Genomic analyses in kelp species

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Shivani Rana

aus

Meerut, Indien

Köln, 2020

Vorsitzender:	Professor Dr. Burkhard Becker
1. Berichterstatter:	Professor Dr. Michael Bonkowski
2. Berichterstatter:	Professor Dr. Ludwig Eichinger
Tag der mündlichen Prüfung:	27.11.2020

The present research work was carried out from February 2017 to August 2020 at the Centre for Biochemistry, Institute for Biochemistry I, Medical Faculty, University of Cologne, Cologne, Germany, under the supervision of Professor Dr. Gernot Glöckner.

Die vorliegende Arbeit wurde in der Zeit von Februar 2017 bis August 2020 unter der Anleitung von Professor Dr. Gernot Glöckner am Biochemischen Institut I der Medizinischen Fakultät der Universität zu Köln, Köln, Deutschland, angefertigt.

This work is dedicated to my family and friends

Acknowledgement

First of all, I would like to thank Prof. em. Dr. Angelika A. Noegel and Prof Dr. Henning Walczak for providing me the opportunity to carry out my PhD at their institute. Their helpful feedback and constant encouragement have contributed greatly to the quality of this work.

I want to express my deepest and sincere gratitude to my supervisor Prof. Dr. Gernot Glöckner for his sound guidance and continuous support throughout my PhD. work. His comprehensive and vast knowledge as well as his constructive feedback helped me to overcome numerous arising problems. Without his scientific advice, this work would not have been a success. He promptly helped me to learn various tools and tricks in the area of bioinformatics.

I would also like to thank Prof. Dr. Ludwig Eichinger for his technical suggestions during lab seminars. I extend my thanks to the various collaborators especially Dr. Inka Bartsch, Prof. Dr. Klaus Valentin and Daniel Liesner (AWI, Bremerhaven) for providing samples, valuable feedback on numerous progress updates and help me understand the kelp biology.

A very special thanks goes to my thesis committee members, Prof. Dr. Michael Bonkowski and Prof. Dr. Burkhard Becker for evaluating my dissertation.

I also want to thank Maria Stumpf, Martina Munck, Rolf Müller, Sonja Winninger and Berthold Gaßen for their excellent practical support and help in procurement and preparation of reagents and media as required for my work. I also thank Gudrun Konertz for her brilliant IT support.

My heartily gratitude goes to all former and current member of the institute, who helped me in several situations throughout this work and created a very friendly environment while working at the institute: Dr. Bhagyashri Burgute, Dr. Vivek Peche, Dr. Kathrin Schrage, Dr. Sarah Fischer, Dr. Roxana Solga, Dr. Malte Karow, Dr. Sussane Meßling, Jana Riehl, Maria Iqbal and Seema Waseem. Last but not the least, this work would not have been possible without the support of my parents, my dearly husband and friends, who gave me encouragement in the critical moments during this work, often associated with sacrifices on their end. They have been a constant source of unconditional love, faith, support, and endurance. The past year has been specially challenging for me being a first time mother, and I would not have made it this far without their backing.

Table of contents

Summary

Kelps are large brown algae in the order Laminariales. Kelp species have different distribution ranges along temperate to Polar rocky coastal lines. We sequenced DNA from three Laminaria species namely *Laminaria digitata*, *Laminaria solidungula* and *Laminaria rodriguezii. Laminaria digitata* is found in the Northern Atlantic region with a southern boundary in Brittany (France) or Massachusetts (USA) and a northern boundary in the Artic. *Laminaria solidungula* is endemic to the Artic and *Laminaria rodriguezii* is restricted to deep waters of Mediterranean Sea. Currently, not much is known about the nuclear and organellar genomes of kelp species. To initiate the analysis of sequencing data in kelp species the organellar genomes of Laminaria species were generated. The mitochondrial and chloroplast genomes of *Laminaria rodriguezii* and *Laminaria solidungula,* and chloroplast genome of *Laminaria digitata* were analysed and compared with phylogenetic trees derived from publicly available complete mitochondrial and chloroplast kelp genomes. All analysed kelp organellar genomes were found collinear, where large insertion, deletion (indels) or rearrangements were rare with some essential exceptions. *Laminaria rodriguezii* is very closely related to the North Atlantic temperate to Arctic *Laminaria digitata* according to the chloroplast and mitochondrial phylogeny. In the mitochondrial genome of *Laminaria rodriguezii* a stretch of more than 700 base pairs was found, which was not present in any other kelp sequenced so far. The translated Open Reading Frame (ORF) matches a protein coding region in the mitochondrial genome from *Desmarestia viridis,* a brown seaweed with a cold-temperate to Arctic distribution in the order Desmarestiales, which is closely related to the Laminariales. The high similarity of overlapping parts of two ORFs suggests that it originated through independent introduction, potentially by infection with similar mitoviruses, which is currently known in fungi and plants only. In the chloroplast genomes of *Laminaria solidungula* a small rearrangement at the inverted repeat regions was found. These rearrangements led to the pseudogenisation of ycf37 gene in *Laminaria solidungula,* a gene possibly required under high light conditions. This defunct gene might be one of the reasons why the habitat ranges of *Laminaria solidungula* is restricted to lowlight sublittoral sites in the incomplete lineage sorting of chloroplast genomes in kelp species. This work laid the foundation for analysis of nuclear genome (ca. 400Mb) of *Laminaria digitata*. The Single nucleotide polymorphism analysis yielded a first glimpse into the population diversity of this species. The draft genome analysis of *Laminaria digitata* will be part of the comprehensive analysis of brown algal genomes in the framework of the international Phaeoexplorer project led by the Biologigue de Roscoff in France.

Zusammenfassung

Braunalgen sind große Seetange in der Ordnung der Laminariales. Die Arten der Seetange haben verschiedene Verbreitungsgebiete von gemäßigten Zonen zu polaren felsigen Küstenlinien. Wir sequenzierten DNA von drei Braunalgenarten, nämlich *Laminaria digitata*, *Laminaria solidungula* und *Laminaria rodriguezii. Laminaria digitata* findet sich in einer Region im nördlichen Atlantik mit der Bretagne (Frankreich) oder Massachusetts (USA) als südlichen Grenze und der Arktis als nördlichen Grenze. *Laminaria solidungula* ist endemisch in der Arktis und *Laminaria rodriguezii* ist beschränkt auf die Tiefenwasser des Mittelmeers. Derzeit ist wenig über das Kern- und Organellengenom bekannt. Um die Sequenzierungsdaten in Braunalgenarten initial zugänglich zu machen und erste Resultate zu generieren, wurden Organellengenome von Laminaria-Arten erzeugt. Mitochondriale Genome und Chloroplastengenome von *Laminaria rodriguezii*, *Laminaria solidungula* und Chloroplastengenom von *Laminaria digitata* wurden analysiert und mit vollständigen Genomen der Mitochondrien und Chloroplasten von Seetangen, die öffentlich verfügbar sind verglichen. Dazu wurden phylogenetische Bäume berechnet. Es stellte sich heraus, dass alle analysierten Organellengenome der Braunalgen kollinear waren, wo große Insertionen, Deletionen (Indels) oder Rearrangements sind selten mit einigen wichtigen Ausanahmen. *Laminaria rodriguezii* steht in enger Beziehung zu der im gemäßigten bis arktischen nördlichen Atlantik vorkommenden *Laminaria digitata* gemäß der Chloroplasten- und Mitochondrienphylogenie. Im mitochondrialen Genom von *Laminaria rodriguezii* wurde ein Abschnitt von mehr als 700 Basenpaaren gefunden, der nicht in den anderen bisher sequenzierten Braunalgen vorhanden ist. Der übersetze offene Leserahmen (OLR) stimmt mit einer Proteinkodierenden Region im Mitochondriengenom von *Desmarestia viridis* überein, eine Braunalge mit einem kalt-gemäßigten bis arktischen Verbreitungsgebiet in der Ordnung der Desmarestiales, die eng verwandt mit den Laminariales ist. Die hohe Ähnlichkeit der sich überschneidenden Abschnitte der zwei OLR legt nahe, dass es durch unabhängige Insertionen hervorgebracht wurde, möglicherweise durch die Infektion mit ähnlichen Mitoviren, die gegenwärtig nur in Pilzen und Pflanzen bekannt ist. Im Chloroplastengenom von *Laminaria solidungula* wurde ein kleines Rearrangement in der inverted Repeats-Region gefunden. Dieses Rearrangement führte zur Pseudogenisierung des ycf37-Gens in *Laminaria solidungula*, einem Gen das möglicherweise unter besonders hellen Lichtbedingungen benötigt wird. Dieses nicht mehr bestehende Gen könnte einer der Gründe sein, weshalb der Lebensraum der *Laminaria solidungula* auf sublitorale Zonen mit schwachen Lichtverhältnissen im unvollständige lineage sorting beschränkt ist.

Diese Arbeit legt die Grundlage für die Analyse von Kerngenomen (ca. 400Mb) von *Laminaria digitata*. Die Einzelnukleotide-Polymorphismus Analyse ergab einen ersten Eindruck der Populationsdiversität dieser Arten. Die Genomanalysis von *Laminaria digitata* wird im Rahmen des internalen Projekts Phaeoexplorer Teil einer umfassenden Analyse von Braunlagengenomen sein, welches von Biologigue de Roscoff in Frankreich geführt wird.

1 Introduction

1.1 The key organism Algae

Algae are one of most common living organismal groups on the earth. They are eukaryotic, heterogeneous and photosynthetic organism, which contain chlorophyll and produce oxygen, like plants (Singh and Sharma 2012). Algae can be both single-celled and complex multicellular organism. They can differ in habitat, size, reproduction, physiology, organization, and biochemistry. The morphology and size of algae are very different to each other, unicellular species of algae are measuring 3-10 μm, and large water algae can grow up to 70 meter in length (El Gamal 2010). Algae are autotrophic in nature and acquire their food and energy from the environment, specifically from the sunlight. They play a crucial role in many ecosystems and help maintaining oxygen supply in the world. Algae are found everywhere on the earth, for example in sea, in lake, in river, in ponds, on trees, walls and soils, and also have symbiotic relationship with fungi. They have a wide range of habitats from fresh water, marine and brackish to the soils and rocks (Wehr, Sheath, and Kociolek 2015). Generally, they are found in saturated environment, and attached to the surface or sometimes at the air water interface. Marine algae are like plants, which play an important role in marine community, helping to improve marine primary productivity and also controlling the climate of marine system with processes like biogenic calcification, oceanic sequestration of carbon di-oxide, and release of dimethyl sulfide (Das and Mangwani 2015). Reproduction process in algae can be through asexual (vegetative) and sexual means. Algae produce large range of bioactive secondary metabolites, and important bioactive substances, such as proteins, carbohydrates, lipids, polyunsaturated fatty acids, polysaccharides, polyphenols, and sterols (Borowitzka and Moheimani 2013). The quality of proteins obtained from algae is better than the other plants sources for example- wheat, rice and beans, but not as good as animal proteins such as, milk and meat (Mendes, Lopes da Silva, and Reis 2007). Algae are used in several commercial activities (Figure 1-1), they are useful in the chemical & cosmetic industry, and are a good source of bactericidal substances. Algae are used to remove the organic compound, heavy metals and pathogens from the environment (Muñoz and Guieysse 2006). They are also used in animal feed and fertilizers industry, for example brown algae *Ecklonia radiate* (Charoensiddhi et al. 2018) playing an important role in production of functional food and dietary supplements (Borowitzka and Moheimani 2013). According to size algae are divided into two categories: microalgae and macroalgae.

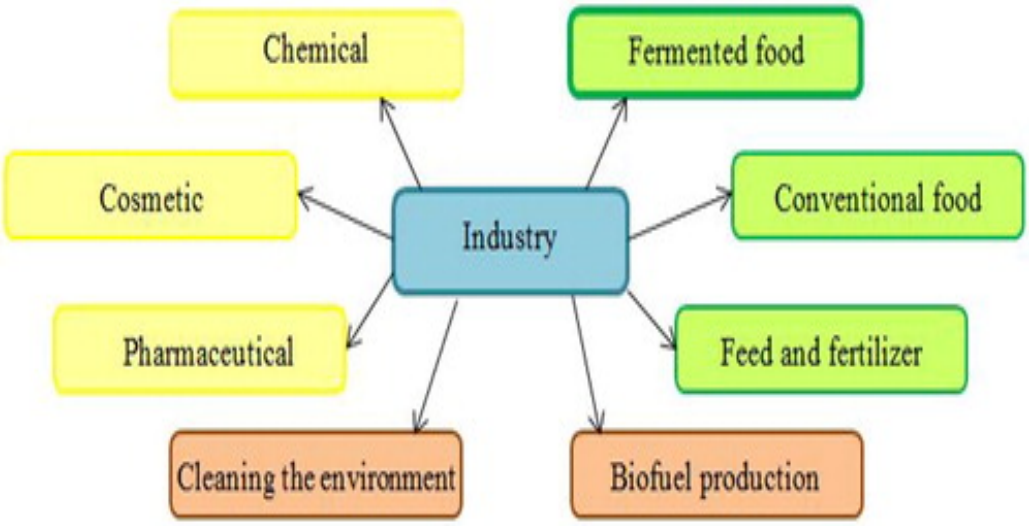


Figure 1- 1: Applications of algae. Taken from Algae in food: a general review (Ścieszka and Klewicka 2019)

1.2 Microalgae

Microalgae are small, microscopic and unicellular photosynthetic organism, for example diatoms, phytoplankton, and zooxanthellae which live in coral tissue, and need a microscope to be observed (Great Barrier Reef Marine Park Authority 2003). They are found in fresh water and in marine systems. Microalgae have no true roots, leaves and stems, they are present individually and also contain symbiotic relationship with bacteria (Barsanti and Gualtieri 2014). Microalgae are capable of producing biomass, and absorbing significant amount of carbon dioxide, because of these properties microalgae can be used as renewable source of energy (Francisco et al. 2010). Different variety of biofuel can be obtained from the different components of the algal cell wall, for example anaerobic fermentation of microalgae biomass allows production of biodiesel (Brennan and Owende 2010; Chisti 2007). Production of fuel from microalgae is not a new idea (Nagle and Lemke 1990), today it is one of the seriously considered solution for biofuel production because of the increasing price of oil, climate change and global warming, which are associated with combustion of fossil fuels (El Arroussi et al. 2017; Le Quéré et al. 2017). Microalgae can be divided into different categories according to their cytological and morphological properties, reserve metabolites, cell wall components and pigments such as marine diatoms, which are golden brown color because of xanthophyll pigments and blue green algae, which contain chlorophyll a and blue phycocyanins pigments

(Levasseur, Perré, and Pozzobon 2020). The microalgae such as *Chlorella* species, *Dunaliella* species, and *Scenedesmus* species are used as nutrient rich foods and are also sources of some chemicals.

1.3 Macroalgae (Seaweeds)

Macroalgae are multicellular, eukaryotes and photosynthetic organism with a great diversity of forms, sizes, varying pigmentation and morphology (Littler and Littler 2011) (example in Figure 1-2 and Table 1-1). Macroalgae are also known as Seaweeds, they are mostly benthic plants, and are attached to the seabed or to the reef substrate, as exception, two species of *Sargassum* (*S. natans* and *S. fluitans*) are free floating (planktonic) (Smetacek and Zingone 2013). Macroalgae can grow intertidally and subtidally. Similar to plants they execute photosynthesis process- producing Oxygen and taking carbon dioxide from the water (Littler and Littler 2011). Macroalgae being aquatic organism play an important role in marine ecosystem, where they are primary producers and provide shelter to other marine organisms for example fish, plankton, invertebrates and microorganism. Macroalgae are one of the most important organism in the world because they are the largest oxygen producer, consuming huge amount of CO_2 from the environment, and have an important range of applications in different industries all over the world (Gomez-Zavaglia et al. 2019). Macroalgae are commonly part of one of three major branches of algal evolution: Rhodophyta (red algae), Chlorophyta (green algae) and Phaeophyta (brown algae). These three categories of macroalgae differ from each other in various manners, for example the composition of cell walls, the chloroplast structure and they all have different ultrastructural, biochemical features with different photosynthetic pigments. They are an important and valuable food source, rich in proteins, polysaccharides, vitamins and minerals (Anantharaman et al. 2010; Wong and Cheung 2000), and also have antibacterial, antiviral, anti-inflammatory and antioxidant activity, which are particularly useful in medical and pharma industry to make different type of drugs. Protein content are different in each species of macroalgae such as, brown algae have low protein content in comparison of red and green algae (Mamatha et al. 2007). Some red algae for instance, *Palmaria palmate* (dulse) and *Porphyra tenera* (nori) have protein level around 37% to 47% (w/w) (Patarra et al. 2011), and green algae such as *Ulva* species around 7% to 33% (Fleurence et al. 2012), and brown algae are known to contain maximum 15% protein content, with exceptions of *Undaria pinnatifida* (wakame) and *Alaria esculenta* which have 11% to 24% and 9% to 20% (w/w) protein, respectively (Burtin 2003; Fleurence 1999; Fleurence, Morançais, and Dumay 2018). Additionally, these days

seaweed based food additives are used in the preparation of fast food (Dhargalkar and Verlecar 2009).

Figure 1- 2: Example of red, green and brown macroalgae, Taken from (de.wikipedia.org and seaweed.ie)

Table 1- 1: Different features of red, green and brown macroalgae, modified from (peddia.com)

Features	**Red algae**	**Green algae**	**Brown algae**
Habitat	Mainly marine	Mainly fresh water	Mainly marine
Phylum	Rhodophyta	Chlorophyta	Phaeophyta
Photosynthesis Pigments	Chlorophyll a, d, and phycobilins	Chlorophyll a, b, and xanthophylls	Chlorophyll a, c, fucoxanthin and xanthophylls
Organisation of cells	Unicellular forms very few	Unicellular forms abundant	Unicellular forms absent
Thylakoids	Unstacked	Stacks of 2-20 thylakoids	Three of them stacked
Motility	Sessile	Mainly motile and contain flagella	Sessile
Reproduction	Not produce motile stages during their life cycle	Produce motile sperms with multiple flagella	Produce motile sperms
Cell wall components	Cellulose and sulphated phycocolloids	Cellulose	Cellulose and non-sulphated phycocolloids

Examples	Irish moss, coralline algae, dulse (Palmaria palmate), etc.	Sea lettuce (*Ulva* species), *Codium* species, etc.	Kelp, Fucus, Sargassum, etc.

1.3.1 Rhodophyta (red algae)

Red algae or Rhodophyta are unicellular and multicellular aquatic photoautotrophic algae. Rhodophytes are red in colour, because of the r- phycoerythrin and r- phycocyanin pigments, hidden in the chlorophyll a and chlorophyll d. These pigments absorb blue light and reflect red light which makes them appear red (Cole and Sheath 1990). Red algae are mostly marine although few fresh water (for example- *Compsopogon caeruleus, Kumanoa mahlacensis, and Batrachospermum turfosum),* (Sheath and Vis 2015) and terrestrial red algae are also known (Guiry 2012). Red algae are found in the lower littoral and sublittoral zone, and sometimes at great depths in tropical seas. They have eukaryotic cells, lacking in flagella and centrioles, and their chloroplast are without endoplasmic reticulum and contain unstacked thylakoid (Stiller and Hall 1997). Over 7000 species of red algae are present in the world. Many economically important species of red algae are found in different genera for example – *Chondrus, Eucheuma, Gelidium, Gigartina, Gracilaria, Porphyra* and *Pterocladia*. Red algae can reproduce themselves both sexually and asexually. Asexual reproduction occurs through spore formation and also through vegetative like fragmentation and cell division (Norall, Mathieson, and Kilar 1981). Rhodophytes are important source of food (Nori) in Asia, because they contain high amounts of vitamins and proteins (Shibata, Jin, and Morita 1990). Many species are widely used for commercial applications since they are an important source of colloids. These colloids are used for culturing bacteria, and as suspending agents, stabilizers and moisture retainer in food industry (ice cream, chocolates and milk). Some red algae are coralline and make calcium carbonate structures which play a crucial role in formation of coral reefs (Chaudhury, Sanghvi, and Jain 2018).

1.3.2 Chlorophyta (green algae)

Green algae are photosynthetic and eukaryotic organism. They share common ancestry as plants. Green algae have carotenoid, chlorophyll a and b pigments and produced same carbohydrate in photosynthesis process like higher plants do (Rasala and Mayfield 2015). Starch is their major storage product, situated in chloroplast. There are around 9000 to 12000 types of green algae species present (Guiry 2012). Most of them (90%) are found

in fresh water and are attached to submerged rocks and wood or scum on stagnant water, and others are found in terrestrial and marine systems. They can differ in size and shape, for example single-celled, colonial, filamentous, and tubular form (Alves, Sousa, and Reis 2013). They can have sexual and asexual reproduction, sexual reproduction with gametes that have two or four flagella and asexual reproduction through cell division, fragmentation and with motile and non-motile spores (Sekimoto 2017). They are important food source for herbivorous in marine ecosystem for example fish, gastropods and sea snail. Green algae are also substantial for human being, used as food because it contains high amount of minerals like calcium, iodine, copper, magnesium, potassium and zinc (Wells et al. 2017). Economically, green algae are sustainable biomass feedstocks for the biotech industries, also including bioremediation integrated aquaculture systems and biofuel production (Neori et al. 1996, 2004; Nisizawa et al. 1987). Additionally, green algae's pigment beta- carotene is also used in food colouring.

