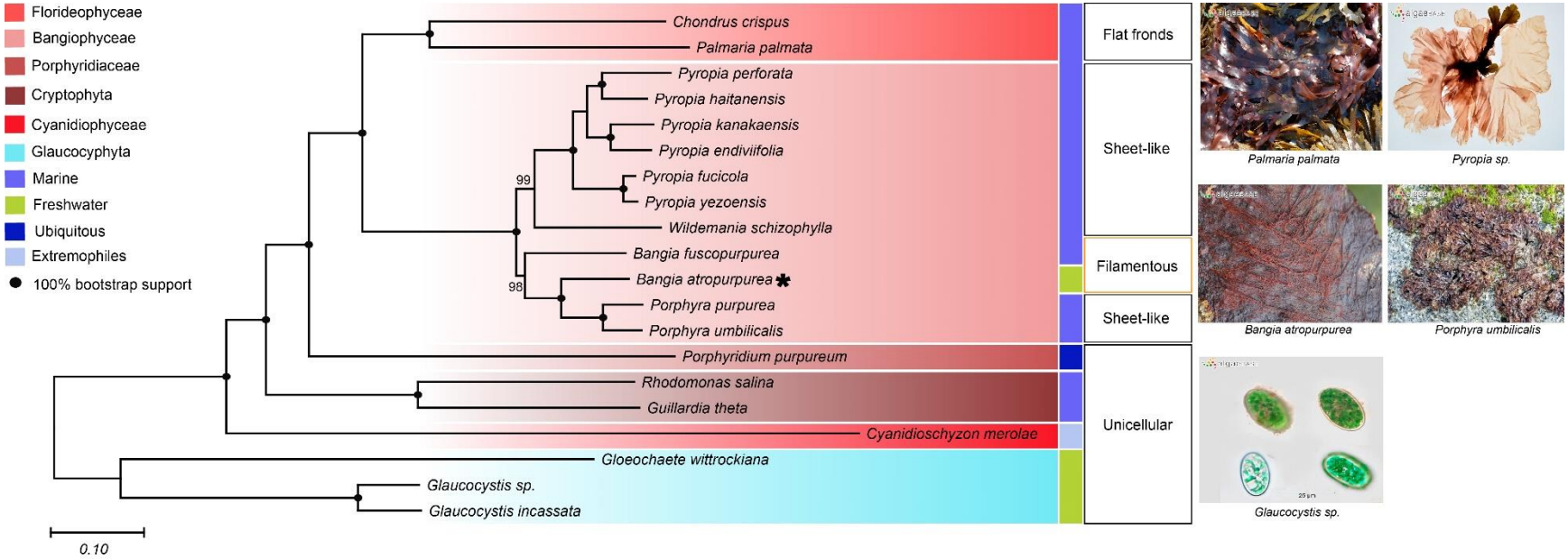

Figure 2.6. Phylogenetic tree based on a concatenated 35,032 bp multiple alignment of ortholog blocks, placing the *B. atropurpurea* chloroplast genome within the Bangiales species. The pictures that exemplify the different types of thalli were retrieved from the AlgaeBase database (http://www.algaebase.org).

2.4 Discussion

The draft chloroplast genome

This thesis presents for the first time a pipeline created to retrieve a complete draft of the chloroplast genome using an environmental sample and MiSeq technology. The focus for this pipeline was on the only freshwater and invasive species within the order Bangiales, *Bangia atropurpurea*. I validated the use of this new method to obtain an organelle genome from an organism that is difficult to grow or to grow enough in abundance for other types of genome sequencing. Because the sequencing was based on an environmental sample, the DNA sequences from other organisms, including archaea, bacteria and other eukaryotes, were also generated. The identification of the *B. atropurpurea* contigs was possible due to the large conservative blocks present in other Rhodophyta species, that acted as reference genomes. These blocks also worked as references for the alignment and genome organization. The excellent genome coverage obtained in this process allowed further investigation into gene identification in the chloroplast of *B. atropurpurea*. However, novel and unique genes in the chloroplast genome were not detected, and this may be associated with the methodology used, establishing the alignment, annotation and gene identification of *B. atropurpurea* exclusively based on the available data, leaving novel genes yet to be discovered.

Chloroplast genomes in Rhodophyta

The gene content of chloroplasts in Rhodophyta can be grouped by function or synthases into Photosystem I, Photosystem II, the cytochrome b6f complex and ATP synthase (Appendix B). The available chloroplast genomes of Rhodophyta species (Table 2.1) are represented by a circular DNA molecule with a size varying from 149 to 217 kb, with GC% from 28 to 38%, rRNA genes from 2 to 6 and 22–37 tRNA genes. In addition, the coding genes that are responsible for the translation of all amino acids necessary to fulfill the chloroplast functions in Rhodophyta can vary from 209 to 250. The genome of *B. atropurpurea* follows in between the same values. Agreeing with previous literature (Kessler et al., 1992; Green, 2011), all the Rhodophyta species, including *B. atropurpurea*, also have four subunits of a prokaryotic-type RNA polymerase (*rpo*A, B, C1 and C2). Although chloroplast genomes in Rhodophyta are known for having large conserved blocks among species (Ilha et al., 2018; Yesson et al., 2020),

with gene content and genome organization well understood, some red algal species have specific genes that vary even within the same taxonomic family or genus (Martin et al., 1998). This variance is observed mainly in the ORFs and y*cf*'s genes (Appendix A), where the *ycf* gene family encodes conserved protein sequences that are still identified as uncharacterized protein, therefore, lacking in-depth information for most of them, especially regarding their role. Because of this lack of information, it is difficult to conclude how these differences within the genomes can alter the organelle functionality and evolution of each species or family. The habitat that an alga occupies is one explanation for the variance in the chloroplast gene content since the differences in the environment is strong evolutionary pressure for all algal organisms (Glöckner et al., 2000; Cho et al., 2018). The rates of substitution of chloroplast genes taking into consideration the habitat of the algal species will be addressed in Chapter 3 of this thesis.

Chloroplast genomes in Bangiales

All Bangiales species, including the new *B. atropurpurea* data presents the same genome organization and gene content: 8 ATPase proteins, 11 for photosystem I, 18 for the photosystem II, 6 photosynthetic electron transport proteins - Cytochrome b6f complex, both ribulose-l,5-bisphosphate carboxylase (Rubisco) subunits and 5 phycobilisome polypeptides (*apc* gene family) (Table 2.2). Their genome size varies from 186 to 195 kbp, with an average GC content of 33%, and an average of 215 protein-coding genes. The number of rRNAs varies from 2 to 6 and tRNAs from 34 to 37. The similarities of the genomes are also shown in the synteny plot (Fig. 2.5), where only minor or no rearrangements are present within the Bangiales species. However, to understand how the chloroplasts in Bangiales species evolved and its mechanisms to strive in both marine and freshwater environments, the differences within the genomes must be considered.

When Bangiales genomes are compared on their gene content basis, the variance between the species becomes more evident. Like in the Rhodophyta comparison, the *ycf* gene family is inconsistent throughout the Bangiales species. The gene *ycf*27, for example, is absent in *B. atropurpurea, P. umbilicalis, B. fuscopurpurea, Py. yezoensis, Py. perforata, Py. endiviifolia, Py. fucicola and W. schizophylla.* This gene is homologous to an OmpR-like gene (Kessler et al., 1992; Martin et al., 1998; Ashby et al., 2002; Puthiyaveetil and Allen, 2009), which has a prokaryotic origin (Ashby and Mullineaux, 1999; Puthiyaveetil and Allen, 2009) and is known to

be responsible for a two-component sensory transduction pathway that regulates the coupling of phycobilisomes to the Photosystem I (PSI) and Photosystem II (PSII) (Hakenbeck and Stock, 1996; Ashby and Mullineaux, 1999; Martin et al., 1998; Ashby et al, 2002). The homologous chloroplast gene was first described in the unicellular red algae species *Porphyridium aerugineum* and *Cyanidium caldarium* (Kessler et al., 1992), but is also known to be present in several different taxa, such as in haptophyte, cryptophyte, glaucophyte, charophyte and cyanophyte (Duplessis et al. 2007). A study with cyanobacteria species that analyzed gene loss in chloroplast sequences, has shown that the *ycf*27 gene was lost in chloroplast genomes on two occasions throughout the evolution of the chloroplast, possibly being transferred to the nucleus (Martin, 1998). Curiously, a second chloroplast response regulator was identified as a by-product of chloroplast genome sequencing, the *ycf*29 gene (Maris et al., 2002; Puthiyaveetil and Allen, 2009). The *ycf*29 gene is phylogenetically distributed among the non-green algae (Duplessis et al., 2007), and is present in *B. atropurpurea* and all other Bangiales, with exception of *Pyropia yezoensis* and *Porphyridium purpureum*. It has been hypothesized that the loss of the *ycf*27 and *ycf*29 genes are correlated with the loss of phycobilisomes (Ashby et al., 2002), however, the same genes were found in genera that never had phycobilisomes, therefore, the genes are not linked exclusively to the phycobilisomes regulation (Duplessis et al., 2007, Puthiyaveetil and Allen, 2009). The presence of the *ycf*29 gene in *B. atropurpurea* may suggest that the regulatory function of the *ycf27* gene is being done by the *ycf*29 gene, but there are still no answers on how *Py. yezoensis*, species without both copies, does its regulatory function, and if there are other targets for these genes in the chloroplast.

Another gene that is missing in *B. atropurpurea, P. umbilicalis, P. purpurea, Py. haitanensis, Py. yezoensis and Por. purpureum* is the 2-Cys peroxiredoxins gene (*bas*1). This gene is known for its role in cell protection against oxidative stress, acting in the developing shoot and photosynthesizing leaves of plants (Broin et al., 2002). Other studies suggest that the chloroplast *bas*1 gene is associated with a*hp*C gene (alkyl hydroperoxide reductase), a homodimer related to the stromal surface of the thylakoids (Baier and Dietz, 1999). Both genes act to protect the photosynthetic apparatus against oxidative damage during leaf development (Baier and Dietz, 1999) and are an integral part of the organellar antioxidant network (Baier et al., 2000). Regardless of the importance of thioredoxins proteins in oxidative stress responses in

plants, their role is largely unexplored (Broin et al., 2002) and there is no literature connecting this protein with any function in Rhodophyta or Bangiales.

In addition to the *ycf*27 and *bas*1 genes, the *psb*Y gene is absent in the chloroplast genomes of *B. atropurpurea, P. umbilicalis, P. purpurea, Pyropia. haitanensis and Pyropia yezoensis*. The *psb*Y gene is part of the PSII (Debus, 1992; Gau et al., 1995; Ke 2001; Neufeld et al., 2004; Biswas and Eaton, 2018), along with several genes on the same family. In cyanobacteria species there are 26 *psb* genes (*psb*A – Z) (Neufeld et al., 2004), while in Bangiales species there are 18 commonly shared *psb* genes (present study – Appendix A) with only *psb*Y gene is missing from the above-mentioned species. A study focused on the *psb*Y gene observed that the protein is present in cyanobacteria has the role of preventing photodamage to PSII (Biswas and Eaton, 2018), while in red algae species, there is no defined role, but the gene responsible for the protein translation was the *ycf32* gene (Gau et al., 1995; Gau et al., 1998). Curiously, the *ycf*32 gene, homologous to the *psb*Y gene, is encoded in the chloroplast by *B. atropurpurea, P. umbilicalis, Py. haitanensis* and *P. purpurea.* Since the *psbY* gene plays an important role in optimizing the process of water oxidation (Meetam et al., 1999; Neufeld et al., 2004) it is peculiar that *Py. yezoensis* does not present either *psbY* or *ycf32* genes. Because the *psbY* gene is known to be nuclear-encoded in Viridaeplantae (Plöchinger et al., 2016), it is possible that a gene duplication event may have occurred and the gene was translocated from the organelle to the nucleus (Gau et al., 1998; Thompson et al., 1999). Even though the role of the PSII in the photosynthetic process is well understood, the individual role of many low molecular weight proteins that form the PSII remains elusive, with research being almost exclusive to cyanobacteria (Kashino et al., 2002; Neufeld et al., 2004; Schwenkert et al., 2007).

Lastly, a gene not observed in the *B. atropurpurea* chloroplast genome but present in all the other Bangiales species was the *pet*N gene. This gene is located in the PSI and encodes the Cytochrome b_6f (Cyt b_6f) complex subunit 8 protein (Kurisu et al., 2003; Schwenkert et al., 2007; Mohanta et al., 2019). Along with other subunits of part of the Cyt b_6f *complex (petG and petL),* the *pet*N gene is the smallest in the plastid genome and encodes a low molecular weight protein *(*Schwenkert et al., 2007). The protein complex acts within the thylakoid membranes, more specifically, providing the electronic connection between PSI and PSII in oxygenic photosynthesis (Kurisu et al., 2003; Allen, 2004; Oliver et al., 2010). The research done exclusively around the *pet*N gene functions is with Viridiplantae species, where it shows to have

a crucial role in the assembly and stability of the cytochrome complex (Schwenkert et al., 2007; Oliver et al., 2010; Mohanta et al., 2019) and some cases the gene has been transferred to the nucleus (Park et al., 2018). In cyanobacteria, the *pet*N gene is present in all studied species (Oliver et al., 2010) and in micro and macro green algae there is a gene loss in several species (Mohanta et al., 2019). Due to the high importance role of the *pet*N gene concerning desiccation tolerance (Oliver et al., 2010) and with Cyt b_6f complex, it is likely that the gene is present in *B. atropurpurea* chloroplast genome, but it was not covered by the sequencing, possibly due to its small size. To verify the possibility of translocation to the nuclear genome, a BLAST search using the *pet*N gene nucleotide sequences was performed against *B. atropurpurea* complete data, but with no matching results. Further assembly will be needed to determine the presence or detect the gene loss of the *pet*N gene in *B. atropurpurea*.

Bangiales taxonomy

The class Bangiales has had many changes in taxonomy throughout the years due to the morphological plasticity of species and the lack of molecular data, which makes it difficult to understand the relationship and evolution of the group. As shown in Fig. 5 there is significant conservation of gene order and large-scale synteny between all the Bangiales chloroplasts genomes. This provides additional support for the high quality of our draft genome and assembly method. Despite strong synteny overall, it is observed that the *B. atropurpurea* chloroplast exhibits stronger similarity to the chloroplast genome from *P. purpurea* and *P. umbilicalis* than from the same genus species *B. fuscopurpurea*. In this study, the phylogenetic analyses were done based on the concatenate ortholog blocks produced in the synteny plot, produced with whole chloroplast genomes data. The use of ortholog blocks for the phylogeny analyzes was a great alternative since the *B. atropurpurea* chloroplast genome still contained small non-sequenced gaps, and the blocks optimize the array of the analysis. Due to the amount of phylogenetic information present in the whole-genome alignment, Maximum Likelihood analysis was capable of resolving the phylogeny with significant confidence (i.e., most nodes achieved 100% bootstrap support). The placement of organisms in the phylogeny (Fig. 6) is consistent with previous literature. The chloroplast genome from *B. atropurpurea* grouped with a clade including *Porphyra* species and *B. fuscopurpurea*. According to the tree, the chloroplast genome from *B. atropurpurea* shares a more recent common ancestry with the chloroplast genome from

P. purpurea and *P. umbilicalis* than to *B. fuscopurpurea*. This is consistent with the patterns of synteny observed in Fig. 5 and supports earlier work done by Müller et al. (1998, 2001a, 2001b) and Sutherland et al. (2011), studies based on phylogenetic analyses of red algal molecular markers (nuclear SSU rRNA and *rbc*L genes).

It is noteworthy that the order Bangiales have a heteromorphic life history, alternating macroscopic gametophyte and microscopic sporophyte (conchocelis stage) (Nelson et al., 1999; Blouin et al., 2011); in addition, there are significant differences in the morphology of mature thallus, where *B. atropurpurea* is a distally multiseriate filament, while the related *Porphyra* species have a sheet-like thallus (Sheath and Cole, 1984; Lindstrom and Cole, 1992; Stiller and Waaland., 1993; Brodie et al., 1996; Woolcott et al., 1998; Oliveira and Bhattacharya, 2000; Müller et al., 2003; Broom et al., 2004; Brawley et al., 2017). A summary of sheet-like and filamentous thallus within Bangiales species is present in Fig 6. Moreover, the species *B. maxima* differs from *B. atropurpurea* and *Porphyra* by its long filaments and a multiseriate tube-like growth form (Lynch et al., 2008), which displays a morphology that appears to be intermediate between filamentous and sheet-like thallus. Future molecular analyses using *B. maxima* and comparing them to other Bangiales species is needed to shed light on the thallus morphological transitions in this order. Given that ancient related red algal species also possess a filamentous form (Butterfield et al., 1990 and 2000; Bengtson et al., 2017), this implies evolutionary transitions in morphological form (Hayden et al., 2003; Lynch et al., 2008), that do not simply mirror patterns of speciation and genome divergence.

Additional investigation of the genes that might contribute to the change in morphology of the thallus in Bangiales (Stiller and Waaland, 1993; Asamizu et al., 2003; Lynch et al., 2008) can hopefully be answered with further analyses of the nuclear genome of *B. atropurpurea*, also retrieved in this study (Poletto Borges et al. unpublished). Also of interest, is that *B. fuscopurpurea* and *Porphyra* species are both marine red algal species (Geesink, 1973; Woolcott et al., 1998; Cao et al., 2018), while *B. atropurpurea* is observed in freshwater (Lin & Blum, 1977; Jackson, 1985; Seath, 1987; Müller et al., 1998; Shea at al., 2014; Müller et al., 2003; Kucera and Saunders, 2012). This implies an evolutionary transition from an ancestral marine species related to *P. purpurea* and to the freshwater occupying lineage that gave rise to *B. atropurpurea* (den Hartog, 1972; Sheath and Cole, 1980; Müller et al. 1998, 2001b), and ultimately a freshwater invasive species of the Great Lakes (Sheath and Cole, 1984; Müller et al.

1998, 2001b; Shea at al., 2014). This raises an interesting question as to whether or not the transition to a filamentous form happened prior to or after the ancestral marine organism moved to freshwater habitat.

2.5 Conclusion

The present study suggests a new way to retrieve molecular data from difficult-to-culture organisms from metagenomic data, not only retrieve random or highly conserved genes but a complete chloroplast draft genome. The organization of *B. atropurpurea* chloroplast genome exhibits similar gene content, order and translation orientation when compared with other available chloroplast genomes from the Bangiales, and a very strong similarity with *Porphyra* species. Overall, the phylogenetic analysis in this chapter created based on all the conservative blocks of complete genomes of Bangiales species agreed with the literature (Sutherland et al. 2011; Müller et al. 1998, 2003, 2005), which was done based on marker genes. The matching results indicate that the use of traditional methods of sequencing a few conserved genes such as *rbc*L, SSU and copies of PSA complex, can generate relatable phylogeny for red algae, which can be useful for studies towards the relationship between species. However, other in-depth sequencing and analysis, like the one performed in this chapter, will help answer questions related to the evolutionary history between the species, such as shared common ancestor and gene content variation.

It is interesting to confirm the close relationship between *B. atropurpurea P. purpurea* and *P. umbilicalis*, a freshwater filamentous species and two marine with foliose thallus species. Although there is no confirmation on the morphology of the thallus of their common ancestor, it is important to remember that *Porphyra* species pass through a filamentous stage during their lifecycle, and perhaps the ancestral thallus form was also a filament that develops the foliose thallus, maybe to survive in a marine environment. Curiously, *B. atropurpurea* is the only filamentous Bangiales species that migrated to freshwater, invading the Great Lakes and striving through both harsh winter and summer. The metagenomic data obtained in this study provided substantial data for further research to characterize the remaining genomes (e.g., nuclear and mitochondrial) of *B. atropurpurea* which can reveal specific gene gains/losses, substitutions, or

genomic rearrangements that coincide with the changes in morphology. Furthermore, additional analysis of the macroalgal microbiome content of the *B. atropurpurea* metagenome will be important to uncover additional factors responsible for its adaptation to a new ecological niche (Egan et al., 2013; Hollants et al., 2013; Singh & Reddy, 2014; Saha et al., 2016; Brawley et al., 2017; Ihua et al., 2019). These data may reveal potential symbiotic relations and uncover a broader pan-genome for *B. atropurpurea* that better reflects its metabolic and functional dependencies, capabilities and its adaptation to freshwater habitats.

Chapter 3

Analysis of substitution rates of the chloroplast genes in Rhodophyta species

3.1 Introduction

To examine the possible selective pressures acting at the protein of species, tests of rates of substitutions among genes or genomes are applied. Several tests are available to date, but the most common test consists of the analysis of the synonymous (K_s) and non-synonymous substitutions (K_a), or K_s and K_a mutations, and their ratios ($\omega = K_a / K_s$) (Kimura, 1983; Ohta, 1993; Yang & Bielawski, 2000; De la Fuente et al., 2003; Yang et al., 2005; Han et al., 2021). The K_s substitutions are generally considered silent since they are usually in the third position of the codon where substitutions do not cause changes in the amino acid sequences, hence not acted by natural selection (Akashi, 1995; Han et al., 2021). K_a on the other hand, often occur in the first or second position of the codon, altering the amino acid and, consequently, the protein product (Yang et al., 2000; Yang et al., 2005; Doron-Faigenboim & Pupko, 2007). Traditionally, the model used to study rates of substitution follows the Miyata & Yasunaga (1980) methodology, where ω lower than 1 (< 1) indicates purifying selection, ω equal to 1 indicates neutral selection, and ω higher than 1 (> 1), indicates positive selection. Although some authors have expressed concern about the number of false-purifying results that can be generated by the Miyata & Yasunaga method, and suggested different methods to double-check the accuracy of the results precise (Yang et al., 2000; Yang & Nielsen, 2002; Filip & Mundy, 2004; Moury, 2004; Wong et al., 2004), this method is a valid test to perform for an overview of the possible rates of selection.

The Rhodophyta are an ancient lineage, being one of the primary plastids-bearing hosts, and the origin of plastids for the secondary and tertiary endosymbiosis (Yoon et al., 2002; Nan et al., 2017; Chapter 2 of this Thesis). Species within the Rhodophyta are primarily marine, and only a small proportion, around 3%, is observed in freshwater (Sheath, 1984; Wehr et al., 2015; Nan et al., 2017; Fangru et al., 2020; Han et al., 2021). The probable origin of freshwater red algae is still debated, with little molecular evidence to determine if the species are migrants from the sea or have arisen from original inhabitants of inland waters (Skuja, 1938; Nan et al., 2017). Freshwater red algae species are an important constituent of stream floras, usually striving in clean water with a lower temperature than those in the ocean, they are benthic and macroscopic, but smaller than marine species (Sheath & Hambrook, 1990; Kumano, 2002; Han et al., 2019). Red algae species living in different environments are under different stress factors, both biotic

and abiotic. The abiotic factors can influence the species is genome-wide rate of molecular evolution (Mitterboeck et al., 2016). One example of an abiotic factor is light exposure, which includes changes in quality, intensity and photoperiod, being a key factor affecting the distribution and successful survival of the species (Sheath & Hambrook, 1990). This correlation happens because illumination affects algal growth via photosynthesis (Sheath & Vis, 2015). Therefore, it is expected that gene groups related to the photosynthetic activity will have differences at a molecular level that corresponds to the habitat where the species is observed, and the stresses that are upon the organisms will also reflect in differences in the rates of substitution.

The species *Bangia atropurpurea* is the only freshwater species of the Bangiales (Rhodophyta) that invaded the Laurentian Great Lakes and is likely derived from a freshwater species in Europe (Müller et al., 1998; Shea et al., 2014). The unexpected appearance of *B. atropurpurea* in the Great Lakes is probably due to vector-assisted transport from Europe, indicating a single invasion of the Great Lakes (Sheath, 1984; Müller et al., 1998). This hypothesis is supported by the disjunctive distribution of Bangia in the Great Lakes and molecular data (Müller et al., 1998). Phylogenetic studies confirm a closer relationship to the marine *Porphyra* species than to the same genus species *B. fuscopurpurea*, which indicates a marine common ancestor (Chapter 2 of this thesis). Interestingly, however, *B. atropurpurea* and *B. fuscopurpurea* have the same filamentous thallus morphology. The *Porphyra* species have foliose thallus but pass through a filamentous phase during the early stage of their life cycle. In addition, *B. atropurpurea* is known for being an asexual species, while all the other species of the Bangiales are known to reproduce sexually. These characteristics make *B. atropurpurea* an excellent model for rates of substitution studies focusing on habitat adaptation and resproductive strategies.

Studies about the adaptative evolution of the chloroplast genes among Rhodophyta that inhabit different environments are non-existent. The available literature that mentions gene evolution typically focuses on a few chloroplast genes, mostly the ones used as gene markers in phylogenetic studies (Moreno & Spreitzer, 1999; Mitterboeck et al., 2016; Xu et al., 2018; Han et al., 2021), excluding all the other genes that are part of the chloroplast genome resulting in a large gap of knowledge about these genes. Hence, examining the rates of substitution of the chloroplast genomes of red algae, making use of a large dataset of chloroplast genes is needed. With the chloroplast data of the only freshwater species within the Bangiales (*B. atropurpurea*,

Chapter 2 of this thesis) the analysis will provide an opportunity to examine the potential impact of habitat on the evolution of photosynthetic genes within Rhodophyta orders. Since little has been done to examine the selection pressures on chloroplast genes within the red algae, this is also an opportunity to examine the differences among various taxonomic lineages.

Chloroplast genomes of red algae have high gene capacity and compact structure (Yoon et al., 2006; Janouškovec et al., 2013; Wang et al., 2013; Brawley et al., 2017; Nan et al., 2017), being a tool to resolve phylogenetic relationships in deep terminal branches, reflecting the evolutionary history among Rhodophyta species (Tajima et al., 2014; Lee et al., 2016). Moreover, the chloroplast genomes of red algae can be useful for genome evolution studies due to the low variability in gene content and high similarity in gene arrangement (Costa et al., 2016). The present chapter aims to investigate how the chloroplast genes of Rhodophyta species vary regarding rates of substitution, and to elucidate these differences among the species based on their habitat and taxonomy. The substitution rates of 71 chloroplast genes of 30 Rhodophyta species are organized and compared not only by their taxonomic position but also by the habitat where the species are usually found. Due to the lack of research on selection pressures among Rhodophyta species, our results will help the understanding of the chloroplast evolution process of red algae, possible molecular pathways used for adaptation to different environmental pressures, and the relationship between nucleotides and the molecular mechanism of evolution.

3.2 Material and Methods

3.2.1 Sequence data

The sequence of the chloroplast genome of *Bangia atropurpurea* was retrieved as mentioned in Chapter 2 and analyzed amongst 30 other Rhodophyta species with complete chloroplast genomes available on NCBI, and one Chlorophyta species, *Bryopsis plumosa* (Table 1). Once all the chloroplast genomes were retrieved, all coding genes were plotted in a table (Appendix D1, E1, F1, G1 and H1) to verify the presence or absence of each gene in every genome. Traditionally, the chloroplast genes are separated into 12 groups according to their function (Chapter 2), and this study focused on eight of these groups: ATP synthase (8 genes), Photosystem I and II (11 and 17 genes), Ribosomal proteins LSU and SSU (27 and 19 genes), RNA polymerase (4 genes) and RubisCO (Du et al., 2016). The nucleotide sequences of each gene were retrieved manually and combined into fasta files for alignment, performed using ClustalW (Thompson et al., 1994) within the Geneious Prime software (www.geneious.com). Gene sequences from *B. atropurpurea* that were considerably smaller than the sequences from other Rhodophyta species were not selected to avoid false results and/or extreme values. The aligned sequences were then analyzed for substitution rates using DnaSP 6 (Rozas et al., 2017), using the chloroplast data format option and assigned coding regions for the 1st, 2nd and 3rd positions. To estimate the balance between neutral mutations, positive or purifying selection acting on the chloroplasts genes, the values of the nonsynonymous substitutions per non-synonymous site (K_a) and the synonymous substitutions per synonymous site (K_s) were calculated and its ratio (ω) compared.

3.2.2 Analysis with DnaSP

Once the DNA sequences were aligned, they were imported to DnaSP software and analyzed with the command Substitution Rate Analysis. This command estimates K_a and K_s for any pair of sequences, using the Nei and Gojobori (1986, equations 1-3) branch methodology. The software can also estimate the substitution rates on noncoding protein regions, but this feature was not used in our final analysis. For computing K_a and K_s sites, DnaSP excludes all

pathways that go through stop codons. Although DnaSP allows the user to choose the input data format for chloroplast data, there are no pre-defined genetic codes specific for chloroplasts, and the universal nuclear code was chosen. Sites (or codons) with alignment gaps or missing data are not used, being completely excluded. The substitution analysis is done by calculating the average number of nucleotide differences per site between two or more sequences, or nucleotide diversity (Nei 1987, equations 10.5 or 10.6), with additional correction using the Jukes and Cantor (1969) (Lynch and Crease 1990, equations 1- 2). The correction is performed in each pairwise comparison of two or more sequences (Nei and Gojobori 1986, equations 1-3).

It is important to note that the software calculates the average of differences, or mutations, in the whole sequence, not providing an in-depth analysis of the number of substitutions per site. Therefore, the result of the ω ratio (K_a/K_s) is an average of all the substitutions present in the gene, which can be positive, purifying, and neutral at the same time. Since this is the first research work looking into rates of substitution of a significant number of chloroplast genes (71 in total) in Rhodophyta species, the analysis with DnaSP is a good option because it provides an overview of the mutation rates. The results obtained here are the grounds for future studies, pointing out genes that can be evolutionary interesting for in-depth analysis and to verify the presence of selection forces. For that, other software that can analyze the substitutions per site, methodologies that compare the specific sites among sequences, and different tests for verifying false positives, such as PAML, are possibly more appropriate (Yang, 2007).

The well-studied species *Cyanidiozchyzon merolae* was selected as the reference genome, allowing the comparison of *B. atropurpurea* in the context of its own group. Therefore, all genes of all species were first compared with *C. merolae*, and all the rates of substitution values are a reflection of this comparison. The species *C. merolae* belongs to the Cyanidiales order, a paraphyletic group to the Bangiales and is found in hot springs.

3.2.3 Data comparison

To understand if selective pressures are acting on the chloroplast genes of *B. atropurpurea* and other red algae species, the chloroplast genes of every species were first analyzed for high ($\omega > 1.25$), neutral (ω from 0.75 to 1.25) or low ω ratios ($\omega < 0.75$).

Afterwards, the same ω ratio dataset was arranged in two different ways for comparison. First, the species were arranged according to the habitats where they are usually found, and the average ω ratio of each habitat was compared. For this chapter, 4 different habitats were selected: freshwater, salt marshes, hot springs and marine. This comparison can provide insights into the impact that different habitat constraints have on chloroplast genes by pointing to the group/habitat with more differences/mutations in the genes. Secondly, the species were rearranged according to the phylogenetic position (taxonomy) they are currently classified, and the same ω ratio dataset from DnaSP was used to compare the average rate of substitution on each taxonomic class or order. In this study, 6 classes and 12 orders are compared: Bangiophyceae (Bangiales), Florideophyceae (Thoreales, Batrachospermales, Palmariales, Gigartinales, Caramiales, Gracilariales, Hildenbrandiales), Porphyridiaceae (Porphyridiales), Compsopogonophyceae (Compsopogonales), Cyanidiophyceae (Cyanidiales) and Ulvophyceae (Bryopsidales). To confirm the phylogenetic position of the species, reference phylogenetic trees were used (Yoon et al., 2006; Sutherland et al., 2011) and a new phylogenetic tree was built based on the *rbc*L gene sequences from the species used in this study (Fig. 3.1). The evolutionary analyses of the *rbc*L gene were conducted in MEGA X (Kumar et al., 2018) and the evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible + G +I as the evolutionary model (Posada & Crandall, 1998; Nei & Kumar, 2000). The bootstrap consensus inferred was 1000 replicates (Felsenstein, 1985). By comparing the average substitution ω ratio among different phylogenetic groups it is possible to detect and observe trends of selection forces acting on chloroplast genes of specific taxonomic groups.

Because the *rbc*L gene is known to be a conserved gene among Rhodophyta and used extensively for phylogenetic studies, the other gene families present in this study most likely will not exhibit the same phylogenetic tree patterns. To observe the phylogeny of each gene family, the complete gene sequences that were retrieved manually were concatenated and aligned using ClustalW (Thompson et al. 1994) within the Geneious Prime software (www.geneious.com). The aligned sequences were then analyzed with Seaview (version 4, Gouy et al., 2010) and gblocks of conserved areas were selected. The gblocks were used to build phylogenetic trees of each gene family on MEGA X (Kumar et al., 2018). The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible + G +I as the evolutionary model (Posada & Crandall, 1998; Nei & Kumar, 2000). The bootstrap consensus inferred was 1000

replicates (Felsenstein, 1985). The trees resulting from this analysis are in the Appendix section of this chapter (Appendix Figures D3, E5 and F5).

3.2.3 Substitution rates interpretation

To determine possible purifying, neutral or positive selection forces acting on the selected genes, the ω ratios calculated using the DnaSP were plotted in Microsoft Excel for statistical analysis. The traditional interpretation of the ratio states that a ω ratio equal to or around 1 represents neutral selection, $\omega < 1$ represents purifying selection and, $\omega > 1$ indicates positive selection (Yang, 2007). However, the ratio results are in most cases very subjective and can greatly vary from 0 up, with values around 1 in many cases, leading to questions on how strict the value of 1 should be taken into consideration to validate an evolutionary selection type. Because of the ratio variation, this study will follow a couple of more recent publications that better interpret the ratio values and specify a cut-off value for ω ratios (Wang et al., 2011; Hussain & Rasool, 2017), where ω with values from 0.75 to 1.25 as neutral selection; $\omega >$ 1.25 as positive selection and $\omega < 0.75$ as purifying selection.

3.2.3 Statistical analyzes and data visualization

Once the ω ratios were calculated, the rates of substitution were tested with the Kolmogorov-Smirnov test (KS test) for a normal distribution. If the data followed a normal distribution, the value of the KS would be 0 and a parametric test would be performed. A p-value threshold of 0.05 was used to determine whether the difference is significant to reject the null hypothesis (H_0), which assumes a normal distribution. If the p-value was > 0.05, the distribution is not normal (H_1) (Massey, 1951). Because most of the results obtained by the normality test rejected the null hypothesis, a non-parametric test, the Kruskal-Wallis test (h-test), was chosen to verify significant differences within the ω ratios (Theodorsson-Norheim, 1986) for both phylogenetic position and habitat comparisons.

For the h-test, the H_0 is that the rates of substitution will not have statistical differences among the species. The data analysis will be verified by an h-value that is lower than the critical value, or a p-value < 0.05, indicating significant differences among the compared groups. On the

other hand, if the h-value is higher than the critical value, or the p-value > 0.05, there is not support for significant differences among the compared groups (H_1). As a non-parametric test, it is worth noting that the h-test performs a statistical comparison of ranks instead of comparing the actual values of the observations. Given the nature of the h-test analysis, it is important to note that if the rates of substitution agree with H_1, the different groups will be flagged as having differences, but not pointing out specifically which species are different. The observations that point out which group is different from the other were possible due to the data arrangement into the habitats or phylogenetic groups, but no additional statistical analysis was performed to detect the significance of these differences. Following most of the rates of substitution studies previously published the p-value of 0.05 is satisfactory for an overall perspective of possible selection pressures acting on a gene. The complete list of the h-test and p-values obtained in this analysis are in Appendix J.

To minimize false-positive selection results a Bonferroni correction was initially performed; however, the correction test was excessively conservative and, the accuracy appeared to be altered when the different habitats were compared. The Bonferroni correction is usually applied in experiments that compare different groups at baseline and study the relationship between variables (Streiner & Norman, 2011; Dmitrienko & D'Agostino, 2013). However, despite the test's popularity, the routine use of this test has been criticized as deleterious to sound statistical judgment, testing the wrong hypothesis (Armstrong, 2014). A similar result for the Bonferroni correction was obtained by Hartmann & Golding (1998), and other authors share the same opinion regarding the extreme conservative results from this test (Narum, 2006; Streiner & Norman, 2011). Due to these undesirable properties, the Bonferroni method of correction was not performed further. Graphs reflecting the ω ratios of each chloroplast gene for both habitat and phylogenetic position arrangements were generated on Excel and edited on Adobe Illustrator.

Table 3.1. Chloroplast genomes of Rhodophyta and Chlorophyta species used in this study.

Species	Class	Order	Morphology	Genome size (bp)	GC%	Coding genes	Accession number	References
Bangia atropurpurea	Bangiophyceae	Bangiales	Filament	186.84	32.9	207	MN484624	This study
Porphyra umbilicalis	Bangiophyceae	Bangiales	Foliose	190.173	32.9	207	NC_035573	Brawley et al., 2017
Porphyra purpurea	Bangiophyceae	Bangiales	Foliose	191.028	33	250	NC_000925	Reith and Munholland, 1995
Bangia fuscopurpurea	Bangiophyceae	Bangiales	Filament	196.913	33.5	205	KP714733	Cao et al., 2018
Pyropia fucicola	Bangiophyceae	Bangiales	Foliose	191.982	32.7	206	KJ776837	Hughey et al., 2014
Pyropia endiviifolia	Bangiophyceae	Bangiales	Foliose	195.784	33.3	210	KT716756	Xu et al., 2018
Pyropia haitanensis	Bangiophyceae	Bangiales	Foliose	195.597	32.98	213	NC_021189	Wang L, 2013
Pyropia kanakaensis	Bangiophyceae	Bangiales	Foliose	194.631	32	209	KJ776836	Hughey et al., 2014
Pyropia perforata	Bangiophyceae	Bangiales	Foliose	189.788	32.9	209	KF515973	Hughey et al., 2014
Pyropia yezoensis	Bangiophyceae	Bangiales	Foliose	191.975	32.09	213	KC517072	Wang L, 2013
Wildemania schizophylla	Bangiophyceae	Bangiales	Foliose	193.008	34.4	211	NC_029576	Hughey et al., 2016
Thorea hispida	Florideophyceae	Thoreales	Filament	175.193	28.3	194	NC_031171.1	Cho et al., 2018
Palmaria palmata	Florideophyceae	Palmariales	Flat fronds	192.961	33.9	203	NC_031147	Cho et al., 2018
Chondrus crispus	Florideophyceae	Gigartinales	Flat fronds	180.09	28.7	204	NC_020795	Collen et al., 2013
Caloglossa beccarii	Florideophyceae	Ceramiales	Foliose	165.038	26.9	201	NC_035269.1	Díaz-Tapia et al., 2015
Caloglossa intermedia	Florideophyceae	Ceramiales	strap-like	166.397	31	209	NC_035265.1	Díaz-Tapia et al., 2015
Kumanoa americana	Florideophyceae	Batrachospermales	Pedicilate	184.026	29.3	201	NC_031178.1	Cho et al., 2018
Sheathia arcuata	Florideophyceae	Batrachospermales	*Branched*	187.354	29.9	187	NC_035231.1	Paiano et al., 2018
Hildenbrandia rubra	Florideophyceae	Hildenbrandiales	Crust	180.141	31.4	191	NC_031146	Lee et al., 2016
Hildenbrandia rivularis	Florideophyceae	Hildenbrandiales	Crust	189.725	32.4	186	NC_031177	Lee et al., 2016
Gracilaria changii	Florideophyceae	Gracilariales	Branched	183.555	28.1	201	NC_038051	Ho et al., 2018
Gracilaria chorda	Florideophyceae	Gracilariales	Branched	182.459	27.4	203	NC_031149	Lee et al., 2016
Gracilaria edulis	Florideophyceae	Gracilariales	Branched	179.410	30	201	NC_046041	Liu et al., 2019
Gracilaria ferox	Florideophyceae	Gracilariales	Branched	180.255	27.4	204	NC_039140	Ilha et al., 2018
Gracilaria salicornia	Florideophyceae	Gracilariales	Solid cylindrical	179.757	28.8	206	NC_023785	Campbell et al., 2014
Porphyridium purpureum	Porphyridiaceae	Porphyridiales	Unicellular	217.69	30.3	224	NC_023133.1	Tajima et al., 2014
Compsopogon caeruleus	Compsopogonaceae	Compsopogonales	Filament	221.013	29.9	195	NC_035350.1	Nan et al., 2017
Cyanidioschyzon merolae	Cyanidiophyceae	Cyanidiales	Unicellular	149.987	37.6	243	AB002583	Otha et al., 2003
Cyanidium caldarium	Cyanidiophyceae	Cyanidiales	Unicellular	164.921	32.7	197	NC_001840.1	Glöckner et al., 2000
Galdieria sulphuraria	Cyanidiophyceae	Cyanidiales	Unicellular	167.741	28.5	182	NC_024665.1	Jain et al., 2015
Bryopsis plumosa	Bryopsidaceae	Bryopsidales	Sinophonus	106.859	30.8	115	NC_026795.1	Leliaert & Lopez-Bautista, 2015

The highlighted species are found in freshwater.

Table 3.2. Chloroplast genes encoded in Rhodophyta species that were used for the substitution rates comparison study.

Gene name	Group of Genes
*atp*A	ATP synthase
*atp*B	
*atp*D	
*atp*E	
*atp*F	
*atp*G	
*atp*H	
*atp*I	
*psa*A	Photosystem I
*psa*B	
*psa*C	
*psa*D	
*psa*E	
*psa*F	
*psa*J	
*psa*K	
*psa*L	
*psa*M	
*psa*I	
*psb*A	Photosystem II
*psb*B	
*psb*C	
*psb*D	
*psb*E	
*psb*F	
*psb*I	
*psb*J	
*psb*K	
*psb*L	
*psb*H	
*psb*N	
*psb*T	
*psb*V	
*psb*W	
*psb*X	
*psb*Y	
*psb*Z	
*rps*1	Ribosomal proteins(SSU)
*rps*10	
*rps*11	
*rps*12	
*rps*13	
*rps*14	
*rps*16	
*rps*17	
*rps*18	
*rps*19	
*rps*20	
*rps*2	
*rps*3	
*rps*4	
*rps*5	
*rps*6	
*rps*7	
*rps*8	
*rps*9	

Gene name	Group of Genes
*rpl*1	Ribosomal proteins(LSU)
*rpl*2	
*rpl*3	
*rpl*4	
*rpl*5	
*rpl*6	
*rpl*9	
*rpl*11	
*rpl*12	
*rpl*13	
*rpl*14	
*rpl*16	
*rpl*18	
*rpl*19	
*rpl*20	
*rpl*21	
*rpl*22	
*rpl*23	
*rpl*24	
*rpl*27	
*rpl*29	
*rpl*28	
*rpl*31	
*rpl*32	
*rpl*33	
*rpl*34	
*rpl*35	
*rpl*36	
*rbc*L	RubisCO

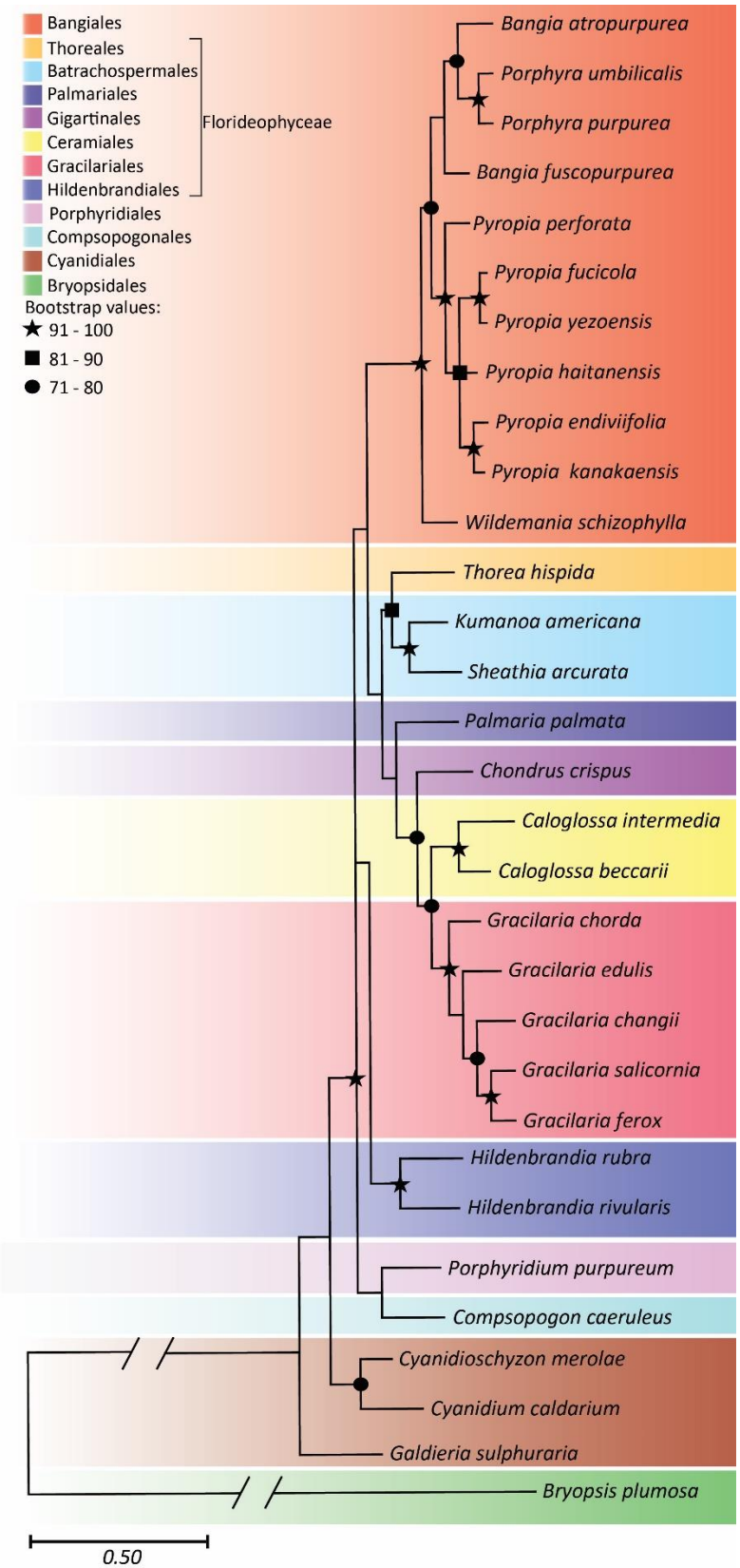

Figure 3.1. Maximum likelihood phylogenetic tree based on the *rbc*L gene sequences. All sequences, with exception of *B. atropurpurea*, were retrieved from the NCBI database.

3.3 Results and Discussion

The investigation of the rates of substitution of 71 chloroplast genes among Rhodophyta species is presented in this chapter. As a preview of the detailed data analysis of each gene, Fig. 3.2 shows a summary of the result of the average ω ratios observed in every gene and the comparison among the species habitats as well as the phylogenetic position comparison. The detailed analysis of each comparison is described next. Since the ω ratios are the same for both comparisons, the difference between the analysis is due to the rearrangement of these values. Examination of Fig. 3.2, reveals that different habitats and taxonomic relationships act on the molecular evolution of the chloroplast genes which exhibit different patterns of rates of substitution.

When the species ω ratios were averaged and grouped by habitat, the comparison shows that most genes are under similar selection pressures, but some genes with divergent ratios are also detected. In species observed in hot springs, the *atp*H gene appears to be under strong positive selection pressure which is different from the selection pressure over other species habitats. Also, the habitat comparison of the *atp* genes shows that species observed in salt marshes and freshwater are under similar selection forces for the *atp*B gene, being different from the selection forces over species observed in hot springs and the marine environment. Moreover, the comparison by habitat shows that the *psb* genes have different selection pressures over the *psb*F and *psb*L genes, with emphasis on the salt marshes species which are under strong positive selection forces. The *psa*, *rpl*, and *rps* genes are under similar selection pressure among species independently of the habitats they are found. Although the graphs seem to show that different habitat constraints have little impact on the rates of substitution overall (Fig. 3.2), the statistical analysis shows that 24 of the 71 genes analyzed exhibit significantly different patterns of selection forces among the species (p-value < 0.05, h-test).

Following the result obtained by the *rbc*L tree (Fig 3.1), the species ω ratios were also averaged and grouped by taxonomical orders. Like in the habitat comparison, the graphs show evidence that different taxonomic orders have specific selection forces patterns acting over their chloroplast genes (Fig. 3.2). All six orders exhibit a different rate of substitution for the *atp*B gene, and the green algae order, *Bryopsidale*, has the highest ratio which indicates positive selection. A similar scenario is shown on the *atp*H gene, but with the *Cyanidiales* having the

highest average rate. The Bangiales on the other hand is the order with the lowest averages and is considered under purifying selection for the *atp*H gene. The *psa* genes exhibit a very different pattern from the habitat comparison when the species are compared by their phylogenetic position. Where the rate of each order seems to follow a particular trend. The exception is the *psa*I gene, where all Rhodophyta orders are under strong positive selection. Interestingly, the average of the Bangiales species is again the lowest among the orders for the *psa* gene family. The *psb* genes are again the gene family with a more uniform average of rates of substitution. For the *psb*F, however, the Bangiales exhibit the highest rate, and along with the Florideophyceae and Cyanidales are under strong positive selection. The *psb*L and *psb*T genes exhibit high substitution rates and the Porphyridiales have the highest rate, indicating strong positive selection. The *rpl* and *rps* genes have a similar average of substitution ratios among the orders, yet, the Bangiales contrast with the other orders by having the highest (*rpl*11, *rpl*13, *rps*14) and lowest ratios (*rps*5, *rps*10, *rps*18) rates of substitution for several genes. Although these are the genes that visually have differences, the statistical analysis shows that 31 of the 71 chloroplast genes have significant differences in the rates of substitution (p-value < 0.05, h-test).

Undoubtedly, different selective forces are acting on the chloroplast genes of the Rhodophyta species based on their taxonomy, even among species within the same orders. The same is observed among different habitats, even when species share a similar environment. Nonetheless, a comprehensive analysis of the rates of substitutions of every species is needed to answer if there are exclusive trends of mutation in any of the 30 species and to determine which habitat has more impact over the molecular changes. Thus, a detailed and complete analysis will be presented next.

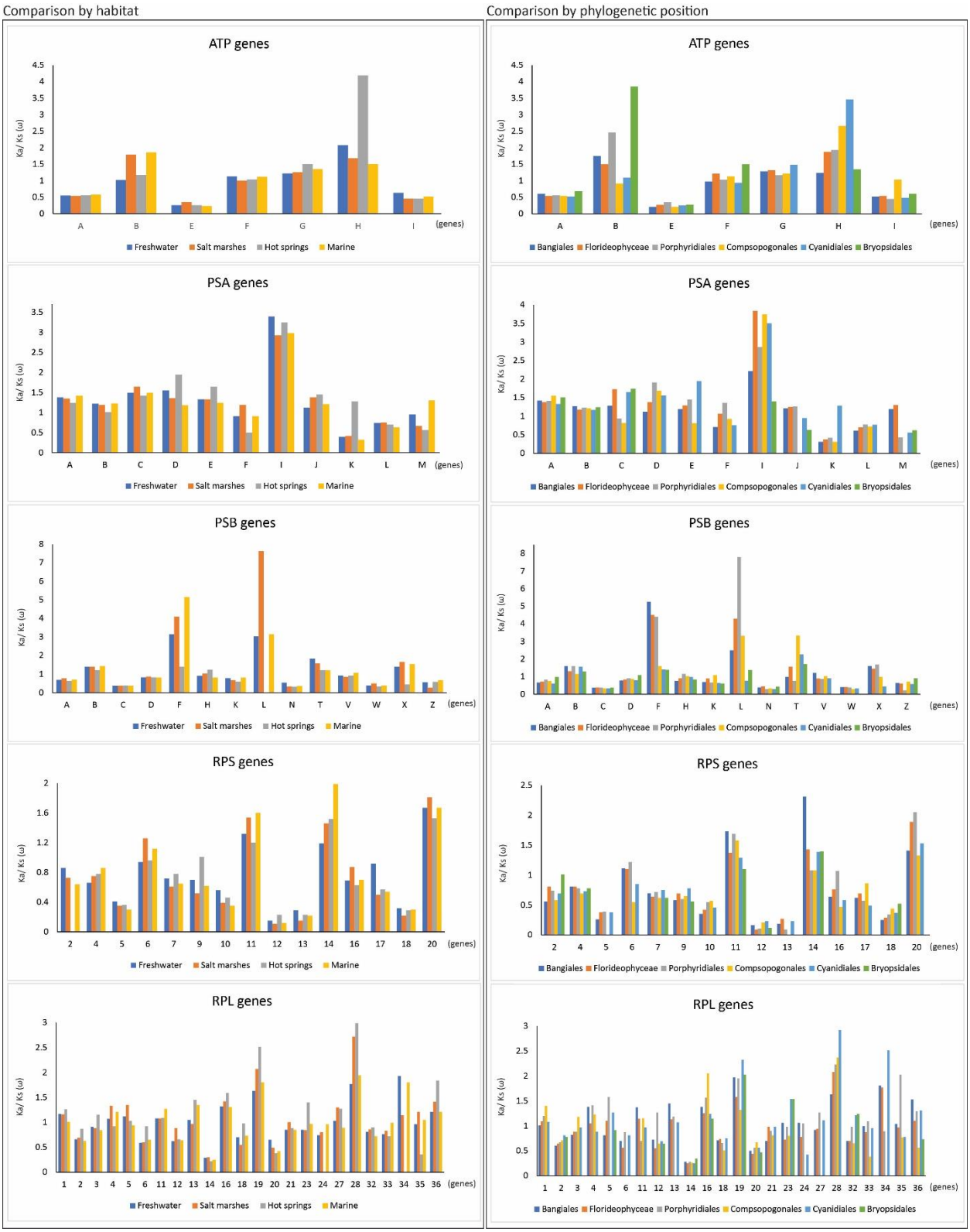

Figure 3.2. Summary of the analysis of the average ω ratios of the chloroplast genes, which were grouped according to the species habitats (graphs on the left) and according to the species phylogenetic position (graphs on the right). The X-axis values are the chloroplast genes used in the analysis. Every bar represents the average ω value that the species in that habitat or taxonomic order have.

3.3.1 Rubisco (*rbc*L) gene

The *rbc*L gene, which encodes the large subunit of ribulose- 1,5-bisphosphate carboxylase oxygenase (RubisCO), has been widely sequenced from numerous plant taxa, resulting in a large database that has been applied to several plant phylogeny studies (Palmer et al. 1988; Clegg and Zurawski 1991; Chase et al. 1993; Gielly & Taberlet, 1994). Since this gene is known for being conserved, the sequences were used in this study as a reference for the substitution rates among the species of the different phylogenetic groups. In Fig. 1, the phylogenetic tree matches the trees used as a reference, supporting the methodology chosen for data retrieval and for the analysis that was performed for all the other genes.

When observing ω ratios of the *rbc*L gene on the species individually, most are under possible positive selection (> 1.25). Only 6 of the 30 species have values that support neutral selection (0.75 to 1.25): *B. atropurpurea, S. arcuata, Chondrus crispus, Pyropia kanakaensis, Porphyra purpurea* and the green alga *Bryopsis plumosa.* Once the data was organized by the phylogenetic order based on the tree on Fig. 1, statistical support for significant differences between the phylogenetic groups was confirmed (p-value = 0.01), and the higher ratios were observed for Florideophyceae species, with emphasis on the *Gracilaria* genus (Fig. 2). The class has a ratio average of 1.43 which is indicative of positive selection. The highest ratio in the *rbc*L gene, however, belongs to the Cyanidiales species *Galdieria sulphuraria* (1.68), which is a species found in hot springs and has strong support for positive selection. The Bangiales species have an average of 1.3 and agree with the literature that states that Bangiales species have a highly conserved *rbc*L gene, therefore, with low rates of mutation (Müller et al., 1998; Sutherland et al., 2011; Xu et al., 2018). There were no ratios below the cut-off value (< 0.75), therefore, no species have the *rbc*L gene under purifying selection. When species are arranged based on habitat, the ω ratios do not have statistical support for significant differences between the groups (Appendix Figure C1).

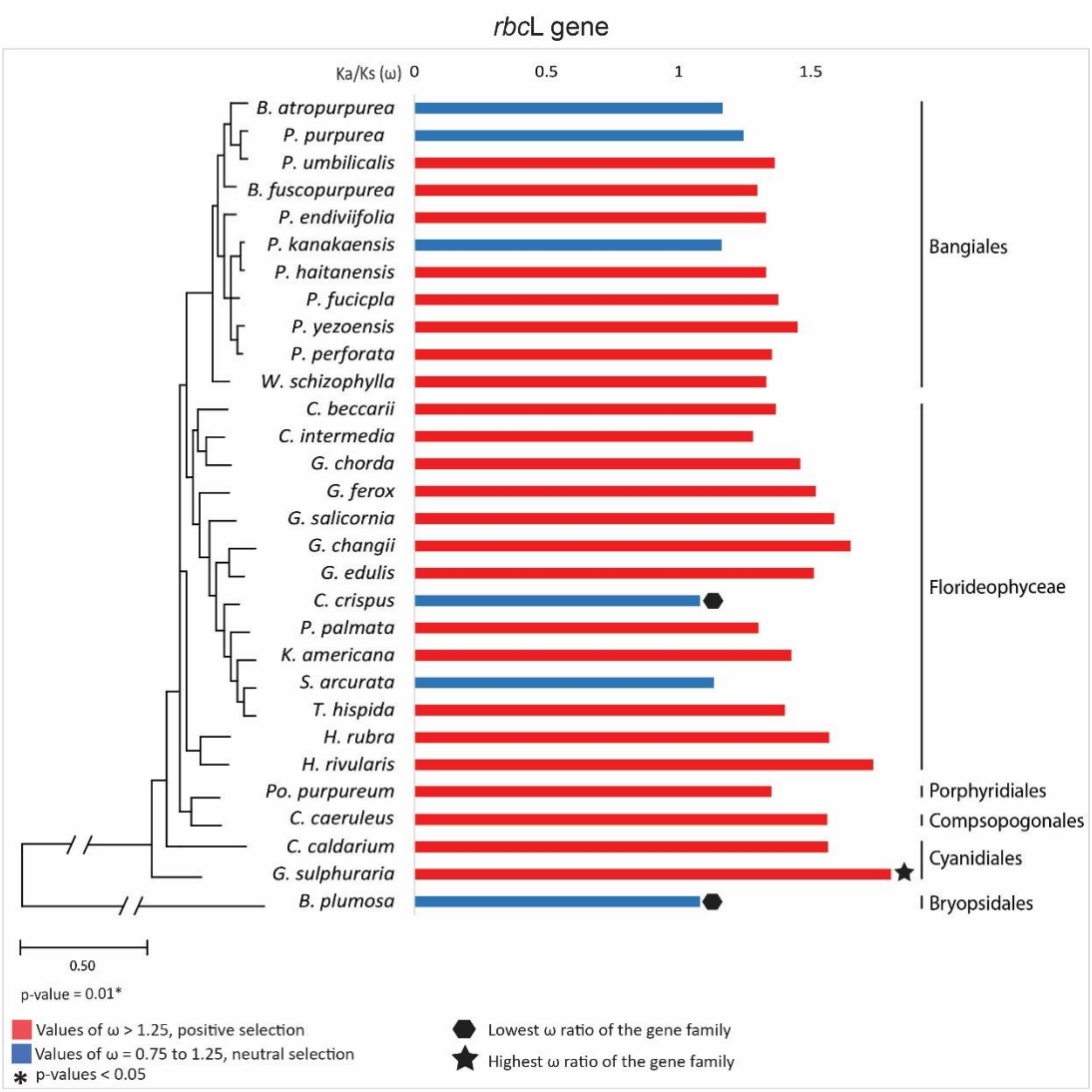

Figure 3.3. Substitution ω ratios of red and green algae in comparison to *C. merolae rbc*L gene, and then compared by phylogenetic position. The phylogenetic position of each species follows the *rbc*L gene tree.

3.3.1.1 Rubisco (*rbc*L) gene - Discussion

The *rbc*L gene encodes the Rubisco protein, which is within the chloroplast matrix of plants and is a crucial enzyme involved in plant photosynthesis (Young et al., 2012; Fangru et al., 2020; Han et al., 2021). A large amount of Rubisco protein is needed to maintain the normal metabolic process of cells, and as result, is the most abundant protein in plant cells, accounting for about 50% of the total soluble protein (Kapralov et al., 2007). Due to the critical role of the protein, it is often postulated that Rubisco is a highly conserved enzyme. Data availability about

adaptive evolution of the *rbc*L gene is mostly in Viridiplantae species, such as in fern, sweet wormwood and small trees species (Zhang et al., 2010; Liu et al. 2010; Zhou et al.,2011; Xiong et al. 2014). These studies detected multiple positive selection loci, that are essential to the functionality of the *rbc*L gene. Despite the importance of the *rbc*L, there are only a few studies on the adaptive evolution of this gene in Rhodophyta species.

Young et al. (2012) investigated the history of evolution adaptation of Rubisco protein in red and Chromista algae and observed that there was no signal of positive selection in recent lineages, and the positive selection that did occur in the *rbc*L gene happened prior to the divergence of large algal taxonomic groups, but not within the Rhodophyta and Chromista species. However, other studies that focused on red algae species disagree with the previous statement and noted adaptive positive evolution on the *rbc*L gene in species within the order Batrachospermales (Gong et al., 2017), and in *Galdieria* species (Han et al., 2021). The sites in the sequence that exhibited positive selection agree with our results, where the *G. sulphuraria* represents the highest substitution rates among all the other 29 Rhodophyta species analyzed. The high ω rates in *Galdieria* can be explained by the unique characteristics of the species, such as the extreme environments in which they are observed, typically high-temperature and high-acid hotsprings (Pinto et al., 2007; Ciniglia et al., 2014). The nucleotide substitutions related to Rubisco's function are likely subjected to selection pressures due to their direct relationship to the biological fitness of the species (Nozaki & Morita, 2002).

Another study examining the adaptive evolution of the *rbc*L gene in Rhodophyta focused on *Hildenbrandia* species (Fangru et al., 2020). In this study, positive selection sites were not detected, implying that the *rbc*L gene in *Hildenbrandia* species evolved neutrally. This genus includes the freshwater species *H. rivularis*, and the marine species *H. rubra*, and both species were part of our dataset as well. However, our study shows conflicting results, where a positive selection was detected in the *rbc*L gene of both species. One explanation for the different results is the lack of representation of other Rhodophyta species in the Fangru et al. (2020) analyses, where the evolution rates were compared only within *Hildenbrandia* species. Therefore, among *Hildenbrandia* species the *rbc*L gene can be evolving neutrally, yet, when the species rates are compared among other Rhodophyta, they are under positive selection. This emphasizes the importance of integrating as many species as possible in a comparison study and the need for more literature around the adaptive evolution in Rhodophyta.

Investigations about the connection between the *rbc*L gene evolution in response to the freshwater habitat in comparison to species found in the ocean noted that environmental stresses such as oxidative stress and osmotic stress affect the activity of Rubisco (Moreno & Spreitzer, 1999). This theory was confirmed by Mitterboeck et al. (2016), which hypothesized that freshwater species have higher evolutionary rates within protein-coding genes when compared with marine relatives. Although the theory is not based on algal species, it states that a different evolutionary rate is expected when the *rbc*L gene of species found in freshwater is compared to the ones found in the ocean.

Contradicting the above literature, other studies that exclusively analyzed red algae genes argue that freshwater species, in most cases, have no signal of adaptive evolution and/or are under strong negative selection pressure (Gong et al., 2019; Han et al., 2021). In our data 6 of the 8 freshwater species have ω that support positive selection. However, it is important to note that *B. atropurpurea* and *S. arcuata* showed no signal of positive adaptive evolution for the *rbc*L gene, instead showing ω that support neutral selection. Moreover, most of the marine species used in our study also show ω that support positive selection, having a very similar ω trend when compared with the freshwater species. The incongruence of results regarding the evolution of the *rbc*L gene can be explained by several important points, such as the sample size, sequence quality, method of analyses, and most importantly the species being compared. Reinforcing the importance of more data analysis including Rhodophyta species with a focus on the adaptive evolution of its genes.

Rates of substitutions that support positive selection indicate beneficial changes for the species and can be reflecting different impact factors and underlying variables that result in higher mutation rates. One of these factors can be the population size. Moreover, the substitution values are an average, and although there is support for positive selection, a combination of neutral and purifying selection can be acting on different gene sites. Another point to be considered given the high substitution rates of the *rbc*L gene for Rhodophyta species is sequence saturation. Given the common use of the *rbc*L gene in phylogenetic analysis of red algae, it is important to verify for sequence saturation, since if confirmed, might not reflect the true relationships among species.

3.3.2 *atp* genes

There are 8 genes present in the *atp* gene family and for the present study, seven were analyzed for its ω ratios in Rhodophyta and Chlorophyta species: *atp*A, *atp*B, *atp*E, *atp*F, *atp*G, *atp*H and *atp*I. The green algae species *B. plumosa* does not have a copy of the *atp*D and *atp*G genes (Appendix Table D1), therefore were excluded from the *atp* gene family analysis.

The individual species ratio analysis shows that all 30 species have at least 3 *atp* genes with low ratios, with exception of the *C. caeruleus*, which has only 2 low ratio genes. These genes vary according to the species. However, the ω ratios of *atp*A and *atp*E genes are lower than the cut-off value (0.75) in all the species, which is strong support for purifying selection (Appendix Fig. D1 and D2). The number of *atp* genes with high and neutral ratios are similar throughout the species, while the number of genes with low ratios vary. Curiously, the species *B. atropurpurea* does not have any *atp* genes with high ratios, but yet has most of its genes under low and neutral values. The highest ω ratio from the *atp* gene family is represented by *G. sulphuraria* for the *atp*H gene (Fig. 3.3), and the lowest ratio is represented by *W. schizophylla* for the *atp*E gene.

3.3.2.1 *atp* genes – Habitat

The comparison of ω ratios by species and their ecological habitat show that 5 of the 7 genes had statistical support: *atp*B, *atp*E, *atp*H, *atp*G and *atp*I (Appendix Fig. C2). Of the 19 marine species with a copy of the *atp*B gene, 15 species have ω ratios above 1.25, resulting in a group ratio average of 1.86, indicating possible positive selection in the marine species. This result contrasts with the values for the freshwater species, where the ratio average is 1.2, representing a possible neutral selection. For the *atp*E gene, the difference is not that evident when examining the graph (Appendix Fig. C2), where all the species have low ratios and express strong evidence for this gene to be under purifying selection. However, the data shows significant differences in the ratios for the compared habitats. The salt marshes species *P. purpureum*, with a 0.36 ratio, is significantly different than the freshwater and hot spring species, which have an equal average of 0.26. The marine species, for the *atp*E gene, represents the lowest values of the group, with an average of 0.24. The *atp*G gene, similarly to the *atp*B gene,

has a mix of high and neutral ratios among the species (Appendix Fig. C2). The higher ratio in this gene is observed in the species found in hot springs, *G. sulphuraria* (1.51), followed by the species in the marine environment, which have a significantly lower average of 1.35. In addition, the ratio for species inhabiting salt marshes is also above the maximum cut-off value (1.26) and in all these three habitats, the high average supports a possible positive selection in the *atp*G gene of these species. On the other hand, the species found in freshwater have an average ratio is below the maximum cut-off value (1.22), and it is possibly under neutral selection.

Two *atp* genes stand out when observing the ω ratios of the species, the *atp*H and *atp*I genes. The *atp*H gene has high ratios (> 1.25) in most of the species and across all the observed habitats (Fig. 3.3A), which indicates that positive selection forces are acting in most of the species. Moreover, the ω ratio of *G. sulphuraria* in the *atp*H gene represents the highest ratio in the entire *atp* gene family (4.19), being strong support for positive selection forces. Although the ratios are very similar throughout the species, there is a large difference from the average ω ratio in this gene among all species,1.81, to *B. atropurpurea* ratio, 0.29. This extremely low ratio is strong support of purifying selection forces action and represents the lowest ratio of the gene. When observing the other freshwater species, the average ω ratio is 2.33, and the only species that indicate a possible neutral selection is *H. rivularis* (1.14) which is still a high rate when comparing it to *B. atropurpurea*. The *atp*I gene expressed ratios that are the opposite of the *atp*H gene ratios. All the species in different habitats have ratios below the minimum cut-off value (0.75), with an average of 0.55 which is strong support for purifying selection (Fig. 3.3B). The *atp*I gene ratios of *B. atropurpurea* and *C. caeruleus*, however, are different from the rest of the species, having neutral values, which is indicative of neutral selection.

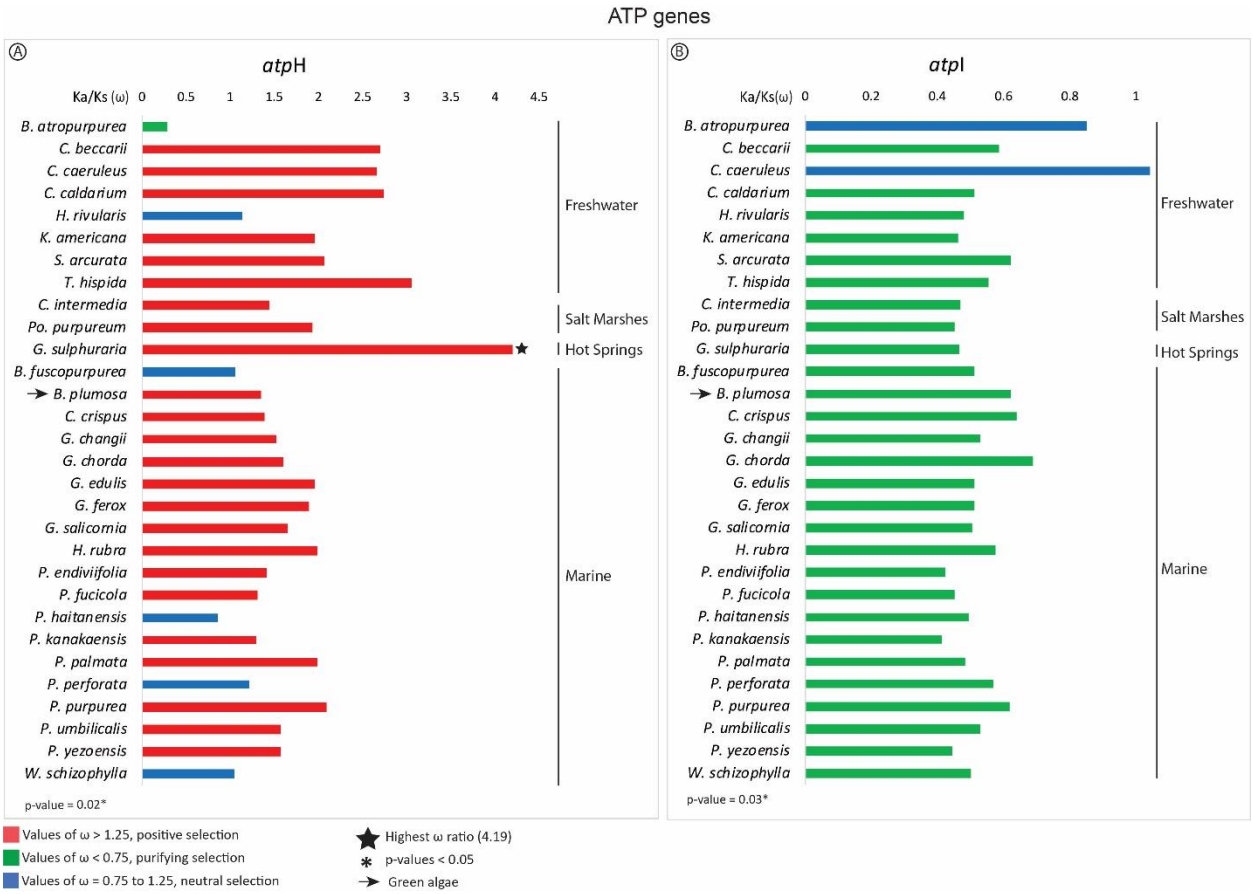

Figure 3.4. Substitution ω ratios of red and green algae in comparison to *C. merolae atp*H and *atp*I gene, and then compared by the ecological habitats of each species.

3.3.2.2 *atp* genes – Phylogeny

The dataset for this analysis reflects the ω ratios of the *atp* genes compared by the species phylogenetic positions. Different genes to the ones mentioned in the *atp* habitat comparison had statistical significance (p-value < 0.05), the *atp*A, *atp*B, *atp*E, *atp*F and *atp*H (Appendix Fig. D1). The ω ratios of the *atp*A gene indicate a possible purifying selection acting in all species since all species gene had ratios below the minimum cut-off value (0.75). The members of the class Florideophyceae, along with the order Porphyridiales, Compsopogonales and Cyanidiales are either identical or have a very similar low ω ratio average (0.54, 0.56, 0.54 and 0.52 respectively). The species within the orders Bangiales (0.61) and Bryopsidales (0.69), represent the highest ω of the *atp*A gene analysis. There is also strong evidence for purifying selection in the *atp*E gene, where the Bangiales and Compsopogonales have the same ω ratio averages (0.21), being the lowest ratios of the *atp*E gene.

However, compared to the above-mentioned genes, the *atp*B gene shows evidence of neutral selection forces acting on the Compsopogonales and Cyanidiales (0.92 and 1.09), and an indication of positive selection for Bangiales, Florideophyceae, Porphyridiales and Bryopsidales (1.75, 1.5, 2.46, 3.85 respectively). The highest average ratio for the *atp*F gene belongs to the Bryopsidales species (1.5), which is strong evidence for positive selection. The species in the class Florideophyceae (1.22) and the orders Compsopogonales (1.13), Porphyridiales (1.03), Bangiales (0.98) and Cyanidiales (0.94) show evidence of being under neutral selection forces (Fig. 3.5). For the *atp*H gene, all lineages have ratios that support positive selection with exception of the Bangiales, which average (1.24) is evidence of neutral selection (Fig. 3.5). It is important to note that the concatenated *atp* gene sequences phylogeny (Appendix, Figure D.3) shows that *B. atropurpurea* genes are closely related to *Porphyra* species, but is represented in a long branch.