1.3.3 Phaeophyta (brown algae)

Brown algae are multicellular, photosynthetic and large complex organism. No unicellular brown algae species is known so far. They are one of the six eukaryotes lineages that have developed true multicellularity and in terms of cell biology, they have typical features of both animals (centrosomes) and green plants (centrifugal cell formation) (Arun et al. 2019). Brown algae have secondary plastids, and they are mainly brown, yellowish-brown and olive colored due to the presence of fucoxanthin pigment (Andersen 2004). This pigment is only present in brown algae, for this reason they differ from another algae and plants, hence, they come into different kingdom called *Chromist*. Together with oomycetes and diatoms, they establish the eukaryotic lineage of Heterokonta or Stramenopiles (Yoon HS et al. 2008). This class of macroalgae (Phaeophytes) differ from most other heterokont groups, as they are comprised of- cell walls made of cellulose, alginic acid and polysaccharides, cellular inclusion of polyphenolic polymers, chloroplasts with thylakoids in stacks of three, which are enclosed by a girdle lamella and main storage product in Laminarin, a ß- 1, 3-glucan (Pueschel and Stein, 1983). Brown algae are very useful for animal and human being, as they are major source of iodine and potash. An important substrate from brown algae (cell wall) is Alginate (Barnes 1988; Draget and Taylor 2011), a colloidal gel used as stabilizer in commercial produced ice cream, beer, salad dressing, textile, paper and toothpaste industries (Gómez-Díaz and Navaza 2003; Podkorytova et al. 2007). Most brown algae are marine algae living in cold water, and shallow ocean water, very few are present in freshwater (for example- *Ectocarpus siliculosus, Pleurocladia,*

Heribaudiella fluviatilis, Porterinema fluviatile). Many are attached to the substrata. Brown algae are organized in thalli (bodies) which range from filamentous to sophisticated body plans in kelps with blades, stipes and holdfast. Since, the thallus of brown algae lacks vascular tissues, they have no true root, stem and leaves by definition (Kawai and Henry 2017). Brown algae reproduces both sexually and asexually (example in Figure 1-3). Asexual reproduction is by fragmentation and by spores. Spores are attached to rocky bottom and start to grow into haploid gametophyte plants, which undergo sexual reproduction. Sexual reproduction in brown algae are of three types- isogamous, anisogamous and oogamous (Fu et al. 2014). They are largest and fastest growing algae in the world and have lengths ranging from 1 meter (*Ectocarpus*) to 100 meters (kelps). They are dominant organism in the marine ecosystem, in terms of biomass and they often form kelp forests, that host a high level of diversity (Charrier et al. 2008).

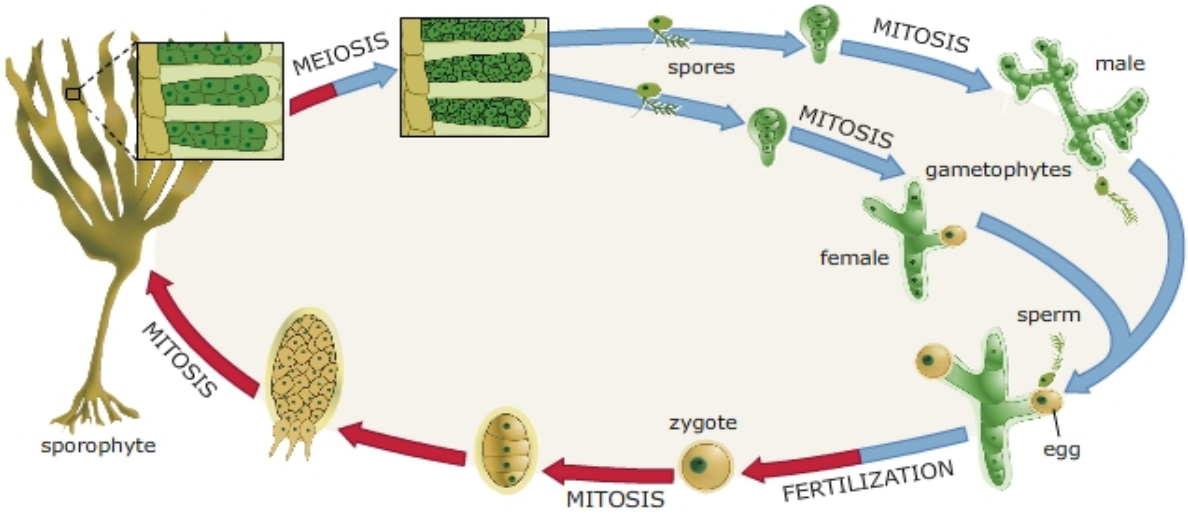


Figure 1- 3: The life cycle of brown algae is complex. The life cycle of Laminaria as example, a brown alga, which involves alternation of generations. In Laminaria, the sporophyte is a large seaweed with several leaf like blades. The gametophytes are short, branched filaments. Taken from (biological science by freeman, © 2008 Pearsons education)

1.4 Model organism Kelp

Kelps are large and complex brown seaweed in the order Laminariales, they are found in cold water. Habitat of kelps are shallow subtidal and intertidal rocks (Dayton 1985). Kelps use sunlight for photosynthesis and change carbohydrate into sugar (Bruhn et al. 2019). Kelps extract important nutrients from the surrounding water (Gerard 1982). They have the most differentiated thalli of all brown algae: they have hold fast, although it functions in a way like a root, stipe grow upward from the holdfast and blade (fronds), which are attached to the stipe (Schmid and Stengel 2015). Kelp have a biphasic life cycle, with alteration of generations among microscopic haploid gametophytes to broad diploid sporophytes (Figure 1-3) (Oppliger et al. 2014). Kelps can grow up to 60 centimeters per day and can reach up to 50 meters in length (Jackson 1987). Kelps are of great use, they are important source of iodine and potash (Cameron 1913). Commercially, kelps are used in cosmetics, food, paints, paper and skin care product industries (Hasselström et al. 2020). Numerous kelp species hold major economic importance as an essential food product in Asian countries (Cherry et al. 2019). Most of the kelp species are perennial, but some are annuals (Wernberg, Kendrick, and Toohey 2005). Many different species of kelp make up huge coastal forest, broad underwater habitats of large seaweed, which can form large canopies over the seafloor and provide food, shelter, nutrients and essential habitat to large communities of marine organisms like animals, microbes, algae and terrestrial food web (STENECK et al. 2002; Teagle et al. 2017). Kelp forest manifest an interface between ocean, environment and land masses and dominate along one-quarter of the world's coastlines, in Arctic and temperate latitudes in both hemispheres (Krumhansl et al. 2016). Kelp forest ecosystem are biogenic structure, found in the benthic marine system, their great diversity of plant and animal species support many fisheries all over the world. At each part of the kelp forest (present at ocean floor, middle part, upper part) the canopy formed by blades provide residence to many marine species of fish and mammals such as sea lion and whales (Dayton 1985). Order *Laminariales* are main kelp, which make kelp forest, currently 112 species of Laminarian kelp in 33 genera are present (Bolton 2010), one of the most important kelp is *Macrocystis pyrifera* (Macaya and Zuccarello 2010) commonly known as giant kelp, found in rocky and shallow seafloor, they grow up to 65-meter long, whereas Laminaria species only grow up to 1 to 3-meter long and growth mainly occurs from meristem region to the stipes and then blade. Laminarian kelp species have heteromorphic diplohaplontic life cycle, one is a microscopic haploid gametophyte generation and another one is a macroscopic diploid sporophyte generation, which form the kelp forests (Schiel and Foster 2006). Reproduction process starts with the formation

of reproductive tissue sporangia in reproductive blades sporophylls (*Macrocystic pyrifera, Alaria esculenta*), and flagellated zoospores are formed. The zoospores are released, scatter and settle to develop into male and female gametophytes, later, produce gametes. The sperm leave the male gametophyte and fertilize the female egg attached to the female gametophytes, and then juvenile sporophyte produce. Reproduction in kelp species depend on environment conditions, such as light, nutrients and water motion and some kelp species show seasonality in reproduction (Bartsch et al. 2013). Kelps help in purification of marine water and remove waste product which are produced by animals and living organism in the kelp forest (Smale et al. 2013).

1.4.1 Laminaria digitata

Laminaria digitata (Hudson) J.V. Lamouroux is a large brown alga, belongs to the order Laminariales, found at low water shore area, and in the sublittoral zone (Lüning and tom Dieck 1989) of the Atlantic Ocean (Dankworth et al. 2020). *Laminaria digitata* grows in the Northern Atlantic with a southern distribution boundary in France and USA (Schaffelke and Lüning 1994). They can be found growing up to 10-meter in depth, 3-meter long and 60 cm across the frond. Frond is split into 3-8 blades and *Laminaria digitata* is firmly attached to bedrock by a holdfast, and has flexible and smooth stipes (Figure 1-4). *Laminaria digitata* has haplodiplophasic life cycle, which is alternating between microscopic stages such as meiospores, filamentous gametophyte, and microscopic sporophytes, and macroscopic sporophytes (Liu et al. 2017), which can grow several meters in length. The sporophyte phase of *Laminaria digitata* grows finest at 10°C to 15° C (Dieck (Bartsch) 1992) although it reproduces at 5°C to 10°C (Bartsch et al. 2013). *Laminaria digitata* produces alginate, which acts as thickener (Vauchel et al. 2008), widely used in food, pharmaceutical, medical and paper industries. Blade of Laminaria can be used as salad and eaten directly, for example in Asian countries it is widely used as soup and food called Dashi (Kolb et al. 2004). It is also an important source of iodine, in 19th century it was commonly used as a supplement of iodine (Gall, Küpper, and Kloareg 2004) which in turn was used in goiter treatment, later used as an organic fertilizer and also in glass industries (Verhaeghe et al. 2008).

Figure 1- 4: *Laminaria digitata* with holdfast, stipes and blades. *Laminaria digitata* (Hudson) J.V. Lamouroux, 1813. Taken from (aphotomarine.com)

1.4.2 Laminaria solidungula

Laminaria solidungula is heterotrophic brown algae, found in lower depths of Arctic region. *Laminaria solidungula* is very important brown algae in Arctic community (Figure 1-5), it provides food and shelter to many marine faunas (Chapman and Lindley 1980). The thallus of *Laminaria solidungula* is medium brown to dark brown in color, stipes are 1-meter, holdfast is like suction cup and blades are split. Reproduction is the same as in *Laminaria digitata* such as alteration in generation, they release spores which get attached to the rocky bottom and germinate into male and female gametophytes. Oogamous and oogania are the sexual ways of reproduction where male gametophytes produce male gametes (antherozoids), and female gametophytes produce female egg. When favorable conditions occur male and female gametes fertilize and form zygote to produce sporophytes (Roleda 2016).

Figure 1- 5: *Laminaria solidungula* Arctic region kelp. Taken from (commons.wikimedia.org)

1.4.3 Laminaria rodriguezii

Laminaria rodriguezii is deep water Mediterranean kelp. They are found in warm deep water, down to 260-meter deep (Lüning 1990) at 13-14°C constant temperature. The knowledge of the distribution of *Laminaria rodriguezii* was reported during the Hvar fisheries biology expedition (1948-1949) (Ercegović 1960). They have a branched holdfast which are strictly attached to the rocky surface, stipes and united blades (Figure 1-6). They are closely related to Laminaria species in the Atlantic. They are endemic species with a reproduction same as in other kelp species (Žuljević et al. 2016).

Figure 1- 6: *Laminaria rodriguezii*. The Mediterranean deep-water kelp *Laminaria rodriguezii* is an endangered species in the Adriatic Sea. Taken from (ocena4future.org, Ante Žuljević et. al 2016).

1.5 Organelle genomes in kelp

1.5.1 Mitochondria genomes

Mitochondria are an important cellular structure, often called as cellular power factory. Few of the key functions of mitochondria include generating energy for various cell-functions, metabolic pathways, cell ageing and cell death. Origin of mitochondria has variations on two different theories of endosymbiosis, which are referred as the “archezoan scenario” and the “symbiogenesis scenario” (Koonin 2010). In the archezoan scenario the host of the proto-mitochondria endosymbiont was a hypothetical primitive amitochondrioal eukaryote and in the symbiogenesis scenario, a single endosymbiotic event involving the uptake of an alphaproteobacterium by an archaeal host cell which lead to the generation of mitochondria. It is clear that mitochondria came from the endosymbiosis between an archaeon and an alphaproteobacterium (Ku et al. 2015) and it is widely accepted that the origin of mitochondria was a single event that happened between 1.5 and 1.8 billion years ago, according to fossil record (Javaux and Lepot 2018) (Figure 1-7). Apart from the nucleus and chloroplast, mitochondria also contain its own genome- mitochondria DNA (mtDNA), to be found inside mitochondrial matrix. Mitochondrial DNA are only transmitted by female germ line. Mitochondria have circular and linear genome structure. Eukaryotic cells have more than hundred copies of mitochondria. These mitochondria copies have their own replication, transcription and translation (Zheng et al. 2019). Nuclear encoded gene products contribute the majority of functions in mitochondria. These gene products are produced from the nuclear genes, translated in the cytosol and later transferred into the mitochondria. The evolutionary course of mitochondria genomes among different groups of eukaryotes vary substantially and results in a large diversity of genome sizes and gene contents, which is the striking difference between mitochondria genome of plants and animals. Mitochondrial genome of animals (Wolstenholme 1992) are well organized and have compact gene content and very limited intergenic regions, in contrast plants mitochondrial genomes (Barraclough and Savolainen 2001) contain more genes and large proportion of noncoding DNA. Evolutionary rates of plants and animal mitochondrial sequences differ as much as 100-fold (Cole et al. 2018). The genome of mitochondria varies in size in different species, in human it contains 16,600 base pairs which encode 37 genes (Anderson et al. 1981), and in kelp it is circular and contain 37,000 base pairs to 38,000 base pairs. Mitochondria genome have one non coding region, which can be very useful for studying genetic variation and population structure in the control region. Mitochondria genome sequences play vital role in phylogenetic study of closely related species. Maternal lineages can be determining by mtDNA, as most of them are passed by

mother to offspring and show maternal genetics. Mitochondrial DNA structure and organization are highly conserved in mammals (Tineo et al. 2019) .

1.5.1 Chloroplast genomes

Chloroplast is semi-autonomous, intracellular eukaryotic cell organelle. It is a chlorophyll containing plastid, found in plant cells and algae, which convert sunlight into oxygen and carbohydrates in the process called photosynthesis, additionally, chloroplast is involved in the biosynthesis of fatty acids, vitamins, pigments and amino acids (Prabhudas et al. 2016). Chloroplasts came later between 1.5 and 1.2 billion years ago, but they developed through very same process like mitochondria symbiotic event (Dyall, Brown, and Johnson 2004) (Figure 1-7). In case of chloroplast, the endosymbiotic relationship was between a heterotrophic protist and a cyanobacterium. This single event emerged the eukaryotic photosynthesis and the monophyletic lineage Archaeoplastida (Gould, Waller, and McFadden 2008). It has been proven that, all plastid (chloroplast) came from primary endosymbiotic relationship, where cyanobacteria were engulfed by heterotrophic eukaryotes (Figure 1-7). They are called primary plastid and this kind of chloroplast is mainly found in red and green algae, but exceptions do exist (*Paulinella chromatophora*). In case of brown algae, the chloroplast is secondary since it is derived from the uptake of a photosynthetic red alga (Keeling 2010). Consequently, kelp have very composite chloroplasts, and they are surrounded by four membranes. Due to their origin from bacteria, chloroplasts have their own distinct genome, which is made up of single and circular DNA molecules. Plastid DNA is compact and has fast evolutionary rate. Chloroplast genome is totally different from nuclear genome in aspect of replication and mode of inheritance (Birky 2001). Chloroplast genomes mainly have a quadripartite structure and contain protein coding sequences, intergenic spacers and two inverted repeat regions, which can divide chloroplast's circular genome into long single copy region and short single copy region (Glöckner, Rosenthal, and Valentin 2000). Chloroplast genome of kelp mostly contain two copies of ribosomal RNA (rRNA) genes, tRNA genes, some prokaryotic RNA polymerase, and protein coding genes, for example Rubisco subunit, ribosomal proteins and thylakoids. The chloroplast gene content is highly conserved among the kelp, and most of them are involved in photosynthesis, transcription and translation (Wang et al. 2013). Nucleotide substitution rate in plastid genome provide the information about phylogeny in kelp at evolutionary level. These characteristics make the chloroplast genome an attractive tool for phylogeny studies. So far, only for some kelps species the analyzed and annotated chloroplast genome are available, such as

Saccharina japonica, *Undaria pinnatifida, Laminaria digitata* and *Laminaria solidungula.* The size of chloroplast genome in kelp are around 120 kbp to 150 kbp (Rana et al. 2019).

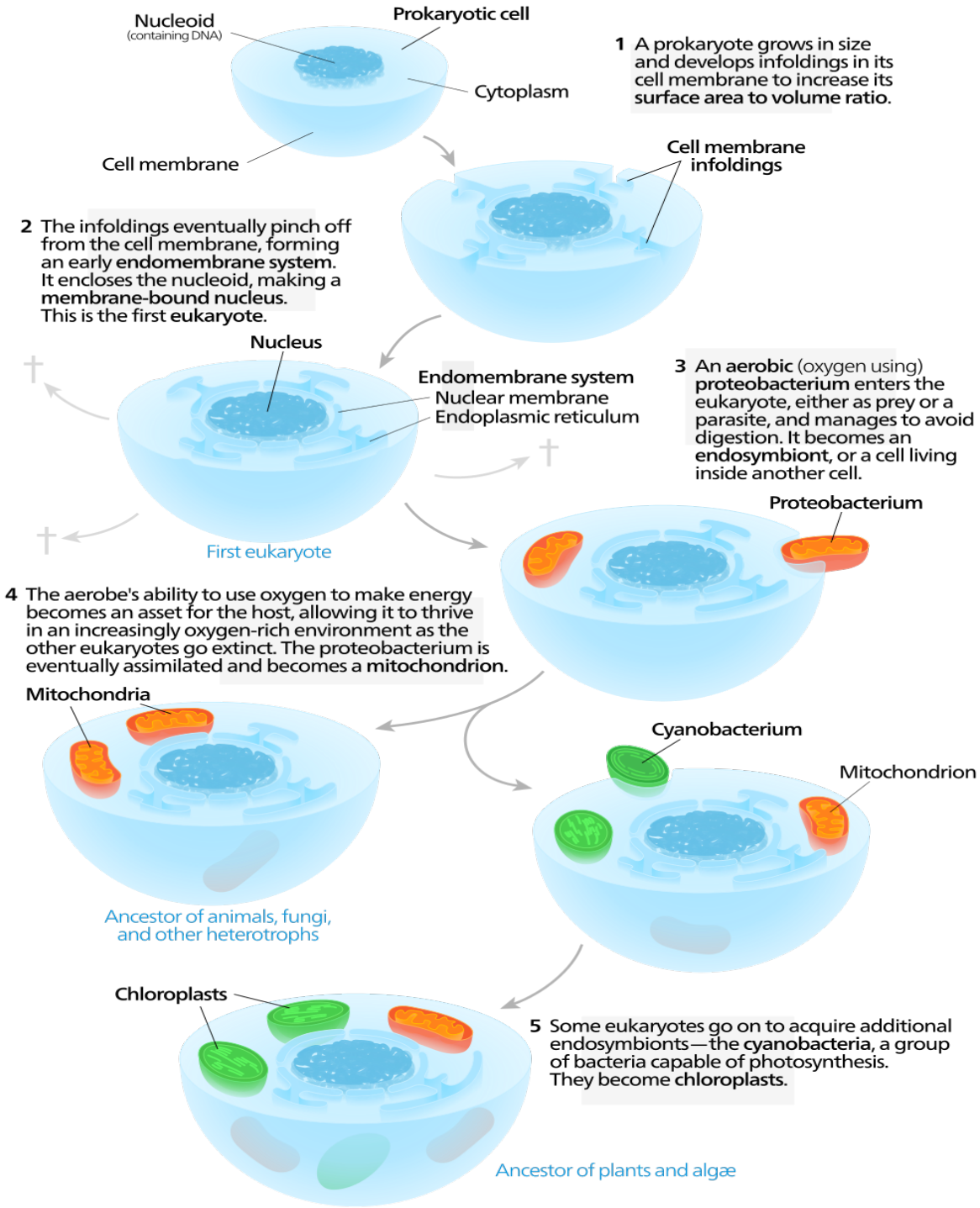


Figure 1- 7: Endosymbiotic theory: origin of mitochondria and chloroplast in eukaryotic cells. Taken from (en.wikipedia.org)

1.6 Organelle genome sequencing

In eukaryotes, the genome comprises all heritable traits of an organism. It provides essential information which are required to determine morphological and physiological traits. Sequencing of organelle genome was once a long and difficult task. Now it is commonplace with the arrival of next-generation sequencing (NGS) technologies (Bahassi and Stambrook 2014) and user-friendly bioinformatics software have made it possible to gather mitochondrial and chloroplast DNA sequences and can be used in organelle genome evolution (Gan, Schultz, and Austin 2014). Genome sequencing reveal the presence of DNA nucleotides or bases in the entire genome of organism (Sanger, Nicklen, and Coulson 1977), and this knowledge can be used in study of genomes and proteins which they encode, detection of genetic variants, which are associated with the complex molecular disease in human, mutation detection, carrier screening, detection of inherited disorders and can identified possible drug targets (Berglund, Kiialainen, and Syvänen 2011). These days, when it comes to the field of human and animal genomic research, next generation sequencing is the most efficient sequencing method. With the help of this technology 100 times more data can be produced in comparison of Sanger method based capillary sequencer. NGS technology has revolutionized the analysis of genome, transcriptomes and epigenome (Schuster 2008). In next generation Illumina sequencing process, first step is library preparation or genomics template preparation, where ultrasonic fragmentation divides genomic DNA into fragments of 200-500bp long. Adapters are added to the both 5' and 3' ends of these segments, the process called tagmentation joins both fragmentation and ligation steps in one thus, increases the efficiency of library preparation process (Figure 1-8). These adapter-ligated fragments are then amplified using PCR and purified with gel, resulting in a sequencing library. Next, is cluster generation, where flow cell acts as a channel for adsorbing mobile DNA fragments as they pass through the lanes. There are eight lanes in every flow cell and these lanes have many adapters linked to the surface, which then match to the adapters added at the ends of the DNA fragment in the building process. This is how flow cell absorbs the DNA and supports the amplification of the bridge PCR on DNA surface. Afterwards, bridge PCR is performed using the adaptors on the surface flow cell as template, as a result of continuous amplification and mutation cycles, DNA fragments are clustered at their respective locations, where each cluster holds multiple copies of a single DNA template, making them ready for next step sequencing (Shendure and Ji 2008). The sequencing method is based on sequencing-by-synthesis (SBS), where DNA polymerase, connector primers and four dNTP with base–specific fluorescent markers are added to reaction system. At the end of

the reaction all unused DNA polymerase and dNTP are eluted. Then, fluorescence excitation buffer is added, signal is excited by laser and consequently recorded by optical equipment. This optical signal is converted into sequencing base by computer analysis. For the next round of sequencing reaction the fluorescence signal is quenched by adding a chemical reagents. Next is alignment and data analysis (Figure 1-8), where all newly identified sequences are mapped to a reference genome (Carver et al. 2010). With the help of information's system the sequencing differences are identified such as single nucleotide polymorphism (SNPs), or insertion-deletion (indel) (Li, Ruan, and Durbin 2008), phylogenetic and metagenomics (Medvedev, Stanciu, and Brudno 2009). Sequencing output for genome sequencing is limited to short reads of up to several hundred base pairs. Thus, genome assembly requires further computational processing (Ariyaratne and Sung 2011). The assembly of short reads into contiguous sequences depends on an overlapping read design and high single nucleotide coverage. Computational approaches are used to create long sequences by connecting overlapping reads. Next generation Illumina sequencing gives information about gene structure, normalized coverage of all genes, information about regulatory genes, genome organization and coding sequences which determine the protein produced by each gene and deliver high accuracy, high yield of error free reads, high percentage of base calls and allows unlimited ranges for reads counting methods such as gene expression analysis (Marguerat, Wilhelm, and Bähler 2008). Genome analysis of more kelp species is a promising approach for further understanding the evolutionary history of this eukaryotic lineage.

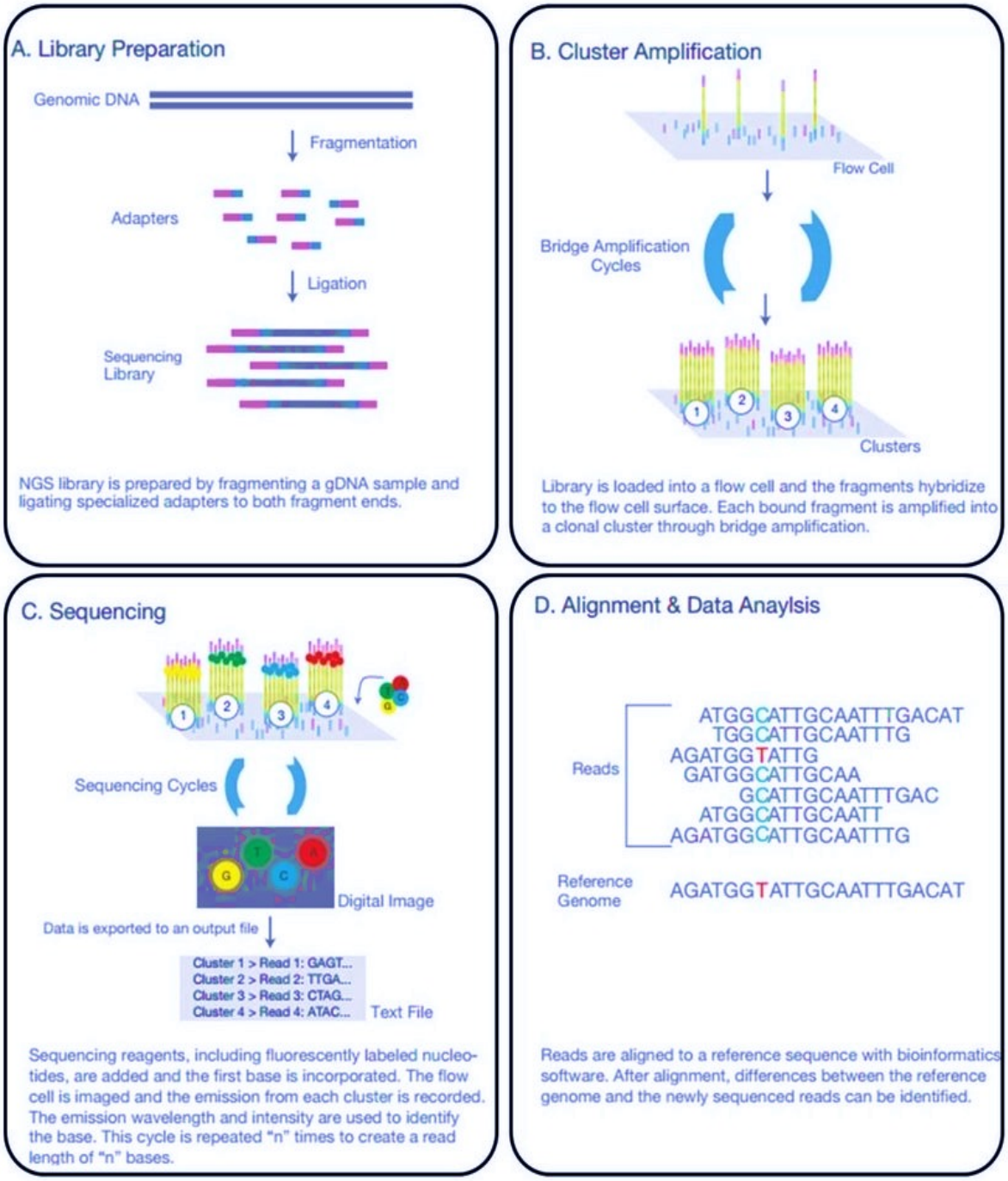


Figure 1- 8: Next-Generation Sequencing Steps—Main steps of Illumina NGS **(A)** library preparation, **(B)** cluster generation, **(C)** sequencing, and **(D)** alignment and data analysis. Taken from (Illumina.com)

1.7 Key aim of this work

The key aim of this study was to interpret the analysed sequencing data, elucidate the organellar genome architecture, phylogeny and evolutionary study in kelp species as part of the MARFOR (Functional variability and dynamics of responses of marine forests to global change) project.

The specific objectives were:

1. Detection of single nucleotide polymorphism (SNP) differences between organellar genomes of kelp species.
2. Defining the population level single nucleotide polymorphism (SNP) density in *Laminaria digitata* (together with Alfred Wagner Institute)
3. And contributing to *Laminaria digitata* genome sequencing (in the frame work of Phaeoexplorer project)

1.8 Publications

Significant part of this work has been published as research article and another is currently in pipeline to be published as listed below-

Rana, S., Riehl, J., Valentine, K., Blanfuné, A., Reynes, L., Thibaut, T., Bartsch, I., Eichinger, L., Glöckner, G., 2020. Analysis of Organellar genomes in brown algae reveals an independent introduction of similar foreign sequences into the mitochondrial genome. *Genomics* journal, accepted 19.01.2021.

Rana, S., Valentine, K., Bartsch, I., Glöckner, G. 2019. Loss of chloroplast encoded function could influence species range in kelp. *Ecology and Evolution*.2019;9:8759-8770.

2 Materials and methods

2.1 Materials

2.1.1 Kelp strains

Laminaria digitata

Clonal male gametophytes of *Laminaria digitata* were maintained at the laboratory of Alfred Wegener Institute (AWI seaweed culture number 3157), these gametophytes were originally isolated from Helgoland (North Sea). Gametophytes were cultivated at 8 - 15°C in sterilized filtered sea water in the presence of red light, which avoids differentiation and helps generate sufficient vegetative biomass for DNA extraction. Another isolates of *Laminaria digitata* were used for the analysis of population difference, which came from Connecticuts, USA (AWI culture number 3380), and Halifax, Canada (AWI culture number 3259), and non-clonal vegetative gametophyte material (mixture from both sexes), which had been derived from spores collected in September 2018 at Roscoff and Quiberon (France) were used for DNA extraction. Additional, gametophytes of *Laminaria digitata* (for transcriptomics data, AWI) were also obtained from Helgoland, Germany (AWI seaweed culture: ♀ 3436, ♂ 3435) and Kongsfjorden, Spitsbergen, Norway (AWI seaweed culture collection: ♀ 3472, ♂ 3471).

Laminaria solidungula

Laminaria solidungula sporophytes produced from gametophytes cultures, were maintained at the key laboratory of Alfred Wegener Institute (culture number 3130), which were isolated from Kongsfjorden, Spitsbergen. Fertilization of the gametophytes occurred in short day lengths 5:19 hour LD at 0°C. For further cultivation, gametophytes were transferred into 16:8 hour LD conditions at 5°C, and at a 40 µmol m^{-2} s^{-1} photon fluence rate. After reaching a size of 5 cm, sporophytes were sampled for DNA extraction.

Laminaria rodriguezii

Laminaria rodriguezii was taken from the Mediterranean Sea. Silica dried, In-situ sporophytes of *Laminaria rodriguezii* were used for DNA extraction.

2.1.2 Chemicals

1 kb Plus DNA Ladder	Promega GmbH
Agarose (Electrophoresis Grade)	Life Technologies™Corp.
β-Mercaptoethanol	Sigma-Aldrich Corp.
Cetyl trimethylammonium bromide (CTAB)	Sigma-Aldrich Crop.
Deoxyribonucleotide triphosphates (dNTPs)	Sigma-Aldrich Corp.
Ethanol (98-100 %)	Sigma-Aldrich Corp.
Ethidium bromide	Sigma-Aldrich Corp.
Ethylenediaminetetraacetic acid (EDTA)	Merck KGaA
Hydrochloric acid (37 %)	Sigma-Aldrich Corp.
Isopropanol	Sigma-Aldrich Corp.
Phenol: Chloroform: Isoamyl alcohol	Thermo Scientific Inc.
Sodium acetate	Merck KGaA
Sodium chloride	Sigma-Aldrich Corp.

2.1.3 Enzymes

Proteinase K	Merck KGaA
Ribonuclease A (RNase A)	Sigma-Aldrich Corp.

2.1.4 Kit

DNeasy PowerSoil kit	Qiagen N.V.
RNeasy® Mini Kit	Qiagen N.V.

2.1.5 Media and buffers

2.1.5.1 Media for *Laminaria digitata* gametophytes, *Laminaria solidungula* and *Laminaria rodriguezii* sporophytes culture

Provasoli Enriched Seawater Medium (PES) (Andersen and America 2005)

Enrichment Stock Solution

Tris base	5.0 g
$NaNO_3$	3.5 g
Na_2 ß-glycerophosphate • H_2O	0.5 g
Iron-EDTA solution	250 ml
Trace Metal Solution	25 ml
Thiamine • HCl (vitamin B_1)	0.500 mg
Biotin (vitamin H)	1 ml
Cyanocobalamin (vitamin B_{12})	1 ml

Iron- EDTA Solution

The below mentioned compounds were dissolved in 900 ml distilled water and solution was made to 1 liter by adding more distilled water. Later, it was autoclaved and stored at 4°C

$Na_2EDTA • 2H_2O$	0.841 g
$Fe(NH_4)_2(SO_4)_2 • 6H_2O$	0.702 g

Trace Metals Solution (from Provasoli 1968)

Na2EDTA • 2H2O	12.74 g
FeCl3 • 6H2O	0.484 g
H3BO3	11.439 g
MnSO4 • 4H2O	1.624 g
ZnSO4 • 7H2O	0.220 g
CoSO4 • 7H2O	0.048 g

2.2 Methods

2.2.1 Genomic DNA isolation of *Laminaria digitata* gametophytes, *Laminaria solidungula* and *Laminaria rodriguezii* sporophytes

The CTAB (cetyl trimethylammonium bromide) method for isolating algal DNA (Doyle 1991) was applied. Before DNA isolation, the gametophytes were washed three to six times with sterilized filtered seawater every second day to reduce the amount of bacteria in the culture. Algae samples (gametophytes and sporophytes) were prepared by cryogenically grinding the tissues in a motor pestle, then made into fine powder. 1 ml of lysis buffer was added to the homogenized sample, mixed and thoroughly vortexed and then was incubated at 70°C for 1 hour. After incubation, the homogenate was centrifuged for 10 minutes at 20000 x g at room temperature to pellet the cell debris. Supernatant was transferred into a new tube, 1 volume of Phenol: Chloroform: Isoamyl alcohol (25: 24:1) was added to it and again vortexed for 5 seconds. The sample was then centrifuged for 10 minutes at 20000 x g at 4°C. The aqueous phase was transferred into a fresh tube and DNA was precipitated with 1/10 volume of 3 M sodium acetate (pH 4.8) and 2/3 volume of isopropanol, followed by 20 minutes of centrifugation at 20000 x g at 4°C. Afterwards, pellet was washed with 70% ethanol. Consequently, the pellet was dried and alcohol was removed without completely drying the DNA. The pellet was then re-suspended in 25µl TE buffer (10mM Tris, pH 8.1, 1 mM EDTA) or in nuclease free water.

Lysis buffer

Tris/HCl, pH 8.0	100mM
NaCl	1.4 M
Na_2EDTA	20 mM
ß- Mercaptoethanol	0.2% (v/v)
CTAB	1%(w/v)

2.2.2 Sequencing

Extracted DNA of *Laminaria digitata*, *Laminaria solidungula* and *Laminaria rodriguezii* were sequenced through Illumina next generation sequencing by an Illumina HiSeq run. This technology has the advantage of using a terminator-based method which can recognize single bases incorporated into DNA template strand resulting in an accurate base-by-base sequencing without sequence context-specific errors. Approximately, 5µg of total DNA was converted to an Illumina sequencing library and analyzed on an Illumina HiSeq machine. The next step was library preparation- where all three genomic DNA of Laminaria species were first fragmented and later adaptors were attached to both 5´and 3´ end. Further, sequencing was performed, which amplified the DNA fragments by PCR process. Both, library preparation and sequencing were performed by the Cologne Center for Genomics (CCG). Furthermore, trimming and alternative clipping strategies were used to remove the lower quality bases and also to get rid of adapters with Illumina software suit.

2.2.3 Assembly and Organellar genome extraction

The raw reads were assembled by ABySS-pe (assembly by short sequences) (Simpson et al. 2009) using *k*-mer 40, 45 and 55. The Abyss algorithm contains two stages. First, all possible sub-strings of length *k*-mer are developed from sequence reads. The *k*-mer dataset removes the read errors and builds initial contigs. Secondly, mate pair information is used to enlarge contigs by determining the uncertainty in contig overlaps. The reads of *Laminaria digitata*, *Laminaria solidungula* and *Laminaria rodriguezii* contain non-plastid DNA, which are further encountered by Basic Local Alignment Search Tool (BLAST) with known plastid genome sequences of *Saccharina japonica (*JQ405663). Following that, contigs were used to generate complete circular chloroplast genomes of *Laminaria digitata* and *Laminaria solidungula.* In case of *Laminaria rodriguezii,* the known coding sequences of chloroplast (NC_044689) (Rana et al. 2019) and mitochondria (NC_004024) (Secq, Kloareg, and Goër 2002) genome of *Laminaria digitata* were used. Gaps between the contigs were filled by Gapfiller (Boetzer and Pirovano 2012). Gapfiller has satisfactory high short-reads coverage, and it produces high coverage of sufficient longer sequences. For population differences study in *Laminaria digitata*, PCR was done with forward primer TTCATCAATAAATAAAAGACCACCATTGC AT POSITION 75,636 to 75,665 and reverse primer TTCATCAATAAATAAAAGACCACCCATTGC at position 76,426 to 76,455. The resulting PCR products were ligated into pGem-T Easy vectors. To be able to discern

between polymerase errors and true SNPs, these clones from each ligation were sequenced.

2.2.4 Phylogenetic Analysis

The chloroplast coding sequences *of Laminaria digitata* and *Laminaria solidungula* were obtained from coding sequences of *Saccharina japonica* chloroplast genome, using BLAST algorithm. And, chloroplast and mitochondria coding sequences of *Laminaria rodriguezii* were obtained from coding sequences of organelle genome of *Laminaria digitata* (Rana et al. 2019; Secq et al. 2002). Mitochondrial and chloroplast nucleotide sequences were derived, and aligned gene-wise, using muscle tool (Edgar 2004). For the second publication whole organellar genomes were aligned and checked manually. The single alignment was checked manually and edited, wherever necessary. Single alignments were concatenated by SCaFoS (Roure, Rodriguez-Ezpeleta, and Philippe 2007). Afterwards, the data set was used in Molecular Evolutionary Genetics Analysis (MEGAX) software (Tamura et al. 2013) using maximum-likelihood approach and the General Time Reversible model (Nei and Kumar 2000) and 500 bootstrap repetitions were performed on the data sets. MEGAX program is useful for determining the spacing and ordering of sequence diversity in species and gene family tree. Further, MrBayes was used to evaluate the outcome of the maximum likelihood analysis (Ronquist et al. 2012a), in which evolutionary model was set to GTR with gamma distributed rate variation and a proportion of invariable sites, and used 80000 generations to get the standard deviation of split frequencies below 0.01 and a burn-in of 1000 was used.

2.2.5 Organellar genome analyses

Assembled genome of *Laminaria digitata*, *Laminaria solidungula* and *Laminaria rodriguezii* were tested with the help of nucmer tool of MUMmer (Kurtz et al. 2004), which could rapidly align the entire genomes. Further, alignment was done with the help of MAFFT multiple sequence alignment tool (Katoh and Standley 2013). The annotation of assembled genomes of all Laminaria species was done by using available kelp chloroplast genomes in BLAST tool. t-RNA were detected via t-RNA-scan-SE (Lowe and Eddy 1997) using organelle t-RNA detection method. Single nucleotide polymorphism (SNPs), which are the basic form of sequence variations, originated by single base substitutions, and insertion deletion polymorphism (indels), where one or more nucleotides were inserted and deleted

in the genome sequences, this way both SNPs and small insertions-deletions were defined for the analysis of allelic differences in diploid eukaryote genomes.

The raw sequence reads from both *Laminaria digitata* and *Laminaria solidungula* were mapped to the *Saccharina japonica* genome as a reference. Another chloroplast genome of kelp species were taken from NCBI data base (Table 2-1), and artificial reads were produced through ArtificialFastqGenerator (Frampton and Houlston 2012). The reads from all chloroplast genomes were mapped to the reference genome (*Saccharina japonica*) with the help of Bowtie2 tool (Langmead and Salzberg 2012) and sorted bam file was generated. Bowtie 2 is very efficient tool, which can align sequence reads to the long reference sequences. Furthermore, sequence variations were analyzed using The Genome Analysis Toolkit (Auwera et al. 2013) and obtained SNP library was further examined manually. Afterwards, the sequences portion of *Laminaria rodriguezii*, which were larger than 100 bases were examined for presence of additional open reading frames (ORFs) without any annotation. The entire annotated organellar genomes were converted to a GenBank file and later checked with OGDRAW (Greiner, Lehwark, and Bock 2019). Subsequently, all sequences were edited to start at the same position and afterwards, all were aligned by Clustal Omega (McWilliam et al. 2013; Sievers et al. 2011).