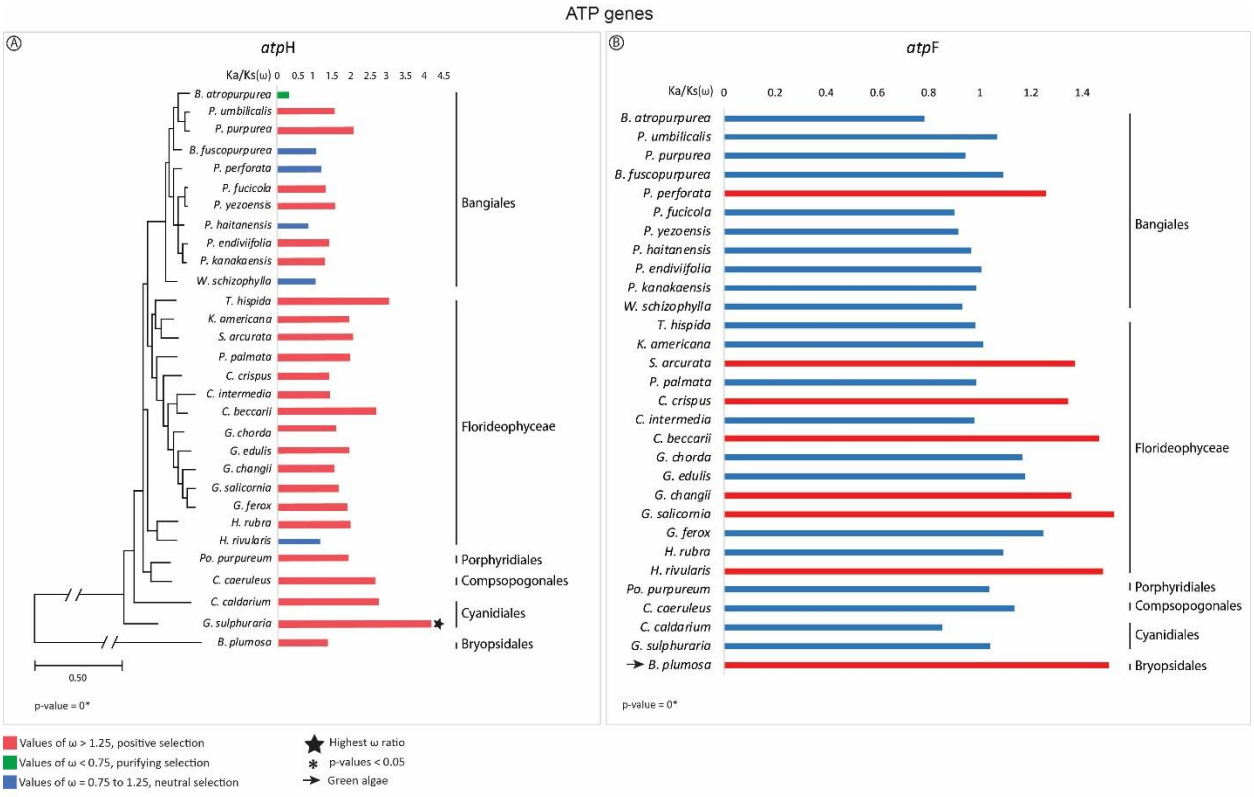

Figure 3.5. Substitution ω ratios of red and green algae in comparison to *C. merolae atp*H and *atp*F gene and compared by phylogenetic position. The phylogenetic position of each species follows the *rbc*L gene tree.

Overall, the species ω ratios comparison by phylogenetic position complemented the observations on the species ω comparison by habitat, where 42% of the ratios are low (< 0.75), being strong evidence of purifying selection for the *atp* gene family. The ratios support purifying selection for the *atp*A, *atp*E and *atp*I genes. Possible positive selection is observed in the *atp*H gene, since it has high averages (> 1.25) for all lineages, with exception of Bangiales, which have a neutral average (1.24). High average ratios are also in most lineages for the *atp*B gene. For the *atp*F gene, all lineages are neutral with exception of the Bryopsidales order, which has a high average (1.5). Lastly, the *atp*G gene has high averages in the Bangiales, Florideophyceae and Cyanidiales, which are indicative of positive selection.

3.3.2.3 *atp* genes – Discussion

The statistical support (p-value < 0.05) for 5 of the 7 *atp* genes confirms that different habitats have an impact on these genes and that different selection forces are acting on the species based on the habitat they are usually found in. The species observed in salt marshes, *C. intermedia* and *P. purpureum*, exhibit the highest averages of ω ratios for the *atp*B, *atpE* and *atp*H genes. These high ratios are strong evidence for positive selection and can be linked to the light and water level variations the salt marshes species experience in their habitat. For example, Pomeroy (1959) noted that algal species inhabiting salt marshes have different photosynthetic rates according to the season (optimal during summer and spring, below optimal during winter and fall), tide (high is optimal) and light availability. Therefore, mutations in the *atp* genes are passed to other generations to adapt to their habitat and obtain better fitness.

The *atp*B and *atp*G genes ω ratios for the freshwater species are considered neutral. This indicates that neutral selection forces are acting on these genes in freshwater species and that mutations that have occurred are not affecting the fitness of these species (Duret, 2008). There is strong evidence of purifying selection for the *atp*I gene, where all species exhibit low ratios except for two freshwater species: *B. atropurpurea* and *C. caeruleus*. These two species have ratios that indicate neutral selection. It is difficult to interpret at present how the *atp*I gene and its protein affect red algae chloroplast function and the reason for it to be under different selection forces in different species since there is no research about the specific role of the gene in Rhodophyta, and no research studies that confirm or refute the adaptive molecular evolution for this gene to be purifying.

The ω ratios of the species arranged by phylogenetic position resulted in 5 *atp* genes with statistical support. For the *atp*A gene, the Chlorophyta species *Bryopsis plumosa* has the highest ω ratio (positive selection) and all Rhodophyta lineages have a low ratio average (purifying selection). The same pattern in ratio averages from green and red species was observed in the *atp*B and *atp*F genes. The difference between the substitution ratios between the chloroplast genes of green and red algae was expected since there is a large evolutionary distance between Chlorophyta and Rhodophyta lineages (Parfrey et al., 2011; Sánchez-Baracaldo et al., 2017; Gibson et al., 2018; Tang et al., 2020). However, the class Florideophyceae and the order Bangiales are sister groups (Müller et al., 2001 and 2010; Saunders and Hommersand, 2004;

Yoon et al., 2006), and because of the closeness of these lineages, it was expected that the ω ratios would be similar when compared to the other lineages. Yet, the Bangiales exhibit a different trend when compared to not only the Florideophyceae but to all the other Rhodophyta orders. For example, the Bangiales *atp*B gene has an average ratio of 1.75, whereas the ratio in the Florideophyceae is 1.5, which indicates that this gene is evolving under different selection pressures in the different lineages. The *atp*B gene codes the beta (β) subunit of ATP synthase enzyme and is vastly used as a sequence marker, and on many occasions is analyzed together with *rbc*L and 18S rRNA genes to help resolve phylogenetic relationships (Hoot et al., 1995; Chiang et al., 1998; Magee et al., 2010; Iqbal et al., 2019; Savolainen et al., 2000). The evolution rate of the *atp*B gene is known for being similar to the *rbc*L rate (Hoot et al., 1995), and our data corroborates this conclusion.

Within the Bangiales species, differences in evolution rates were also noted, as in the *atp*H gene. The species *B. atropurpurea* had an extremely low ω ratio, indicating being under purifying selection, while the other Bangiales species have ratios that support being under neutral or, in most species, positive selection. In addition, the evolutionary trend in Rhodophyta species observed for the *atp*H gene indicates that most of the Florideophyceae, Porphyridiales, Compsopogonales, Cyanidiales and Bryopsidales species are under positive selection forces as well. Like the *atp*I gene, there is no research about the *atp*H gene and its role in Rhodophyta, nor in any of the orders compared in the present study. Our results show that *B. atropurpurea* is the only species that has kept a conserved *atp*H gene, enforcing a stasis in terms of sequence and function of the protein (Prztycka et al., 2008). Other Rhodophyta species, from the Bangiales or different orders, freshwater or marine, had a fast evolution of the gene. More investigation about this specific gene is necessary to conclude the reason for the contrasting results.

As described in the first chapter of this thesis, during the evolution of chloroplasts from cyanobacteria prokaryotic ancestors, the gene clusters and main arrangements of the genes remain mostly unchanged. It is known that a couple of *atp* genes were transferred to the nucleus in the course of evolution, the *atp*C for example is a nuclear gene on Rhodophyta species such as *Porphyra purpurea* (Reith & Munholland, 1995) and in Glaucophytes (Stirewalt et al. 1995). In Viridiplantae species, the *atp*D and *atp*G are also known to be found in the nucleus (Herrmann et al. 1985; Leliaert & Lopez-Bautista, 2015), which agrees with the retrieved data from the green alga used in this analysis, *Bryopsis plumosa*.

When the substitution rates of the *atp* genes are compared, the species *B. atropurpurea* stands out for exhibiting the lowest values and/or having a very different trend in terms of ratio values, being the opposite of most species. The only ratio that is similar to the other species is on the *atp*B gene, and the lower ratios can be observed on the *atp*E, *atp*G and *atpH* genes. Neutral ratios are observed for the *apt*I gene, however, all the other species with exception of *C. caeruleus* exhibit low ratios. These results are supported by the ML phylogenetic tree of the concatenated *atp* genes, which shows *B. atropurpurea* within the Bangiales, but in a very long branch. Long branches represent a large number of evolutionary changes (Parks & Goldman, 2014), and therefore, how many mutations occurred in the evolutionary time between lineages. As mentioned in Chapter 1 of this thesis, the population of *B. atropurpurea* most likely derived from freshwaters in Europe, and due to an excellent adaptation mechanism invaded successfully into the Laurentian Great Lakes (Müller et al., 1998; Shea et al., 2014). Because the ATP complex is responsible for the photosynthetic electron transport (Gillham & Boynton, 1986; Groth & Stromann, 1999), it can be relevant for the adaptation of *B. atropurpurea* into new freshwater habitats, since a photosynthetic mechanism change must have occurred to survive in such a different environment.

During the evolution of chloroplasts from cyanobacteria prokaryotic ancestors, the gene clusters and main arrangements of the genes have remained mostly unchanged (Groth & Stromann, 1999). There is evidence that some of the *atp* genes were transferred to the nucleus in the course of evolution, the *atp*C for example, is a nuclear gene on Rhodophyta species such as *Porphyra purpurea* (Reith & Munholland, 1995) and in Glaucophytes (Stirewalt et al. 1995), justifying the absence of the *atp*C gene on this study. In Viridiplantae species, the *atp*D and *atp*G are also found in the nucleus (Herrmann et al., 1985), which agrees with the retrieved data from the green alga used in this analysis, *Bryopsis plumosa*.

When the substitution rates of the *atp* genes are compared, the species *B. atropurpurea* stands out for exhibiting the lowest values and/or having a very different trend in terms of ratio values, being the opposite of most species. The only ratio that is similar to the other species is on the *atp*B gene, and the lower ratios can be observed on the *atp*E, *atp*G and *atpH* genes. Neutral ratios are observed for the *apt*I gene, however, all the other species with exception of *C. caeruleus* exhibit low ratios. These results are supported by the ML phylogenetic tree of the concatenated *atp* genes, which shows *B. atropurpurea* within the Bangiales, but in a very long

branch. Long branches represent a large number of evolutionary changes (Parks & Goldman, 2014), and therefore, how many mutations occurred in the evolutionary time between lineages. It is important to consider that for the analysis performed in this chapter, all the substitution rates are a result of the comparison of the species gene sequence against *C. merolae* (as detailed in the material and methods). This, together with the long branches observed in the concatenated sequence phylogeny, must be taken into consideration given the unique characteristics of *C. merolae*, such as the known accelerated evolution due to the extremophile habitat the organism is observed (Kuroiwa, 1998; Kobayashi et al., 2014; Antoshvili et al., 2019). Interestingly, previous observations of the *Bangia* species in long branches were noted by Sutherland et al. (2011), with the use of the marker genes *rbc*L and SSU rRNA to create the phylogenetic analysis. The long branches and the discrepancy on the rates of substitution of *B. atropurpurea atp* genes, could be an indication of fast evolution or sequence saturation, and should be investigated further. As mentioned in Chapter 1 of this thesis, the population of *B. atropurpurea* most likely derived from freshwaters in Europe, and due to an excellent adaptation mechanism invaded successfully into the Great Lakes in Canada (Müller et al., 1998; Shea et al., 2014). Because the ATP complex is responsible for the photosynthetic electron transport (Gillham & Boynton, 1986; Groth & Stromann, 1999), it can be relevant for the adaptation of *B. atropurpurea* into Canadian waters, since a photosynthetic mechanism change must have occurred to survive in such different environment.

3.3.4 *psa* genes

The *psa* genes are part of the Photosystem I complex (PSI), built by two large subunits (*psa*A and *psa*B) and some low-molecular-weight polypeptides (*psa*F, *psa*L, *psa*K, *psa*G, *psa*I, *psa*J, and *psa*M), that are integral membrane proteins with transmembrane domains (Shimada and Sugiura, 1991; Chitnis, 1996). Moreover, the *psa*C, *psa*D, *psa*E, and *psa*H are peripheral proteins on the reducing side of PSI, whereas *psa*N associates with the luminal side (Chitnis, 1996). The functions of the small PSI proteins have been studied for their biochemical mechanisms, such as the *psaC*, *psaI*, and *psaJ*, that make up the protein P700 and are responsible for the cyclic electron flow around in the thylakoid membranes (Arnon *et al.*, 1954; Shimada and Sugiura, 1991); the *psa*D, is part of the photosynthesis reaction center (Bolle et al., 1996; Chitnis et al., 1996); the *psa*E act on stabilizing the interaction between *psa*C and *psa*I (Sonoike, et al., 1993); the *psa*F is essential for the thylakoid viability and part of the electron transfer (Xu et al., 1994; Farah et al., 1995; Hippler et al., 1997); the *psa*L is responsible for mRNA binding (Chitnis et al., 1993; Schluchter et al., 1996); the *psa*M is involved in the coordination of some antenna pigments and stabilization of PSI trimers (Naithani et al., 2000). There are no studies regarding the role of the protein produced by the *psa*J gene, but it is a recognized part of the PSI (Chitnis, 1996). The PSI is essential for photosynthesis because it contributes to the ATP synthesis and regulates the ATP: NADPH ratio (Munekage *et al.*, 2004), which is crucial for preventing stroma over-reduction (Backhausen et al., 2000; Kramer et al., 2004*;* Munekage et al., 2004; Gao & Wang, 2012). In addition, studies demonstrated that the PSI is a key element in plant responses to drought or desiccation stress (Horváth et al., 2000; Golding and Johnson, 2003; Gao *et al.*, 2011; Gao & Wang, 2012).

Although this pathway is known to be present in Viridiplantae, cyanobacteria and eukaryotic algae, research around the PSI functionality have focused on model organisms such *Arabidopsis thaliana*, *Chlamydomonas reinhardtii*, and tobacco (Horváth et al., 2000; Joët et al., 2001, 2002; Munekage et al., 2002, 2004; DalCorso et al., 2008; Iwai et al., 2010). The main difference between the photosynthetic apparatus of green algae and Viridiplantae species to Rhodophyta is the presence of phycobilisomes instead of chlorophyll a/b complexes in red algae species (Gantt, 1980). Unfortunately, there is only a small amount of literature available about the PSI complex in Rhodophyta and most studies use only marine species, focusing on the

cycling electron flow and photosynthesis analyzes of the red species undergoing stress (Levy & Gantt, 1990; Marquardt & Rehm, 1995; Sagert & Schubert, 1995; Figueroa et al., 1995; Lin et al., 2009; Gao & Wang, 2012). Therefore, the knowledge about the PSI and the evolution rates on each PSA gene in Rhodophyta is very limited.

For the *psa* gene family, 11 genes were analyzed for their ω ratios in Rhodophyta and Chlorophyta species: *psa*A, *psa*B, *psa*C, *psa*D, *psa*E, *psa*F, *psa*I, *psa*J, *psa*K, *psa*L and *psa*M. The green algae *Bryopsis plumosa* does not have a copy of the *psa*C, *psa*D, *psa*E, *psa*J, *psa*K and *psa*L genes. The freshwater species *C. caeruleus* (Compsopogonales), lack the *psa*J and *psa*M genes. Another freshwater species, *Sheathia arcuata* (Batrachopsermales), does not have the *psa*F and *psa*I genes (Appendix Table E1). The calculation error of the number of synonymous substitutions per synonymous site (K_s) noted previously on the *atp*D gene, was also observed with the *psa*K gene, in the following species: *Hildenbrandia rivularis, Cyanidium caldarium, Caloglossa intermedia, Gracialaria salicornia, G. ferox, G. changii, G. edulis, Pyropia yezoensis, P. fucicola, P. endiviifolia, Bangia fuscopurpurea, Porphyra umbilicalis and P. purpurea.* Hence, the analysis of the *psa*K gene was concluded without these species since the ratio of *B. atropurpurea* was not affected.

The ω ratios for the individual species show that 15 species have at least one *psa* gene with a ratio below 0.75, which is indicative of purifying selection. The species *B. atropurpurea*, with 3 genes with low ratios (*psa*C, *psa*L and *psa*M), has strong support for being under purifying selection for the *psa* genes. In addition, the ratios of *B. atropurpurea* are opposite to the trend of the other species in the *psa*C, *psa*F and *psa*M genes. Of the 30 species analyzed, *H. rivularis, Kumanoa Americana* and *Thorea hispida* don't have any low ratios within the *psa* genes, having ratios that suggest neutral or positive selection only. The species *G. ferox* have 8 of the 15 *psa* genes with high ratios (> 1.25), being the species with the highest ω average and is possibly under positive selection for the *psa* genes. Neutral ratios (0.75 to 1.25) were mostly observed in C. *beccarii*, 8 *psa* genes, indicating these genes to be under neutral selection. Of the whole analysis, the lowest ω ratio is observed on the *psa*K of the species *Pyropia haitanensis* (0.16), and the highest ratio is observed on the *psa*I of the species *Kumanoa americana* (6.95).

3.3.4.1 *psa* genes – Habitat

The comparison of ω ratios by species and their ecological habitat show that four of the eleven genes have statistical support (p-value < 0.05), confirming significant evolutionary differences in different environments: *psa*A, *psa*D, *psa*L and *psa*M (Appendix Figures E1 and E2). For the *psa*A gene, all species but *C. beccarii* (freshwater) and *Galdieria sulphuraria* (hot springs), have high ω ratios. These two species have ratios of 1.21 and 1.24 respectively, which according to the followed methodology, are under neutral selection. All the other species fall under possible positive selection (ω > 1.25). The species observed in the marine environment have the highest average (1.42) for the *psa*A gene and are significantly different from the species in freshwater (1.38) and the salt marshes group (1.35).

Among the different environments, the species have high and neutral ω ratios for the *psa*D gene. The graph (Fig. 5a) shows that the species observed in hot springs have the highest ratio of the *psa*D gene (1.94), being significantly different from the second-highest average, the freshwater group (1.38). High ratios were also noted for the species observed in salt marshes, with an average of 1.36. The high ratios for the *psa*D gene in hot springs, freshwater and salt marshes species indicate possible positive selection forces occurring in these habitats. On the other hand, the average ratio of the species observed in the marine environment (1.18), is the lowest average of the *psa*D gene and indicates neutral selection.

Differently from the *psa*A and *psa*D genes, the ratios of the *psa*L gene are below the minimum cut-off value (< 0.75) in 23 species. The remaining 6 species have ratios that are considered neutral (0.75 – 1.25) and are observed in freshwater and salt marshes. The only marine representative with neutral ratios for the *psa*L gene is the species *Bangia fuscopurpurea* (Fig.5b). The species observed in hot springs and the marine habitat have an average of 0.7 and 0.63, the lowest ratios for the *psa*L gene, which is strong evidence for purifying selection forces (Fig. 5b).

Lastly, the *psa*M gene has ratios that vary from high (> 1.25), neutral (0.75 to 1.25) and low (< 0.75). The marine environment species concentrate the highest ratios observed, with an average of 1.3 (Fig. 5c). This high ratio suggests that the species found in the marine environment have positive selection forces acting on the *psa*M gene. The species observed in hot springs and salt marshes have an average below the minimum cut-off (0.56 and 0.67), indicating

possible purifying selection. The species found in freshwater are possibly under neutral selection, with an average of 0.95.

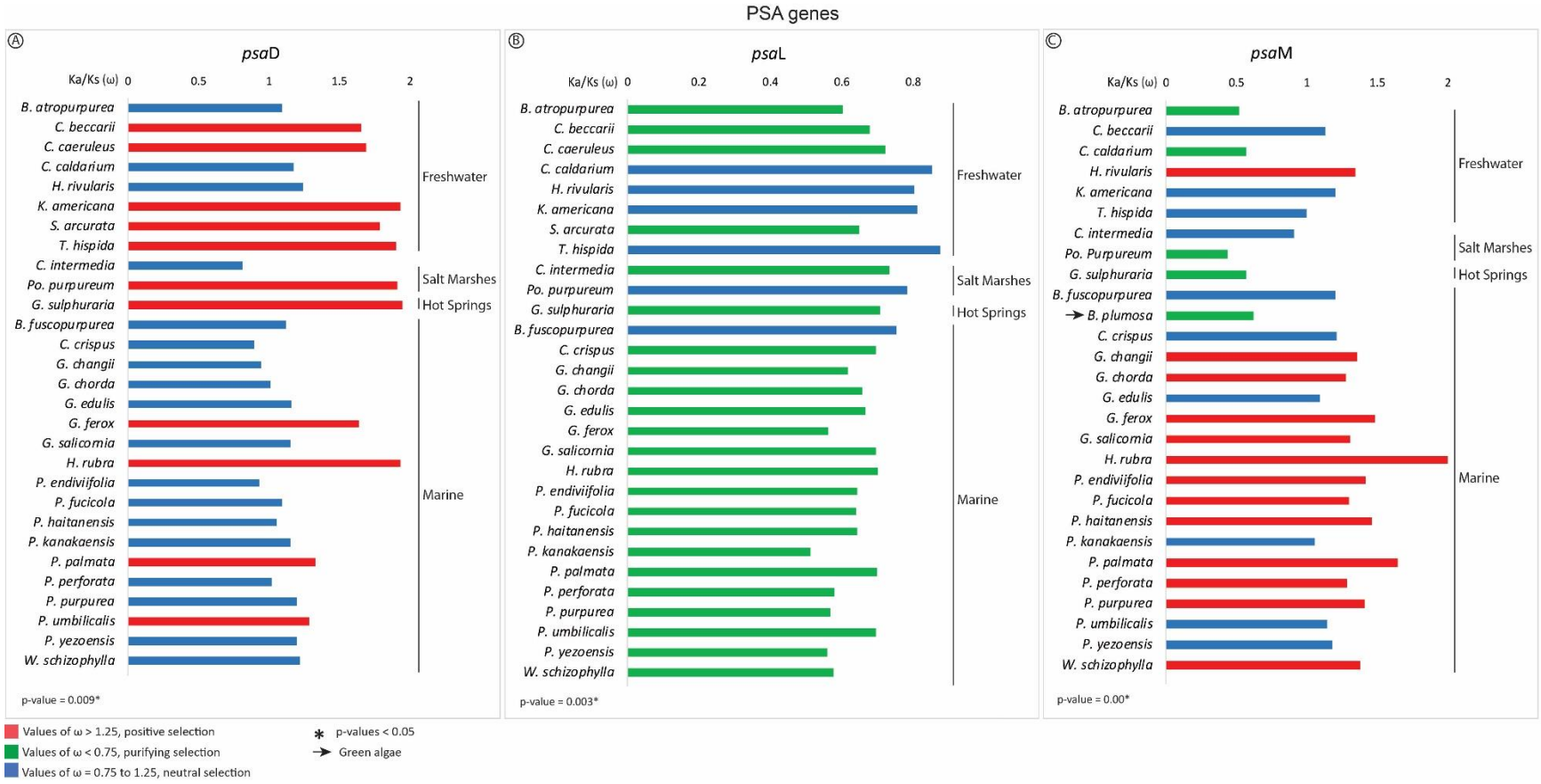

Figure 3. 6. Substitution ω ratios of red and green algae in comparison to *C. merolae psa*D, *psa*M and *psa*L gene, and then compared by the ecological habitats of each species.

3.3.4.2 *psa* genes – Phylogeny

The comparison of ω ratios by species and their phylogenetic position shows that 7 of the 11 PSA (*psa*B, *psa*C, *psa*E, *psa*F, *psa*I, *psa*L and *psa*M) have statistical support for different selection forces acting on these genes according to the phylogenetic position of the species, confirming evolutionary differences among the lineages. For the *psa*B gene, the species exhibit a mix of high (> 1.25) and neutral ratios (0.75 to 1.25), with Bangiales species concentrating the highest ratios (Fig. 6a). Therefore, the gene is possibly under positive selection in the Bangiales species. All the other lineages have average ratios are under 1.25, indicating neutral selection.

For the *psa*C gene, most species have high ω ratios, with only four species having neutral ratios: *Pyropia perforata, Wildemania schizophylla, Porphyridium purpureum* and *Compsopogon caeruleus*. Figure 6b shows that the ratio of *B. atropurpurea* (Bangiales) is 0.62, which is below the minimum cut-off value and the opposite of all the other species. This is strong evidence for purifying selection on the *psa*C gene of *B. atropurpurea*. However, all the other Bangiales species have ratios that support positive selection, with a ω average of 1.35. The Florideophyceae lineage has the highest ratio for the *psa*C gene, and the group is considered to be under positive selection as well, along with the Cyanidiales and Bryopsidales lineages. Neutral ratios are observed for the Porphyridiales and Compsopogonales, suggesting that the species of these lineages are under neutral selection. Likewise, the compared species have high and neutral substitution ω ratios for the *psa*E gene. The neutral ratios indicate possible neutral selection forces acting on species of the Florideophyceae (1.29), Bangiales (1.19) and Compsopogonales (0.81). The Porphyridiales and Cyanidiales have high average ratios (1.45 and 1.95 respectively), indicating strong evidence for being under positive selection.

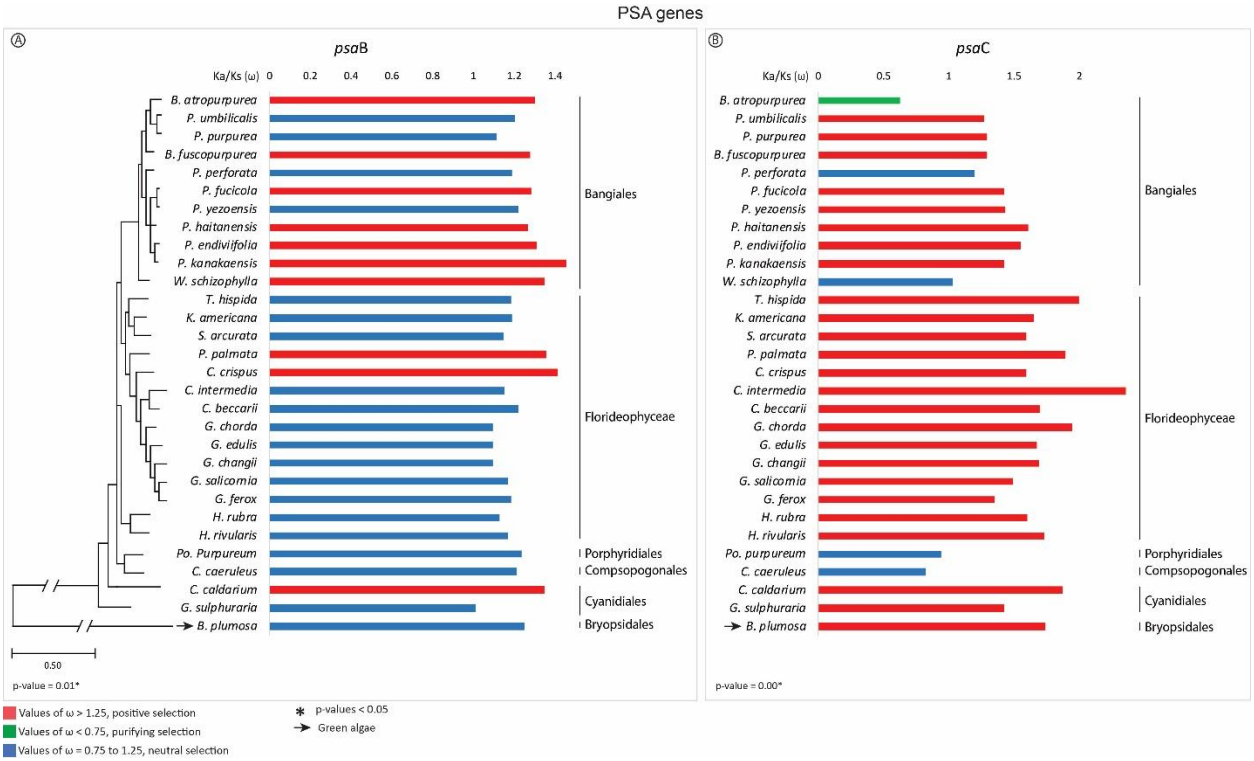

Figure 3. 7. Substitution ω ratios of red and green algae in comparison to *C. merolae psa*B and *psa*C gene, and then compared by the phylogenetic position of each species. The phylogenetic position of each species follows the *rbc*L gene tree.

There is a considerable difference between the Bangiales and the other lineages on the substitution rates of the *psa*F gene (Fig. 7b), where 7 of the 11 species have low ω ratios (< 0.75), suggesting strong evidence for purifying selection forces acting on these species. The species *B. atropurpurea*, *B. fuscopurpurea*, *P. fucicola* and *W. schizophylla* have neutral ratios for the *psa*F gene and are possibly under neutral selection. The Porphyridiales is the only order with average ratios higher than the maximum cut-off value (1.36) which is support for positive selection.

All 30 species, and consequently all lineages, have high ω ratios for the *psa*I gene (Fig. 7a). Although all species have high ratios and have strong support for being under positive selection, there are substantial differences between the lineages. The Bryopsidales have the lowest ratio (1.4), followed by the Bangiales (2.22) and Porphyridiales (2.86). The ratios of Cyanidiales, Compsopogonales and Florideophyceae (3.5, 3.74 and 3.84 respectively) represent the highest of the gene.

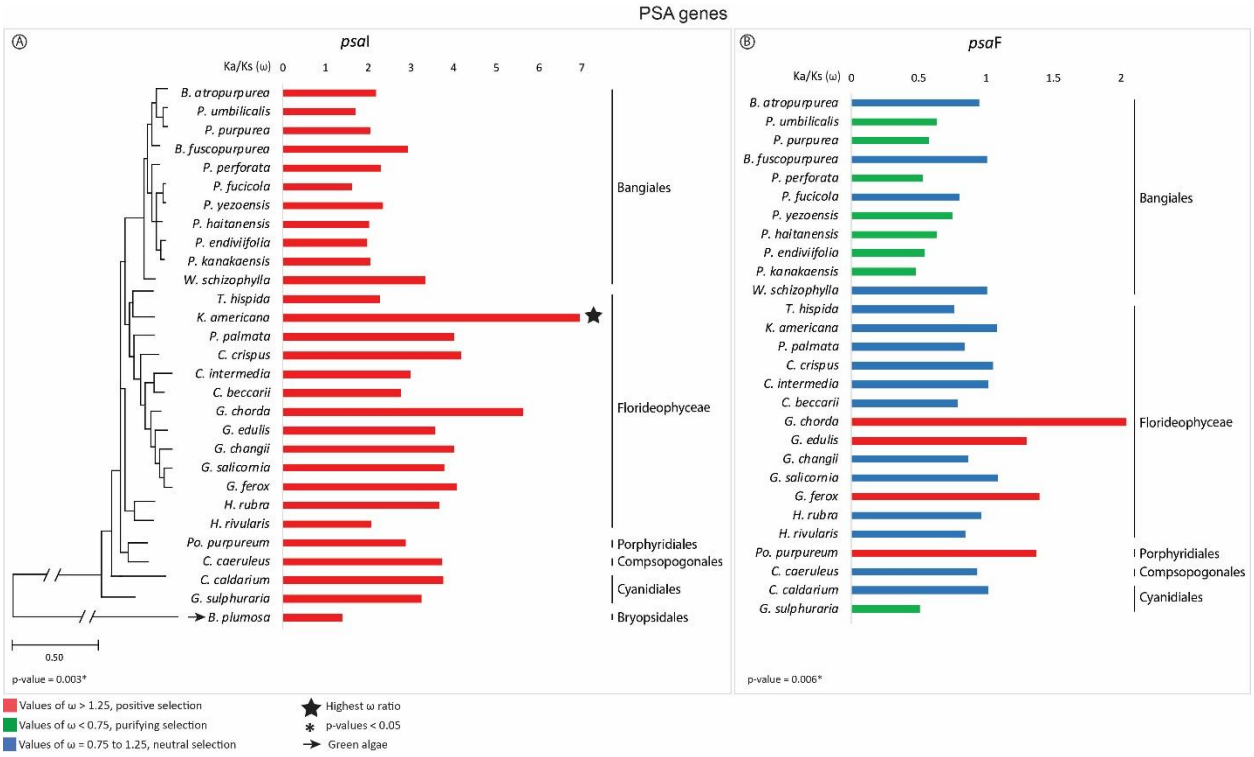

Figure 3. 8. Substitution ω ratios of red and green algae in comparison to *C. merolae psa*I and *psa*F gene, and then compared by the phylogenetic position of each species. The phylogenetic position of each species follows the *rbc*L gene tree. The green algae *B. plumosa* does not have a copy of the *psa*F gene and was not used in this analysis.

Of the 29 species that have a copy of the *psa*L gene, 23 have ω below the minimum cut-off ratios (< 0.75), indicating being under purifying selection (Fig. 8a). The other species have neutral ω ratios and are possibly under neutral selection. The species with neutral ratios are spread throughout the lineages, therefore, the ω ratio average of the lineages is low. The Compsopogonales and Florideophyceae have similar ratios (0.72 and 0.7), while Bangiales represents the lowest average of the *psa*L gene (0.61).

The substitution ratios of the *psa*M gene are high, neutral and low. The Bangiales order is the only lineage with species expressing all three types of ratios (Fig. 8b), and despite the high ratio of some of the Bangiales species, the order's ω ratio average is 1.19, supporting neutral selection. Curiously, like on the ratio for the *psa*C gene, *B. atropurpurea* (0.52) has similar ratios to the ones expressed by the Porphyridiales, Cyanidiales and Bryopsidales orders (0.43, 0.56, 0.62), and are possibly under purifying selection. The Florideophyce species have neutral and high ratios, where the dominance of high values is observed in the *Gracilaria* and *Hildenbrandia*

species. Because of these high ratios, the Florideophyceae is the only lineage with an ω ratio average above the maximum cut-off (1.3), indicating positive selection for the *psa*M gene.

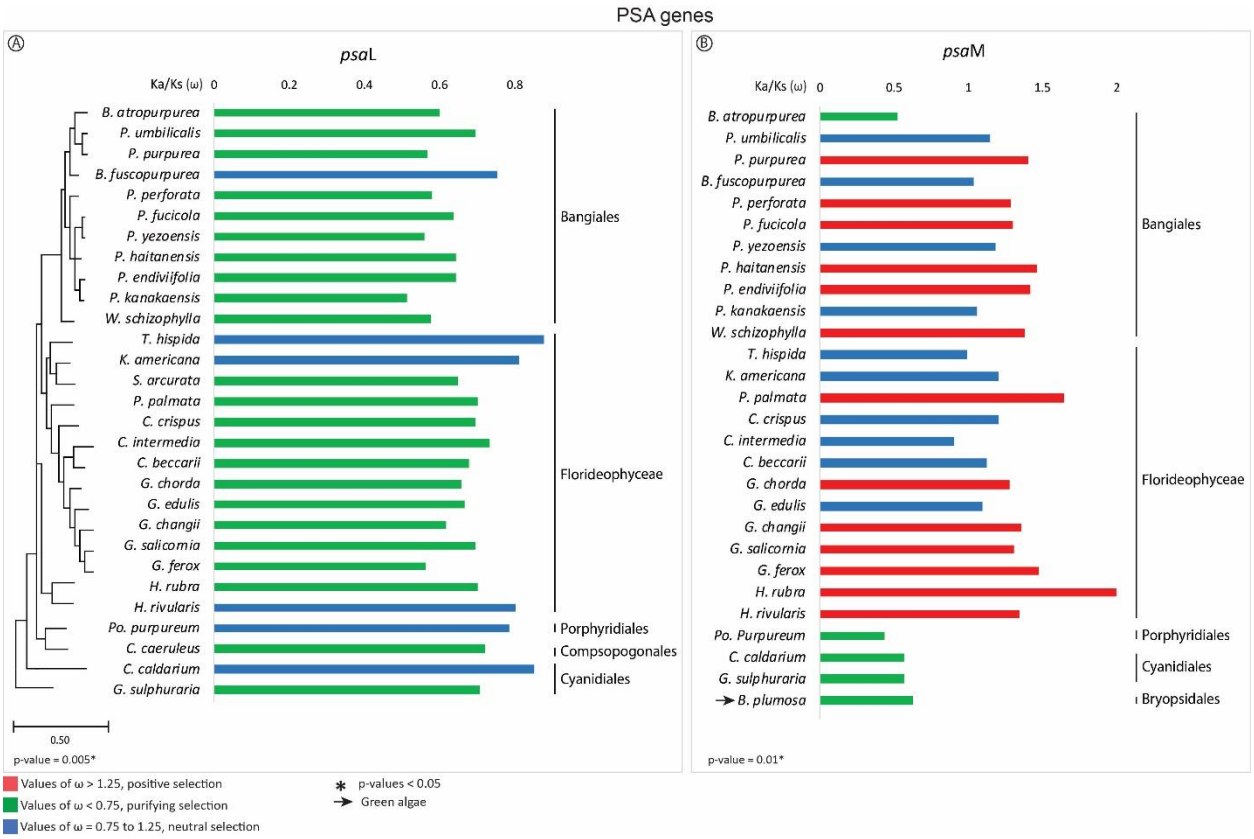

Figure 3. 9. Substitution ω ratios of red and green algae in comparison to *C. merolae psa*L and *psa*M gene, and then compared by the phylogenetic position of each species. The phylogenetic position of each species follows the *rbc*L gene tree. The green algae *B. plumosa* does not have a copy of the *psa*L gene and was not used in this analysis.

The comparison of ω ratios by species and their phylogenetic position shows that 46% of the ratios are considered high (ω > 1.25), being strong evidence of positive selection forces acting in almost half of the compared species. Higher ω averages are particularly noted for the Florideophyceae, Porphyridiales and Cyanidiales lineages, each having 6 *psa* genes with ratios above 1.25. The Bangiales species has a distinct ω pattern for the *psa*B, *psa*D, *psa*F and *psa*M, where: it is high in *psa*B while all the other lineages are neutral; neutral in *psa*D while all the other lineages are high; low in *psa*F where all but Porphyridiales (which is high) are neutral; and neutral in *psa*M, where all lineages but Florideophyceae (high) are low.

3.3.4.3 *psa* genes – Discussion

In our analysis, 4 of the 11 *psa* genes exhibit statistical support when ω ratios are compared by species and their ecological habitat. The species observed in the marine environment have either the highest (*psa*A and *psa*M) or the lowest (*psa*D and *psa*L) ω ratios. There is no research about the substitution rates of these particular genes in Rhodophyta, but because they are part of the PSI, we can speculate that the marine species have different ω ratios for some of the *psa* genes due to environmental adaptation. The species used in the present study are predominantly marine intertidal species (Reith & Munholland, 1995; Wang, 2013; Hughey et al., 2014; Gao &Wang, 2012; Brawley et al., 2017; Cao et al., 2018; Xu et al., 2018). Marine intertidal species must adapt to the environment to survive desiccation during low tides, re-hydration, and drought during high tides (Gao & Wang, 2012). A study about the physiology of the PSI in Rhodophyta was performed with the species *Pyropia yezoensis*, which has elevated PSI electron flow during desiccation, suggesting that the PSI has an important role during desiccation and re-hydration of its blades (Gao & Wang, 2012). More research about specific *psa* genes is needed to conclude how the PSI in marine Rhodophyta evolved, and how they are connected to the adaptation mechanisms of the species.

Curiously, of the 8 freshwater species, *B. atropurpurea* has a unique pattern of ω ratios pattern for the *psa* genes, having more genes under purifying selection than the other 29 species being compared. More specifically, the *psa*C, *psa*F and *psa*M genes. It is worth noting the deficit of literature around these genes and their evolution ratios even for Viridiplantae species. Studies focusing on Rhodophyta species are limited to describing the presence or absence of these genes in their genomes. Since the PSI is a vital part of the photosynthetic cell (Fromme & Mathis, 2004) and plays an important role in response to stress (Lin et al., 2009; Gao & Wang, 2012), one can hypothesize that *B. atropurpurea* has a different gene evolution pattern due to the adaptation in freshwater. The species is the only freshwater within the Bangiales, and its successful adaptation to different environments has been tested by Geesink (1973) and Sheath & Cole (1980). The studies show that *B. atropurpurea* can quickly adapt from freshwater to saltwater and vice-versa, reinforcing how peculiar the physiology and adaptation mechanisms of this species are. Moreover, the ω ratios of *B. atropurpurea* are significantly different from the ω ratios of other freshwater species, which indicates that *B. atropurpurea* has an exclusive

substitution rates pattern, therefore, making use of its PSI complex to overcome environmental stress and strive in harsh environments such as the Laurentian Great Lakes.

The comparison of ω ratios by species and their phylogenetic position shows statistical support in 7 of the 11 *psa* genes, indicating that different selection forces are acting on the species based on their taxa (Appendix E3 and E4). The most distinctive results are for the Bangiales order, where the ω ratios are the opposite of all the other lineages in 4 *psa* genes: *psa*B, *psa*D, *psa*F and *psa*M. The *psa*B gene is essential for the biogenesis in the PSI since the suppression of the gene results in the disruption of the complex (Smart & McIntosh, 1993). This gene is vastly used in phylogenetic studies, usually in addition to *rbc*L and other marker genes (Tullberg et al., 2000; Cameron, 2004; Yoon et al., 2006; Saunders & Moore, 2013; Zhan et al., 2020), is considered suitable for intrafamilial investigations within Viridiplant species (Nishiyama & Kato, 1999; Graham & Olmstead, 2000; Sanderson et al., 2000).

The higher average substitution ratio for the *psa*B gene in Bangiales is due to the high ω ratios of the *Pyropia* and *Bangia* species and is support for positive selection. The *Porphyra* species have ω that indicates neutral selection, which is very similar to all the other lineages. Curiously, *Pyropia* and *Bangia* species are more distant phylogenetically than *Bangia* and *Porphyra* species, which share a common ancestor. Moreover, *B. atropurpurea* is the only freshwater species in this analysis with a high ω ratio for *psa*B gene. The mutations occurring on the *psa*B genes of *Pyropia* and *Bangia* are accumulating faster than in the other species. However, the gene is conserved and evolving at a neutral rate in most Rhodophyta, and can continue being used for phylogenetic studies along with other marker genes.

The other genes with a different ω pattern in Bangiales, *psa*D, *psa*F and *psa*M, are small protein subunits that are associated and assist in the functionality of the PSI complex (Fromme & Mathis, 2004). With exception of the *psa*F gene, which is known for its critical role on the thylakoid structure and viability (Xu et al., 1994; Farah et al., 1995; Hippler et al., 1997), there are not many studies about the other genes besides their main roles. Unquestionably, more research about these genes in Rhodophyta and their evolution mechanisms is necessary to better understand how and why the evolutionary pattern of the PSI genes in Bangiales are so different from the other orders, even among species that share the same habitat.

3.3.5 *psb* genes

The photosystem II (PSII), located in the thylakoid membrane of algae, cyanobacteria, and plants, is responsible for the photochemical and electron transport reaction of oxygenic photosynthesis, performed along with the PSI, cytochrome b_6f complex and ATP synthase (Whitmarsh & Govindjee, 1999; Freeman, 2006; Kern et al., 2010; Mizusawa & Wada, 2012). During photosynthesis, these organisms fix carbon dioxide (CO_2) into carbohydrates, resulting in the release of molecular oxygen into the atmosphere; this process is vital to support life on Earth (Murata et al., 2007; Kern et al., 2010; Shevela et al., 2012). Therefore, the production of oxygen and the evolution of oxidative respiration depends on this unique protein complex that is the PSII. The structure and function of PSII in Viridiplantae species, algae and bacteria are extremely similar (Kern et al., 2010). Sadekar et al. (2006), notes that the PSI and PSII reaction centres of bacterial, algae and land plants share several structural features, indicating ancient evolutionary links. However, the antenna system, responsible for the light capture, will vary its structure in different organisms, which indicates multiple origins along with the evolution of the photosynthesis (Green et al., 2003; Kern et al., 2010).

The PSII is known as a multiprotein super complex which requires the production of over 30 proteins subunits (Baena–González & Aro, 2002; Minagawa & Takahashi, 2004). It is also known that part of the critical proteins of the PSII is encoded not only in the chloroplast but also in the nucleus (Baena–González & Aro, 2002). The PSII is built by two types of proteins, intrinsic and extrinsic. The intrinsic proteins are conserved in most photosynthetic organisms and are commonly found in prokaryotic cyanobacteria and eukaryotic plants (Bricker & Ghanotakis, 1996; Enami et al., 1998; Gardian et al., 2007). These intrinsic proteins are the D1/D2 heterodimer proteins – encoded by the *psb*A and *psb*D genes; the α and β subunits of the cytochrome b559 – encoded by the *psb*E and *psb*F genes; and the *psb*I and *psb*W gene products (Lorković et al., 1995; Baena–González & Aro, 2002; Green et al., 2003). Other subunits include the chlorophyll-binding proteins CP43 and CP47, which are encoded by the *psb*B and *psb*C genes (Mullet et al., 1990; Bricker & Frankel, 2002; Minagawa & Takahashi, 2004). Research shows that the inactivation of the genes responsible for the translation of the core subunits proteins leads to a loss of normal PSII assembly and a loss of the ability to evolve oxygen (Kiss

et al., 2012). Therefore, all photosynthetic organisms are expected to have a copy of these core genes.

The extrinsic proteins are significantly different among the photosynthetic organisms. For example, in green algal and Viridiplantae species, the PSII has 3 extrinsic proteins: 33, 23, and 17 kDa, which maintain the stability and activity of the oxygen-evolving complex (Enami et al., 1998; Shen et al., 1995). These proteins are encoded by the *psb*O, *psb*P, *psb*Q genes respectively (Seidler, 1996; Ohta et al, 2003). In cyanobacteria, only the 33 kDa protein is present, along with two different extrinsic proteins, the cytochrome c_{550} and a 12 kDa, which have similar roles as the 23 and 17 kDa proteins (Shen et al., 1992; Enami et al., 1998; Gardian et al., 2007), and are encoded by the *psb*U and *psb*V genes respectively (Shen et al., 1993; Ohta et al., 2003). The difference between the extrinsic proteins in cyanobacteria and Viridiplantae species elucidates the evolutional process of the oxygen-evolving complex from prokaryotes to eukaryotes (Enami et al., 1998). Studies with the species *Cyanidium caldarium* (Rhodophyta), show that red algae have cyanobacterial-type extrinsic proteins - cytochrome c_{550} and the 12 kDa protein, in addition to the 33 kDa protein (Enami et al., 1995; Enami et al., 1998; Ohta et al., 2003). This implies that the red algae PSII is closely related to cyanobacterial PSII and that the replacement of the two extrinsic proteins by the 23 and 17 kDa proteins occurred beyond the development of red algae (Enami et al., 1998; Ohta et al., 2003; Gardian et al., 2007; Busch et al., 2010; Tian et al., 2017).

In Rhodophyta, the PSII contains more than 25 subunits (coded by *psb*A-*psb*Z genes) (Gardian et al., 2007), but the functions of the majority of these subunits remain obscure (Kiss et al., 2012). Research about the photosynthetic apparatus indicates how unique the red algae PSII is. The red algae species have a novel extrinsic protein, the 20 kDa, which is not present in cyanobacterial nor Viridiplantae species (Enami et al., 1995; Ohta et al., 2003; Enami et al., 2005; Zheng et al., 2020). Moreover, the light-harvesting antenna, which is distinct to Rhodophyta, Cyanophyta, Glaucophyta and Cryptophyta, contain phycobilisomes, similar to those found in prokaryotic cyanobacteria, rather than chlorophylls to capture light like in Viridiplantae (Bricker & Frankel, 2002; Green & Gantt 2005; Kern et al., 2010; Zheng et al., 2020). However, the connection between the red algae PSI and PSII follows what is observed in all the major groups of photosynthetic eukaryotes (Jansson, 1994; Wolfe et al., 1994; Scheller et al., 2001; Yokono et al., 2011). There might be other peculiarities among the PSII genes of

Rhodophyta species, especially regarding differences in evolution, but more research is needed to have a conclusion.

In our study, a total of 15 genes part of the *psb* gene family were analyzed: *psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psbF*, *psb*H, *psb*K, *psb*L, *psb*N, *psb*T, *psb*V, *psbW*, *psb*X and *psb*Z. The marine species, *Pyropia fucicola, P. kanakaensis, P. perforata* and *Wildemania schizophylla* have its *psb*B gene named as *psi* gene. Other species lacked one or few copies of PSB genes: *Sheathia arcuata* lacks *psb*L and *psb*X genes; *Kumanoa americana* and *Compsopogon caeruleus* lack *psb*T gene; *Chondrus crispus* lack *psb*W; *Bryopsis plumosa* lacks *psb*V, *psb*W and *psb*X; *Cyanidium caldarium* lacks *psb*X and *psb*Z (Appendix Table F1). A total of ten species (*B. atropurpurea, B. fuscopurpurea, P. fucicola, P. yezoensis, P. haitanensis, P. kanakaensis, P. perforata, P. endiviifolia, W. schizophylla*, and *K. americana*) could not have their K_s value calculated by the DNAsp software because the Jukes and Cantor (1969) correction could not be computed for the *psb*J gene, affecting the further analysis of the gene.

3.3.5.1 *psb* genes - Habitat

From the 15 *psb* genes, three had statistical support for significant differences in the various habitat groups: *psb*E, *psb*H and *psb*Z (p-value < 0.05) (Appendix Figures F1 and F2). Of the 7 gene families analyzed in this study (*rbc*L, *atp*, *psa*, *psb*, *rpl*, *rps*, *rpo*), the comparison of the *psb* genes of species in different habitats has the lower number of significant differences. This reinforces that the *psb* genes are conserved in most species independently of their habitats.

The *psb*E gene has high ω ratios for all species, with a total average of 2.3, which is strong evidence for positive selection (Figure 8a). Curiously, *B. atropurpurea* is the only species with a low ω ratio, 0.33, one of the lowest of the whole *psb* gene analysis, which is strong support of purifying selection. When analyzing the habitats, the hot springs have the lowest ratio of the *psb*E gene (1.5), which is already above the maximum cut-off (1.25) being under positive selection. Also, under possible positive selection, the other groups have higher averages such as the freshwater group, with an average of 2.18, followed by the marine (2.34) and salt marshes (2.84) (Fig. 8).

Contrasting with the high ratios of the *psb*E gene, the *psb*H has a mix of low ratios (< 0.75) and average ratios (0.75 to 1.25) with a total average of 0.87. Although all habitats have

ratios that suggest the gene to be under neutral selection, there is more similarity between the freshwater and marine groups (0.91 and 0.84) than between salt marshes (1.03) and hot springs (1.24) groups (Appendix F1 and F2). With more species with lower-than-average ratios, the *psb*Z gene is probably under purifying selection. The difference between habitats is visible when observing the graph in Fig. 8b, where all the freshwater species have ratios below the minimum cut-off (0.57), as well as the hot springs (0.58) and salt marshes species (0.27). These ratios indicate that the *psb*Z gene is possibly under purifying selection. The marine habitat is the only group with species with average ratios, but the group average is 0.68 which is still below the minimum cut-off and suggests purifying selection.

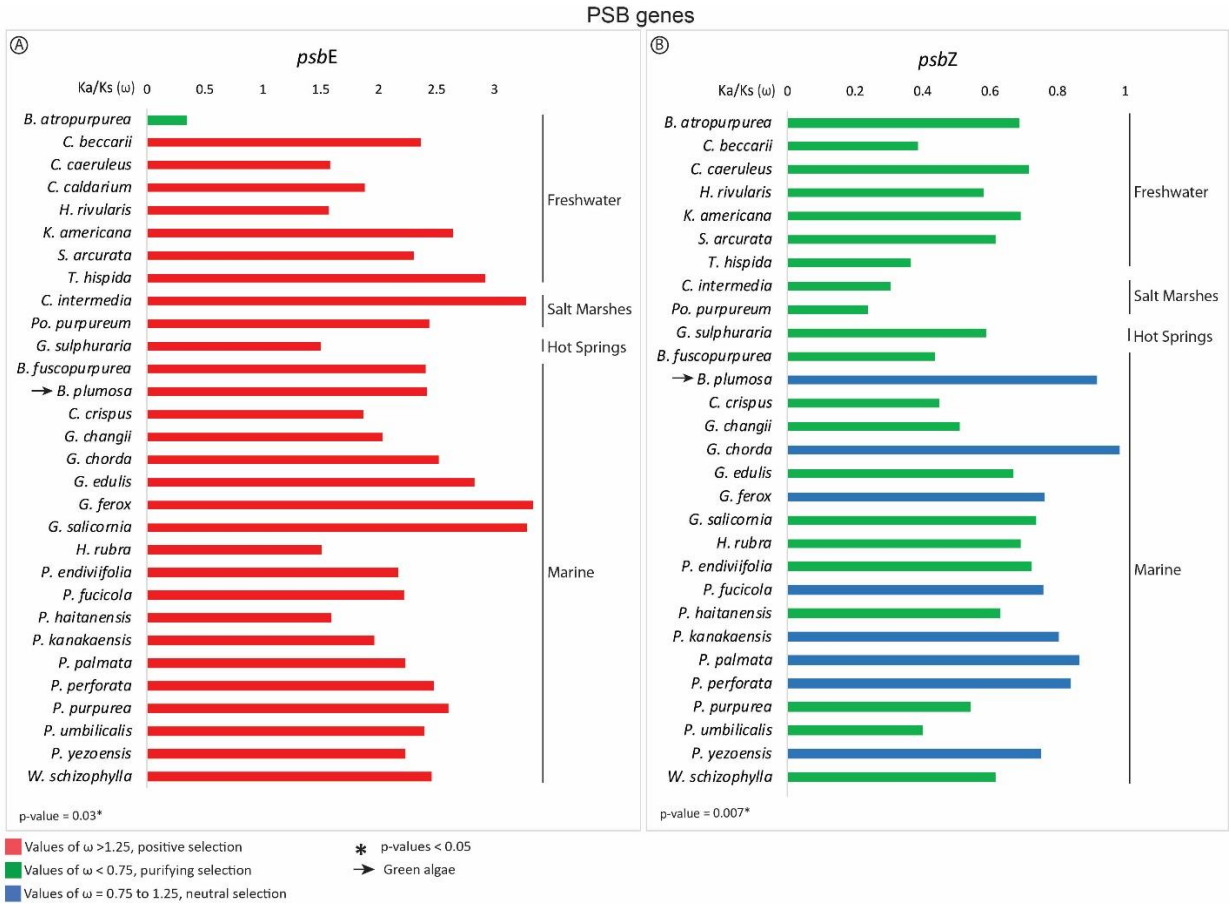

Figure 3. 10. Substitution ω ratios of red and green algae in comparison to *C. merolae psb*E and *psb*Z gene, and the ecological habitat of each species.

3.3.5.2 *psb* genes – Phylogeny

The comparison of ω ratios by species and their phylogenetic position show 4 genes with statistical support: *psb*A, *psb*B, *psb*K and *psb*V (Appendix F4 and F5). It is important to note the phylogeny of the concatenated sequences of 15 *psb* genes: (*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*K, *psb*L, *psb*N, *psb*T, *psb*V, *psb*W, *psb*X and *psb*Z) shown on Appendix Figure F5. Although being located in the expected place in the phylogenetic tree, closer to *Porphyra* species, the species *B. atropurpurea* is represented by a long branch.

Most species exhibit low ω ratios (< 0.75) for the *psb*A gene. Only 5 of the 30 compared species have neutral ω ratios (0.75 to 1.25), however, not all 5 are in the same lineage. The Cyanidiales and Bangiales orders represent the lowest ω ratios of the *psb*A gene (0.61 and 0.66), with all species in these orders with ω ratios below 0.75, which is strong evidence of purifying selection. The highest ω ratio was observed by the green algae *Bryopsis plumosa*, 0.98, indicating neutral selection (Fig. 9a). On the other hand, for the *psb*B gene, many species exhibit high ω ratios (Fig. 9b). However, the Bangiales is the only lineage with average ω ratios supporting positive selection (1.27), while all the other lineages have average ω ratios supporting neutral selection.

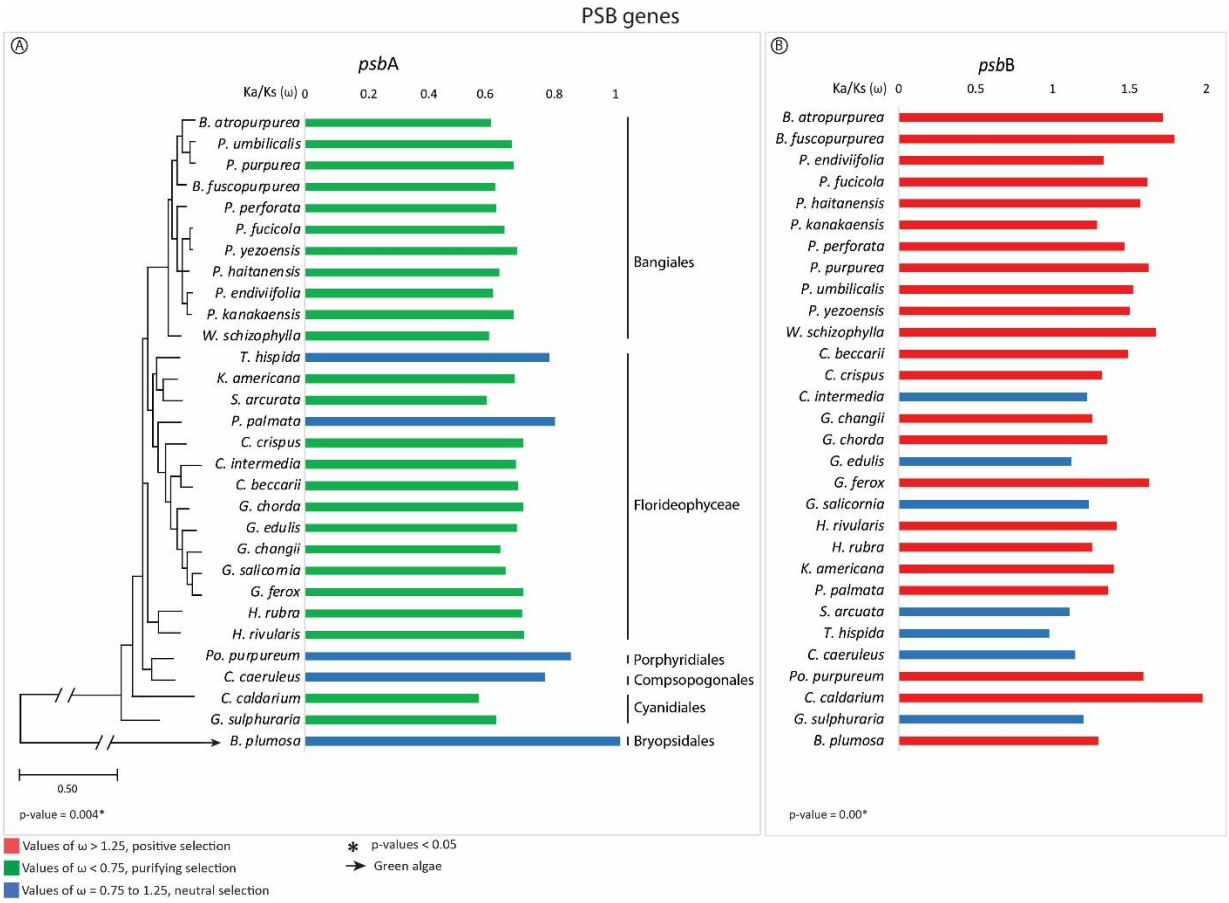

Figure 3. 11. Substitution ω ratios of red and green algae in comparison to *C. merolae psb*A and *psb*B gene, and then compared by the phylogenetic position of each species. The phylogenetic position of each species follows the *rbc*L gene tree.

For the *psb*K gene, the species have ω ratios that vary among the lineages. For example, all the Bangiales species that are in the first clade of the phylogenetic tree have low ratios and support purifying selection, while the other Bangiales species have neutral ratios and are indicative of neutral selection (Fig. 10a). The first clade includes the species *B. atropurpurea*, *P. umbilicalis*, *P. purpurea* and *B. fuscopurpurea.* Similarly, the *Gracilaria* species, part of the Florideophyceae, have neutral ω ratios, while the other species exhibit low or even high ratios. The ω ratio average of the Florideophyceae (0.89), supports neutral selection as well.

The Bangiales species stands out for its ω ratios for the *psb*V gene, which is very different from the other species (Figure 10b). Of the 29 species that have a copy of the gene, 4 Bangiales species exhibit high ω ratios, all *Pyropia*. The high ω ratios (> 1.25) indicate positive selection forces acting on the *psb*V gene of *Pyropia* species. Although *Pyropia* has high ω ratios,

the Bangiales as an order has an average of 1.21, indicative of possible neutral selection. All the other lineages have lower averages but also fall under possible neutral selection.

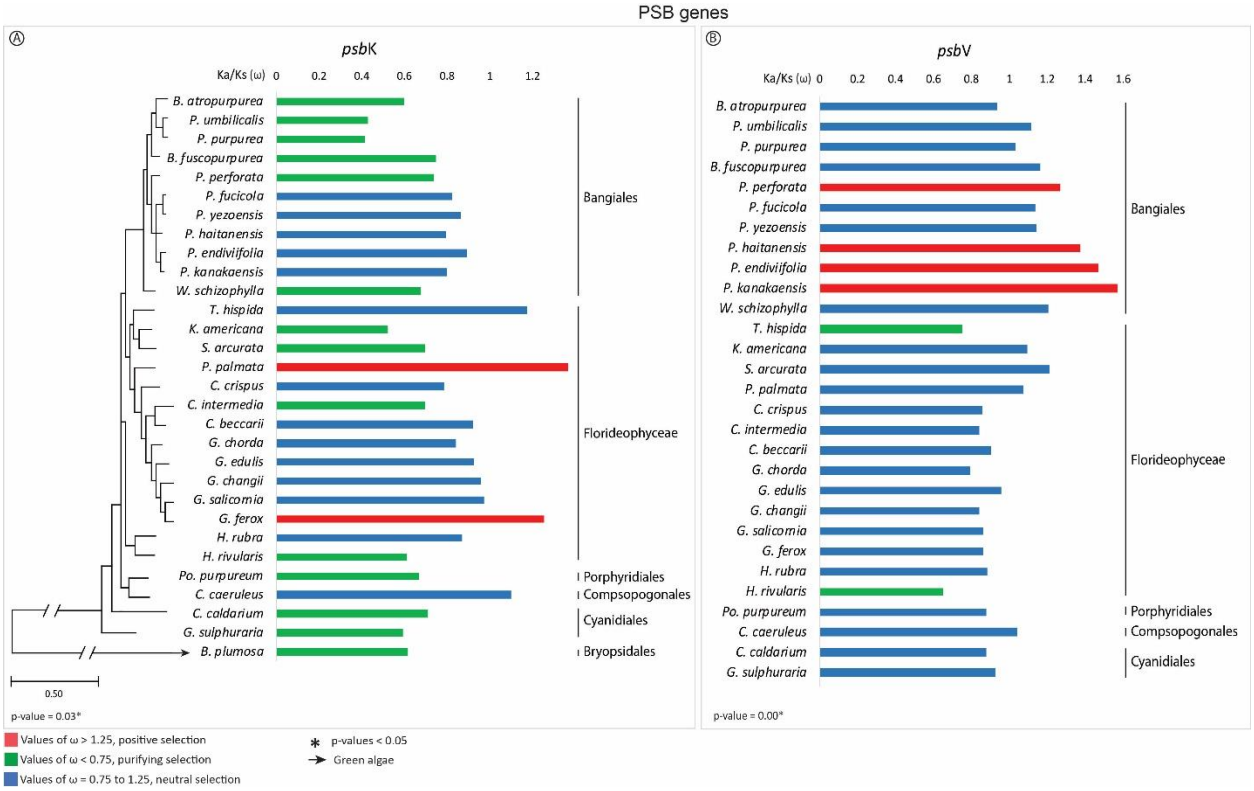

Figure 3. 12. Substitution ω ratios of red and green algae in comparison to *C. merolae* psbK and psbV gene, and then compared by the phylogenetic position of each species. The phylogenetic position of each species follows the *rbc*L gene tree. The green algae *B. plumosa* does not have a *psb*V gene copy, therefore, not used in this analysis.

The overall results of the substitution ratios for the *psb* genes compared by phylogenetic position shows that 37% of the average ω ratios are low (< 0.75), 34% are neutral (0.75-1.25) and 28% are high (> 1.25). This is a more linear distribution of ω ratios than what was observed in the *atp* and *psa* genes. Moreover, the average results of the habitat comparison are congruent to the phylogeny comparison, where a small number of genes have statistical support for differences between species. The *psb*C, *psb*N and *psb*W genes have low ratios among all the lineages, which is strong evidence of purifying selection forces acting in all Rhodophyta species. Contrary, the *psb*F has high averages in all lineages, which confirms the positive selection of the gene in Rhodophyta species.

3.3.5.3 *psb* genes – Discussion

Our analysis shows that 20% of the *psb* genes of the compared species are under selection forces based on the habitat they are usually found, represented by the *psb*E, *psb*H and *psb*Z. The *psb*E and *psb*F genes transcribe the subunits of the cytochrome b_{559}, α and β proteins respectively (Herrmann et al., 1984; Mor et al., 1995; Kiss et al., 2012). In all photosynthetic organisms studied up to date, the *psbE* and *psbF* genes form a cluster with the *psbJ* and *psbL* genes, known as *psb*EFLJ operon (Swiatek et al., 2003; Kiss et al., 2012). Studies suggest that the cytochrome b_{559} is essential for the PSII assembly (Tae & Cramer,1992; Mor et al., 1995; Swiatek et al., 2003), and the genes play a role in protecting the PSII from photoinhibition by providing the PSII with an alternative cyclic route of electron flow (Thompson & Brudvig, 1988; Nedbal et al., 1992; Alizadeh et al., 1994). We identified that the *psb*E gene in Rhodophyta species is under positive selection and is highly conserved, with exception of *B. atropurpurea,* which has a substitution rate that supports purifying selection. This implies that the α protein of the cytochrome b_{559} is unique in *B. atropurpurea,* and consequently, the *psb*EFLJ operon as well. Thus, *B. atropurpurea* has a different repair system for its PSII and consequently a distinct mechanism to avoid photoinhibition.

Photoinhibition is the decline in photosynthetic viability due to excessive illumination and it occurs in all oxygen-evolving photosynthetic organisms exposed to light (Adir et al., 2003). The process is determined by the balance between the rate of photodamage to the PSII and the rate of its repair (Murata et al., 2007). Environmental stress factors can enhance the extent of photoinhibition, such as salinity, low temperature, light levels and oxidative stress, by inhibiting the repair of the PSII (Murata et al., 2007; Nishiyama & Murata, 2014; Campbell & Serôdio, 2020). The species *B. atropurpurea* is extremely adaptive, striving in niches where many stress factors are present, such as different accessibilities of light, as well as variable temperatures and salinity levels (Geesink, 1973; Sheath & Cole, 1980; Graham & Graham, 1987). The intertidal marine habitat, where all the closest related species of *B. atropurpurea* are observed, is extremely different from the Great Lakes, therefore, the marine species face different environmental stressors and different repair systems are needed to control photoinhibition. Since purifying selection is a mechanism on which there is a purging of changes that cause deleterious impacts on the fitness of the host (Buckling et al., 2006), the

purifying selection in the *psb*E gene of *B. atropurpurea* might be linked to the adaptations needed to survive in the Great Lakes despite the environmental stress factors. Until now, there were no studies that help elucidate how does the photoacclimation *of B. atropurpurea* takes place on a molecular basis. Although more research about this topic is needed, the purifying selection forces on the *psb*E gene shed light on what can be a novel PSII mechanism connected to the fitness of *B. atropurpurea*.

The *psb*H gene is located in another highly conserved gene cluster including *psb*B-*psb*T-*psb*H-*pet*B-*pet*D, and the gene product is essential for the PSII assembly (Minagawa & Takahashi, 2004; Iwai et al., 2006). No literature explains more about the role of the *psb*H gene in Rhodophyta. Our data shows that all 30 Rhodophyta species have a copy of this core gene, with substitution rates that have a statistical difference between species from freshwater and marine environment to the species from salt marshes and hot springs. Although the statistical support, the substitution ratios indicate that all species have neutral selection forces acting on the *psb*H gene, therefore, evolving neutrally throughout the species.

The *psb*Z gene, first identified as *ycf*9, encodes a protein that is part of the stabilization of the PSII supercomplex (Swiatek et al., 2001; Minagawa & Takahashi, 2004). Our results show that the substitution rates of the *psb*Z gene are under purifying selection for 21 of the 29 compared species. The species from salt marshes have a significantly lower ratio (almost 50% lower) than the species observed in other habitats. The species *Porphyridium purpureum* has been a model organism since the 1980s for its systematics and photosynthetic apparatus (Gantt, 1980; Ley & Butler, 1980, Levy & Gantt, 1988; Levy & Gantt, 1990), but not much is known about its rates of substitution. The only *psb* gene analyzed in *P. purpureum* is the *psb*D gene (Scherer et al., 1993). As well, there is little information about other species observed in salt marshes used in our study, *Caloglossa intermedia,* which the only substitution rate study made use of only one gene - the nuclear-encoded large-subunit ribosomal RNA (LSU rDNA- to help determine the species phylogeny (Kamiya et al., 2003). There are no specific studies about the *psb*Z gene in Rhodophyta species to the present date, however, many substitution rates analyses in Viridiplantae species agree with our results for purifying selection forces on the *psb*Z gene (Choi & Park, 2015; Yu et al., 2019; Du et al., 2020). Still, more investigation is needed to determine how the purifying selection of the *psb*Z gene is connected to the stabilization of the PSII.

The comparison of ω ratios by species and their phylogenetic position shows statistical support in 27% of *psb* genes, for the *psb*A, *psb*B, *psb*K and *psb*V genes. For the *psb*A gene, Bangiales, Cyanidiales and Florideophyceae species have substitution ratios that indicate purifying selection, which is significantly different from Porphyridiales and Compsopogonales. The *psb*A gene encodes the D1 protein, and together with the D2 protein (*psb*D gene), forms the reaction center core that binds most of the PSII electron transport components (Adir et al., 2003; Bittner et al., 2011; Wegener et al., 2015). Due to its important role, the *psb*A gene has been extensively studied among many organisms and is considered a marker gene that can help solve phylogenetic relationships (Yang & Boo, 2004; Nelson et al., 2015). In Rhodophyta species, the *psb*A gene is used simultaneously with the *psa*A, *rbc*L and SSU genes in multi-gene sequences analyzes for taxonomic studies. For example, in the evolutionary history of the *Corallina* species (Bittner et al., 2011; Brodie et al., 2021); to identify lineages of *Griffithsia* (Ceramiaceae) (Yang & Boo, 2004); and the divergence time and the evolution of major lineages in the Florideophyceae (Yang et al., 2016; Yang et al., 2020).

Research about the *psb*A gene in Viridiplantae species has been done more in-depth, not only for taxonomy but also for describing its nucleotide substitution rate pattern. According to Wolfe et al. (1989), in a study on Solanaceae and Brassicaceae species, the *psb*A gene is identified as a low ω ratio gene. A different study confirms that the *psb*A gene has the lowest rate of synonymous substitution among plant chloroplast genes, suggesting that purifying selection is acting on this gene (Morton, 1997). In agreement, Li et al. (2016), which studied the molecular evolution of ferns, show that the *psb*A gene has a significant 2- to 3-fold deceleration in their substitution rate among species. A study with Cryptophyte algae species, done by Yang et al. (2020), tested the *psb*A gene for possible selection pressure, which resulted in a very low synonymous substitution rate, like the results of our analyzes. Despite the extensive use of the *psb*A gene in many taxonomic and evolutionary studies, no literature focus on the nucleotide substitution rates of the gene in Rhodophyta.

For the *psb*B gene, our results show that Rhodophyta species have high substitution ratios, with Bangiales, Porphyridiales and Cyanidiales lineages supporting positive selection, and the Florideophyceae and Bryopsidales exhibiting significantly lower ratios and supporting neutral selection. The *psb*B gene encodes the CP-47 protein, a major intrinsic protein, which constitutes an important operon along with the *psb*H, *pet*B and *pet*D genes (Johnson & Schmidt,

1993; Bricker et al., 1998; de Santana Lopes et al., 2018; Khattab et al., 2021). Many authors describe the *psb*B gene as a highly conserved gene, which allows adequate inference of relationships at deep phylogenetic levels (Graham & Olmstead, 2000; Sanderson et al., 2000; Magallón & Sanderson, 2002). However, like the *psb*A gene, not many substitution ratios analyses were made with the *psb*B gene, especially with Rhodophyta species. Hu et al. (2019), in a study with Chlaminomonales species (green algae), argues that the *psb*B gene is under positive selection, contradicting previous results that show the *psb*B gene as being the lowest synonymous substitution rates among single-copy plastid genes (Olmstead & Palmer, 1994). Moreover, de Santana Lopes et al. (2018) analyzed the selection forces on the macaw palm (*Acronomia acueata*) chloroplast genes, and also described the *psb*B gene as being under positive selection. Other studies with Brassicaceae and grasses (both Viridiplantae species) have the same results of positive selection (Hu et al., 2015; Piot et al., 2017), and our results agree with these studies. More analyses about the substitution rates of the *psb*B gene in Rhodopophyta species are needed to better understand the selective forces acting on the gene and to clarify why sister groups like Bangiales and Florideophyceae have different substitution rates for a highly conserved gene.

Our results for the *psb*K gene show that the Florideophyceae and Compsopogonales are distinct from the other lineages for exhibiting high ω ratios, indicating that the gene is under positive selection for its species. The Porphyridiales, Cyanidiales and Bryopsidales have the opposite trend, with ω ratios supporting purifying selection. The Bangiales is the only lineage that the average substitution ratio indicates neutral selection, although half of the analyzed species ratios support purifying selection. The available literature about the *psb*K gene and its low molecular weight protein is limited when compared to the *psb*A and *psb*B genes. It is known that the gene is part of the structure assembly and stabilization of the PSII (Koike et al., 1989; Takahashi et al., 1994; Sugimoto & Takahashi 2003; Minagawa & Takahashi, 2004), and although the gene product is needed for optimal functions, it is not essential for the PSII activity (Ikeuchi et al., 1991; Löffelhardt & Bohnert, 1994; Rochaix, 1997). Studies suggest that the *psb*K gene is under positive selection pressure in many different species, such as in palm cultivars (Enan & Ahmed, 2016), in the tree species *Robinia* (*Yu et al., 2019) and Calligonum (*Duan et al., 2020), and the flower *Echinacanthus* (Gao et al., 2019). The *psb*K gene is highly conserved in some lineages and was proposed as a marker gene to construct reliable phylogenetic

trees (Enan & Ahmed, 2016; Li et al., 2018). Contradicting these results, oak (*Quercus* sp.), and a leguminous species (*Vicia sepium*) have their *psb*K under strong purifying selection (Yin et al., 2018; Li et al., 2020). Therefore, the *psb*K gene in Viridiplantae has a similar pattern to the one observed in Rhodophyta, with lineages varying between positive and purifying selection. There are no reviews that indicate the *psb*K gene as evolving neutrally as suggested by our results for the Bangiales, and although the gene is conserved in the repertoire of red algal chloroplast genomes (Ng, et al., 2017), this is the first study that shows the difference of adaptive evolution of the gene among the reds.

Our data reveal that the *psb*V is under neutral selection in all lineages, but there is a considerable difference between the Bangiales and the other orders. The extrinsic protein encoded by the *psb*V gene, also called cytochrome c_{550}, is found in cyanobacteria and red algae, with function linked to the optimization of oxygen evolution (Enami et al., 1998; Ohta et al., 2003; Bricker et al., 2012). The gene is not present in higher plants evolution (Bricker et al., 2012), and our data shows that the green algae *Bryopsis plumosa* does not have a copy for the gene as well. Despite the extensive structural analyses about the *psb*V gene from different cyanobacteria (Frazão et al., 2001; Sawaya et al., 2001; Kerfeld et al., 2003), there are no studies that report the adaptive evolution of the gene. For Rhodophyta, a study done by Kimura et al. (2002) demonstrates that the removal of the *psb*V gene from *Porphyra* species results in development failure to cellular thermotolerance when grown at moderately high temperatures. This might be the reason why there are high substitution ratios in Porphyra species for the *psb*V gene.

It is known that the PSII reaction centre shares several structural features, indicating ancient evolutionary links (Sadekar et al., 2006), and our analysis agrees with this statement, where most Rhodophyta species share the same genes. Furthermore, the substitution ratio analyses of Rhodophyta species show that almost 50% of the PSII genes are under positive selection (*psb*B, *psb*E, *psb*F, *psb*L, *psb*T and *psb*X), contributing to the conservation of the genes. Similarly to what was observed within the *atp* gene family, the *psb* concatenated sequence phylogeny shows *B. atropurpurea* in a long branch, corroborating with the possibility of the organism to be under rapid evolution and/or sequence saturation. The available literature about the photosynthetic apparatuses in red algae includes the species *Cyanidium caldarium, C. merolae, Galdieria sulphuraria, Porphyridium cruentum*, and *Pyropia yezoensis* (Wolfe et al.,

1994; Busch et al., 2010; Gardian et al., 2007; Thangaraj et al., 2011; Zheng et al., 2020); in addition to a physiological study that focuses on the chloroplast pigmentation of *P. umbilicalis* (Figueroa et al., 1995). The structure of the PSII is, therefore, well-investigated, however, the adaptive evolution of the genes that build the structure is yet to be analyzed in more depth.

3.3.6 *rps* genes

Of the 19 *rps* genes commonly present in Rhodophyta species, 15 were analyzed for its ω ratios: *rps*2, *rps*4, *rps*5, *rps*6, *rps*7, *rps*9, *rps*10, *rps*11, *rps*12, *rps*13, *rps*14, *rps*16, *rps*17, *rps*18, *rps*20. Although most species have a copy of all the 15 genes, some species do not have all of the selected genes, for example, *Wildemania schizophylla* does not have a copy of the *rsp*20 gene; *Palmaria. palmata rps*17 gene; *Compsopogon caeruleus rps*5 and *rps*8 genes; *Cyanidium caldarium rps*1 gene. The green algae, *Bryopsis plumosa*, is the most different species in terms of gene content, missing a copy of the *rps*1, *rps*5, *rps*6, *rps*10, *rps*13, *rps*16, *rps*18 and *rps*20 (Appendix Table G1).

The individual species ratio analyses show the prevalence of low ω ratios (57%), more than what was observed in all the other gene families analyzed in this study. Of the 15 *rps* genes, 8 species have 10 genes each with ratios below 0.75 (*Bangia atropurpurea, Pyropia perforata, P. fucicola, P. haitanensis, W. schizophylla, Chondrus crispus, Caloglossa intermedia* and *Hildenbrandia rubra*). This is strong evidence that the *rps* complex genes are under purifying selection in these species. Neutral ratios were observed in all species, with exception of *C. caldarium*, which does not have any genes within neutral values.

3.3.6.1 *rps* genes – Habitat

The ω ratios comparison by species and their natural habitat had statistical support in five of the 15 *rps* genes: *rps*2, *rps*4, *rps*10, *rps*14 and *rps*17. The freshwater species have ω ratios that indicate neutral selection and are the highest ratios of the *rps*2 gene. Curiously, the freshwater species *H. rivularis*, has a high ω ratio (> 1.25) and is possibly under positive selection (Appendix Figures G1 and G2). The species observed in salt marshes and the marine environment have low average ω ratios (< 0.75), which indicates purifying selection and are statistically different from the species present in the other habitats.

The freshwater habitat comprises the lowest ω ratios for the *rps*4 gene (0.66), which is support for purifying selection. Of all low ratios observed in *rps*4 gene, *B. atropurpurea* has the lowest one (0.27), which is almost 60% is lower than the other species with low ratios, being strong evidence for purifying selection forces acting in this species gene. The species observed in salt marshes (0.75) and hot springs (0.78) are possibly under neutral selection, and although the

species found in the marine environment also have support for neutral selection, they are statistically different (0.86) (Fig. 14a).

Purifying selection forces can also be suggested for the *rps*10 gene, where all species have ω ratios is lower than 0.75. There are, however, two species with neutral ratios, the species observed in freshwater *B. atropurpurea*, and the species observed in the marine environment, *G. chorda*, which indicates a possible neutral selection happening on the *rps*10 gene of these species. Although most species are under purifying selection, the species observed in freshwater exhibit the highest ratios (0.56) and are statistically different from species in different habitats. There is no statistical difference between the ratios of species observed in salt marshes (0.39) and species observed in the marine environment (0.35).

The *rps*14 gene has a different pattern from what was observed previously on the RPS genes, where most species have high ω ratios (> 1.25). The spices found in the marine environment have the highest average ω ratios (1.99), which is strong support for positive selection (Fig. 14b). The species observed in freshwater have high and neutral ω ratios, but curiously, the species *B. atropurpurea* is the only species with a low ω ratio (0.31) for the *rps*14 gene, which is strong support for purifying selection. On the other hand, the marine species that share the same genus, *B. fuscopurpurea*, has the highest ratio not only for the *rps*14 gene but is the highest ω ratio of all the *rps* gene family (3.26).

For the *rps*17 gene, the freshwater species *B. atropurpurea* also has a peculiar ω ratio, this time being high (1.71), and possibly being under positive selection (Fig. 14c). All the other species have low or neutral ω ratios for the *rps*17 gene, where the neutral is observed predominantly in freshwater species. This factor resulted in the average ω ratios of the freshwater habitat being higher than the other habitats (0.8), indicating a possible neutral selection. The species found in salt marshes (0.5), hot springs (0.5), and marine environment (0.54) have similar low averages which are strong support for purifying selection.

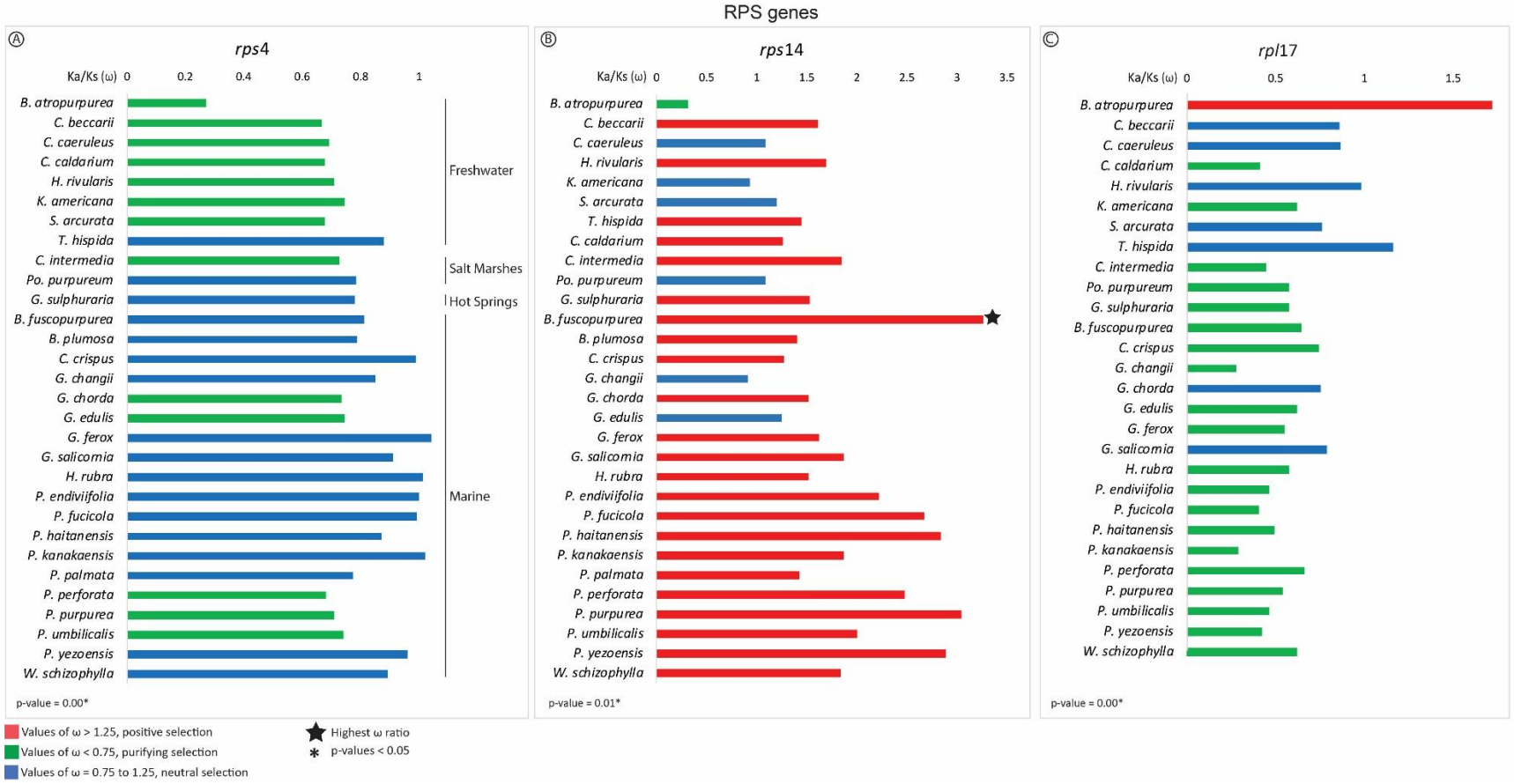

Figure 3. 13. Substitution ω ratios of red and green algae in comparison to *C. merolae rps*4, *rps*14 and *rps*17 gene, and then compared by the ecological habitat of each species.

Summarizing, when Rhodophyta species *rps* genes ω ratios were compared by habitat, the analyzes resulted in mostly low ratios. The species observed in the marine environment have 10 of the 15 *rps* genes with ratios below 0.75, which is strong support for purifying selection. Not so different, the species observed in freshwater and salt marshes, have 9 *rps* genes with low ratios. The most particular pattern was observed in the species found in hot springs, that have 7 genes with low ratios, and 5 genes with neutral ratios (0.75 to 1.25). These values indicate that like in the marine environment, a possible purifying selection can be happening within the *rps* genes of species in different environments.

3.3.6.2 *rps* genes - Phylogeny

The comparison of ω ratios by species and their phylogenetic position shows that five of the 15 *rps* genes have statistical support: *rps*2, *rps*5, *rps*11, *rps*14 and *rps*16. For the *rps*2 gene, which also exhibited statistical differences for ω ratios among different habitats, the phylogenetic comparison shows that the Bryopsidales and Florideophyceae have neutral ω ratios (1.01, 0.81), and are different from the other lineages. The Bangiales, Porphyridiales and Cyanidiales have ω ratios below 0.75, having strong support for purifying selection (Appendix Figures G3 and G4).

Although all lineages have low average ω ratios for the *rps*5 gene and possibly have purifying selection forces acting on the gene. There is a statistical difference (p-value < 0.005) between Bangiales and the other lineages, where the Bangiales have the lowest ω ratios average (0.26). Curiously, the species *B. atropurpurea* has a ω ratio that is almost double of the ω ratios observed in the other Bangiales species (0.59). The class Florideophyceae (0.38), the order Porphyridiales (0.39) and Cyanidiales (0.38) also have strong support to have purifying selection forces acting on this particular gene.

Most of the species exhibit high ω ratios for the *rps*11 gene (> 1.25) (Fig. 15a). All the species part of the Bangiales have high ω ratios and have the highest ω average ratio for the gene (1.73), which is strong evidence for positive selection. Once again, the species *B. atropurpurea* has a particular result and is the only species with a low ω ratio (0.4), which is indicative of purifying selection forces acting on the *rps*11 gene of this species. It can also be observed that 4 of the 5 *Gracilaria* species (Florideophyceae) have neutral ω ratios, and similar ω ratios are

noticed on the clade including the *Hildenbrangia* species, which is strong support for neutral selection forces.

The Bangiales also exhibit the highest ω ratios for the *rps*14 gene, with an average ω ratio of 2.31, which is almost double the average of the other lineages, which is strong evidence of positive selection. As noted in the habitat comparison, *B. atropurpurea* has a different pattern than the other Bangiales species regarding its substitution rates, exhibiting a low ω ratio (Fig. 15b). The low ω ratio of *B. atropurpurea* is one of the lowest registered for the entire *rps* gene family (0.3), which is strong support for purifying selection. Interestingly, the same genus species *B. fuscopurpurea* has the highest ω ratio (3.26), which is the highest substitution rate registered for the *rps* genes. The Rhodophyta species exhibit neutral and low ω ratios for the *rps*16 gene. Neutral ω ratios are predominantly observed on species that belong to the class Florideophyceae (1.07) and Porphyridiales (0.76). These lineages are statically different from Bangiales (0.64), Compsopogpnales (0.47) and Cyanidiales (0.58), which have support for purifying selection.

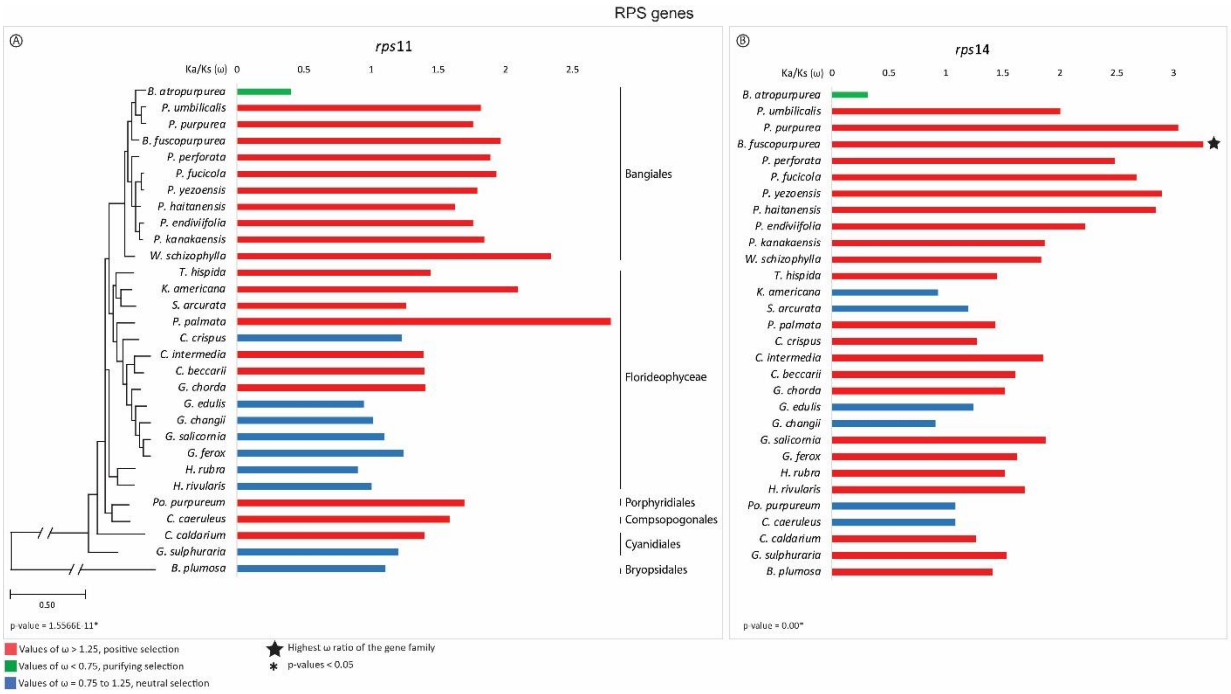

Figure 3. 14. Substitution ω ratios of red and green algae in comparison to *C. merolae rps*11 and *rps*14 genes, and then compared by the phylogenetic position of each species. The phylogenetic position of each species follows the *rbc*L gene tree.

In conclusion, the order Bangiales has the most particular pattern in substitution rates among the compared lineages. The order has 10 genes with low ω ratios (< 0.75), which is strong support for purifying selection forces acting in these species' *rps* genes. Our analyses show that the *rps* gene family has the lowest ω ratios in comparison to all the other gene families previously analyzed (*atp*, *psa*, *psb* and *rpl*). The average of the substitution rates by lineages shows that 55% of the genes are low (< 0.75), the neutral ratios (0.75 to 1.25) are present in 20% of the genes, and the remaining, 15%, correspond to genes with high ratios (> 1.25). These values are strong support for the *rps* genes being under purifying selection in Rhodophyta species.

3.3.6.3 *rps* genes – Discussion

The analyses of the substitution rates show that around 34% of *rps* genes in Rhodophyta are under different selection forces, for the habitat (*rps*2, *rps*4, *rps*10, *rps*14 and *rps*17) as well as for the phylogenetic position comparison (*rps*2, *rps*5, *rps*11, *rps*14 and *rps*16). It is worth noting that two genes, *rps*2 and *rps*14, have significant differences ($p < 0.05$) in both comparisons. The *rps*2 gene has been used by many plant research as a marker gene for phylogeny studies, analyzed alone or with other plastid genes like *rbc*L and *mat*K (DePahmphilis et al., 1997; Young et al., 1999; Tomari et al., 2002; Young & Claude, 2005; Park et al., 2008). Although considered a highly conserved gene in Viridiplantae species (DePamphilis et al., 1997; Young et al., 1999; Park et al., 2008), there is no information about the role of the *rps*2 gene. Larkum et al. (2007) noted that the cluster containing the *rps*2 gene and the *atp*A operon has not been reported from cyanobacteria but is present in multiple plastid lineages.

Our data agree that the *rps*2 gene is conserved since is present in all the 30 Rhodophyta species and green algae as well. Regarding the substitution rates, the *rps*2 gene has been categorized as one of the more rapidly evolving plastid genes (DePamphilis et al., 1997; Young & Claude, 2005; Park et al., 2008), therefore, under positive selection. Our analyses of the *rps*2 gene in Rhodophyta show the opposite result, as 21 of 29 species have low substitution rates, indicating purifying selection. In addition, species have significant differences (p-value < 0.05) in rates when compared by habitat and by phylogenetic position, adding to the support for purifying selection. As mentioned previously, differences in substitution rate patterns among

Viridiplantae and Rhodophyta can easily happen due to the evolutionary distance between the lineages.

The *rps*4 gene encodes the ribosomal protein S4, it is present in all plant sequenced plastid genomes, including the most reduced genomes (Bellot & Renner, 2016). Some knockout mutants' studies in tobacco show misshapen leaves if the *rps*4 gene is excluded, therefore, having an essential role in leaf development (Rogalski et al., 2008; Tang et al., 2018). The gene is considered a marker gene and used for many phylogenetic analyses in plants, successfully resolving taxonomic problems at the intra-familial level (Nadot et al., 1994; Souza-Chies et al., 1997; Horn, 2009; Chen et al., 2020). These studies have a nucleotide substitution rate analysis, but all data is concatenated, which does not express the results for the *rps*4 gene alone. At the moment, the only study that looked exclusively at the substitution rates of the gene suggests that the myco-heterotrophic plant has purifying selection forces acting on its *rps*4 gene (Yudina et al., 2021). Because of the large evolutionary distance between these plants and Rhodophyta species, the comparison of results is unjustified, and more investigations about the *rps*4 gene in red algae are needed to better understand why freshwater species have such distinct substitution rates.

The *rps*5 gene is important for plastid ribosome function and is involved in photosynthesis, plant development, and cold stress resistance in Arabidopsis (Zhang et al., 2016). According to Sugira (1992), the *rps*5 gene is not present in some of the completely sequenced chloroplast genomes of Viridiplantae species. There is limited research about the *rps*5 gene, and no report about substitution rates, which our analyses suggest being under purifying selection for Rhodophyta species. It is worth noting that the Rhodophyta species *Compsopogon caeruleus* and the green algae *Bryopsis plumosa* do not have a copy of the *rps*5 gene, and future analyses of other Compsopogonales species might confirm if the gene was lost in all order' species.

The plastid *rps*10 gene encodes the ribosomal protein S10 and is part of a highly conserved cluster of genes in Rhodophyta species which was studied in depth for three species: *Porphyra purpurea*, *Cyanidium caldarium* and *Cyanidioschyzon merolae* (Provan et al., 2004). Our analyses also show that all Rhodophyta species have a copy of the *rpl*10 gene. There are no studies at the present date, however, that discuss the substitution rates of the gene in Rhodophyta. In a study with *Cyanophora paradoxa*, a freshwater species of Glaucophyte, it was noted that the species does not have a copy of the plastid *rps*10 gene, which was confirmed to be a nuclear gene, like in Viridiplantae and green algae species (Neumann-Spallart et al., 1991). Because we

also used the green algae *B. plumosa* in our comparison, we confirm that the green algae do not have a copy of the *rps*10 gene. As mentioned previously, although green plants and red algae share a common ancestor, the evolutionary distance between the lineages is significant (Rodríguez-Ezpeleta et al. 2005; Gould et al. 2008), which can explain the plastid gene to have moved to the nucleus along the course of evolution.

The *rps*11 gene has been vastly used for phylogenetic studies within Viridiplantae species because is a conserved gene (Mahmood et al., 2011; Jabeen et al., 2012), however, to obtain a better result at an infrageneric level, the authors suggest the use of other marker genes alongside the *rps*11. On the other hand, a study based on *rps*11 gene shows that future studies can reliably use the gene for assessing genetic kinship among wheat cultivars and to possibly change the direction of breeding programs in future (Rehman et al., 2015). The only study with *rps*11 gene in Rhodophyta investigated the inheritance of chloroplasts genes in the marine red alga *Porphyra yezoensis*, which used the *rps*11 gene as a molecular marker (Choi et al., 2008). To date, the substitution rates of the *rps*11 genes are known only for a parasite of the red algae *Pterocladiophila hemisphaerica*, which shows increased selection when compared to the algae itself (Preuss et al., 2020). Our data shows that 20 Rhodophyta species are also under positive selection pressures for the *rps*11 gene, and curiously, *B. atropurpurea* is the only species with purifying selection pressures over the gene. For this peculiar reason, more investigations are needed to understand the mechanism behind the rps11 gene in *B. atropurpurea*, more in-depth analysis of the DNA sequence to observe different gene parts and see where is the change that differs so drastically from the other Rhodophyta.

The *rps*14 gene, which encodes ribosomal protein S14, was used to assess the phylogenetic relationship among Viridiplantae species such as date palm, plantago and citrus species (Akhtar et al., 2011; Saeed et al., 2011; Wali & Mahmood, 2013). All these studies identified the *rps*14 gene sequence as highly conserved among the species with a low genetic divergence, not providing much information for establishing phylogeny at a genus level. There are no studies to date regarding the substitution rates of the *rps*14 gene among Viridiplantae nor Rhodophyta. Our data agree with the gene being conserved among species and shows that 23 of the 30 compared species are under positive selection for the *rps*14 gene. Moreover, there are significant differences (p-value < 0.05) for the comparisons among lineages and habitats. The highest averages of substitution rates are observed for the Bangiales, which is statistically

different from the other lineages. Also, the species occupying the marine environment represent the highest average of substitutions, being represented mostly by the Bangiales species. Because of the strong positive selection within Bangiales species, future studies might benefit from the use of the *rps*14 gene associated with other marker genes to resolve phylogenetic relationships among species at an ordinal and familial level. The substitution rates of the *rps*14 gene in *B. atropurpurea* are very interesting, given that the species is the only freshwater Bangiales and is under strong purifying selection. More investigation is needed to understand why this organism gene is so different from the other red algae, and to verify what parts of the *rps*14 gene are being affected by the purifying selection, and what parts are under positive and neutral selection.

The *rps*16 gene encodes the S16 protein is considered essential for plant survival (Keller et al., 2017), and is present in most of the Viridiplantae species (Ueda et al., 2008). However, some species do not have a copy of the gene, and studies show that the loss of the chloroplast copy was compensated by having a mitochondrial *rps*16 (Ueda et al., 2008). A study with *Lupinus* species shows that the loss of the *rps*16 gene was most likely replaced by the nuclear *rps*16 gene (Logacheva et al., 2016; Keller et al., 2017). Our analysis shows that the green algae *B. plumosa* does not have a copy of the chloroplast *rps*16 gene, and further investigations might be able to answer if the species has a mitochondrial and/or nucleus copy of the gene that can substitute the chloroplast gene, like in some Viridiplantae species. The loss of the *rps*16 gene was not observed in any of the 30 Rhodophyta species compared in this study. Regarding the substitution rates of the gene, only one study was found, in Angiosperm species (Keller et al., 2017). The analyses revealed that both gene copies (chloroplast and mitochondria) are under purifying selection. Our data show the same results, where most Rhodophyta species are under purifying selection as well. The lineage that has a statistical difference (p-value < 0.05) from the others is the Porphyridiales, which is under neutral selection. Lastly, the *rps*16 gene is also considered a rapidly evolving gene, which can be used for phylogenetic studies, increasing resolution and statistical support when compared to previous studies that used only the *rbc*L gene (Mouly et al., 2009; Martirosyan et al., 2009; Schäferhoff et al., 2010). Since the *rps*16 gene is conserved in Rhodophyta, future taxonomic studies might benefit from using this gene as well.

In studies with Viridiplantae species, the *rps*17 gene appears not to be required for basal ribosome activity (Gong et al., 2013), and plants can complete their entire life cycle in their absence (Romani et al., 2012). An analysis of chloroplast genomes study shows that the *rps*17

gene is specifically derived from non-green algae (De Las Rivas et al., 2002), which justifies why the green algae, *B. plumosa*, in our study lacks the copy of the gene. To date, there are no studies about the *rps*17 gene in Rhodophyta, nor for rates of substitution. Our data shows that the gene is conserved among red algae. Strong purifying selection is acting on the *rps*17 genes of species observed in the marine environment, and organisms that occupy freshwaters have a significant difference (p-value < 0.05) in substitution rates from the other habitats, being under neutral selection. More research about the *rps*17 gene is needed for Rhodophyta species since is a conserved gene that was lost in green algae and might bring insights about red algae chloroplast evolution. The only Rhodophyta species that has a contradicting substitution rate is *B. atropurpurea*, which is under strong positive selection. As was noted for the *rps*11 and *rps*14 genes, the *rps*17 gene will need future investigations are needed to clarify why these genes are under selection pressures that are the opposite of other red algae.

The 15 *rps* genes analyzed in this study are conserved among Rhodophyta, while in the green alga species *B. plumosa* many gene copies are missing: *rps*1, *rps*5, *rps*6, *rps*10, *rps*13, *rps*16, *rps*18 and *rps*20. Given that green algae and Viridiplantae species are closer related than red algae, this was expected. Previous studies explain that many Viridiplantae species have a high deletion rate for chloroplast ribosome genes and proteins, which most likely, are taken on by genes encoded in the nuclear genome. (De Las Rivas et al., 2002; Gong et al., 2013). Although the *rps* genes encode proteins that are essential components of protein synthesis machinery with a vast range of roles in plant growth and development, the understanding of how the chloroplast proteins regulate these processes is not fully understood (Lumsden et al., 2010; Zhang et al., 2016), and the functions of some individual plastid ribosomal proteins remain largely unknown. The results from our analyses can be used as reference data for future studies regarding rates of evolution in Rhodophyta in hopes to increase the knowledge about these important genes and proteins.

3.3.7 *rpl* genes

As previously mentioned in the 3.3.6 section of this thesis, the ribosomal proteins are well-known for their role in ribosome biogenesis and protein synthesis, maintaining the stability of the ribosomal complex, which includes small (*rps*) and large (*rpl*) subunits (Moin et al., 2016; Saha et al., 2017). The expression of ribosomal protein genes has also been shown to be regulated by signalling molecules and environmental stresses (Warner & McIntosh, 2009; Sormani et al., 2011; Moin et al., 2016). Therefore, the ribosomal proteins genes substitution rates can help us understand if there is a connection between these chloroplast genes' evolution rates and the environment where Rhodophyta species inhabit.

The *rpl* genes have been studied in more depth using Viridiplantae species such as tobacco, rice and Geraniaceae species (flowering plant) (Guisinger et al., 2008; Schwarz et al., 2017; Sloan et al., 2012; Dong et al., 2013; Park et al., 2017; Weng et al., 2013), where the authors agree with accelerated substitution rates in the ribosomal protein genes. According to Ren et al. (2020), the diatom plastids also have higher substitution rates in their transcription-translation apparatus (*rps*, *rpl* and *rpo* genes) than to the ones involved in photosynthesis (*atp*, *psa*, *psb* genes). However, all these studies have made a selection of the *rpl* genes, not using the whole gene family. Moreover, the selected genes are different from the ones observed in Rhodophyta and the data were concatenated, not being able to serve as a reference to our work that is doing an individual analysis of each *rpl* gene.

From the 26 *rpl* genes, 24 were analyzed for its ω ratios: *rpl*1, *rpl*2, *rpl*3, *rpl*4, *rpl*5, *rpl*6, *rpl*11, *rpl*12, *rpl*13, *rpl*14, *rpl*16, *rpl*18, *rpl*19, *rpl*20, *rpl*21, *rpl*23, *rpl*24, *rpl*27, *rpl*28, *rpl*32, *rpl*33, *rpl*34, *rpl*35 and *rpl*36. More variance of gene content was observed for the *rpl* gene family in comparison to the *atp*, *psa* and *psb* gene families, where most species have a gene copy of all genes. The absence of some ***rpl*** genes resulted in the exclusion of some species from the substitution ratio analyses, for example, *B. plumosa* without a copy of *rpl*1, *rpl*3, *rpl*4, *rpl*6, *rpl*11, *rpl*13, *rpl*18, *rpl*21, *rpl*24, *rpl*27 and *rpl*28 genes; *C. caeruleus* without a copy of *rpl*5, *rpl*6, *rpl*13, *rpl*24, *rpl*27 and *rpl*34 genes; *G. sulphuraria* without a copy of *rpl*24 and *rpl*34 genes; the *rpl*4 gene is absent of *P. kanakaensis* and *P. perforata*; the *rpl*5 gene is absent of *P. fucicola; G. edulis* do not have a copy of the *rpl*14 gene and *S. arcuata* do not have a copy of the *rpl*34 gene (Appendix Table H1).

When examining the ω ratios of the species, it is noticeable that some have significantly lower ratios than others (p-value < 0.05). For example, the species *S. arcuata* has 12 of the 24 genes with ω ratios below 0.75, indicating that the species possibly have purifying selection forces acting on its ribosomal complex. Other species, such as *G. sulphuraria* (11 genes), *B. atropurpurea* and *G. changii* (10 genes), also have the majority of their *rpl* genes with low ratios. Neutral ratios genes (0.75 to 1.25) are observed in all species, but in *G. chorda*, *B. fuscopurpurea* and *P. palmata* the neutral ratios are predominant, with the first having 13 neutral ratios genes and the others with 12 neutral genes each. Ratios above 1.25 were observed in all species, with exception of *P. kanakaensis*, which does not have any *rpl* gene with high ω ratios. The species *Porphyridium purpureum*, *P. yezoensis* and *W. schisophylla* are the ones with more genes (9) with high ω ratios, indicating a possible positive selection on the *rpl* gene family of these species.

3.3.7.1 *rpl* genes – habitat

The comparison of ω ratios by species and their ecological habitat, show statistical support (p-value < 0.05) for 8 *rpl* genes: *rpl*1, *rpl*2, *rpl*5, *rpl*13, *rpl*19, *rpl*20, *rpl*24 and *rpl*27. Our analysis show that most red algae species have ω ratios under the neutral selection range (0.75 to 1.25), 5 species have high ω ratios (> 1.25) and 3 of these are species observed in freshwater: *C. beccarii*, *C. caeruleus* and *K. americana* (Appendix Figures H1, H2 and H3). The species observed in hot springs have the highest ratio of the *rpl*1 gene, 1.26, which is indicative of positive selection, being significantly different from the other habitats (p-value < 0.05). For the *rpl*2 gene, most species have low ω ratios (< 0.75), but 5 species have neutral values. From these 5 species, 2 are observed in freshwater, 2 are observed in the marine environment, and 1 is observed in hot springs. The average of the ratios among the habitats are similar: freshwater 0.66, salt marshes 0.69 and marine 0.63. Because of the low ratios, these are considered to be under purifying selection. The hot springs, however, with a ratio of 0.87, fall into the neutral selection range, being different from the other habitats.

For the *rpl*5 gene, the species found in freshwater exhibit a diverse range of ratios, where 3 species have neutral, 3 species have high and 1 species, *B. atropurpurea*, have low ω ratio (0.44). The species observed in the marine environment *P. perforata* is the only other species

with a low ω ratio (0.71), however, is still significantly different from *B. atropurpurea*. As ecological habitats groups, the average of freshwater, marine and hot springs follow under the neutral selection, which is significantly different (p-value < 0.05) from the salt marshes, considered to be under positive selection (1.35).

Most of the species have high ω ratios (> 1.25) for the *rpl*13 gene, with only 2 species with low ω ratios (< 0.75) (Fig. 11b). Like on *rpl*5, the species found in freshwater represent the most mixed environment, with 4 species have neutral, 2 species have high and 1 species have low ω ratios. The average ω ratios of the freshwater and salt marshes are similar, and both groups are possibly under neutral selection. These groups are significantly different from the marine environment and hot springs, which exhibit average ω ratios that support positive selection.

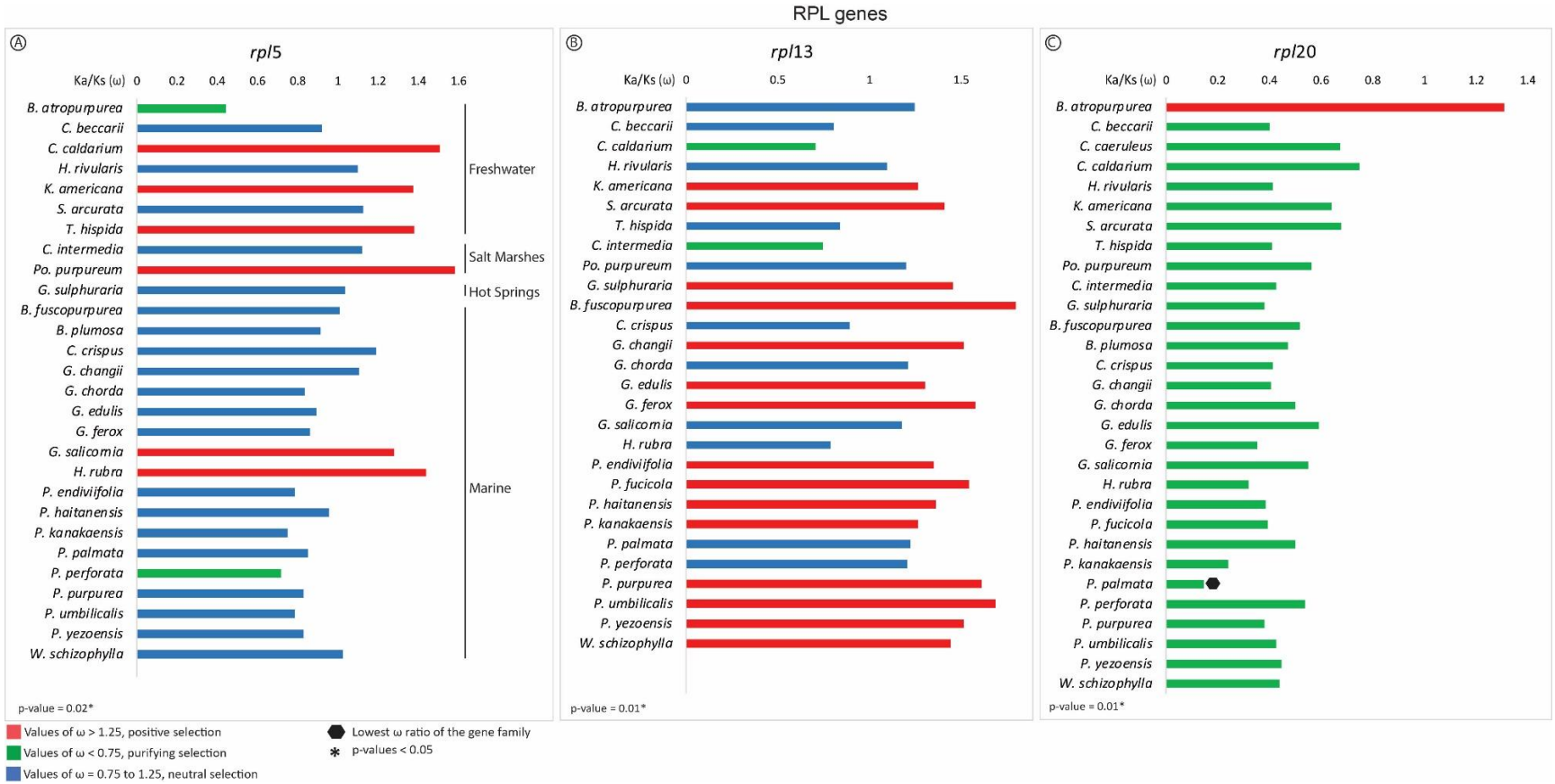

Figure 3. 15. Substitution ω ratios of red and green algae in comparison to *C. merolae rpl*5, *rpl*13 and *rpl*20 genes, and then compared by the ecological habitat of each species.

All but two species exhibit high ω ratios for the *rpl*19 gene, the species observed in freshwater *H. rivularis*, and the species observed in the marine environment *G. salicornia*, which have neutral ω ratios. Despite the average high ratios for all the compared habitats, which support positive selection, the freshwater and marine are statistically different from the salt marshes and hot springs. The opposite scenario is observed for the *rpl*20 gene (Fig. 11c), where all the species have low ω ratios (< 0.75) and are possibly under purifying selection.
The species observed in the marine environment *P. palmata* has the lowest ω ratio for the *rpl*20 gene of the entire *rpl* gene family. Contrasting with the low ratios, the freshwater species *B. atropurpurea* has a high ω ratio (1.3) which indicates positive selection. Statistically, the hot springs' average ω ratio is significantly different from the other habitats.

Neutral, high, and low ω ratios are observed for the *rpl*24 genes in species found in marine environments. There are no high ω ratios for the other habitats. The average ω ratios of the species found in the marine environment (0.96) are similar to the average of the species observed in salt marshes (0.8) and are possibly under neutral selection. The freshwater group has a lower average (0.74), which indicates a possibility of purifying selection. Most species have neutral ω ratios for the *rpl27* gene and support neutral selection, including the average ω ratios of the freshwater, hot springs and marine environment. However, the species observed in salt marshes exhibit high ω ratios, which indicates positive selection. Overall, the ω ratios averages within the RPL gene family among the different habitats, there are more species with neutral ratios than species with high or low ratios. The species observed in freshwater, for example, have 13 RPL neutral genes, against 7 low and 4 high ω ratios, indicating that species in this habitat are likely under neutral selection. The possible neutral selection is also observed for the salt marshes species, marine and hot springs.

3.3.7.2 *rpl* genes – phylogeny

The comparison of ω ratios by species and their phylogenetic position show that 12 genes have statistical support for significant differences among lineages: *rpl*2, *rpl*4, *rpl*5, *rpl*6, *rpl*11, *rpl*12, *rpl*13, *rpl*19, *rpl*21, *rpl*23, *rpl*24, *rpl*28 (p-value < 0.05). For the *rpl*2 gene, the Bangiales species have only low ω ratios, while two species part of the Florideophyceae class has neutral ω ratios (*Kumanoa americana* and *Gracilaria chorda*) and all the other species also have low ω

ratios. Both Cyanidiale and Briopsydale orders have only neutral ω ratios (0.75 to 1.25). The averages of the ratios among the lineages indicate that the Bangiales (0.6), Florideophyceae (0.64) and Compsopogonales (0.71) are possibly under purifying selection. The average of the Cyanidiales (0.81) and Bryopsidales (0.78), have support for being under neutral selection (Appendix Figures H4, H5 and H6).

For the *rpl*4 gene, the Bangiales have six out of nine species with high ω ratios (Fig. 12a). Curiously, the high ω ratios are exclusive by the second clade species of the Bangiales and for *Bangia fuscopurpurea*. The species part of the first clade of the Bangiales has neutral ω ratios. All the other orders are predominantly neutral, with exception of one Floridiophyceae species, *Caloglossa intermedia*, and the Porphyridiales species, *Porphyridium purpureum*. Interestingly, both of these species are observed in salt marshes, but when the ω ratios for the *rpl*4 gene were compared by habitat it did not reveal a significant difference (p-value > 0.05). The averages of the Florideophyceae (1.05) and Compsopogonales (1.23) species indicate that both groups are possibly under neutral selection, while the averages of Bangiales (1.38) and Porphyridiales (1.41) have support for positive selection.

The *rpl*5 gene ω ratios show that the Bangiales order is the only order with species that exhibit low ω ratios: *B. atropurpurea* and *Porphyra perforata* (Fig. 12b). All the other species in Bagiales have neutral ratios. Species with high ratios are spread out in all other orders, along with species with neutral ω ratios. The average ω ratios of Bangiales (0.81), Bryopsidales (091), and Florideochyceae (1.1) show support for being under neutral selection, while the average ω ratios of Cyanidiales (1.27) and Porphyridium (1.58) are higher and indicate positive selection. Only low and neutral ω ratios are observed for the *rpl*6 gene. The Floridiophyceae species have ω ratios that are visibly is lower than the other lineages, and its average ω ratio (0.56) indicates that the group is under purifying selection, along with the Bangiales (0.7). The Porphyridiales and Cyanidiales exhibit average ω ratios that support neutral selection.

By observing the *rpl*11 gene graph (Fig. 12c), it is evident that the Bangiales is the order has mostly high ω ratios. Similar to the *rpl*4 gene results, the Bangiales first clade has a different trend from the second clade, with neutral ω ratios, while the other species of the order have high ω ratios. The only exception is the species *B. fuscopurpurea*, which is part of the first clade but has a high ω ratio. The Florideophyceae class also has species with high ω ratios but most species have neutral ω ratios. Differently, the Porphyridiales species have a low ω ratio and have

support to be under purifying selection for the *rpl*11 gene. The average of the ω lineages shows that Florideophyceae (1.14), Compsopogonales (1.15) and Cyanidiales (0.97) are possibly under neutral selection. The Bangiales average ω ratios show evidence for positive selection.

The ω ratios for the *rpl*12 gene are mostly low (< 0.75), with exception of the Porphyridiales, which exhibit high ω ratios and have support for positive selection, being statistically different from the other lineages (p-value < 0.05). The ratios observed in the Bangiales (0.72), Compsopogonales (0.64), Cyanidiales (0.69) and Bryopsidales (0.64) have an average ω ratio that supports purifying selection. Moreover, similarly to the *rpl*6 gene ω ratios, the Florideophyceae species have significantly lower ratios in comparison to the other lineages (0.55), being the lowest average for the *rpl*12 gene and also under purifying selection.

For the *rpl*13 gene, high ω ratios are exhibited for the Bangiales species, where only two species have neutral ratios: *B. atropurpurea* and *P. perforata* (Fig. 12d), therefore, the *rpl*13 gene is considered to be under positive selection for Bangiales species. Two species with low ω ratios, indicative of purifying selection, are observed, one being part of the Florideophyceae class, *C. intermed*ia, and another species that belongs to the Cyanidiales order, *C. caldarium*. Curiously, the other Cyanidiale species analyzed in our study, *G. sulphuraria*, has a high ω ratio, indicating positive selection. The Floridiophyceae and Porphyridiales have ω ratios averages that support neutral selection.

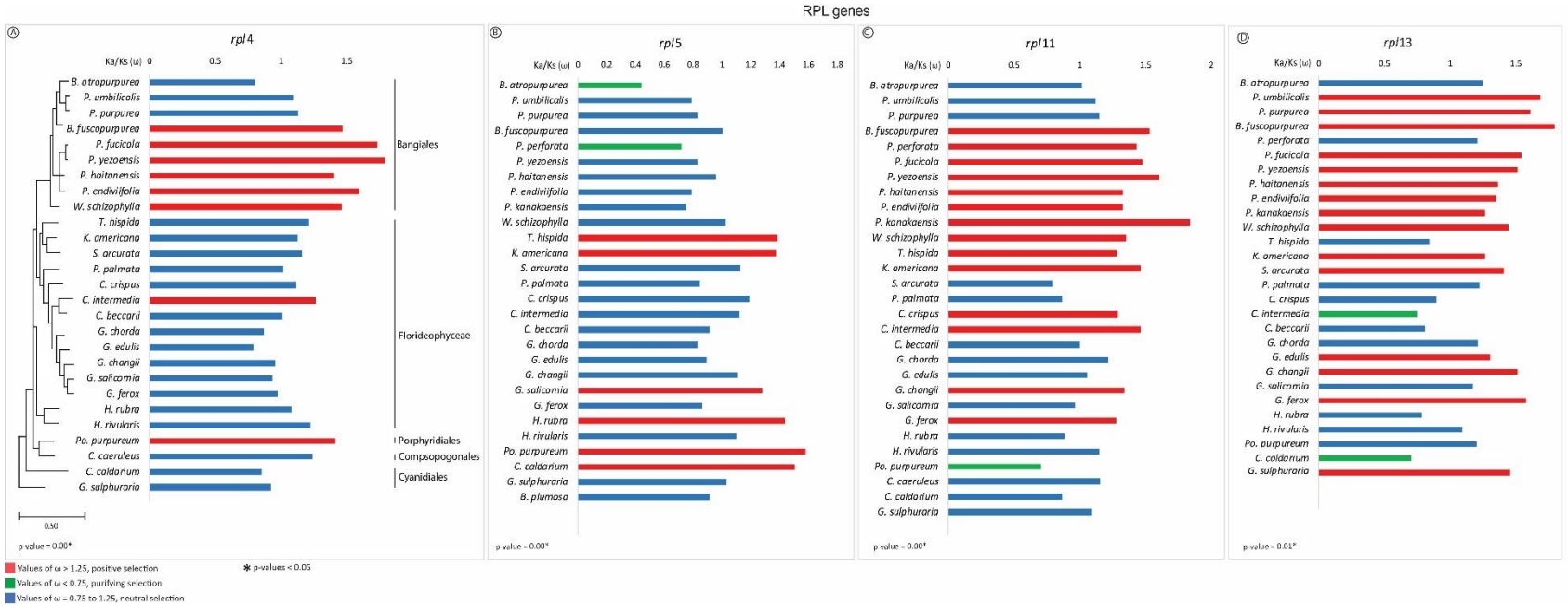

Figure 3. 16. Substitution ω ratios of red and green algae in comparison to *C. merolae rpl*4, *rpl*5, *rpl*11 and *rpl*13 gene, and then compared by the phylogenetic position of each species. The phylogenetic position of each species follows the *rbc*L gene tree.

All species exhibit high ω ratios for the *rpl*19 gene, with exception of two species that have neutral ratios, *G. salicornia* and *H. rivularis*. Although the average ω ratio for all lineages supports positive selection, there is a statistical difference (p-value < 0.05) between the Florideophyceae (1.58) and all the other lineages (~1.91). Differently, the species exhibit low, neutral and high ω ratios of the *rpl*21 gene (Fig. 13a), and the statistical difference is between the Bangiales and the other orders. By observing the graph, of the 11 Bangiales species, 9 have low ω ratios (< 0.75) and two are neutral (0.75 - 1.25). This results in a low average for the Bangiales (0.7), which indicates a possible purifying selection acting on the *rpl*21 gene. The ω ratios average of the Floridiaophyceae, Porphyridiales, Compsopogonales and Cyanidiales is similar and supports neutral selection.

The Bangiales order also shows a different ω ratio trend for the *rpl*23 gene, with most species with ratios supporting neutral selection. However, *P. yezoensis* has a high ω ratio (positive selection) and *B. atropurpurea* exhibits a low ω ratio (purifying selection) (Fig. 13b). The Floridiophyece species exhibit mostly low ω ratios, similar to the *B. atropurpurea* ratios, which is indicative of purifying selection. Both species of the genus *Hildenbrandia* have high ω ratios, which indicates that species within this genus might be under positive selection. The Cyanidiales and Bryopsidales also have high ω ratios (> 1.25), and both orders have support to be under positive selection.

The Bangiales is also particular for the rpl24 gene ω ratios, being the only order with two species exhibiting high ω ratios, *P. fucicola* and *W. schizophylla* (Fig. 13c). Moreover, *B. atropurpurea* has the lowest ω ratio for the *rpl*24 gene within the Bangiales. These species have also the lowest ω ratio whiting the Bangiales for the *rpl*23 and *rpl*35 genes. Overall, the ω ratios average for the Bangiales supports neutral selection. For Floridiophyceae, 8 species have low ω ratios, similar to *B. atropurpurea*, but the ω ratios average suggest the class be under neutral selection. The Porphyridiales is the only order with the low ω ratio average, indicating purifying selection. The species *C. caeruleus* (Compsopogonales), *B. plumosa* (Bryopsidales) and *G. sulphuraria* (Cyanidiales) do not have a copy of this gene and were excluded from this analysis.

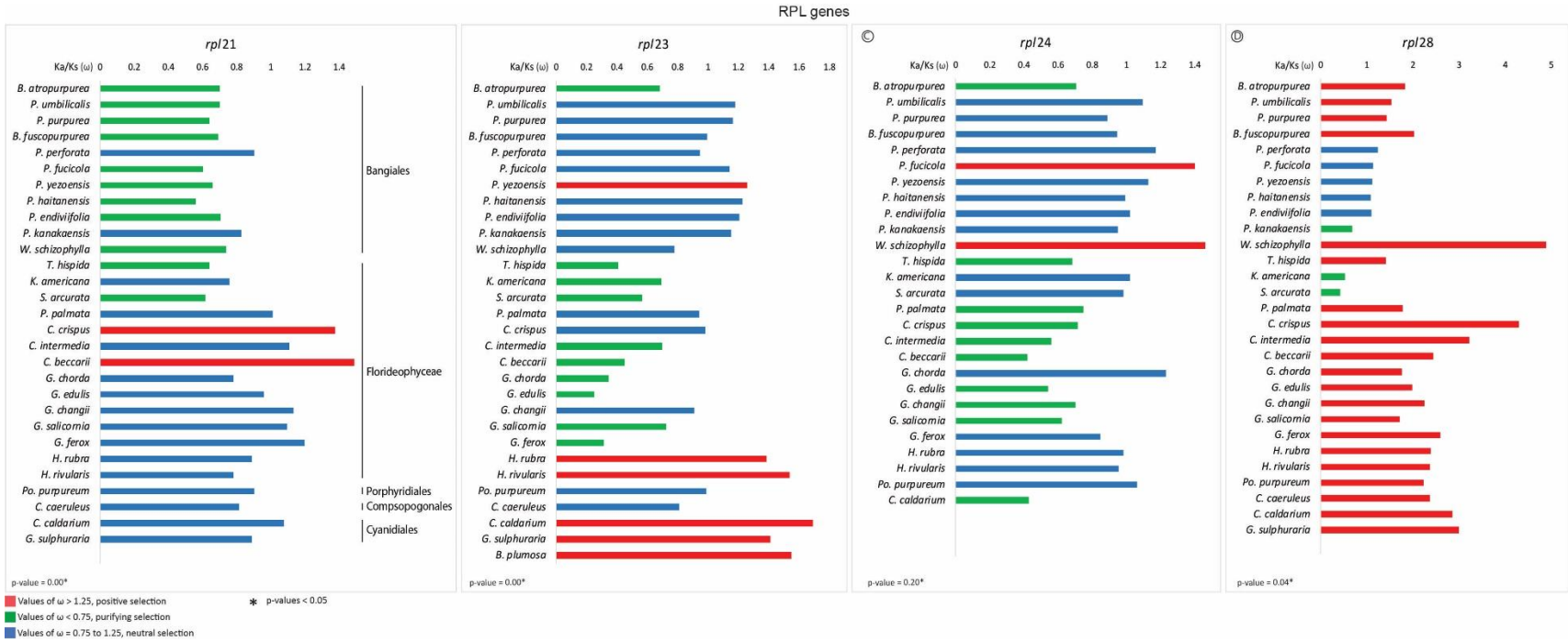

Figure 3. 17. Substitution ω ratios of red and green algae in comparison to *C. merolae rpl*21, *rpl*23, *rpl*24 and *rpl*28 gene, and then compared by the phylogenetic position of each species. The phylogenetic position of each species follows the *rbc*L gene tree.

Distinct ω ratios pattern is again observed in the Bangiales for the *rpl*28 gene. The order is divided by the species on the top clade with high ω ratios and the bottom clade species with low, neutral and high ω ratios (Fig. 13d). Interestingly, the species of the genus *Pyropia* have neutral ratios, with exception of *P. kanakaensis*, which has a low ω ratio, which indicates that neutral selection can be happening for this species. Overall, the Bangiales average ω ratios support positive selection, but due to the neutral ratios observed for Pyropia species, the order is statistically different from the other lineages. The Florideophyceae, Porphyridiales, Compsopogonales, and Cyanidiales exhibit similar high ω ratios, with no significant difference among them (p-value > 0.05), and are possibly under positive selection as well.

In summary, the comparison of the ω ratios of Rhodophyta species organized by their phylogenetic position, show that the RPL gene family have 45% of the genes ratios with support for being under neutral selection, 29% are considered low and indicate purifying selection, and 26% of the genes have high ratios, supporting positive selection. The Florideophyceae and Cyanidiales have a similar ω ratios pattern, where the number of neutral ratios genes is 13 and 14, respectively, which indicates that both groups are under possible neutral selection. The Bangiales have statistical support for a different ω ratios pattern in 6 of the 12 *rpl* genes, being distinct from the other lineages. Moreover, the species average ω ratios within the Bangiales show that from the 24 *rpl* genes, 8 genes are under positive, 8 genes are under neutral and 8 genes are under purifying selection, therefore not being able to have a conclusion about the overall *rpl* substitution rates of these species.

3.3.7.3 RPL genes – Discussion

The analyses of the substitution rates of the *rpl* genes of 30 Rhodophyta species show that 8 RPL genes have statistical differences when the comparison takes into consideration the habitat where the species are usually found. For example, the species observed in hot springs have 6 *rpl* genes with distinct high substitution rate patterns (*rpl*1, *rpl*2, rpl5, rpl13, *rpl*19 and *rpl*27) when compared to the species observed in freshwater, salt marshes and marine environment. For the *rpl*19, the hot spring species is also the highest ω ratio observed in all of the 24 *rpl* genes analyzed. This pattern is a strong indication that species in hot springs are under

positive selection for their *rpl* genes, while species in the other habitats have neutral ω ratios and are possibly under neutral selection.

Our study used the species *Galdieria sulphuraria* (Cyanidiales) as representative of the hot springs' habitat. Species of the genus *Galdieria* also occupies endolithic and interlithic habitats, being more exposed to microenvironmental fluctuations such as desiccation, a broad range of acidity, temperature, and salinity (Cho et al., 2020). Although *G. sulphuraria* is considered a model organism for studies with Rhodophyta species, there is no literature about chloroplast genes evolution rates for the genus. The chloroplast genomic structure and content of *G. sulphuraria* is considered typical when compared to other red algae (Jain et al., 2015), however, studies regarding the genome evolution of this species mitochondrial data, shows the highest substitution rates among all red algae, likely due to their polyextremophilic lifestyle (Jain et al., 2015; Cho et al., 2020). Because there are no data about *G. sulphuraria* chloroplast substitution rates, an in-depth study and future comparisons are necessary to better understand the complexity of the substitution rates and adaptive evolution of the Cyanidiales species, especially *G. sulphuraria*.

The species observed in freshwater have a considerable difference in substitution rates for the *rpl*20 gene, being the habitat with the highest ω ratios, even though all rates support purifying selection. Another singularity of the *rpl*20 gene is that the ω ratio of the freshwater species *B. atropurpurea* is very distinct, indicating positive selection, while all other species, independently of the habitat are under purifying selection. There are only a few studies with Viridiplantae species that focus on the *rpl*20 gene, one being with peppers, *Capsicum annuum*, which revealed large indels insertions on the gene (Jo et al., 2011); and another study with an endemic Chinese plant, *Urophysa*, which indicates that the *rpl*20 gene is under strong positive selection (Xie et al., 2018), and confirm the use of the gene for phylogenetic purposes when studying the genus *Urophysa* (Muto & Ushida, 2002, 1995; Xie et al., 2018). These studies contradict the results from our analyses, but the different substitution ratios pattern can be justified by the large evolutionary distance among Rhodophyta and Viridiplante chloroplast genomes. The *rpl*20 gene is involved in translation, which is an important part of protein synthesis (Xie et al.,2018), however, more functional studies about the *rpl*20 gene are needed to better understand its role (Jo et al., 2011), particularly in Rhodophyta. Moreover, a detailed

analysis of the gene codon bias and identification of mutations in the 1st, 2nd and 3rd positions can help to explain why the *rpl*20 gene in *B. atropurpurea* is under purifying selection.

The substitution rates of the *rpl*24 gene of species observed in freshwater support purifying selection, while other habitats' species are under neutral selection. Studies about the *rpl*24 gene claim that its protein, L24, is responsible for the stabilization of the peptidyl transferase activity (Reyes et al., 1977; Balogu et al., 2015). Moreover, Balogu et al. (2015) suggest that *rpl*24 gene has an overexpression when the plants are under drought stress. There are no studies that explain details about the *rpl*24 gene in Rhodophyta species, independently of its habitat. Therefore, more investigation is needed to learn how and why red algae freshwater species have the *rpl*24 gene under purifying selection.

The comparison of the ω ratios among species and their phylogenetic position show that the Bangiales is the most distinct lineage, having 6 *rpl* genes with statistical differences when compared to the other lineages: *rpl*4, *rpl*11, *rpl*13, *rpl*21, *rpl*24 and *rpl*28. The *rpl*4 gene protein is crucial for the maintenance of ribosomal translational efficiency and fidelity; moreover, the proteins are involved in the delivery of vacuolar targeted proteins (O'Connor et al., 2004; Rosado et al., 2010). According to Wu et al. (2014), who used peanut species (*Arachis hypogaea)* and several other Viridiplantae species for comparison, the presence of the *rpl*4 gene in all species and the alignment of the gene sequences indicated a highly conserved gene within the plants. In addition, the substitution rates analysis identified the *rpl*4 gene to be under purifying selection among the compared species. Our study agrees that the *rpl*4 gene is conserved since it is present in all Rhodophyta species, and the alignment shows highly conserved blocks. However, our substitution rate analysis of the gene shows that Bangiales and Porphyridiales' *rpl*4 gene is under positive selection. The other species have ratios that categorize them as under neutral selection, but no purifying selection was identified. Future investigations about the *rpl*4 gene in Rhodophyta can help to clarify why so many Bangiales species are under positive selection, and why the first clade of the order (*B. atropurpurea, P. umbilicalis* and *P. purpurea*) have different substitution rate patterns to the other Bangiales species.

For the *rpl*11 gene, the Bangiales substitution rates are high, indicating positive selection, while Porphyridiales is under purifying selection. All the other lineages are under neutral selection for the gene. Contrary to the *rpl*4 gene, there are no studies regarding the *rpl*11 gene

substitution rates and no reference about the role of the *rpl*11 gene in Rhodophyta or any other algal group. Curiously, the Bangiales species have the exact substitution rate pattern observed for the *rpl*4 gene, where the species *B. atropurpurea, P. umbilicalis* and *P. purpurea* (Bangiales top clade) are under neutral selection, while the other species are under positive selection. An in-depth study of Bangiales species' *rpl*4 and *rpl*11 genes can elucidate why most species are exhibiting higher substitution rates than these three species.

The *rpl*13 gene of Bangiales species has the highest ω ratios of the gene analysis, therefore, under positive selection, while the other lineages are neutral. Once again, there are not much information about the gene and even less about the substitution rate pattern in Rhodophyta species. Song et al. (2014) reported that the ribosome protein L13, a product of the *rpl*13 gene, is essential for the normal development of rice, and is very similar to the L13 protein in *E. coli*. Studies with *E. coli* suggest the *rpl*13 gene to play an important role in chloroplast development, especially under low-temperature conditions (Chandra & Liljas, 2000; Sharma et al., 2007). Taking into consideration the role of the *rpl*13 gene under different temperature stresses, the high rates might be reflecting the importance of this gene in Rhodophyta species. In addition, the *rpl*13 gene also had statistical support in our analysis of compared habitats, where the hot springs and marine environment species exhibited high substitution rates. Thus, further investigations about the role of the *rpl*13 gene and the importance of the L13 protein in Bangiales species can reveal an important part of the species response mechanism to temperature stress.

The *rpl*21 gene of Bangiales species differs from the other lineages for exhibiting the lowest average substitution ratio, indicating purifying selection, while the other species are under neutral selection. For the *rpl*24, the Bangiales substitution ratio is similar to Porphyridiales, both are under neutral selection - the similarity between these lineages was also noted for the *rpl*4 gene - while the other lineages are under purifying selection. Lastly, for the *rpl*28 gene, all lineages show support for positive selection, but the Bangiales show statistical support and have the most particular substitution rates of all lineages. Unfortunately, there is no available literature about these 3 genes regarding their role within the cell and possible substitution rates, not only for Rhodophyta species but for Viridiplantae species as well. It is worth noting that the green algae *B. plumosa* does not have a copy of the *rpl*21 and *rpl*28 genes, which are also absent from other Viridiplantae species and were possibly lost throughout evolution. Both of these genes, however, are conserved in red algae species, are considered a crucial part of the photosynthetic

apparatus (Green, 2011). Although some *rpl* genes information is available regarding its protein product and possible roles within the cell, studies including the substitution rates of individual *rpl* genes are extremely limited. The data are usually from Viridiplantae species, which are helpful for a broad understanding, but due to the large distance between Viridiplantae and Rhodophyta, it is difficult to have a conclusion about the evolution of the *rpl* genes in red algae.

.

3.4 Conclusion

The present study used a non-conventional methodology to investigate how the substitution rates of chloroplast genes vary among Rhodophyta species and is the most detailed study examining rates of change of individual chloroplast genes in this group. The methodology used in this study demonstrates that the traditional use of concatenated data and conserved gene blocks for substitution rates analyses, dismiss the particularities of each gene and do not reflect the actual substitution rates that occur within a single gene. Because of the nature of the substitution rate analysis, I recommend that future studies make use of entire gene sequences instead of conserved gene blocks, generating results that better reflect real values. The analyses conducted in this chapter analyzed the substitution rates of 71 vital photosynthetic housekeeping genes for photosynthesis of 30 red algal species and compared the results by phylogenetic position and different habitats in which the species evolved. This is a unique way to examine potential selection pressures that are upon species, and how they may be reflected in the DNA sequence.

As expected, the chloroplast genes of a taxon will exhibit trends in substitution rates (and differences among each lineage) according to its phylogenetic position. My analyses demonstrated that the order Bangiales is the most distinct lineage among the red algae and exhibits patterns of molecular adaptations based on the species phylogenetic position. For instance, this order shows that taxa have 31% of the genes under positive selection and 43% under purifying selection. The other Rhodophyta lineages have around 28% of the genes under positive, and around 35% under purifying selection. Hence, selection forces are strongly acting in the chloroplast genomes of species within the Bangiales, maintaining the DNA sequence and protein products conserved. The Bangiales substitution rates are significantly different from the other red algal orders for the *atp*, *psa*, *rps* and *rpl* gene families. Hence, these differences may affect essential plastid function and should be investigated further to verify if the chloroplast machinery and functions in the Bangiales are also affected. In addition, future taxonomic studies that are looking to resolve phylogenetic relationships among Bangiales species at a familial, ordinal and genus level might benefit from the use of genes part of the *atp*, *psa*, *rps* and *rpl* families.

In addition, I noted that species occupying freshwater have contrasting rates, having the highest or lowest rates of substitution when compared to other habitats. These results clearly illustrate that there are different selection pressures on red algae chloroplast genes that appear to be due to the habitat that species occupy. This is the first study to examine this in the red algae. Among Rhodophyta chloroplast genomes examined, the rates of change in the genes for species observed in freshwater are significantly different from the species occupying other habitats (e.g. marine, hot springs) for both the *atp* and *rps* gene families. The rates of 3 *atp* genes are under neutral selection (*atp*B, *atp*F and *atp*G) and 1 gene is under positive selection (*atp*H). The species in other habitats differ in the number of *atp* genes under neutral (around 2) and positive (around 3) selection. The rates of 4 *rps* genes are under neutral selection (*rps*2, *rps*6, *rps*14 and *rps*17), while species in other habitats have around 3 genes. The rates of substitution in species that occupy the marine environment are significantly different for the PSA gene family, with 2 genes under purifying forces (*psa*K and *psa*L), 5 genes under neutral (*psa*B, *psa*D, *psa*E, *psa*F and *psa*J) and 4 genes under positive selection (*psa*A, *psa*C, *psa*I and *psa*M). The species in other habitats differ in the number of *psa* genes under neutral (around 3) and positive (around 6) selection. The rates of species observed in hot springs showed differences for the *rpl* gene family, where 5 genes are under purifying (*rpl*12, *rpl*14, *rpl*20, *rpl*33 and rpl35), 9 genes are under neutral (*rpl*2-6, *rpl*11, *rpl*18, *rpl*21 and *rpl*32), and 8 genes are under positive selection (*rpl*1, *rpl*13, *rpl*16, *rpl*19, *rpl*23, *rpl*27-28, *rpl*36). The species in other habitats differ in the number of *rpl* genes under purifying (6), neutral (12) and positive (5) selection. By investigating further details of the organism's molecular evolution due to habitat constraints, researchers will be able to understand more about the adaptation of organisms living in almost identical environmental selection pressure, and finally, observe the effects of habitat shifts reflected in the DNA sequence.

Within the order Bangiales, it appears that the individual chloroplast genes in the invasive red alga *B. atropurpurea* (Bangiales) are the most distinct rates among the Rhodophyta examined. The species has several genes under purifying and positive selection pressures that contradict the ones observed in other red algae, even within Bangiales species (*atp*H, *atp*I, *psa*C, *psb*E, *psb*F, *psb*L, *psb*N, *rps*10, *rps*11, *rps*14, *rps*17, *rps*20 and *rpl*20), confirming how particular this organism is despite sharing a more recent common ancestor with various *Porphyra* species that are only known to be marine. This shift in habitat, from marine to freshwater, required

adaptations which may be reflected in the substitution rates seen in the present study. This also raises the possibility of further invasions into other freshwater habitats by *B. atropurpurea*, including further lakes and rivers in North America. More recently (Müller unpublished), this species has for the first time been reported in the finger lakes in New York state, outside of its known distribution in the Laurentian Great Lakes. Another important point to be considered given the unique rates of substitution, is the possibility of *B. atropupurea* chloroplast genes with distinct substitution rates to be indicating sequence saturation. The sequences and results obtained in this chapter can be used for future studies on specific sequence saturation tests to confirm the saturation and understand more about the evolution of the chloroplast of *B. atropurpurea*.

As more data on relative rates of substitution become available, our understanding of the mutational patterns among chloroplast genomes across major groups will improve, which will allow us to test how much variation occurs in eukaryotic genome evolution and the impact of habitat on these organisms. This study not only increased the chloroplast genomic information of *B. atropurpurea* but also provided useful information to understand the evolutionary relationship among Rhodophyta species. The findings showed here via molecular analyses have important implications that lay a vigorous foundation for further studies in the areas of genetics, evolution, conservation, and biotechnology of red algae.

Chapter 4

The evolution of meiotic toolkit genes in *Bangia atropurpurea* (Bangiales) and the Rhodophyta

4.1 Introduction

Reproduction in Bangiales

The Rhodophyta comprises a monophyletic lineage of approximately 7,200 species, which belong to the Archaeplastida (Plantae) that contain plastids derived from primary endosymbiosis (Yoon et al., 2006). They are an ancient eukaryotic group with representation in the fossil record dating 1.6–1.0 billion years ago (Butterfield, 1990; Butterfield, 2000; Bengtson et al., 2017). Red algae species, particularly within the Florideophyceae, have sexual reproductive structures known to be more elaborate than any other group of algae (Shim et al., 2021), with unique life histories that include a diploid phase (carposporophyte) that is contained on the gametophyte (Kamiya & Kawai, 2002). In most red algae, the male and female sexual structures are observed on separate male and female gametophytes (dioecious), but many species are also monoecious (Nam-Gil, 1999; Shim et al., 2021). In addition, sexual reproduction in the red algae has been linked to external factors such as light, temperature, and nutrients, which regulate their sexual reproduction (Liu et al. 2017). However, the molecular mechanisms underlying the control of their sexual reproduction are still poorly understood. The mechanisms regulating multiple patterns of gamete formation appear to represent ancestral features of eukaryotic sexual reproduction because the fossil records (1200 Mya) of *Bangiomorpha*, which closely resemble the extant filamentous Bangiales, provide the oldest evidence of sexually reproducing eukaryotes (Butterfield, 2000). Despite the evolutionary and ecological importance of the group, little is known about the genetic pathways that lead to sex determination and the process of meiosis in red algae species (Shim et al., 2021).

Members of the red algal order Bangiales (Bangiophyceae) have a heteromorphic life cycle, that alternates between a haploid gametophytic thallus (foliose in *Pyropia*, *Porphyra* or filamentous in *Bangia*) phase and a diploid sporophytic branched filamentous phase (conchocelis) (Mikami, 2019; Song et al., 2020; Li et al., 2021; Uji & Mizuta, 2021) (Fig. 3 of Chapter 1 of this thesis; Appendix K). The conchocelis stage occupies an unusual ecological niche, such as boring into shells and calcareous rocks, which require specialized proteins and an exclusive cell wall structure that is different from the gametophyte foliose thallus (Mukai et al., 1981; Brawley et al., 2017). In addition to the sexual life cycle, many species within Bangiales undergo an asexual life cycle by producing asexual spores, archaeospores, which can develop

into the gametophytic thallus (Drew, 1956; Kornmann, 1994; Tanaka, 1952; Uji & Mizuta, 2021; Shimizu et al. 2007; Mikami et al., 2019; Chapter 1 of this thesis).

The first description of the *Porphyra* life cycle (Drew, 1949) revolutionized the farming of this seaweed in mass cultivation and the development of the aquaculture industry of *Porphyra*. Nowadays, most of the seaweed cultivation industries take advantage of the asexual reproduction of Bangiales species, since this type of reproduction results in fast yield growth, being much faster than by sexual reproduction (Li & Wang 1984; Yan et al. 2004; Song et al., 2020; Uji & Mizuta, 2021). The transitions from gametophyte to sporophyte and from sporophyte to gametophyte are triggered by fertilization of male and female gametes and meiosis, respectively (Coelho et al., 2007; Horst & Reski, 2016; Liu et al., 2017; Mikami, 2019). It is known that in the Bangiales, meiotic cell division occurs early during the development of gametophytes (Burzycki & Waaland, 1987; Ohme & Miura, 1988; Wang et al., 2006; Shimizu et al., 2007; Mikami et al., 2012; Takahashi & Mikami, 2017), but currently, there has been little research on the molecular mechanisms by which red algae switch from sexual to asexual reproductive stages (Endo et al., 2021; Uji et al., 2016, 2021; Yanagisawa et al., 2019). Moreover, much of the genetic mechanisms involved in archaeospore formation and asexual reproduction in Bangiales is still unknown (Xu et al., 2003; Kitade et al., 2008; He et al., 2013; Song et al., 2020; Li et al., 2021). The differentiation of males and females in *Porphyra/Pyropia/Bangia* is hypothesized to be controlled by a pair of Mendelian alleles designated as *mtm* and *mtf (*van der Meer and Todd 1977; Mitman and van der Meer, 1994; Zhang et al. 2013). Unfortunately, studies using current molecular tools have not examined the underlying genetics of the genes involved in these processes and there are still many questions regarding the molecular details of sexual and asexual reproduction in this order.

Bangia species are known for being asexual and dioecious, but their reproduction is not well studied as other Bangiales species. (Notoya & Iijima, 2003; Hwang et al., 2008; Sutherland et al., 2011; Khoa et al., 2021). The *Porphyra* and *Pyropia* species, are the most studied species within the Bangiales, and are known for reproducing sexually or asexually, being monoecious or dioecious (Brodie et al., 1998; Yanagisawa et al., 2019). These species are closely related to *Bangia* and knowing their reproduction system can help understand *Bangia* reproduction and point out stages that might be very similar, or very different, among them. *Porphyra* and *Pyropia* species have similar morphology but a wide variety of life-history strategies, and, even though

the gametophyte generation is the dominant component and most studied part of their life cycle, each species has a different mating system (Sutherland et al., 2011; Varela-Álvarez et al., 2018). *Porphyra dioica*, for example, has mostly individuals with separate sexes (dioecious) (Pereira et al., 2004); *P. umbilicalis* can be hermaphroditic with both sexes in the same individual and gametangia separated in different sections of the thallus (monoecious) or also dioecious (Brodie & Irvine 2003), however, some populations also reproduce only asexually (Bird & McLachlan 1992; Blouin et al., 2007; Gantt et al., 2010); *Porphyra linearis* also has individuals with both systems, dioecious and hermaphroditic (Bird, 1973; Varela-Alvarez et al., 2004). For *Pyropia* species, it is known that *P. yezoensis* and *P. tenera*, produce male and female gametes in mixed on the same gametophyte (Kim, 2011), while other species are monoecious, such as *P. katadae* (Neefus et al., 2008). There are also dioecious species, such as *P. pseudolinearis* and *P. dentata* (Kim, 2011). Due to the various types of reproduction observed in *Porphyra* and *Pyropia* species, the taxonomic classification and identification of the species before molecular tests were misleading (Sutherland et al., 2011).

To solve the misleading classification of the Bangiales species a taxonomy revision based on phylogenetic analyses of *rbc*L and SSU rRNA genes was conducted by Sutherland et al. (2011). The results of this important study were later supported by other studies that agree with the revised taxonomic position of *Bangia* species (Sánchez et al., 2014; Li et al., 2019). To date, the taxonomy of *Bangia* comprises 3 filamentous clades ('*Bangia*' 1, '*Bangia*' 2, and '*Bangia*' 3), which include *B. atropurpurea*, *B. fuscopurpurea*, and *B. gloiopeltidicola* respectively (Sutherland et al., 2011). *Porphyra* and *Pyropia* species are phylogenetically close related to *Bangia* species (Chapters 2 and 3 of this thesis), but the number of studies about *Bangia* reproduction and life cycles is considerably less available. Two studies focused on *Bangia* species reproduction, one being with a *Bangia* species that is part of the '*Bangia*' 2 clade (marine) (Khoa et al., 2021). In this research, the authors state that *B. fuscopurpurea* can be induced to reproduce asexually when under heat stress. The study also mentions that it was difficult to induce male and female gamete development under laboratory culture conditions, and only samples that were already producing conchocelis in the field continued through sexual reproduction. The second study with a focus on *Bangia* reproduction is with *B. atropurpurea* ('*Bangia*' 1 clade) (Gargiulo et al., 2001). Freshwater populations of *Bangia* have been reported to reproduce only asexually and no sexual reproduction was ever detected during four years of

observations of natural populations (Lin & Blum 1977; Sheath & Cole 1980; Sheath, 1987; Müller et al., 1998, 2003). Gargiulo et al., (2001) however, claimed that *B. atropurpurea* from a region in Italy was reproducing sexually in laboratory with the use of enriched freshwater cultures. This was the first and only time that *B. atropurpurea* was observed to produce gametangia, and no subsequent studies have reproduced this type of reproduction under laboratory conditions.