2.2.6 Physiology and Single nucleotide polymorphisms (SNPs) analyses of transcript data on the draft *Laminaria digitata* nuclear genome

2.2.6.1 Physiological Experiment

Laminaria digitata gametophytes, which were isolated from Spitsbergen and Helgoland were used for physiological experiments (performed at Alfred-Wegener-Institute, Bremerhaven). Prior to the start of the experiment, cultures were maintained vegetatively under red light (approx. 3 µmol photons m^{-2} s^{-1}; ProfiLux 3 with LED Mitras daylight 150, GHL Advanced Technology, Kaiserslautern, Germany) in a 16:8 h L:D cycle at 15°C in a temperature-controlled cooling chamber (error ± 1°C) in sterile Provasoli-enriched seawater (PES; Provasoli, 1968; modifications: HEPES-buffer instead of TRIS, double concentration of disodium glycerol phosphate; iodine enrichment following Tatewaki 1966). To perform fertilization of selfings and crosses, stock suspensions of each gametophyte culture were prepared by gently fragmenting gametophyte material using mortar and pestle. Gametophytes were added to petri dishes (Ø 5 cm) containing four glass cover slips and filled with 12 ml half-strength PES to a desired density of each 250 male and female gametophyte filaments cm^{-2} in all combinations. This created the “parent” treatments of HfHm, Helgoland female x Helgoland male; HfSm, Helgoland female x

Spitsbergen male; SfHm, Spitsbergen female x Helgoland male; SfSm, Spitsbergen female x Spitsbergen male (n = 4). Gametogenesis was induced at 10°C and 15–18 µmol photons m^{-2} s^{-1} white light in a temperature-controlled cooling chamber. Macroscopic sporophytes were subsequently cultivated in 1-litre glass beakers and 5-litre bottles with gentle aeration at 10°C under increased irradiance of 30–35 µmol photons m^{-2} s^{-1} with weekly changes of half-strength PES medium. In the main experiment, 4–7 cm long sporophytes were subjected to temperatures of 10°C (control), 19°C and 20.5°C to assess heat stress responses between crosses. Seven sporophytes were assigned to one replicate plastic container (Wide neck containers series 310 PETG, 2000 ml, Kautex GmbH & Co. KG, Bonn-Holzlar, Germany) filled with 1.8 litre of half-strength PES (n = 4). Samples were acclimated for one day at 13.5°C and one day at 17°C before reaching the experimental temperatures of 19°C and 20.5°C on day 0 of the experiment, while the control treatments remained at 10°C. Two sporophytes per replicate were marked to be used for growth and fluorometry measurements throughout the experiment by punching small holes in the distal thallus with a Pasteur pipette. Of the unmarked five sporophytes per replicate, three were frozen in liquid nitrogen throughout the experiment (before acclimation, day 1 of temperature treatment, day 18 of temperature treatment). Samples frozen before acclimation and after 18 days of temperature treatment were used for transcriptomic analysis, and stored at -80°C and processed within three weeks. The remaining two sporophytes served as backup.

2.2.6.2 RNA isolation

Frozen sporophytes (500mg) were grounded in liquid nitrogen with pre cooled mortar and pestle, and homogenized was transferred to 2.0 ml Eppendorf tubes. 1 ml of extraction buffer and 20 µl ß-Mercaptoethanol were added and mixed well. The mixture was incubated for 10 min at 45°C in a thermoblock, then 1 ml of Chloroform:Isoamyl alcohol (24:1 v/v) was added, and vortexed vigorously for 10 min followed by centrifugation of sample at 12,000g for 20 min at 20°C (Heinrich et al, 2012). Later, aqueous phase was collected (maximal 750 µl) into a cleaned 2 ml microcentrifuge tube. Afterwards, 0.3 volume of ethanol (96-100%) was added and mixed gently. Once again, 1 ml Chloroform:Isoamyl alcohol (24:1 v/v) was mixed, and vortexed for 10 min and later centrifuged at 12,000g for 20 min at 20°C (Mundt, Heinrich, and Hanelt 2019). Supernatant was collected (maximal 500 µl) into a cleaned 2 ml microcentifuge tube and then, total RNA extraction was carried out using the Qiagen RNeasy® Mini Kit according to the manufacturer's instructions. RNA was eluted from the spin columns with 50 µl of

RNase-free water, and the quantity and purity of the extracted RNA were determined by absorbance analysis of the samples using NanoDrop ND-100 spectrometer and the quality of RNA were checked by Bioanalyzer.

Lysis buffer

CTAB	2% (w/v)
Tris/HCl, pH 8.0	100mM
NaCl	1.0 M
Na_2EDTA	50mM
ß- Mercaptoethanol	0.2% (v/v)

2.2.6.3 RNA sequencing

For measurement of gene expression changes, to reveal a snapshot of the whole transcriptome at a particular time period, Illumina next generation sequencing was performed by an Illumina HiSeq run. Both RNA library preparation and sequencing were performed by the Cologne Center for Genomics.

2.2.6.4 Data analysis with DESeq2

Bioinformatic transcriptomics and data processing were carried out by Daniel (AWI) and principal component analysis was generated in the statistical environment software R (v.2.15.0, R Development Core Team, 2012) and RStudio (v.0.99.465, RStudio Team, 2015). Further data visualisation was done with CorelDRAW® 2017 (in collaboration with AWI).

2.2.6.5 Mapping and SNPs detection

RNA sequencing reads were aligned to the reference genomic sequences with the help of Bowtie2 tool (implemented in tophat2). Single nucleotide polymorphisms (SNPs) were detected using Genome Analysis Toolkit (GATK), and all potential variation of allele expression was analysed. Furthermore, the analyses of SNPs in different crosses were done in excel.

3 Results

DNA isolation in kelp species is very difficult due to the presence of high level of lipids, polysaccharides or phenolic compound. After a number of unsuccessful attempts, we finally produced enough DNA for genomic assembly of Laminaria species and also high molecular weight DNA (Figure 2-1) for nanopore sequencing (this chapter is a compilation of the corresponding part of reference (Rana et al. 2019)).

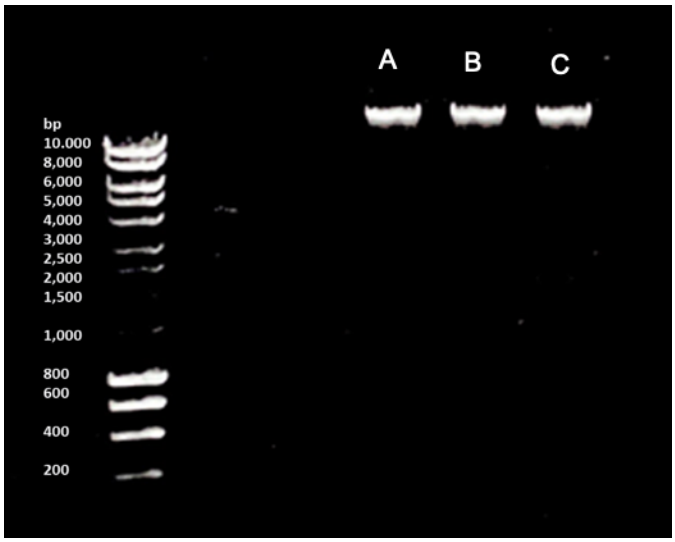
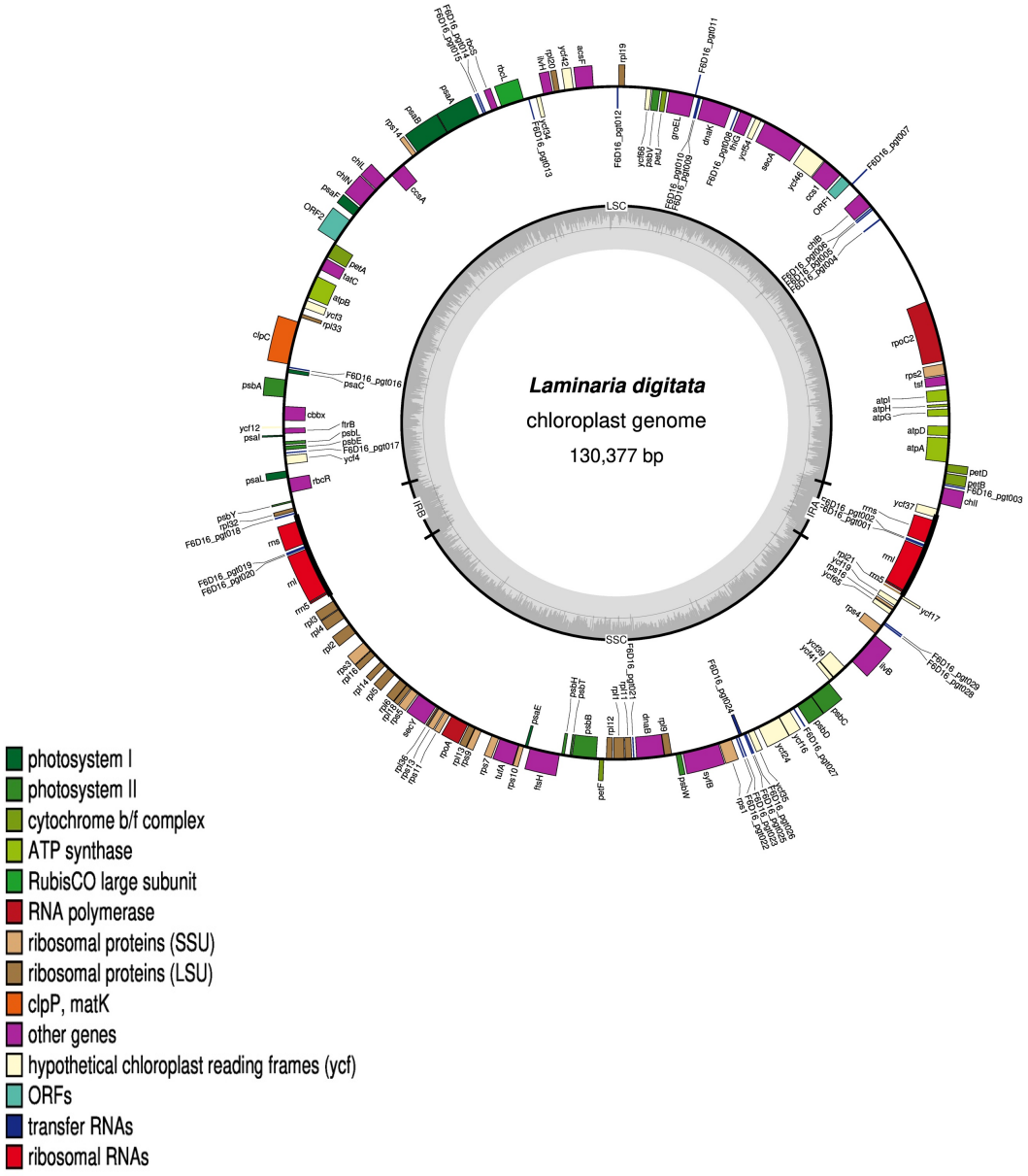


Figure 2- 1: Genomic DNA, gel picture of Laminaria species A: *Laminaria digitata* **B:** *Laminaria solidungula* **C:** *Laminaria rodriguezii*

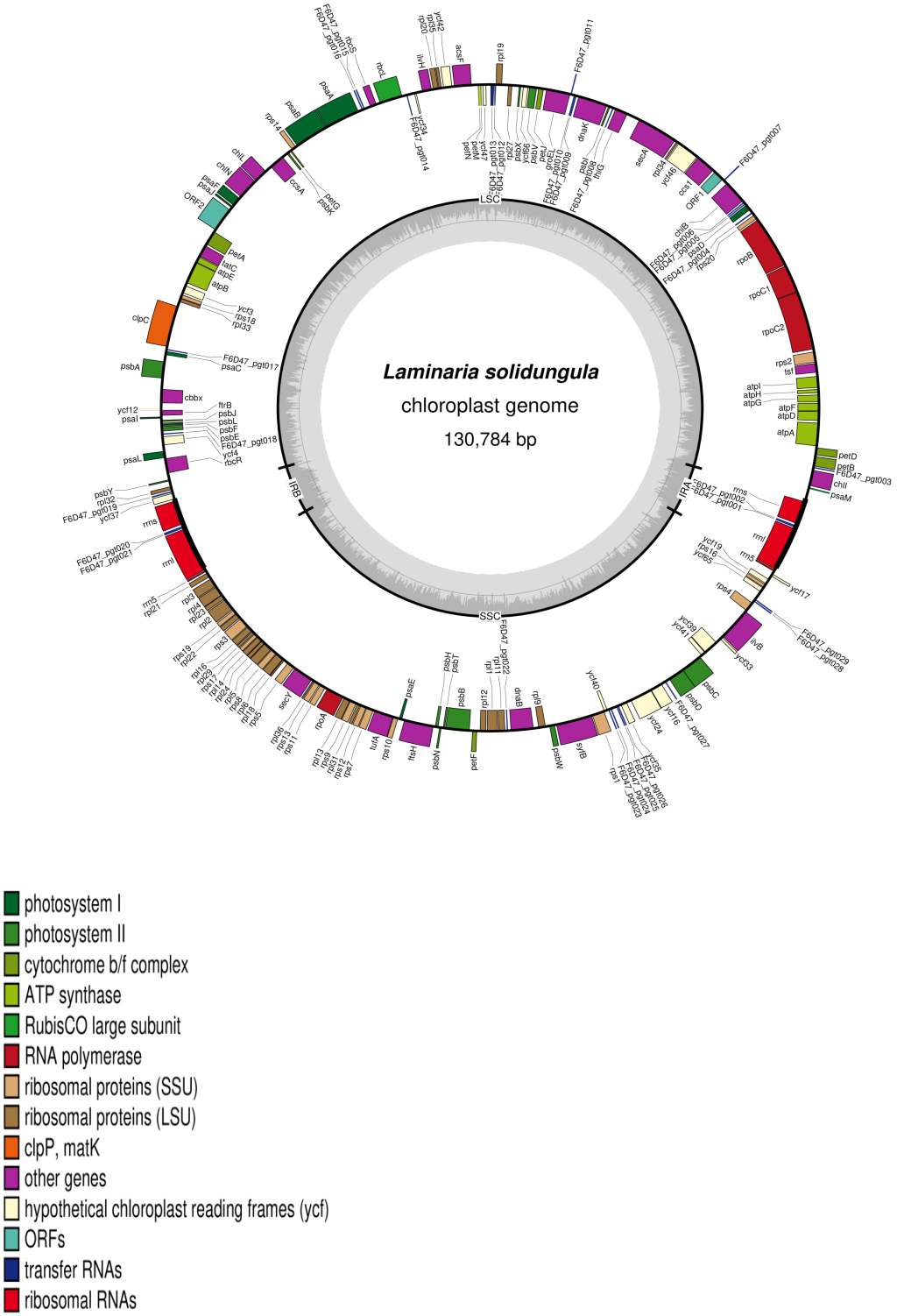
3.1 The organellar genomes of *Laminaria digitata*, *Laminaria solidungula* and *Laminaria rodriguezii*

The sequencing of total DNA generated 179 million reads for *Laminaria digitata* and 150 million reads for *Laminaria solidungula* and 134 million reads for *Laminaria rodriguezii*. Which were amounting to around 12.3, 11.3 and 9.8 gigabases, respectively. The mitochondrial genomes of *L. solidungula* and *L. rodriguezii* and chloroplast genome of *L. digitata, L. solidungula* and *L. rodriguezii* were completely reconstructed from short read sequencing (Figure 2-2 A, B, C, D and E) using the draft assembly of all reads and extracting the organellar parts with BLAST tool (Method 2.2.3)

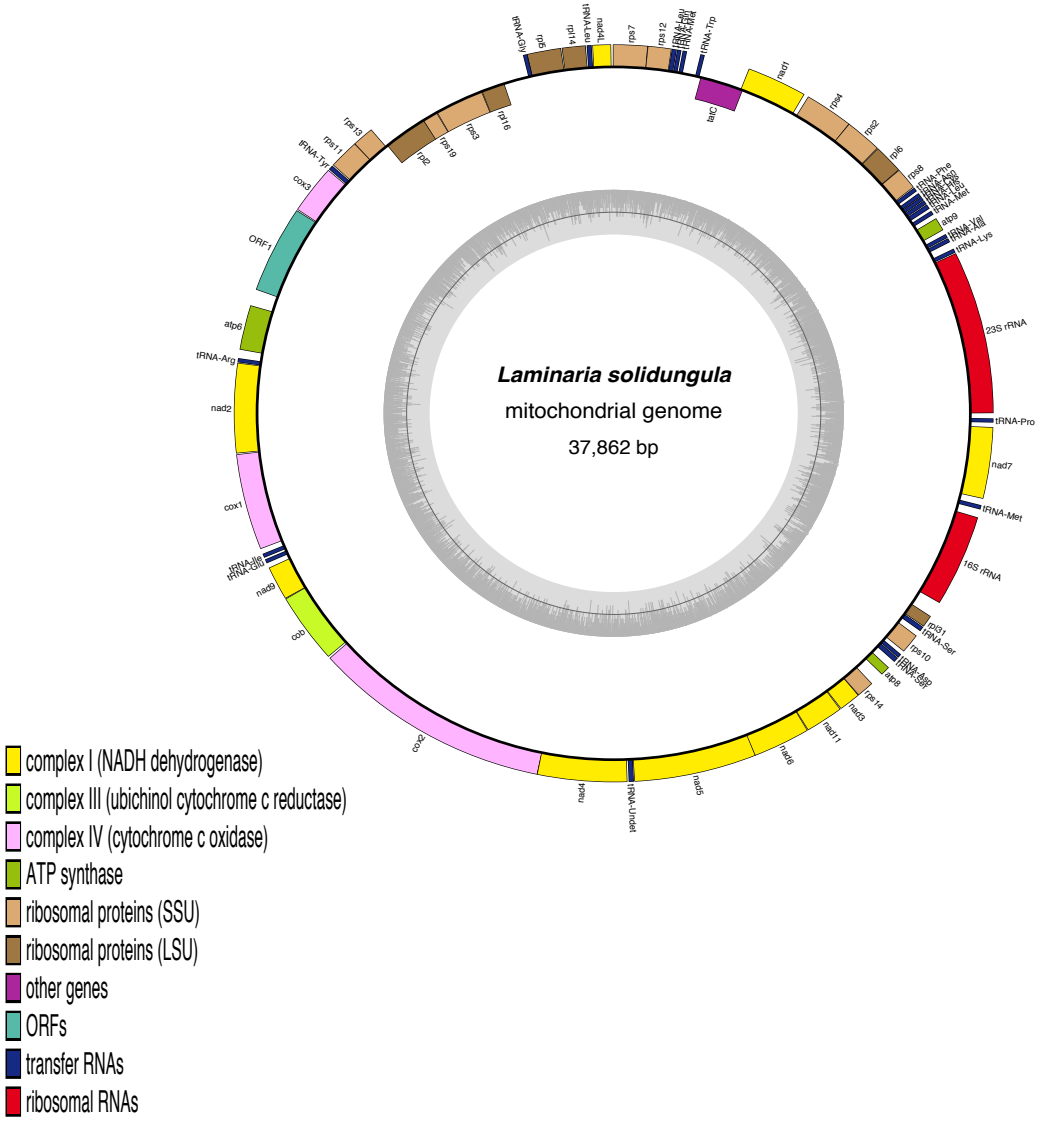
A.


Laminaria digitata
chloroplast genome
130,377 bp
LSC
SSC
IRA
IRB
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
ORFs
transfer RNAs
ribosomal RNAs

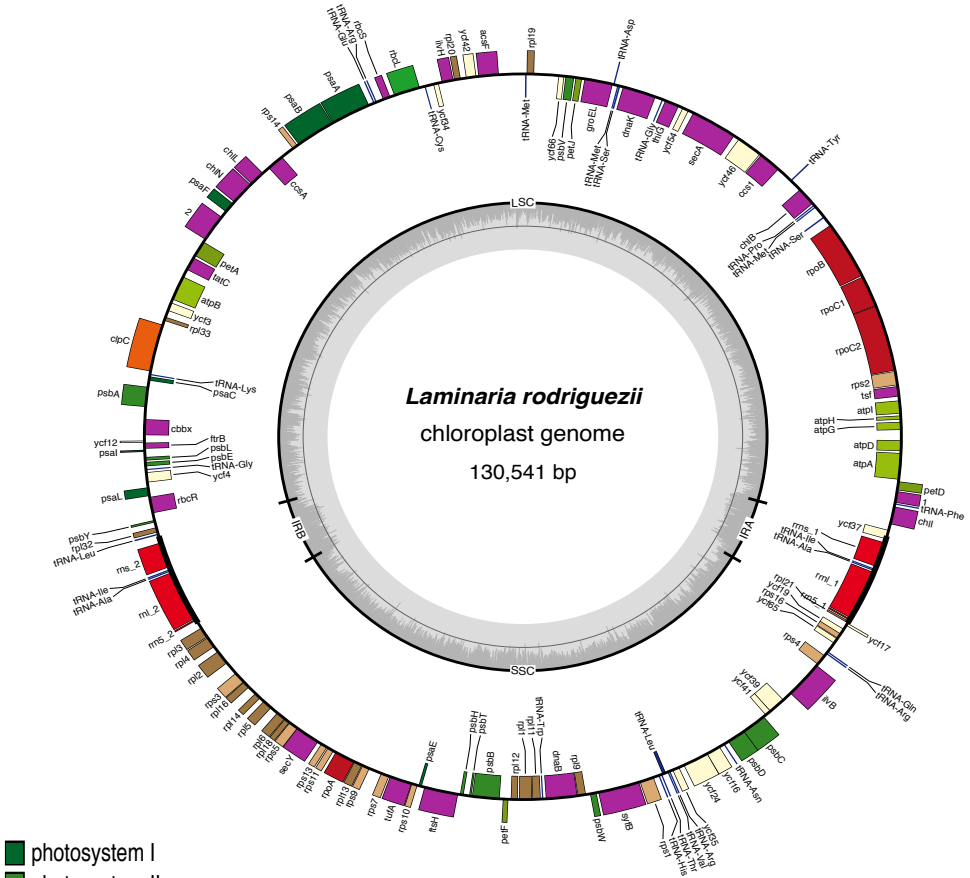
B.