The meiotic toolkit

As mentioned in Chapter 1 of this thesis, meiosis is a key feature of sexual reproduction. During meiosis, homologous chromosomes replicate, recombine, and randomly segregate, followed by the segregation of sister chromatids to produce haploid cells (Wang et al., 2021) (Fig. 4.1). The use of a eukaryotic meiotic toolbox that includes the core genes related to meiosis is used for comparison studies, and for Rhodophyta species, the data for the toolbox is based on model organisms such as *Arabidopsis thaliana*, and *Mus musculus* (Schurko & Logsdon, 2008; Malik et al., 2008) (Table 4.1). Although these are very different organisms, sex inheritance, sex chromosomes and meiotic genes in land plants are very similar to those in animals and the same toolbox can be used as reference among eukaryotic organisms (Critchlow et al., 2004; Simanovsky & Bogdanov, 2018). It is important to note that although the meiosis stages are very well established (Fig. 4.1), there is still a large gap in knowledge regarding the red algae genes and proteins that orchestrate the meiosis process. All the genes listed in the eukaryotic meiotic toolbox (Table 4.1) are known for their role during prophase I, and only a couple of genes are known to date to play a role in other stages of meiosis (Hamant et al., 2006; Wang et al., 2021).

Due to the possible changes in genome throughout evolution, there is a debate whether the meiotic toolbox based on Viridiplantae species genes can be used for algae as well, especially for the ancient lineage of the Bangiales species (Schurko & Logsdon, 2008; Butterfield, 2000; Bengtson et al., 2017). However, Brawley et al. (2017), when describing the complete genome of *Porphyra umbilicalis* (Bangiophyceae) also noted the presence of meiotic-related genes using the meiotic eukaryotic toolbox as a reference, and to date this is the most complete dataset for Rhodophyta species. More recently, a transcriptome study with the red algae *Bostrychia moritziana* (Florideophyceae) was able to determine sex-specific genes that could be involved in sex determination and sexual differentiation (Shim et al., 2021). Another transcriptome study of

the red alga *Thorea hispida* (Florideophyceae), identified meiosis-specific genes, including *spo*11, *mnd*1, *rad*51, *msh*4, *msh*5, *rec*8, and *mer*3 (Nan et al., 2020). It is intriguing how even red algae species with high economic value have very little information available regarding their reproduction. Further molecular, physiological, and phenotypic studies are required to better understand the asexual reproduction in Rhodophyta, since this type of reproduction can be very useful for genetic studies and applications, such as mapping populations (Chaotian et al., 2010), and quantitative trait loci analysis (Xu et al. 2012), which are prerequisites for breeding new cultivars (Yan et al. 2010; Jiang et al. 2010). Regarding sexual reproduction, there are also many questions to be answered, and even though the cell cycle in plants is relatively well characterized, the genetic control of meiosis progression is still largely unknown and will need more research efforts in the coming years.

The goal of this thesis chapter is to verify the presence of meiotic genes from metagenomic sequencing in *B. atropurpurea*, the only freshwater taxon within the Bangiales. As noted in Chapter 1, this species has only been observed to be asexual in nature and reported once to be induced into sexual reproduction in the laboratory (Gargiulo et al., 2001). Furthermore, global populations sequenced by Müller et al. (1998, 2001, 2003) were observed to be identical in 18S rRNA and nearly identical in chloroplast genes; indicating little divergence despite broad geographic distribution and perhaps a lack of sexual reproduction in nature. The genes obtained from metagenomic sequencing were then compared to a total of 29 meiotic-related genes from available Rhodophyta species. All these genes are part of the meiotic Prophase I, more specifically, part of stages 1, 3 and 4 (Fig. 4.1), and a description of these genes is presented below:

Table 4. 1 List of meiosis-related genes (eukaryotic meiotic toolkit) according to Malik et al., 2008.

Gene	Function	Present in *Porphyra umbilicalis*	Present in *Bangia atropurpurea* *
dmc1	Meiosis-specific recombinase	No[1]	No
hop1	Binds double stranded breaksand oligomerizes during meiotic prophase	Yes[1]	No
hop2	Works with Mnd1, homology search downstream Rad51 during pachytene	Yes[1]	Yes
mer3	Promotes Holliday junction with ZMM proteins, Msh4, and Msh5	Yes[1]	yes
mlh1	Di-and trinucleotide sequence mismatch repair; forms a heterodimer with Mlh2, Mlh3, & Pms1; cooperates with Msh2	Yes	Yes[3]
mlh2	Heterodimer formation with Mlh1, removes cisplatin adducts with Msh 2/3 or 2/6	No	No
mlh3	Heterodimer formation with Mlh1, conducts frameshift repairs with Msh2/3 or Msh2/6, promotes meiotic crossover with Msh4/5	No	Yes[3]
mnd1	Works with Hop2, required for stable heteroduplex DNA formation	Yes[1]	Yes
mre11	dsDNA exonuclease, ssDNA endonuclease; trims hairpins and DNA ends from 3'-5'	Yes	No
msh2	Heterodimer formation with Msh3 or Msh6	Yes	No
msh4	Heterodimer formation with Msh5, directs the Holliday junction towards crossover with interference	Yes[1]	No
msh5	Heterodimer formation with Msh4, directs Holliday junction towards crossover with interference	Yes[1]	No
msh6	Heterodimer formation with Msh2, binds to mismatches	No	Yes
pds5	Maintains interactions between sister chromatids in late prophase	No	No
pms1	Heterodimer formation with Mlh1, mismatch repair with Msh2/3 or Msh2/6	Yes	Yes[3]
rad1	DNA Exonuclease, meiotic checkpoint control	Yes	No
rad21	Holds Smc1/3 by binding terminal domains to hold sister chromatids during mitosis and meiosis	No	No
rad50	Double-strand break repair protein; Complexes with Mre11 and Nbs1	Yes	No
rad51	Facilitates intrahomologouss recombination and cause formation of helices on ss-and dsDNA	Yes[2]	Yes
rad52	Iniciate doublestrend repaie by homologous recombination	No	No
rec8	Paralogue of Rad21; only expressed during meiosis	No[1]	No
scc3	Repairs double stranded breaks; interacts with Smc1/3 and Rec8/Rad21 for maintaining cohesin ring	No	Yes
smc1	Heterodimer formation with Smc3, cohesion ring for sister chromatids	Yes[2]	Yes
smc2	Heterodimer formation with Smc4, condensing ring for chromosome assembly and segregation	Yes[2]	No
smc3	Heterodimer formation with Smc1, cohesion ring for sister chromatids	Yes[2]	No
smc4	Heterodimer formation with Smc2, condensing ring for chromosome assembly and segregation	Yes[2]	Yes
smc5	Heterodimer formation with Smc6, checkpoint and DNA repair	Yes[2]	Yes
smc6	Heterodimer formation with Smc5; after replication, binds to ssDNA	Yes[2]	No
spo11	Causes the double-strand breaks for homologous chromosomes	Yes[2]	Yes

Genes in **bold**: considered core miosis genes; 1: As described by Brawley et al., 2017; 2: Genes found by BLAST search; 3: Multiple contigs hits for the same gene; *Described by this study.

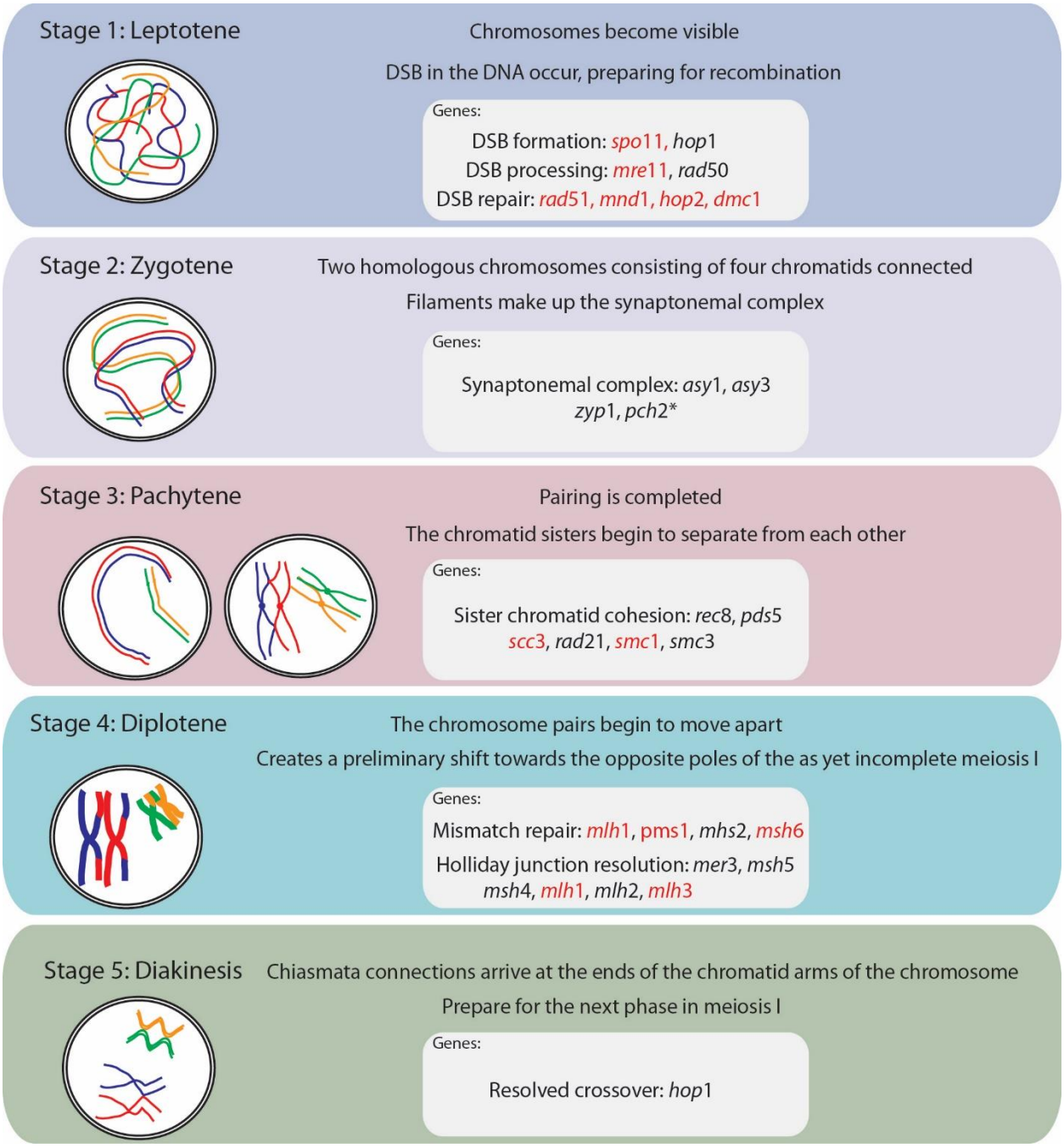

Figure 4. 1. Stages of Meiosis Prophase I, chromosome pairing, condensation and crossing over. Main events that occur in every stage are inside the coloured box, and the genes identified for each event are inside the grey boxes. The genes in red are the ones observed in *Bangia atropurpurea*. DSB: Double-Strand Breaks. *: Only identified in rice and maize to date. Figure based on the reviews of Malik et al. (2008) and Wang et al. (2021).

Stage 1. Leptotene genes

spo11 gene:

The Initiator Of Meiotic Double Stranded Breaks (*spo*11) gene was one of the first meiotic recombination genes to be identified, with records from 1969 (Esposito & Esposito) and 1985 (Klapholz, et al. 1985), using *Saccharomyces* and other yeast species as study models. This gene is known for being essential to meiosis since its protein product initiates the meiotic recombination by forming double-strand breaks (DSBs) and for chromosome pairing synapsis (Keeney et al., 1997; Romanienko & Camerini-Otero, 2000; Benyahya et al., 2020). Through the years, extensive research has been done with meiotic genes of many model organisms (Keeney et al., 1997; Metzler-Guillemain & de Massy, 2000; Romanienko & Camerini-Otero, 2000), and because the gene is observed from yeast to humans and plants, it is considered a widely conserved gene (Keeney et al., 1997; Hartung & Puchta, 2000).

While the *spo*11 gene product in the initiation of meiotic recombination appears to be highly conserved among most eukaryotes, some organisms differ in the molecular requirements for synapsis (Hartung & Puchta, 2000; Walker & Hawley 2000). Cervantes et al., (2000) demonstrated that the *spo*11 protein in *Saccharomyces cerevisiae* and *S. pombe* does not participate in the synapsis, and the species have a distinct system that includes other genes to bypass the requirement for the *spo*11 protein (Cervantes et al., 2000). Surprisingly, a study with *Arabidopsis thaliana* shows that the plant has two *spo*11 homologues genes (*spo*11-2 and *spo*11-3) which are closely related to each other than to other eukaryotic homologues (Hartung & Puchta, 2000). This is a strong indication that both genes arose through a recent duplication event during plant evolution. Hartung & Puchta (2000) also argue that in plants, the *spo*11-like proteins possibly have other biological functions than originally described for yeast and animals. The studies that followed identified yet another *spo*11-like protein (*spo*11-3), but only the *spo*11-1 and *spo*11-2 are involved in meiosis (Grelon et al., 2001; Hartung & Puchta 2001; Hartung et al., 2002; Sugimoto-Shirasu et al., 2002; Yin et al. 2002; Stacey et al., 2006; Hartung et al. 2007; Sprink & Hartung, 2014; Benyahya et al., 2020). Although there are several studies about the *spo*11 gene and its protein roles in different plant model organisms, information about how the gene is involved in the initiation of meiotic recombination in red algae is still non-existent.

rad 51 and dmc1 genes

During the first meiotic division, homologous chromosomes must separate to opposite poles of the cell to ensure the right complement in the progeny (Petukhova et al., 2005). The homologous recombination provides a mechanism for a genome-wide homology search and the formation of homologous joint molecules (JM) (Petukhova et al., 2005; Cloud et al., 2012). The Radiation Sensitive 51 (*rad*51) and the Disrupted Meiotic cDNA 1 (*dmc*1) genes play a crucial role in these processes (Shinohara and Shinohara, 2004; Petukhova et al., 2005; Cloud et al., 2012). Specifically, these genes play a role in the removal of single-stranded DNA (ssDNA) binding proteins from ssDNA and facilitate the invasion of the 3′-extended strand into the duplex of the homolog (Cloud et al., 2012; Kurzbauer et al., 2012, Shinohara and Shinohara, 2004; Takemoto et al., 2020). To understand if both genes are equally responsible for JM formation, studies with *Arabidopsis thaliana* and other flowering plants, have suppressed the genes separately (Cloud et al., 2012; Da Ines et al., 2013). As result, it was noted that the *rad*51 gene product is fully dispensable for meiotic recombination, being classified as a multifunctional protein that catalyzes recombination directly in mitosis, and a meiosis accessory factor (Cloud et al., 2012; Da Ines et al., 2013). On the other hand, the suppression of *dmc*1 caused a profound recombination defect (Cloud et al., 2012), indicating that the gene works independently from rad51 during meiosis, and classifying the gene as an integral part of the regular meiotic toolbox (Kurzbauer et al., 2012). Despite the increasing number of genome datasets available for Rhodophyta species, the number of studies that explain details about the working mechanisms and connections among meiotic genes in these organisms is nonexistent.

mnd1 and hop2 genes

The *mnd*1 (Meiotic Nuclear Division Protein 1) and the *hop*2 (Homologous-pairing Protein 2) genes are responsible for translating proteins that are critical for homologous recombination (Tsubouchi & Roeder, 2002; Chi et al, 2007; Pezza et al., 2007, Uanschou et al., 2013; Farahani-Tafreshi et al., 2021). The protein products from the *mnd*1 and *hop*2 genes form a heterodimer that physically interacts with *rad*51 and *dmc*1 proteins, stimulating their activity and forming the first recombination during meiosis (Chen et al. 2004; Petukhova et al. 2005; Pezza et al. 2006; Chi et al., 2007; Zhao et al., 2013; Farahani-Tafreshi et al., 2021). The

homologous recombination has the function to repair the DNA double-strand breaks (DSBs), ensuring the proper segregation of homologous chromosomes at the first division of meiosis (Kleckner, 1996; Roeder, 1997, Uanschou et al., 2013). More recently, it was established that *hop*2 also plays a role in promoting homologous chromosome synapsis and preventing nonhomologous chromosome exchanges (Farahani-Tafreshi et al., 2021). The *mnd*1/*hop*2 heterodimer is recognized as the main force shaping genetic variation among genomes (Pezza et al., 2007).

Several studies used model organisms to investigate if the *mnd*1 and *hop*2 genes are essential for meiotic events. The results of the studies are very similar, where there are severe defects in homologous synapsis and during the early-stage recombination when the genes are disrupted, resulting in sterility (Petukhova et al., 2005; Pezza et al., 2007). Curiously, without the *mnd*1 and *hop*2 proteins, the recombination proceeds normally until the point when *rad*51 and *dmc*1 proteins are loaded on the ends of meiotic DSBs, without further progression (Zierhut et al., 2004; Enomoto et al., 2004; Chen et al., 2004; Uanschou et al., 2013). Studies with yeast species (Leu et al. 1998; Gerton & DeRisi, 2002; Tsubouchi & Roeder 2002; Zierhut et al. 2004; Henry et al. 2006), *Arabidopsis thaliana* (Domenichini et al., 2006; Kerzendorfer et al., 2006; Panoli et al., 2006), and mouse (Petukhova et al., 2003) agree that the synchronous work between *mnd*1/*hop*2 and *rad*51/*dmc*1 is likely to be crucial *in vivo* (Pezza et al., 2007). Unfortunately, there are no studies that explain the functionality of the *mnd*1 and *hop*2 protein heterodimer, and if the genes are essential for meiosis in red algae species.

Stage 3. Pachytene

SMC gene family

The Structural Maintenance of Chromosomes (SMC) genes and proteins are highly conserved from bacteria to humans and play fundamental roles in many aspects of chromosome organization and dynamics, such as the chromosome architecture via sister chromatid cohesion and chromosome condensation, in addition to their role in transcription, and DNA repair and recombination (Revenkova et al., 2001; Cobbe & Heck, 2004; Losada & Hirano, 2005; Schubert, 2009). The discovery of the cohesion, condensation, DNA recombination and repair, and gene-

dosage compensation functions of the SMC proteins provided clues on how the DNA is folded, organized, and segregated in the cell (Jessberger, 2002; Losada & Hirano, 2005). In eukaryotes, there are six members of the SMC gene family, forming heterodimers: the *smc*1-*smc*3, *smc*2-*smc*4 and *smc*5-*smc*6 (Revenkova et al., 2001; Jessberger, 2002; Bolaños-Villegas, 2021) (Fig. 4.2). Based on phylogenetic analysis, Cobbe & Heck (2004) proposed that each of the six *smc* genes originated through a series of ancient gene duplication events, with the condensins evolving more rapidly than the cohensins. The study also shows that *smc*1 and *smc*4 are evolutionarily related to one another, as are *smc*2 and *smc*3. The *smc*5 and *smc*6 genes are part of a more ancestral family (Cobbe & Heck, 2000; Jessberger et al., 1998; Jones & Sgouros, 2001; Jessberger, 2002).

Within several studies with animal models, it was established that the *smc*1-*smc*3 proteins, along with other polypeptides, act as the core of the cohesin complexes that mediate sister chromatid cohesion (Molnar et al., 1995; Guacci et al., 1997; Cobbe & Heck; 2000; Nasmyth et al., 2000; Bolaños-Villegas, 2021). The *smc*2-*smc*4 protein heterodimer regulates the proper clustering of rDNA arrays during interphase (Bolaños-Villegas, 2021), involving the multiprotein complex condensing which is critical for chromosome assembly and segregation (Haering & Nasmyth 2003; Losada & Hirano, 2005; Hirano, 2005). In addition, the *smc*2-*smc*4 is considered key for chromosome condensation in *S. cerevisiae* (Strunnikov et al., 1995), *S. pombe* (Sutani & Yanagida, 1997), *Xenopus leavis* (Hirano et al., 1997) and humans (Schmiesing et al., 2000; Revenkova et al., 2001). Lastly, the *smc*5-*smc*6 protein heterodimer is required for homologous DNA recombination during the S-phase and proper meiotic synapsis (Bolaños-Villegas, 2021), and act as a DNA repair and checkpoint response (Lehmann, 2005). Although being conserved among species, some variations do exist. For example, there are two *smc*4 proteins in the namtode *Caenorhabditis elegans* (Hagstrom et al., 2002), and a meiosis-specific SMC isoform from higher eukaryotes has been described which is highly homologous to *smc*1, and hence was named *smc*1β (Revenkova et al., 2001).

The available literature about the SMC genes, proteins and functions in plants is not as well documented as in animal models. Studies with *Arabidopsis thaliana* and maize are the ones published to date, and state that SMC proteins form complexes that mediate dynamic chromosome architecture in meristematic and differentiated tissues (Schubert, 2009). Yet, it remains uncertain if plant SMC participates in regulating gene expression and development as

described for yeast, *Caenorhabditis* (nematode), *Drosophila* and mammals, the SMC proteins in *A. thaliana* (Meyer, 2005; Gause et al., 2008; Uhlmann, 2008). It was recently established that in *Arabidopsis* and maize, the *smc*1-*smc*3 heterodimer is a key determinant of meiosis and is an indispensable regulator of meiotic recombination in eukaryotes (Bolaños-Villegas, 2021). The literature is non-existent for SMC genes and proteins functions within Rhodophyta species, and a better understanding of the establishment of cohesion and genome organization in crops with economically important red algae may contribute to increases in yield under less-than-ideal environmental conditions (Bolaños-Villegas, 2021).

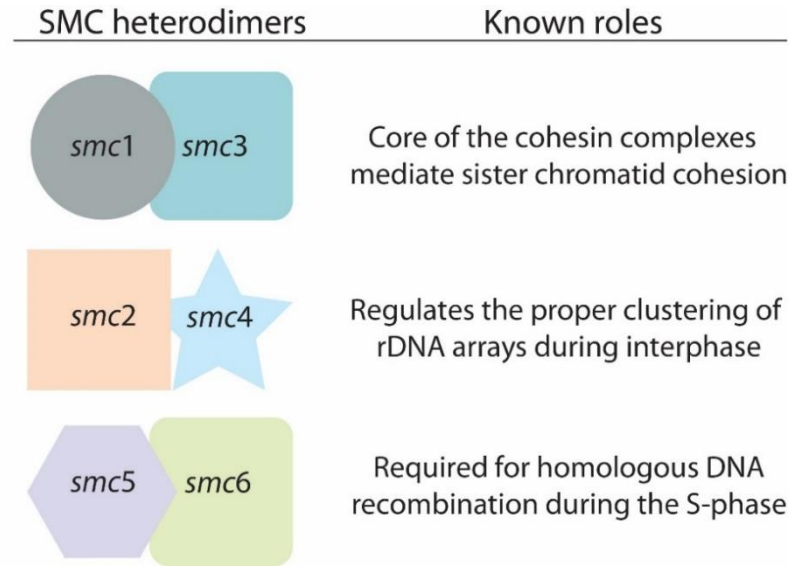

Figure 4. 2. Summary of roles for the SMC proteins in their heterodimer formation

scc3 gene

Chromosomes replicate during the S-phase of the cell cycle and from that time on consist of two sister chromatids (Pasierbek et al., 2003). Sister chromatid cohesion is fundamental for the faithful transmission of chromosomes during both meiosis and mitosis, ensuring that each daughter cell receives an equal complement of genetic information through replication (Pasierbek et al., 2003; Li et al., 2018). This cohesion is facilitated by cohesin, a conserved four-subunit protein complex that binds distinct regions of chromatin either intra- or inter-molecular (Rollins et al., 2004; Orgil et al., 2015; Li et al., 2018; Pathania et al., 2021). Two of the four

core cohesin subunits, *smc*1 and *smc*3, belong to the SMC gene family; the other two are the *scc*3 and *scc*1 (Ball & Yokomori, 2001; Uhlmaan, 2001; Pasierbek et al., 2003; Pathania et al., 2021). The cohesin is essential for mediating proper chromosome structure and dynamics, maintaining genome stability (Guacci et al., 1997; Michaelis et al., 1997; Tóth et al., 1999; Jessberger, 2002; Collier et al., 2020). In addition, the cohesin regulates gene expression and is involved in DNA repair (Donze et al., 1999; Dorset, 2007; Ünal et al., 2007; Schaaf et al., 2009). Although the Sister-Chromatid Cohesion 3 (*scc*3) gene was long ago identified (Ptashne & Gann, 1997; Tóth et al., 1999), the role of the gene and its protein was more recently established, with evidence for playing a major role in the initial step of chromosome entrapment and/or possibly DNA translocation (Li et al., 2018; Collier et al., 2020). The authors suggest that the *scc*3-*scc*1 proteins are responsible for the recruitment of cohesin complexes to chromosomes and, therefore, for cohesin to faithfully execute its functions during cell division (Li et al., 2018). In the absence of *scc*3, there is no sister-chromatid formation in meiosis (Pasierbek et al., 2003), which is considered essential for cell viability (Orgil et al., 2015) and the most important of cohesin regulatory subunits (Roig et al., 2014).

The available studies about the *scc*3 gene derive largely from studies with *Saccharomyces cerevisiae* (yeast) (Ptashne & Gann, 1997), but over the years this gene and homologs were described in other organisms. For example, in mammals and *Schizosaccharomyces pombe* (yeast), the Stromal Antigen 3 (*stag*3) and *rec*11 genes are *scc*3 homologue and are specific to meiosis (Pezzi et al., 2000; Pasierbek et al., 2003; Chelysheva et al., 2005) and involved in sister chromatid cohesion (Prieto et al., 2001). In yeast species, the *scc*3 homologs are *psc*3 and *rec*11 genes (Krawchuk et al., 1999; Tomonaga et al., 2000); in vertebrates, there are three known *scc*3 homologs, the *sa*1, *sa*2 and *sa*3 (Carramolino et al., 1997; Losada et al., 2000; Pezzi et al., 2000). In *Arabidopsis thaliana*, *scc*3 homologue is the AtSCC3 gene, which is involved in both meiotic and mitotic divisions (Chelysheva et al., 2005). It is worth noting that although some organisms possess a single copy of the *scc*3 gene (*S. cerevisiae, C. elegans* and *A. thaliana*), others have additional *scc*3 homologues (*S. pombe,* mammals and *Drosophila melanogaster*) with diversified functions (Pathania et al., 2021). Despite the importance of the *scc*3-*scc*1 heterodimer during cell replication, there are no studies to date that describe the *scc*3 gene function or in Rhodophyta species.

Stage 4. Diplotene

mutL, mlh, and pms1 genes

The DNA mismatch repair system (MMR) in eukaryotic cells is responsible for the repair of DNA mismatches that can result from several different mechanisms including DNA replication, such as genetic recombination, chemical modification of DNA or nucleotide pools, and control of meiotic recombination (Kneitz et al., 2000). The MMR enhances the overall fidelity of DNA replication up to 1000-fold (Lee et al., 2014). The main components of the MMR are known to be highly conserved throughout evolution (Modrich & Lahue, 1996). The DNA Mismatch Repair (*mut*L) family of DNA MMR plays a critical role in removing and repairing misincorporation errors during DNA replication, maintaining genomic integrity in somatic and meiotic cells by promoting accurate segregation of chromosomes during the Meiosis I division (Kaydrov et al., 2007; Furman et al., 2021; Pannafino & Alani, 2021). In eukaryotes, the *mut*L homolog (Mlh) genes are the *pms*1, *mlh*1, *mlh*2, and *mlh*3 and most eukaryotes encode at least two Mlh heterodimers complexes (Kaydrov et al., 2007; Reyes et al., 2020). Studies with yeast show that the Mlh form the *mut*Lα heterodimer (*mlh*1–*pms*1 genes), the *mut*Lβ heterodimer (*mlh*1–*mlh*2 genes), and the *mut*Lγ heterodimer (*mlh*1–*mlh*3 genes) (Fig. 4.3) (Wang et al., 1999; Reyes et al., 2020; Pannafino & Alani, 2021).

Evolutionary analysis of the *mlh* genes suggests that the *mlh*1 diverged prior to the splits that led to the *pms*1, *mlh*2, and *mlh*3 paralogs (Wolfe & Shields, 1997; Campbell et al., 2014; Furman et al., 2021). These studies mention that the *mlh*1 serving as a common partner for *pms*1, *mlh*2, and *mlh*3 is still one of the long-lasting paradigms of the eukaryotic cell life (Furman et al., 2021). Each heterodimer has a different role within the meiosis process and those are well established by the literature. The *mlh*1-*pms*1, analogous to its role in MMR, repairs mismatches that form in heteroduplex DNA during genetic recombination; the *mlh*1-*mlh*2 regulates gene conversion tract lengths; and the *mlh*1-*mlh*3 acts in the biased cleavage of double Holliday junctions to yield crossovers that are critical for the formation of gametes (Hunter &Borts, 1997; Harfe et al., 2000; Abdullah et al., 2004; Zakharyevich, et al., 2012; Campbell et al., 2014; Manhart & Alani, 2016; Al-Sweel et al., 2017; Duroc et al., 2017; Furman et al., 2021). Although the roles of each *mlh* gene are well studied in eukaryotes, most of the work has been done using animal models and yeast and literature on algae is lacking.

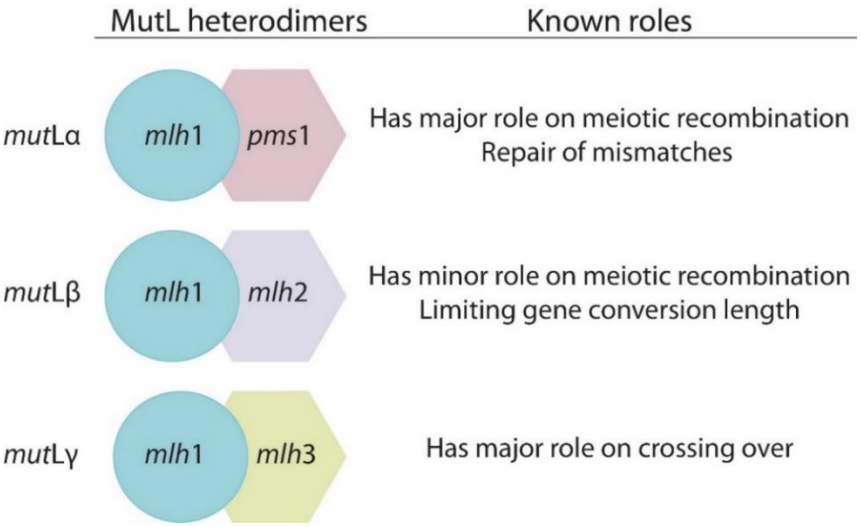

Figure 4. 3. Summary of roles for the MLH heterodimers proteins in MMR and meiotic recombination (Figure based on Reyes et al., 2020, and Pannafino & Alani, 2021).

mutS and msh gene family

During meiosis, the interactions between homologous chromosomes have the potential to generate mismatched DNA structures. Several proteins play a crucial role in DNA mismatch repair (MMR), such as the *mut*S homologs (*msh* genes) which are known to influence meiotic recombination in several eukaryotes (Blackwell, 2020). They initiate the MMR by recognizing mismatched nucleotides and then transmitting this discovery to downstream editing machinery, playing a key role in post-replicative mutation correction and rejection of heteroduplex DNA during recombination (Lee et al., 2014; Blackwell et al., 2020). Most eukaryotes contain up to seven *msh* genes (*msh*1–7), although the mitochondrial-localized *Saccharomyces cerevisiae msh*1 and the *Arabidopsis thaliana msh*7 do not appear to be conserved (Snowden et al., 2004). Eukaryotes have duplicated and refined the *mut*S gene to form two *msh* heterodimers (*mut*Sβ: *msh*2-*msh*3, and *mut*Sα: *msh*2-*msh*6) (Fig. 4.4), which are known for being widely conserved complexes (Fishel & Wilson, 1997; Kolodner & Marsischky, 1999; Blackwell et al., 2020; Reyes et al., 2020).

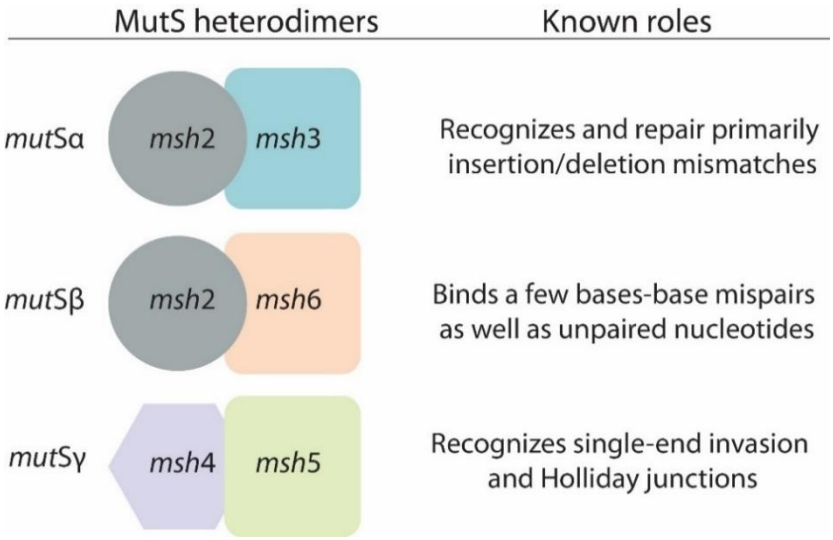

Figure 4. 4. Summary of roles for canonical MSH protein heterodimers in MMR and meiotic recombination (Figure based on Manhart & Alani, 2016, and Reyes et al., 2020).

The *msh* heterodimers are well studied, and their functions are established in the literature for animals and some Viridiplantae species. The *msh*2–*msh*3 complex is primarily responsible for binding and repair of larger insertion/deletion mispairs as well as unpaired nucleotides during meiosis (Marsischky et al., 1996; Flores-Rozas & Kolodner, 1998; Kolodner & Marsischky, 1999; Wilson et al., 1999; Gradia et al., 2000; reviewed in Manhart & Alani, 2016). The *msh*2-*msh*6 heterodimer has a broader ability to function in the repair of mispaired bases than the *msh*2-*msh*3, recognizing base-base mispairs and single insertion/deletion unpaired nucleotides (Marsischky et al., 1996; Genschel et al., 1998; Kolodner & Marsischky, 1999; Wilson at al., 1999; Gradia et al., 2000). Some eukaryotes also have *msh*4 and *msh*5 genes, which form the heterodimer *mut*Sγ and play essential roles in crossover formation in diverse eukaryotes (Paquis-Flucklinger et al., 1997; Her & Doggett, 1998; Haversat et al, 2021). The *mut*Sγ recognizes single-end invasion intermediates and Holliday junctions *in vitro* (Lahiri et al., 2018; Snowden et al., 2004). In addition, MutSγ recruits and activates the resolvase activity of *mut*Lγ, enabling biased processing of Holliday junctions into crossovers during meiosis (Cannavo et al., 2020; Kulkarni et al., 2020). In yeast, *S. cerevisiae*, there is the *msh*1 gene, which protein has a function in mitochondrial genome stability, but the gene was not observed in any eukaryote to date (Reenan & Kolodner, 1992; Chi & Kolodner, 1994). In *A. thaliana* there is also the *msh*2-

*msh*7 heterodimer, which recognizes mismatched nucleotides and has demonstrated roles in repressing meiotic crossovers in hybrid plants (Blackwell, 2020). In the Rhodophyta, the *msh*4 and *msh*5 genes have been identified in *Thorea hispida* (Nan et al., 2020) and are the only study to date that has observed *mut*S genes homologs in Rhodophyta.

Marsischky et al. (1996) identified that in the cellular nucleus, the recognition of mismatched nucleotides requires *msh*2–*msh*3 and *msh*2–*msh*6 complexes, along with the *mlh*1-*pms*1 heterodimer. This discovery led to phylogenetic studies of the *mut*S gene, which demonstrates that the gene is an example of a single bacterial gene that duplicated and evolved early in eukaryogenesis (the evolutionary process leading to the origin of the eukaryotic cell) to yield gene families with distinct vegetative growth and meiotic functions (Furman et al., 2021). The gene is thought to have entered archaea and eukaryotes through horizontal gene transfer, followed by a series of gene duplications (Eisen, 1998; Lin et al., 2007).

Although no phylogenetic relationship has been shown between the *mut*S and *mut*L homologs, they are considered part of the same pathway for their cellular functions by some authors (MutHLS pathway) (Paquis-Flucklinger et al., 1997; Kneitz et al., 2000), and together the genes are responsible for DNA integrity and represent a barrier to genetic recombination during interspecific sexual conjugation in prokaryotes and eukaryotes (Paquis-Flucklinger et al., 1997). The link between *msh* and *mlh* proteins is newly discovered (Furman et al., 2021), where the *msh* proteins recognize base-base and insertion/deletion mismatches that escape DNA polymerase proofreading and transmit the recognition signal to downstream repair proteins such as the *mlh* protein family, which coordinate excision of the replication error and DNA resynthesis using the parental DNA strand as a repair template (Fig. 4.5). Despite the known presence of some of the *msh* genes in *Thorea hispida* (Nan et al., 2020), there is no information about how and if the *msh* and *mlh* genes are connected in red algae.

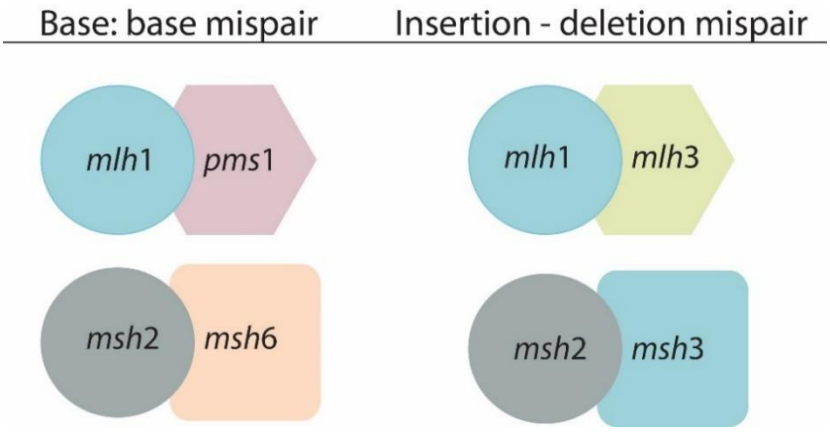

Figure 4. 5. Interaction between complexes of *mut*L and *mut*S proteins during MMR (Figure based on Kolodner & Marsischky, 1999, and Furman et al., 2021).

4.3 Material and methods

4.3.1 Sequence data

The genes of interest within the set of eukaryotic meiotic genes were identified based on literature (Table 4.1) and the correct queries were selected based on the available genomic data of model organisms (*Arabidopsis thaliana*, rice) from UniProt (The UniProt Consortium, 2017) and NCBI. Using a representative protein sequence for each gene, a blastp search was done against the Rhodophyta proteomes and the results were carefully examined. The sequences that had significant matches/hits (*E*-value < 0.001) were retrieved for further analysis (Figure 4.6).

To search for meiotic-related genes in *B. atropurpurea*, the contigs obtained by the environmental material sequence described on Chapter 2 was used. The protein sequences of the model organisms were also used to search for the meiotic genes in the *B. atropurpurea* contigs through tblastn. The homologous contigs with 90%, or higher, similarity were flagged and the DNA sequence corresponding to the query protein sequence was retrieved and translated using ExPASy (Artimo et al., 2012). The translation was then blastp searched against the NCBI's non-redundant protein sequences database as well as the Model Organisms (landmark) database to verify the authenticity of the newly translated sequence (Figure 4.6). A similar method was used to identify and retrieve the meiotic genes in the Rhodophyta species *Digenea simplex*, *Pyropia haitanensis* and *Galdieria phlegrea* since their partial or complete genomes are available, but the

proteins in questions are not published and do not show in the BLAST result. The novel data produced by this study will be submitted to GenBank.

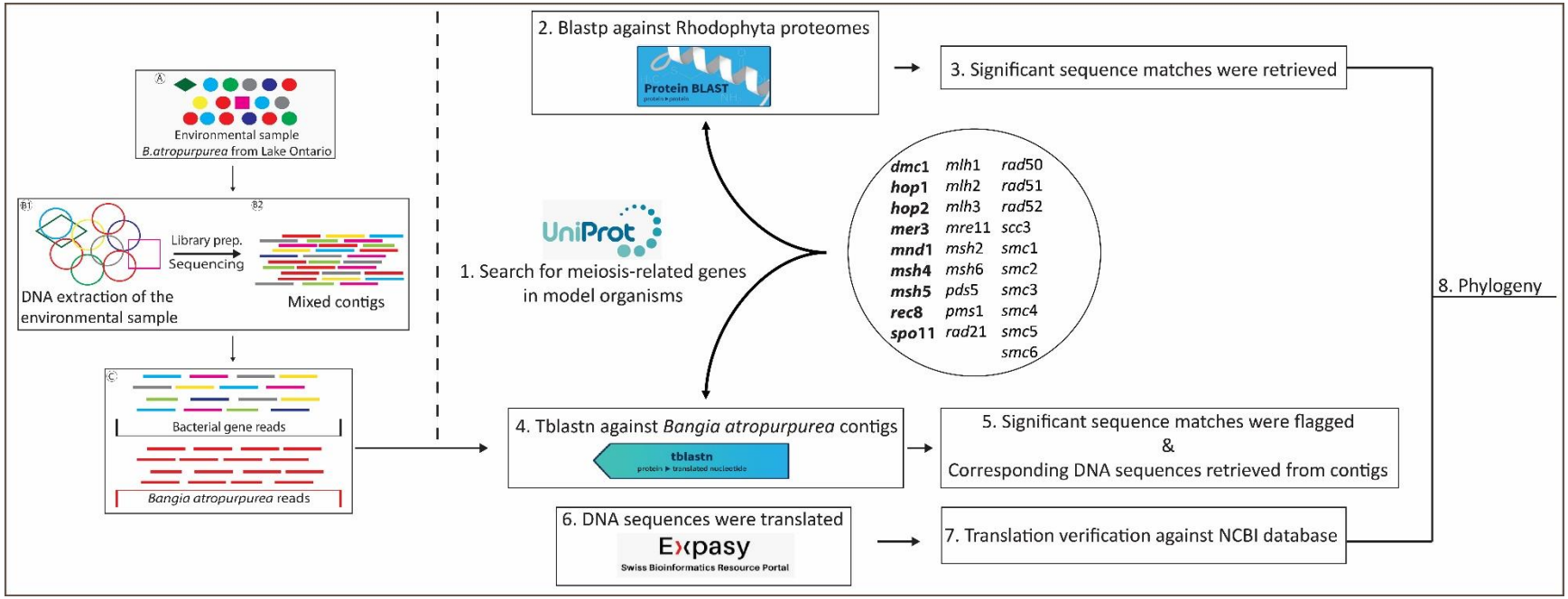

Figure 4. 6. Bioinformatic workflow to retrieve meiosis-related genes present in *B. atropurpurea* contigs. Steps on the left of the dashed line (A, B1, B2 and C) follow the steps described on Chapter 2, prior to the analysis of the genes. The meiosis toolkit in *A. thaliana*, corn and rice is composed of the genes inside the circle, where genes in bold are consider exclusive for meiosis. Identification and retrieval of these protein sequences was done on UniProt (1) and using the blastp option on NCBI the model organisms' sequences were blasted against Rhodophyta proteomes (2). Sequences with an e-value < 0.001 were considered significant and retrieved for phylogenetic analysis (3). The same sequences from the model organisms were used to search for meiosis genes in *B. atropurpurea* contigs with the tblastn option on NCBI (4). Homologous sequences with 90% or higher were flagged, and the corresponding DNA sequences were retrieved from the contigs and translated (5,6). To verify the accuracy of the translated sequences, the protein sequences of *B. atropurpurea* were blasted against NCBI database. All the complete sequences from A. thaliana and Rhodophyta species, along with the partial sequences of *B. atropurpurea* were used to built phylogenetic trees (8).

4.3.2 Phylogenetic analyses of Meiotic Genes and Gene Families

The evolutionary inferences among the eukaryotic meiotic genes were described by Ramesh & Malik (2005) and Malik et al. (2008), and the same was applied in this study (Fig. 4.6). The protein sequences of model organisms, Rhodophyta species and *B. atropurpurea* were uploaded to the Geneious Prime software (www.geneious.com) and aligned using ClustalW (Higgins et al., 1996). The alignment was exported to Seaview (Gouy et al., 2010), where the GBlocks option was used to select the most conserved sites within the protein sequences for phylogenetic analysis purposes (Castresana 2000; Talavera and Castresana, 2007). A maximum-likelihood (ML) tree was constructed based on the Gblocks using MegaX (Kumar et al., 2018); all the trees were built having 1000 bootstrap replicates and following the default parameters. The trees were edited and finalized with Adobe Illustrator (Adobe Inc. 2019).

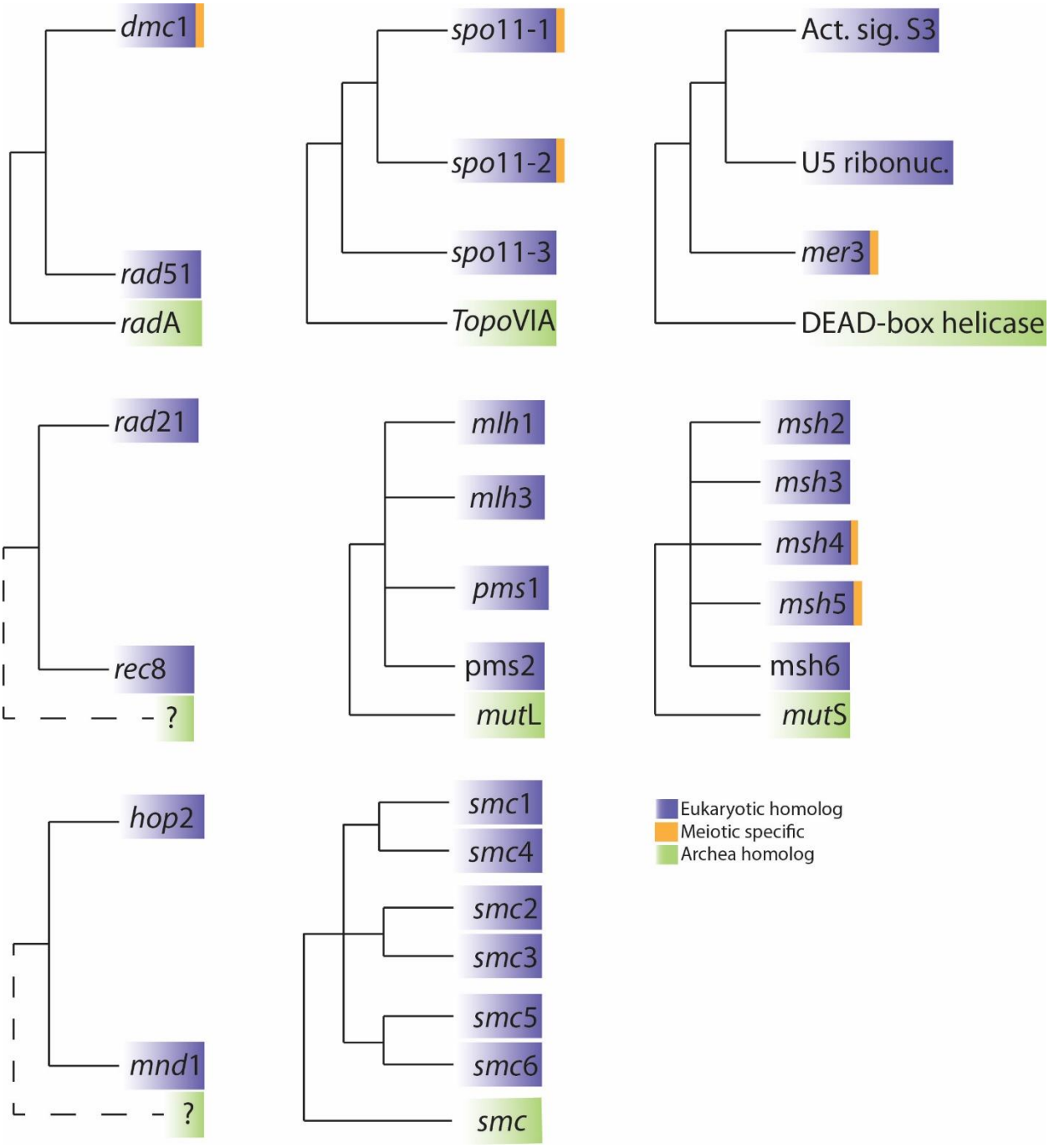

Figure 4. 7. Phylogenetic relationship of miosis-related genes according to existent literature (Cobbe & Heck, 2004; Ramesh & Malik, 2005).

4.4 Results and Discussion

In this study, a total of 29 meiotic related genes in Viridiplantae species (*rad*51, *dmc*1, *mnd*1, *hop*1, *hop*2, *smc*1, *smc*2, *smc*3, *smc*4, *smc*5, *smc*6, *spo*11, *scc*3, *pms*1, *mlh*1, *mlh*2, *mlh*3, *msh*2, *msh*4, *msh*5, *msh*6, *mer*3, *mre*11, *pds*5, *rad*1, *rad*21, *rad*50, *rad*52, *rec*8) were analyzed to verify the presence of these genes in the species *Bangia atropurpurea*. Of these genes, 13 are present in *B. atropurpurea* (*rad*51, *mnd*1, *hop*2, mer3, *smc*1, *smc*4, *smc*5, *spo*11, *scc*3, *pms*1, *mlh*1, *mlh*3, *msh*6). In addition, *B. atropurpurea* also has a copy of the meiotic-related genes *msh*3, *mut*S, *mut*L genes, which are considered ancient homologs and were identified in species of Archaea. A detailed discussion of these genes will be presented next, with exception of genes part of the *mlh* (*mlh*1, *mlh*2, *mlh*3, *pms*1 and *mut*L) gene family. Our results indicate that further investigations are needed to conclude how the *mlh* genes are related among species and to verify the copies present in *B. atropurpurea* (Appendix L).

4.4.1 *spo*11 gene

The blastp search against Rhodophyta resulted in 11 matches with *E*-value < 0.001 for the *spo*11 gene. The tblastn search against the metagenomic data of *B. atropurpurea* had one match that was confirmed to be part of the *spo*11 gene. The phylogenetic tree of the *spo*11 gene has well-supported branches among the different taxonomic groups within Rhodophyta species, including the Bangiales species *P. umbilicalis*, and *P. haitanensis* (Fig. 4.7). A copy of the gene is also present in *B. atropurpurea* and follows the expected grouping within the Bangiales. However, the analysis shows that the gene of *B. atropurpurea* homolog is closer related to *P. haitanensis*, sharing a common ancestor, than to *P. umbilicalis*. Future analysis with the addition of other *Porphyra* and *Pyropia spo*11 genes may help elucidate the closest related species with *B. atropurpurea*.

In Rhodophyta, the studies that mention meiotic genes are usually transcriptomes analyses and bring interesting information because of the red algae biphasic and triphasic life history. The transcriptome of the marine species *Porphyra purpurea* revealed that certain genes are regulated differently according to the life phase, conchocelis and blade, from sexual to asexual reproduction (Chan et al., 2012). More recently, a transcriptome study of the freshwater

species *Thorea hispida* detected an active *spo*11 gene during the sexual phase of the organism (Nan et al., 2020). Therefore, although there is a limited number of studies with red algae that elucidate nuclear transcriptome, there is an indication that *spo*11 is part of the meiosis event. Moreover, there are no studies that confirm or deny that red algae species have other *spo*11-like genes, but a more in-depth analysis, with more species, is needed to rule out this possibility. The literature about the biochemistry and functionality of the *spo*11 gene in Rhodophyta is non-existent to date but will be interesting to see new research investigating where in the eukaryotic evolution does the *spo*11 genes of red algae are located. The presence of this gene in Rhodophyta, especially in a species that is known for being asexual, leaves more questions than answers, such as, does the protein act exclusively to start miosis? Does it also play a role in chromosome synapsis? Do some species have other *spo*-like proteins like in Viridiplantae?

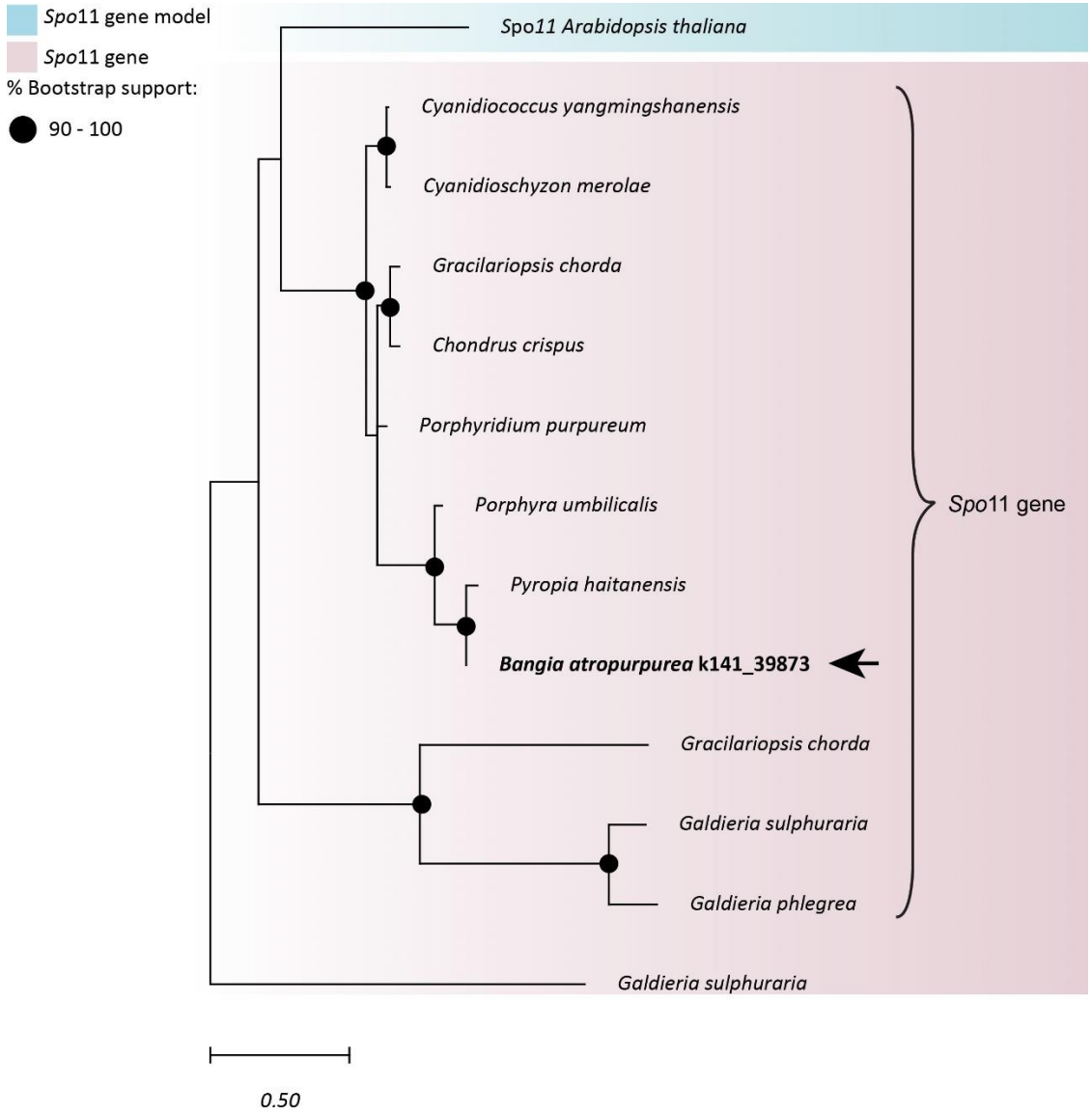

Figure 4. 8. Evolutionary analysis by Maximum Likelihood method of the *spo*11 genes. The arrow indicates the presence of the gene in *Bangia atropurpurea* metagenome, and the number is the reference of the sequence fragment (contig) that was used for this analysis.

4.4.2 *rad*51 and *dmc*1 genes

The blastp search against Rhodophyta species resulted in 9 hits for the *rad*51 gene, and 7 hits for the *dmc*1 gene (all hits with *E*-value < 0.001). The tblastn search against the metagenomic data of *B. atropurpurea* resulted in one hit for *rad*51 gene, and no hits for *dmc*1 gene. The maximum likelihood tree of the paralog genes *rad*51 and *dmc*1 shows the node that differentiates both genes to be well-supported, as well as for most Rhodophyta species (Fig. 4.8). The gene *rad*51 is present in *B. atropurpurea* and this phylogenetic tree shows to be closely related to the *rad*51 gene of *Porphyra umbilicalis* and *Pyropia haitanensis*. This result agrees with the phylogenetic analysis of Bangiales species that indicates the close relationship of *B. atropurpurea* and *Porphyra/Pyropia* species (Müller et al., 1998; Yoon et al, 2006; Sutherland et al., 2011; Chapter 3 of this thesis). The *dmc*1 gene, however, is not present in *B. atropurpurea* and, to date, there is no published evidence of the gene to be present in other Bangiales species. Other Rhodophyta species do have a copy of the gene, including the ones observed in hot springs (*Cyanidioschyzon merolae* and *Galdieria sulphuraria*), the microalgae *Porphyridium purpureum*, and the macroalgae *Gracilariopsis chorda*, *Chondrus crispus* and *Digenea simplex*.

It is interesting to observe that Bangiales species have a copy of the *rad*51 gene (support gene) and do not have a copy of the *dmc*1 gene (essential to meiosis). Given that Bangiales species have a biphasic life cycle, going from asexual to sexual, the *dmc*1 gene would play a vital role during meiosis. Since *B. atropurpurea* is only known for its asexual reproduction, it could be expected for the absence of the dmc1 gene, but the other Bangiales species that have a sexual reproduction should have a copy of the gene. The order Bangiales is part of the Bangiophyceae class, which is a sister group from Florideophycea. The Florideophyceae species are also known for having a triphasic life cycle and, curiously, my analysis shows that 3 Florideophyceae species have a copy of the *dmc*1 gene (*G.chorda*, *C. crispus* and *D. simplex*) (Fig. 4.8). The possible loss of the *dmc*1 gene in Bangiales has not been studied yet and future analyses are needed to conclude if and where in the evolution the *dmc*1 gene was lost, and how the species overcome not having such a critical gene during their sexual phase. It is important to note that the non-observation of the *dmc*1 gene in *B. atropurpurea* can be due to the methodology used in this study, metagenomic sequencing dataset, and a search using primers for

the gene of full genome sequencing could be a useful technique to answer definitively if the gene is present or not in *B. atropurpurea.*

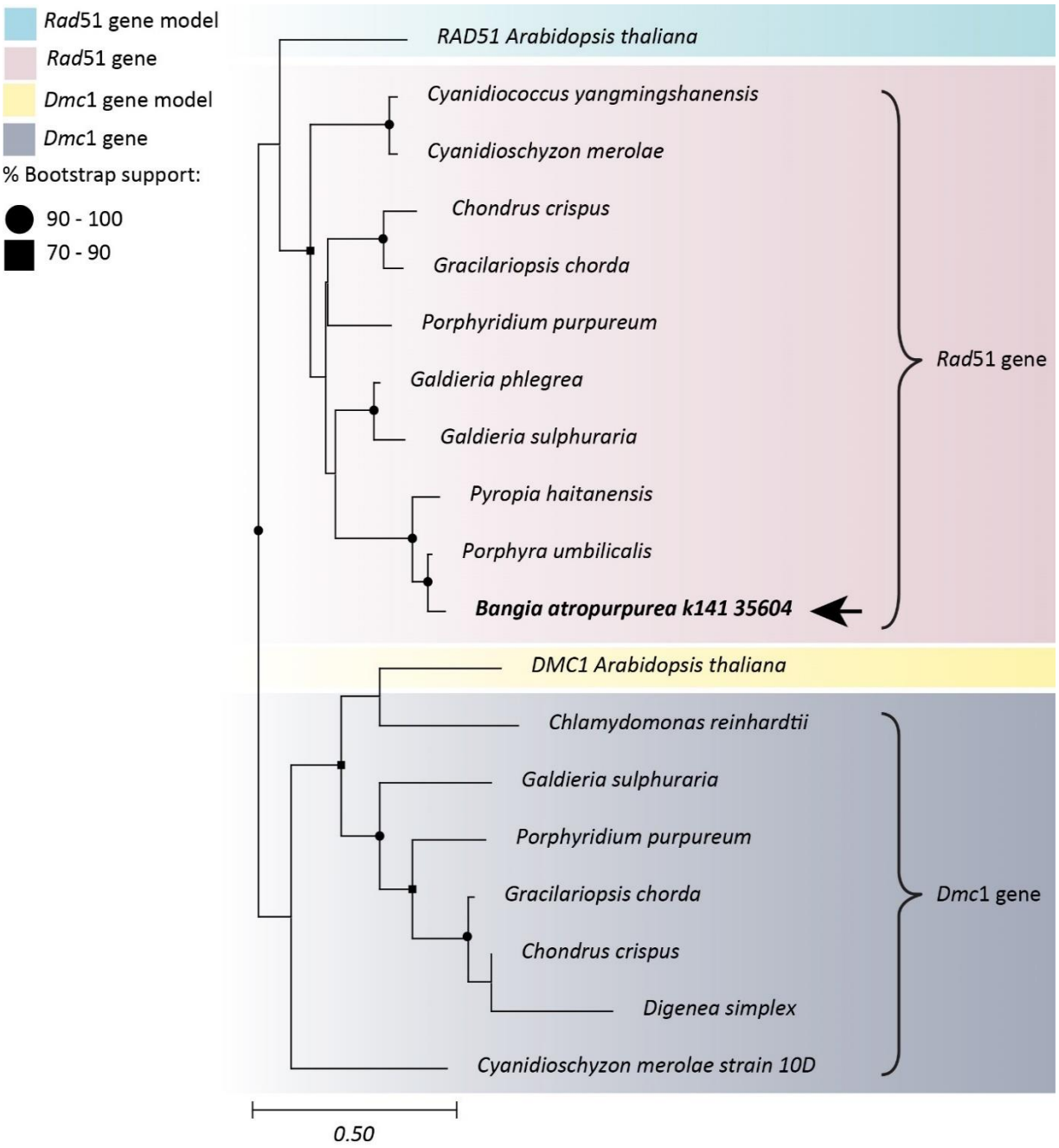

Figure 4. 9. Evolutionary analysis by Maximum Likelihood method of the *rad*51 and *dmc*1 genes. The arrow indicates the presence of the gene in *Bangia atropurpurea* metagenome, and the number is the reference of the sequence fragment (contig) that was used for this analysis.

4.4.3 *mnd*1 and *hop*2

The blastp search against Rhodophyta resulted in 10 hits for *mnd*1 gene, and 9 hits for the *hop*2 gene (all hits with *E*-value < 0.01). The tblastn search against the *B. atropurpurea* metagenomic sequence data had hits for both *mnd*1 and *hop*2 genes. The maximum likelihood tree of the genes *mnd*1 and *hop2* shows that both genes are present in *B. atropurpurea* and in both cases the genes are closely related to *Porphyra umbilicalis* and *Pyropia haitanensis*, reinforcing the taxonomic relationship of the species with well-supported branches (Fig. 4.9). The observation of these homolog genes in many other organisms strongly suggests that the function of the genes is conserved among eukaryotes (Tsubouchi & Roeder, 2002), and our data contributes to this affirmation since the genes are present in Rhodophyta species of different taxonomic orders. There is a gap in knowledge around these meiotic genes in Rhodophyta species, where the literature about their evolution and function within reproduction are non-existent. The presence of both genes in *B. atropurpurea* is intriguing since the species is generally known to be asexual and both genes are meiosis-specific genes. Future studies with other asexual Rhodophyta species can help elucidate how *mnd*1 and *hop*2 genes evolved among red algae and if other asexual species also have a copy of these genes in their genome. Moreover, future sequencing data of the complete nuclear genome of Rhodophyta species will be able to determine the phylogeny and conservation of *mnd*1 and *hop2*.

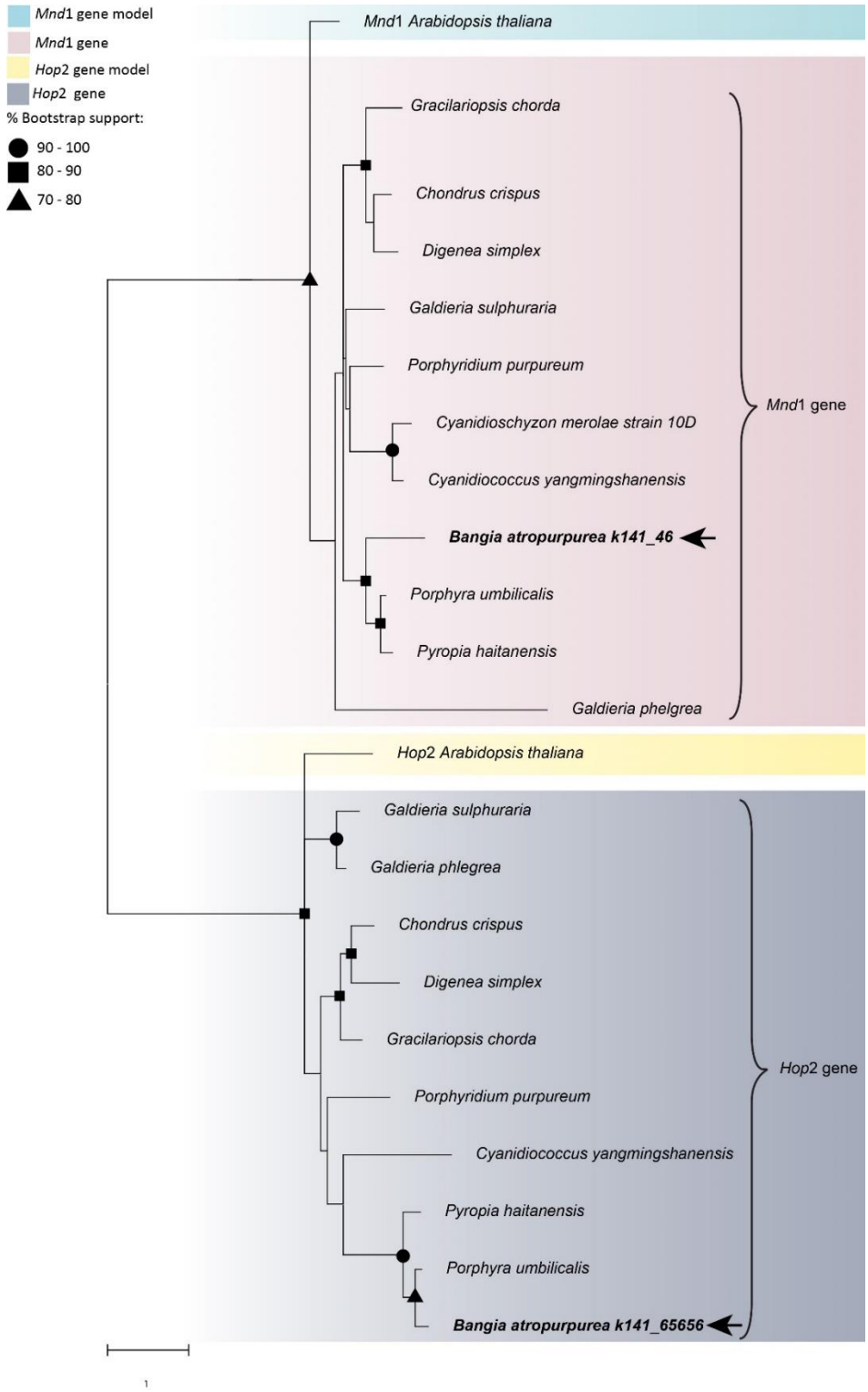

Figure 4. 10. Evolutionary analysis by Maximum Likelihood method of the *mnd*1 and *hop*2 genes. The arrow indicates the presence of the gene in *Bangia atropurpurea* metagenome, and the number is the reference of the sequence fragment (contig) that was used for this analysis.

4.4.4 *smc* gene family

The blastp search against Rhodophyta resulted in 8 hits for the *smc*1 gene, 6 hits for the *smc*2 and *smc*3 genes, 7 hits for the *smc*4 gene, 5 hits for the *smc*5 gene and 7 hits for the *smc*6 gene (all with E-values < 0.001). The tblastn search against the *B. atropurpurea* metagenomic data of had 3 hits, for the *smc*1, *smc*4 and *smc*5 genes. The phylogenetic analysis of these 6 SMC genes shows that the gene family is conserved among Rhodophyta species, where *Gracilariopsis chorda, Chondrus crispus, Porphyra umbilicalis, Porphyridium purpureum* and *Cyanidioschyzon merolae* have at least one copy of each gene (Fig. 4.10). Curiously, there are two *smc*1-like genes in *Porphyra haitanensis*, but no other *smc* gene was observed in the BLAST search. These two *smc*1 genes are different from each other, forming another branch in the phylogenetic tree. The species *Galdieria phlegrea* also only has a copy of the *smc*1 gene and no other SMC gene was detected during blasting against the Rhodophyta group. On the other hand, the species *G. sulphuraria* does not have a copy of the *smc*1 and *smc*5 genes. All of the analyzed rhodophyte species likely have the 6 *smc* genes in their nuclear genome and are well known for sexual reproduction, but due to the lack of data within nuclear genes of red algae species, an in-depth comparison is still not possible. Future investigations are needed to confirm the existence of these missing genes in sexual red algae species.

The species *Bangia atropurpurea* has a copy of the *smc*1, *smc*4, and *smc*5 genes. These are exactly one gene of each known gene that form heterodimers of the *smc* gene family (*smc*1-*smc*3; *smc*2-*smc*4, and *smc*5-*smc*6). It is possible that *B. atropurpurea* has a copy of the *smc*2, *smc*5 and/or *smc*6 but was unable to be identified with the methodology used in this study, metagenomic sequencing dataset, and without a complete nuclear genome available for analysis. In addition, *P. umbilicalis,* species that is closely related to *B. atropurpurea* has all the *smc* heterodimers, which supports the idea of *B. atropurpurea* to also have all the copies of the genes. Hence the importance of whole-genome sequencing and molecular studies with red algae, which have a relevant role in Eukaryotic evolution and still many details of their reproduction and cytology remain unclear.

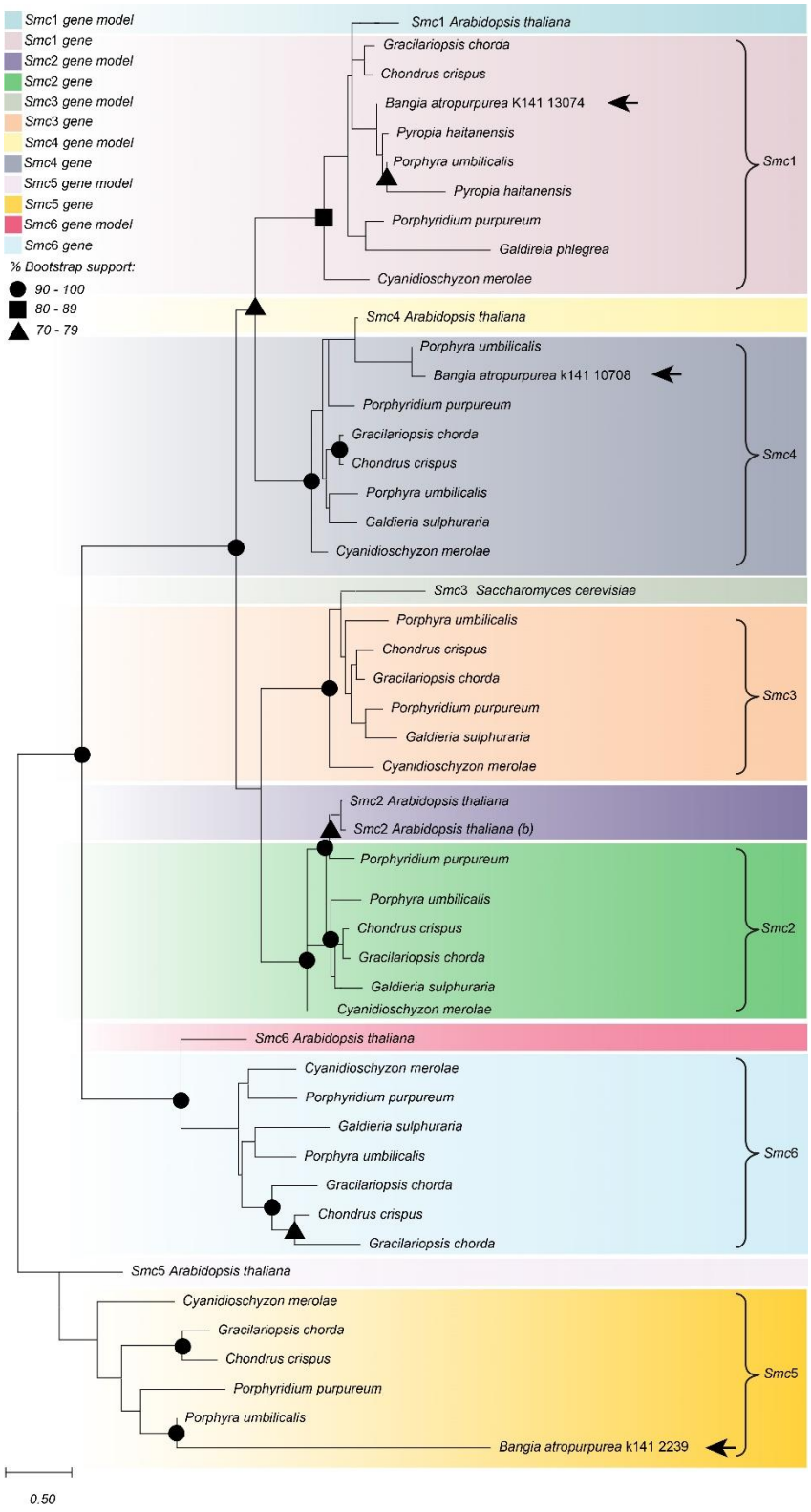

Figure 4. 11. Evolutionary analysis by Maximum Likelihood method of the *smc* gene family. The arrow indicates the presence of the gene in *Bangia atropurpurea* metagenome, and the number is the reference of the sequence fragment (contig) that was used for this analysis.

4.4.6 *scc*3 gene

The blastp search for the *scc*3 gene was done against Rhodophyta, with 11 hits, and against Chlorophyta, with the top 3 hits selected for this study (all hits had *E*-values < 0.001). The inclusion of Chlorophyta species helped not only to visualize division between the green and red organisms but also to confirm the relationship among the species. The tblastn search against the *B. atropurpurea* metagenomic sequence data resulted in one hit. The phylogenetic tree of the *scc*3 gene has well-supported branches among the different taxonomic groups within Rhodophyta species. Three species of green algae (*Chlorella sorokuniana, Ostreococcus lucimarinus* and *Ostreococcus tauri*) group together and share the same ancestral to the Viridiplantae species *A. thaliana,* confirming the course of evolution from green algae to land plants. A copy of the gene is present in *B. atropurpurea* and the species follows the expected grouping within the Bangiales species *Pyropia haitanensis* (Fig. 4.11). Curiously, the species *Porphyra umbilicalis,* which is closer related taxonomically to *B. atropurpurea,* does not have a copy of this gene. The *scc*3 gene present in *P. haitanensis* is the only Bangiales species with available data for comparison to date. The species *Gracilariopsis chorda* (Florideophyceae) has more than one copy of the gene. It will be interesting to observe if other Bangiales have a copy of this gene and if other Florideophyceae species have more copies of the gene in future studies. It is known that the absence or mutation of the *scc*3 gene in other organisms results in cohesion defects (Rollins et al., 1999; Losada et al., 2000; Remeseiro et al., 2012), but it remains to be seen if organisms lacking meiotic *scc*3 homologs have selected another unrelated protein to provide some cohesin function in place of meiotic *scc*3 or if the sole *scc*3 protein deals with the specific meiotic *scc*1 homolog (*rec*8) (Chelysheva et al., 2005).