Laminaria solidungula
chloroplast genome
130,784 bp
LSC
IRB
IRA
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
ORFs
transfer RNAs
ribosomal RNAs

C.


Laminaria solidungula
mitochondrial genome
37,862 bp
complex I (NADH dehydrogenase)
complex III (ubichinol cytochrome c reductase)
complex IV (cytochrome c oxidase)
ATP synthase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
other genes
ORFs
transfer RNAs
ribosomal RNAs

D.


Laminaria rodriguezii
chloroplast genome
130,541 bp
LSC
IRB
SSC
IRA
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
transfer RNAs
ribosomal RNAs

E.

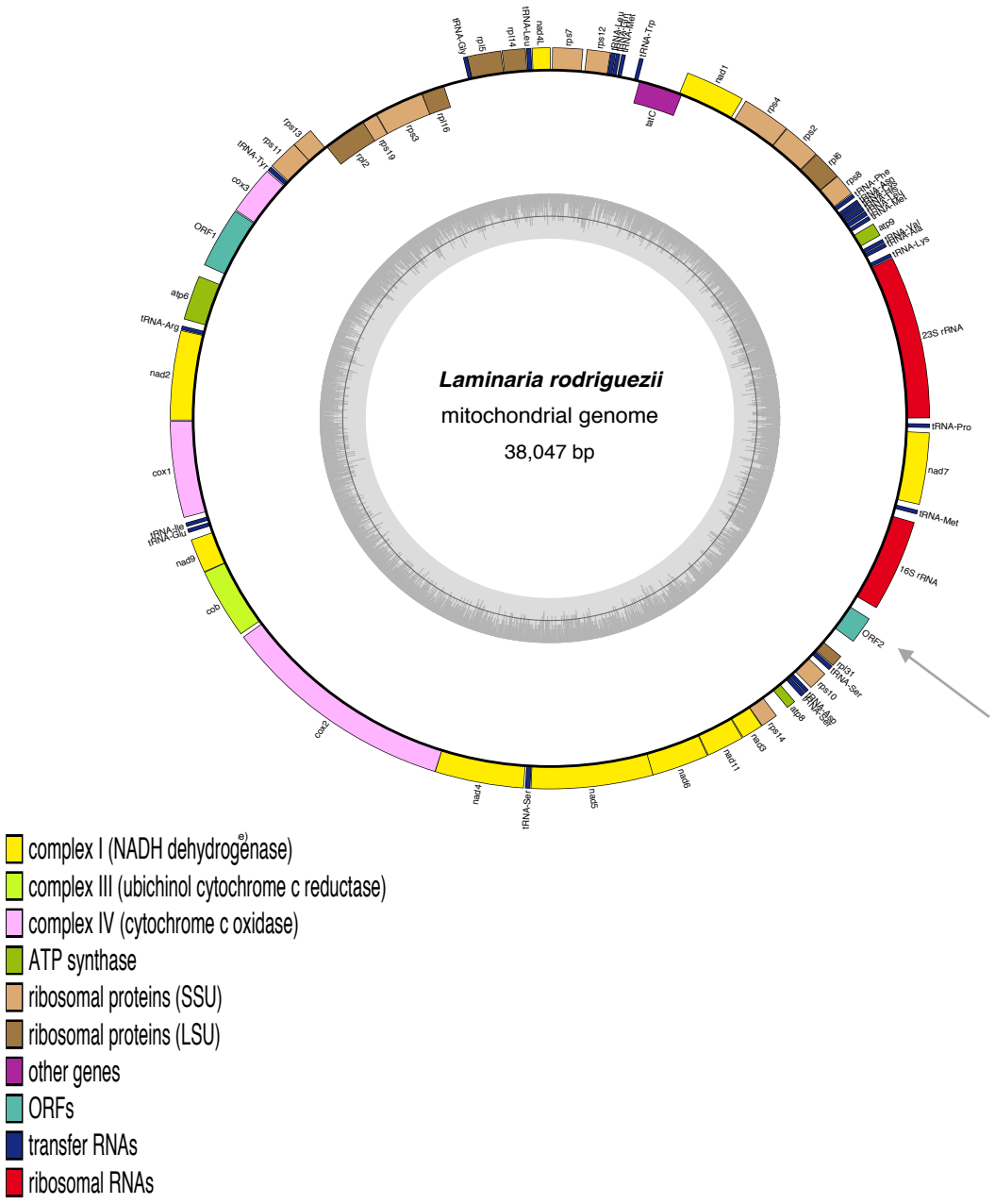


Figure 2- 2: Newly sequenced organellar genomes. A: *Laminaria digitata* chloroplast genome **B:** *Laminaria solidungula* chloroplast genome **C:** *Laminaria solidungula* mitochondrial genome **D:** *Laminaria rodriguezii* chloroplast genome **E:** *Laminaria rodriguezii* mitochondrial genome & the arrow indicates the position of the ORF as discussed in the text. The figures were made with OGdraw (Greiner et al. 2019) from the annotated sequences files in GenBank.

After the assembly of all reads, we extracted the chloroplast contigs from the total assembly using the *Ectocarpus siliculosus* chloroplast coding sequences as a bait, and mitochondria contigs of *Laminaria rodriguezii* and *Laminaria solidungula* were retrieved from the known mitochondrial sequences of *Laminaria digitata*. The coverage of the organelle genome is much higher than the nuclear genomes, which were estimated around ~3000x each for *Laminaria digitata*, *Laminaria solidungula* and *Laminaria rodriguezii,* the assembly of large amount of reads results in a very fragmented organellar genome. Consequently, the extracted chloroplast and mitochondria contigs were extended, and scaffolded and the gaps between them were filled by using the original raw reads information with the help of Gapfiller (Boetzer and Pirovano 2012). We annotated organellar genomes using the other kelp genomes (Table 2-1). All these genomes had similar lengths as their related counterparts (Table 2-1).

Table 2- 1: The organellar genomes used in this study. The IDs are to the NCBI accession numbers.

Name	ID	Length	Comments
Mitochondrial genomes			
Desmarestia viridis	AY500367.1	39049	
Lessonia spicata	NC_044181.1	37097	
Nereocystisluetkeana	NC_042395.1	37399	
Macrocystis integrifolia	NC_042669.1	37366	
Laminaria solidungula	MT732098	37862	Included in PhD work
Laminaria rodriguezii	MT732097	38047	Included in PhD work
Laminaria hyperborea	JN099683.1	37976	
Laminaria digitata	AJ344328.1	38007	
Saccharina latissima	KM675818.1	37659	
Saccharina longissima	JN099684.1	37628	
Saccharina japonica	MG712776.1	37657	
Saccharina japonica	AP011493	37657	
Saccharina religiosa	AP011494.1	37657	
Saccharina longipedalis	AP011497.1	37657	
Saccharina diabolica	AP011496.1	37657	
Saccharina ochotensis	AP011495.1	37656	
Saccharina coriacea	AP011499.1	37500	
Saccharina angustata	AP011498.1	37605	
Saccharina sculpera	KR350664.1	37627	
Costaria costata	KF384641.1	37461	
Undaria pinnatifida	KF319031.1	37402	
Chloroplast genomes			
Lessonia spicata	NC_044182.1	130305	
Laminaria solidungula	MH784528.1	130784	
Costaria costata	NC_028502.1	129947	
Undaria pinnatifida	NC_028503.1	130383	
Laminaria digitata	MH784527	130376	Included in PhD work
Saccharina japonica	JQ405663.1	130584	
Laminaria rodriguezii	MT732096	131092	Included in PhD work

Annotation revealed that the slightly longer *Laminaria rodriguezii* mitochondrial genome contain an additional open reading frame, which are not yet known from any other kelp. We also included de novo detection of tRNAs in chloroplast genome. With this approach, we defined 139 coding sequences each in the genomes and 29 (*L. digitata* and *L. rodriguezii*) and 30 (*L. solidungula*) tRNAs together with three rRNA species (16S, 23S, and 5S) located in the inverted repeats. Because of the difference in tRNAs number in chloroplast genomes of kelp species, we further analysed, which tRNA were affected by potential evolutionary processes. In total, we defined 36 tRNA locations on the chloroplast genomes of which 27 are located on the same position in five chloroplast genomes, which we used for chloroplast analyses (*Undaria pinnatifida, Costaria costata, Saccharina japonica, Laminaria digitata and Laminaria solidungula*) (Table 2-2). Out of the nine tRNAs, seven are present in only one species, one can be found in two species, and the remaining one is missing in *Costaria costata* only. Interestingly, six or seven orphan tRNA occurring in two genomes are predicted to contain type II introns.

Table 2- 2: tRNAs and their positions in the kelp genomes. LD**= *Laminaria digitata***, LS**= *L. solidungula,*** SJ**= *Saccharina japonica,*** CC**= *Costaria costata***, UP**= *Undaria pinnatifida***

Number of tRNAs	**31**		**28**		**31**		**29**		**30**	
Species	**UP**		**CC**		**SJ**		**LD**		**LS**	
tRNA	**start**	**stop**	**start**	**stop**	**start**	**stop**	**start**	**stop**	**start**	**stop**
Met					5832	5899				
Phe	7288	7360	7293	7365	7148	7220	7126	7198	7143	7215
Tyr	27354	27434	27433	27513	27249	27329	27279	27359	27236	27316
Sup (intron)	34028	34266							33945	34190
Asp	38590	38664	38727	38801	38498	38572	38600	38674	38529	38603
Ile (intron)	50988	51192								
Arg	51998	52070	52161	52233	51870	51942	51929	52001	51872	51944
Glu	52127	52199	52290	52362	51999	52071	52058	52130	52001	52073
Ile (intron)	59171	59416								
Leu	81865	81946	81703	81784	81909	81990	81785	81866	81854	81935
Ile	83943	84016	83860	83933	84124	84197	83915	83988	84195	84268
Ala	84020	84092	83937	84009	84201	84273	83992	84064	84272	84344
Gly (intron)			96734	96976						
His	116468	116540	116392	116463	116650	116722	116377	116449	116776	116848
Thr	116618	116690	116545	116617	116804	116876	116531	116603	116930	117002
Val	117062	117133	116874	116945	117160	117231	116914	116985	117317	117388
Arg	117156	117228			117254	117326	117008	117080	117411	117483
Phe (intron)					117852	117957				
Asn	120522	120593	120233	120304	120563	120634	120381	120452	120789	120860
Arg	128196	128269	127843	127916	128307	128380	128066	128139	128471	128544
Gln	128311	128382	127951	128022	128414	128485	128181	128252	128591	128662
Leu (intron)							116823	116617		
Trp	109963	109891	109885	109813	110141	110069	109912	109840	110272	110200
Gly	78187	78117	78029	77959	78220	78150	78094	78024	78169	78099
Lys	72723	72652	72767	72696	72690	72619	72778	72707	72697	72626

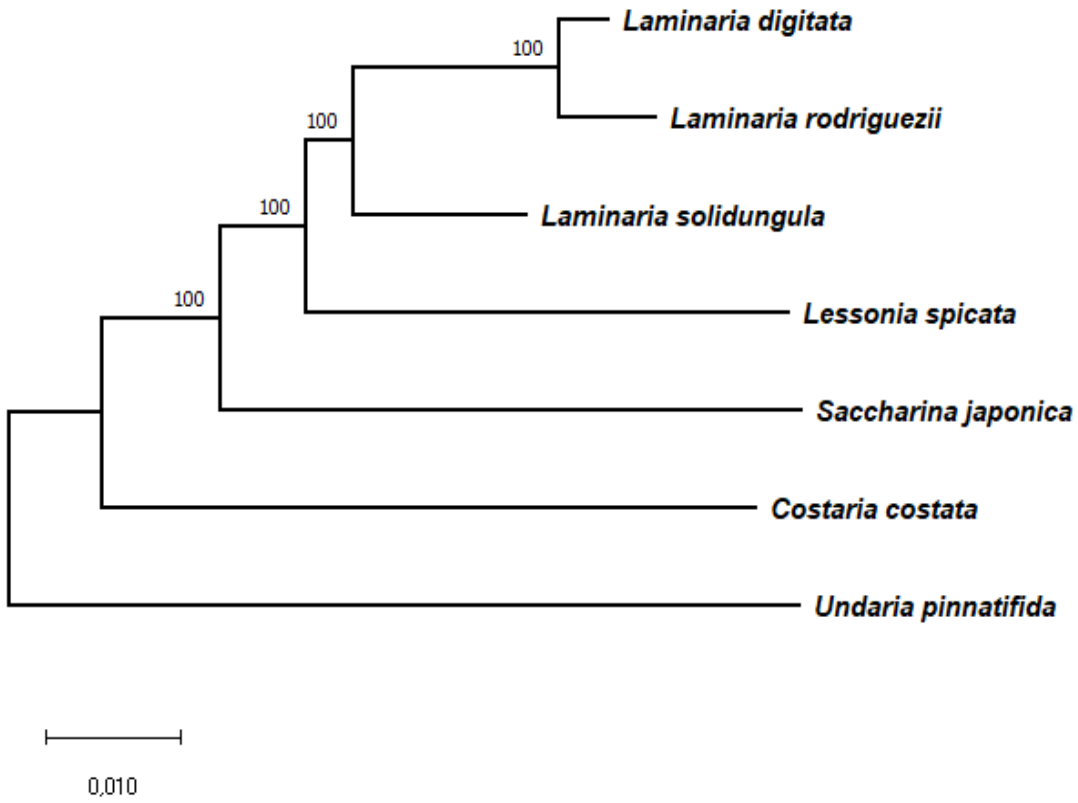
Cys	49214	49144	49362	49292	49084	49014	49174	49104	49106	49036
Lys (intron)					43701	43498			43708	43505
Met	43556	43471	43710	43625	43453	43368	43562	43477	43461	43376
Met	38427	38355	38569	38497	38339	38267	38441	38369	38371	38299
Ser	38347	38258	38489	38400	38259	38170	38361	38272	38291	38202
Gly	35888	35817	35999	35928	35770	35699	35865	35794	35798	35727
Pro	25512	25439	25613	25540	25430	25357	25458	25385	25416	25343
Met	25383	25310	25492	25419	25309	25236	25337	25264	25295	25222
Ser	24675	24588	24754	24667	24564	24477	24588	24501	24575	24488
Ile	3464	3391	3466	3393	3286	3213	3288	3215	3287	3214
Ala	3387	3315	3389	3317	3209	3137	3211	3139	3210	3138

3.2 The phylogeny of kelp species using complete mitochondria and chloroplast genomes

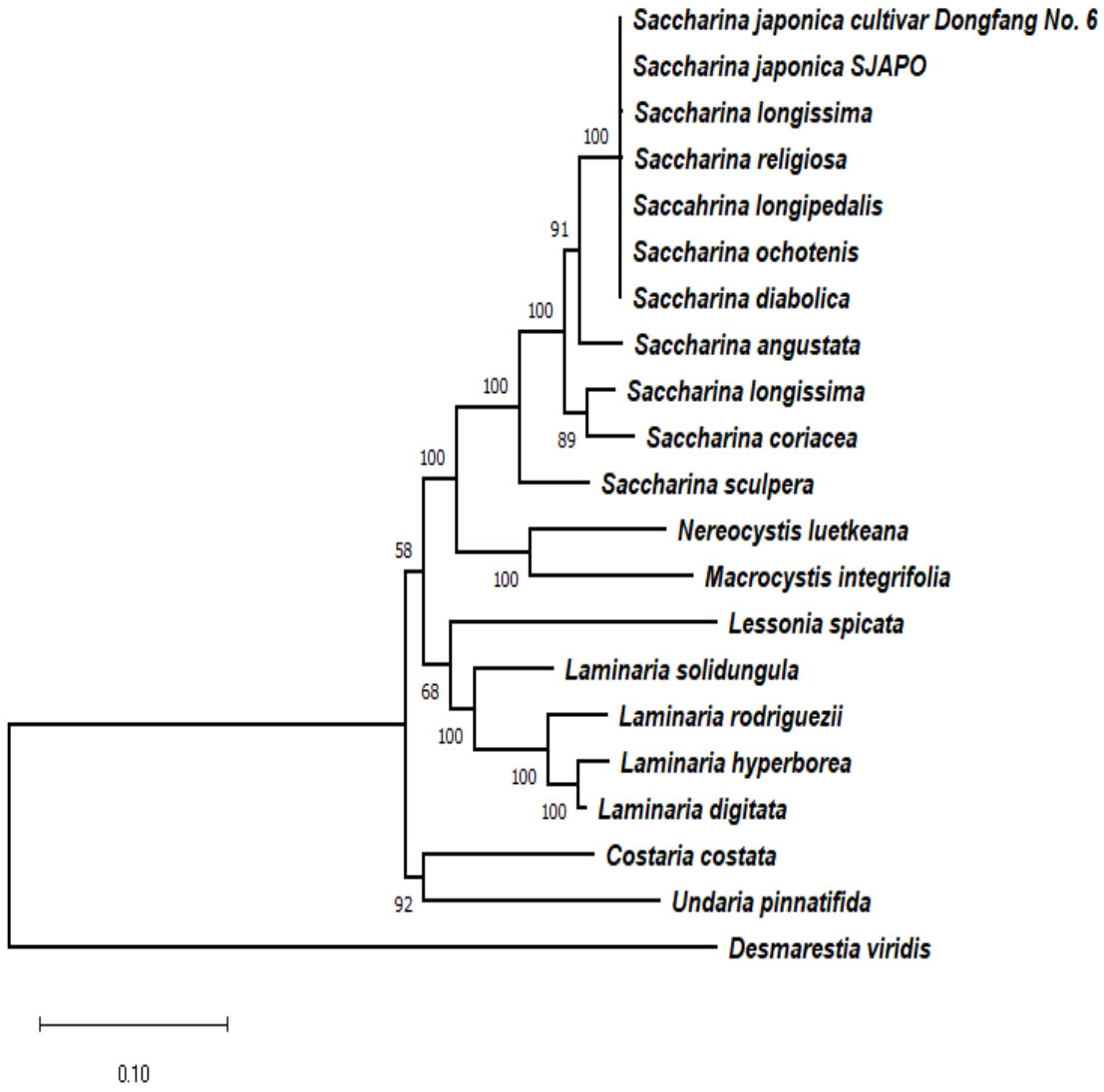
To be able to trace back the evolution of kelp species we needed a robust phylogeny of the analysed species. Hence, we checked the sequences for collinearity and retrieved all kelp species organellar genome from NCBI database to our reconstructed organellar genomes (Table 2-1). Additionally, the mitochondrial genome of the brown alga *Desmarestia viridis* (Secq et al., 2006), which is not a kelp species, but turned out to be collinear to kelp mitochondrial genome, was providing a suitable outgroup for the phylogenetic analyses. The collinearity of organellar genomes in brown algal species was used to employ complete organellar genome alignments directly for a phylogenetic analysis irrespective of their coding potential, thus also including tRNAs, rRNAs, and intergenic regions. We used 18 complete mitochondrial genomes of kelps for the calculation of the mitochondrial tree together with the newly sequenced mitochondrial genomes from *L. solidungula* and *L. rodriguezii* and *D. viridis* as an outgroup. We found no differences in species placements between the organellar trees (Fig 2-3 A, B and C). For the chloroplast tree, we used chloroplast genomes from *Saccharina japonica* (Wang et al. 2013), *Undaria pinnatifida* (Zhang, Wang, Liu, G. Wang, et al. 2015a), *Costaria costata* (Zhang, Wang, Liu, H. Wang, et al. 2015), *Lessonia spiculata* (Tineo et al. 2019) and some of the Laminaria species analysed here (Figure 2-3 A). The mitochondrial maximum likelihood tree is not well resolved in the Saccharina species complex due to inadequate phylogenetic signal. We also used MrBayes (Method 2.2.4) to compare its outcome with the maximum likelihood tree. This returned the same tree topology as with the maximum likelihood method, with even higher support at the deeper branches with cumulative probability of 100 % at all nodes (Figure 2-3 C). MrBayes also yielded two alternative topologies with lower cumulative probability affecting the placement of *S. diabolica* and *S. cochotenis*. It is clear from the phylogeny tree that the Laminaria species group together, and the bootstrap values of the kelp trees indicate that the phylogenetic

relationships of the species are well resolved. Sequence variations not following the species tree were also observed, but the phylogenetic signal over the whole plastid genome seems to be strong enough to be not influenced by them. This phylogeny was then the basis for further analysis of the noticeable trends in kelp chloroplast genome evolution.

A.



B.


Saccharina japonica cultivar Dongfang No. 6
Saccharina japonica SJAPO
Saccharina longissima
Saccharina religiosa
Saccahrina longipedalis
Saccharina ochotenis
Saccharina diabolica
Saccharina angustata
Saccharina longissima
Saccharina coriacea
Saccharina sculpera
Nereocystis luetkeana
Macrocystis integrifolia
Lessonia spicata
Laminaria solidungula
Laminaria rodriguezii
Laminaria hyperborea
Laminaria digitata
Costaria costata
Undaria pinnatifida
Desmarestia viridis
100
91
100
100
89
100
58
100
68
100
100
100
92
0.10

C.

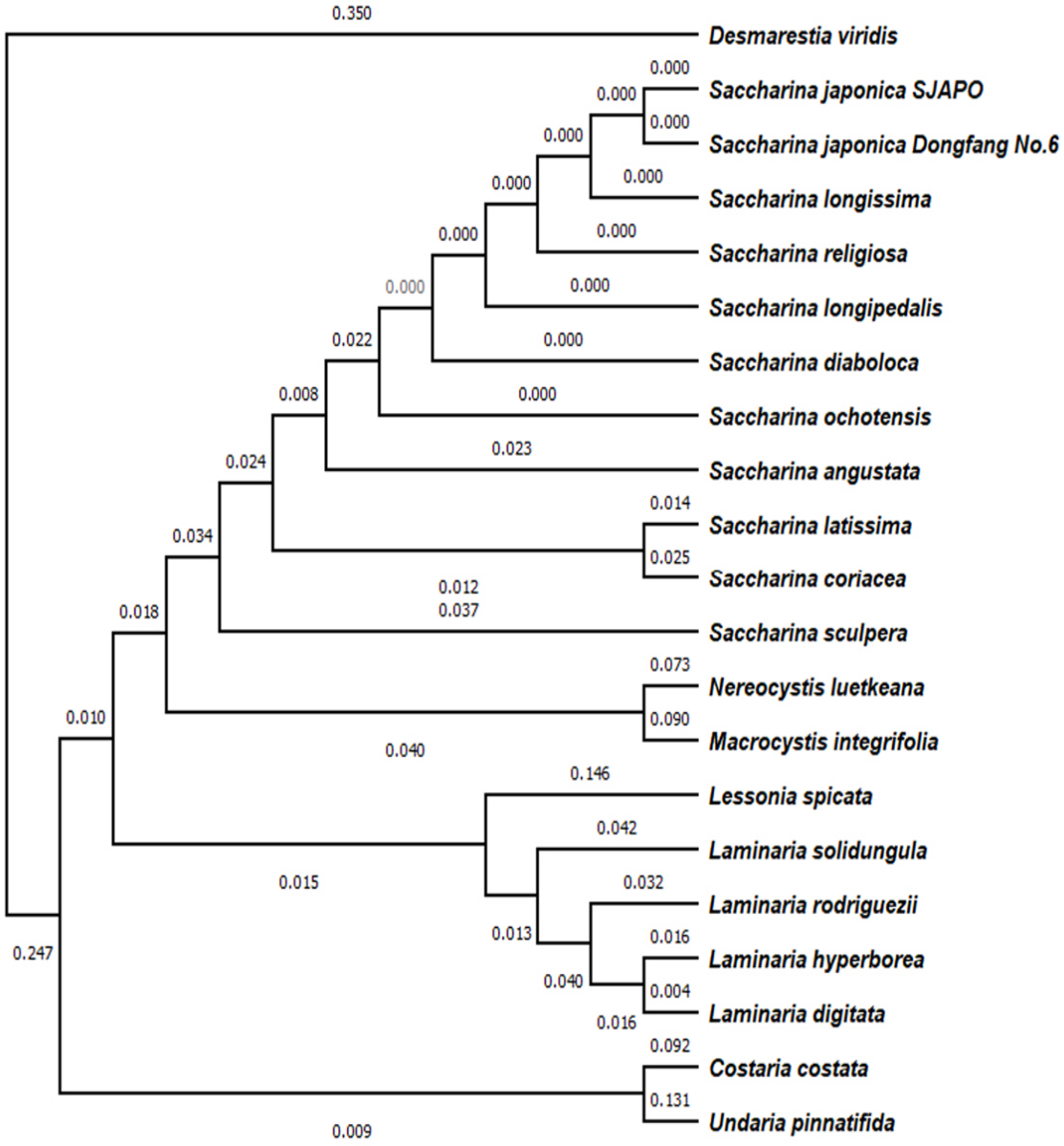


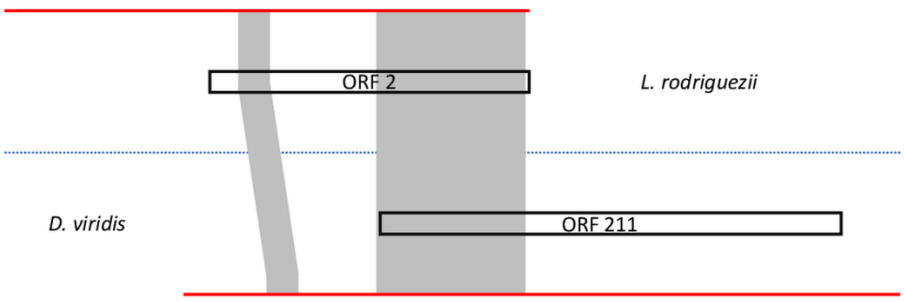
Figure 2- 3: Phylogenetic analysis of kelp species with complete chloroplast and mitochondrial genomes. A: Maximun likelihood tree of aligned chloroplast genomes **B:** Maximum likelihood tree of aligned mitochondrial genomes. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model (Nei and Kumar 2000). The tree with the highest log likelihood (-306179.15) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches calculated from 500 bootstrap repetitions. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. Evolutionary analyses were conducted in MEGA X (Kumar et al. 2018) **C:** Topology of the kelp mitochondrial tree calculated with MrBayes (Ronquist et al. 2012) using 80000 generations and a burn in of 1000. All splits have a posterior probability of 100 %. The branch lengths are shown even if they are 0 in case of the Saccharina species cluster.

3.3 Alignment and special feature in organelle genomes

3.3.1 Mitochondrial genome

All kelp species organelles contain the same genes, that means, no gene was missing or additionally present in any species with one exception: only in *Laminaria rodriguezii* we found an additional open reading frame (ORF) in the mitochondrial genome. This ORF is embedded in a ca. 700 base-pairs stretch of DNA not represented in any of the analysed mitochondrial kelp genomes so far (Table 2-1). The ORF in *Laminaria rodriguezii* is 456 bases or 152 amino acids long and, according to InterProScan (Jones et al. 2014), contains no recognizable domain. We found, however, significant similarity to a mitochondrial ORF in *Desmarestia viridis* (Secq et al. 2006), where the similarity covers the 3' end of the *L. rodriguezii* ORF and at the 5' end of the *D. viridis* ORF over a length of 234 bases or 78 amino acids (Figure 2-4 B). Also a short stretch of 45 bases upstream of orf211 in *D.viridis* is similar to ORF 2 in *L. rodriguezii* (Figure 2-4 A, C). Despite overall collinearity of the two mitochondrial genomes the two ORFs do not share the same relative position within the mitochondrial genome. While the *D. viridis* ORF is located between tRNA-K and tRNA-V adjacent to the 23S rRNA gene replacing tRNA-A, the *L. rodriguezii* ORF is placed between the 16S rRNA gene and rpl31 (Figure 2-4 B). The foreign sequences in both*, L. rodriguezii* and *D. viridis*, show no detectable similarity to any known sequences including brown algal or kelp nuclear genomes, bacteria associated with brown algae, or viruses. We also did not find any evidence for transcriptional activity of the foreign sequences as no similar sequences are present in any brown algal transcriptome datasets.

A.



B.


L. rodriguezii MGTYHPQGCITSSGLSSPGCIVGITNVLNIHPSGMLSKFLFIQILEQLTE
D. viridis ----------
L. rodriguezii SILGIIHKYETTFPSSIFKNKPSSDPILISLLEQIVQSQEDQRHFVREMS
D. viridis ----------MEQVVQSQEDQRRFVKEIS
L. rodriguezii NRQNILLESLNDVVMRQDKISLDLKQMFYKDNNSWYYSHLSTTLQYLQII
D. viridis NRQNILLESLNDVVIRQDKMSLDLKHLSSGKSHSWHY-LSTTLQYLQVI
L. rodriguezii NV----------
D. viridis SIFTPYIAKILPTTLIENIPLVNKVVWFFSPGQPEWVGPMRNMETNINTL
L. rodriguezii ----------
D. viridis AAALNQTNQGLAGVHTQVTALNETARGLVESQTNIQRIHENGLAEQLARD
L. rodriguezii ----------
D. viridis LNTLATQTPYTPPVPTAEVTSAVDNSTPMDRPRLQNTAHLFRRT

C.

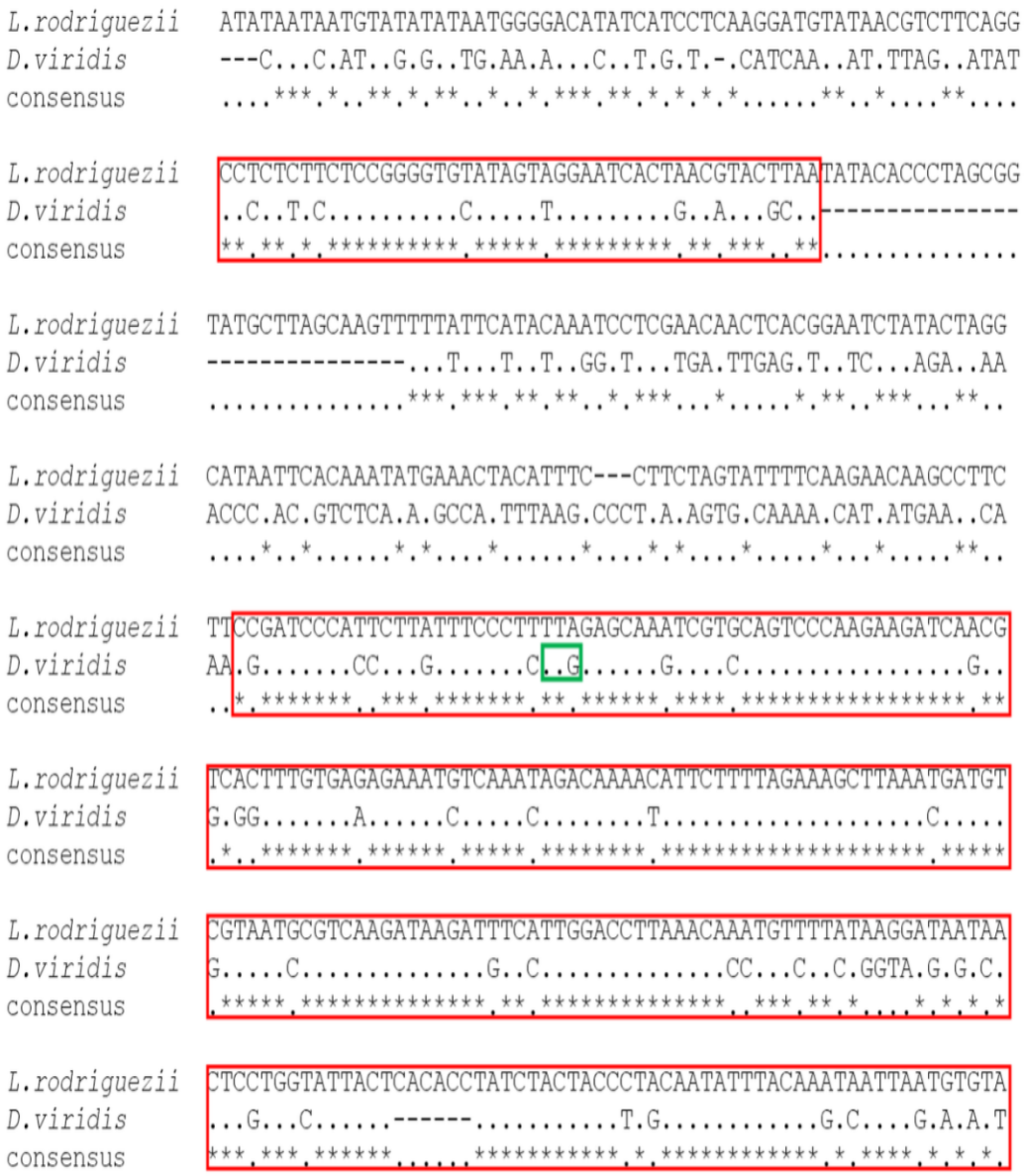


Figure 2- 4 Alignment of the gene product of ORF2 of *L. rodriguezii* with its partial counterpart from *D. viridis* (orf211 in NC_007684). A: A sketch of the two ORFs. ORFs are drawn as open rectangles, the unique regions of both species are represented by red lines. the two overlaps are depicted as grey boxes. **B:** The alignment of the translated protein sequences of the two ORFs. Asterisks denote conserved residues, colons conserved substitutions, and dots semi conserved substitutions. **C:** The section comprising the two overlaps aligned on nucleotide level. The overlapping parts from A are framed with red lines, the start codon of orf211 is highlighted by a green open box. Asterisks denote conserved residues in the consensus. Identical bases in *D. viridis* are represented by dots. The alignments were done with clustalw (Larkin et al. 2007).

3.3.2 Chloroplast genome

Collinearity of chloroplast genome were tested with the help of nucmer tool (NUCleotide MUMer, part of MUMmer tool) (Kurtz et al. 2004) alignment, using *Undaria pinnatifida* genome as reference, which showed that the large segments of all chloroplast genomes could indeed be aligned (Figure 2-5). Hence, only a few regions appear to be rearranged or contain larger insertions and deletions, as a result similarity dropped below the 90% threshold.

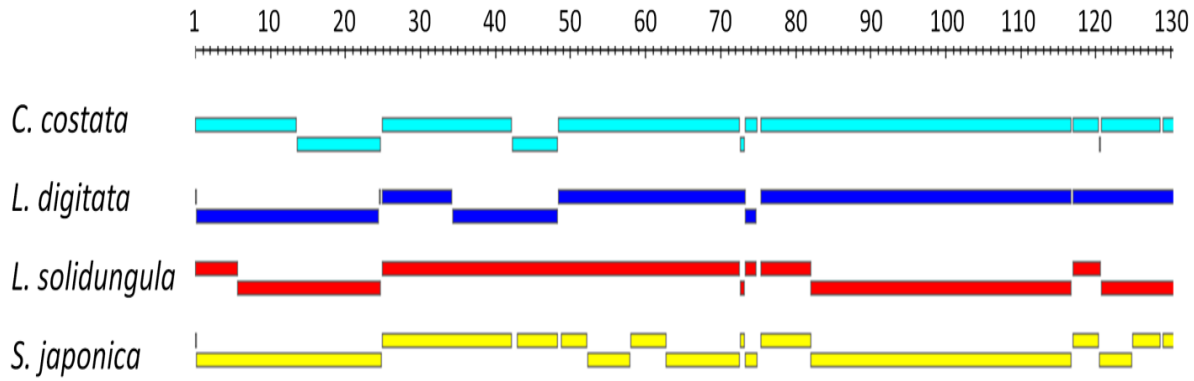


Figure 2- 5: Synteny of the four kelp chloroplast genomes. The assembled genomes were mapped against the *Undaria pinnatifida* genome using nucmer (Kurtz et al. 2004) and visualized with Bio:: Graphics (https://metacpan.org/release/LDS/Bio-Graphics-2.37). Colours for the different chloroplast genomes were chosen arbitrarily. The identity threshold for each segment was 90 % and small hits contained within a larger one were removed including the matches of the second repeat region. The scale represents the *U. pinnatifida* base positions in kb. The breaks indicate nucmer alignment breaks Table 2-3. When gaps between alignments are small the graphics software shifted the next alignment block to a lower position to emphasize the alignment gap positions.

Missing or additional tRNAs are very small to cause such similarity breakpoints as the comparison of tRNA positions (Table 2-2) and nucmer similarity breakpoint positions show (Table 2-3).

Table 2- 3: **nucmer segments of kelp genomes** mapped to the U. pinnitafida genome (corresponds to Figure 2-5)

CC	Costaria	segments	9	13403	.	+	1
CC	Costaria	segments	13533	24692	.	+	1

CC	Costaria	segments	24859	42156	.	+	1
CC	Costaria	segments	42290	48198	.	+	1
CC	Costaria	segments	48384	72508	.	+	1
CC	Costaria	segments	72648	73200	.	+	1
CC	Costaria	segments	73336	74800	.	+	1
CC	Costaria	segments	75314	116698	.	+	1
CC	Costaria	segments	116987	120312	.	+	1
CC	Costaria	segments	120439	120642	.	+	1
CC	Costaria	segments	120746	128507	.	+	1
CC	Costaria	segments	128991	130383	.	+	1
SJ	Saccharina	segments	1	177	.	+	1
SJ	Saccharina	segments	178	24745	.	+	1
SJ	Saccharina	segments	24854	42083	.	+	1
SJ	Saccharina	segments	42946	48206	.	+	1
SJ	Saccharina	segments	48778	52199	.	+	1
SJ	Saccharina	segments	52297	57837	.	+	1
SJ	Saccharina	segments	58025	62661	.	+	1
SJ	Saccharina	segments	62748	72506	.	+	1
SJ	Saccharina	segments	72643	73187	.	+	1
SJ	Saccharina	segments	73341	74776	.	+	1
SJ	Saccharina	segments	75310	81960	.	+	1
SJ	Saccharina	segments	82003	116698	.	+	1
SJ	Saccharina	segments	116995	120286	.	+	1
SJ	Saccharina	segments	120474	124772	.	+	1
SJ	Saccharina	segments	124883	128515	.	+	1
SJ	Saccharina	segments	128991	130383	.	+	1
LD	L_digitata	segments	1	177	.	+	1
LD	L_digitata	segments	178	24331	.	+	1
LD	L_digitata	segments	24575	24692	.	+	1
LD	L_digitata	segments	24853	34190	.	+	1

LD	L_digitata	segments	34292	48199	.	+	1
LD	L_digitata	segments	48416	73211	.	+	1
LD	L_digitata	segments	73341	74756	.	+	1
LD	L_digitata	segments	75313	116698	.	+	1
LD	L_digitata	segments	116971	130383	.	+	1
LS	L_solidungula	segments	5639	24692	.	+	1
LS	L_solidungula	segments	5669	1	.	+	1
LS	L_solidungula	segments	24853	72534	.	+	1
LS	L_solidungula	segments	72644	73206	.	+	1
LS	L_solidungula	segments	73336	74760	.	+	1
LS	L_solidungula	segments	75313	81953	.	+	1
LS	L_solidungula	segments	82003	116697	.	+	1
LS	L_solidungula	segments	116971	120642	.	+	1
LS	L_solidungula	segments	120746	130383	.	+	1

Furthermore, we aligned the chloroplast genomes with MAFFT (multiple sequence alignment program), which proved that the nucmer segments are aligned in the same order in all chloroplast genomes and that therefore all kelp chloroplast genomes are collinear. However, closer inspection revealed that small rearrangements occurred involving the inverted repeat (IR) regions (Table 2-4). In comparison to *C. costata*, *S. japonica* and *U. pinnatifida* both *Laminaria* species *L. digitata* and *L. solidungula* have a gene directly adjacent of the IRs translocated to the other copy of the IR (Table 2-4). In *L. digitata rpl*21 is affected and in *L. solidungula ycf*37. Interestingly, *ycf*37 was presumably pseudogenized during this process in *L. solidungula,* since the N terminal part of the protein is no longer encoded in this gene (Table 2-5).

Table 2- 4: Chloroplast genome features of kelp species. The inverted repeat (IR) consists of the genes in the order 16S ribosomal RNA, tRNA-Ile, tRNA-Ala, 23S ribosomal RNA, 5S ribosomal RNA. The first row shows the vicinity of the forward repeat and the second row the reverse repeat for each species row.