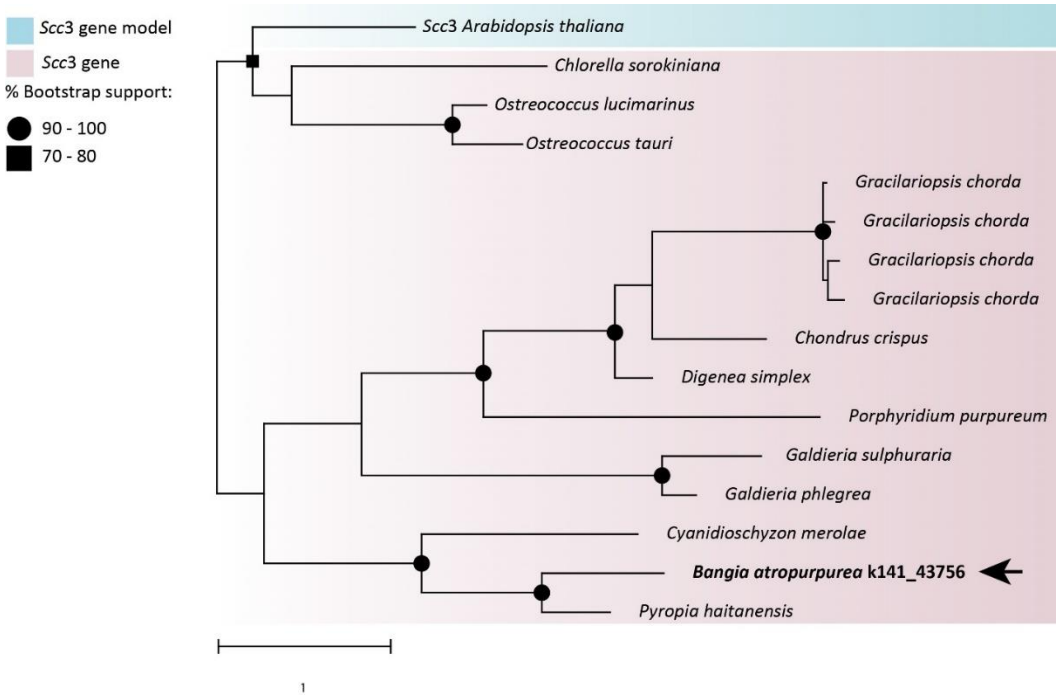

Figure 4. 12. Evolutionary analysis using Maximum Likelihood method of the *scc*3 gene. The arrow indicates the presence of the gene in *Bangia atropurpurea* metagenome, and the number is the reference of the sequence fragment (contig) that was used for this analysis.

4.4.8 *msh* and *mut*S gene family

The blastp search against Rhodophyta resulted in 10 matches for the *mut*S gene, 3 matches for the *msh*2 gene, 4 matches for *msh3* gene, 6 matches for *msh*4 gene, 10 matches for the *msh*5 gene, and 6 matches for the *msh*6 gene (all matches with *E*-value < 0.001). Due to the complexity of the sequences and the number of genes part of this gene family, a couple of genes were also blastp searched against Chlorophyta. The species with E-values < 0.001 were selected to be part of the analysis, which resulted in 6 matches for the *msh*4 gene, and 4 hits for the *msh*6 gene. The tblastn search against the *B. atropurpurea* metagenomic sequence data resulted in one match for the *mut*S gene, 2 matches for the *msh*3 gene, and one match for the *msh*6 gene.

The phylogenetic analysis of the *mut*S gene and 5 *msh* genes in red algae shows that they are conserved among Rhodophyta species, with branches having high support in all of the analyzed genes (Fig. 4.12). The species *P. umbilicalis* has a copy of all 6 genes and is grouped with the red algae species *G. chorda* and *C. crispus* for *mut*S, *msh*5, *msh*2, and *msh*4 genes. The grouping of a Bangiales species (*P. umbilicalis*) with other Rhodophyta orders is a result of the lack of available data among the Bangiales species. It can be observed that *B. atropurpurea* has a copy of the *mut*S, *msh*3 and *msh*6 genes. The *mut*S gene of *B. atropurpurea* appears to be very different from the other species and does not group with any Rhodophyta species. Most likely, this unexpected result is due to the size of the contig extracted from the data from *B. atropurpurea*, which does not represent a whole gene sequence. Further analysis is needed to verify if the complete sequence of the *mut*S gene in *B. atropurpurea* falls within the Bangiales and Rhodophyta, or if the gene is indeed very different from the other organisms. The tblastn results also indicate two contigs with similarity higher than 90% that match the *msh*3 gene from other red algae, one (k141_31547), that grouped with *P. umbilicalis* with a high support node; And another contig (k141_37764) that like *mut*S gene, is separated by the other Rhodophyta species in a low support node. An in-depth look at these contigs can maybe answer if only the first is indeed part of the *msh*3 gene, or even if both contigs are complementary. The *msh*6 gene of *B. atropurpurea* is well resolved, and the gene groups with *P. umbilicalis* on a high supported node.

The structure of the *msh* heterodimers and roles of the genes that are part of the MMR system in meiosis is well established for both animals and plants (Lee et al., 2014; Blackwell,

2020; Blackwell et al., 2020). For Rhodophyta species, little is known of the MMR and there is a considerable lack of available data for comparative studies. Intriguingly, an asexual species would have sexual genes such as the *msh*, and even though the presence of some of *msh* genes in *B. atropurpurea* must be investigated further, this study presents strong evidence that these genes are present in its genome and could be active during the Bangiales evolution.

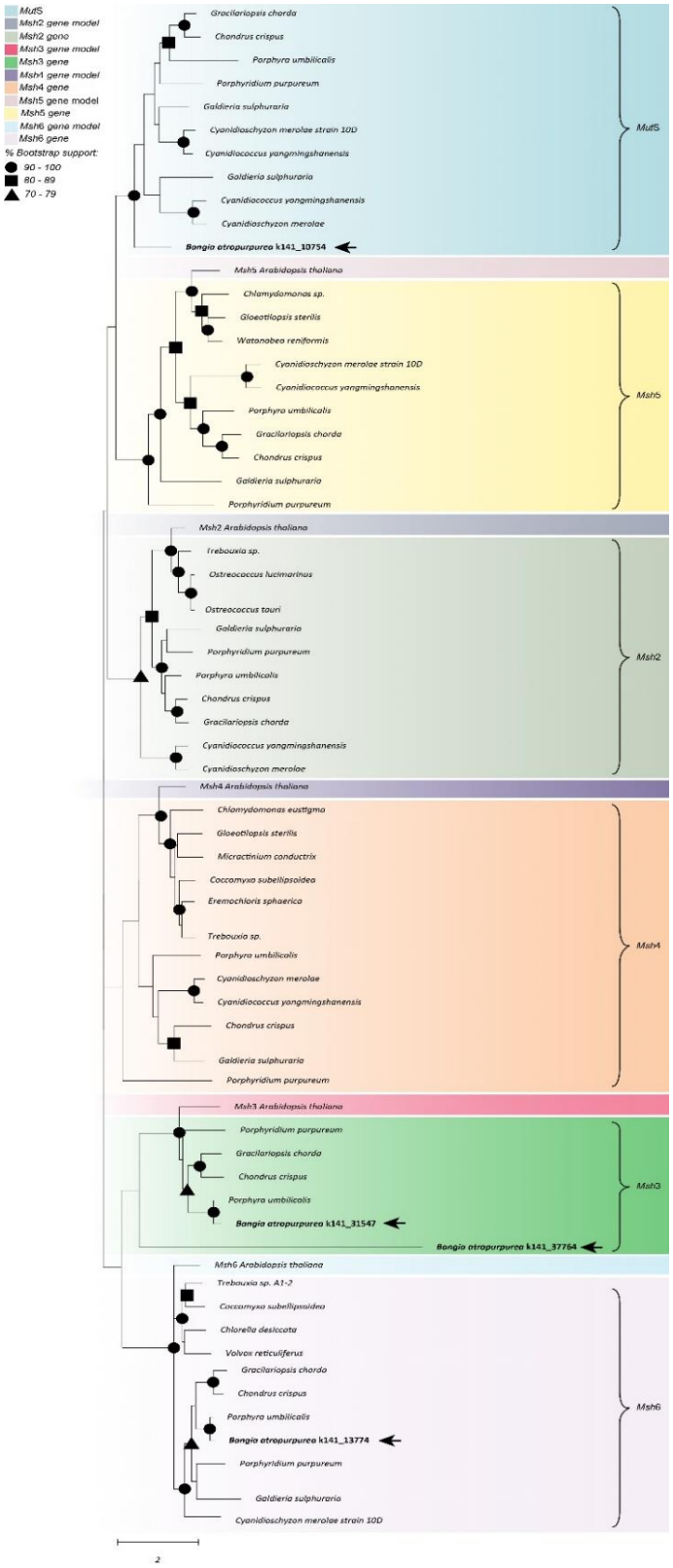

Figure 4. 13. Evolutionary analysis by Maximum Likelihood method of the *msh* gene family and *mut*S genes. The arrow indicates the presence of the gene in *Bangia atropurpurea* metagenome, and the number is the reference of the sequence fragment (contig) that was used for this analysis.

4.5 Conclusion

With the Eukaryotic Meiotic Toolbox as a reference, the present study made use of different BLAST approaches to successfully identify and retrieve meiotic genes from an environmental metagenomic sequencing sample of *B. atropurpurea*. From the 22 meiotic-exclusive genes known to be part of most eukaryotic meiosis machinery, 13 are present in *B. atropurpurea* which raises interesting questions about this organism that is suspected of only undergoing asexual reproduction. The meiotic genes observed in *B. atropurpurea* may be functional since no premature stop codons were detected in the analysis, and future investigations may need to examine if sexual reproduction is possible (e.g., Garguilo et al. 2001) and potentially address if it is occurring rarely in this species. It is also possible that these genes are functional but play a role in a different mechanism other than reproduction. To verify if the meiotic genes in *B. atropurpurea* are functional, or if they are pseudogenes, future analysis of the codon usage can be performed using the dataset of this chapter. The emphasis on the type of substitutions happening in the 1st, 2nd and 3rd positions of the gene sequences will be able to answer if the mutations are non-functional (1^{st} and 2^{nd}), or functional (3^{rd}), and if the substitutions are considered normal or radical. Thus, there is a need for other red algae meiotic genes to be sequenced and the nucleotide sequence to become available to proceed with the analysis of rates of substitution and codon usage, which will then show the differences and similarities of these genes in closely related species.

The use of metagenomic sequence is recommended for future studies searching for meiotic genes and has a few advantages over full genome sequencing, especially for red algae species that are difficult to grow in the lab. This methodology allows the successful search, retrieval, and identification of genes of interest based on a reference genome, with the addition of a large dataset, besides the genes of interest, that can be investigated as well (Chapter 2 of this thesis). There are, however, questions that cannot be answered solely by the metagenomic sequencing and the additional information generated by a whole-genome sequence might be beneficial. Not only verifying the completeness of the gene sequences but also determining the presence or absence of certain genes. The metagenomic sequencing has gaps that can be exactly where a small gene of interest might be, subsequently, the gene appears absent in BLAST searches causing misleading information if not verified by a different sequencing method.

Transcriptome data can also add significant details to the metagenomic sequencing, verifying the over/under-expression of the meiotic genes during the various reproductive stages observed in the Bangiales species. Overall, the *B. atropurpurea* metagenomic generated great results for the identification of meiotic genes, and a different sequence approach might help to better understand the reproduction and life history of the species.

The phylogenetic analysis of each gene or gene family present in *B. atropurpurea* was also shown in this chapter, confirming the relationship between the meiotic genes from *B. atropurpurea* and other Bangiales species whenever other Bangiales species data was available. Unfortunately, there are no other species known for being exclusively asexual within the Bangiales, that are close to *B. atropurpurea*, for comparison. To date, there are no studies exclusively about meiotic genes in red algae and in-depth functional analyses of these genes are restricted to model organisms' from the Viridiplantae. The advance of whole-genome sequencing and the addition of other Bangiales meiotic genes data, as well as data from the *Bangia*les sister group Florideophyecae, will help solve the evolution of these genes among the groups and elucidate more about the evolution of the Rhodophyta species. This study provides novel insights that extend our understanding of the molecular mechanisms in *B. atropurpurea*, being able to answer many questions regarding the nuclear gene content of the species and create intriguing new questions that are to be solved by future research.

Chapter 5

Research Summary and Future Considerations

5.1 Research Summary

The order Bangiales is considered the starting point of divergence for other algal orders, with the oldest recognized microfossil record dating 1.2 and 1.6 billion years ago that resemble the cellular structure of the modern red algal taxon *Bangia* (Butterfield, 2000; Müller et al., 2001; 2010; Yoon et al., 2006; Bengtson et al., 2017). Given the evolutionary importance of the Bangiales and *Bangia* species, phylogenetic studies, traditionally using marker genes such as *rbc*L, aimed to resolve the relationship among the Bangiales species, (Müller et al., 2001; Yoon et al., 2006; Sutherland et al., 2011). In this thesis, I described for the first time the chloroplast genome of *Bangia atropurpurea*, using environmental metagenomic sequence data. This methodology allowed me to obtain a chloroplast genome with 97% estimated coverage, and to analyze the gene content and organization in comparison to other Bangiales species. In addition, a whole chloroplast genome phylogenetic analysis was performed, validating what was previously postulated by Müller et al. (2005) and Sutherland et al. (2011) – that *B. atropurpurea* shares a common ancestor with *Porphyra* species, and is more closely related to these species than to *Bangia fuscopurpurea*. The chloroplast genome of *B. atropurpurea* was also used to examine the rates of substitution of chloroplast genes among Rhodophyta species. This detailed study showed that the Bangiales is the most distinct red algal group regarding rates of substitution and that *B. atropurpurea* is the most distinct species among the red algae. Furthermore, I was able to successfully recover nuclear meiotic genes from the metagenomic sequence data of *B. atropurpurea*. Given that the species is only known to be asexual in nature, these novel data open new avenues for a wide range of future studies. This complete thesis not only made use of new and non-conventional methodology but also built solid grounds for future research to discover more about the order Bangiales and the uniqueness of reproduction, adaptation and evolution in *B. atropurpurea*.

Chapter 2

Chapter 2 of this thesis bridges a considerable knowledge gap of the Bangiales order, with the sequence and description of *Bangia atropurpurea* chloroplast genome and provides a

detailed comparison of the first freshwater chloroplast genome of the order Bangiales. The order Bangiales is of significant evolutionary and ecological interest given their morphological similarities to ancient red algal fossils, and their recent invasion into the Great Lakes. The order consists of filamentous and foliose forms that are commonly observed in the mid-to the high intertidal zone of the marine environment. The single exception to this is the species *Bangia atropurpurea,* which is a recent (1964) introduction into the Laurentian Great Lakes and has subsequently spread to all of the lakes, except for Lake Superior. However, little is known about the ecological impact of this invader and genomic data is lacking. Here, I present the first chloroplast genome extracted from a mixed metagenomic sample isolated directly from the environment and outline a metagenomic workflow for assembling chloroplast genomes from difficult to grow organisms. The resulting chloroplast genome from *B. atropurpurea* has approximately 97% coverage and is 186,840 bp with 32.9% GC content, 38 tRNAs and 2 rRNAs. Comparison with other red algal chloroplast genomes reveals broad synteny and indicates that this species is more closely related to the genus *Porphyra* than to marine *Bangia* species.

The synteny analysis also shows that the genome organization and gene content of *B. atropurpurea* chloroplast genome is highly similar to other Bangiales species. The novel whole chloroplast sequence, produced in this research project, and its whole chloroplast phylogeny, corroborate with what has been shown by previous researchers, providing definitive evidence that *B. atropurpurea* is more closely related to *Porphyra purpurea* and *P. umbilicalis* than to any other species of red algae studied to date, including *B. fuscopurpurea*. The new chloroplast genome sequence of *B. atropurpurea* also allowed the analysis of the rates of substitution among chloroplast genes of Rhodophyta species observed in different environments. The comparison of 71 genes from 30 red algae species to *B. atropurpurea* shows that not only *B. atropurpurea* has unique rates of substitution within the Bangiales, but also how species that are observed in freshwater differs from the species observed in other habitats even when the species are closely related taxonomically. The work presented in this thesis represents an important foundation for comparison and taxonomic studies, and as new chloroplast genomes become available, we will be able to better understand the mutational patterns among red algae chloroplast genes evolution.

Chapter 3

Substitutions in the DNA sequence are the raw material for evolutionary processes and can be caused by genetic and/or environmental variations. The substitutions in the coding regions among species can be neutral, positive, and purifying. The analysis of the synonymous (K_s) and non-synonymous substitutions (K_a), and their ratios ($\omega = K_a / K_s$) can determine the average of selection pressures acting on a particular gene. Red algae species are an ancient lineage and one of the primary plastids-bearing hosts, also being the origin of plastids for the secondary and tertiary endosymbiosis. This thesis chapter identified the rates of substitutions for 71 non-concatenated chloroplast genes of 30 Rhodophyta species, pointing to different environmental selection pressures acting on the species genes. The analysis of non-concatenated gene sequences instead of conserved blocks reflected with more accuracy the substitution rates that occur in every single gene, not dismissing particularities that could be deleted in the conserved block formation. There is extensive literature in plant model organisms (*Arabidopsis thaliana*, maize, tobacco, rice) noting that mutations in chloroplast genes affect not only the plant phenotype but, in many cases, the gene variations can promote plant growth in adverse environmental conditions such as drought, high UV radiation and salinity. Future studies are needed to provide supportive evidence that similar selection pressures also influence *B. atropurpurea*, contributing to the survival of the species in freshwater.

The analysis of the rates of substitution among 6 algal orders (Bangiales, Florideophyceae, Porphyridiales, Compsopogonales, Cyanidiales and Bryopsidales) shows that the order Bangiales is the most distinct lineage among the red algae and exhibits patterns of molecular adaptations based on the species phylogenetic position. The differences in rates of substitution in the Bangiales species genes are significantly different from the other taxonomic groups for the *atp*, *psa*, *rps* and *rpl* gene families. In addition, the comparison of the substitution rates of genes of species observed in the marine environment, freshwater, hot springs and salt marshes, clearly shows that species occupying freshwater have contrasting rates. The freshwater species presented the highest or lowest rates of substitution for genes part of the *atp* and *rps* gene families, having a larger difference from the genes of species occupying the marine and hot springs environments. Thus, we can speculate that the shift in habitat, from marine to freshwater,

of *B. atropurpurea* likely required molecular adaptations, which may be reflected in the rates of substitution seen in the present study.

Chapter 4

As noted previously, the red algae have great evolutionary and ecological importance, yet little is known about the genetic pathways that lead to sex determination and the process of meiosis in these organisms. Typically, the members of the red algal order Bangiales (Bangiophyceae) have a heteromorphic life cycle, that alternates between a haploid gametophytic thallus phase (foliose in *Pyropia*, *Porphyra* or filamentous in *Bangia*) and a diploid sporophytic branched filamentous phase (conchocelis). In addition to the sexual life cycle, many species within Bangiales undergo an asexual life cycle by producing asexual spores, the archaeospores, which can later develop into the gametophytic thallus. However, there is little to no research on the molecular mechanisms by which red algae switch from sexual to asexual reproductive stages or the genes involved in either life history stage. The search for meiotic related genes is done using the eukaryotic meiotic toolbox, which has as reference model organisms such as *Arabidopsis thaliana*, and *Mus musculus* and includes genes that play a role during prophase I. Although the toolbox is a great starting point to detect meiotic genes, there is still a large gap in knowledge regarding specific red algae meiotic genes and proteins.

Bangia atropurpurea has only been observed to be asexual in nature and reported once to be induced into sexual reproduction in the laboratory (Gargiulo et al., 2001). This thesis chapter made use of the eukaryotic meiotic toolbox and metagenomic sequencing to investigate the presence of specific genes related to meiosis in *B. atropurpurea*, providing new information on the evolution of the meiotic genes in the Bangiales. From the 22 meiotic-exclusive genes known to be part of most eukaryotic meiosis machinery, 13 are present in *B. atropurpurea*. All these genes are part of the meiotic Prophase I, more specifically, part of stages 1, 3 and 4. The results obtained in this study are the grounds for future investigations, that can examine if sexual reproduction in *B. atropurpurea* is occurring rarely in this species and if the meiotic genes are functional or not. The metagenomic sequence approach has generated great results for the identification of meiotic genes, and this methodology is recommended for future studies,

especially for red algae species that are difficult to grow in the lab. The confirmation that *B. atropurpurea* has copies of several meiotic genes while being an asexual organism, and the fact that the phylogeny of the genes shows a close relationship to species that are sexual (*Porphyra umbilicalis*), raises questions about how the change in habitat could have impact *Bangia* reproduction. Assuredly, this thesis chapter extends our understanding of the molecular mechanisms in *B. atropurpurea*, nuclear gene content and phylogenetic relationships of this organism, being the groundwork for future research.

5.2 Future Considerations

I) Fill the gaps of the chloroplast genome of B. atropurpurea

The sequencing and assembly of the metagenome from an environmental sample was an effective strategy to reconstruct 97% of the chloroplast genome of *B. atropurpurea*. Further sequencing and assembly using long-reads technologies as well as improved computational methods for metagenome assembly may help to spam the gaps in the existing assembly and potentially identify genes missing in the current version. The chloroplast genes that are not present in *B. atropurpurea* and also absent in other Bangiales species (with available data to date) include *ycf*27, *bas*1, and *psb*Y. The *pet*N gene must also be examined further since the gene is present in other Bangiales species but was not observed in *B. atropurpurea*. Future sequencing and analysis will help to verify whether these genes are truly absent due to gene loss or whether they are present in *B. atropurpurea* and simply lacked sequencing coverage in this study.

II) Nuclear and mitochondrial genome sequencing

Given the unique characteristics of *B. atropurpurea*, including the transition from a marine environment to freshwater, the filamentous thallus closely related to sheet-like thallus species, and its asexual reproduction while having sexual genes, whole genome sequencing might be able to uncover more details about the organism that are still unknown. A complete nuclear and mitochondrial sequence could reveal specific gene gains/losses, substitutions, or genomic rearrangements that coincide with the changes in morphology. Considering the difficulty to grow *B. atropurpurea* in lab due to contamination by other organisms and the slow

growth rate, new metagenomic sequencing is a valid alternative for future studies to obtain good coverage for the mitochondrial and nuclear genomes.

III) Codon usage bias analysis for chloroplast genes

The rates of substitution analysis of the chloroplast genes of Rhodophyta species showed how species within the same taxonomic groups are under different selection pressures even for genes that are considered critical for the chloroplast machinery. Selection forces may also influence the codon usage of an organism and the analysis of the codon usage bias (CUB) can provide further explanation on the species evolution, gene function, gene expression and lateral gene transfer (Bodilis & Barray, 2006; Wang et al., 2018).

IV) Substitution saturation

Given the results obtained on rates of substitution of *B. atropurpurea* chloroplast genes, the data produced on chapter 3 also allows future investigations to verify if the chloroplast genes are under substitution saturation and perhaps answer why the species is the most distinct among the Bangiales when comparing rates of substitution. Substitution saturation decreases phylogenetic information contained in the sequences (Xie et al., 2003), resulting in conflicting rates. As synonymous substitutions seem to saturate faster than could be handled by nucleotide models in mitochondrial DNA, resulting in purifying selection (Suchard and Rambaut, 2009), similar pattern may be observed in some of *B. atropurpurea* chloroplast sequences. Substitution saturation might also explain why *B. atropurpurea* has long branches in conserved gene families such as *atp* and *psb*. In addition, the results shown for the *rbc*L gene (evidence for positive selection) also suggest sequence saturation and must be investigated further, specially given the common use of the gene sequence in phylogenetic analysis in Rhodophyta.

IV) Rates of evolution and codon usage bias for meiotic genes

With new red algae genomes becoming available, the study of the meiotic genes can be done in more depth. The nucleotide sequences of meiotic genes can be used for rates of substitution analysis, which can provide interesting results about how different selection pressures act on the meiotic genes of each species. The Codon Usage Bias (CUB) analysis can be performed to recognize protein-coding genes and pseudo-genes, where open reading frames with

high incidence of rare codons are unlikely to encode a protein (Ghaemmaghami et al., 2003; Roth et al., 2012). The data described in Chapter 4 can be used for CUB studies and used to compare with the CUB of meiotic genes of other *Bangia* species, increasing the chances to understand why *B. atropurpurea* has meiotic genes, and if they are functional.

References

Abdullah, M. F., Hoffmann, E. R., Cotton, V. E., & Borts, R. H. (2004). A role for the *mut*L homolog *mlh*2 in controlling heteroduplex formation and in regulating between two different crossover pathways in budding yeast. *Cytogenetic Genome Research*,107, 180-190.

Adams, J., Mansfield, M. J., Richard, D. J., & Doxey, A. C. (2017). Lineage-specific mutational clustering in protein structures predicts evolutionary shifts in function. *Bioinformatics*, *33*(9), 1338-1345.

Adl, S. M., Simpson, A. G., Farmer, M. A., Andersen, R. A., Anderson, O. R., Barta, J. R., ... & Taylor, M. F. (2005). The new higher-level classification of eukaryotes with emphasis on the taxonomy of protists. *Journal of Eukaryotic Microbiology*, *52*(5), 399-451.

Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., ... & Spiegel, F. W. (2012). The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, *59*(5), 429-514.

Agardh, C. A. (1824). *Systema Algarum*. literis Berlingianis.

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at silent sites in Drosophila DNA. Genetics, 139(2), 1067-1076.

Akhtar, W., Rasheed, A., Shinwari, Z. K., Naqvi, S. M. S., & Mahmood, T. (2014). Genetic characterization of different Pakistani date palm varieties. *Pak J Bot*, *46*(6), 2095-100.

Alberts, B., Bray, D., Hopkin, K., Johnson, A. D., Lewis, J., Raff, M., ... & Walter, P. (2015). *Essential Cell Biology*. Garland Science.

Allen, J. F. (2004). Cytochrome b6f: structure for signalling and vectorial metabolism. *Trends in Plant Science*, *9*(3), 130-137.

Al-Sweel, N., Raghavan, V., Dutta, A., Ajith, V. P., Di Vietro, L., Khondakar, N., ... & Alani, E. (2017). mlh3 mutations in baker's yeast alter meiotic recombination outcomes by increasing noncrossover events genome-wide. *PLoS Genetics*, *13*(8), e1006974.

Antoshvili, M., Caspy, I., Hippler, M., & Nelson, N. (2019). Structure and function of photosystem I in *Cyanidioschyzon merolae*. *Photosynthesis research*, *139*(1), 499-508.

Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Current Biology*, *25*(19), R911-R921.

Archibald, J. M., & Keeling, P. J. (2002). Recycled plastids: a 'green movement' in eukaryotic evolution. *Trends in Genetics*, *18*(11), 577-584.

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502-508.

Arnon, D. I., ALLEN, M. B., & Whatley, F. R. (1954). Photosynthesis by isolated chloroplasts. *Nature*, *174*(4426), 394-396.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., ... & Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*(W1), W597-W603.

Asamizu, E., Nakajima, M., Kitade, Y., Saga, N., Nakamura, Y., & Tabata, S. (2003). Comparison of RNA expression profiles between the two generations of *Porphyra yezoensis* (Rhodophyta), based on expressed sequence tag frequency analysis. *Journal of Phycology*, *39*(5), 923-930.

Ashby, M. K., & Mullineaux, C. W. (1999). Cyanobacterial *ycf*27 gene products regulate energy transfer from phycobilisomes to photosystems I and II. *FEMS Microbiology Letters*, *181*(2), 253-260.

Ashby, M. K., Houmard, J., & Mullineaux, C. W. (2002). The *ycf*27 genes from cyanobacteria and eukaryotic algae: distribution and implications for chloroplast evolution. *FEMS Microbiology Letters*, *214*(1), 25-30.

Backhausen, J. E., Kitzmann, C., Horton, P., & Scheibe, R. (2000). Electron acceptors in isolated intact spinach chloroplasts act hierarchically to prevent over-reduction and competition for electrons. *Photosynthesis Research*, *64*(1), 1-13.

Baier, M., and Dietz, K.-J. (1999). Protective function of chloroplast 2-cysteine peroxiredoxin in photosynthesis: Evidence from transgenic *Arabidopsis*. *Plant Physiol*. 119, 1407–1414

Baier, M., Noctor, G., Foyer, C., and Dietz, K.-J. (2000). Antisense suppression of 2-cysteine peroxiredoxin in *Arabidopsis* specifically enhances the activities and expression of enzymes

associated with ascorbate metabolism but not glutathione metabolism. *Plant Physiol*. 124, 823–832.

Baker, N. R. (1991). A possible role for photosystem II in environmental perturbations of photosynthesis. *Physiologia Plantarum*, *81*(4), 563-570.

Ball, A. R., & Yokomori, K. (2001). The structural maintenance of chromosomes (SMC) family of proteins in mammals. *Chromosome Research*, *9*(2), 85-96.

Baloglu, M. C., Ulu, F., Altunoglu, Y. C., Pekol, S., Alagoz, G., & Ese, O. (2015). Identification, molecular characterization and expression analysis of *rpl*24 genes in three Cucurbitaceae family members: cucumber, melon and watermelon. *Biotechnology & Biotechnological Equipment*, *29*(6), 1024-1034.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D. and Pyshkin, A.V. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology, 19*(5), 455-477.

Barghoorn, E. S., & Tyler, S. A. (1965). Microorganisms from the Gunflint chert. *Science*, *147*(3658), 563-577.

Baweja, P., Sahoo, D., García-Jiménez, P., & Robaina, R. R. (2009). Seaweed tissue culture as applied to biotechnology: Problems, achievements and prospects. *Phycological Research, 57*(1), 45-58.

Bedbrook, J. R., & Kolodner, R. (1979). The structure of chloroplast DNA. *Annual Review of Plant Physiology*, *30*(1), 593-620.

Bellot, S., & Renner, S. S. (2016). The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution*, *8*(1), 189-201.

Bengtson, S., Sallstedt, T., Belivanova, V., & Whitehouse, M. (2017). Three-dimensional preservation of cellular and subcellular structures suggests 1.6-billion-year-old crown-group red algae. *PLoS Biology, 15*(3), e2000735.

Benyahya, F., Nadaud, I., Da Ines, O., Rimbert, H., White, C., & Sourdille, P. (2020). *spo*11. 2 is essential for programmed double strand break formation during meiosis in bread wheat (*Triticum aestivum* L.). *Plant Journal*, *104*(1), 30-43.

Bhattacharya, D., & Medlin, L. (1995). The phylogeny of plastids: A review based on comparisons of small-subunit ribosomal RNA coding regions. *Journal of Phycology*, *31*(4), 489-498.

Bhattacharya, D., Price, D. C., Chan, C. X., Qiu, H., Rose, N., Ball, S., ... & Yoon, H. S. (2013). Genome of the red alga *Porphyridium purpureum*. *Nature Communications*, *4*(1), 1-10.

Bhattacharya, D., Price, D.C., Chan, C.X., Qiu, H., Rose, N., Ball, S., Weber, A.P., Arias, M.C., Henrissat, B., Coutinho, P.M. and Krishnan, A. (2013). Genome of the red alga *Porphyridium purpureum*. *Nature Communications*, *4*, 1941.

Bhattacharya, D., Yoon, H. S., & Hackett, J. D. (2004). Photosynthetic eukaryotes unite: Endosymbiosis connects the dots. *Bioessays*, *26*(1), 50-60.

Bidoux, C., & Magne, F. (1989). Étude de quelques Acrochaetiales (Rhodophyta) devant être rapportées au genre Rhodothamniella. *Cryptogamie. Algologie*, *10*(1), 33-55.

Bird, C. J. (1973). Aspects of the life history and ecology of *Porphyra linearis* (Bangiales, Rhodophyceae) in nature. *Canadian Journal of Botany*, *51*(12), 2371-2379.

Bird, C. J. (1992). Seaweed flora of the Maritimes 1. Rhodophyta-Red algae. *Bipotess, Bristol*.

Bird, C. J., Murphy, C. A., Rice, E. L., & Ragan, M. A. (1992). The 18S rRNA gene sequences of four commercially important seaweeds. *Journal of Applied Phycology*, *4*(4), 379-384.

Bischoff-Bäsmann, B., & Wiencke, C. (1996). Temperature requirements for growth and survival of Antarctica Rhodophyta 1. *Journal of Phycology*, *32*(4), 525-535.

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics*, *22*(8), 437-446.

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics*, *22*(8), 437-446.

Biswas, S., & Eaton-Rye, J. J. (2018). PsbY is required for prevention of photodamage to photosystem II in a PsbM-lacking mutant of Synechocystis sp. PCC 6803. *Photosynthetica*, *56*(1), 200-209.

Blackwell, A. R. (2020). *The Role of Arabidopsis thaliana Mismatch Repair Proteins in Meiotic Recombination* (Doctoral dissertation, University of Cambridge).

Blackwell, A. R., Dluzewska, J., Szymanska-Lejman, M., Desjardins, S., Tock, A. J., Kbiri, N., ... & Henderson, I. R. (2020). *msh*2 shapes the meiotic crossover landscape in relation to interhomolog polymorphism in *Arabidopsis*. *The EMBO Journal*, *39*(21), e104858.

Blouin, N. A., & Brawley, S. H. (2012). An AFLP-based test of clonality in widespread, putatively asexual populations of *Porphyra umbilicalis* (Rhodophyta) in the Northwest Atlantic with an in silico analysis for bacterial contamination. *Marine Biology*, *159*(12), 2723-2729.

Blouin, N. A., Brodie, J. A., Grossman, A. C., Xu, P., & Brawley, S. H. (2011). *Porphyra*: a marine crop shaped by stress. *Trends in plant science*, *16*(1), 29-37.

Blouin, N. A., Brodie, J. A., Grossman, A. C., Xu, P., & Brawley, S. H. (2011). *Porphyra*: A marine crop shaped by stress. *Trends in Plant Science*, *16*(1), 29-37.

Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgő, B., ... & Pál, C. (2017). Phenotypic heterogeneity promotes adaptive evolution. *PLoS Biology*, *15*(5), e2000644.

Bodilis, J., & Barray, S. (2006). Molecular evolution of the major outer-membrane protein gene (*opr*F) of Pseudomonas. *Microbiology*, *152*(4), 1075-1088.

Bolaños-Villegas, P. (2021). The Role of Structural Maintenance of Chromosomes Complexes in Meiosis and Genome Maintenance: Translating Biomedical and Model Plant Research Into Crop Breeding Opportunities. *Frontiers in Plant Science*, *12*, 563.

Bolle, C., Herrmann, R. G., & Oelmüller, R. (1996). Intron sequences are involved in the plastid-and light-dependent expression of the spinach psaD gene. The Plant Journal, 10(5), 919-924.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.F., Lió, P., Crescenzi, P., Fani, R. and Fondi, M. (2015). MeDuSa: A multi-draft based scaffolder. *Bioinformatics*, *31*(15), 2443-2451.

Brawley, S.H., Blouin, N.A., Ficko-Blean, E., Wheeler, G.L., Lohr, M., Goodson, H.V., Jenkins, J.W., Blaby-Haas, C.E., Helliwell, K.E., Chan, C.X. and Marriage, T.N. (2017). Insights into the red algae and eukaryotic evolution from the genome of *Porphyra umbilicalis*

(Bangiophyceae, Rhodophyta). *Proceedings of the National Academy of Sciences of the United States of America*, *114*(31), E6361-E6370.

Brodie J, Hayes PK, Barker GL, Irvine LM, Bartsch I (1998). A reappraisal of *Porphyra* and *Bangia* (Bangiophycidae, Rhodophyta) in the Northeast Atlantic based on the *rbc*L-*rbc*S intergenic spacer. *J. Phycol*. 34: 1069–1074.

Brodie, J. A., & Irvine, L. M. (2003). Seaweeds of the British Isles, Vol. I: Bangiophycidae. Part 3b. *British Museum, Natural History, London*.

Brodie, J., Hayes, P. K., Barker, G. L., & Irvine, L. M. (1996). Molecular and morphological characters distinguishing two *Porphyra* species (Rhodophyta: Bangiophycidae). *European Journal of Phycology*, *31*(4), 303-308.

Brodie, J., Mortensen, A. M., Ramirez, M. E., Russell, S., & Rinkel, B. (2008). Making the links: towards a global taxonomy for the red algal genus *Porphyra* (Bangiales, Rhodophyta). In *Nineteenth International Seaweed Symposium* (pp. 489-499). Springer, Dordrecht.

Broin, M., Cuiné, S., Eymery, F., & Rey, P. (2002). The plastidic 2-cysteine peroxiredoxin is a target for a thioredoxin involved in the protection of the photosynthetic apparatus against oxidative damage. *The Plant Cell*, *14*(6), 1417-1432.

Broom, J. E., Jones, W. A., Hill, D. F., Knight, G. A., & Nelson, W. A. (1999). Species recognition in New Zealand *Porphyra* using 18S rDNA sequencing. *Journal of Applied Phycology*, *11*(5), 421-428.

Broom, J. E., Nelson, W. A., Yarish, C. A. J. W., Jones, W. A., Rosas, R. A., & Rosas, L. A. (2002). A reassessment of the taxonomic status of *Porphyra suborbiculata*, *Porphyra carolinensis* and *Porphyra lilliputiana* (Bangiales, Rhodophyta) based on molecular and morphological data. *European Journal of Phycology*, *37*(2), 227-235.

Broom, J., Farr, T., & Nelson, W. (2004). Phylogeny of the *Bangia* flora of New Zealand suggests a southern origin for *Porphyra* and *Bangia* (Bangiales, Rhodophyta). *Molecular Phylogenetics and Evolution, 31*(3), 1197-1207.

Broom, J., Nelson, W., Yarish, C., Jones, W., Rosas, R. A., & Rosas, L. A. (2002). A reassessment of the taxonomic status of *Porphyra suborbiculata*, *Porphyra carolinensis*

and *Porphyra lilliputiana* (Bangiales, Rhodophyta) based on molecular and morphological data. *European Journal of Phycology, 37*(2), 227-235.

Brunet, T. D., Doolittle, W. F., & Bielawski, J. P. (2021). The role of purifying selection in the origin and maintenance of complex function. *Studies in History and Philosophy of Science Part A*, *87*, 125-135.

Buckling, A., Wei, Y., Massey, R. C., Brockhurst, M. A., & Hochberg, M. E. (2006). Antagonistic coevolution with parasites increases the cost of host deleterious mutations. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1582), 45-49.

Burki, F. (2014). The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harbor Perspectives in Biology*, *6*(5), a016147.

Burzycki GM, Waaland, JR (1987) On the position of meiosis in the life history of *Porphyra torta* (Rhodophyta). *Botanica marina* 30: 5–10.

Bush, R. M. (2001). Predicting adaptive evolution. *Nature Reviews Genetics*, *2*(5), 387-392.

Butterfield, N. J. (2000). *Bangiomorpha pubescens* n. gen., n. sp.: Implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology, 26*(3), 386-404.

Butterfield, N. J. (2005). Probable Proterozoic fungi. *Paleobiology*, *31*(1), 165-182.

Butterfield, N. J., Knoll, A. H., & Swett, K. (1990). A bangiophyte red alga from the Proterozoic of arctic Canada. *Science*, *250*(4977), 104-108.

Butterfield, N. J., Knoll, A. H., & Swett, K. (1990). A bangiophyte red alga from the Proterozoic of Arctic Canada. *Science*, *250*(4977), 104-107.

Cameron, K. M. (2004). Utility of plastid *psa*B gene sequences for investigating intrafamilial relationships within Orchidaceae. *Molecular Phylogenetics and Evolution*, *31*(3), 1157-1180.

Campbell, C. S., Hombauer, H., Srivatsan, A., Bowen, N., Gries, K., Desai, A., ... & Kolodner, R. D. (2014). *mlh*2 is an accessory factor for DNA mismatch repair *in Saccharomyces cerevisiae*. *PLoS genetics*, *10*(5), e1004327.

Campbell, S. E. (1980). *Palaeoconchocelis starmachii*, a carbonate boring microfossil from the Upper Silurian of Poland (425 million years old): implications for the evolution of the Bangiaceae (Rhodophyta). *Phycologia*, *19*(1), 25-36.

Cannavo, E., Sanchez, A., Anand, R., Ranjha, L., Hugener, J., Adam, C., ... & Cejka, P. (2020). Regulation of the *mlh*1–*mlh*3 endonuclease in meiosis. *Nature*, *586*(7830), 618-622.

Cao, M., Bi, G., Mao, Y., Li, G., & Kong, F. (2018). The first plastid genome of a filamentous taxon '*Bangia'*sp. OUCPT-01 in the Bangiales. *Scientific Reports, 8*(1), 10688.

Carramolino, L., Lee, B. C., Zaballos, A., Peled, A., Barthelemy, I., Shav-Tal, Y., ... & Zipori, D. (1997). SA-1, a nuclear protein encoded by one member of a novel gene family: molecular cloning and detection in hemopoietic organs. *Gene*, *195*(2), 151-159.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540-552.

Cavalier-Smith, T. (2002). Origins of the machinery of recombination and sex. *Heredity*, *88*(2), 125-141.

Cavalier-Smith, T. O. M. (1999). Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree 1, 2. *Journal of Eukaryotic Microbiology*, *46*(4), 347-366.

Cervantes, M. D., Farah, J. A., & Smith, G. R. (2000). Meiotic DNA breaks associated with recombination in *S. pombe*. *Molecular cell*, *5*(5), 883-888.

Chan, C. X., Blouin, N. A., Zhuang, Y., Zäuner, S., Prochnik, S. E., Lindquist, E., ... & Bhattacharya, D. (2012). *Porphyra* (Bangiophyceae) transcriptomes provide insights into red algal development and metabolism. *Journal of Phycology*, *48*(6), 1328-1342.

Chan, C. X., Gross, J., Yoon, H. S., & Bhattacharya, D. (2011). Plastid origin and evolution: new models provide insights into old problems. *Plant physiology*, *155*(4), 1552-1560.

Chandra Sanyal S, Liljas A (2000) The end of the beginning: structural studies of ribosomal proteins. *Curr Opin Struct Biol* 10:633–636

Chaotian, X., Changsheng, C., Yan, X., & Dehua, J. (2010). Construction of a genetic linkage map for *Porphyra haitanensis* (*Bangia*les, Rhodophyta) based on sequence-related

amplified polymorphism and simple sequence repeat markers 1. *Journal of Phycology*, *46*(4), 780-787.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., ... & Albert, V. A. (1993). Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden*, 528-580.

Chelysheva, L., Diallo, S., Vezon, D., Gendrot, G., Vrielynck, N., Belcram, K., ... & Grelon, M. (2005). AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis. *Journal of Cell Science*, *118*(20), 4621-4632.

Chen, C. C., Hyvönen, J., & Schneider, H. (2020). Exploring phylogeny of the microsoroid ferns (Polypodiaceae) based on six plastid DNA markers. *Molecular Phylogenetics and Evolution*, *143*, 106665.

Chen, Y. K., Leng, C. H., Olivares, H., Lee, M. H., Chang, Y. C., Kung, W. M., ... & Wang, T. F. (2004). Heterodimeric complexes of *hop*2 and Mnd1 function with *dmc*1 to promote meiotic homolog juxtaposition and strand assimilation. *Proceedings of the National Academy of Sciences*, *101*(29), 10572-10577.

Chi, N. W., & Kolodner, R. D. (1994). Purification and characterization of *msh*1, a yeast mitochondrial protein that binds to DNA mismatches. *Journal of Biological Chemistry*, *269*(47), 29984-29992.

Chi, P., San Filippo, J., Sehorn, M. G., Petukhova, G. V., & Sung, P. (2007). Bipartite stimulatory action of the *hop*2–*mnd*1 complex on the *rad*51 recombinase. *Genes & Development*, *21*(14), 1747-1757.

Chiang, T. Y., Schaal, B. A., & Peng, C. I. (1998). Universal primers for amplification and sequencing a noncoding spacer between the *atp*B and *rbc*L genes of chloroplast DNA. *Botanical Bulletin of Academia Sinica*, *39*.

Chitnis, P. R. (1996). Photosystem I. *Plant physiology*, *111*(3), 661.

Chitnis, V. P., & Chitnis, P. R. (1993). *psa*L subunit is required for the formation of photosystem I trimers in the cyanobacterium *Synechocystis* sp. PCC 6803. FEBS letters, 336(2), 330-334.

Chitnis, V. P., Xu, Qiang G., Yu, LianN., Golbeck, J. H., Nakamoto, H., Xie, D. L., & Chitnis, P. R. (1993). Targeted inactivation of the gene *psa*L encoding a subunit of photosystem I of the cyanobacterium *Synechocystis* sp. PCC 6803. *Journal of Biological Chemistry*, *268*(16), 11678-11684.

Cho, C. H., Choi, J. W., Lam, D. W., Kim, K. M., & Yoon, H. S. (2018). Plastid genome analysis of three Nemaliophycidae red algal species suggests environmental adaptation for iron limited habitats. *Plos One*, *13*(5), e0196995.

Cho, C. H., Park, S. I., Ciniglia, C., Yang, E. C., Graf, L., Bhattacharya, D., & Yoon, H. S. (2020). Potential causes and consequences of rapid mitochondrial genome evolution in thermoacidophilic *Galdieria* (Rhodophyta). *BMC Evolutionary Biology*, *20*(1), 1-15.

Choi, S. J., Kim, Y., & Choi, C. (2019). Chloroplast genome-based hypervariable markers for rapid authentication of six Korean *Pyropia* species. *Diversity*, *11*(12), 220.

Choi, S. J., Park, E. J., Endo, H., Kitade, Y., & Saga, N. (2008). Inheritance pattern of chloroplast and mitochondrial genomes in artificial hybrids of *Porphyra yezoensis* (Rhodophyta). *Fisheries Science*, *74*(4), 822-829.

Chopin, T., Sharp, G., Belyea, E., Semple, R., & Jones, D. (1999). Open-water aquaculture of the red alga *Chondrus crispus* in Prince Edward Island, Canada. In *Sixteenth International Seaweed Symposium* (pp. 417-425). Springer, Dordrecht.

Ciniglia, C., Yang, E. C., Pollio, A., Pinto, G., Iovinella, M., Vitale, L. & Su, H. 2014. Cyanidiophyceae in Iceland: Plastid rbcL gene elucidates origin and dispersal of extremophilic Galdieria sulphuraria and G. maxima (Galdieriaceae, Rhodophyta). Phycologia, 53:542–551.

Clegg, M. T., & Zurawski, G. (1992). Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In *Molecular systematics of plants* (pp. 1-13). Springer, Boston, MA.

Cloud, V., Chan, Y. L., Grubb, J., Budke, B., & Bishop, D. K. (2012). *rad*51 is an accessory factor for *dmc*1-mediated joint molecule formation during meiosis. *Science*, *337*(6099), 1222-1225.

Cobbe, N., & Heck, M. M. (2004). The evolution of SMC proteins: phylogenetic analysis and structural implications. *Molecular biology and evolution*, *21*(2), 332-347.

Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G., ... & Wincker, P. (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature*, *465*(7298), 617-621.

Coelho, S. M., Peters, A. F., Charrier, B., Roze, D., Destombe, C., Valero, M., & Cock, J. M. (2007). Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. *Gene*, *406*(1-2), 152-170.

Coleman, A. W., & Goff, L. J. (1991). DNA analysis of Eukaryotic algal species 1. *Journal of Phycology*, *27*(4), 463-473.

Collén, J., Porcel, B., Carré, W., Ball, S. G., Chaparro, C., Tonon, T., ... & Boyen, C. (2013). Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences*, *110*(13), 5247-5252.

Collier, J. E., Lee, B. G., Roig, M. B., Yatskevich, S., Petela, N. J., Metson, J., ... & Nasmyth, K. A. (2020). Transport of DNA within cohesin involves clamping on top of engaged heads by *scc*2 and entrapment within the ring by *scc*3. *Elife*, *9*, e59560.

Comeron, J. M. (2014). Background selection as baseline for nucleotide variation across the *Drosophila* genome. *PLoS Genetics*, *10*(6), e1004434.

Conde, F. R., Churio, M. S., & Previtali, C. M. (2000). The photoprotector mechanism of mycosporine-like amino acids. Excited-state properties and photostability of *Porphyra*-334 in aqueous solution. *Journal of Photochemistry and Photobiology B: Biology*, *56*(2-3), 139-144.

Costa, J. F., Lin, S. M., Macaya, E. C., Fernández-García, C., & Verbruggen, H. (2016). Chloroplast genomes as a tool to resolve red algal phylogenies: a case study in the Nemaliales. BMC Evolutionary Biology, 16(1), 1-13.

Critchlow, H. M., Payne, A., & Griffin, D. K. (2004). Genes and proteins involved in the control of meiosis. *Cytogenetic and Genome Research*, *105*(1), 4-10.

Csuros, M., Rogozin, I. B., & Koonin, E. V. (2011). A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. *PLoS computational biology*, *7*(9), e1002150.

Cvijović, I., Good, B. H., & Desai, M. M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics*, *209*(4), 1235-1278.

Da Ines, O., Degroote, F., Goubely, C., Amiard, S., Gallego, M. E., & White, C. I. (2013). Meiotic recombination in Arabidopsis is catalysed by *dmc*1, with *rad*51 playing a supporting role. *PLoS Genetics*, *9*(9), e1003787.

DalCorso, G., Pesaresi, P., Masiero, S., Aseeva, E., Schünemann, D., Finazzi, G., ... & Leister, D. (2008). A complex containing PGRL1 and PGR5 is involved in the switch between linear and cyclic electron flow in *Arabidopsis*. *Cell*, *132*(2), 273-285.

Damann, K. E., & KE, D. (1979). Occurrence of the red alga *Bangia atropurpurea* in Lake Ontario.

Darling, A. E., Mau, B., & Perna, N. T. (2010). ProgressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PloS One, 5*(6), e11147.

De la Fuente, J., Van Den Bussche, R. A., Prado, T. M., & Kocan, K. M. (2003). Anaplasma marginale msp1 genotypes evolved under positive selection pressure but are not markers for geographic isolates. J. Clin. Microbiol, 41(4), 1609-1616.

De Las Rivas, J., Lozano, J. J., & Ortiz, A. R. (2002). Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Rresearch*, *12*(4), 567-583.

de Vries, J., & Archibald, J. M. (2017). Endosymbiosis: did plastids evolve from a freshwater cyanobacterium. *Current Biology*, *27*(3), R103-R105.

Debus RJ (1992) The manganese and calcium ions of photosynthetic oxygen evolution. *Biochim Biophys Acta* 1102:269–352

Delport, W., Scheffler, K., & Seoighe, C. (2009). Models of coding sequence evolution. Briefings in Bioinformatics, 10(1), 97-109.

Delwiche, C. F., & Palmer, J. D. (1997). The origin of plastids and their spread via secondary symbiosis. *Origins of algae and their plastids*, 53-86.

den Hartog, C. (1972). Substratum: Multicellular plants. *Marine Ecology, 1*, 1277-1289.

DePamphilis, C. W., Young, N. D., & Wolfe, A. D. (1997). Evolution of plastid gene *rps*2 in a lineage of hemiparasitic and holoparasitic plants: many losses of photosynthesis and complex patterns of rate variation. *Proceedings of the National Academy of Sciences*, *94*(14), 7367-7372.

DePriest, M. S., Bhattacharya, D., & López-Bautista, J. M. (2013). The plastid genome of the red macroalga *Grateloupia taiwanensis* (Halymeniaceae). *PLoS One, 8*(7), e68246.

Dmitrienko, A., & D'Agostino Sr, R. (2013). Traditional multiplicity adjustment methods in clinical trials. *Statistics in Medicine*, *32*(29), 5172-5218.

Domenichini, S., Raynaud, C., Ni, D. A., Henry, Y., & Bergounioux, C. (2006). Atmnd1-Δ1 is sensitive to gamma-irradiation and defective in meiotic DNA repair. *DNA Repair*, *5*(4), 455-464.

Dong, W., Xu, C., Cheng, T., & Zhou, S. (2013). Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. *PloS One*, *8*(10), e77965.

Donze, D., Adams, C. R., Rine, J., & Kamakaka, R. T. (1999). The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. *Genes & Development*, *13*(6), 698-708.

Doron-Faigenboim, A., & Pupko, T. (2007). A combined empirical and mechanistic codon model. Molecular Biology and Evolution, 24(2), 388-397.

Dorsett, D. (2007). Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. *Chromosoma*, *116*(1), 1-13.

Douglas, S. E., & Penny, S. L. (1999). The plastid genome of the Cryptophyte alga, *Guillardia theta*: Complete sequence and conserved synteny groups confirm its common ancestry with red algae. *Journal of Molecular Evolution, 48*(2), 236-244.

Douglas, S., Zauner, S., Fraunholz, M., Beaton, M., Penny, S., Deng, L. T., ... & Maier, U. G. (2001). The highly reduced genome of an enslaved algal nucleus. *Nature*, *410*(6832), 1091-1096.

Drager, R. G., & Hallick, R. B. (1993). A novel *Euglena gracilis* chloroplast operon encoding four ATP synthase subunits and two ribosomal proteins contains 17 introns. *Current Genetics*, 23(3), 271-280.

Drapier, D., Suzuki, H., Levy, H., Rimbault, B., Kindle, K. L., Stern, D. B., & Wollman, F. A. (1998). The chloroplast atpA gene cluster in *Chlamydomonas reinhardtii*: Functional analysis of a polycistronic transcription unit. *Plant Physiology*, 117(2), 629-641.

Drew, K. M. (1949). Conchocelis-phase in the life-history of *Porphyra umbilicalis* (L.) Kütz. *Nature*, *164*(4174), 748-749.

Drew, K. M. (1956). Reproduction in the Bangiophyceae. *The Botanical Review*, 22, 553–611.

Du, Q., Bi, G., Mao, Y., & Sui, Z. (2016). The complete chloroplast genome of Gracilariopsis lemaneiformis (Rhodophyta) gives new insight into the evolution of family Gracilariaceae. Journal of Phycology, 52(3), 441-450.

Duplessis, M. R., Karol, K. G., Adman, E. T., Choi, L. Y., Jacobs, M. A., & Cattolico, R. A. (2007). Chloroplast His-to-Asp signal transduction: a potential mechanism for plastid gene regulation in *Heterosigma akashiwo* (Raphidophyceae). *BMC Evolutionary Biology*, *7*(1), 70.

Duret, L. (2008). Neutral theory: the null hypothesis of molecular evolution. *Nature Education*, *1*, 803-806.

Duroc, Y., Kumar, R., Ranjha, L., Adam, C., Guérois, R., Md Muntaz, K., MarsolierKergoat, M. C., Dingli, F., Laureau, R., Loew, D., Llorente, B., Charbonnier, J. B., Cejka, P., &Borde, V. (2017). Concerted action of the *mut*Lβ heterodimer and *mer*3 helicase regulates the global extent of meiotic gene conversion. *eLife*, 6, e21900.

Egan, S., Harder, T., Burke, C., Steinberg, P., Kjelleberg, S., & Thomas, T. (2013). The seaweed holobiont: Understanding seaweed–bacteria interactions. *FEMS Microbiology Reviews, 37*(3), 462-476.

Eisen, J. A. (1998). A phylogenomic study of the *mut*S family of proteins. *Nucleic acids research*, *26*(18), 4291-4300.

Elyashiv, E., Sattath, S., Hu, T. T., Strutsovsky, A., McVicker, G., Andolfatto, P., ... & Sella, G. (2016). A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genetics*, *12*(8), e1006130.

Endo, H., Sugie, T., Yonemori, Y., Nishikido, Y., Moriyama, H., Ito, R., & Okunishi, S. (2021). Vegetative Reproduction Is More Advantageous Than Sexual Reproduction in a Canopy-Forming Clonal Macroalga under Ocean Warming Accompanied by Oligotrophication and Intensive Herbivory. *Plants*, *10*(8), 1522.

Enomoto, R., Kinebuchi, T., Sato, M., Yagi, H., Shibata, T., Kurumizaka, H., & Yokoyama, S. (2004). Positive role of the mammalian TBPIP/HOP2 protein in *dmc*1-mediated homologous pairing. *Journal of Biological Chemistry*, *279*(34), 35263-35272.

Esposito, M. S., & Esposito, R. E. (1969). The genetic control of sporulation in *Saccharomyces* I. The isolation of temperature-sensitive sporulation-deficient mutants. *Genetics*, *61*(1), 79.

Fangru, N., Yuxin, H., Xudong, L., Jia, F., Junping, L., Qi, L., & Shulian, X. (2020). Analysis of Adaptive Evolution and Coevolution of *rbc*L Gene in the Genus *Hildenbrandia* (Rhodophyta). *Evolutionary Bioinformatics*, *16*, 1176934320977862.

Farah, J., Rappaport, F., Choquet, Y., Joliot, P., & Rochaix, J. D. (1995). Isolation of a *psa*F-deficient mutant of *Chlamydomonas reinhardtii*: efficient interaction of plastocyanin with the photosystem I reaction center is mediated by the *psa*F subunit. *The EMBO Journal*, *14*(20), 4976-4984.

Farahani-Tafreshi, Y., Wei, C., Gan, P., Daradur, J., Riggs, C. D., & Hasenkampf, C. A. (2021). The *Arabidopsis hop*2 Gene Has a Role in Preventing illegitimate Exchanges between Nonhomologous Chromosomes. *bioRxiv*.

Feller, U., Anders, I., & Mae, T. (2008). Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. *Journal of Experimental Botany*, *59*(7), 1615-1624.

Felsenstein, J., & Felenstein, J. (2004). Inferring phylogenies (Vol. 2, p. 664). Sunderland, MA: Sinauer associates.

Figueroa, F. L., Aguilera, J., & Niell, F. X. (1995). Red and blue light regulation of growth and photosynthetic metabolism in *Porphyra umbilicalis* (Bangiales, Rhodophyta). *European Journal of Phycology*, *30*(1), 11-18.

Filip, L. C., & Mundy, N. I. (2004). Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates. Molecular Biology and Evolution, 21(8), 1504-1511.

Fishel, R., & Wilson, T. (1997). *mut*S homologs in mammalian cells. *Current Opinion in Genetics & Development*, *7*(1), 105-113.

Flores-Rozas, H., &Kolodner, R. D. (1998). The *Saccharomyces cerevisiae mlh*3 gene functions in *msh*3-dependent suppression of frameshift mutations. *Proceedings of the National Academy of Science*s, USA, 95, 12404-12409.

Freshwater, D. W., Fredericq, S., Butler, B. S., Hommersand, M. H., & Chase, M. W. (1994). A gene phylogeny of the red algae (Rhodophyta) based on plastid *rbc*L. *Proceedings of the National Academy of Sciences*, *91*(15), 7281-7285.

Fries, L. (1980). Axenic tissue cultures from the sporophytes of *Laminaria digitata* and *Laminaria hyperborea* (Phaeophyta). *Journal of Phycology*, *16*(3), 475-477.

Fries, L. (1985). Propagation of *Fucus* (Fucales, Phaeophyta) by hairs with trichothallic growth. *Phycologia*, *24*(4), 481-484.

Fromme, P., & Mathis, P. (2004). Unraveling the Photosystem I reaction center: a history, or the sum of many efforts. Photosynthesis Research, 80(1), 109-124.

Furman, C. M., Elbashir, R., & Alani, E. (2021). Expanded roles for the *mut*L family of DNA mismatch repair proteins. *Yeast*, *38*(1), 39-53.

Gabrielson, P. W., Garbary, D. J., & Scagel, R. F. (1985). The nature of the ancestral red alga: inferences from a cladistic analysis. *BioSystems*, *18*(3-4), 335-346.

Gabrielson, P. W., Garbary, D. J., Sommerfeld, R. A., & Townsend, R. A. (1990). Rhodophyta. Jones and Bartlett Publishers.

Gantt, E. (1980). Structure and function of phycobilisomes: light harvesting pigment complexes in red and blue-green algae. In *International Review of Cytology* (Vol. 66, pp. 45-80). Academic Press.

Gantt, E., Berg, G. M., Bhattacharya, D., Blouin, N. A., Brodie, J. A., Chan, C. X., ... & Brawley, S. H. (2010). *Porphyra*: complex life histories in a harsh environment: *P.*

umbilicalis, an intertidal red alga for genomic analysis. In *Red algae in the genomic age* (pp. 129-148). Springer, Dordrecht.

Gao, F., Nan, F., Song, W., Feng, J., Lv, J., & Xie, S. (2016). Identification and characterization of miRNAs in *Chondrus crispus* by high-throughput sequencing and bioinformatics analysis. *Scientific reports*, *6*(1), 1-10.

Gao, S., & Wang, G. (2012). The enhancement of cyclic electron flow around photosystem I improves the recovery of severely desiccated *Porphyra yezoensis* (Bangiales, Rhodophyta). *Journal of Experimental Botany*, *63*(12), 4349-4358.

Gao, S., Shen, S., Wang, G., Niu, J., Lin, A., & Pan, G. (2011). PSI-driven cyclic electron flow allows intertidal macro-algae *Ulva* sp.(Chlorophyta) to survive in desiccated conditions. *Plant and Cell Physiology*, *52*(5), 885-893.

Garbary, D, J. & Gabrielson, P.W. (1990). Taxonomy and evolution. In Biology of the Red Algae (Cole, K.M. & Sheath, R.G., editors), 477-498. Cambridge University Press, Cambridge.

Garbary, D. J., DJ, G., GI, H., & RF, S. (1980). A revised classification of the Bangiophyceae (Rhodophyta).

Gargiulo, G. M., Genovese, G., Morabito, M., Culoso, F., & De Masi, F. (2001). Sexual and asexual reproduction in a freshwater population of *Bangia atropurpurea* (Bangiales, Rhodophyta) from eastern Sicily (Italy). *Phycologia*, *40*(1), 88-96.

Gau AE, Thole HH, Pistorius EK (1995) Isolation and partial characterization of a manganese requiring L-arginine metabolizing enzyme being present in photosystem II complexes of spinach and tobacco. *Z Naturforsch* 50c:638–651

Gau AE, Thole HH, Sokolenko A, Altschmied L, Herrmann RG, Pistorius EK (1998) *psb*Y, a novel manganese-binding, low-molecular-mass protein associated with photosystem II. Mol Gen Genet 260:56–68

Gause, M., Schaaf, C. A., & Dorsett, D. (2008). Cohesin and CTCF: cooperating to control chromosome conformation?. *Bioessays*, *30*(8), 715-718.

Geesink, R. (1973). Experimental investigations on marine and freshwater *Bangia* (rhodophyta) from the Netherlands. *Journal of Experimental Marine Biology and Ecology*, *11*(3), 239-247.

Genschel, G. T., Filosi, N., Kane, M. F., & Kolodner, R. (1996). Redundancy of *Saccharomyces cerevisiae msh*3 and *msh*6 in *msh*2-dependent mismatch repair. *Genes & Development*, *10*(4), 407-420.

Genschel, J., Littman, S. J., Drummond, J. T., & Modrich, P. (1998). Isolation of *mut*Sβ from human cells and comparison of the mismatch repair specificities of *mut*Sβ and *mut*Sα. *Journal of Biological Chemistry*, *273*(31), 19895-19901.

Gerton, J. L., & DeRisi, J. L. (2002). *mnd*1p: an evolutionarily conserved protein required for meiotic recombination. *Proceedings of the National Academy of Sciences*, *99*(10), 6895-6900.

Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., ... & Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature*, *425*(6959), 737-741.

Gibson, T. M., Shih, P. M., Cumming, V. M., Fischer, W. W., Crockford, P. W., Hodgskiss, M. S., ... & Halverson, G. P. (2018). Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. *Geology*, *46*(2), 135-138.

Gibson, T. M., Shih, P. M., Cumming, V. M., Fischer, W. W., Crockford, P. W., Hodgskiss, M. S., ... & Halverson, G. P. (2018). Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. *Geology*, *46*(2), 135-138.

Gielly, L., & Taberlet, P. (1994). The use of chloroplast DNA to resolve plant phylogenies: noncoding versus *rbc*L sequences. *Molecular Biology and Evolution*, *11*(5), 769-777.

Gillespie, J. H. (1994). The causes of molecular evolution (Vol. 2). Oxford University Press On Demand.

Gillespie, J. H. (1994). The causes of molecular evolution (Vol. 2). Oxford University Press On Demand.

Gillham, N. W., & Boynton, J. E. (1986). The sequence of the chloroplast *atp*B gene and its flanking regions in *Chlamydomonas reinhardtii*. *Gene*, *44*(1), 17-28.

Glöckner, G., Rosenthal, A., & Valentin, K. (2000). The structure and gene repertoire of an ancient red algal plastid genome. *Journal of Molecular Evolution*, *51*(4), 382-390.

Golding, A. J., & Johnson, G. N. (2003). Down-regulation of linear and activation of cyclic electron transport during drought. *Planta*, *218*(1), 107-114.

Goldman, N., & Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, *11*(5), 725-736.

Gong, C. Y., Nan, F. R., Feng, J., Lv, J. P., Liu, Q., Liu, X. D. & Xie, S. L. (2019). Adaptive evolutionary and co-evolutionary analysis on *psa*A gene of Batrachospermales. *Journal of Shanxi University* (Natural Sciense Edition), 42(3): 662-672.

Gong, C. Y., Nan, F. R., Feng, J., Lv, J. P., Liu, Q., & Xie, S. L. (2017). Adaptive evolutionary analysis on rbcL gene of Batrachospermales. Oceanologia Limnol Sinica, 48, 527-535.

Gong, X., Jiang, Q., Xu, J., Zhang, J., Teng, S., Lin, D., & Dong, Y. (2013). Disruption of the rice plastid ribosomal protein S20 leads to chloroplast developmental defects and seedling lethality. *G3: Genes, Genomes, Genetics*, *3*(10), 1769-1777.

Gould, S. B., Maier, U., & Martin, W. F. (2015). Protein import and the origin of red complex plastids. *Current Biology*, *25*(12), R515-R521.

Gould, S. B., Waller, R. F., & McFadden, G. I. (2008). Plastid evolution. *Annu. Rev. Plant Biol.*, *59*, 491-517.

Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, *27*(2), 221-224.

Gradia, S., Acharya, S., & Fishel, R. (2000). The role of mismatched nucleotides in activating the hmsh2-hmsh6 molecular switch. *Journal of Biological Chemistry*, *275*(6), 3922-3930.

Graham, L. E., Cook, M. E., & Busse, J. S. (2000). The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proceedings of the National Academy of Sciences*, *97*(9), 4535-4540.

Graham, S. W., & Olmstead, R. G. (2000). Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany*, *87*(11), 1712-1730.

Gray, M. W. (1993). Origin and evolution of organelle genomes. *Current Opinion in Genetics and Development*, *3*(6), 884-890.

Gray, M. W., Burger, G., & Lang, B. F. (1999). Mitochondrial evolution. *Science*, *283*(5407), 1476-1481.

Green, B. R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *The Plant Journal*, *66*(1), 34-44.

Grelon, M., Vezon, D., Gendrot, G., Pelletier, G. (2001) AtSPO11-1 is necessary for efficient meiotic recombination in plants. *EMBO J.* 20, 589-600.

Gröniger, A., Hallier, C., & Häder, D. P. (1999). Influence of UV radiation and visible light on *Porphyra umbilicalis*: photoinhibition and MAA concentration. *Journal of applied phycology*, *11*(5), 437-445.

Groth, G., & Strotmann, H. (1999). New results about structure, function and regulation of the chloroplast ATP synthase (CF0CF1). *Physiologia Plantarum*, *106*(1), 142-148.

Guacci, V., Koshland, D., & Strunnikov, A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of *mcd*1 in *S. cerevisiae*. *Cell*, *91*(1), 47-57.

Guillon, J. M. (2004). Phylogeny of horsetails (Equisetum) based on the chloroplast *rps*4 gene and adjacent noncoding sequences. *Systematic Botany*, *29*(2), 251-259.

Guiry, M. D., & Guiry, G. 2. (2017). AlgaeBase. *AlgaeBase*.

Guisinger, M. M., Kuehl, J. V., Boore, J. L. & Jansen, R. K. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc. Natl. Acad. Sci. USA* 105, 18424–18429 (2008).

Güven, K. C., Özsoy, Y., & Ulutin, O. N. (2021). Anticoagulant, fibrinolytic and antiaggregant activity of carrageenans and alginic acid. In *Volume 34, 1991* (pp. 429-432). De Gruyter.

Haehnel, W., Jansen, T., Gause, K., Klösgen, R. B., Stahl, B., Michl, D., ... & Herrmann, R. G. (1994). Electron transfer from plastocyanin to photosystem I. *The EMBO Journal*, 13(5), 1028-1038.

Haering, C. H., & Nasmyth, K. (2003). Building and breaking bridges between sister chromatids. *Bioessays*, *25*(12), 1178-1191.

Hagstrom, K. A., Holmes, V. F., Cozzarelli, N. R., & Meyer, B. J. (2002). *C. elegans* condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes & Development*, *16*(6), 729-742.

Hakenbeck, R., & Stock, J. B. (1996). Analysis of two-component signal transduction systems involved in transcriptional regulation. In *Methods in Enzymology* (Vol. 273, pp. 281-300). Academic Press.

Hallick, R. B., & Bairoch, A. (1994). Proposals for the naming of chloroplast genes. III. Nomenclature for open reading frames encoded in chloroplast genomes. *Plant Molecular Biology Reporter*, *12*(2), S29-S30.

Hamant, O., Ma, H., & Cande, W. Z. (2006). Genetics of meiotic prophase I in plants. *Annu. Rev. Plant Biol.*, *57*, 267-302.

Hamilton, W. D., & Hamilton, W. D. (1996). *Narrow roads of gene land: volume 2: evolution of sex* (Vol. 2). Oxford University Press.

Han, J. F., Nan, F. R., Feng, J., Lv, J. P., Liu, Q., Liu, X. D., & Xie, S. L. (2019). *Sheathia matouensis* (Batrachospermales, Rhodophyta), a new freshwater red algal species from North China. *Phytotaxa*, *415*(5), 255-263.

Han, Y., Liu, X., Nan, F., Feng, J., Lv, J., Liu, Q., & Xie, S. (2021). Analysis of Adaptive Evolution and Coevolution of rbcL Gene in the Genus Galdieria (Rhodophyta). Journal of Eukaryotic Microbiology, 68(2), e12838.

Harfe, B. D., Minesinger, B. K., &Jinks-Robertson, S. (2000). Discrete in vivo roles for the *mut*L homologs *mlh*2p and *mlh* 3p in the removal of frameshift intermediates in budding yeast. *Current Biology*, 10, 145-148.

Harper, J. T., & Saunders, G. W. (2001). Molecular systematics of the Florideophyceae (Rhodophyta) using nuclear large and small subunit rDNA sequence data. *Journal of phycology*, *37*(6), 1073-1082.

Hartmann, M., & Golding, G. B. (1998). Searching for substitution rate heterogeneity. *Molecular phylogenetics and evolution*, *9*(1), 64-71.

Hartung, F. and Puchta, H. (2001). Molecular characterization of homologues of both subunits A (*spo*11) and B of the archaebacterial topoisomerase 6 in plants. *Gene*, 271, 21-26.

Hartung, F., & Puchta, H. (2000). Molecular characterisation of two paralogous *spo*11 homologues in *Arabidopsis thaliana*. *Nucleic Acids Research*, *28*(7), 1548-1554.

Hartung, F., Angelis, K. J., Meister, A., Schubert, I., Melzer, M., & Puchta, H. (2002). An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Current Biology*, *12*(20), 1787-1791.

Hartung, F., Wurz-Wildersinn, R., Fuchs, J., Schubert, I., Suer, S., & Puchta, H. (2007). The catalytically active tyrosine residues of both *spo*11-1 and *spo*11-2 are required for meiotic double-strand break induction in *Arabidopsis*. *The Plant Cell*, *19*(10), 3090-3099.

Haversat, J., Woglar, A., Klatt, K., Akerib, C. C., Roberts, V., Salazar, C. C., ... & Kim, Y. (2021). Robust designation of meiotic crossover sites by CDK-2 through phosphorylation of the MutSγ complex. *bioRxiv*.

Hayashi, L., & Reis, R. P. (2012). Cultivation of the red algae *Kappaphycus alvarezii* in Brazil and its pharmacological potential. *Revista Brasileira de Farmacognosia*, *22*, 748-752.

Hayden, H. S., Blomster, J., Maggs, C. A., Silva, P. C., Stanhope, M. J., & Waaland, J. R. (2003). Linnaeus was right all along: *Ulva* and *Enteromorpha* are not distinct genera. *European Journal of Phycology*, *38*(3), 277-294.

He, L., Zhu, J., Lu, Q., Niu, J., Zhang, B., Lin, A., & Wang, G. (2013). Genetic similarity analysis within *Pyropia yezoensis* blades developed from both conchospores and blade archeospores using AFLP. *Journal of Phycology*, *49*(3), 517-522.

Hebert, P. D., & Gregory, T. R. (2005). The promise of DNA barcoding for taxonomy. *Systematic biology*, *54*(5), 852-859.

Hejase, H. A., Dukler, N., & Siepel, A. (2020). From summary statistics to gene trees: methods for inferring positive selection. *Trends in Genetics*, *36*(4), 243-258.

Henry, J. M., Camahort, R., Rice, D. A., Florens, L., Swanson, S. K., Washburn, M. P., & Gerton, J. L. (2006). *mnd*1/*hop*2 facilitates Dmc1-dependent interhomolog crossover formation in meiosis of budding yeast. *Molecular and Cellular Biology*, *26*(8), 2913-2923.

Her, C., & Doggett, N. A. (1998). Cloning, Structural Characterization, and Chromosomal Localization of the Human Orthologue of *Saccharomyces cerevisiae msh*5 Gene. *Genomics*, *52*(1), 50-61.

Herrmann, R. G., Westoff, P., Alt, J., Tittgen, J., & Nelson, N. (1985). Thylakoid membrane proteins and their genes. *NATO Advanced Science Institutes Series: Series A: Life Sciences (USA)*.

Higgins, D. G., Thompson, J. D., & Gibson, T. J. (1996). [22] Using CLUSTAL for multiple sequence alignments. *Methods in Enzymology*, *266*, 383-402.

Hippler, M., Drepper, F., Farah, J., & Rochaix, J. D. (1997). Fast electron transfer from cytochrome c 6 and plastocyanin to photosystem I of Chlamydomonas reinhardtii requires psaF. Biochemistry, 36(21), 6343-6349.

Hirano, T. (2005). Condensins: organizing and segregating the genome. *Current Biology*, *15*(7), R265-R275.

Hirano, T., Kobayashi, R., & Hirano, M. (1997). Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a Xenopus homolog of the *Drosophila barren* protein. *Cell*, *89*(4), 511-521.

Hofstatter, P. G., Ribeiro, G. M., Porfírio-Sousa, A. L., & Lahr, D. J. (2020). The Sexual Ancestor of all Eukaryotes: A Defense of the “Meiosis Toolkit” A Rigorous Survey Supports the Obligate Link between Meiosis Machinery and Sexual Recombination. *BioEssays*, *42*(9), 2000037.

Hollants, J., Leliaert, F., De Clerck, O., & Willems, A. (2013). What we can learn from sushi: A review on seaweed–bacterial associations. *FEMS Microbiology Ecology*, *83*(1), 1-16.

Hommersand, M. H., Fredericq, S., & Freshwater, D. W. (1994). Phylogenetic systematics and biogeography of the Gigartinaceae (Gigartinales, Rhodophyta) based on sequence analysis of *rbc*L.

Hong, L., Stevenson, J. K., Roth, W. B., & Hallick, R. B. (1995). *Euglena gracilis* chloroplast *psb*B, *psb*T, *psb*H and *psb*N gene cluster: regulation of *psb*B-*psb*T pre-mRNA processing. *Molecular and General Genetics* MGG, 247(2), 180-188.

Hoot, S. B., Culham, A., & Crane, P. R. (1995). The utility of *atp*B gene sequences in resolving phylogenetic relationships: comparison with rbcL and 18S ribosomal DNA sequences in the Lardizabalaceae. *Annals of the Missouri Botanical Garden*, 194-207.

Hope, A. B. (1993). The chloroplast cytochrome bf complex A critical focus on function. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1143*(1), 1-22.

Horn, J. W. (2009). Phylogenetics of Dilleniaceae using sequence data from four plastid loci (*rbc*L, *inf*A, *rps*4, *rpl*16 intron). *International Journal of Plant Sciences*, *170*(6), 794-813.

Horst, N. A., & Reski, R. (2016). Alternation of generations–unravelling the underlying molecular mechanism of a 165-year-old botanical observation. *Plant Biology*, *18*(4), 549-551.

Horváth, E. M., Peter, S. O., Joët, T., Rumeau, D., Cournac, L., Horváth, G. V., ... & Medgyesy, P. (2000). Targeted inactivation of the plastid *ndh*B gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. *Plant Physiology*, *123*(4), 1337-1350.

Hu, Z. M., Duan, D. L., & Lopez-Bautista, J. (2016). Seaweed phylogeography from 1994 to 2014: an overview. *Seaweed phylogeography*, 3-22.

Hudson, G. S., & Mason, J. G. (1988). The chloroplast genes encoding subunits of the H+-ATP synthase. *Molecular Biology of Photosynthesis*, 565-582.

Huelsenbeck, J. P., Larget, B., Miller, R. E., & Ronquist, F. (2002). Potential applications and pitfalls of Bayesian inference of phylogeny. Systematic Biology, 51(5), 673-688.

Hughey, J. R. (2016). Genomic and phylogenetic analysis of the complete plastid genome of the California endemic seaweed *Wildemania schizophylla* (Bangiaceae). *Madrono, 63*(1), 34-38.

Hughey, J. R., Gabrielson, P. W., Rohmer, L., Tortolani, J., Silva, M., Miller, K. A., ... & Ruediger, E. (2014). Minimally destructive sampling of type specimens of *Pyropia* (Bangiales, Rhodophyta) recovers complete plastid and mitochondrial genomes. *Scientific Reports*, *4*(1), 1-10.

Hughey, J.R., Gabrielson, P.W., Rohmer, L., Tortolani, J., Silva, M., Miller, K.A., Young, J.D., Martell, C. and Ruediger, E. (2014). Minimally destructive sampling of type specimens of *Pyropia* (Bangiales, Rhodophyta) recovers complete plastid and mitochondrial genomes. *Scientific Reports*, *4*, 5113.

Hunter, N., & Borts, R. H. (1997). *mlh*1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis. *Genes & Development*, *11*(12), 1573-1582.

Hurtado, A. Q., Gerung, G. S., Yasir, S., & Critchley, A. T. (2014). Cultivation of tropical red seaweeds in the BIMP-EAGA region. *Journal of Applied Phycology*, *26*(2), 707-718.

Hussain, S., & Rasool, S. T. (2017). Analysis of synonymous codon usage in Zika virus. *Acta Tropica*, *173*, 136-146.

Hwang, M. S., Kim, S. M., Jung, M. Y., Choi, H. G., & Kim, H. S. (2008). A revision of *Porphyra* and *Bangia* (Bangiales, Rhodophyta) from Korea based on morphological and molecular data.

Ihnatowicz, A., Pesaresi, P., Varotto, C., Richly, E., Schneider, A., Jahns, P., ... & Leister, D. (2004). Mutants for photosystem I subunit D of *Arabidopsis thaliana*: effects on photosynthesis, photosystem I stability and expression of nuclear genes for chloroplast functions. *The Plant Journal*, 37(6), 839-852.

Ihua, M., Dobson, A., & Stengel, D. (2019). Metagenomic analysis reveals significant seasonal variations in the epiphytic bacterial communities associated with different parts of the brown seaweed *Laminaria digitata*. *Access Microbiology*, *1*(1A)

Iqbal, J., Shinwari, Z. K., & Mahmood, T. (2019). Phylogenetic relationships within the cosmopolitan family rhamnaceae using *atp*B gene promoter. *Pakistan Journal of Botany*, *51*(3), 1027-1040.

Iwai, M., Takizawa, K., Tokutsu, R., Okamuro, A., Takahashi, Y., & Minagawa, J. (2010). Isolation of the elusive supercomplex that drives cyclic electron flow in photosynthesis. *Nature*, *464*(7292), 1210-1213.

Jabeen, A., Guo, B., Abbasi, B. H., Shinwari, Z. K., & Mahmood, T. (2012). Phylogenetics of selected Mentha species on the basis of *rps*8, *rps*11 and *rps*14 chloroplast genes. *Journal of Medicinal Plants Research*, *6*(1), 30-36.

Jackson, M. B. (1985). The red alga *Bangia* in lake Simoe. *Journal of Great Lakes Research, 11*(2), 179-181.

Jain, K., Krause, K., Grewe, F., Nelson, G. F., Weber, A. P., Christensen, A. C., & Mower, J. P. (2015). Extreme features of the *Galdieria sulphuraria* organellar genomes: a consequence of polyextremophily?. *Genome biology and evolution*, *7*(1), 367-380.

Janouškovec, J., Horák, A., Oborník, M., Lukeš, J., & Keeling, P. J. (2010). A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. *Proceedings of the National Academy of Sciences*, *107*(24), 10949-10954.

Janouškovec, J., Liu, S. L., Martone, P. T., Carré, W., Leblanc, C., Collén, J., & Keeling, P. J. (2013). Evolution of red algal plastid genomes: ancient architectures, introns, horizontal gene transfer, and taxonomic utility of plastid markers. *PloS one*, *8*(3), e59001.

Javaux, E. J., Knoll, A. H., & Walter, M. R. (2001). Morphological and ecological complexity in early eukaryotic ecosystems. *Nature*, *412*(6842), 66-69.

Jessberger, R. (2002). The many functions of SMC proteins in chromosome dynamics. *Nature Reviews Molecular Cell Biology*, *3*(10), 767-778.

Jessberger, R., Frei, C., & Gasser, S. M. (1998). Chromosome dynamics: the SMC protein family. *Current opinion in Genetics & Development*, *8*(2), 254-259.

Jiang, Y., Yan, X., & Liu, C. (2010). Selection and characterization of a fast growing strain of *Porphyra haitanensis* (Bangiales, Rhodophyta). *Journal of Fisheries of China*, *34*(9), 1363-1370.

Jo, Y. D., Park, J., Kim, J., Song, W., Hur, C. G., Lee, Y. H., & Kang, B. C. (2011). Complete sequencing and comparative analyses of the pepper (*Capsicum annuum* L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. *Plant Cell Reports*, *30*(2), 217-229.

Joët, T., Cournac, L., Peltier, G., & Havaux, M. (2002). Cyclic Electron Flow around Photosystem I in C3Plants. In Vivo Control by the Redox State of Chloroplasts and Involvement of the NADH-Dehydrogenase Complex. *Plant Physiology*, *128*(2), 760-769.

Johnson, C. H., & Schmidt, G. W. (1993). The *psb*B gene cluster of the *Chlamydomonas reinhardtii* chloroplast: sequence and transcriptional analyses of *psb*N and *psb*H. *Plant Molecular Biology*, 22(4), 645-658.

Jones, S., & Sgouros, J. (2001). The cohesin complex: sequence homologies, interaction networks and shared motifs. *Genome Biology*, *2*(3), 1-12.

Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. Mammalian protein Metabolism, 3, 21-132.

Kadyrov, F. A., Holmes, S. F., Arana, M. E., Lukianova, O. A., O'Donnell, M., Kunkel, T. A., & Modrich, P. (2007). *Saccharomyces cerevisiae mut*Lα is a mismatch repair endonuclease. *Journal of Biological Chemistry*, *282*(51), 37181-37190.

Kalitnik, A. A., Marcov, P. A., Anastyuk, S. D., Barabanova, A. O. B., Glazunov, V. P., Popov, S. V., ... & Yermak, I. M. (2015). Gelling polysaccharide from *Chondrus armatus* and its oligosaccharides: The structural peculiarities and anti-inflammatory activity. *Carbohydrate polymers*, *115*, 768-775.

Kamiya, M., & Kawai, H. (2002). Dependence of the carposporophyte on the maternal gametophyte in three ceramiacean algae (Rhodophyta), with respect to carposporophyte development, spore production and germination success. *Phycologia*, *41*(2), 107-115.

Kapralov, M. V., & Filatov, D. A. (2007). Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evolutionary Biology, 7(1), 1-10.

Karsten, U. (2012). Seaweed acclimation to salinity and desiccation stress. In *Seaweed biology* (pp. 87-107). Springer, Berlin, Heidelberg.

Karsten, U., & West, J. A. (2000). Living in the intertidal zone—seasonal effects on heterosides and sun-screen compounds in the red alga *Bangia atropurpurea* (Bangiales). *Journal of Experimental Marine Biology and Ecology*, *254*(2), 221-234.

Kasanah, N., Triyanto, T., Seto, D. S., Amelia, W., & Isnansetyo, A. (2015). Antibacterial compounds from red seaweeds (Rhodophyta). *Indonesian Journal of Chemistry*, *15*(2), 201-209.

Kashino Y, Lauber WM, Carroll JA, Wang Q, Whitmarsh J, Satoh K, Pakrasi HB. (2002). Proteomic analysis of a highly active photosystem II preparation from the cyanobacterium *Synechocystis* sp. PCC 6803 reveals the presence of novel polypeptides. *Biochemistry* 41:8004–8012

Ke, B. (2001). Oxygen Evolution Introduction. *Photosynthesis: Photobiochemistry and Photobiophysics*, 323-336.

Keeling, P. J., & Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, *9*(8), 605-618.

Keeling, P. J., Burger, G., Durnford, D. G., Lang, B. F., Lee, R. W., Pearlman, R. E., ... & Gray, M. W. (2005). The tree of eukaryotes. *Trends in ecology & evolution*, *20*(12), 670-676.

Keeney, S., Baudat, F., Angeles, M., Zhou, Z. H., Copeland, N. G., Jenkins, N. A., ... & Jasin, M. (1999). A mouse homolog of the *Saccharomyces cerevisiae* meiotic recombination DNA transesterase *spo*11p. *Genomics*, *61*(2), 170-182.

Keller, J., Rousseau-Gueutin, M., Martin, G. E., Morice, J., Boutte, J., Coissac, E., ... & Aïnouche, A. (2017). The evolutionary fate of the chloroplast and nuclear *rps*16 genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from *Lupinus*. *DNA Research*, *24*(4), 343-358.

Kerzendorfer, C., Vignard, J., Pedrosa-Harand, A., Siwiec, T., Akimcheva, S., Jolivet, S., ... & Schlogelhofer, P. (2006). The *Arabidopsis thaliana mnd*1 homologue plays a key role in meiotic homologous pairing, synapsis and recombination. *Journal of Cell Science*, *119*(12), 2486-2496.

Kessler, U., Maid, U., & Zetsche, K. (1992). An equivalent to bacterial *omp*R genes is encoded on the plastid genome of red algae. *Plant Molecular Biology*, *18*(4), 777-780.

Khan, H., Parks, N., Kozera, C., Curtis, B. A., Parsons, B. J., Bowman, S., et al. (2007). Plastid genome sequence of the Cryptophyte alga *Rhodomonas salina* CCMP1319: Lateral transfer of putative DNA replication machinery and a test of chromist plastid phylogeny. *Molecular Biology and Evolution, 24*(8), 1832-1842.

Khoa, H. V., Kumari, P., Uchida, H., Murakami, A., Shimada, S., & Mikami, K. (2021). Heat-Stress Responses Differ among Species from Different '*Bangia*'Clades of Bangiales (Rhodophyta). *Plants, 10*(8), 1733.

Kim, E., Archibald, J. M., Aronsson, H., & Sandelius, A. S. (2009*).* Diversity and evolution of plastids and their genomes. In *The chloroplast—interactions with the environment* (p. 1-39). Berlin (Germany) Springer-Verlag.

Kim, K. M., Park, J., Bhattacharya, D., & Yoon, H. S. (2014). Applications of next-generation sequencing to unravelling the evolutionary history of algae. *International Journal of Systematic and Evolutionary Microbiology*, *64*(2), 333-345.

Kim, S. H., & Kim, G. H. (1999). Cell-cell recognition during fertilization in the red alga, *Aglaothamnion oosumiense* (Ceramiaceae, Rhodophyta). In *Sixteenth International Seaweed Symposium* (pp. 81-89). Springer, Dordrecht.

Kimura, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetics research*, *11*(3), 247-270.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution, 16(2), 111-120.

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.

King, J. L., & Jukes, T. H. (1969). Non-Darwinian evolution. *University of California, Berkeley*, *94720*, 735.

King, M. C. (1975). AC Wilson. *Evolution at two levels in humans and chimpanzees*, 107-16.

King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science*, *188*(4184), 107-116.

Kirst, G. O. (1990). Salinity tolerance of eukaryotic marine algae. *Annual review of plant biology*, *41*(1), 21-53.

Kitade, Y., Asamizu, E., Fukuda, S., Nakajima, M., Ootsuka, S., Endo, H., ... & Saga, N. (2008). Identification of genes preferentially expressed during sexual sporulation in *Porphyra yezoensis* gametophytes (Bangia*a*les, Rhodophyta) 1. *Journal of Phycology*, *44*(1), 113-123.

Klapholz, S., Waddell, C. S., & Esposito, R. E. (1985). The role of the *spo*11 gene in meiotic recombination in yeast. *Genetics*, *110*(2), 187-216.

Kleckner, N. (1996). Meiosis: how could it work?. *Proceedings of the National Academy of Sciences*, *93*(16), 8167-8174.

Klein, A. S., Mathieson, A. C., Neefus, C. D., Cain, D. F., Taylor, H. A., Teasdale, B. W., ... & Wallace, A. L. (2003). Identification of north-western Atlantic *Porphyra* (Bangiaceae, Bangiales) based on sequence variation in nuclear SSU and plastid *rbc*L genes. *Phycologia*, *42*(2), 109-122.

Klein, R. R., & Mullet, J. E. (1987). Control of gene expression during higher plant chloroplast biogenesis. Protein synthesis and transcript levels of *psb*A, *psa*A-*psa*B, and *rbc*L in dark-grown and illuminated barley seedlings. *Journal of Biological Chemistry*, 262(9), 4341-4348.

Kneitz, B., Cohen, P. E., Avdievich, E., Zhu, L., Kane, M. F., Hou, H., ... & Edelmann, W. (2000). *mut*S homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes & Development*, *14*(9), 1085-1097.

Kobayashi, Y., Harada, N., Nishimura, Y., Saito, T., Nakamura, M., Fujiwara, T., ... & Misumi, O. (2014). Algae sense exact temperatures: small heat shock proteins are expressed at the survival threshold temperature in *Cyanidioschyzon merolae* and *Chlamydomonas reinhardtii*. *Genome biology and evolution*, *6*(10), 2731-2740.

Kogame, K., Uwai, S., Anderson, R., Choi, H., & Bolton, J. (2017). DNA barcoding of south African geniculate coralline red algae (Corallinales, Rhodophyta). *South African Journal of Botany*, *108*, 337-341.

Kohzuma, Kaori, John E. Froehlich, Geoffry A. Davis, Joshua A. Temple, Deepika Minhas, Amit Dhingra, Jeffrey A. Cruz, and David M. Kramer. (2017). The role of light–dark regulation of the chloroplast ATP synthase. *Frontiers in plant science* 8 1248.

Kolodner, R. D., & Marsischky, G. T. (1999). Eukaryotic DNA mismatch repair. *Current opinion in genetics & development*, *9*(1), 89-96.

Kong, M., Wang, F., Yang, Z., & Mi, H. (2013). *atp*G is required for the accumulation and function of chloroplast ATP synthase in *Arabidopsis*. *Chinese Science Bulletin*, 58(26), 3224-3232.

Kornmann, P. (1994). Life histories of monostromatic *Porphyra* species as a basis for taxonomy and classification. *European Journal of Phycology*, *29*(2), 69-71.

Kosiol, C., Vinař, T., da Fonseca, R. R., Hubisz, M. J., Bustamante, C. D., Nielsen, R., & Siepel, A. (2008). Patterns of positive selection in six mammalian genomes. *PLoS genetics*, *4*(8), e1000144.

Kostrzewa, M., Valentin, K., Maid, U., Radetzky, R., & Zetsche, K. (1990). Structure of the Rubisco Operon from the multicellular red alga *Antithamnion* spec. *Current Genetics*, *18*(5), 465-469.

Kramer, D. M., Avenson, T. J., & Edwards, G. E. (2004). Dynamic flexibility in the light reactions of photosynthesis governed by both electron and proton transfer reactions. *Trends in Plant Science*, *9*(7), 349-357.

Krawchuk, M. D., DeVeaux, L. C., & Wahls, W. P. (1999). Meiotic chromosome dynamics dependent upon the rec8+, rec10+ and rec11+ genes of the fission yeast *Schizosaccharomyces pombe*. *Genetics*, *153*(1), 57-68.

Kucera, H., & Saunders, G. W. (2012). a survey of Bangiales (Rhodophyta) based on multiple molecular markers reveals cryptic diversity1. *Journal of Phycology*, *48*(4), 869-882.

Kulkarni, D. S., Owens, S. N., Honda, M., Ito, M., Yang, Y., Corrigan, M. W., ... & Hunter, N. (2020). PCNA activates the *mut*Lγ endonuclease to promote meiotic crossing over. *Nature*, *586*(7830), 623-627.

Kumano, S. (1980). On the distribution of some freshwater red algae in Japan and Southeast Asia. Organizing Committee 21 SIL Congress.

Kumano, S. (2002). Freshwater red algae of the world. Biopress.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547-1549.

Kunimoto, M., Kito, H., Kaminishi, Y., Mizukami, Y., & Murase, N. (1999). Molecular divergence of the SSU rRNA gene and internal transcribed spacer 1 in *Porphyra yezoensis* (Rhodophyta). *Journal of Applied Phycology*, *11*(2), 211-216.

Künstner, P., Guardiola, A., Takahashi, Y., & Rochaix, J. D. (1995). A Mutant Strain of *Chlamydomonas reinhardtii* Lacking the Chloroplast Photosystem II *psb*I Gene Grows Photoautotrophically (∗). *Journal of Biological Chemistry*, 270(16), 9651-9654.

Kurisu, G., Zhang, H., Smith, J. L., & Cramer, W. A. (2003). Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. *Science*, *302*(5647), 1009-1014.

Kuroiwa, T. (1998). The primitive red algae *Cyanidium caldarium* and *Cyanidioschyzon merolae* as model system for investigating the dividing apparatus of mitochondria and plastids. *Bioessays*, *20*(4), 344-354.

Kurzbauer, M. T., Uanschou, C., Chen, D., & Schlögelhofer, P. (2012). The recombinases *dmc*1 and *rad*51 are functionally and spatially separated during meiosis in *Arabidopsis*. *The Plant Cell*, *24*(5), 2058-2070.

Lahiri, S., Li, Y., Hingorani, M. M., & Mukerji, I. (2018). *mut*Sγ-induced DNA conformational changes provide insights into its role in meiotic recombination. *Biophysical Journal*, *115*(11), 2087-2101.

Lang, B. F., & Nedelcu, A. M. (2012). Plastid genomes of algae. In *Genomics of chloroplasts and mitochondria* (pp. 59-87). Springer, Dordrecht.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357.

Larkum, A. W., Lockhart, P. J., & Howe, C. J. (2007). Shopping for plastids. *Trends in Plant Science*, *12*(5), 189-195.

Lee, H. N., Kim, S. H., Han, Y. J., Im, S., Jeong, W. J., Park, E. J., ... & Choi, D. W. (2017). PsCYP1 of marine red alga *Pyropia seriata* (Bangiales, Rhodophyta) confers salt and heat tolerance in Chlamydomonas. *Journal of Applied Phycology*, *29*(1), 617-625.

Lee, J. B., Cho, W. K., Park, J., Jeon, Y., Kim, D., Lee, S. H., & Fishel, R. (2014). Single-molecule views of *mut*S on mismatched DNA. *DNA Repair*, *20*, 82-93.

Lee, J., Cho, C. H., Park, S. I., Choi, J. W., Song, H. S., West, J. A., ... & Yoon, H. S. (2016). Parallel evolution of highly conserved plastid genome architecture in red seaweeds and seed plants. *BMC biology*, *14*(1), 1-16.

Lehmann, A. R. (2005). The role of SMC proteins in the responses to DNA damage. *DNA Repair*, *4*(3), 309-314.

Leliaert, F., & Lopez-Bautista, J. M. (2015). The chloroplast genomes of Bryopsis plumosa and Tydemania expeditiones (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. BMC Genomics, 16(1), 1-20.

Leu, J. Y., Chua, P. R., & Roeder, G. S. (1998). The meiosis-specific *hop*2 protein of *S. cerevisiae* ensures synapsis between homologous chromosomes. *Cell*, *94*(3), 375-386.

Levy, I., & Gantt, E. (1990). Development of Photosynthetic Activity in *Porphyridium purpureum* (Rhodophyta) Following Nitrogen Starvation 1, 2. *Journal of Phycology*, *26*(1), 62-68.

Li, C., Irie, R., Shimada, S., & Mikami, K. (2019). Requirement of different normalization genes for quantitative gene expression analysis under abiotic stress conditions in '*Bangia*'sp. *ESS1. J. Aquat. Res. Mar. Sci*, *2019*, 194-205.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589-595.

Li, H., Gao, D., & Yan, X. (2021). Identification and expression analysis of a calmodulin gene in *Phycocalida chauhanii* (Rhodophyceae, *Bangia*les). *Journal of Applied Phycology*, *33*(3), 1915-1923.

Li, S. Y., & Wang, J. C. (1984). The influence of the light intensity on the formation, discharge and adherence of monospores. *Mar Sci*, *8*, 41-43.

Li, W. (1997). Molecular evolution. Sinauer associates incorporated.

Li, Y., Muir, K. W., Bowler, M. W., Metz, J., Haering, C. H., & Panne, D. (2018). Structural basis for *scc*3-dependent cohesin recruitment to chromatin. *Elife*, *7*, e38356.

Lim, P. E., Yang, L. E., Tan, J., Maggs, C. A., & Brodie, J. (2017). Advancing the taxonomy of economically important red seaweeds (Rhodophyta). *European Journal of Phycology*, *52*(4), 438-451.

Lin, A. P., Wang, G. C., Yang, F., & Pan, G. H. (2009). Photosynthetic parameters of sexually different parts of *Porphyra katadai* var. hemiphylla (Bangiales, Rhodophyta) during dehydration and re-hydration. *Planta*, *229*(4), 803-810.

Lin, C. K., & Blum, J. L. (1977). Recent invasion of a red alga (*Bangia atropurpurea*) in Lake Michigan. *Journal of the Fisheries Board of Canada*, *34*(12), 2413-2416.

Lin, Z., Nei, M., & Ma, H. (2007). The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution. *Nucleic acids research*, *35*(22), 7591-7603.

Lindstrom, S. C., & Cole, K. M. (1992). A revision of the species of *Porphyra* (Rhodophyta: Bangiales) occurring in British Columbia and adjacent waters. *Canadian Journal of Botany*, *70*(10), 2066-2075.

Lindstrom, S. C., & Fredericq, S. (2003). *rbc*L gene sequences reveal relationships among north-east Pacific species of Porphyra (Bangiales, Rhodophyta) and a new species, *P. aestivalis*. *Phycological Research*, *51*(3), 211-211.

Liu, N., Wang, Q., Chen, J., Zhu, Y., Tashi, T., Hu, Y., ... & Zhong, Y. (2010). Adaptive evolution and structure modeling of rbcL gene in Ephedra. Chinese Science Bulletin, 55(22), 2341-2346.

Liu, X., Bogaert, K., Engelen, A. H., Leliaert, F., Roleda, M. Y., & De Clerck, O. (2017). Seaweed reproductive biology: environmental and genetic controls. *Botanica Marina*, *60*(2), 89-108.

Losada, A., & Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes & Development*, *19*(11), 1269-1287.

Losada, A., Yokochi, T., Kobayashi, R., & Hirano, T. (2000). Identification and characterization of SA/*scc*3p subunits in the Xenopus and human cohesin complexes. *The Journal of Cell Biology*, *150*(3), 405-416.

Lumsden, T., Bentley, A. A., Beutler, W., Ghosh, A., Galkin, O., & Komar, A. A. (2010). Yeast strains with N-terminally truncated ribosomal protein S5: implications for the evolution, structure and function of the *rps*5/*rps*7 proteins. *Nucleic Acids Research*, *38*(4), 1261-1272.

Lynch, M. D., Sheath, R. G., & Müller, K. M. (2008). Phylogenetic position and ISSR-estimated intraspecific genetic variation of *Bangia maxima* (Bangiales, Rhodophyta). *Phycologia*, *47*(6), 599-613.

Lyngbye, H. C. (1819). *Tentamen hydrophytologiae danicae continens omnia hydrophyta cryptogama Daniae, Holsatiae, Faeroae, Islandiae, Groenlandiae hucusque cognita... adjectis simul speciebus Norvegicis*. Schultz.

Magee, A. M., Aspinall, S., Rice, D. W., Cusack, B. P., Sémon, M., Perry, A. S., ... & Wolfe, K. H. (2010). Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Research*, *20*(12), 1700-1710.

Mahmood, T., Hassan, N., Nazar, N., & Naveed, I. (2011). Phylogenetic analysis of different *Artemisia* species based on chloroplast gene *rps*11. *Archives of Biological Sciences*, *63*(3), 661-665.

Maiwald, D., Dietzmann, A., Jahns, P., Pesaresi, P., Joliot, P., Joliot, A., ... & Leister, D. (2003). Knock-out of the genes coding for the Rieske protein and the ATP-synthase δ-subunit of *Arabidopsis*. Effects on photosynthesis, thylakoid protein composition, and nuclear chloroplast gene expression. *PlantPhysiology*, 133(1), 191-202.

Malik, S. B., Pightling, A. W., Stefaniak, L. M., Schurko, A. M., & Logsdon Jr, J. M. (2008). An expanded inventory of conserved meiotic genes provides evidence for sex in *Trichomonas vaginalis*. *PloS One*, *3*(8), e2879.

Manhart, C. M., Alani, E. (2016). Roles for mismatch repair family proteins in promoting meiotic crossing over. *DNA Repair*, 38, 84-93.

Maris, Ann E., Michael R. Sawaya, Maria Kaczor-Grzeskowiak, Michael R. Jarvis, Shawn MD Bearson, Mary L. Kopka, Imke Schröder, Robert P. Gunsalus, and Richard E. Dickerson. (2002). Dimerization allows DNA target site recognition by the NarL response regulator. *Nature structural biology* 9(10), 771-778.

Marquardt, J., & Rehm, A. M. (1995). *Porphyridium purpureum* (Rhodophyta) from red and green light: characterization of photosystem I and determination of in situ fluorescence spectra of the photosystems. *Journal of Photochemistry and Photobiology B: Biology*, *30*(1), 49-56.

Martin, W. F., Tielens, A. G., Mentel, M., Garg, S. G., & Gould, S. B. (2017). The physiology of phagocytosis in the context of mitochondrial origin. *Microbiology and Molecular Biology Reviews*, *81*(3), e00008-17.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M., & Kowallik, K. V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature*, *393*(6681), 162-165.

Martins, L. O., Empadinhas, N., Marugg, J. D., Miguel, C., Ferreira, C., da Costa, M. S., & Santos, H. (1999). Biosynthesis of mannosylglycerate in the thermophilic bacterium *Rhodothermus marinus*: biochemical and genetic characterization of a mannosylglycerate synthase. *Journal of Biological Chemistry*, *274*(50), 35407-35414.

Martirosyan, E. V., Ryzhova, N. N., Kochieva, E. Z., & Skryabin, K. G. (2009). Analysis of chloroplast *rps*16 intron sequences in Lemnaceae. *Molecular Biology*, *43*(1), 32-38.

Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. Journal of the American statistical Association, 46(253), 68-78.

Maul, J. E., Lilly, J. W., Cui, L., DePamphilis, C. W., Miller, W., Harris, E. H., & Stern, D. B. (2002). The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. *The Plant Cell*, *14*(11), 2659-2679.

McFadden, G. I. (2014). Origin and evolution of plastids and photosynthesis in eukaryotes. *Cold Spring Harbor perspectives in biology*, *6*(4), a016105.

Meetam M, Keren N, Ohad I, Pakrasi HB. (1999). The *psb*Y protein is not essential for oxygenic photosynthesis in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Physiol* 121:1267–1272

Merola, A., Castaldo, R., Luca, P. D., Gambardella, R., Musacchio, A., & Taddei, R. (1981). Revision of *Cyanidium caldarium*. Three species of acidophilic algae. *Plant Biosystem*, *115*(4-5), 189-195.

Metzler-Guillemain, C., & de Massy, B. (2000). Identification and characterization of an *spo*11 homolog in the mouse. *Chromosoma*, *109*(1), 133-138.

Meyer, B. J. (2005). X-Chromosome dosage compensation. *WormBook: The Online Review of C. elegans Biology [Internet]*.

Meynard, A., Zapata, J., Salas, N., Betancourtt, C., Pérez-Lara, G., Castañeda, F., ... & Contreras-Porcia, L. (2019). Genetic and morphological differentiation of *Porphyra* and *Pyropia* species (Bangiales, Rhodophyta) coexisting in a rocky intertidal in Central Chile. *Journal of Phycology*, *55*(2), 297-313.

Michaelis, C., Ciosk, R., & Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, *91*(1), 35-45.

Mikami, K. (2019). Diploid apogamy in red algal species of the genus *Pyropia*. *J. Aquat. Res. Mar. Sci*, *2019*, 206-208.

Mikami, K., Li, L., & Takahashi, M. (2012). Monospore-based asexual life cycle in *Porphyra yezoensis*. *Porphyra yezoensis: Frontiers in Physiological and Molecular Biological Research*, 15-37.

Milstein, D., & de Oliveira, M. C. (2005). Molecular phylogeny of Bangiales (Rhodophyta) based on small subunit rDNA sequencing: emphasis on Brazilian *Porphyra* species. *Phycologia*, *44*(2), 212-221.

Mitman, G. G., & van der Meer, J. P. (1994). Meiosis, blade development, and sex determination in *Porphyra purpurea* (Rhodophyta) 1. *Journal of phycology*, *30*(1), 147-159.

Mitterboeck, T. F., Chen, A. Y., Zaheer, O. A., Ma, E. Y., & Adamowicz, S. J. (2016). Do saline taxa evolve faster? Comparing relative rates of molecular evolution between freshwater and marine eukaryotes. Evolution, 70(9), 1960-1978.

Miyata, T., & Yasunaga, T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. Journal of molecular evolution, 16(1), 23-36.

Modrich, P., & Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annual Review of Biochemistry*, *65*(1), 101-133.

Moenne, A., González, A., & Sáez, C. A. (2016). Mechanisms of metal tolerance in marine macroalgae, with emphasis on copper tolerance in Chlorophyta and Rhodophyta. *Aquatic Toxicology*, *176*, 30-37.

Mohanta, T. K., Khan, A. L., Hashem, A., Allah, E. F. A., Yadav, D., & Al-Harrasi, A. (2019). Genomic and evolutionary aspects of chloroplast tRNA in monocot plants. *BMC Plant Biology, 19*(1), 39.

Mohanta, T. K., Khan, A., Khan, A., Abd_Allah, E. F., & Al-Harrasi, A. (2019). Gene Loss and Evolution of the Plastome. B*ioRxiv*, 676304.

Mohanta, T. K., Mishra, A. K., Khan, A., Hashem, A., Abd_Allah, E. F., & Al-Harrasi, A. (2020). Gene loss and evolution of the plastome. *Genes*, 11(10), 1133.

Moin, M., Bakshi, A., Saha, A., Dutta, M., Madhav, S. M., & Kirti, P. B. (2016). Rice ribosomal protein large subunit genes and their spatio-temporal and stress regulation. *Frontiers in Plant Science*, *7*, 1284.

Molnar, M., Bähler, J., Sipiczki, M., & Kohli, J. (1995). The *rec*8 gene of *Schizosaccharomyces pombe* is involved in linear element formation, chromosome pairing and sister-chromatid cohesion during meiosis. *Genetics*, *141*(1), 61-73.

Monod, C., Takahashi, Y., Goldschmidt-Clermont, M., & Rochaix, J. D. (1994). The chloroplast *ycf*8 open reading frame encodes a photosystem II polypeptide which maintains photosynthetic activity under adverse growth conditions. *The EMBO Journal*, 13(12), 2747-2754.

Moreno, J., & Spreitzer, R. J. (1999). C172S substitution in the chloroplast-encoded large subunit affects stability and stress-induced turnover of ribulose-1, 5-bisphosphate carboxylase/oxygenase. Journal of Biological Chemistry, 274(38), 26789-26793.

Mouly, A., Razafimandimbison, S. G., Florence, J., Jérémie, J., & Bremer, B. (2009). Paraphyly of Ixora and New Tribal Delimitation of Ixoreae (Rubiaceae): Inference from Combined Chloroplast (*rps*16, *rbc*L, and *trn*T-F) Sequence Data1. *Annals of the Missouri Botanical Garden*, *96*(1), 146-160.

Moury, B. (2004). Differential selection of genes of Cucumber mosaic virus subgroups. *Molecular Biology and Evolution*, *21*(8), 1602-1611.

Mukai, L. S., Craigie, J. S., & Brown, R. G. (1981). Chemical composition and structure of the cell walls of the conchocelis and thallus phases of *Porphyra tenera* (Rhodophyceae) 1, 2. *Journal of Phycology*, *17*(2), 192-198.

Müller, K. M., Cannone, J. J., & Sheath, R. G. (2005). A molecular phylogenetic analysis of the Bangiales (Rhodophyta) and description of a new genus and species, *Pseudobangia kaycoleia*. *Phycologia*, *44*(2), 146-155.

Müller, K. M., Cannone, J. J., Gutell, R. R., & Sheath, R. G. (2001a). A structural and phylogenetic analysis of the group IC1 introns in the order Bangiales (Rhodophyta). *Molecular Biology and Evolution*, *18*(9), 1654-1667.

Müller, K. M., Cole, K. M., & Sheath, R. G. (2003). Systematics of *Bangia* (Bangiales, Rhodophyta) in North America. II. Biogeographical trends in karyology: Chromosome numbers and linkage with gene sequence phylogenetic trees. *Phycologia, 42*(3), 209-219.

Müller, K. M., Lynch, M. D., & Sheath, R. G. (2010). Bangiophytes: from one class to six; where do we go from here?. In *Red Algae in the Genomic Age* (pp. 241-259). Springer, Dordrecht.

Müller, K. M., Oliveira, M. C., Sheath, R. G., & Bhattacharya, D. (2001b). Ribosomal DNA phylogeny of the Bangiophycidae (Rhodophyta) and the origin of secondary plastids. *American Journal of Botany*, *88*(8), 1390-1400

Müller, K. M., Sheath, R. G., Vis, M. L., Crease, T. J., & Cole, K. M. (1998). Biogeography and systematics of *Bangia* (Bangiales, Rhodophyta) based on the rubisco spacer, *rbc*L gene and 18S rRNA gene sequences and morphometric analyses. 1. North America. *Phycologia*, 37(3), 195-207.

Munekage, Y., Hashimoto, M., Miyake, C., Tomizawa, K. I., Endo, T., Tasaka, M., & Shikanai, T. (2004). Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature*, *429*(6991), 579-582.

Muñoz-Gómez, S. A., Mejía-Franco, F. G., Durnin, K., Colp, M., Grisdale, C. J., Archibald, J. M., & Slamovits, C. H. (2017). The new red algal subphylum Proteorhodophytina comprises the largest and most divergent plastid genomes known. *Current Biology*, *27*(11), 1677-1684.

Muto, A., & Ushida, C. (2002). Transcription and translation. In *Molecular Biology and Pathogenicity of Mycoplasmas* (pp. 323-345). Springer, Boston, MA.Nadot, S., Bajon, R., & Lejeune, B. (1994). The chloroplast gene *rps*4 as a tool for the study of Poaceae phylogeny. *Plant Systematics and Evolution*, *191*(1), 27-38.

Naithani, S., Hou, J. M., & Chitnis, P. R. (2000). Targeted inactivation of the *psa*K1, *psa*K2 and *psa*M genes encoding subunits of Photosystem I in the cyanobacterium *Synechocystis* sp. PCC 6803. *Photosynthesis Research*, *63*(3), 225-236.

Nakamura, Y., Sasaki, N., Kobayashi, M., Ojima, N., Yasuike, M., Shigenobu, Y., ... & Ikeo, K. (2013). The first symbiont-free genome sequence of marine red alga, Susabi-nori (Pyropia yezoensis). *PloS one*, *8*(3), e57122.

Nam-Gil, K. (1999). Culture studdies of *Porphyra dentata* and *P. pseudolinearis* (Bangiales, Rhodophyta), two dioecious species from Korea. In *Sixteenth International Seaweed Symposium* (pp. 127-135). Springer, Dordrecht.

Nan, F. R., Feng, J., Lv, J. P., Liu, Q., Liu, X. D., Gao, F., & Xie, S. L. (2020). Comparison of the transcriptomes of different life history stages of the freshwater Rhodophyte *Thorea hispida*. *Genomics*, *112*(6), 3978-3990.

Nan, F., Feng, J., Lv, J., Liu, Q., Fang, K., Gong, C., & Xie, S. (2017). Origin and evolutionary history of freshwater Rhodophyta: further insights based on phylogenomic evidence. Scientific Reports, 7(1), 1-12.

Narum, S. R. (2006). Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation genetics*, *7*(5), 783-787.

Nasmyth, K., Peters, J. M., & Uhlmann, F. (2000). Splitting the chromosome: cutting the ties that bind sister chromatids. *Science*, *288*(5470), 1379-1384.

Neefus, C. D., Mathieson, A. C., Bray, T. L., & Yarish, C. (2008). The distribution, morphology, and ecology of three introduced Asiatic species of *Porphyra* (*Bangia*les, Rhodophyta) in the Northwestern Atlantic 1. *Journal of Phycology*, *44*(6), 1399-1414.

Neefus, C. D., Mathieson, A. C., Yarish, C., Klein, A., West, A., Teasdale, B., & Hehre, E. J. (2000). Five cryptic species of *Porphyra* from the Northwest Atlantic. *Journal of Phycology*, *36*, 73-73.

Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics*. Oxford university press.

Nei, M., Rogozin, I. B., & Piontkivska, H. (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proceedings of the National Academy of Sciences*, *97*(20), 10866-10871.

Nei, M., Suzuki, Y., & Nozawa, M. (2010). The neutral theory of molecular evolution in the genomic era. *Annual Review of Genomics and Human Genetics*, *11*, 265-289.

Nelson, W. A., Brodie, J., & Guiry, M. D. (1999). Terminology used to describe reproduction and life history stages in the genus *Porphyra* (Bangiales, Rhodophyta). *Journal of Applied Phycology*, *11*(5), 407-410.

Nelson, W. A., Broom, J. E., & Farr, T. J. (2001). Four new species of *Porphyra* (Bangiales, Rhodophyta) from the New Zealand region described using traditional characters and 18S rDNA sequence data. *Cryptogamie Algologie*, *22*(3), 263-284.

Nelson, W. A., Farr, T. J., & Broom, J. E. (2006). Phylogenetic relationships and generic concepts in the red order Bangiales: Challenges ahead. *Phycologia, 45*(3), 249-259.

Nelson, W., Brodie, J., & Guiry, M. (1999). Terminology used to describe reproduction and life history stages in the genus *Porphyra* (Bangiales, Rhodophyta). *Journal of Applied Phycology, 11*(5), 407-410.

Neufeld, S., Zinchenko, V., Stephan, D. P., Bader, K. P., & Pistorius, E. K. (2004). On the functional significance of the polypeptide *psb*Y for photosynthetic water oxidation in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Molecular Genetics and Genomics*, *271*(4), 458-467.

Neumann-Spallart, C., Jakowitsch, J., Kraus, M., Brandtner, M., Bohnert, H. J., & Löffelhardt, W. (1991). *rps*10, unreported for plastid DNAs, is located on the cyanelle genome of Cyanophora paradoxa and is contranscribed with the str operon genes. *Current Genetics*, *19*(4), 313-315.

Nickelsen, J., Fleischmann, M., Boudreau, E., Rahire, M., & Rochaix, J. D. (1999). Identification of cis-acting RNA leader elements required for chloroplast *psb*D gene expression in *Chlamydomonas*. *The Plant Cell*, 11(5), 957-970.

Nielsen, R., & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 148(3), 929-936.

Nishiyama, T., & Kato, M. (1999). Molecular phylogenetic analysis among bryophytes and tracheophytes based on combined data of plastid coded genes and the 18S rRNA gene. *Molecular Biology and Evolution*, *16*(8), 1027-1036.

Notoya, M., & Iijima, N. (2003). Life history and sexuality of archeospore and apogamy of *Bangia atropurpurea* (Roth) Lyngbye (*Bangia*les, Rhodophyta) from Fukaura and Enoshima, Japan. *Fisheries Science*, *69*(4), 799-805.

Nowick, K., Walter Costa, M. B., Höner zu Siederdissen, C., & Stadler, P. F. (2019). Selection pressures on RNA sequences and structures. Evolutionary Bioinformatics, 15, 1176934319871919.

Nozaki, H., Onishi, K., & Morita, E. (2002). Differences in pyrenoid morphology are correlated with differences in the rbcL genes of members of the Chloromonas lineage (Volvocales, Chlorophyceae). Journal of Molecular Evolution, 55(4), 414-430.

O'Connor, H. E., Ruffle, S. V., Cain, A. J., Deak, Z., Vass, I., Nugent, J. H., & Purton, S. (1998). The 9-kDa phosphoprotein of photosystem: II. Generation and characterisation of *Chlamydomonas* mutants lacking PSII-H and a site-directed mutant lacking the phosphorylation site. *Biochimica et Biophysica Acta* (BBA)-Bioenergetics, 1364(1), 63-72.

O'Connor, M., Gregory, S. T., & Dahlberg, A. E. (2004). Multiple defects in translation associated with altered ribosomal protein L4. *Nucleic Acids Research*, *32*(19), 5750-5756.

Ohme, M., & Miura, A. (1988). Tetrad analysis in conchospore germlings of *Porphyra* yezoensis (Rhodophyta, *Bangia*les). *Plant Science*, *57*(2), 135-140.

Ohta, N., Matsuzaki, M., Misumi, O., Miyagishima, S. Y., Nozaki, H., Tanaka, K., ... & Kuroiwa, T. (2003). Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Research*, *10*(2), 67-77.

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*, *246*(5428), 96-98.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, *23*(1), 263-286.

Oliveira, M. C., & Bhattacharya, D. (2000). Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. *Journal of Phycology*, *36*(s3), 52-52.

Oliveira, M. C., & Ragan, M. A. (1994). Variant forms of a group I intron in nuclear small-subunit rRNA genes of the marine red alga *Porphyra spiralis var. amplifolia*. *Molecular Biology and Evolution*, *11*(2), 195-207.

Oliveira, M. C., Kurniawan, J., Bird, C. J., Rice, E. L., Murphy, C. A., Singh, R. K., ... & Ragan, M. A. (1995). A preliminary investigation of the order Bangiales (Bangiophycidae, Rhodophyta) based on sequences of nuclear small-subunit ribosomal RNA genes. *Phycological Research*, *43*(2), 71-79.

Oliver, M.J., Murdock, A.G., Mishler, B.D., Kuehl, J.V., Boore, J.L., Mandoli, D.F., Everett, K.D., Wolf, P.G., Duffy, A.M. and Karol, K.G. (2010). Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. *BMC genomics*, *11*(1), 143.

Orgil, O., Matityahu, A., Eng, T., Guacci, V., Koshland, D., & Onn, I. (2015). A conserved domain in the *scc*3 subunit of cohesin mediates the interaction with both mcd1 and the cohesin loader complex. *PLoS Genetics*, *11*(3), e1005036.

Ott, F. D., & Seckbach, J. (1994). New classification for the genus *Cyanidium* Geitler 1933. In *Evolutionary pathways and enigmatic algae: Cyanidium caldarium (Rhodophyta) and related cells* (pp. 145-152). Springer, Dordrecht.

Pakrasi, H. B., Williams, J. G., & Arntzen, C. J. (1988). Targeted mutagenesis of the *psb*E and *psb*F genes blocks photosynthetic electron transport: evidence for a functional role of cytochrome b559 in photosystem II. *The EMBO Journal*, 7(2), 325-332.

Palmer, J. D. (2000). Molecular evolution: A single birth of all plastids? *Nature*, *405*(6782), 32-33.

Palmer, J. D. (2003). The symbiotic birth and spread of plastids: How many times and whodunit? *Journal of Phycology*, *39*(1), 4-12.

Palmer, J. D., Jansen, R. K., Michaels, H. J., Chase, M. W., & Manhart, J. R. (1988). Chloroplast DNA variation and plant phylogeny. *Annals of the Missouri Botanical Garden*, 1180-1206.

Pannafino, G., & Alani, E. (2021). Coordinated and Independent Roles for *mlh* Subunits in DNA Repair. *Cells*, *10*(4), 948.

Panoli, A. P., Ravi, M., Sebastian, J., Nishal, B., Reddy, T. V., Marimuthu, M. P., ... & Siddiqi, I. (2006). At *mnd*1 is required for homologous pairing during meiosis in *Arabidopsis*. *BMC Molecular Biology*, *7*(1), 1-12.

Paquis-Flucklinger, V., Santucci-Darmanin, S., Paul, R., Saunieres, A., Turc-Carel, C., & Desnuelle, C. (1997). Cloning and expression analysis of a meiosis-specific *mut*S homolog: the HumanMSH4Gene. *Genomics*, *44*(2), 188-194.

Parfrey, L. W., Lahr, D. J., Knoll, A. H., & Katz, L. A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proceedings of the National Academy of Sciences*, *108*(33), 13624-13629.

Park, J. M., Manen, J. F., Colwell, A. E., & Schneeweiss, G. M. (2008). A plastid gene phylogeny of the non-photosynthetic parasitic *Orobanche* (Orobanchaceae) and related genera. *Journal of Plant Research*, *121*(4), 365-376.

Park, M., Park, H., Lee, H., Lee, B. H., & Lee, J. (2018). The complete plastome sequence of an Antarctic bryophyte *Sanionia uncinata* (Hedw.) Loeske. *International Journal of Molecular Sciences*, *19*(3), 709.

Park, S., Ruhlman, T. A., Weng, M. L., Hajrah, N. H., Sabir, J. S., & Jansen, R. K. (2017). Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. *Genome Biology and Evolution*, *9*(6), 1766-1780.

Parks, S. L., & Goldman, N. (2014). Maximum likelihood inference of small trees in the presence of long branches. *Systematic Biology*, *63*(5), 798-811.

Pasierbek, P., Födermayr, M., Jantsch, V., Jantsch, M., Schweizer, D., & Loidl, J. (2003). The *Caenorhabditis elegans scc*3 homologue is required for meiotic synapsis and for proper chromosome disjunction in mitosis and meiosis. *Experimental Cell Research*, *289*(2), 245-255.

Pathania, A., Liu, W., Matityahu, A., Irudayaraj, J., & Onn, I. (2021). Chromosome loading of cohesin depends on conserved residues in *scc*3. *Current Genetics*, *67*(3), 447-459.

Pearson, G., Kautsky, L., & Serrão, E. (2000). Recent evolution in Baltic *Fucus vesiculosus*: reduced tolerance to emersion stresses compared to intertidal (North Sea) populations. *Marine Ecology Progress Series*, *202*, 67-79.

Pereira, R., Sousa-Pinto, I., & Yarish, C. (2004). Field and culture studies of the life history of *Porphyra dioica* (Bangiales, Rhodophyta) from Portugal. *Phycologia*, *43*(6), 756-767.

Petukhova, G. V., Pezza, R. J., Vanevski, F., Ploquin, M., Masson, J. Y., & Camerini-Otero, R. D. (2005). The *hop*2 and *mnd*1 proteins act in concert with *rad*51 and *dmc*1 in meiotic recombination. *Nature Structural & Molecular Biology*, *12*(5), 449-453.

Pezza, R. J., Petukhova, G. V., Ghirlando, R., & Camerini-Otero, R. D. (2006). Molecular activities of meiosis-specific proteins *hop*2, *mnd*1, and the *hop*2-*mnd*1 complex. *Journal of Biological Chemistry*, *281*(27), 18426-18434.

Pezza, R. J., Voloshin, O. N., Vanevski, F., & Camerini-Otero, R. D. (2007). *hop*2/*mnd*1 acts on two critical steps in *dmc*1-promoted homologous pairing. *Genes & Development*, *21*(14), 1758-1766.

Pezzi, N., Prieto, I., Kremer, L., Jurado, L. A. P., Valero, C., Del Mazo, J., ... & Barbero, J. L. (2000). STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion. *The FASEB Journal*, *14*(3), 581-592.

Pinheiro, F. L., Santos, F. C., & Pacheco, J. M. (2012). How selection pressure changes the nature of social dilemmas in structured populations. *New Journal of Physics*, *14*(7), 073035.

Pinto, G., Ciniglia, C., Cascone, C. & Pollio, A. (2007). Species composition of Cyanidiales assemblages in Pisciarelli (Campi Flegrei, Italy) and description of Galdieria phlegrea sp. NOV. In: Seckbach, J. (eds) Algae and Cyanobacteria in Extreme Environments. Cellular Origin, Life in Extreme Habitats and Astrobiology, vol. 11, p. 487–502. Springer, Dordrecht

Plöchinger, M., Schwenkert, S., von Sydow, L., Schröder, W. P., & Meurer, J. (2016). Functional update of the auxiliary proteins *psb*W, *psb*Y, HCF136, *psb*N, *ter*C and ALB3 in maintenance and assembly of PSII. *Frontiers in Plant Science*, *7*, 423.

Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., ... & Haussler, D. (2006). Forces shaping the fastest evolving regions in the human genome. PLoS Genet, 2(10), e168.

Polne-Fuller, M. (1988). The past, present and future of tissue culture and biotechnology of seaweeds. *Algal Biotechnology,*111-118.

Pomeroy, L. R. (1959). Algal Productivity in Salt Marshes of Georgia 1. *Limnology and Oceanography*, *4*(4), 386-397.

Ponce-Toledo, R. I., Deschamps, P., López-García, P., Zivanovic, Y., Benzerara, K., & Moreira, D. (2017). An early-branching freshwater cyanobacterium at the origin of plastids. *Current Biology*, *27*(3), 386-391.

Porter, S. M. (2006). The Proterozoic fossil record of heterotrophic eukaryotes. In *Neoproterozoic geobiology and paleobiology* (pp. 1-21). Springer, Dordrecht.

Posada, D., & Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics (Oxford, England)*, *14*(9), 817-818.

Preuss, M., Verbruggen, H., & Zuccarello, G. C. (2020). The organelle genomes in the photosynthetic red algal parasite *Pterocladiophila hemisphaerica* (Florideophyceae, Rhodophyta) have elevated substitution rates and extreme gene loss in the plastid genome. *Journal of Phycology*, *56*(4), 1006-1018.

Prieto, I., Suja, J. A., Pezzi, N., Kremer, L., Martínez-A, C., Rufas, J. S., & Barbero, J. L. (2001). Mammalian STAG3 is a cohesin specific to sister chromatid arms in meiosis I. *Nature cell biology*, *3*(8), 761-766.

Provan, J., Murphy, S., & Maggs, C. A. (2004). Universal plastid primers for Chlorophyta and Rhodophyta. *European Journal of Phycology*, *39*(1), 43-50.

Przytycka, T. M., Jothi, R., Aravind, L., & Lipman, D. J. (2008). Differences in evolutionary pressure acting within highly conserved ortholog groups. BMC Evolutionary Biology, 8(1), 1-12.

Ptashne, M., & Gann, A. (1997). Transcriptional activation by recruitment. *Nature*, *386*(6625), 569-577.

Pueschel, C. M., & Cole, K. M. (1982). Rhodophycean pit plugs: an ultrastructural survey with taxonomic implications. *American Journal of Botany*, *69*(5), 703-720.

Puthiyaveetil, S., & Allen, J. F. (2009). Chloroplast two-component systems: evolution of the link between photosynthesis and gene expression. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1665), 2133-2145.

Qiu, H., Price, D. C., Yang, E. C., Yoon, H. S., & Bhattacharya, D. (2015). Evidence of ancient genome reduction in red algae (Rhodophyta). *Journal of Phycology*, *51*(4), 624-636.

Qiu, H., Yoon, H. S., & Bhattacharya, D. (2013). Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Frontiers in plant science*, *4*, 366.

Ragan, M. A., Bird, C. J., RicE, E. L., Gutell, R. R., Murphy, C. A., & Singh, R. K. (1994). A molecular phylogeny of the marine red algae (Rhodophyta) based on the nuclear small-subunit rRNA gene. *Proceedings of the National Academy of Sciences*, *91*(15), 7276-7280.

Ramesh, M. A., Malik, S. B., & Logsdon Jr, J. M. (2005). A phylogenomic inventory of meiotic genes: evidence for sex in *Giardia* and an early eukaryotic origin of meiosis. *Current biology*, *15*(2), 185-191.

Reed, R. H. (1980). On the conspecificity of marine and freshwater *Bangia* in Britain. *British Phycological Journal, 15*(4), 411-416.

Reenan, R. A., & Kolodner, R. D. (1992). Characterization of insertion mutations in the *Saccharomyces cerevisiae msh*1 and *msh*2 genes: evidence for separate mitochondrial and nuclear functions. *Genetics*, *132*(4), 975-985.

Rehman, S., Chaudhary, H. J., Rasheed, A., & Mahmood, T. (2015). Phylogenetic relationship of selected Pakistani wheat varieties based on a chloroplast *rps*11 Gene. *J. Ani. Plant Sci*, *25*(2), 442-447.

Reith, M., & Munholland, J. (1995). Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Plant Molecular Biology Reporter*, *13*(4), 333-335.

Remeseiro, S., Cuadrado, A., Carretero, M., Martínez, P., Drosopoulos, W. C., Cañamero, M., ... & Losada, A. (2012). Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. *The EMBO Journal*, *31*(9), 2076-2089.

Ren, Y., Yu, M., Low, W. Y., Ruhlman, T. A., Hajrah, N. H., El Omri, A., ... & Rather, I. A. (2020). Nucleotide substitution rates of diatom plastid encoded protein genes are positively correlated with genome architecture. *Scientific Reports*, *10*(1), 1-11.

Revenkova, E. M. E. C. H. B. G., Eijpe, M., Heyting, C., Gross, B., & Jessberger, R. (2001). Novel meiosis-specific isoform of mammalian *smc*1. *Molecular and Cellular Biology*, *21*(20), 6984-6998.

Reyes, G. X., Zhao, B., Schmidt, T. T., Gries, K., Kloor, M., & Hombauer, H. (2020). Identification of *mlh*2/h*pms*1 dominant mutations that prevent DNA mismatch repair function. *Communications biology*, *3*(1), 1-14.

Reyes, R., Vázquez, D., & Ballesta, J. P. (1977). Peptidyl Transferase Center of Rat-Liver Ribosome Cores. *European Journal of Biochemistry*, *73*(1), 25-31.

Reyes-Prieto, A., Russell, S., Figueroa-Martinez, F., & Jackson, C. (2018). Comparative plastid genomics of Glaucophytes. In *Advances in Botanical Research* (Vol. 85, pp. 95-127). Academic Press.

Reyes-Prieto, A., Weber, A. P., & Bhattacharya, D. (2007). The origin and establishment of the plastid in algae and plants. *Annu. Rev. Genet.*, *41*, 147-168.

Robertson, D., Boynton, J. E., & Gillham, N. W. (1990). Cotranscription of the wild-type chloroplast *atp*E gene encoding the CF 1/CF 0 epsilon subunit with the 3′ half of the rps7 gene in *Chlamydomonas reinhardtii* and characterization of frameshift mutations in *atp*E. *Molecular and General Genetics* MGG, 221(2), 155-163.

Rochaix, J. D., Kuchka, M., Mayfield, S., Schirmer-Rahire, M., Girard-Bascou, J., & Bennoun, P. (1989). Nuclear and chloroplast mutations affect the synthesis or stability of the chloroplast *psb*C gene product in *Chlamydomonas reinhardtii*. *The EMBO Journal*, 8(4), 1013-1021.

Rockwell, N. C., Lagarias, J. C., & Bhattacharya, D. (2014). Primary endosymbiosis and the evolution of light and oxygen sensing in photosynthetic eukaryotes. *Frontiers in Ecology and Evolution*, *2*, 66.

Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S. C., Roure, B., Burger, G., Löffelhardt, W., ... & Lang, B. F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Current Biology*, *15*(14), 1325-1330.

Roeder, G. S. (1997). Meiotic chromosomes: it takes two to tango. *Genes & Development*, *11*(20), 2600-2621.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends in biochemical sciences*, *21*(9), 327-335.

Rogalski, M., Schöttler, M. A., Thiele, W., Schulze, W. X., & Bock, R. (2008). *rpl*33, a nonessential plastid-encoded ribosomal protein in tobacco, is required under cold stress conditions. *The Plant Cell*, *20*(8), 2221-2237.

Roig, M. B., Löwe, J., Chan, K. L., Beckouët, F., Metson, J., & Nasmyth, K. (2014). Structure and function of cohesin's *scc*3/SA regulatory subunit. *FEBS Letters*, *588*(20), 3692-3702.

Rollins, R. A., Korom, M., Aulner, N., Martens, A., & Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/*scc*3 cohesion factor to facilitate long-range activation of the cut gene. *Molecular and Cellular Biology*, *24*(8), 3100-3111.

Rollins, R. A., Morcillo, P., & Dorsett, D. (1999). Nipped-B, a *Drosophila* homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics*, *152*(2), 577-593.

Romani, I., Tadini, L., Rossi, F., Masiero, S., Pribil, M., Jahns, P., ... & Pesaresi, P. (2012). Versatile roles of *Arabidopsis* plastid ribosomal proteins in plant growth and development. *The Plant Journal*, *72*(6), 922-934.

Romanienko, P. J., & Camerini-Otero, R. D. (2000). The mouse *spo*11 gene is required for meiotic chromosome synapsis. *Molecular Cell*, *6*(5), 975-987.

Rosado, A., Sohn, E. J., Drakakaki, G., Pan, S., Swidergal, A., Xiong, Y., ... & Raikhel, N. V. (2010). Auxin-mediated ribosomal biogenesis regulates vacuolar trafficking in *Arabidopsis*. *The Plant Cell*, *22*(1), 143-158.

Roth, A. W. (1806). *Catalecta botanica, quibus plantae novae et minus eognitae describuntur atque illustrantur*

Roth, A., Anisimova, M., & Cannarozzi, G. M. (2012). Measuring codon usage bias. *Codon Evolution Mechanisms and Models*, 189-217.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution, 34(12), 3299-3302.

Russell, G. (1985). Recent evolutionary changes in the algae of the Baltic Sea. *British Phycological Journal*, *20*(2), 87-104.

Saeed, S., Munir, F., Naveed, I., Raja, G. K., & Mahmood, T. (2011). Phylogenetics of selected Plantago species on the basis of *rps*14 chloroplast gene. *Journal of Medicinal Plants Research*, *5*(19), 4888-4891.

Sagert, S., & Schubert, H. (1995). Acclimation of the Photosynthetic Apparatus of *Palmaria Palmata* (Rhodophyta) To Light Qualities That Preferentially Excite Photosystem I or Photosystem Ii1. *Journal of Phycology*, *31*(4), 547-554.

Saha, A., Das, S., Moin, M., Dutta, M., Bakshi, A., Madhav, M. S., & Kirti, P. B. (2017). Genome-wide identification and comprehensive expression profiling of Ribosomal Protein Small Subunit (*rps*) genes and their comparative analysis with the Large Subunit (*rpl*) genes in rice. *Frontiers in Plant Science*, *8*, 1553.

Saha, M., Wiese, J., Weinberger, F., & Wahl, M. (2016). Rapid adaptation to controlling new microbial epibionts in the invaded range promotes invasiveness of an exotic seaweed. *Journal of Ecology, 104*(4), 969-978.

Saha, M., Wiese, J., Weinberger, F., & Wahl, M. (2016). Rapid adaptation to controlling new microbial epibionts in the invaded range promotes invasiveness of an exotic seaweed. *Journal of Ecology*, *104*(4), 969-978.

Sánchez, N., Vergés, A., Peteiro, C., Sutherland, J. E., & Brodie, J. (2014). Diversity of bladed Bangiales (Rhodophyta) in western Mediterranean: recognition of the genus *Themis* and descriptions of *T. ballesterosii* sp. nov., *T. iberica* sp. nov., and *Pyropia parva* sp. nov. *Journal of Phycology*, *50*(5), 908-929.

Sánchez-Baracaldo, P., Raven, J. A., Pisani, D., & Knoll, A. H. (2017). Early photosynthetic eukaryotes inhabited low-salinity habitats. Proceedings of the National Academy of Sciences, 114(37), E7737-E7745.

Sanderson, M. J., Wojciechowski, M. F., Hu, J. M., Khan, T. S., & Brady, S. G. (2000). Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution*, *17*(5), 782-797.

Saunders, G. W. (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1462), 1879-1888.

Saunders, G. W., & Hommersand, M. H. (2004). Assessing red algal supraordinal diversity and taxonomy in the context of contemporary systematic data. *American Journal of Botany*, *91*(10), 1494-1507.

Saunders, G. W., & Kraft, G. T. (1997). A molecular perspective on red algal evolution: focus on the Florideophycidae. *Origins of Algae and their Plastids*, 115-138.

Saunders, G. W., & Moore, T. E. (2013). Refinements for the amplification and sequencing of red algal DNA barcode and RedToL phylogenetic markers: a summary of current primers, profiles and strategies. *Algae*, *28*(1), 31-43.

Savolainen, V., Chase, M. W., Hoot, S. B., Morton, C. M., Soltis, D. E., Bayer, C., ... & Qiu, Y. L. (2000). Phylogenetics of flowering plants based on combined analysis of plastid atpB and rbcL gene sequences. Systematic Biology, 49(2), 306-362.

Schaaf, C. A., Misulovin, Z., Sahota, G., Siddiqui, A. M., Schwartz, Y. B., Kahn, T. G., ... & Dorsett, D. (2009). Regulation of the *Drosophila* Enhancer of split and invected-engrailed gene complexes by sister chromatid cohesion proteins. *PloS One*, *4*(7), e6202.

Schäferhoff, B., Fleischmann, A., Fischer, E., Albach, D. C., Borsch, T., Heubl, G., & Müller, K. F. (2010). Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. *BMC Evolutionary Biology*, *10*(1), 1-22.

Scheller, H. V., Jensen, P. E., Haldrup, A., Lunde, C., & Knoetzel, J. (2001). Role of subunits in eukaryotic Photosystem I. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, *1507*(1-3), 41-60.

Schluchter, W. M., Shen, G., Zhao, J., & Bryant, D. A. (1996). Characterization of psal and psaL mutants of Synechococcus sp. strain PCC 7002: a new model for state transitions in cyanobacteria. Photochemistry and Photobiology, 64(1), 53-66.

Schmiesing, J. A., Gregson, H. C., Zhou, S., & Yokomori, K. (2000). A human condensin complex containing hCAP-C–hCAP-E and CNAP1, a homolog of Xenopus XCAP-D2, colocalizes with phosphorylated histone H3 during the early stage of mitotic chromosome condensation. *Molecular and Cellular Biology*, *20*(18), 6996-7006.

Schönknecht, G., Chen, W. H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., ... & Weber, A. P. (2013). Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science*, *339*(6124), 1207-1210.

Schubert, V. (2009). SMC proteins and their multiple functions in higher plants. *Cytogenetic and Genome Research*, *124*(3-4), 202-214.

Schurko, A. M., & Logsdon Jr, J. M. (2008). Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. *Bioessays*, *30*(6), 579-589.

Schwarz, E. N. *et al.* Plastome-wide nucleotide substitution rates reveal accelerated rates in Papilionoideae and correlations with genome features across legume subfamilies. *J. Mol. Evol.* 84, 187–203 (2017).

Schwenkert, S., Legen, J., Takami, T., Shikanai, T., Herrmann, R. G., & Meurer, J. (2007). Role of the low-molecular-weight subunits *pet*L, *pet*G, and *pet*N in assembly, stability, and

dimerization of the cytochrome b6f complex in tobacco. *Plant Physiology*, *144*(4), 1924-1935.

Seckbach, J. (1991). Systematic problems with *Cyanidium caldarium* and *Galdieria sulphuraria* and their implications for molecular biology studies 1. *Journal of Phycology*, *27*(6), 794-796.

Seo, T. K., & Kishino, H. (2008). Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. Systematic biology, 57(3), 367-377.

Shapiro, B., Rambaut, A., & Drummond, A. J. (2006). Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Molecular Biology and Evolution, 23(1), 7-9.

Sharma, M. R., Wilson, D. N., Datta, P. P., Barat, C., Schluenzen, F., Fucini, P., & Agrawal, R. K. (2007). Cryo-EM study of the spinach chloroplast ribosome reveals the structural and functional roles of plastid-specific ribosomal proteins. *Proceedings of the National Academy of Sciences*, *104*(49), 19315-19320.

Shea, T. B., Sheath, R. G., Chhun, A., Vis, M. L., Chiasson, W. B., & Müller, K. M. (2014). Distribution, seasonality and putative origin of the non-native red alga *Bangia atropurpurea* (Bangiales, Rhodophyta) in the Laurentian Great Lakes. *Journal of Great Lakes Research, 40*(1), 27-34.

Sheath R.G. & Cole K.M. (1980). Distribution and salinity adaptations of *Bangia atropurpurea* (Rhodophyta), a putative migrant into the Laurentian Great Lakes*). Journal of Phycology* 16: 412-420.

Sheath, R. (1987). Invasions into the Laurentian Great Lakes by marine algae. *Archiv Hydrobiologie Beih*, *25*, 165-186.

Sheath, R. G. & Vis, M. L. (2015). Red algae. In Wehr, J. D., Sheath, R. G., & Kociolek, J. P. (Eds.). *Freshwater algae of North America: ecology and classification*. Elsevier.

Sheath, R. G. & Hambrook, J. A. Freshwater ecology. In Biology of the Red Algae (eds Cole, K. M. et al.), 423–453 (Cambridge: Cambridge University Press, 1990).

Sheath, R. G. (1984). The biology of freshwater red algae. Progress in phycological research, 3, 89-157.

Sheath, R. G. (1987). Invasions into the Laurentian Great Lakes by marine algae. *Ergebnisse der Limnologie/Advances in Limnology (Stuttgart)*, *25*, 165-186.

Sheath, R. G., & Cole, K. M. (1980). Distribution and salinity adaptations of *Bangia atropurpurea* (Rhodophyta), a putative migrant into the Laurantian Great Lakes 1. *Journal of Phycology*, *16*(3), 412-420.

Sheath, R. G., & Cole, K. M. (1984). Systematics of *Bangia* (Rhodophyta) in North America. I. Biogeographic trends in morphology. *Phycologia, 23*(3), 383-396.

Shikanai, T., Shimizu, K., Ueda, K., Nishimura, Y., Kuroiwa, T., & Hashimoto, T. (2001). The chloroplast *clp*P gene, encoding a proteolytic subunit of ATP-dependent protease, is indispensable for chloroplast development in tobacco. *Plant and Cell Physiology*, *42*(3), 264-273.

Shim, E., Zuccarello, G. C., & Kim, G. H. (2021). Sex-Specific Genes and their Expression in the Life History of the Red Alga *Bostrychia moritziana* (Ceramiales, Rhodomelaceae). *Journal of Phycology*, *57*(2), 528-540.

Shimda, H., & Sugiuro, M. (1991). Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. *Nucleic acids research*, *19*(5), 983-995.

Shimizu, A., Morishima, K., Kobayashi, M., Kunimoto, M., & Nakayama, I. (2008). Identification of *Porphyra yezoensis* (Rhodophyta) meiosis by DNA quantification using confocal laser scanning microscopy. *Journal of Applied Phycology*, *20*(1), 83-88.

Shinohara, A., & Shinohara, M. (2004). Roles of RecA homologues *rad*51 and *dmc*1 during meiotic recombination. *Cytogenetic and Genome Research*, *107*(3-4), 201-207.

Shrager, J., Hauser, C., Chang, C. W., Harris, E. H., Davies, J., McDermott, J., ... & Grossman, A. R. (2003). *Chlamydomonas reinhardtii* genome project. A guide to the generation and use of the cDNA information. *Plant Physiology*, *131*(2), 401-408.

Siepel, A., & Haussler, D. (2004). Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. Molecular Biology and Evolution, 21(3), 468-488.

Signor, P. W., & Lipps, J. H. (1992). Origin and early radiation of the Metazoa. In *Origin and early evolution of the Metazoa* (pp. 3-23). Springer, Boston, MA.

Silva, M. Y., & Hughey, J. R. (2016). Complete mitochondrial genome of the holotype specimen of *Wildemania schizophylla* (Bangiales: Rhodophyta). *Mitochondrial DNA Part A*, *27*(2), 1001-1002.

Simanovsky, S. A., & Bogdanov, Y. F. (2018). Genetic control of meiosis in plants. *Russian Journal of Genetics*, *54*(4), 389-402.

Simchen, G., & Hugerat, Y. (1993). What determines whether chromosomes segregate reductionally or equationally in meiosis?. *BioEssays*, *15*(1), 1-8.

Simpson, C. L., & Stern, D. B. (2002). The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. *Plant Physiology*, *129*(3), 957-966.

Singh, R. P., & Reddy, C. (2014). Seaweed–microbial interactions: Key functions of seaweed-associated bacteria. *FEMS Microbiology Ecology, 88*(2), 213-230.

Skuja, H. (1938). Comments on fresh-water rhodophyceae. *The Botanical Review*, *4*(12), 665-676.

Sloan, D. B., Alverson, A. J., Wu, M., Palmer, J. D. & Taylor, D. R. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol. Evol.* 4, 294–306 (2012).

Smart, L. B., & McIntosh, L. (1993). Genetic inactivation of the *psa*B gene in *Synechocystis* sp. PCC 6803 disrupts assembly of photosystem I. *Plant Molecular Biology*, *21*(1), 177-180.

Smith, J. M., & Maynard-Smith, J. (1978). The evolution of sex (Vol. 4). *Cambridge: Cambridge University Press.*

Snowden, T., Acharya, S., Butz, C., Berardini, M., & Fishel, R. (2004). h*msh*4-h*msh*5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. *Molecular Cell*, *15*(3), 437-451.

Song, J., Wei, X., Shao, G., Sheng, Z., Chen, D., Liu, C., ... & Hu, P. (2014). The rice nuclear gene WLP1 encoding a chloroplast ribosome L13 protein is needed for chloroplast

development in rice grown under low temperature conditions. *Plant Molecular Biology*, *84*(3), 301-314.

Song, S., Gao, D., & Yan, X. (2020). Transcriptomic exploration of genes related to the formation of archeospores in *Pyropia yezoensis* (Rhodophyta). *Journal of Applied Phycology*, *32*(5), 3295-3304.

Sormani, R., Masclaux-Daubresse, C., Daniele-Vedele, F., & Chardon, F. (2011). Transcriptional regulation of ribosome components are determined by stress according to cellular compartments in *Arabidopsis thaliana*. *PloS One*, *6*(12), e28070.

Souza-Chies, T. T., Bittar, G., Nadot, S., Carter, L., Besin, E., & Lejeune, B. (1997). Phylogenetic analysis of Iridaceae with parsimony and distance methods using the plastid gene *rps*4. *Plant Systematics and Evolution*, *204*(1), 109-123.

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., ... & Ettema, T. J. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, *521*(7551), 173-179.

Sprink, T., & Hartung, F. (2014). The splicing fate of plant *spo*11 genes. *Frontiers in Plant Science*, *5*, 214.

Stacey, N.J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K., Sugimoto-Shirasu, K. (2006) *Arabidopsis spo*11-2 functions with *spo*11-1 in meiotic recombination. *The Plant Journal*, 48, 206–216

Stelzl, U., Connell, S., Nierhaus, K. H., & Wittmann-Liebold, B. (2001). Ribosomal proteins: role in ribosomal function. *Encyclopedia Life Sci*, *2001*, 1-12.

Stiller, J. W., & Waaland, J. R. (1993). Molecular analysis reveals cryptic diversity in *Porphyra* (Rhodophyta). *Journal of Phycology*, *29*(4), 506-517.

Stirewalt, V. L., Michalowski, C. B., Löffelhardt, W., Bohnert, H. J., & Bryant, D. A. (1995). Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Plant Molecular Biology Reporter*, *13*(4), 327-332.

Streiner, D. L., & Norman, G. R. (2011). Correction for multiple testing: is there a resolution?. *Chest*, *140*(1), 16-18.

Streiner, D. L., & Norman, G. R. (2011). Correction for multiple testing: is there a resolution?. *Chest*, *140*(1), 16-18.

Strunnikov, A. V., Hogan, E., & Koshland, D. (1995). SMC2, a *Saccharomyces cerevisiae* gene essential for chromosome segregation and condensation, defines a subgroup within the SMC family. *Genes & Development*, *9*(5), 587-599.

Sturm, S., Engelken, J., Gruber, A., Vugrinec, S., Kroth, P. G., Adamska, I., & Lavaud, J. (2013). A novel type of light-harvesting antenna protein of red algal origin in algae with secondary plastids. *BMC evolutionary biology*, *13*(1), 1-14.

Suchard, M. A., & Rambaut, A. (2009). Many-core algorithms for statistical phylogenetics. *Bioinformatics*, *25*(11), 1370-1376.

Sugimoto-Shirasu, K., Stacey, N.J., Corsar, J., Roberts, K., McCann, M.C. (2002). DNA Topoisomerase VI Is Essential for Endoreduplication in *Arabidopsis*. *Current Biology*, 12, 1782– 1786

Sugiura, M. (1992). The chloroplast genome. *10 Years plant molecular biology*, 149-168.

Summer, E. J., Schmid, V. H., Bruns, B. U., & Schmidt, G. W. (1997). Requirement for the H phosphoprotein in photosystem II of *Chlamydomonas reinhardtii*. *Plant Physiology*, 113(4), 1359-1368.