Species	Length	Inverted repeat length (bp)	Gene order found at boundaries of two IR regions	Rearrangements
Costaria costata	129947	5409	rpl32-tRNALeu-IR-rpl21-rpl3 ycf17-IR-ycf37-psaM	
Laminaria digitata	130376	5294	rpl32-tRNALeu-IR- rpl3 ycf17-rpl21-IR-ycf37-psaM	rpl21 at other IR
Laminaria solidungula	130398	5493	rpl32-tRNALeu-ycf37-IR-rpl21-rpl3 ycf17-IR-psaM	ycf37 at other IR; pseudogene
Saccharina japonica	130584	5496	rpl32-tRNALeu-IR-rpl21-rpl3 ycf17-IR-ycf37-psaM	
Undaria pinnatifida	130383	5404	rpl32-tRNALeu-IR-rpl21-rpl3 ycf17-IR-ycf37-psaM	

Table 2- 5: Alignment of chloroplast gene ycf37 from different brown algae. The non-homologous sequence part of *Laminaria solidungula* is colored in red. The amino acid translation was made from the 70% consensus sequence, where a nucleotide was taken if at least seventy percent of the aligned sequences have the same letter. n in consensus denotes any nucleotide at that alignment position, u stands for purine bases at that position.

```
1 Fucus_vesiculosus                 ATGAATTCTTTATTCCCTTTATTATACTCAGCTCTTTTATTTTGTTTGCTTCTTTTAATTAGCTTTTTTATTGTGAAACA
2 Ectocarpus_siliculosus            ATGAATTCTATATTTCCATTATTTTATTCTATTGCTTTATTTTTTTTTCTTTTTTTTATAAGTTTTTATATTTTAAAACA
3 Laminaria_solidungula             ------------------------------TTTAGGTTCTAGTAAGGTTTCTTGTGAGAGTTCAAGTCTCTCCGTTCGTA
4 Undaria_pinnatifida               ATGAATTCTTTATTTCCCTTAATTTACTCAATTGCTTTATTTGTTTTCCTTTTTATAATTAGTTTTTATATCTTAAAACA
5 Saccharina_japonica               ATGAATTCTTTATTTCCGTTAGTTTACTCAATTGCTTTATTTGTTTTTCTTTTTATAATTAGTTTTTATATCTTAAAACA
6 Costaria_costata                  ATGAATTCTTTATTTCCGTTAGTTTACTCAATTGCTTTATTTGTTTTTCTTTTTGTAATTAGTTTTTATATCTTAAAACA
7 Laminaria_digitata                ATGAATTCTTTATTTCCGTTAGTTTACTCAATTGCTTTATTCGTTTTTCTTTTTATAATTAGTTTTTATATATTAAAACA
  consensus/100%                    ..............................nnTnnnTTnTnnnnnnnnnTnnTnnnnunnnnnnnnTnTnTnnnnnnunA
  consensus/70%                     ATGAATTCTTTATTTCCnTTAnTTTACTCAATTGCTTTATTTnTTTTTCTTTTTnTAATTAGTTTTTATATnTTAAAACA
  consensus amino acids             M  N  S  L  F  P  L  X  Y  S  I  A  L  F  X  F  L  F  X  I  S  F  Y  X  L  K  Q

1 Fucus_vesiculosus                 AATACTTAATACTCAAGGCTTAGAAAGAAAAATGTTTGAATTACAGATAATGATAAAAAAAAAATGATGGATCCCATGAAT
2 Ectocarpus_siliculosus            AATAATTAATACTCAAAAGTTAGAAAAAAAAAATATTTTATTTACAGGAGCTTGTTAAAAAAGATGATCTTTACCATGAAG
3 Laminaria_solidungula             AAATTTTAAATATCTTTCTTGACCAACCATTATTTCTAAAAAATAATGGTTGGTTAAGAAAGATAATGTTTCTTATGAAA
4 Undaria_pinnatifida               AATAATTAATACTCAAAAATTAGAAAAAAAGATATTTAGATTACAAGAAAGTATTAAGAAAGATGACGTTTCTTATGAGA
5 Saccharina_japonica               AATAACTAATACCCAAAAATTAGAAAAAAAGATATTTAGGTTACAAGAATCTGTTAAGAAAGATAATGTTTCTTATGAAA
6 Costaria_costata                  AATAACTAATACTCAGAAATTAGAAAAAAAGATATTTAAATTACAGGAAAGTATTAAGAAAGATAATGTTTCTTACGAAA
7 Laminaria_digitata                AATAACTAATACTCAAAAATTAGAAAAAAAGATATTTAAATTACAAGAAAGTGTTAAAAAAGATAATGTTTCTTATGAAA
  consensus/100%                    AAnnnnTAAnnnnCnnnnnTnAnnAAnnAnnATnTnTnunnnAnAunnunnnuTnAAuAAAuATuAnnnnTnnnAnGAun
  consensus/70%                     AATAAnTAATACTCAAAAuTTAGAAAAAAAuATATTTAAATTACAuGAAnnTuTTAAuAAAGATuATGTTTCTTATGAAA
                                     I  X  N  T  Q  N  L  E  K  N  I  F  K  L  H  E  X  F  N  K  D  Y  V  S  Y  E  T

1 Fucus_vesiculosus                 TATATTATAAATTAGGTCAATTATATTTAAAAAAAAAGCTTTTTTCTAAATCAATTTTATTATTTCGTGAAGCGATAAAG
2 Ectocarpus_siliculosus            ATTGTTATCAATTAGGACAATTATATTTAAGAAAAAAACTTTTTTTAAAGGCTATTGTAGTATTTAGAAAGGCTTTAAAA
3 Laminaria_solidungula             CTTTCTATAAATTAGGTCAATTATATTTAAAAAAAAAATTGTTTTATAAAGCTATTTTATTATTTAGACAAGCTTTAAAG
4 Undaria_pinnatifida               CTTTCTATAAACTAGGTCAATTGTATTTAAAAAAAAAATTATTTTATAAAGCTATCTTATTATTTAGACAAGCTTTAAAG
5 Saccharina_japonica               CTTTCTACAAATTAGGTCAATTATATTTAAAAAAAAAATTATTTTATAAAGCTATTTTATTGTTTAGACAAGCTTTAAAA
6 Costaria_costata                  CTTTTTATAAATTAGGTCAATTGTATTTAAAAAAAAAATTATTTTATAAAGCGATTTTATTATTTAGACAAGCTTTAAAG
7 Laminaria_digitata                CTTTTTATAAATTAGGTCAATTATATTTAAAAAAAAAATTGTTTTATAAAGCTATTTTATTATTTAGACAAGCTTTAAAG
  consensus/100%                    nnTnnTAnnAAnTAGGnCAATTuTATTTAAuAAAAAAunTnTTTTnnAAunCnATnnTAnTuTTTnGnnAuGCnnTAAAu
  consensus/70%                     CTTTnTATAAATTAGGTCAATTATATTTAAAAAAAAAATTuTTTTATAAAGCTATTTTATTATTTAGACAAGCTTTAAAG
                                     X  Y  K  L  G  Q  L  Y  L  K  K  K  F  F  Y  K  A  I  L  L  F  R  Q  A  L  K
```

```
1 Fucus_vesiculosus                AACTGGGATATAAATGATAATATTGGACTTGCAAGTTTATATAATAGTATAGGCTTTACTTATTTTACTCTAAAAGAATA
2 Ectocarpus_siliculosus           TTATGGGACCCTAATGATATAATTGGGCTTGGTAACTTATATAATGCTATAGGGTTTACATTTTTTAATTTAGAAGAATA
3 Laminaria_solidungula            GCTTGGAACCCTAATGATAAAATTGGACTTGGAAGTTTATATAATACAATAGGATTTACATATTTTACTCTAAAACAATA
4 Undaria_pinnatifida              GCTTGGAACCCTAATGATAAAATTGGACTTGGTAGTCTATATAATACAATAGGATTTACATATTTTACTTTAAAGCAGTA
5 Saccharina_japonica              GCTTGGAATCCTAATGATAAAATTGGACTTGGAAGTTTATACAATACAATAGGGTTTACCTATTTTACTTTAAAACAATA
6 Costaria_costata                 GCTTGGAACCCTAATGATAAAATTGGACTTGGAAGTTTATATAACACAATAGGATTTACATATTTTACTTTAAAGCAATA
7 Laminaria_digitata               GCTTGGAACCCTAATGATAAAATTGGACTTGGAAGTTTATATAATACAATAGGGTTTACATATTTTACTTTAAAACAATA
  consensus/100%                   nnnTGGuAnnnnAATGATAnnATTGGuCTTGnnAunnTATAnAAnunnATAGGnTTTACnTnTTTTAnTnTAuAunAuTA
  consensus/70%                    GCTTGGAACCCTAATGATAAAATTGGACTTGGAAGTTTATATAATACAATAGGuTTTACATATTTTACTTTAAAACAATA
                                   A  W  N  P  N  D  K  I  G  L  G  S  L  Y  N  T  I  G  F  T  Y  F  T  L  K  Q  X

1 Fucus_vesiculosus                TAATTTAGCAATTTATTATTACAAAATTGCCCTTAAAATTATTCCCGATTATATTGTAGCTTTAATTAATATTGGGTATG
2 Ectocarpus_siliculosus           TGAGTATGCGATCTATTATTATAAAATAGCAGTACAAATTATTCCAGATCATACATTAGCTTTAATAAATCTTGGGTATG
3 Laminaria_solidungula            CAATTTAGCTAATTATTATTATAGTATTGCCATTGAAATTATCCCTGATTATACATTAGCTTTAACCAACCTTGGTTATA
4 Undaria_pinnatifida              CAATTTAGCTAATTATTATTATAGTATCGCTATTGAAATTATACCTGATTATACATTAGCTTTAACAAATCTTGGATATA
5 Saccharina_japonica              CAATTTGGCGAATTATTACTACAGTATTGCTATTGAAATTATCCCTGATTATACATTAGCTTTAACAAATCTTGGTTATA
6 Costaria_costata                 TAGTTTAGCTAATTATTATTACAGTATTGCGATTGAAATTATCCCTGATTATACATTAGCTTTAACCAATCTTGGTTATA
7 Laminaria_digitata               CAATTTAGCTAATTATTATTATAGTATTGCCATTGAAATTATCCCTGATTATACATTAGCTTTAACCAATCTTGGTTATA
  consensus/100%                   nuunTnnGCnAnnTATTAnTAnAunATnGCnnTnnAAATTATnCCnGATnATAnnnTAGCTTTAAnnAAnnTTGGnTATu
  consensus/70%                    nAATTTAGCnAATTATTATTAnAGTATTGCnATTGAAATTATnCCTGATTATACATTAGCTTTAACnAATCTTGGnTATA
                                    N  L  A  N  Y  Y  X  S  I  A  I  E  I  X  P  D  Y  T  L  A  L  T  N  L  G  Y  S

1 Fucus_vesiculosus                CATATGAAAAACAAAATTTATTACTCGAATCTTATAACTCATATAATAAAGTTTTATTTTATAATGCTTATAATAGTTTA
2 Ectocarpus_siliculosus           CCTTTGAAAAAATTAATTCATTTGTAATAGGGTATAATTGTTATAGAGCTGCATTATTTTGGGATACTACTAACGATTTA
3 Laminaria_solidungula            GTTATGAAAAGCTTAACTTATCGGTAGAGTCTTATAATTGTTATAAAAATGCTTTAGCATGGGACCCCAAGAATAGATTA
4 Undaria_pinnatifida              GTTATGAAAAACTTAACTTATCAGTAGAATCTTATAATTGTTATAAGAATGCTTTAGTTTGGGATCCCGAGAATAGATTA
5 Saccharina_japonica              GTTATGAAAAGCTTAACTTATCGGTAGAATCTTATAATTGTTATAAAAATGCTTTAGTATGGGACCCGAAGAATAGATTA
6 Costaria_costata                 GTTATGAAAAACTTAACTTATCTGTAGAGTCTTATAATTGTTATAAGAATGCTTTGGTATGGGATCCTCAAAATAAATTA
7 Laminaria_digitata               GTTATGAAAAACTTAATTTATCGGTAGAGTCTTATAATTGTTATAAAAATGCTTTAGTATGGGACCCTGAGAATAGATTA
  consensus/100%                   nnTnTGAAAAunnnAAnTnATnnnTnunununnnTATAAnTnnTATAununnGnnTTunnnTunuAnnCnnnnAAnuunTTA
  consensus/70%                    GTTATGAAAAACTTAAnTTATCuGTAGAuTCTTATAATTGTTATAAuAATGCTTTAGTnTGGGAnCCnuAuAATAGATTA
                                     Y  E  K  L  X  L  S  V  D  S  Y  N  C  Y  N  N  A  L  V  W  X  P  Y  N  R  L

1 Fucus_vesiculosus                GTTTTAAAAAGAATTAAAATCGT---GAAGAGACTA------TTAATGAGCAAATCGTAA
2 Ectocarpus_siliculosus           GCTTCTACGCGTTTTTTATCGATTGAAAAAAAATTAAGGTATATTCTT---------TAA
3 Laminaria_solidungula            GCTTCGTCACGTATATTAGTTGTTGAAAAGAAACTAAGATATCTAGTTGGTACTAGATAA
4 Undaria_pinnatifida              GCTTCTTCACGTATATTAGTTGTCGAAAAGAAGTTAAGATATTTTGTTGTTACTAAATAA
5 Saccharina_japonica              GCTTCTTCAAGAATATTAGTCGTTGAAAAGAAACTAAGATATCTAGTTGGTACTAGATAA
6 Costaria_costata                 GCTTCTTCACGTATATTAGTGGTTGAAAAGAAGTTAAGATATTTAGTTGGTACTAGATAA
7 Laminaria_digitata               GCCTCTTCACGTATATTAGTCGTTGAAAAGAAACTAAGATATTTAGTTGGTACTAGATAA
  consensus/100%                   GnnTnnnnunGnnTnnnAnnnuT...uAAuAuunTA......nTnnTn.........TAA
  consensus/70%                    GCTTCTTCACGTATATTAGTnGTTGAAAAGAAAnTAAGATATnTAGTTGGTACTAuATAA
                                   A  S  S  R  I  L  V  V  E  K  K  X  R  Y  X  V  G  T  I  *
```

3.3.2.1 Sequence variations across the five chloroplast genomes

The collinearity of the chloroplast genomes allows alignment and definition of sequence variation irrespective of coding, non-coding, or intergenic regions. Since, we however observed small rearrangements in the *Laminaria* species we decided not to use the global alignment for single nucleotide polymorphism (SNP) and insertion or deletion (indel) detection. Instead, we analysed the sequence variations locally using a 100x coverage of artificial reads each, which we mapped to the *S. japonica* genome. In total we found 9,218 SNPs and 164 indels. We counted all SNPs from all species in windows of 1000 bases to examine the SNP distribution over the chloroplast genome (Figure 2-6).

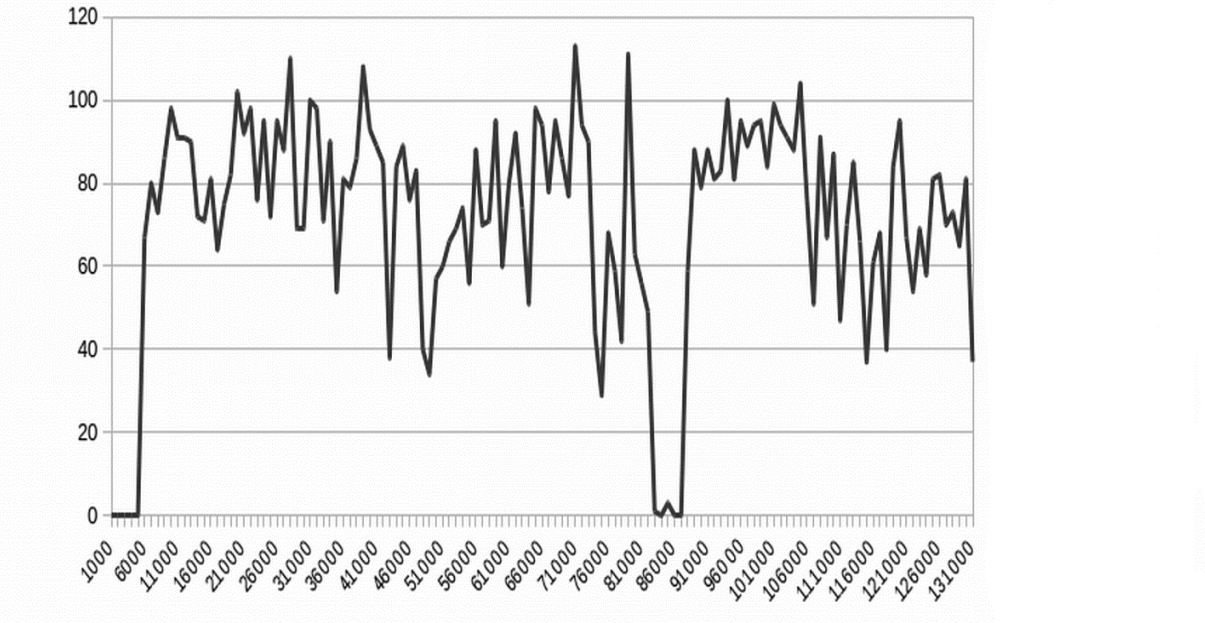


Figure 2- 6: Single nucleotide Polymorphism (SNPs) distribution over the kelp chloroplast genome. SNPs were detected by aligning short reads to the *Saccharina japonica* genome as a reference. All SNPs (Table 2-6) from the aligned reads of the available four kelp species in windows of 1000 bases were counted and plotted. X axis: Base count in the *S. japonica* reference. Y axis: number of SNPs.

SNPs are fairly equally distributed over the whole genome sequence, only the inverted repeat regions are nearly devoid of sequence variation. This phenomenon was already observed in higher plants (Zhu et al. 2016). By far, the highest numbers of unique SNPs are present in the genomes of *U. pinnatifida* and *C. costata* (Figure 2-7). Conversely, the *Laminaria* species have the largest set of SNPs in common (502) which likely evolved with the establishment of this lineage. Not surprisingly, the shared set of both *Laminaria* species with the most distantly related *U. pinnatifida* chloroplast genome is the smallest with 164

(*L. digitata*) and 136 (*L. solidungula*). The 583 SNPs shared between *C. costata* and *U. pinnatifida* likely represent the ancient state of the chloroplast genomes. The overall pattern of SNP evolution indicates that lineage and species-specific SNPs accumulate over time as expected. However, SNPs were frequently observed to be scattered in the phylogeny indicating possible incomplete lineage sorting by, for example recombination of heteroplasmic genomes.

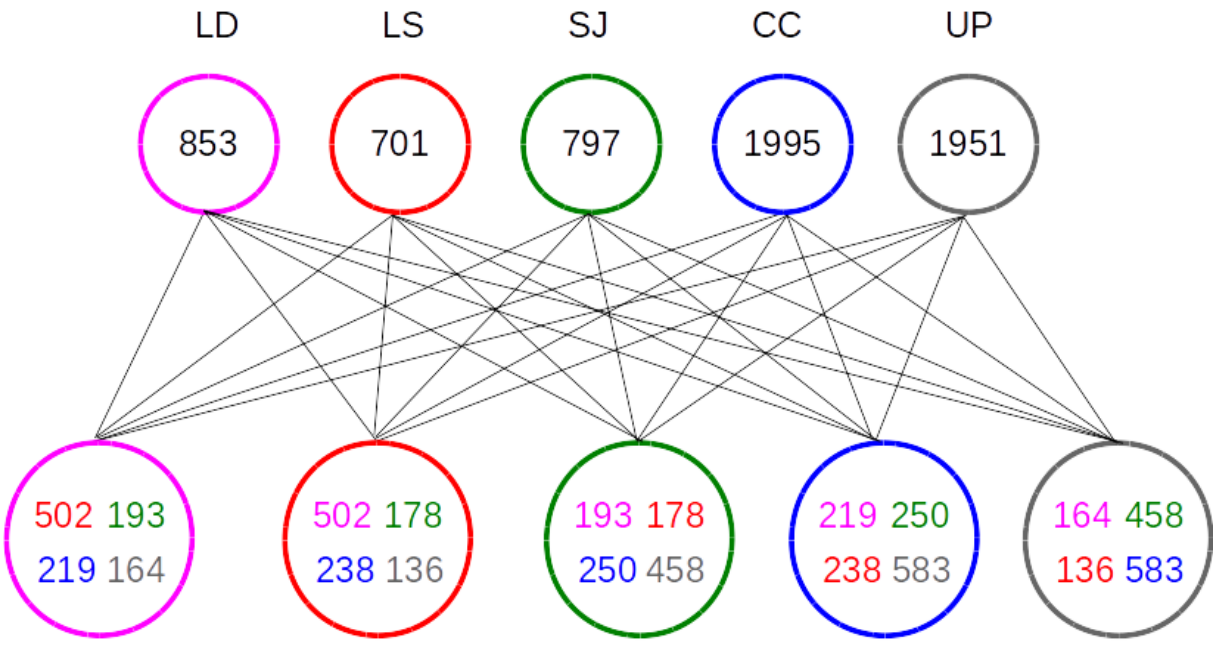


Figure 2- 7: SNPs unique and shared between species. The upper circles show the unique SNPs in each species and the lower row of circles indicates shared SNPs between two species with the numbers in the color of the respective species. To facilitate readability circles are connected by lines. **LD=** *Laminaria digitata* **(magenta); LS=** *L. solidungula* **(red); SJ=** *Saccharina japonica* **(green); CC=***Costaria costata* **(blue); UP=** *Undaria pinnatifida* **(grey).**

Compared to SNPs, indels are rare. In total, we detected 197 indels compared to the *S. japonica* genome. With 59 and 57, the number of indels in *C. costata* and *U. pinnatifida* are highest, whereas *L. solidungula* has only 29 indels and *L. digitata* 36. Indels can only be detected with our method if they are comparably small, that is, in the range of 10 bases. Larger indels exist as the similarity breaks indicate (Figure 2-5). We then examined the ratio of SNPs between genic and intergenic regions (i.e. coding region including RNA genes; Table 2-6). The ratio of genic to intergenic SNPs ranges from 15 to 19 %. The number of detectable SNPs per kb is, however, slightly lower in intergenic compared to genic regions. Since, most larger indels reside in the intergenic regions, the alignability of

these regions is reduced and so the potential to detect SNPs. Overall, the number of SNPs per kb is comparable between intergenic and genic regions in all species (Table 2-6)

Table 2- 6: Number of detected SNPs in genic and intergenic regions. *S. japonica* was used as a reference and artificially generated reads from the other kelp chloroplast genomes were mapped onto this reference.

Total SNPs	Genome length (bp)	Genic regions (bp)	Intergenic regions (bp)	Genic SNPs	Intergenic SNPs	Intergenic/ genic	SNP/kb genic	SNP/kb intergenic
S. japonica	130,584	108,847	21,737					
C. costata	129,947	108,550	21,397	3,615	688	0.2	33.3	32.2
U. pinnatifida	130,383	108,751	21,632	3,633	626	0.2	33.4	28.9
L. solidungula	130,398	108,730	21,668	2,825	429	0.2	26.0	19.8
L. digitata	130,376	108,647	21,729	2,961	566	0.2	27.3	26.0

The distribution of synonymous versus non-synonymous SNPs in coding regions is also of interest (Table 2-7).