Sutani, T., & Yanagida, M. (1997). DNA renaturation activity of the *smc* complex implicated in chromosome condensation. *Nature*, *388*(6644), 798-801.

Sutherland, J. E., Lindstrom, S. C., Nelson, W. A., Brodie, J., Lynch, M. D., Hwang, M. S., ... & Müller, K. M. (2011). A new look at an ancient order: generic revision of the Bangiales (Rhodophyta) 1. *Journal of Phycology*, *47*(5), 1131-1151.

Suzuki, Y., & Gojobori, T. (1999). A method for detecting positive selection at single amino acid sites. Molecular Biology and Evolution, 16(10), 1315-1328.

Tajima, N., Sato, S., Maruyama, F., Kurokawa, K., Ohta, H., Tabata, S., ... & Sato, N. (2014). Analysis of the complete plastid genome of the unicellular red alga Porphyridium purpureum. Journal of Plant Research, 127(3), 389-397.

Takahashi, M., & Mikami, K. (2017). Oxidative stress promotes asexual reproduction and apogamy in the red seaweed *Pyropia yezoensis*. *Frontiers in Plant Science*, *8*, 62.

Takahashi, Y., Goldschmidt-Clermont, M., Soen, S. Y., Franzen, L. G., & Rochaix, J. D. (1991). Directed chloroplast transformation in *Chlamydomonas reinhardtii*: insertional inactivation of the *psa*C gene encoding the iron sulfur protein destabilizes photosystem I. *The EMBO Journal*, 10(8), 2033-2040.

Takahashi, Y., Matsumoto, H., Goldschmidt-Clermont, M., & Rochaix, J. D. (1994). Directed disruption of the *Chlamydomonas* chloroplast *psb*K gene destabilizes the photosystem II reaction center complex. *Plant Molecular Biology*, 24(5), 779-788.

Takemoto, K., Tani, N., Takada-Horisawa, Y., Fujimura, S., Tanno, N., Yamane, M., ... & Ishiguro, K. I. (2020). Meiosis-Specific C19orf57/4930432K21Rik/BRME1 modulates localization of *rad*51 and *dmc*1 to DSBs in mouse meiotic recombination. *Cell Reports*, *31*(8), 107686.

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), 564-577.

Tanaka, T. (1952). The systematic study of the Japanese Protoflorideae. *Memoirs of the Faculty of Fisheries*, Kagoshima University, 2, 1–92

Tang, Q., Pang, K., Yuan, X., & Xiao, S. (2020). A one-billion-year-old multicellular chlorophyte. *Nature Ecology & Evolution*, *4*(4), 543-549.

Tang, X., Wang, Y., Zhang, Y., Huang, S., Liu, Z., Fei, D., & Feng, H. (2018). A missense mutation of plastid *rps*4 is associated with chlorophyll deficiency in Chinese cabbage (*Brassica campestris* ssp. pekinensis). *BMC Plant Biology*, *18*(1), 1-11.

Tappan, H. (1976). Possible eucaryotic algae (Bangiophycidae) among early Proterozoic microfossils. *Geological Society of America Bulletin*, *87*(4), 633-639.

Teasdale, B., West, A., Taylor, H., & Klein, A. (2002). A simple restriction fragment length polymorphism (RFLP) assay to discriminate common *Porphyra* (Bangiophyceae, Rhodophyta) taxa from the Northwest Atlantic. *Journal of Applied Phycology*, *14*(4), 293-298.

Theodorsson-Norheim, E. (1986). Kruskal-Wallis test: BASIC computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples. *Computer Methods and Programs in Biomedicine*, *23*(1), 57-62.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22(22), 4673-4680.

Thompson, J. N. (1999). The raw material for coevolution. *Oikos*, 5-16.

Thompson, S. J., Robinson, C., & Mant, A. (1999). Dual signal peptides mediate the signal recognition particle/Sec-independent insertion of a thylakoid membrane polyprotein, *psb*Y. *Journal of Biological Chemistry*, *274*(7), 4059-4066.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq–versatile and accurate annotation of organelle genomes. *Nucleic acids research*, *45*(W1), W6-W11.

Tomari, N., Ishizuka, Y., Moriya, A., Kojima, S., Deyama, T., Mizukami, H., & Tu, P. (2002). Pharmacognostical studies of Cistanchis herba (III) phylogenetic relationship of the Cistanche plants based on plastid *rps*2 gene and *rpl*16–*rpl*14 intergenic spacer sequences. *Biological and Pharmaceutical Bulletin*, *25*(2), 218-222.

Tomonaga, T., Nagao, K., Kawasaki, Y., Furuya, K., Murakami, A., Morishita, J., ... & Yanagida, M. (2000). Characterization of fission yeast cohesin: essential anaphase proteolysis of *rad*21 phosphorylated in the S phase. *Genes & Development*, *14*(21), 2757-2770.

Torabi, S., Umate, P., Manavski, N., Plöchinger, M., Kleinknecht, L., Bogireddi, H., ... & Meurer, J. (2014). *psb*N is required for assembly of the photosystem II reaction center in *Nicotiana tabacum*. *The Plant Cell*, 26(3), 1183-1199.

Tóth, A., Ciosk, R., Uhlmann, F., Galova, M., Schleiffer, A., & Nasmyth, K. (1999). Yeast cohesin complex requires a conserved protein, Eco1p (Ctf7), to establish cohesion between sister chromatids during DNA replication. *Genes & Development*, *13*(3), 320-333.

Tsubouchi, H., & Roeder, G. S. (2002). The Mnd1 protein forms a complex with *hop*2 to promote homologous chromosome pairing and meiotic double-strand break repair. *Molecular and Cellular Biology*, *22*(9), 3078-3088.

Tullberg, A., Alexciev, K., Pfannschmidt, T., & Allen, J. F. (2000). Photosynthetic electron flow regulates transcription of the *psa*B gene in pea (*Pisum sativum* L.) chloroplasts through the redox state of the plastoquinone pool. *Plant and Cell Physiology*, *41*(9), 1045-1054.

Turner, G., Gershenzon, J., Nielson, E. E., Froehlich, J. E., & Croteau, R. (1999). Limonene synthase, the enzyme responsible for monoterpene biosynthesis in peppermint, is localized to leucoplasts of oil gland secretory cells. *Plant Physiology*, *120*(3), 879-886.

Uanschou, C., Ronceret, A., Von Harder, M., De Muyt, A., Vezon, D., Pereira, L., ... & Grelon, M. (2013). Sufficient amounts of functional *hop*2/*mnd*1 complex promote interhomolog DNA repair but are dispensable for intersister DNA repair during meiosis in *Arabidopsis*. *The Plant Cell*, *25*(12), 4924-4940.

Ueda, M., Nishikawa, T., Fujimoto, M., Takanashi, H., Arimura, S. I., Tsutsumi, N., & Kadowaki, K. I. (2008). Substitution of the gene for chloroplast *rps*16 was assisted by generation of a dual targeting signal. *Molecular Biology and Evolution*, *25*(8), 1566-1575.

Uhlmann, F. (2008). Cohesin branches out. *Nature*, *451*(7180), 777-778.

Uhlmann, F. (2016). SMC complexes: from DNA to chromosomes. *Nature Reviews Molecular Cell Biology*, *17*(7), 399-412.

Uji, T., & Mizuta, H. (2021). Treatment with heat shock protein 90 (Hsp90) inhibitors induces asexual life cycle in the marine red alga *NeoPyropia yezoensis* (Rhodophyta). *Aquaculture Research*.

Uji, T., Matsuda, R., Takechi, K., Takano, H., Mizuta, H., & Takio, S. (2016). Ethylene regulation of sexual reproduction in the marine red alga *Pyropia yezoensis* (Rhodophyta). *Journal of Applied Phycology*, *28*(6), 3501-3509.

Ünal, E., Heidinger-Pauli, J. M., & Koshland, D. (2007). DNA double-strand breaks trigger genome-wide sister-chromatid cohesion through Eco1 (Ctf7). *Science*, *317*(5835), 245-248.

Van de Peer, Y., Neefs, J. M., De Rijk, P., & De Wachter, R. (1993). Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. *Journal of Molecular Evolution*, *37*(2), 221-232.

van der Meer, J. P., & Todd, E. R. (1977). Genetics of *Gracilaria* sp. (Rhodophyceae, Gigartinales). IV. Mitotic recombination and its relationship to mixed phases in the life history. *Canadian Journal of Botany*, *55*(22), 2810-2817

van Dooren, G. G., Schwartzbach, S. D., Osafune, T., & McFadden, G. I. (2001). Translocation of proteins across the multiple membranes of complex plastids. *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research*, *1541*(1), 34-53.

Varela-Álvarez, E., Loureiro, J., Paulino, C., & Serrão, E. A. (2018). Polyploid lineages in the genus *Porphyra*. *Scientific Reports*, *8*(1), 1-15.

Varela-Alvarez, E., Stengel, D. B., & Guiry, M. D. (2004). The use of image processing in assessing conchocelis growth and conchospore production in *Porphyra linearis*. *Phycologia*, *43*(3), 282-287.

Waaland, J. R., Stiller, J. W., & Cheney, D. P. (2004). Macroalgal candidates for genomics. *Journal of Phycology*, *40*(1), 26-33.

Wali, S., Munir, F., & Mahmood, T. (2013). Phylogenetic studies of selected Citrus species based on chloroplast gene, *rps*14. *Int. J. Agric. Biol*, *15*(2), 357-361.

Walker, M. Y., & Hawley, R. S. (2000). Hanging on to your homolog: the roles of pairing, synapsis and recombination in the maintenance of homolog adhesion. *Chromosoma*, *109*(1-2), 3-9.

Waller, R. F., Slamovits, C. H., & Keeling, P. J. (2006). Lateral gene transfer of a multigene region from cyanobacteria to dinoflagellates resulting in a novel plastid-targeted fusion protein. *Molecular Biology and Evolution*, *23*(7), 1437-1443.

Wang, J., Dai, J., & Zhang, Y. (2006). Nuclear division of the vegetative cells, conchosporangial cells and conchospores of *Porphyra yezoensis* (*Bangia*les, Rhodophyta). *Phycological Research*, *54*(3), 201-207.

Wang, L., Mao, Y., Kong, F., Li, G., Ma, F., Zhang, B., ... & Cao, M. (2013). Complete sequence and analysis of plastid genomes of two economically important red algae: *Pyropia haitanensis* and *Pyropia yezoensis*. *PloS one*, *8*(5), e65902.

Wang, L., Xing, H., Yuan, Y., Wang, X., Saeed, M., Tao, J., ... & Sun, X. (2018). Genome-wide analysis of codon usage bias in four sequenced cotton species. *PloS one*, *13*(3), e0194372.

Wang, M., Liu, Y. S., Zhou, J. H., Chen, H. T., Ma, L. N., Ding, Y. Z., ... & Zhang, J. (2011). Analysis of codon usage in Newcastle disease virus. *Virus Genes*, *42*(2), 245-253.

Wang, T. F., Kleckner, N., & Hunter, N. (1999). Functional specificity of *mut*L homologs in yeast: evidence for three *mlh*1-based heterocomplexes with distinct roles during meiosis in recombination and mismatch correction. *Proceedings of the National Academy of Sciences*, *96*(24), 13914-13919.

Wang, Y., van Rengs, W. M., Zaidan, M. W. A. M., & Underwood, C. J. (2021). Meiosis in crops: from genes to genomes. *Journal of Experimental Botany*, *72*(18), 6091-6109.

Warner, J. R., & McIntosh, K. B. (2009). How common are extraribosomal functions of ribosomal proteins?. *Molecular Cell*, *34*(1), 3-11.

Wehr, J. D., Sheath, R. G., & Kociolek, J. P. (Eds.). (2015). Freshwater algae of North America: ecology and classification. Elsevier.

Weng, M. L., Blazier, J. C., Govindu, M. & Jansen, R. K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats and nucleotide substitution rates. *Mol. Biol. Evol.* 31, 645–659 (2013).

White III, R.A., Brown, J., Colby, S., Overall, C.C., Lee, J.Y., Zucker, J., Glaesemann, K.R., Jansson, C. and Jansson, J.K. (2017). ATLAS (automatic tool for local assembly structures)-a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. *PeerJ Preprints*, *5*, e2843v1

Whitehead, A., & Crawford, D. L. (2006). Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, *15*(5), 1197-1211.

Wiencke, C., Clayton, M. N., Gómez, I., Iken, K., Lüder, U. H., Amsler, C. D., ... & Dunton, K. (2007). Life strategy, ecophysiology and ecology of seaweeds in polar waters. *Reviews in Environmental Science and Bio/Technology*, *6*(1-3), 95-126.

Wilkins, A. S., & Holliday, R. (2009). The evolution of meiosis from mitosis. *Genetics*, *181*(1), 3-12.

Wilson, T., Guerrette, S., & Fishel, R. (1999). Dissociation of mismatch recognition and ATPase activity by hMSH2-hMSH3. *Journal of Biological Chemistry*, *274*(31), 21659-21664.

Wolfe, K. H., & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, *387*(6634), 708-713.

Wong, W. S., Yang, Z., Goldman, N., & Nielsen, R. (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics, 168(2), 1041-1051.

Woolcott, G. W., & King, R. J. (1998). *Porphyra* and *Bangia* (Bangiaceae, Rhodophyta) in warm temperate waters of eastern Australia: Morphological and molecular analyses. *Phycological Research, 46*(2), 111-123.

Wray, J. J. (1977). Late Paleozoic calcareous red algae. In *Fossil algae* (pp. 167-176). Springer, Berlin, Heidelberg.

Wu, Q., Wang, X. Z., Tang, Y. Y., Yu, H. T., Cui, F. G., Zhang, J. C., & Wang, C. T. (2014). Molecular cloning, genomic organization and functional analysis of the ribosomal protein L4/L1 (*rpl*4) gene from *Arachis hypogaea*. *Canadian Journal of Plant Science*, *94*(1), 85-97.

Xia, X., Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its application. *Molecular phylogenetics and evolution*, *26*(1), 1-7.

Xiao, S. (2004). New multicellular algal fossils and acritarchs in *Doushantuo chert* nodules (Neoproterozoic; Yangtze Gorges, south China). *Journal of Paleontology*, *78*(2), 393-401.

Xie, C., Li, B., Xu, Y., Ji, D., & Chen, C. (2013). Characterization of the global transcriptome for *Pyropia haitanensis* (Bangiales, Rhodophyta) and development of cSSR markers. *BMC Genomics, 14*(1), 107.

Xie, D. F., Yu, Y., Deng, Y. Q., Li, J., Liu, H. Y., Zhou, S. D., & He, X. J. (2018). Comparative analysis of the chloroplast genomes of the Chinese endemic genus *Urophysa* and their contribution to chloroplast phylogeny and adaptive evolution. *International Journal of Molecular Sciences*, *19*(7), 1847.

Xiong, H. B., Wang, J., Huang, C., Rochaix, J. D., Lin, F. M., Zhang, J. X., ... & Yang, Z. N. (2020). mTERF8, a member of the mitochondrial transcription termination factor family, is involved in the transcription termination of chloroplast Gene *psb*J. *Plant Physiology*, 182(1), 408-423.

Xiong, Y., Zhao, C. Y., Yang, Q. S. & Yang, X. L. 2014. In silico cloning, bioinformatics and adaptive evolution analysis of rbcL gene in Artemisia annua. Biotechnology, 24(6): 50-56

Xu, J., Jiang, B., Chai, S., He, Y., Zhu, J., Shen, Z., & Shen, S. (2016). Complete nuclear ribosomal DNA sequence amplification and molecular analyses of *Bangia* (Bangiales, Rhodophyta) from China. *Chinese Journal of Oceanology and Limnology*, *34*(5), 1044-1053.

Xu, K., Tang, X., Bi, G., Cao, M., Wang, L., & Mao, Y. (2018). The first complete organellar genomes of an Antarctic red alga, *Pyropia endiviifolia*: insights into its genome architecture and phylogenetic position within genus *Pyropia* (Bangiales, Rhodophyta). *Journal of Oceanology and Limnology*, *36*(4), 1315-1328.

Xu, K., Tang, X., Wang, L., Yu, X., Sun, P., & Mao, Y. (2018). Divergence time, historical biogeography and evolutionary rate estimation of the order Bangiales (Rhodophyta) inferred from multilocus data. *Journal of Oceanology and Limnology*, 36(3), 870-881.

Xu, L., Zhou, Y. H., Wang, S. J., Wang, Y. X., Ma, L. B., & Zhou, Z. G. (2003). Ultrastructural and molecular evidence for monospore formation in the thallus *of Porphyra yezoensis*. *J Fish China*, *27*, 519-527.

Xu, Q., Yu, L., Chitnis, V. P., & Chitnis, P. R. (1994). Function and organization of photosystem I in a cyanobacterial mutant strain that lacks psaF and psaJ subunits. Journal of Biological Chemistry, 269(5), 3205-3211.

Xu, Y., Xie, C., Chen, C., Ji, D., & Gao, Y. (2012). Genetic analyses of six quantitative traits of a doubled haploid population of *Porphyra haitanensis* Chang et Zheng (Bangiales, Rhodophyta). *Journal of applied phycology*, *24*(1), 89-96.

Yamaguchi, K., & Subramanian, A. R. (2000). The plastid ribosomal proteins: identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). *Journal of Biological Chemistry*, *275*(37), 28466-28482.

Yan, X. H., Fujita, Y., & Aruga, Y. (2004). High monospore-producing mutants obtained by treatment with MNNG in *Porphyra yezoensis* Ueda (Bangiales, Rhodophyta). In *Asian Pacific Phycology in the 21st Century: Prospects and Challenges* (pp. 133-140). Springer, Dordrecht.

Yan, X. H., Lv, F., Liu, C. J., & Zheng, Y. F. (2010). Selection and characterization of a high-temperature tolerant strain of *Porphyra haitanensis* Chang et Zheng (Bangiales, Rhodophyta). *Journal of Applied Phycology*, *22*(4), 511-516.

Yanagisawa, R., Sekine, N., Mizuta, H., & Uji, T. (2019). Transcriptomic analysis under ethylene precursor treatment uncovers the regulation of gene expression linked to sexual reproduction in the dioecious red alga *Pyropia pseudolinearis*. *Journal of Applied Phycology*, *31*(5), 3317-3329.

Yang, L. E., Lu, Q. Q., & Brodie, J. (2017). A review of the bladed Bangiales (Rhodophyta) in China: history, culture and taxonomy. *European Journal of Phycology*, *52*(3), 251-263.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586-1591.

Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, *15*(12), 496-503.

Yang, Z., & Dos Reis, M. (2010). Statistical properties of the branch-site test of positive selection. Molecular Biology and Evolution, 28(3), 1217-1228.

Yang, Z., & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Molecular Biology and Evolution, 19(6), 908-917.

Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics, 155(1), 431-449.

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. Molecular Biology and Evolution, 22(4), 1107-1118.

Yesson, C., Bian, X., Williamson, C., Briscoe, A. G., & Brodie, J. (2020). Mitochondrial and plastid genome variability of *Corallina officinalis* (Corallinales, Rhodophyta). *Applied Phycology*, *1*(1), 73-79.

Yin, Y., Cheong, H., Friedrichsen, D., Yunde Zhao, Y., Jianping Hu, J., Santiago Mora-Garcia, S., Chory, J. (2002) A crucial role for the putative *Arabidopsis* topoisomerase VI in plant growth and development. PNAS, 99, 10191-10196.

Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G., & Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. *Molecular Biology and Evolution*, *21*(5), 809-818.

Yoon, H. S., Hackett, J. D., Pinto, G., & Bhattacharya, D. (2002). The single, ancient origin of chromist plastids. Journal of Phycology, 38, 40-40.

Yoon, H. S., Müller, K. M., Sheath, R. G., Ott, F. D., & Bhattacharya, D. (2006). Defining the major lineages of red algae (Rhodophyta) 1. *Journal of Phycology*, *42*(2), 482-492.

Yoon, H. S., Müller, K. M., Sheath, R. G., Ott, F. D., & Bhattacharya, D. (2006). Defining the major lineages of red algae (Rhodophyta) 1. *Journal of Phycology*, *42*(2), 482-492.

Young, J. N., Rickaby, R. E. M., Kapralov, M. V., & Filatov, D. A. (2012). Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1588), 483-492.

Young, N. D., & Claude, W. D. (2005). Rate variation in parasitic plants: correlated and uncorrelated patterns among plastid genes of different function. *BMC Evolutionary Biology*, *5*(1), 1-10.

Young, N. D., & DePamphilis, C. W. (2000). Purifying selection detected in the plastid gene *mat*K and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. *Molecular Biology and Evolution*, *17*(12), 1933-1941.

Young, N. D., Steiner, K. E., & dePamphilis, C. W. (1999). The evolution of parasitism in Scrophulariaceae/Orobanchaceae: plastid gene sequences refute an evolutionary transition series. *Annals of the Missouri Botanical Garden*, 876-893.

Yu, J., Li, H., Ge, X., Li, J., & Tang, Y. (2010). Characteristics analysis of complete ribosomal DNA sequence in Chinese mitten crab (*Eriocheir sinensis*). *Journal of Fisheries of China*, *34*(5), 696-703.

Yu, Y., Ouyang, Y., & Yao, W. (2018). shinyCircos: An R/Shiny application for interactive creation of Circos plot. *Bioinformatics*, *34*(7), 1229-1231.

Yudina, S. V., Schelkunov, M. I., Nauheimer, L., Crayn, D., Chantanaorrapint, S., Hroneš, M., ... & Logacheva, M. D. (2021). Comparative analysis of plastid genomes in the non-photosynthetic genus *Thismia* reveals ongoing gene set reduction. *Frontiers in Plant Science*, *12*.

Zakharyevich, K., Tang, S., Ma, Y., & Hunter, N. (2012). Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. *Cell*, 149, 334-347.

Zhan, S. H., Shih, C. C., & Liu, S. L. (2020). Reappraising plastid markers of the red algae for phylogenetic community ecology in the genomic era. *Ecology and Evolution*, *10*(3), 1299-1310.

Zhang, J. (2004). Frequent false detection of positive selection by the likelihood method with branch-site models. *Molecular Biology and Evolution*, 21(7), 1332-1339.

Zhang, J., Liu, N., Meinita, M. D. N., Wang, X., Tang, X., Wang, G., ... & Liu, T. (2020). The complete plastid genomes of *Betaphycus gelatinus*, *Eucheuma denticulatum*, and *Kappaphycus striatus* (Solieriaceae: Rhodophyta) and their phylogenetic analysis. *Journal of Applied Phycology*, *32*, 3521-3532.

Zhang, J., Yuan, H., Yang, Y., Fish, T., Lyi, S. M., Thannhauser, T. W., ... & Li, L. (2016). Plastid ribosomal protein S5 is involved in photosynthesis, plant development, and cold stress tolerance in *Arabidopsis*. *Journal of Experimental Botany*, *67*(9), 2731-2744.

Zhang, L. J., Chen, J., & Wang, T. (2010). Adaptive evolution in the chloroplast gene rps4 in ferns. *Bull. Bot. Res*, 30, 42-50.

Zhang, S., & Scheller, H. V. (2004). Light-harvesting complex II binds to several small subunits of photosystem I. *Journal of Biological Chemistry*, 279(5), 3180-3187.

Zhang, Y., Yan, X. H., & Aruga, Y. (2013). The sex and sex determination in *Pyropia haitanensis* (Bangiales, Rhodophyta). *Plos One*, *8*(8), e73414.

Zhao, J., Snyder, W. B., Mühlenhoff, U., Rhiel, E., Warren, P. V., Golbeck, J. H., & Bryant, D. A. (1993). Cloning and characterization of the *psa*E gene of the cyanobacterium *Synechococcus* sp. PCC 7002: characterization of a *psa*E mutant and overproduction of the protein in *Escherichia coli*. *Molecular microbiology*, 9(1), 183-194.

Zhao, W., Saro, D., Hammel, M., Kwon, Y., Xu, Y., Rambo, R. P., ... & Sung, P. (2013). Mechanistic insights into the role of *hop*2–*mnd*1 in meiotic homologous DNA pairing. *Nucleic Acids Research*, *42*(2), 906-917.

Zhou, G., Sun, Y., Xin, H., Zhang, Y., Li, Z., & Xu, Z. (2004). In vivo antitumor and immunomodulation activities of different molecular weight lambda-carrageenans from *Chondrus ocellatus*. *Pharmacological Research*, *50*(1), 47-53.

Zhou, Y., Wang, B., Gao, L., & Wang, T. (2011). Adaptive evolution and coevolution of the *rbc*L gene in xeric Pteridaceae ferns. *Plant Science Journal*, 29(4), 409-416.

Zierhut, C., Berlinger, M., Rupp, C., Shinohara, A. & Klein, F. *mnd*1 is required for meiotic interhomolog repair. *Curr. Biol.* 14, 752–762 (2004).

Zuckerkandl, E., & Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of Theoretical Biology*, *8*(2), 357-366.

Appendix A

Detailed *Bangia atropurpurea* extraction and sequencing

Bangia cell pellet was frozen in liquid nitrogen and then grinded to fine powder using a clean mortar and pestle. DNA was extracted using Sox DNA Isolation Kit (Metagenom Bio Inc. https://metagenom-bio.myshopify.com/products/sox-dna-extraction-kit) according to the supplier's instruction with following modification. After adding 100 mg of ground cells and buffer Sox1 to a bead tube, the mixture was incubated at 70^{o}C for 10 min. Cells were then homogenized using FastPrep-24 (MP Biomedicals) at 6M/second for 40 second. Following centrifugation of the crude extract, DNA was isolated using a spin column as described in the instruction. The quality of genomic DNA was evaluated with 0.8% TAE agarose gel.

DNA prep was quantified using dsDNA HS Assay Kit (ThermoFisher) and diluted to 0.25 ng/ml. Genomic sequencing library was constructed using Nextera XT kit (Illumina) based on the supplier's guide. Following random tagmentation of the DNA with transposase, fragmented DNA was enriched with two PCR primers (forward primer, AATGATACGGCGACCACCGAGATCTACACGTACTGACTCGTCGGCAGCGTC; reverse primer, CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTCGG). Index sequences (i5, GTACTGAC, underlined; i7, TCATGAGC, underlined) were incorporated into the PCR primers for multiplexing sequenced reads. Library DNA was evaluated with 2% TAE agarose gel and quantified with the dsDNA HS Assay Kit. The DNA of 0.4 kb to 1 kb was denatured and pooled with other sequencing amplicons. Sequencing was performed for 2×250 cycles using MiSeq platform (Illumina). After trimming off adapter sequences, DNA sequences were assembled with N714 CAAGCAGAAGACGGCATACGAGATTCATGAGCGTCTCGTGGGCTCGG

Appendix B

Table B. 1 Chloroplast genes present or absent in Bangiales species

Species & accession number		*Bangia atropurpurea* MN484624	*Porphyra umbilicalis* JQ408795	*Bangia fuscopurpurea* KP714733.1	*Pyropia haitanensis* NC_021189.1	*Porphyra purpurea* NC_000925.1	*Pyropia yezoensis* NC_007932.1	*Pyropia perforata* KF515973	*Pyropia kanakaensis* KJ776836	*Pyropia endiviifolia* KT716756	*Pyropia fucicola* KJ776837	*Wildemania schizophylla* NC_029576
Group of Genes	Gene											
ATP synthase	*atp*A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	*atp*B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	*atp*D	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	*atp*E	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	*atp*F	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	*atp*G	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	*atp*H	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	*atp*I	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ClpP, matK	*clp*C	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cytochrome b6f complex	*petA*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	petB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	petD	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	petG	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	petN	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	petF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hypothetical chloroplast reading frames	*ycf3*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf7	✓	✓	×	✓	✓	×	×	×	✓	×	×
	ycf12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf16	✓	✓	×	✓	✓	✓	×	×	×	×	×
	ycf17	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf18	✓	✓	✓	✓	✓	×	×	×	×	×	×
	ycf19	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf20	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf21	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf22	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf23	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf24	✓	✓	✓	✓	✓	✓	×	×	×	×	×
	ycf26	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
	ycf27	×	×	×	✓	✓	×	×	✓	×	×	×
	ycf28	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
	ycf29	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf31	✓	✓	×	✓	✓	×	×	×	×	×	✓
	ycf32	✓	✓	×	✓	✓	×	×	×	×	×	×
	ycf33	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf34	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf35	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf36	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf37	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf38	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	ycf39	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf46	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf59	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ycf61	✓	✓	✓	✓	✓	✓	✓	×	✓	×	✓
	ycf65	✓	✓	×	✓	✓	✓	✓	✓	✓	×	✓

Genes in blue (*pet*N and *ycf*27) are absent in *B. atropurpurea*, but present in other Bangiales species.

Table B. 2 Chloroplast genes present or absent in Bangiales species

Species & accession number		*Bangia atropurpurea* MN484624	*Porphyra umbilicalis* JQ408795	*Bangia fuscopurpurea* KP714733.1	*Pyropia haitanensis* NC_021189.1	*Porphyra purpurea* NC_000925.1	*Pyropia yezoensis* NC_007932.1	*Pyropia perforata* KF515973	*Pyropia kanakaensis* KJ776836	*Pyropia endiviifolia* KT716756	*Pyropia fucicola* KJ776837	*Wildemania schizophylla* NC_029576
Group of Genes	Gene											
Orfs	*orf27*	✓	×	✓	✓	✓	✓	×	×	×	×	×
	orf38	×	×	✓	×	✓	×	×	×	×	×	×
	orf58	✓	✓	×	✓	✓	×	✓	✓	✓	✓	✓
	orf65	✓	✓	×	✓	×	×	×	×	×	×	×
	orf68	×	×	✓	✓	×	×	✓	✓	✓	✓	✓
	orf71	✓	✓	✓	✓	×	×	✓	✓	×	✓	✓
	orf71a	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
	orf74	✓	✓	×	×	×	×	×	×	×	×	×
	orf75	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
	orf107	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf108	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf110	✓	✓	×	×	×	×	×	×	×	×	×
	orf110a	✓	✓	×	×	×	×	×	×	×	×	×
	orf111	✓	×	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf114	✓	×	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf121	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf148	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf149	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf174	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf198	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf199	✓	✓	×	✓	✓	×	×	×	×	×	×
	orf203	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf238	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf240	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf263	✓	✓	✓	✓	×	×	×	×	×	×	×
	orf263a	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf287	✓	×	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf288	✓	✓	×	×	×	×	×	×	×	×	×
	orf320	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf327	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓
	orf382	✓	×	✓	✓	✓	×	✓	✓	✓	×	✓
	orf383	✓	✓	×	×	×	×	×	×	×	×	×
	orf450	✓	✓	✓	✓	✓	×	✓	✓	×	✓	✓
	orf565	✓	✓	✓	✓	×	×	✓	✓	✓	✓	✓
	orf621	✓	×	✓	✓	✓	×	✓	✓	✓	✓	✓

Genes in blue (*orf*38 and *orf*68) are absent in *B. atropurpurea*, but present in other Bangiales species. Genes in orange (*orf*27, *orf*111, *orf*114, *orf*199, *orf*287, *orf*382, *orf*621) are present in *B. atropurpurea* but absent in *P. umbilicalis.*

Table B. 3 Chloroplast genes present or absent in Bangiales species

Species & accession number		*Bangia atropurpurea* MN484624	*Porphyra umbilicalis* JQ408795	*Bangia fuscopurpurea* KP714733.1	*Pyropia haitanensis* NC_021189.1	*Porphyra purpurea* NC_000925.1	*Pyropia yezoensis* NC_007932.1	*Pyropia perforata* KF515973	*Pyropia kanakaensis* KJ776836	*Pyropia endiviifolia* KT716756	*Pyropia fucicola* KJ776837	*Wildemania schizophylla* NC_029576
Group of Genes	Gene											
Photosystem I	*psaA*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaD	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaJ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaK	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psaI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Photosystem II	*psbA*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbB	✓	✓	✓	✓	✓	✓	×	×	✓	×	×
	psbC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbD	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbJ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbK	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbV	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbW	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	psbY	×	×	✓	×	×	×	✓	✓	✓	✓	✓
	psbZ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Gene in blue (*psbZ*) are absent in *B. atropurpurea*, but present in other Bangiales species

Table B. 4 Chloroplast genes present or absent in Bangiales species

Species & accession number		*Bangia atropurpurea* MN484624	*Porphyra umbilicalis* JQ408795	*Bangia fuscopurpurea* KP714733.1	*Pyropia haitanensis* NC_021189.1	*Porphyra purpurea* NC_000925.1	*Pyropia yezoensis* NC_007932.1	*Pyropia perforata* KF515973	*Pyropia kanakaensis* KJ776836	*Pyropia endiviifolia* KT716756	*Pyropia fucicola* KJ776837	*Wildemania schizophylla* NC_029576
Group of Genes	Gene											
Ribosomal proteins(LSU)	*rpl1*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl9	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl11	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl13	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl14	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl16	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl18	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl19	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl20	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl21	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl22	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl23	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl24	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl27	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl29	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl28	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl31	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl32	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl33	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl34	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl35	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rpl36	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Ribosomal proteins(SSU)	*rps1*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps11	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps13	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps14	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps16	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps17	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps18	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps19	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps20	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps8	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rps9	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Gene in orange (*rpl*29) are present in *B. atropurpurea* but absent in *P. umbilicalis.*

Table B. 5 Chloroplast genes present or absent in Bangiales species

Species & accession number		*Bangia atropurpurea* MN484624	*Porphyra umbilicalis* JQ408795	*Bangia fuscopurpurea* KP714733.1	*Pyropia haitanensis* NC_021189.1	*Porphyra purpurea* NC_000925.1	*Pyropia yezoensis* NC_007932.1	*Pyropia perforata* KF515973	*Pyropia kanakaensis* KJ776836	*Pyropia endiviifolia* KT716756	*Pyropia fucicola* KJ776837	*Wildemania schizophylla* NC_029576
Group of Genes	Gene											
	accA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	accB	✓	✓	✓	✓	✓	✓	×	×	✓	×	✓
	accD	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	acpP	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	apcA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	apcB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	apcD	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	apcE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	apcF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	argB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	carA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	cbbX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ccs1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ccsA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	cem A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	chlB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	chlL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	chlN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	cpcA	✓	✓	✓	✓	✓	✓	✓	×	✓	×	✓
	cpcB	✓	✓	✓	✓	✓	✓	✓	×	✓	×	✓
	cpcG	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	cpeA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	cpeB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	dnaB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	dnaK	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	fabH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	*ftr*B	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓
	fts H	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Other genes	*glnB*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	glt B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	gro EL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ilvB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	ilvH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	inf B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	odp A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	odp B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	pgm A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	pbs A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	rne	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	secA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	secY	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	tat C	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	thi G	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	trp A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	trx A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	tsf	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	tuf A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	syf B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	syh	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	chl I	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	petJ	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	trp G	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	inf C	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	pre A	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓
	bas 1	×	×	✓	×	×	×	✓	✓	✓	✓	✓
	suf C	×	×	✓	×	×	×	✓	✓	✓	✓	✓

Genes in blue (*bas*1 and *suf*C) are absent in *B. atropurpurea*, but present in other Bangiales species. Gene in orange (*pre*A) are present in *B. atropurpurea* but absent in *P. umbilicalis.*

Appendix C

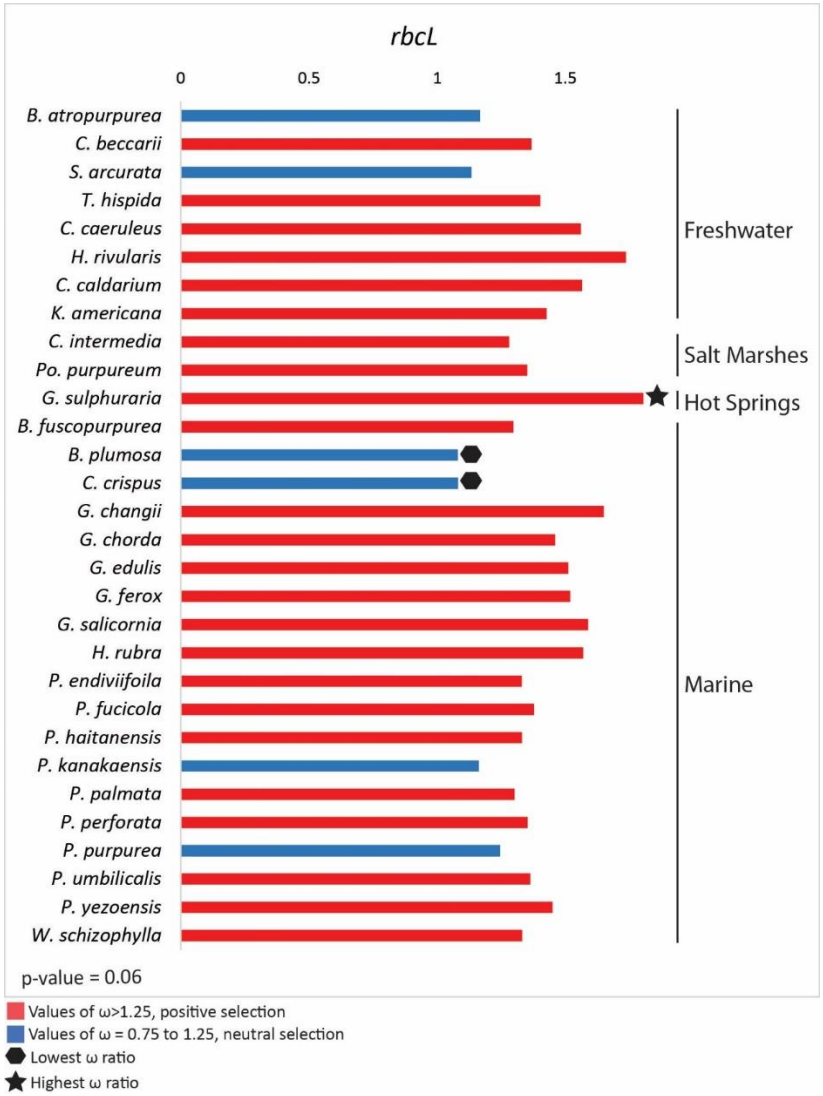

Figure C. 1.Comparison of the substitution ω ratios of the *rbc*L gene of red and green algae taking in consideration the species ecological habitat.

Appendix D

Table D. 1. List of *atp* genes present or absent in Rhodophyta and Chlorophyta species

Species	atpA[+]	atpB[*+]	atpD	atpE[*+]	atpF[+]	atpG	atpH[*+]	atpI[*]
Bangia atropurpurea	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra umbilicalis	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra purpurea	✓	✓	✓	✓	✓	✓	✓	✓
Bangia fuscopurpurea	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia fucicola	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia endiviifolia	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia haitanensis	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia kanakaensis	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia perforata	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia yezoensis	✓	✓	✓	✓	✓	✓	✓	✓
Wildemania schizophylla	✓	✓	✓	✓	✓	✓	✓	✓
Thorea hispida	✓	✓	✓	✓	✓	✓	✓	✓
Palmaria palmata	✓	✓	✓	✓	✓	✓	✓	✓
Chondrus crispus	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa beccarii	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa intermedia	✓	✓	✓	✓	✓	✓	✓	✓
Kumanoa americana	✓	✓	✓	✓	✓	✓	✓	✓
Sheathia arcuata	✓	✓	✓	✓	✓	✓	✓	✓
Hildenbrandia rubra	✓	✓	✓	✓	✓	✓	✓	✓
Hildenbrandia rivularis	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria changii	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria chorda	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria edulis	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria ferox	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria salicornia	✓	✓	✓	✓	✓	✓	✓	✓
Porphyridium purpureum	✓	✓	✓	✓	✓	✓	✓	✓
Compsopogon caeruleus	✓	✓	✓	✓	✓	✓	✓	✓
Cyanidioschyzon merolae	✓	✓	✓	✓	✓	✓	✓	✓
Cyanidium caldarium	✓	✓	✓	✓	✓	✓	✓	✓
Galdieria sulphuraria	✓	✓	✓	✓	✓	✓	✓	✓
Bryopsis plumosa●	✓	✓	✗	✓	✓	✗	✓	✓

* Statistical difference in different habitats; + Statistical difference in phylogenetic groups.
● Chlorophyta species
✓ Gene is present
✗ Gene is absent

Figure D 1. Comparison of the substitution ω ratios of *atp* genes of red and green algae taking in consideration the species ecological habitat.

Figure D. 2. Comparison of the substitution ω ratios of *atp* genes of red and green algae taking in consideration the phylogenetic position of the species.

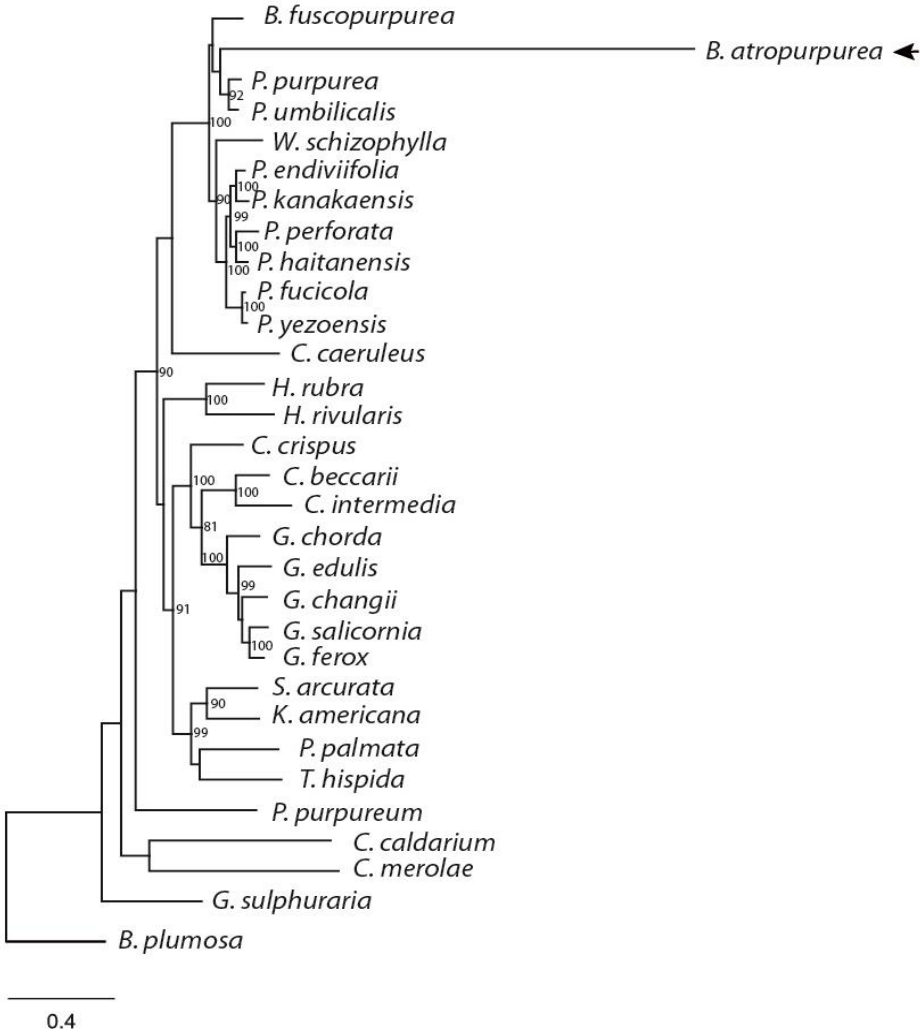

Figure D. 3. Phylogenetic tree based on concatenated sequences of 7 *atp* genes: *atp*A, *atp*B, *atp*E, *atp*F, *atp*G, *atp*H and *atp*I. The tree was reconstructed using the Maximum likelihood method and invoking the GRT+G model. Numbers at internal branches represent bootstrap support (values lower than 80% not included) as a percentage of 1000 replicates.

Appendix E

Table E. 1. List of *psa* genes present or absent in Rhodophyta and Chlorophyta species

Species	psaA*	psaB+	psaC+	psaD*	psaE+	psaF+	psaI+	psaJ	psaK	psaL*+	psaM*+
Bangia atropurpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra umbilicalis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra purpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bangia fuscopurpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia fucicola	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia endiviifolia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia haitanensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia kanakaensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia perforata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia yezoensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wildemania schizophylla	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Thorea hispida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Palmaria palmata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Chondrus crispus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa beccarii	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa intermedia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kumanoa americana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sheathia arcuata	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓
Hildenbrandia rubra	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hildenbrandia rivularis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria changii	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria chorda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria edulis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria ferox	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria salicornia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyridium purpureum	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Compsopogon caeruleus	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗
Cyanidioschyzon merolae	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cyanidium caldarium	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Galdieria sulphuraria	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bryopsis plumosa ●	✓	✓	✗	✗	✗	✓	✓	✗	✗	✗	✓

* Statistical difference among different in habitats; + Statistical difference among phylogenetic groups.
● Chlorophyta species
✓ Gene is present
✗ Gene is absent

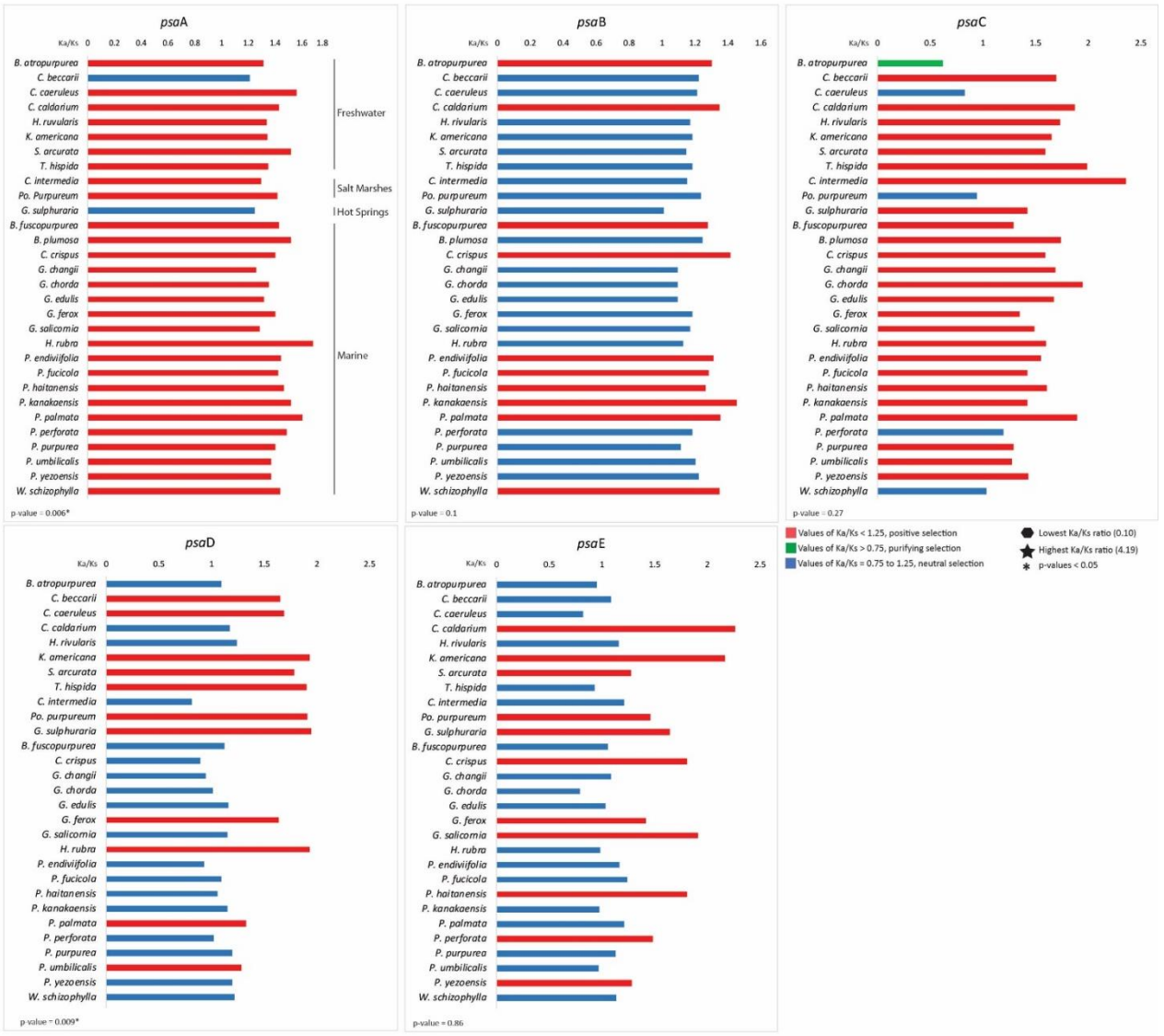

Figure E. 1. Comparison of the substitution ω ratios of *psa* (A – E) genes of red and green algae taking in consideration the species ecological habitat.

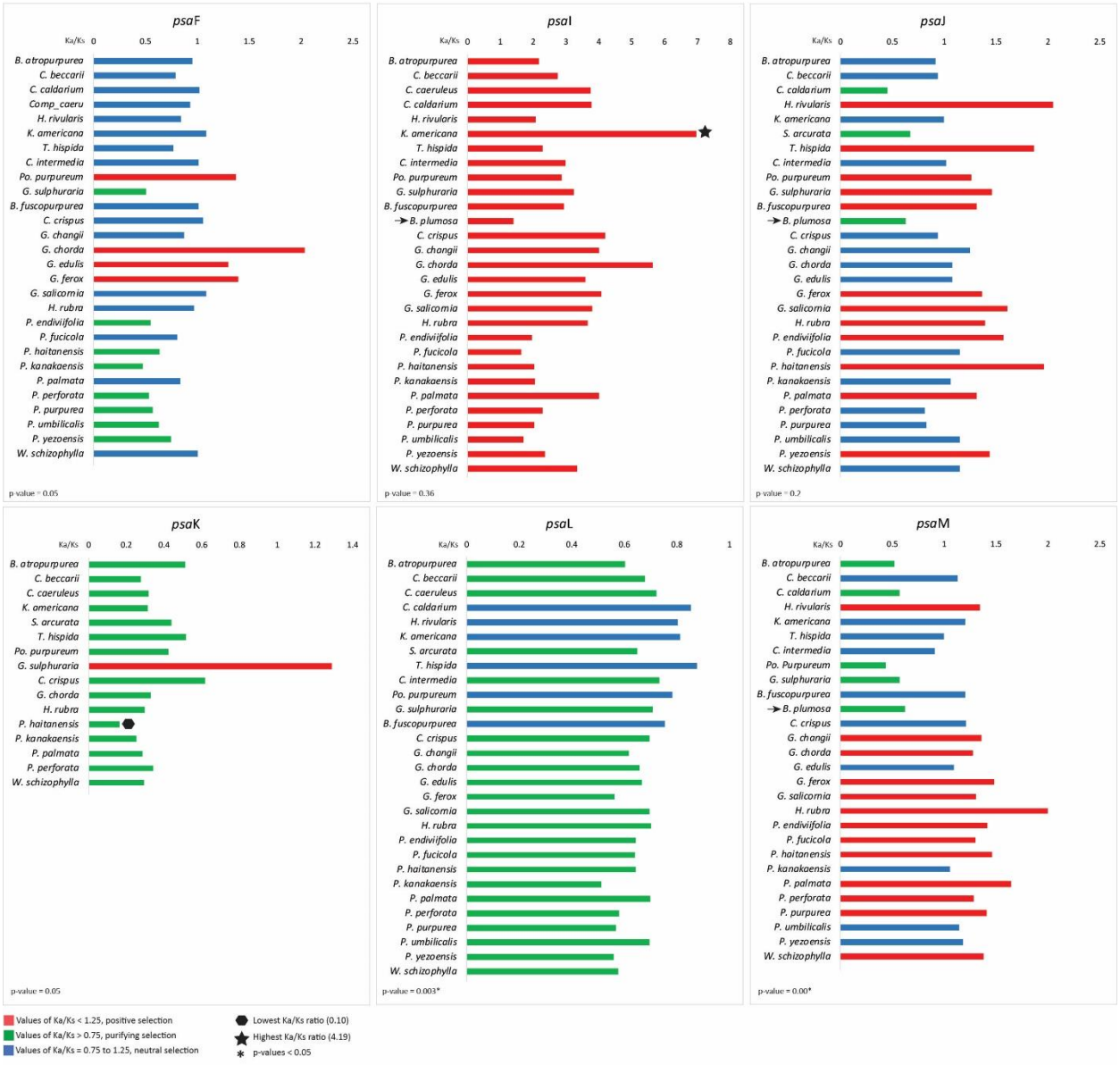

Figure E. 2. Comparison of the substitution ω ratios of *psa* (F – M) genes of red and green algae taking in consideration the species ecological habitat.

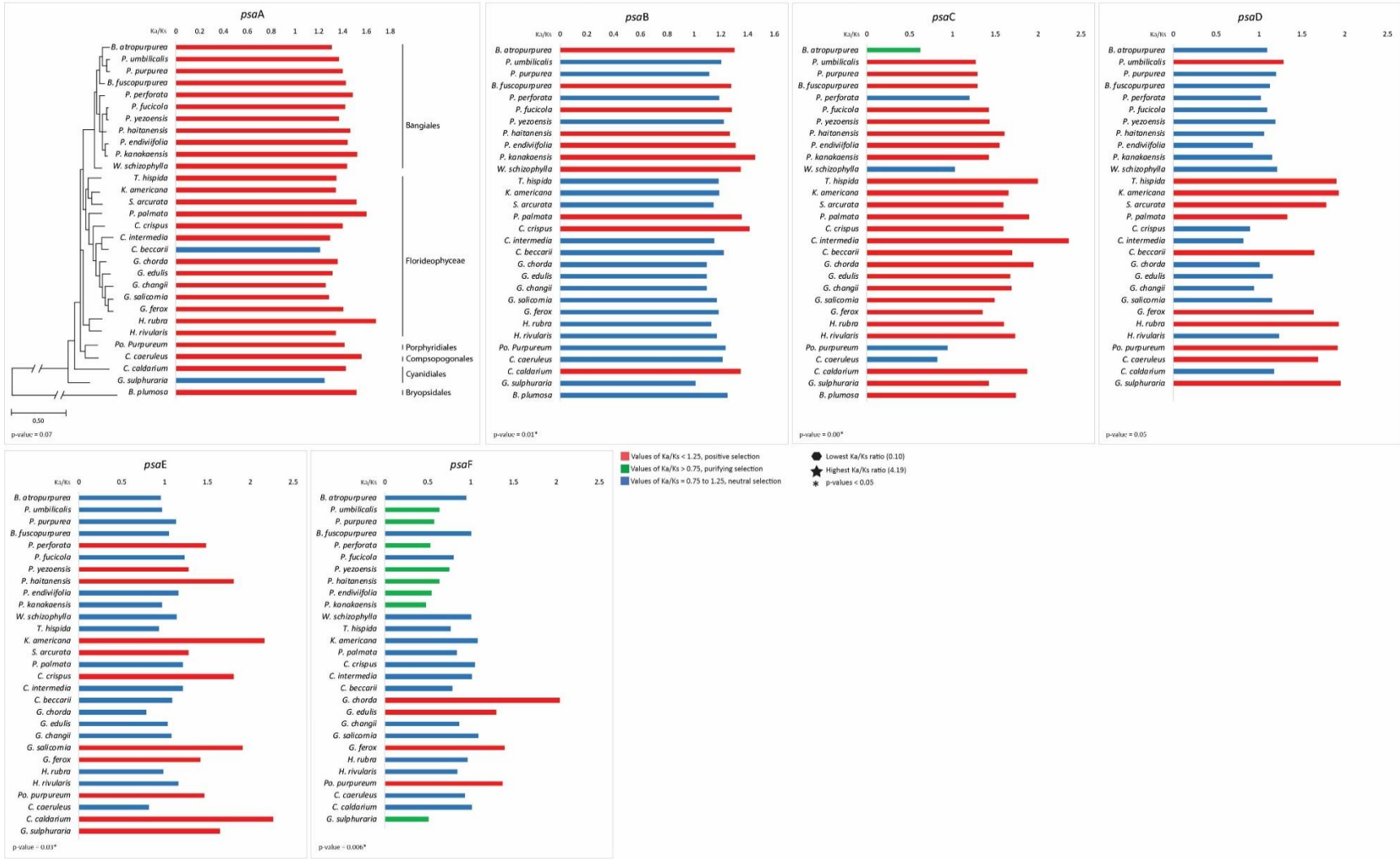

Figure E. 3. Comparison of the substitution ω ratios of *psa* genes (A – F) of red and green algae taking in consideration the phylogenetic position of the species.

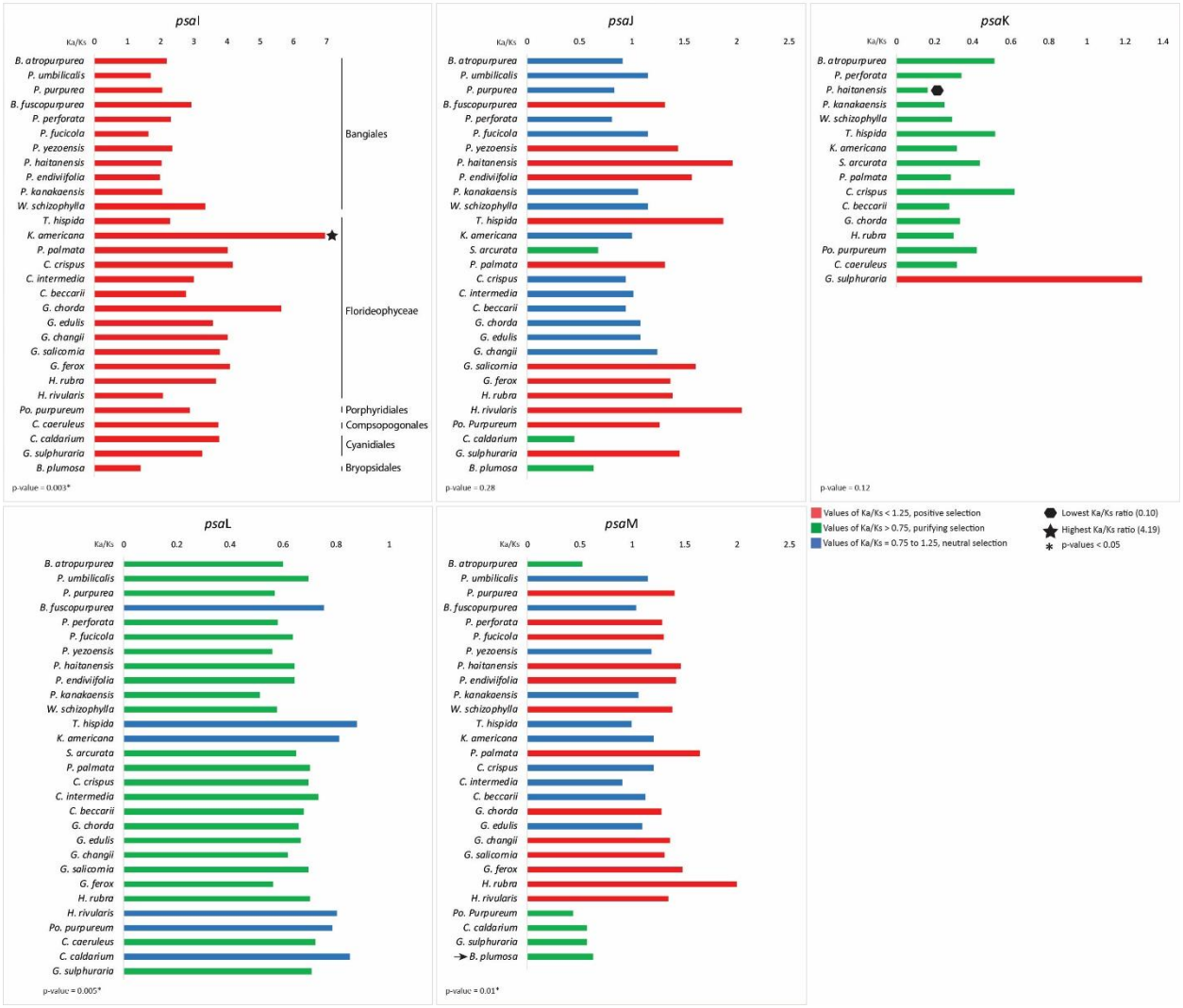

Figure E. 4. Comparison of the substitution ω ratios of *psa* genes (I – M) of red and green algae taking in consideration the phylogenetic position of the species.

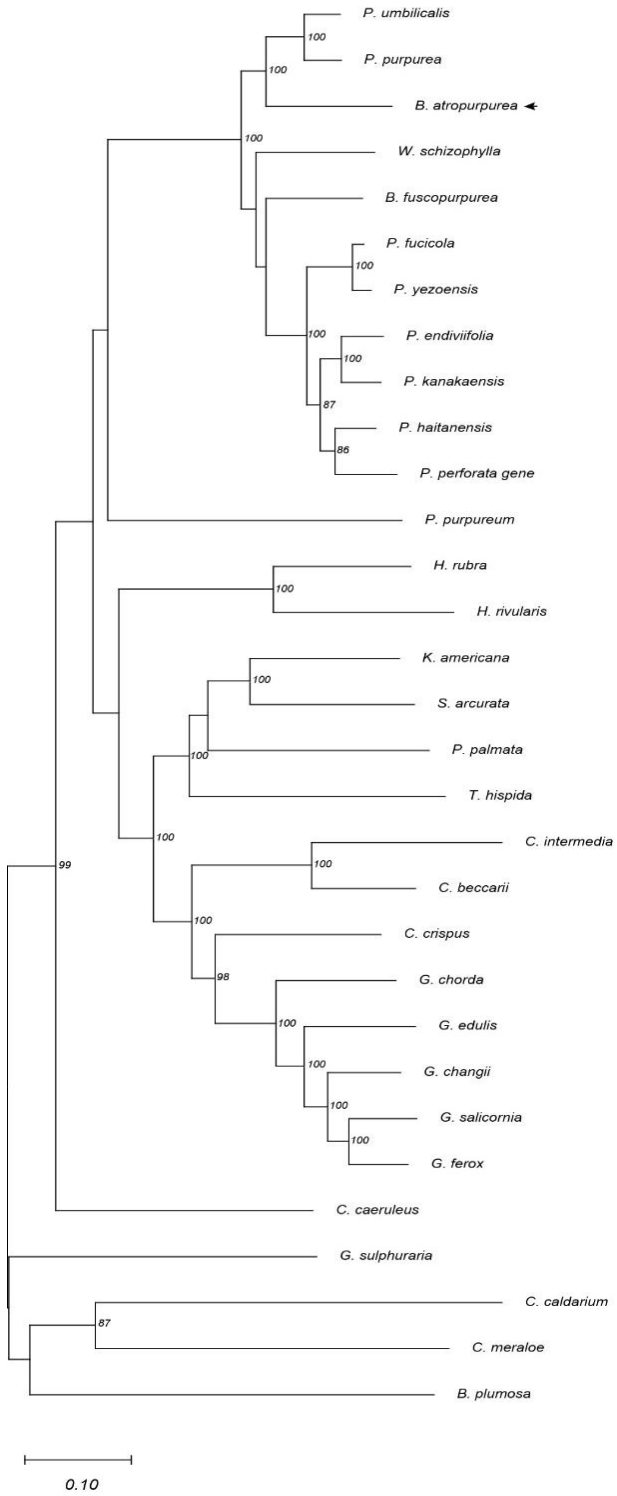

Figure E. 5. Phylogenetic tree based on the concatenated sequences of 11 *psa* genes: *psa*A, *psa*B, *psa*C, *psa*D, *psa*E, *psa*F, *psa*I, *atp*J, *psa*K, *psa*L and *psa*M. The tree was reconstructed using the Maximum likelihood method and invoking the GRT+G+I model. Numbers at internal branches represent bootstrap support (values lower than 80% not included) as a percentage of 1000 replicates.

Appendix F

Table F. 1. List of *psb* genes present or absent in Rhodophyta and Chlorophyta species

Species	psbA^{+}	psbB^{+}	psbC	psbD	psbE*	psbF	psbH*	psbK^{+}	psbL	psbN	psbT	psbV^{+}	psbW	psbX	psbZ*
Bangia atropurpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra umbilicalis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra purpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bangia fuscopurpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia fucicola	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia endiviifolia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia haitanensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia kanakaensis	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia perforata	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia yezoensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wildemania schizophylla	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Thorea hispida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Palmaria palmata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Chondrus crispus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓
Caloglossa beccarii	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa intermedia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kumanoa americana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓
Sheathia arcuata	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	×	✓
Hildenbrandia rubra	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hildenbrandia rivularis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria changii	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria chorda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria edulis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria ferox	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria salicornia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyridium purpureum	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Compsopogon caeruleus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓
Cyanidioschyzon merolae	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cyanidium caldarium	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×
Galdieria sulphuraria	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bryopsis plumosa●	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	✓

* Statistical difference among different in habitats; + Statistical difference among phylogenetic groups.
● Chlorophyta species
✓ Gene is present
× Gene is absent

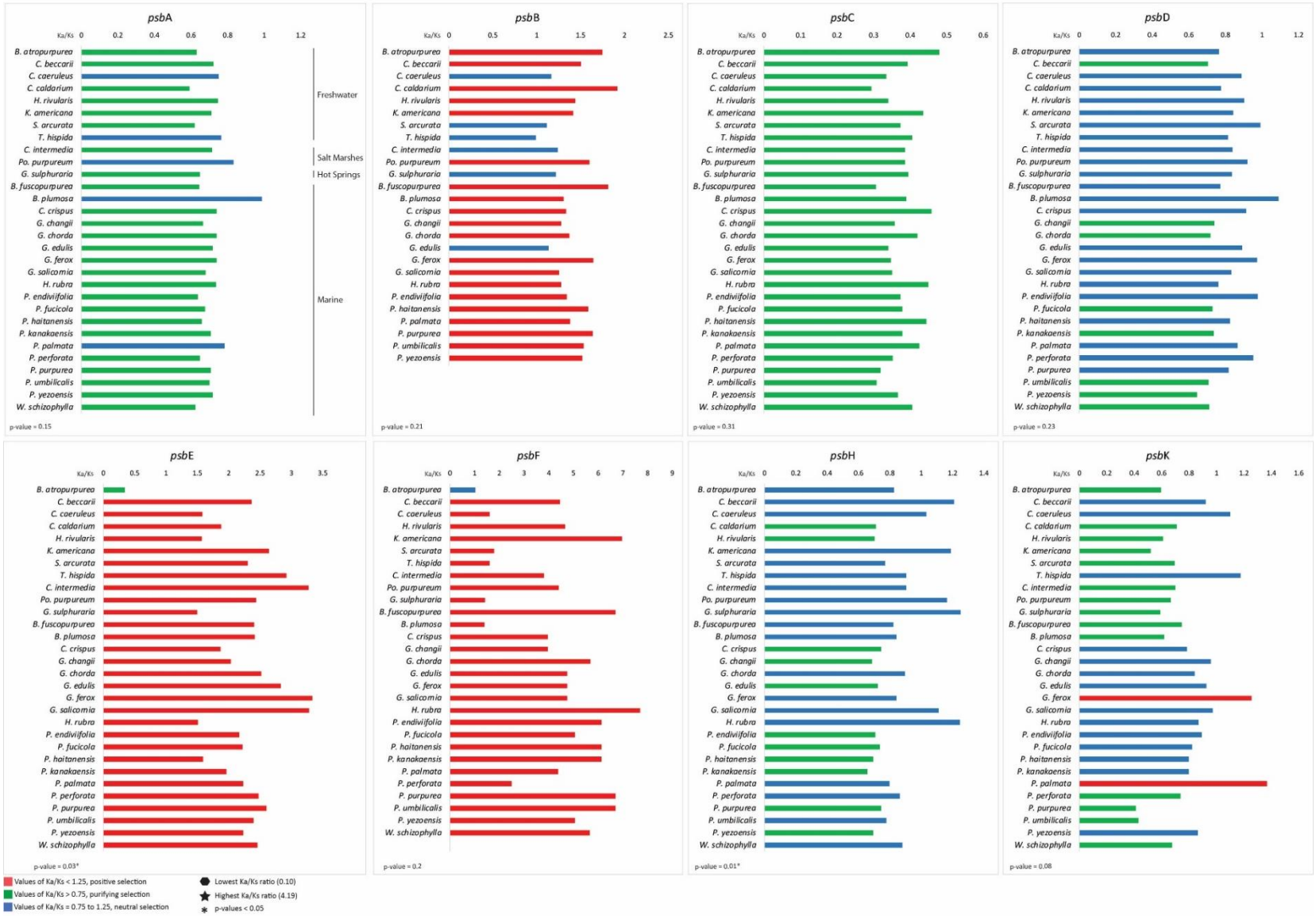

Figure F. 1. Comparison of the substitution ω ratios of *psb* genes (A – K) of red and green algae taking in consideration the species ecological habitat.

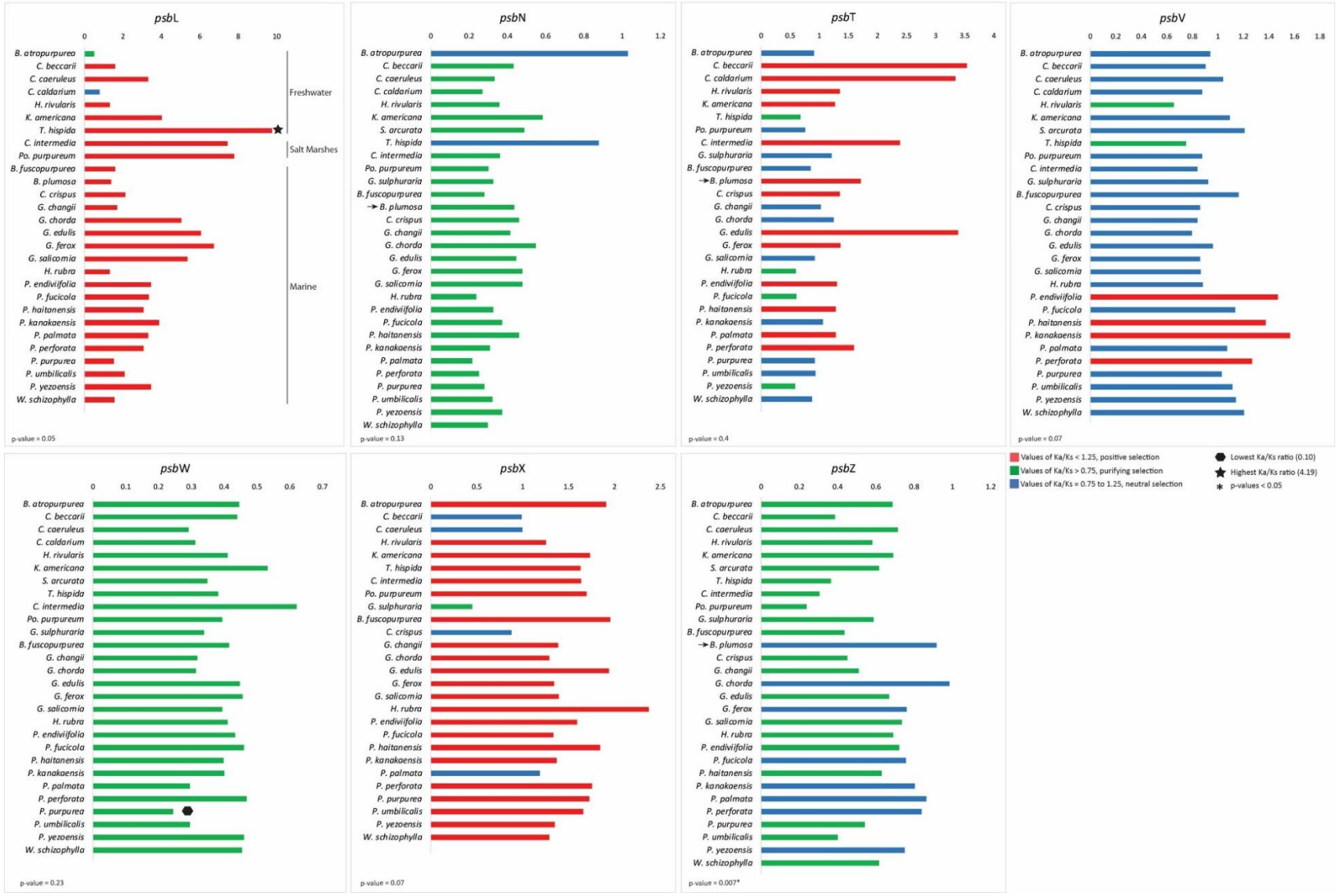

Figure F. 2. Comparison of the substitution ω ratios of *psb* genes (L – Z) of red and green algae taking in consideration the species ecological habitat.

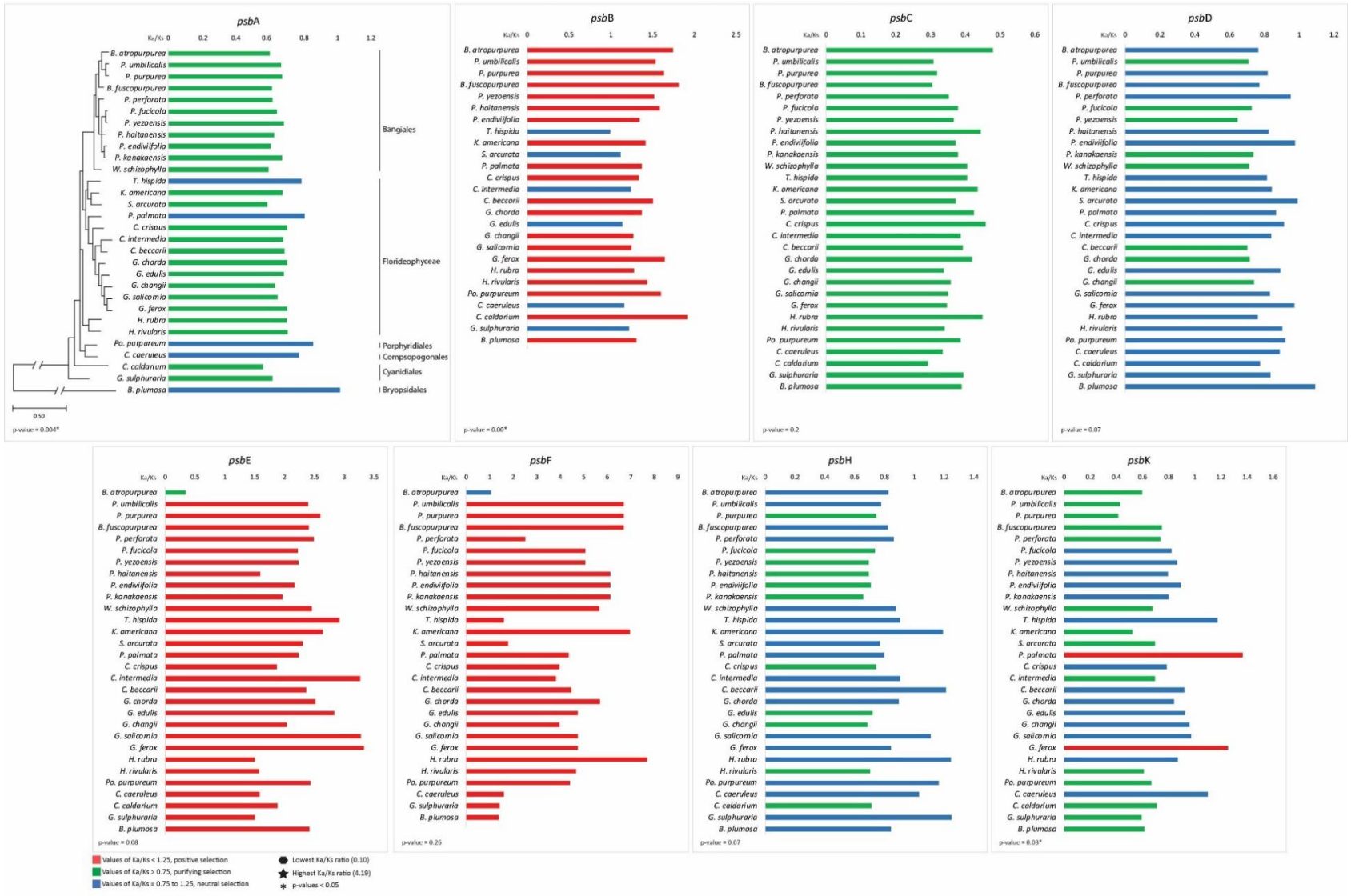

Figure F. 3. Comparison of the substitution ω ratios of *psb* genes (A – K) of red and green algae taking in consideration the phylogenetic position of the species.

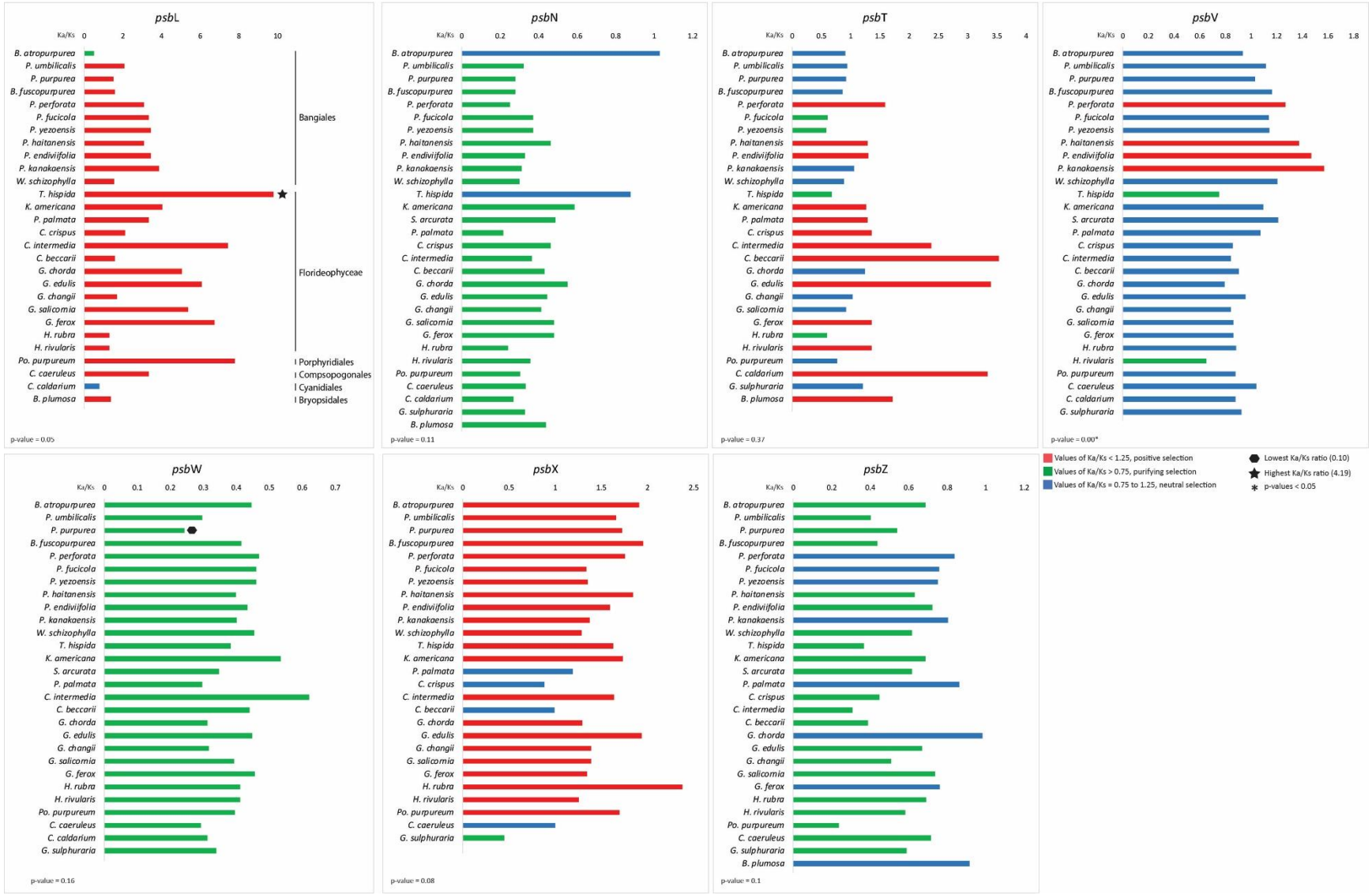

Figure F. 4. Comparison of the substitution ω ratios of *psb* genes (L - Z) of red and green algae taking in consideration the phylogenetic position of the species.

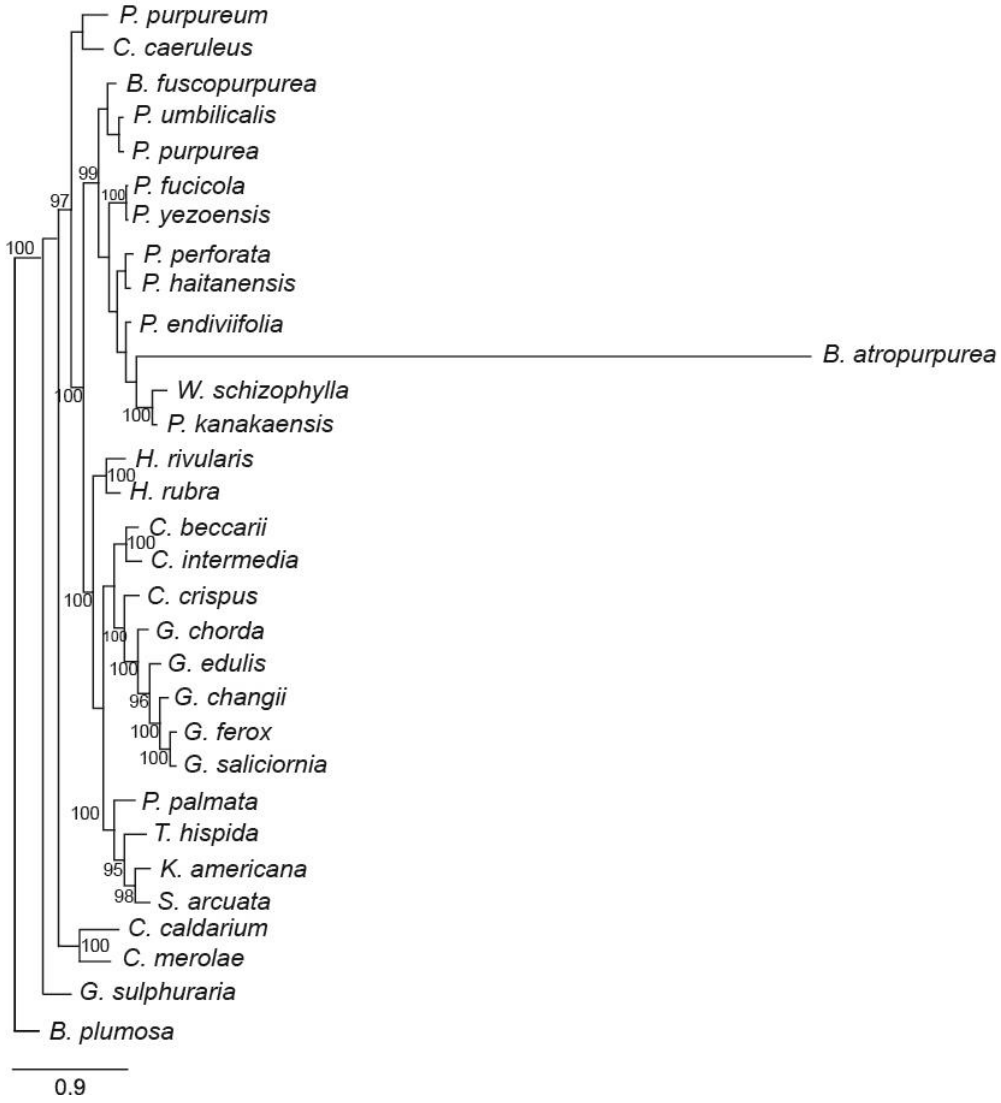

Figure F. 5. Phylogenetic tree based on concatenated sequences of 15 *psb* genes: *psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*K, *psb*L, *psb*N, *psb*T, *psb*V, *psb*W, *psb*X and *psb*Z. The tree was reconstructed using the Maximum likelihood method and invoking the GRT+G model. Numbers at internal branches represent bootstrap support (values lower than 80% not included) as a percentage of 1000 replicates.

Appendix G

Table G. 1. List of *rps* genes present or absent in Rhodophyta and Chlorophyta species

Species	rps1[X]	rps2[*+]	rps3[X]	rps4[*]	rps5[+]	rps6	rps7	rps8[X]	rps9	rps10[*]	rps11[+]	rps12	rps13	rps14[*+]	rps16[+]	rps17[*]	rps18	rps19[X]	rps20
Bangia atropurpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra umbilicalis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra purpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bangia fuscopurpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia fucicola	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia endiviifolia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia haitanensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia kanakaensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia perforata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia yezoensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wildemania schizophylla	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×
Thorea hispida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Palmaria palmata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓
Chondrus crispus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa beccarii	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa intermedia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kumanoa americana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sheathia arcuata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hildenbrandia rubra	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hildenbrandia rivularis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria changii	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria chorda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria edulis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria ferox	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria salicornia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyridium purpureum	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Compsopogon caeruleus	✓	✓	✓	✓	×	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cyanidioschyzon merolae	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cyanidium caldarium	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Galdieria sulphuraria	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bryopsis plumosa●	×	✓	✓	✓	×	×	✓	✓	✓	×	✓	✓	×	✓	×	✓	×	✓	×

* Statistical difference in different habitats; + Statistical difference in phylogenetic groups; X Genes not used in the analysis.

● Chlorophyta species

✓ Gene is present

× Gene is absent

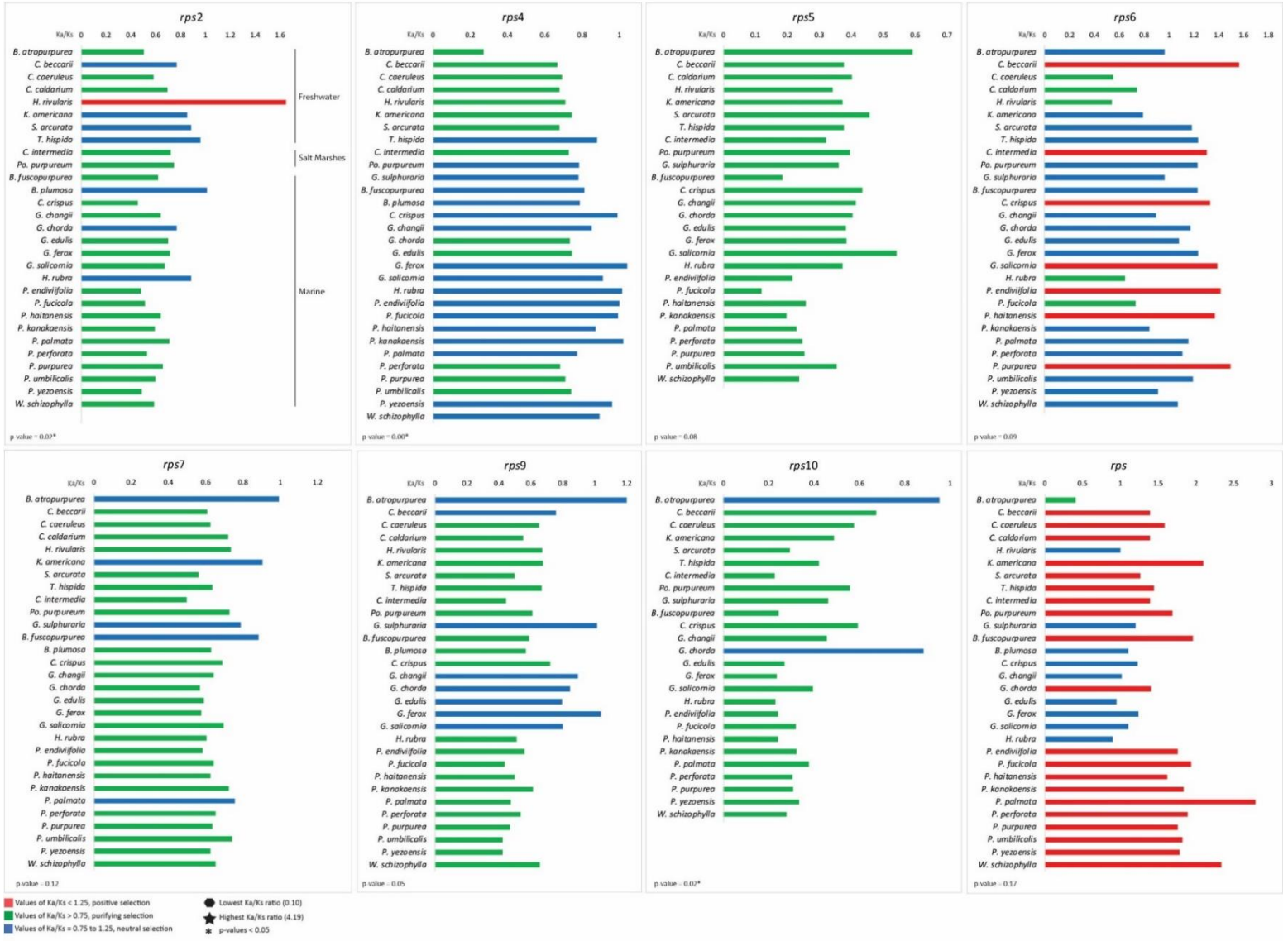

Figure G. 1. Comparison of the substitution ω ratios of *rps* genes (2 – 11) of red and green algae taking in consideration the species ecological habitat.

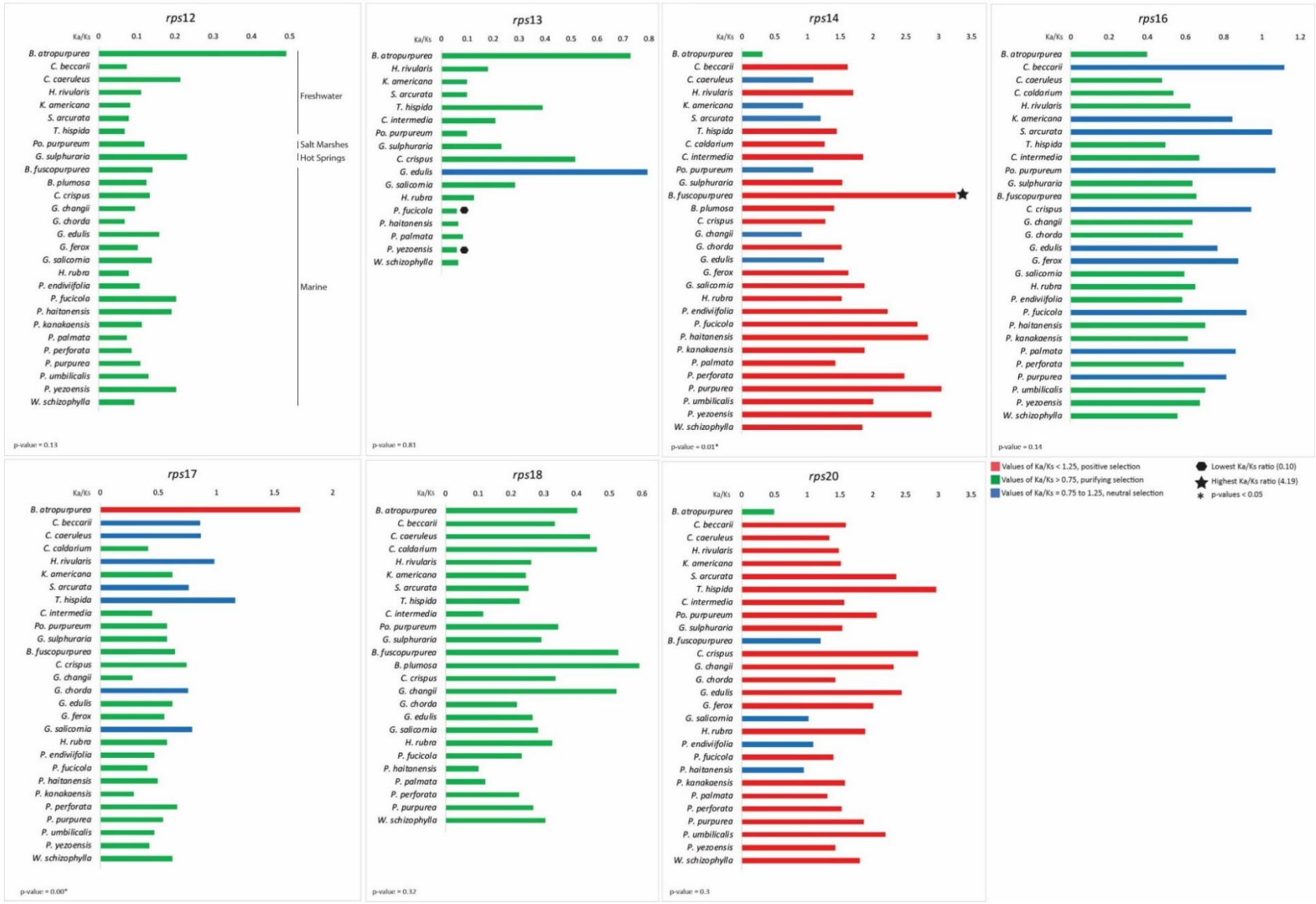

Figure G. 2. Comparison of the substitution ω ratios of *rps* genes (12 - 20) of red and green algae taking in consideration the species ecological habitat.

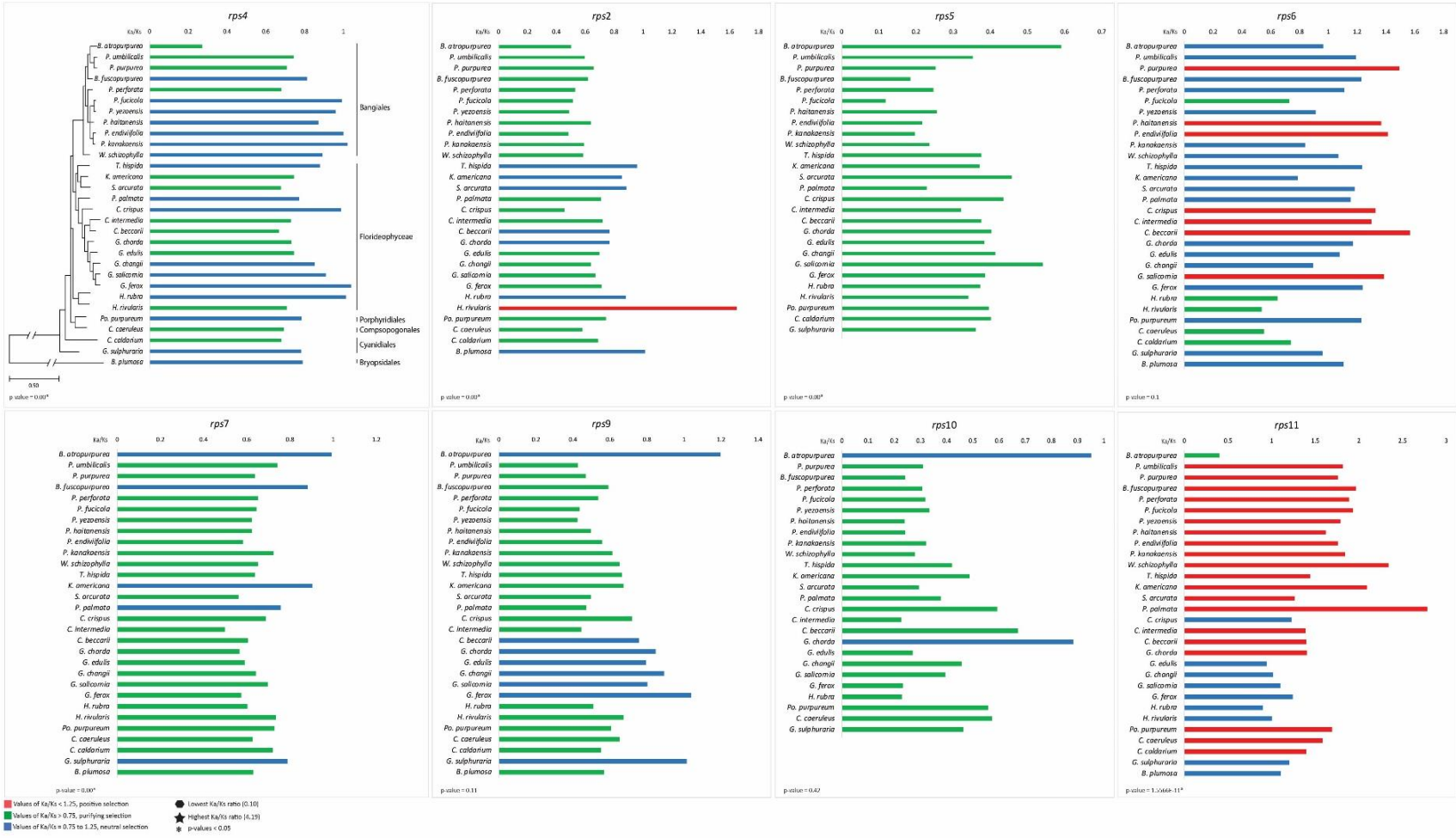

Figure G. 3. Comparison of the substitution ω ratios of *rps* genes (2 – 11) of red and green algae taking in consideration the phylogenetic position of the species.

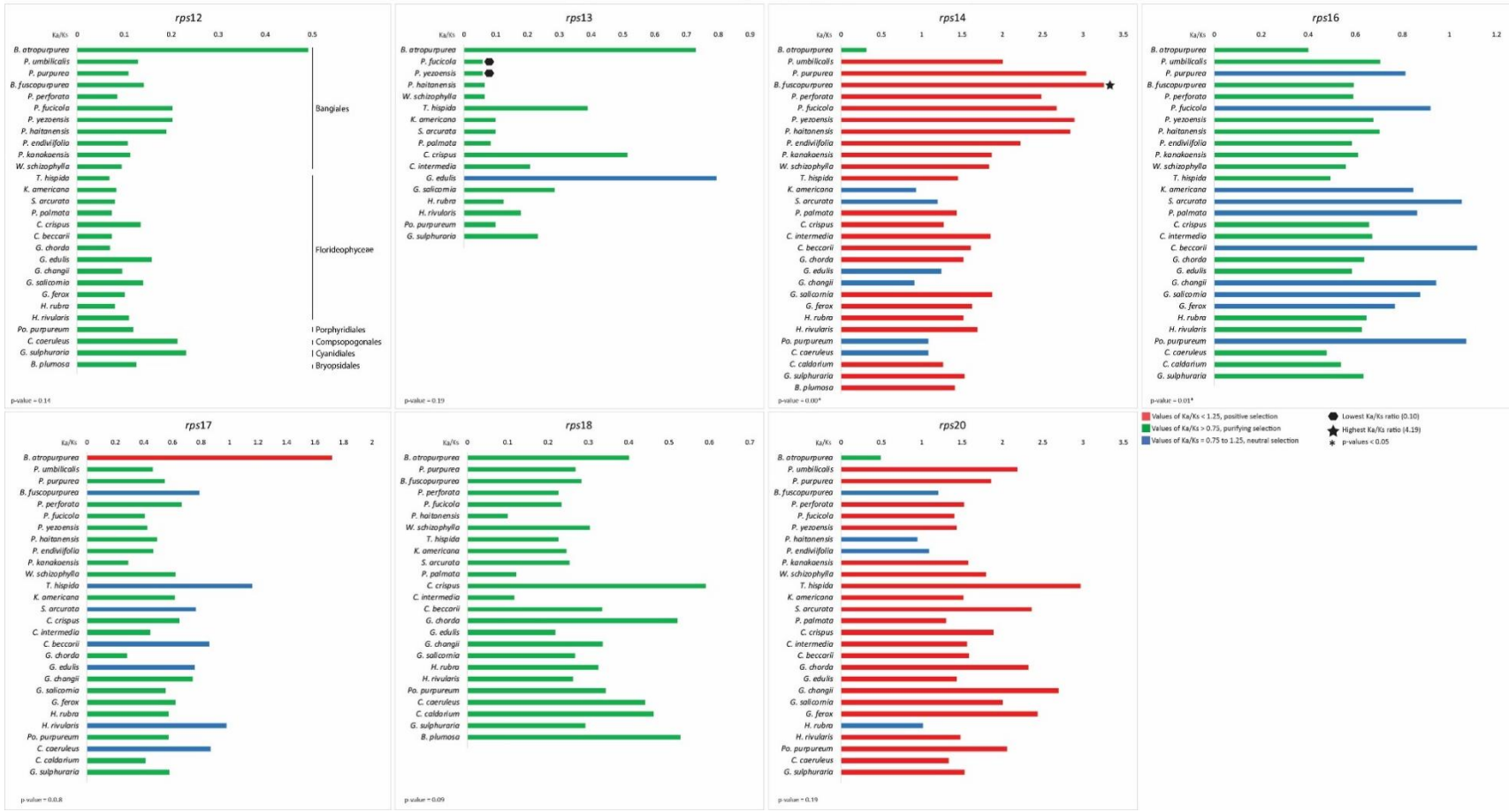

Figure G. 4. Comparison of the substitution ω ratios of *rps* genes (12 – 20) of red and green algae taking in consideration the phylogenetic position of the species

Appendix H

Table H 1. List of *rpl* genes present or absent in Rhodophyta and Chlorophyta species

Species	rpl1*	rpl2*+	rpl3	rpl4+	rpl5*+	rpl6+	rpl11+	rpl12+	rpl13*+	rpl14	rpl16	rpl18	rpl19*+	rpl20*	rpl21+	rpl22	rpl23+	rpl24*+	rpl27*	rpl28+	rpl31	rpl32	rpl33	rpl34	rpl35	rpl36
Bangia atropurpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra umbilicalis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyra purpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Bangia fuscopurpurea	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia fucicola	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia endiviifolia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia haitanensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia kanakaensis	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia perforata	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pyropia yezoensis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wildemania schizophylla	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Thorea hispida	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Palmaria palmata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Chondrus crispus	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa beccarii	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Caloglossa intermedia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kumanoa americana	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sheathia arcuata	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓
Hildenbrandia rubra	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Hildenbrandia rivularis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria changii	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria chorda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria edulis	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria ferox	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gracilaria salicornia	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Porphyridium purpureum	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Compsopogon caeruleus	✓	✓	✓	✓	×	×	✓	✓	×	✓	✓	✓	✓	✓	✓	✓	✓	×	×	✓	✓	✓	✓	×	✓	✓
Cyanidioschyzon merolae	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cyanidium caldarium	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Galdieria sulphuraria	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	×	✓	✓
Bryopsis plumosa●	×	✓	×	×	✓	×	×	✓	×	✓	✓	×	✓	✓	×	✓	✓	×	×	×	✓	✓	×	×	×	✓

* Statistical difference in different habitats; + Statistical difference in phylogenetic groups.
● Chlorophyta species
✓ Gene is present
× Gene is absent

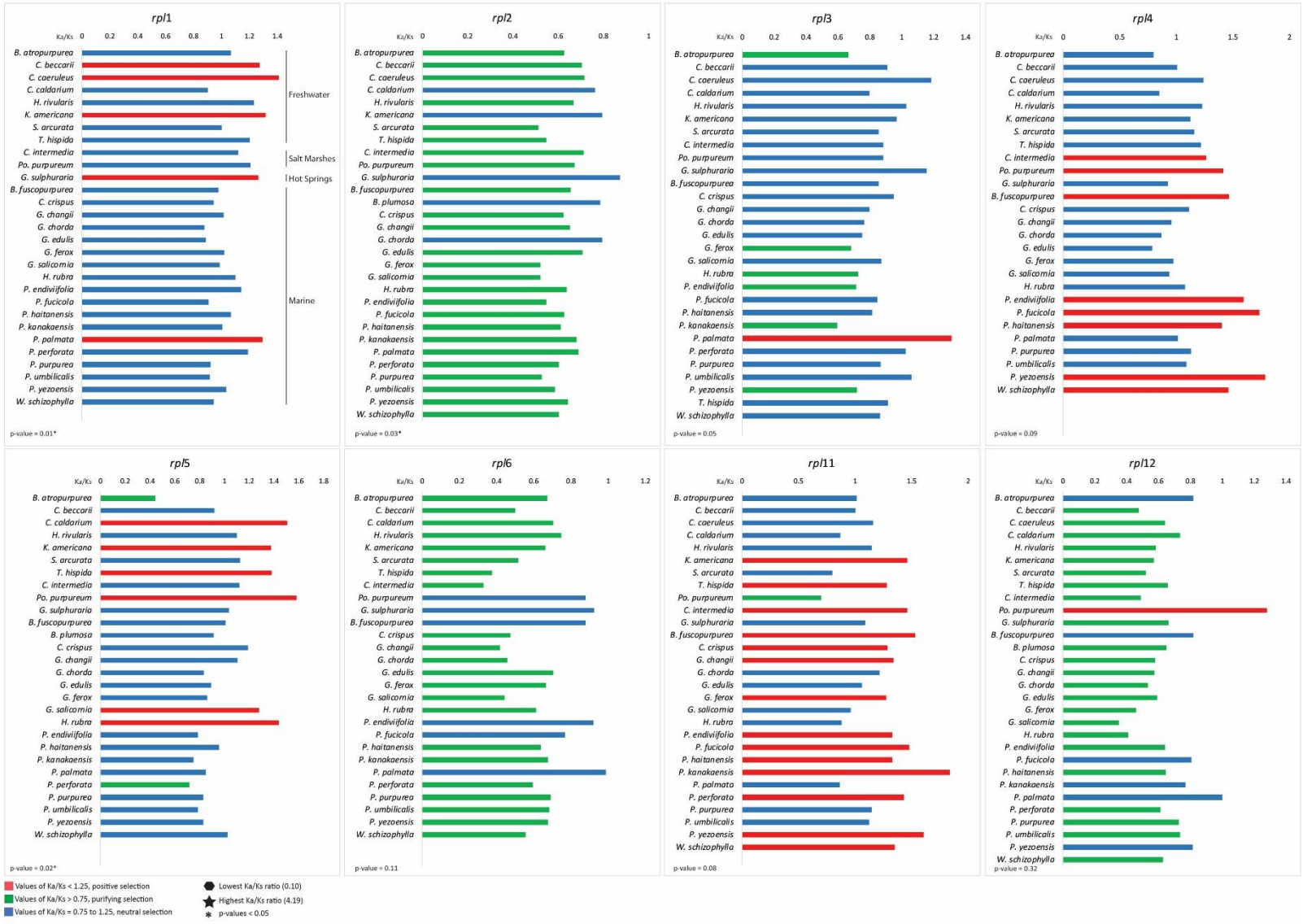

Figure H. 1. Comparison of the substitution ω ratios of *rpl* genes (1 – 12) of red and green algae taking in consideration the species ecological habitat.

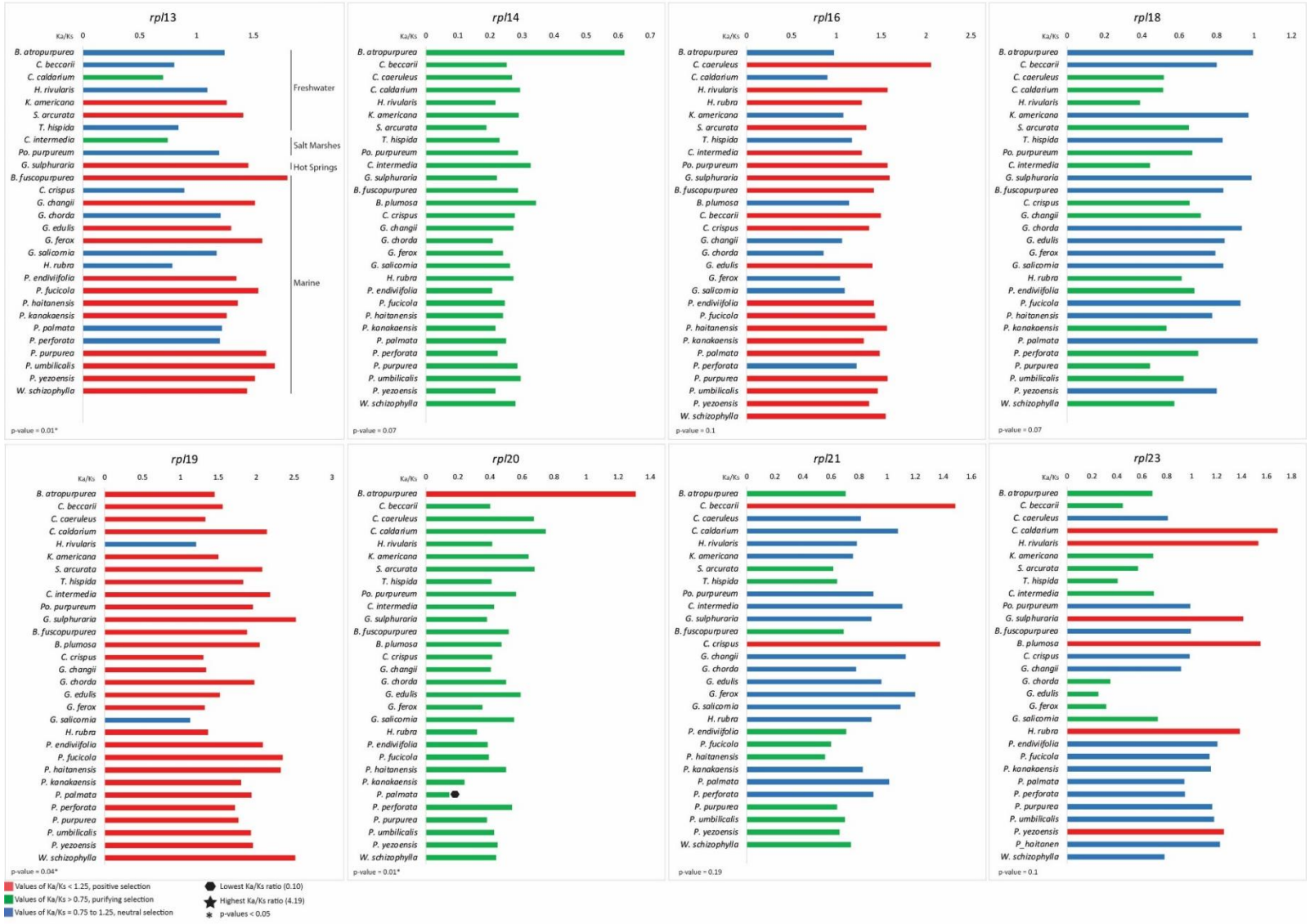

Figure H. 2. Comparison of the substitution ω ratios of *rpl* genes (13 – 23) of red and green algae taking in consideration the species ecological habitat.

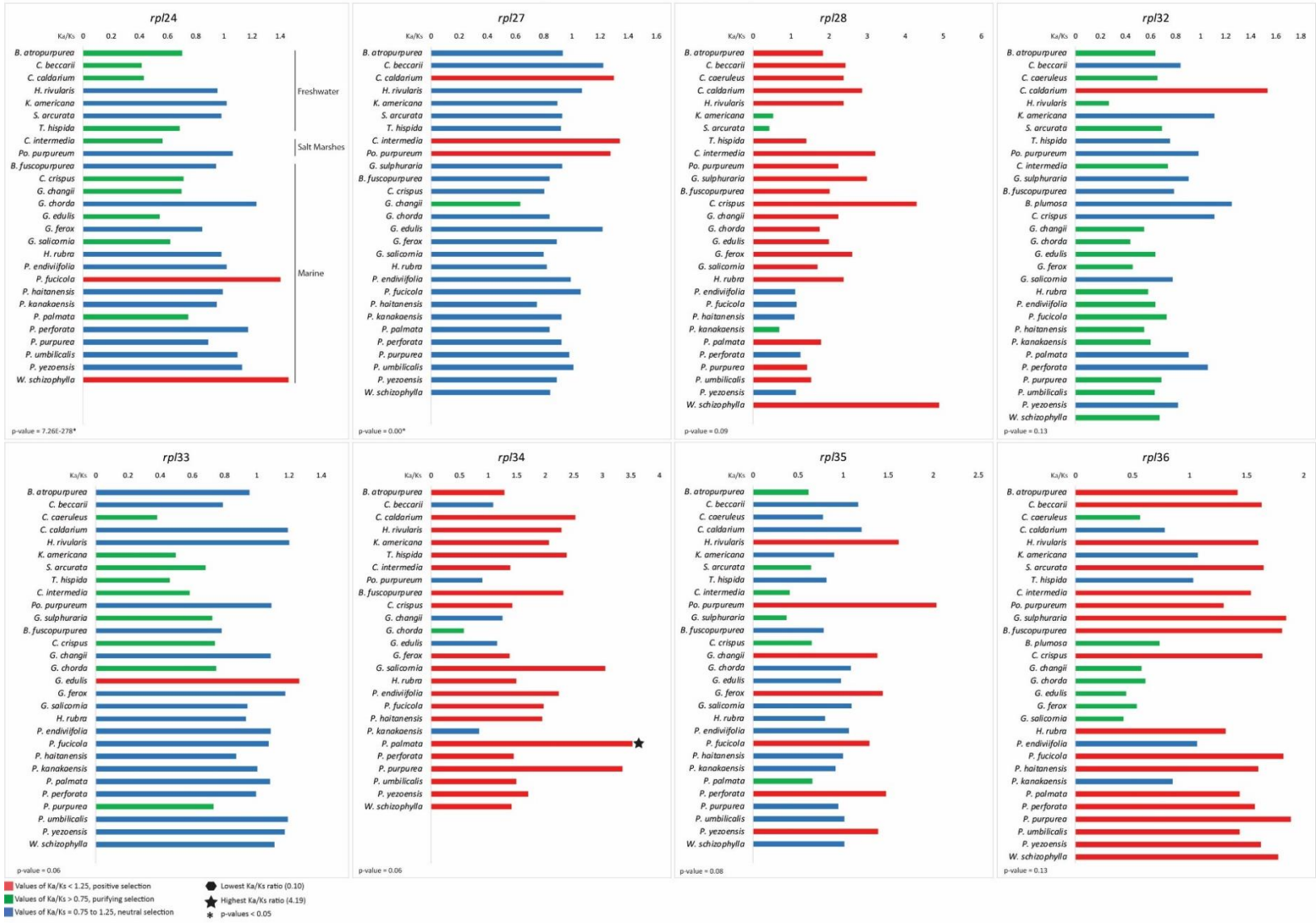

Figure H. 3. Comparison of the substitution ω ratios of *rpl* (24 – 36) genes of red and green algae taking in consideration the species ecological habitat.

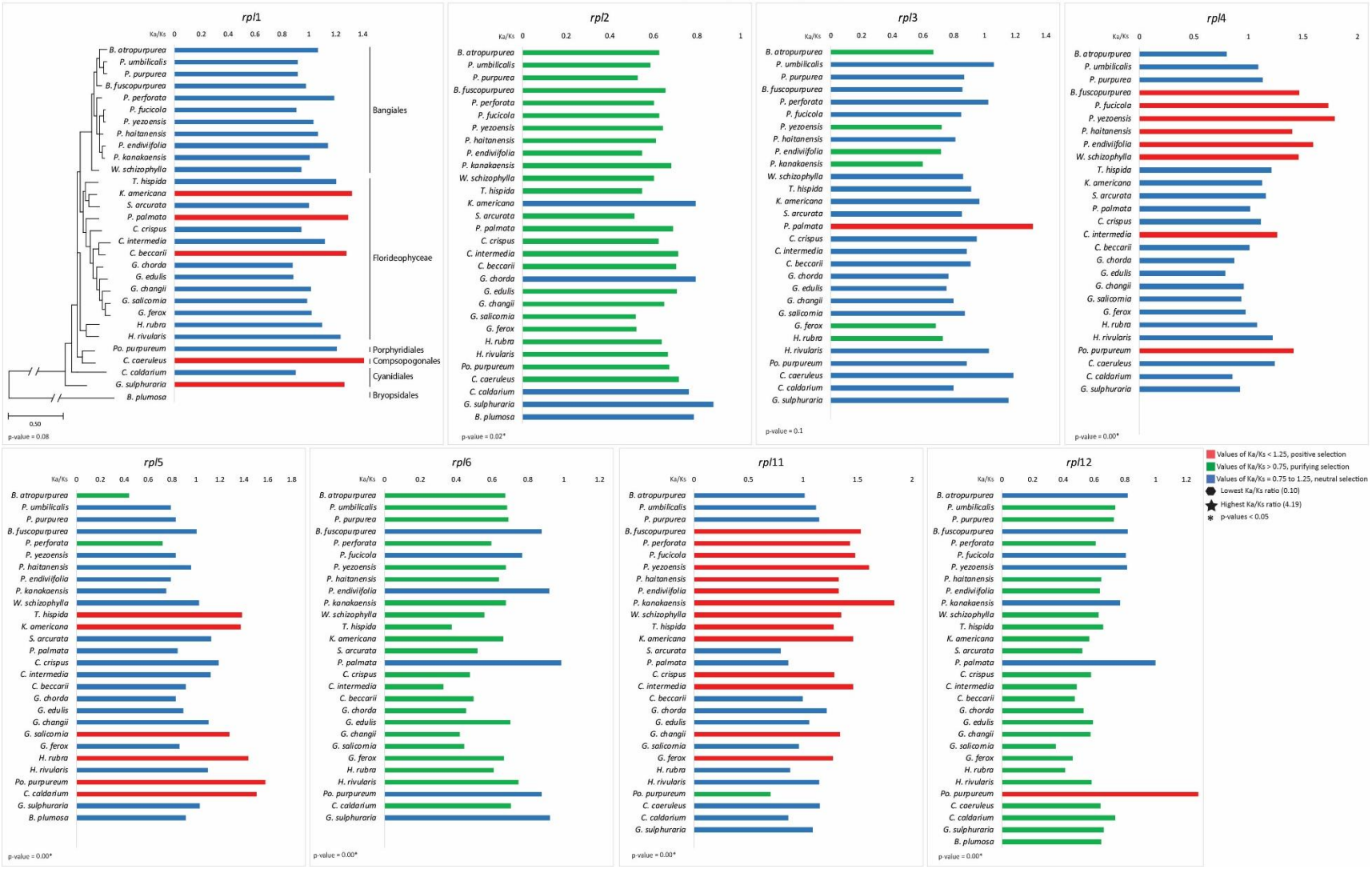

Figure H. 4. Comparison of the substitution ω ratios of *rpl* genes (1 – 12) of red and green algae taking in consideration the phylogenetic position of the species.

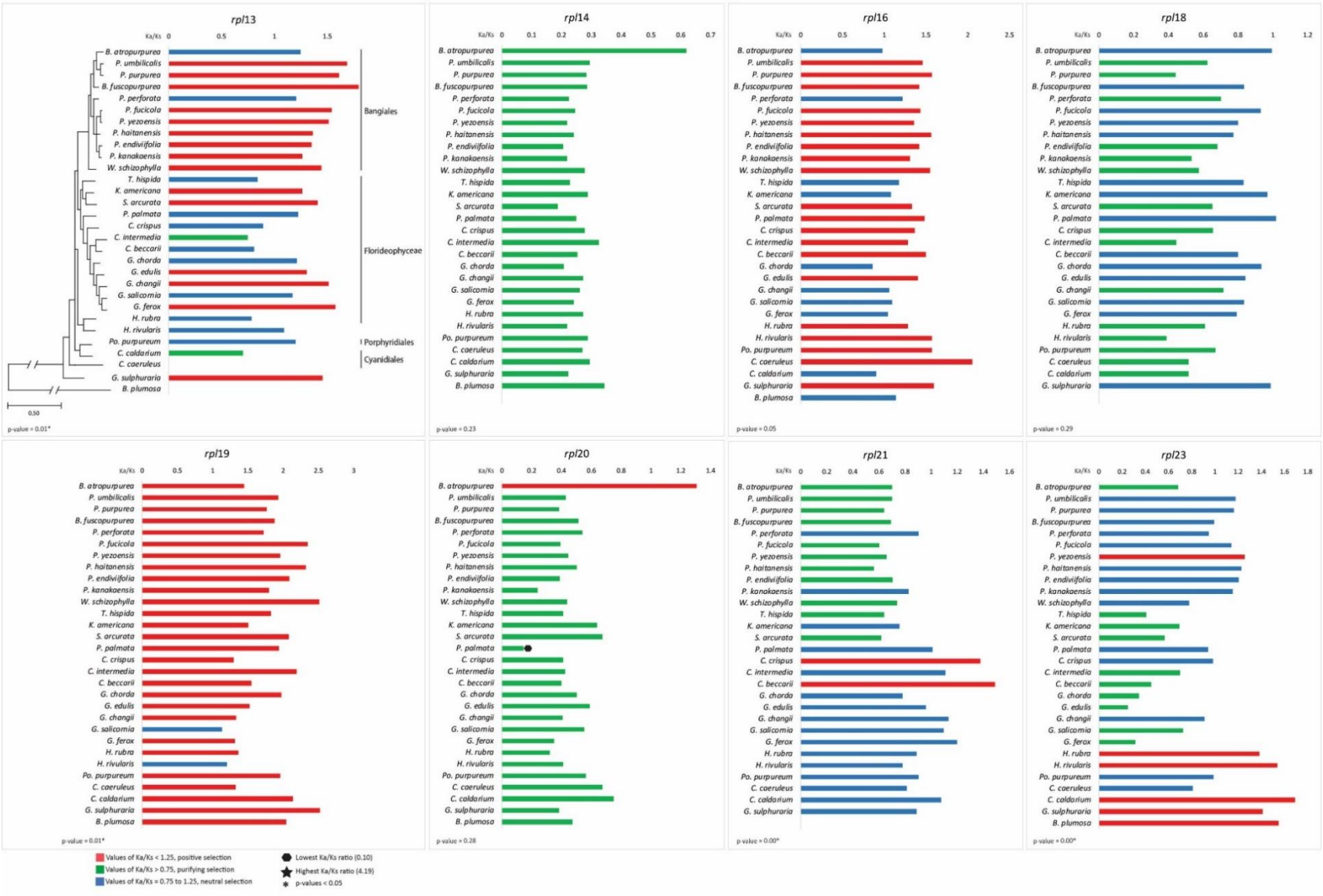

Figure H. 5. Comparison of the substitution ω ratios of *rpl* genes (13 – 23) of red and green algae taking in consideration the phylogenetic position of the species.

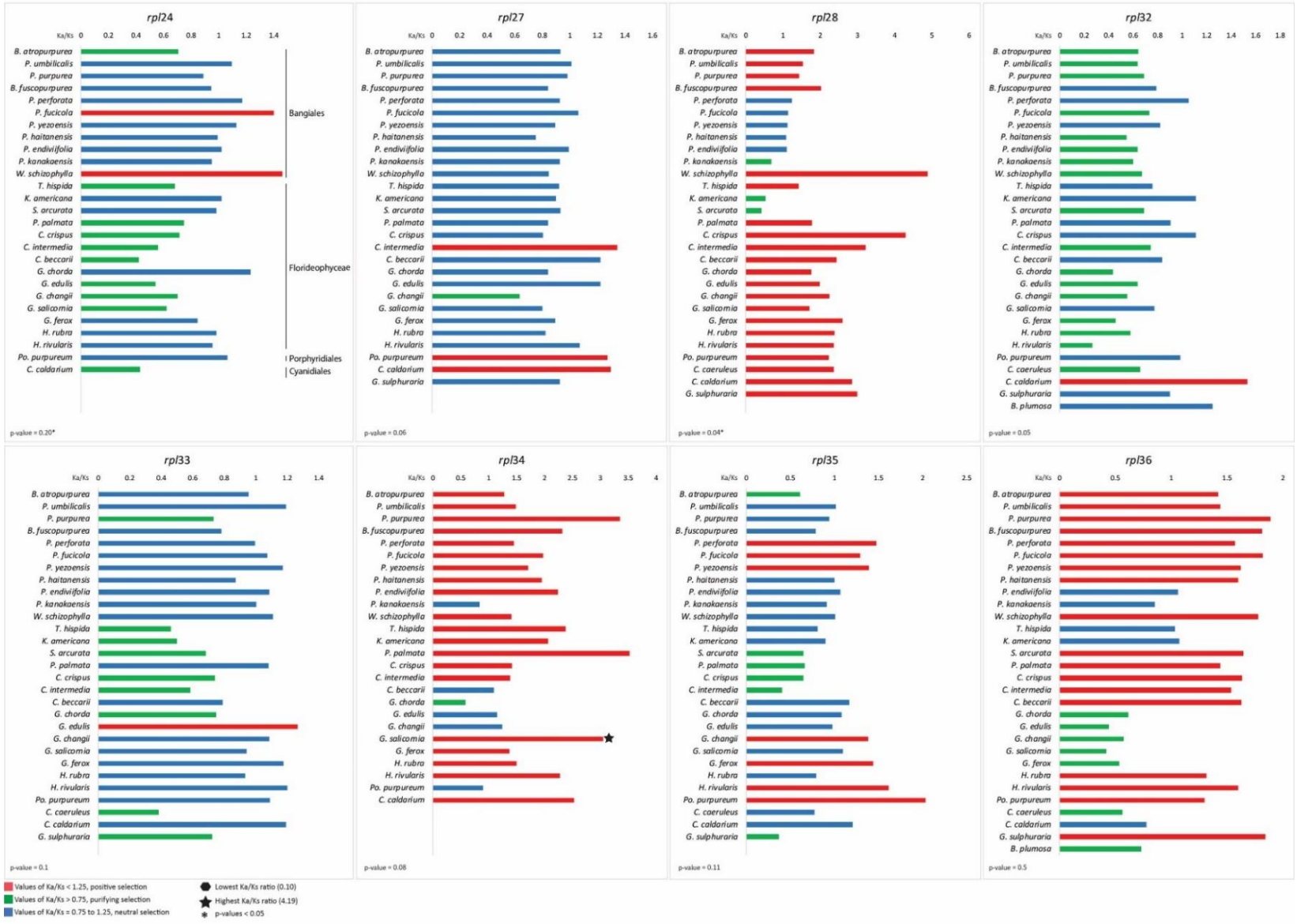

Figure H. 6. Comparison of the substitution ω ratios of *rpl* genes (24 – 36) of red and green algae taking in consideration the phylogenetic position of the species

Appendix I

Table I. 1. Summary of the chloroplast genes analyzed in this chapter for its rates of substitution among the Rhodophyta*, with their respective protein and/or role in the organism.

Gene	Gene family	Protein / Role in plant model	References
atp A		Chloroplast envelope membrane protein	Drapier et al., 1998
atp B		Beta (β) subunit of ATP synthase enzyme; used as marker gene	Hoot et al., 1995
atp D		Delta (δ) subunit of the chloroplast ATP synthase (cpATPase)	Maiwald et al., 2003
atp E	ATP synthase	Epilson (ε) subunit of the chloroplat ATP synthase; required for the CF1 complex in the thylakoid membrane	Robertson et al., 1990
atp F		H+-ATP synthase, essential for electron transport and photophosphorylation during photosynthesis	Hudson & Mason, 1988
atp G		Essential for the accumulation and function of chloroplast ATP synthase	Kong et al., 2013
atp H		Encodes the casein kinase II subunit of the chloroplast ATP synthase	Drager et al., 1993
atp I		Not yet established	
psa A		Photosystem I chlorophyll apoproteins (65 kDa)	Klein & Mullet, 1987
psa B		Photosystem I chlorophyll apoproteins (70 kDa)	Klein & Mullet, 1987
psa C		Encoding the iron sulfur protein of photosystem I (PSI)	Takahashi et al., 1991
psa D		Necessary for the stability of PSI	Ihnatowicz et al., 2004
psa E		7.5 kDa peripheral protein of the photosystem I complex	Zhao et al., 1993
psa F	Photosystem I	Docking site for cytochromeb6f complex	Haehnel
psa J		Not yet established	
psa K		Not yet established	
psa L		Trimerization of the PSI complex; stabilization of *psa* H and *psa* O	Chitnis & Chitnis; Zhang & Scheller
psa M		Subunit XII of PSI	Mohanta et al., 2020
psa I		Not yet established	
psb A		Photosystem II chlorophyll polypeptide (32 kDa)	Klein & Mullet, 1987
psb B		Chlorophyll-a binding protein P5 (CP47); part of the *psb* B-*psb* T-*psb* H gene cluster	Hong et al., 1995; Johnson & Schmit, 1993
psb C		Encodes P6 (43 kDa) protein photosystem II core polypeptide; *psb* C - *psb* D operon formation	Rochaix et al., 1989
psb D		Encodes photosystem II reaction center polypeptide D2	Nickelsen et al., 1999
psb E		Alpha of the cytochrome b559 (cyt b559 - essential component of Photosystem II), catalyzes photosynthetic oxygen evolution	Pakrasi et al., 1988
psb F		Beta cytochrome b559 (cyt b559 - essential component of PSII), catalyzes photosynthetic oxygen evolution	Pakrasi et al., 1988
psb I		Stability of photosystem II and possibly also in modulating electron transport or energy transfer in this complex	Künstner et al., 1995
psb J		The *psbEFLJ* operon encodes low-molecular-weight proteins of PSII	Xion et al., 2020
psb K		Required for the stable assembly and/or stability of the PSII complex and essential for photoautotrophic growth	Takahashi et al., 1994
psb L	Photosystem II	Not yet established	
psb H		Encode small PSII subunits; necessary for phototrophic growth, and it may have a role in the assembly or stability of the PSII complex	O'Connor et al., 1998; Summer et al., 19977
psb N		Assembly factor for PSII	Torabi et al., 2014
psb T		Encode small PSII subunits; equired for maintaining optimal PSII activity under adverse growth conditions	Monod et al., 1994
psb V		Not yet established	
psb W		Important for the assembly and/or stability of the PSII-LHCII supercomplexes in the grana regions of the thylakoid membrane	Plöchinger et al., 2016
psb X		Not yet established	
psb Y		Stabilize the binding of PsbE and PsbF to the heme group by providing a shelter for Cyt *b* 559 from oxidizing compounds	Plöchinger et al., 2016
psb Z		PSII core complexes subunit in the interaction with the peripheral antenna complexes	Swiatek

Gene	Gene family	Protein / Role in plant model	References
rps 1		Required to activate the expression of *HsfA2* (heat stress transcription factor A-2)	Zhang et al., 2020
rps 2		Encodes ribosomal protein S2 of the 30S ribosomal subunit	Plader & Sugira, 2003
rps 3		Not yet established	
rps 4		Is one of the primary rRNA-binding proteins that initiate assembly of the 30S subunit in prokaryotes and plant chloroplast	Zhang et al., 2010
rps 5		Affects proteins involved in photosynthesis and cold stress responses to mediate plant growth and development	Zhang et al., 2016
rps 6		Role in chloroplast development under cold stress	Wang et al., 2017
rps 7		Not yet established	
rps 8		Contributes to cold-adaptability	Zhang et al., 2020
rps 9		Not yet established	
rps 10	Small Subunit Ribosomal Proteins	Not yet established	
rps 11		Marker gene, a useful source for phylogenetic analysis; function not yet established	Jabeen et al.,2012; Mahmood et al., 2011
rps 12		Not yet established	
rps 13		Not yet established	
rps 14		Part of the *psa* A-*psa* B-*rps* 14 operon*;* essential ribosomal subunit for plastid translation	Sun et al., 2018
rps 16		Has lost its functionality many times independently in Viridiplantae	Keller et al., 2017
rps 17		Not yet established	
rps 18		Required for cell survival	Rogalski et al., 2006
rps 19		Not yet established	
rps 20		Not yet established	
rpl 1		Embrio viability and normal embtyo developtment	Hashimoto et al., 2005
rpl 2		Essential for peptidyltransferase activity	Randolph-Anderson et al., 1989
rpl 3		Increase the stability of the inter subunit bridges (along with *rpl* 19 and *rp* l14); shown to be involved in disease resistance	Harms et al., 2001; Di & Turner, 2005
rpl 4		Essential for embryo development	Bryant et al., 2011
rpl 5		Cluster with rpl16, rps3 and rps14. Role not yet established	Yet & Abel, 1993
rpl 6		Essential for embryogenesis	Hsu et al., 2010
rpl 9		Not yet established	
rpl 11		Conformational stabilisation of the GTPase-associated site in the large subunit; related to growth in prokaryotes	Pesaresi et al., 2001
rpl 12		Regulation of protein synthesis; disease resistance; play a minor role in basal resistance against virulent pathogens; form dimer with *rpl* 7	Grebenyuk et al., 2009; Nagaraj et al., 2016
rpl 13		Essential for embryogenesis	Hsu et al., 2010; Romani et al., 2012
rpl 14		Involved in patterning and growth of the *Arabidopsis* ovule	Skinner et al., 2001
rpl 16		Essential for peptidyltransferase activity	Randolph-Anderson et al., 1989
rpl 18		Essential for embryogenesis	Hsu et al., 2010; Romani et al., 2012
rpl 19		Important to interlink the large and small subunits (with *rpl* 14 and *rpl* 3)	Nagaraj et al., 2016
rpl 20	Large Subunit Ribosomal Proteins	Essential for the assembly of the large ribosomal subunit in *E. coli*	Weglöhner & Subramanian, 1992
rpl 21		Essential for embryogenesis	Yin et al., 2012
rpl 22		Essential for cellular viability	Rogalski et al., 2008
rpl 23		Essential for cellular viability	Rogalski et al., 2008
rpl 24		Embryogenesis in *Arabidopsis*	Hashimoto et al., 2005
rpl 27		Essential for embryogenesis in *Arabidopsis*	Romani et al., 2012
rpl 28		Might play a major role in the recruitment of ribosomes to thylakoids.	Romani et al., 2012
rpl 29		Fulfill architectural roles by mediating contacts between the rRNA expansion segments sets	Klinge et al., 2012
rpl 31		Essential for embryogenesis	Hsu et al., 2010; Romani et al., 2012
rpl 32		Essential for cellular viability	Rogalski et al., 2008
rpl 33		Required for resistance to chilling stress	Rogalski et al., 2008
rpl 34		Cytokinin-responsive element in tobacco	Dai et al., 1996
rpl 35		Linked to embrio viability and normal embtyo developtment	Magnard et al., 2004; Hashimoto et al., 2005
rpl 36		Essential for protein biosynthesis in tobacco	Fleischmann et al., 2011

*The highlighted genes indicate 18 *B. atropurpureas*'s genes that have the most different rate of substitution when compared to other Rhodophyta species. The green highlight represents genes under purifying selection, blue represent genes under neutral selection, and red represent genes positive selection.

Appendix J

Table J. 1. Statistical analysis results of the h-test and p-value of the rates of substitution of 74 Rhodophyta chloroplast genes.

Habitat	H-test	Chi square	P-value	Reject Null Hypothesis?
apt A	5.281675	7.814727903	0.152296	Fail to Reject
atp B	10.80957	7.814727903	0.012801	Reject Null
atp D	3.77281	7.814727903	0.287065	Fail to Reject
atp E	63833.88	7.814727903	0	Reject Null
atp F	3.588381	7.814727903	0.309479	Fail to reject
atp G	7.218725	7.814727903	2.19E-19	Fail to reject
atp H	9.738828	7.814727903	0.020922	Reject Null
atp I	8.334073	7.814727903	0.039589	Reject Null
psa A	12.24963	7.814727903	0.006575	Reject Null
psa B	6.11645	7.814727903	0.10608	Fail to Reject
psa C	3.913371	7.814727903	0.270972	Fail to Reject
psa D	11.45309	7.814727903	0.009512	Reject Null
psa E	0.715756	7.814727903	0.869492	Fail to Reject
psa F	7.740795	7.814727903	0.051684	Fail to Reject
psa I	3.210401	7.814727903	0.360309	Fail to Reject
psa J	4.544711	7.814727903	0.208337	Fail to Reject
psa K	7.669261	7.814727903	0.053365	Fail to Reject
psa L	13.46397	7.814727903	0.003733	Reject Null
psa M	16.99886	7.814727903	0.000707	Reject Null
psb A	5.270022	7.814727903	0.15306	Fail to Reject
psb B	4.836468	7.814727903	0.184171	Fail to Reject
psb C	3.581722	7.814727903	0.310317	Fail to Reject
psb D	4.263771	7.814727903	0.234354	Fail to Reject
psb E	8.492406	7.814727903	0.036859	Reject Null
psb F	4.64008	7.814727903	0.200131	Fail to Reject
psb H	10.14986	7.814727903	0.017334	Reject Null
psb K	6.666803	7.814727903	0.083311	Fail to Reject
psb L	7.534036	7.814727903	0.05669	Fail to Reject
psb N	5.528331	7.814727903	0.136954	Fail to Reject
psb T	2.936842	7.814727903	0.401466	Fail to Reject
psb V	6.864014	7.814727903	0.076361	Fail to Reject
psb W	4.247658	7.814727903	0.235933	Fail to Reject
psb X	6.775342	7.814727903	0.079414	Fail to Reject
psb Z	11.85743	7.814727903	0.007888	Reject Null
rps 2	7.820853	7.814727903	0.020032	Reject Null
rps 4	14.49661	7.814727903	0.002302	Reject Null
rps 5	6.716243	7.814727903	0.081514	Fail to Reject
rps 6	6.266825	7.814727903	0.099326	Fail to Reject
rps 7	5.722151	7.814727903	0.125939	Fail to Reject
rps 9	7.529616	7.814727903	0.056802	Fail to Reject
rps 10	9.259449	7.814727903	0.026033	Reject Null
rps 11	4.938548	7.814727903	0.176353	Fail to Reject
rps 12	5.645793	7.814727903	0.130174	Fail to Reject
rps 13	0.954943	7.814727903	0.812153	Fail to Reject
rps 14	11.19546	7.814727903	0.010715	Reject Null
rps 16	5.382621	7.814727903	0.14583	Fail to Reject
rps 17	11.40281	7.814727903	0.009736	Reject Null
rps 18	3.477522	7.814727903	0.323689	Fail to Reject
rps 20	3.601686	7.814727903	0.307811	Fail to Reject
rpl 1	11.09996	7.814727903	0.011197	Reject Null
rpl 2	8.334336	7.814727903	0.039584	Reject Null
rpl 3	7.58749	7.814727903	0.055353	Fail to Reject
rpl 4	6.420205	7.814727903	0.092863	Fail to Reject
rpl 5	9.537701	7.814727903	0.022934	Reject Null
rpl 6	5.861268	7.814727903	0.118558	Fail to Reject
rpl 11	6.570269	7.814727903	0.086932	Fail to Reject
rpl 12	3.504615	7.814727903	0.320164	Fail to Reject
rpl 13	11.22469	7.814727903	0.010571	Reject Null
rpl 14	6.98654	7.814727903	0.072328	Fail to Reject
rpl 16	6.236054	7.814727903	0.100674	Fail to Reject
rpl 18	6.814925	7.814727903	0.078037	Fail to Reject
rpl 19	8.228576	7.814727903	0.041517	Reject Null
rpl 20	10.01186	7.814727903	0.018466	Reject Null
rpl 21	4.653207	7.814727903	0.199025	Fail to Reject
rpl 23	6.094807	7.814727903	0.107088	Fail to Reject
rpl 24	1276.27	5.991464547	7.3E-278	Reject Null
rpl 27	13.31192	7.814727903	0.004008	Reject Null
rpl 28	6.320863	7.814727903	0.097001	Fail to Reject
rpl 32	5.525177	7.814727903	0.137141	Fail to Reject
rpl33	7.156176	7.814727903	0.067083	Fail to Reject
rpl 34	5.488429	5.991464547	0.064299	Fail to Reject
rpl 35	6.719715	7.814727903	0.081389	Fail to Reject
rpl 36	5.520007	7.814727903	0.137447	Fail to Reject

Phylogeny	H-test	Chi square	P-value	Reject Null Hypothesis?
apt A	15.85031	11.07049769	0.007285	Reject Null
atp B	201.438	11.07049769	1.4E-41	Reject Null
atp D	575.8785	9.487729037	2.6E-123	Reject Null
atp E	648.1348	9.487729037	5.9E-139	Reject Null
atp F	16.75078	11.07049769	0.004998	Reject Null
atp G	6.40525	9.487729037	0.170859	Fail to Reject
atp H	18.20449	11.07049769	0.002701	Reject Null
atp I	7.490335	11.07049769	0.186652	Fail to Reject
psa A	10.09221	11.07049769	0.072664	Fail to Reject
psa B	14.80541	11.07049769	0.011227	Reject Null
psa C	21.46139	11.07049769	0.000663	Reject Null
psa D	9.3525	9.487729037	0.052868	Fail to Reject
psa E	10.16494	9.487729037	0.037739	Reject Null
psa F	14.17121	9.487729037	0.006768	Reject Null
psa I	17.85809	11.07049769	0.00313	Reject Null
psa J	5.024771	9.487729037	0.284765	Fail to Reject
psa K	7.164981	9.487729037	0.127422	Fail to Reject
psa L	14.72218	9.487729037	0.005314	Reject Null
psa M	12.27738	9.487729037	0.015404	Reject Null
psb A	16.88492	11.07049769	0.004723	Reject Null
psb B	22.24321	11.07049769	0.000471	Reject Null
psb C	7.160399	11.07049769	0.208982	Fail to Reject
psb D	9.847024	11.07049769	0.079688	Fail to Reject
psb E	9.569151	11.07049769	0.088405	Fail to Reject
psb F	6.420798	11.07049769	0.267399	Fail to Reject
psb H	10.03582	11.07049769	0.074227	Fail to Reject
psb K	11.99555	11.07049769	0.034849	Reject Null
psb L	10.66742	11.07049769	0.058388	Fail to Reject
psb N	8.742278	11.07049769	0.119796	Fail to Reject
psb T	5.314236	11.07049769	0.378746	Fail to Reject
psb V	19.29839	11.07049769	0.001691	Reject Null
psb W	7.806948	11.07049769	0.167201	Fail to Reject
psb X	9.736565	11.07049769	0.083053	Fail to Reject
psb Z	9.076812	11.07049769	0.106039	Fail to Reject
rps 2	20.18802	11.07049769	0.001152	Reject Null
rps 4	9.972599	11.07049769	0.076015	Fail to Reject
rps 5	12.81534	7.814727903	0.005053	Reject Null
rps 6	7.665515	9.487729037	0.104628	Fail to Reject
rps 7	8.159808	9.487729037	0.085897	Fail to Reject
rps 9	8.886274	11.07049769	0.113687	Fail to Reject
rps 10	3.831451	9.487729037	0.429297	Fail to Reject
rps 11	59.48004	11.07049769	1.56E-11	Reject Null
rps 12	8.160572	11.07049769	0.147605	Fail to Reject
rps 13	11.71126	9.487729037	0.019633	Reject Null
rps 14	17.41161	11.07049769	0.003782	Reject Null
rps 16	11.71126	9.487729037	0.019633	Reject Null
rps 17	8.149508	9.487729037	0.086253	Fail to Reject
rps 18	9.498719	11.07049769	0.090751	Fail to Reject
rps 20	7.332765	11.07049769	0.197045	Fail to Reject
rpl 1	8.200142	9.487729037	0.084516	Fail to Reject
rpl 2	13.37064	11.07049769	0.020142	Reject Null
rpl 3	7.61258	9.487729037	0.106846	Fail to Reject
rpl 4	14.19909	9.487729037	0.006686	Reject Null
rpl 5	17.82868	9.487729037	0.001333	Reject Null
rpl 6	12.44723	7.814727903	0.005998	Reject Null
rpl 11	13.36188	9.487729037	0.009637	Reject Null
rpl 12	19.29644	11.07049769	0.001692	Reject Null
rpl 13	11.16117	7.814727903	0.010886	Reject Null
rpl 14	6.758135	11.07049769	0.239258	Fail to Reject
rpl 16	11.0514	11.07049769	0.05037	Fail to Reject
rpl 18	4.929913	9.487729037	0.294565	Fail to Reject
rpl 19	14.26584	11.07049769	0.014006	Reject Null
rpl 20	6.23649	11.07049769	0.283883	Fail to Reject
rpl 21	13.77213	9.487729037	0.008059	Reject Null
rpl 23	16.48822	11.07049769	0.00558	Reject Null
rpl 24	13.06743	7.814727903	0.004493	Reject Null
rpl 27	7.144164	7.814727903	0.067442	Fail to Reject
rpl 28	9.838634	9.487729037	0.043235	Reject Null
rpl 32	10.94909	11.07049769	0.052399	Fail to Reject
rpl33	7.730018	9.487729037	0.101984	Fail to Reject
rpl 34	6.73817	7.814727903	0.080728	Fail to Reject
rpl 35	7.357251	9.487729037	0.118171	Fail to Reject
rpl 36	10.61007	11.07049769	0.059683	Fail to Reject

Appendix K

Table K. 1. Summary of the reproduction stages most common in *Porphyra* (Bangiales) species*

Reproduction phase	Terminology	Description
Foliose/blade phase	Blade archeospore	Formed by differentiation of a vegetative cell which releases a single cell product that germinates into the blade/foliose phase
(Macroscopic, gametophyte stage,assumed that this stage is haploid)	Male gametes	Products of mitosis; Male gametes in Porphyra have been referred to as antherozoids , β-spores and spermatia; male gametes result from the divisions of an original single vegetative cell. This differs from the production of spermatia in Florideophycidae
	Female gametes	Cells of the blade that differentiate to function as gametes prior to fertilization; might produce trichogynes or prototrichogynes
	Zygotospores/zygotosporangia	Mitotic divisions of the zygote result in regularly arranged, packets of diploid cells; The zygotospores are released through breakdown of the cell walls surrounding the packets, and germinate into conchocelis; After fertilization in Porphyra the zygote divides directly to produce spores
	Asexual blade phase spore	In areas of the blade where female gametes are produced, regular packets of spores are formed by a series of mitotic divisions, apparently developing without prior fertilization
	Phyllospore	Spores produced by the blade phase where ploidy level and development are unknown
	Endospores/endosporangia	Irregularly arranged spores of indefinite number encased in a distinct envelope, formed by mitotic divisions of a blade cell
	Two types of asexual spores produced by the blade phase:	
	Agamospores	Formed by mitotic cleavage of blade cells, without fertilization, and germinate into conchocelis
	Neutral spores	Formed by mitotic cleavage of blade cells, without fertilization, and germinate into blades
Conchocelis phase	Conchocelis archeospore	Formed by differentiation of a vegetative cell which releases a single cell product that germinates into the conchocelis phase.
(Microscopic, sporophyte phase, assumed to be diploid)	Conchospores/conchosporangia	Differentiated filaments which can be distinguished by colour, shape, branching. Each sporangium releases a single product which germinates to form the foliose/blade phase. Meiosis has been documented to occur when the conchospore germinates.
	Neutral conchospores	Differentiated filaments which differ from the surrounding vegetative conchocelis. Each sporangium produces a single product which germinates, sometimes in situ, to form further conchocelis.
	Protothalli	Cellular masses which do not exhibit developmental polarity and release protoplasts which develop into foliar plants.

*This table concatenates the information published by Joffé (1896); Conway (1964); Conway & Wylie (1972); Burzycki & Waaland (1987); Cole & Conway (1980); Guiry (1990); Brodie et al. (1998); Müller et al. (1998); Nelson et al., 1999; Knight & Nelson (1999).

Appendix L

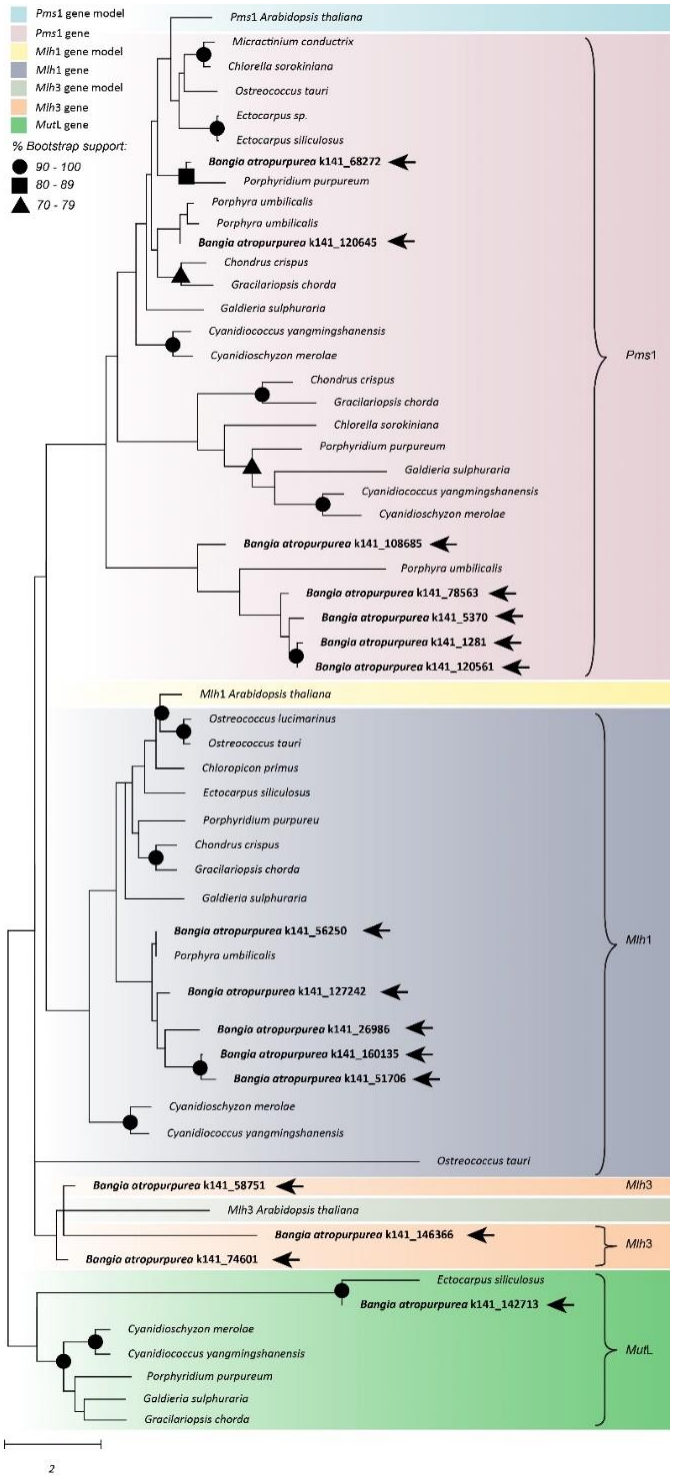

Figure L..1. Evolutionary analysis by Maximum Likelihood method of the *mlh* gene family, *pms*1, and *mut*L genes. The arrow indicates the presence of the gene in *B. atropurpurea* metagenome, and the number is the reference of the sequence fragment (contig) that was used for this analysis. Further analysis is needed to determine if the fragments are part of the same gene and to identify if the fragments are indeed part of this gene family.