Table 2- 7: Synonymous and non-synonymous SNPs in coding regions. The table denotes SNPs occurring in single species versus all others and shared SNPs between two species. Different codon changes denote non synonymous SNPs, which lead to different amino acids in different species.

SNP occurrence	All	Synonymous (s)	Non-synonymous (n)	n/s%
S. japonica	714	610	104	17.0
C. costata	1,596	1,448	148	10.2
U. pinnatifida	1,602	1,352	250	18.5
L. solidungula	569	509	60	11.8
L. digitata	672	600	72	12.0
S. japonica and *C. costata*	224	197	27	13.7
S. japonica and *U. pinnatifida*	404	390	14	3.6
S. japonica and *L. digitata*	159	151	8	5.3
S. japonica and *L. solidungula*	143	142	1	0.7
C. costata and *U. pinnatifida*	492	448	44	9.8
C. costata and *L. solidungula*	207	204	3	1.5
C. costata and *L. digitata*	177	162	15	9.3
U. pinnatifida and *L. digitata*	141	127	14	11.0
U. pinnatifida and *L. solidungula*	118	100	18	18.0

L. digitata and *L. solidungula*	432	391	41	10.5
Total	7,650	6,831	819	12.0
Different codon changes			260	

For this analysis, we calculated for each species the number of SNPs in the two categories and tested, whether those SNPs also occurred in another species. As expected, non-synonymous SNPs are much rarer than synonymous SNPs indicating purifying selection on the coding sequences. Some codons contain different SNPs in different species, resulting sometimes in the encoding of different amino acids. These 260 codons therefore seem to be less constrained in terms of exchangeability. The ratio of non-synonymous to synonymous SNPs ranges from 10.2% to 18.5% in species, and from 0.7% to 18% in species pairs. The somewhat lower values for species pairs might be a result of lower likeliness of maintenance of non-synonymous SNPs in two independent species. Interestingly, *S. japonica* and *U. pinnatifida* have the highest ratio of non-synonymous to synonymous SNPs in their species specific SNPs, which could be due to a less efficient purifying selection or faster accumulation of mutations than in the other species. By calculating the dN/dS ration, we found no evidence for positive selection (i.e., dN/dS > 1) in any of the coding genes of the chloroplast genomes. SNP pairs (i.e. mutations adjacent to each other or multinucleotide polymorphisms [MNPs]) are thought to be not always independent (Prendergast et al. 2019). We analysed such pairs in the kelp chloroplast genomes and found that they are generally rare, but are also partly shared between species (Table 2-8).

Table 2- 8: SNP pairs in kelp chloroplast genomes. Shared pairs between different species are also listed.

Species	**Genic**	**Intergenic**
S. japonica	7	0
C. costata	39	42
U. pinnatifida	31	31
L. solidungula	10	11
L. digitata	9	15
S. japonica and *C. costata*	0	0
S. japonica and *U. pinnatifida*	4	3
S. japonica and *L. digitata*	0	3
S. japonica and *L. solidungula*	2	0
C. costata and *U. pinnatifida*	5	3
C. costata and *L. solidungula*	7	1
C. costata and *L. digitata*	0	2
U. pinnatifida and *L. digitata*	0	0
U. pinnatifida and *L. solidungula*	1	0
L. digitata and *L. solidungula*	2	6
All	117	117

Interestingly, these SNPs are equally distributed between genic and intergenic regions. Since intergenic regions cover a far smaller area of the chloroplast genome, the propensity for this kind of SNPs is to reside in intergenic regions. To exclude the possibility that population structure and sequence variation impact the SNP analyses, we retrieved *L. digitata* samples from 6 different locations (North Sea Helgoland, north east Atlantic Spitsbergen, north west Atlantic Halifax, western Atlantic Connecticut, eastern Atlantic Roscoff and Quiberon). We amplified an 850-bp region containing the cbbx gene and part of the adjacent intergenic region from all samples, cloned the PCR products into vectors, and sequenced three clones each. We could not detect any variation indicating that variation of the chloroplast genome in the whole *L. digitata* population is rare. We then sequenced and assembled the chloroplast plastid genomes from the Helgoland and Spitsbergen isolates and counted the differences to the reference sequence, which was derived from Helgoland. The chloroplast genome of the Spitsbergen isolate contained 27 SNPs and seven small indels, respectively. We therefore conclude that population variation does not impact our SNP analysis across species.

3.4 Analyses of Transcript data

Transcript data analyses was carried out to examine the differential expression genes of the heterologous crosses of *Laminaria digitata* (Spitzbergen and Helgoland), with the help of DESeq2. Corresponding nuclear sequences were extracted from the draft genome, using BLAST tool. To examine SNPs, raw sequencing reads were mapped to the genomic counterparts and the SNPs occurrence was counted in different sets of crosses. SNPs were detected in all 6 crosses (HfHm homozygous Helgoland male and female, HfSm Helgoland female and Spitsbergen male, SfHm Spitsbergen female and Helgoland male and SfSm homozygous Spitsbergen male and female) including day 0 (T0) to day 18 (T18) of *Laminaria digitata*. Total 16285 SNPs were identified by position and occurrence in each sequenced crosses. 911 SNPs were found in Helgoland male and female (HmHf 10°-20°C, T0 & T18). 1836 SNPs were found in Spitsbergen (SfSm 10°-20°C, T0 & T18) crosses, 12435 SNPs in Helgoland and Spitsbergen (HfSm 10°-20°, SfHm 10°-20°C, T0 & T20) crosses and 1103 SNPs were found in all crosses (common) (HfHm,HfSm, SfHm, SfSm 10°-20°C, T0 & T18) (Table 2-9)

Table 2- 9: SNPs counts between all crosses (Helgoland-Spitsbergen 10°-20°C, with T0 &T18) of *Laminaria digitata*

Number of SNPs present in different (any) crosses / Crosses	**HfH m10 °C T0**	**HfH m10 °C T18**	**HfH m20° CT18**	**HfS m10 °C T0**	**HfS m10° CT18**	**HfS m20 °C T18**	**SfH m10 °C T0**	**SfH m10 °C T18**	**SfH m20° CT18**	**SfS m10 °C T0**	**SfS m20 °C T18**
	91	110	0	95	0	109	74	104	121	123	217
	20	47	92	60	92	90	31	55	137	84	134
	27	36	28	21	32	45	24	46	70	48	73
	2	7	11	8	11	9	14	3	13	13	0
	8	10	10	10	10	2	0	0	0	0	0
	31	31	31	31	31	31	0	0	0	0	16
	5	6	6	18	5	20	18	18	16	16	13
	16	15	20	24	20	21	21	21	15	7	49
	165	165	659	660	659	659	660	659	655	502	497
	142	115	152	138	139	139	152	145	99	134	152
Total	507	542	1,009	1,065	999	1,125	994	1,051	1,126	927	1,115

Hf- Helgoland female

Hm- Helgoland male

Sf- Spitsbergen female

Sm- Spitsbergen male

4 Discussion

Kelp species play a crucial role in marine ecosystem providing essential habitat to other marine species and offer valuable economic and ecological resource. Despite of their immense importance, presently for most kelp species very limited knowledge of genetic architecture and phylogeny is available.

4.1 Phylogenetic analyses in kelp organellar genome

Moreover, in the present kelp database, genomic data in particular is very limited, hence, our phylogenetic analysis was constrained to that of the organellar genome. Study of an organelle genome provides essential knowledge of molecular ecology and evolution, furthermore, sequencing technology gives important insight into genome machinery and the phylogenetic history of a species. In general, the number of organellar genomes is found to be higher than the nuclear genome. Therefore, from whole genome sequences the organelle genome can be assembled easily. Here we have determined the complete chloroplast genome sequences of three Laminaria species namely- *Laminaria digitata*, *Laminaria solidungula* and *Laminaria rodriguezii*. And the mitochondrial genome sequences of *Laminaria solidungula* and *Laminaria rodriguezii* were determined and compared with other available kelp organellar genomes (Table 2-1). All chloroplast genomes of kelp species are generally collinear to each other, with some small rearrangements at the inverted repeat (IR) regions. Inverted repeats are sequences of nucleotides which are repeated without any changes, present throughout the genome and as the name suggests have reverse orientation. Mitochondrial genome sequences of all available kelp species (Table 2-1) also show the collinearity to each other. Interestingly, the mitochondrial genome from *D. viridis,* which is a rather a distantly related brown alga from the sister order Desmarestiales, also used as an outgroup in the phylogenetic analysis, was found to contain the same gene order as the kelp species. Owing to that, we could deduce that mitochondrial genomic rearrangements in this lineage are rare, and might even be completely absent. The individual chloroplast and mitochondrial genome sequences of kelp species can be aligned completely, regardless of coding intergenic region. Hence, for the phylogenetic analysis we have more data available than previous studies (Starko et al. 2019; Žuljević et al. 2016), which helped create more robust phylogenetic trees. When compared with previously published comprehensive overview on kelp radiation (Starko et al. 2019), we found no variations in the mitochondrial tree. We therefore are fairly confident of the correct placement of new species within our tree. As

an example, both of the maximum likelihood analyses do place *L. rodriguezii* at the same positions in the phylogenetic tree, and tree topologies of both analyses are also the same (Figure 2-3). In another example, the close relationship between *L. digitata* and *L. rodriguezii* is also reflected in their temperature tolerance profiles. The tree topology within the Laminariaceae, a cytochrome c oxidase (mitochondrial) sequence consisting of only around 2 kb of data (Žuljević et al. 2016), corresponds to that. Thus, the available number of informative sites in organellar genomes, are adequate to determine the kelp phylogenetic tree with high degree of assurance.

4.2 Deviations from kelp collinearity in organellar genomes

4.2.1 Collinearity in mitochondrial genomes

In the process of overall collinearity study in kelp mitochondrial genomes, we detected a long DNA stretch in mitochondrial genome of *L. rodriguezii,* neighbouring to the 16S rRNA gene, with no signs in 17 other kelp counterparts. This DNA stretch comprises of certain piece that can be translated into an ORF, yielding a potential 152 amino acids long protein. This particular protein overlaps with another protein-part encoded by an ORF present in the *D. viridis* mitochondrial genome (Figure 2-4 B) with % identity as high as 72% and a similarity of more than 88%. The underlying nucleotide identity is even higher than the amino acid sequence identity, which points towards either fast evolution or degradation of mitochondrial genome. The mentioned nucleotide sequence of *D. viridis* containing ORF211 is found similar in the region extending from 5' to its start codon (Figure 2-4 A, C), implying that the annotated start codon of this ORF is not the original one. In addition, a 45 bases long stretch in the 5' region of ORF211 also has 80% similarity to ORF2 of *L. rodriguezii*. These found similarities, extending from the originally defined ORF, suggest that the ancient coding gene in *D. viridis* was longer than the existent one. It would be wrong to claim that such similarities on nucleotide level are coincidental rather an outcome of independent lateral gene introduction from same unknown source. Also, vertical transfer between these two species is very unlikely as apparent in the tree (Figure 2-2) they are separated by several other brown algae. Moreover, some remnants of this transfer should have existed in at least some of the other kelp mitochondrial genomes. Additionally, the varying locations of the ORFs in two species disprove a potential vertical transfer in otherwise very strictly collinear genomes. The proposed reason for the nucleotide sequence conservation could either be functional conservation, that is purifying selection, or latest introductions in both species from a similar source, so that, only a handful

mutations could accumulate over time. However, since overlapping stretch showed higher nucleotide conservation than the amino acid conservation, this points rather to functional constraint loss and rapid degradation. That leaves us with the idea, that relatively recent two independent events introduced these similar sequences into the mitochondrial genomes of *L. rodriguezii* and *D. viridis,* species. The origin of such sequences could potentially also be present in the nuclear genome, but we did not find such a conserved gene in the published *Saccharina* species genome (Ye et al. 2015), leaving a horizontal gene transfer (HTG) from external sources the only best explanation. Vectors are often source for such HTG, including so-called mitoviruses- simple RNA(+) viruses consisting of only one reading frame that encodes an RNA dependent RNA polymerase. Until now, mitoviruses are only known from fungi (Hillman and Cai 2013) and plants (Nibert et al. 2018). In fungi generally they are transmitted via spores, mating or cytoplasmatic mixing, but there have been some evidences of transmission between distantly related species by unknown means (Hillman and Cai 2013). These viral RNA polymerases are usually not well preserved (Bartholomäus et al. 2016; Koonin 1991) and therefore, it is not surprising that we did not find any match of the brown algal ORF to published mitovirus sequences. Our hypothesis is that a mitovirus like vector had infested the mitochondria of the two brown algae independently. So, we also sought brown algal genome and transcriptome data for similar sequences, but could not find a positive match. This could be explained by the fact that mitoviruses usually do not integrate into DNA and that's why, are mainly absent in transcript data. Nonetheless, such data are currently limited for brown algae and unfortunately focuses only on polyA+ mRNAs, which limits the likelihood to detect RNA viruses marginal. The gene and virus integration assays tell us that such sequences favourably integrate into highly transcriptionally active genome portions (Christiansen et al. 2015; O'Brien et al. 2018). The mammalian and plant mitochondrial genome analysis also have shown that the rRNA genes are normally expressed at a much higher level than the tRNAs or coding sequences (Finnegan and Brown 1990; Gustafsson, Falkenberg, and Larsson 2016), though sometimes post transcriptional regulation may result in different levels of steady state RNA (Giegé et al. 2000). Considering together, it is intriguing to speculate that the observed locations of the two similar sequences in the vicinity of rRNA genes are due to the higher inclination of DNA pieces to integrate at highly expressed sequences.

4.2.2 Collinearity in chloroplast genomes

Novel plastid genome from kelp species provide important understanding of the diversity and evolution study in kelp phylogeny (Bolton 2010). The chloroplast genomes of different kelp species have almost similar gene contents (Zhang, Wang, Liu, G. Wang, et al. 2015b), making them collinear with some minor additions. In the kelp chloroplast genomes two tRNA genes are inserted, both of them mainly contain introns (Yoshihisa 2014) and are only a second copy of tRNAs species. It might be possible that these tRNAs occur and disappear frequently in evolution and would in this case be dispensable without any effect on collinearity. We observed translocations of genes in *Laminaria* near the IRs (inverted repeat) regions. Such translocation (Li et al. 2016) might be connected to double strand break repair and homologous recombination at IR sites (Raji and Hartsuiker 2006), as it was also witnessed in higher plants (Palmer et al. 1987; Zhu et al. 2016). In *L. solidungula* the translocation of ycf37 in probably led to its defunctionalisation, because the N terminal part including the start codon of the gene is missing as indicated by the alignment (Table 2-5). In the 5' vicinity there was no start codon found, which could have been used as an alternative start from the ribosome. Further work will be needed to confirm whether or not a protein can be created by this truncated gene locus. In another study on *Synechococcus* the functional analysis of a knockout mutant of ycf37 has revealed, that this mutant gene was involved in the specific photosystem I complex formation (Wilde et al. 2001), which seems to be vital under high light conditions (Dühring et al. 2006). It is conceivable that this protein is dispensable under the relative lower light conditions in higher latitudes, for example (Pavlov et al. 2019) the conditions where *L. solidungula* flourishes exclusively (Roleda 2016).

4.2.3 Distribution of single nucleotide polymorphisms (SNPs) across the chloroplast genomes

Single nucleotide polymorphism (SNPs) is a variation, where single or more nucleotides are altered in the genome sequences. SNPs are very common genetic variation distributed throughout the genome. This variation is caused either by a change in single nucleotide or removal (deletion) or addition of new single nucleotide in DNA sequences (Dantas et al. 2009). SNPs can be found in coding region and non-coding region of genes, and also intergenic region between the genes. SNPs variation in human genome can be used to determine the diseases, medication and vaccines, and diversity and evolution when studied in kelp genome (Provan et al. 2013).

The probability of same mutation occurring independently at certain location in different species is very unlikely by evolutionary means (Wray 2007). Hence, if a SNP is located in two species, it should have originated from a single source, that is, one mutation event in the course of evolution. As per our SNP occurrence analysis in chloroplast genomes of kelp species we did not find the correlation to phylogeny, hence we could not trace back the mutation to its source- the first existence of a SNP in the phylogenetic tree. This dispersed occurrence of a SNP, for example, presence in *Undaria pinnatifida* and *Laminaria solidungula* and absence in the other species, does not necessarily point to independent loss of SNP in these lineages. Instead, this particular behaviour can be justified by the presence of heteroplasmic chloroplast genomes with homologous recombination between them (Day and Goldschmidt-Clermont 2011). Consequently, with our study for the first time the incomplete lineage sorting in kelp species has come to light, as it has been known in higher plants (Jakob and Blattner 2006; Sabir et al. 2014). Due to the fact that SNPs are almost equally distributed over the complete chloroplast genomes, the SNP amount per kb cannot be used to distinguish between coding and noncoding regions. This equal scattering of SNPs over the entire chloroplast genome, excluding the inverted repeat regions can be due to the equal constraints on the intergenic and genic regions, if we assume saturation with mutations. This would suggest that intergenic regions encode the regulatory or other functions. The infidelities of the DNA polymerase often leads to the manifestation of multinucleotide mutations, (Schrider, Hourmozdi, and Hahn 2011). In our study we could additionally show that such substitutions are much rarer in coding sequences when compared to intergenic regions. The multinucleotide mutations per kb being scarce in genic regions of the chloroplast genomes is probably due to purifying selection (Su et al. 2018). In kelp chloroplast genomes, we detected a variation of 2.5%-3.3%in pairwise comparisons. For reference, in Gossipium (cotton) species the variation was found to be 0.6% (Xu et al., 2012) with roughly 12.5 million years of divergence time (Wendel, Brubaker, and Seelanan 2010). In Oryza (rice), the variation is 0.36% (Wambugu et al. 2015) with an estimated divergence time of around 10 million years ago (Kellogg 2009). And, around 22 million years ago in Miocene the first kelp forests had found its origin together with grass lands. Thus, evolution of kelp had started much easier in comparison to either rice or cotton species. We hence suggest that evolution of kelp chloroplast genomes have taken the similar course as land plant families.

4.3 Single nucleotide polymorphisms (SNPs) in transcripts of *Laminaria digitata* populations

Different populations of *Laminaria digitata* might harbour different sets of single nucleotide polymorphisms (SNPs). Daniel Liesner from the Alfred Wegener Institute (Bremerhaven) sampled *Laminaria digitata* from Kongsfjorden, Spitsbergen, Norway and Helgoland, Germany, and analysed transcription profiles using DESeq2 to extract the differentially expressed genes, which determines the differences in protein performance or inherent regulation patterns expressed genes (DEG) (Love, Huber, and Anders 2014). Our lab compared the DEGs to draft genome of *Laminaria digitata* and extracted the corresponding genomic regions as a reference. Mapping revealed the allelic differences between the original accessions and the crosses. In total, 200 most highly differentially expressed genes (DEGs) were used for SNPs analyses, which were extracted from all possible parental combinations (HfHm, HfSm, SfHm, SfSm) from each temperature condition (10°C and 20°C) at time point T18. Significantly differentially expressed genes in different populations, and crosses between them might be potentially interesting in terms of SNP distribution across populations and alleles (Avia et al. 2017). SNPs were found across all transcripts including day 0 to day 18, a total of 16285 SNPs are counted. 911 exclusive SNPs are found in Helgoland crosses (HfHm 10°-20°C) and 1836 SNPs in Spitsbergen crosses (SfSm 10°-20°C). Higher Spitsbergen SNPs counts than the Helgoland SNPs might be due to the lower temperature range of Spitsbergen (Arctic region and near to northern distribution) than the Helgoland (North Sea), where in summer temperature goes higher than the other distribution region of *Laminaria digitata* such as, southern distribution in Brittany (Bartsch et al. 2013). 1103 SNPs are found in all crosses (HfHm, HfSm, SfHm and SfSM 10°-20°C) and 12435 SNPs are present in Helgoland and Spitsbergen crosses (HfSm,SfHm 10°-20°C). Interestingly, no allelic expression differences seem to exist within one lineage, which is indicated by 1:1 distribution of all SNPs over all transcripts. This hints to a minor role of allele specific expression in physiological responses. This analysis, however, is ongoing, and might reveal a pattern of adaptation due to SNP combinations. We also counted SNPs individually in different crosses and checked their occurrences (Table 9). Interestingly, no allele specific expression could be found. All alleles are randomly distributed during meiosis, so there are no purely "female" and "male" alleles apart from the sex loci. Therefore, any sex-specific differences in allele expression in sporophytes would have to be linked to the gametophyte sex locus. Physiologically, the (surviving) sporophytes, all behaved very similarly, but there are differences in gene expression among the crosses and selfings, which might be due to local adaptation.

5 References

Alves, Anabela, Rui A. Sousa, and Rui L. Reis. 2013. 'A Practical Perspective on Ulvan Extracted from Green Algae'. *Journal of Applied Phycology* 25(2):407–24.

Anantharaman*, P., G. Karthikaidevi, K. Manivannan, and G. Thirumaran and T. Balasubramanian. 2010. 'Mineral Composition of Marine Macroalgae from Mandapam Coastal Regions; Southeast Coast of India'. *Recent Research in Science and Technology*.

Andersen, Robert A. 2004. 'Biology and Systematics of Heterokont and Haptophyte Algae'. *American Journal of Botany* 91(10):1508–22.

Andersen, Robert A., and Phycological Society of America. 2005. *Algal Culturing Techniques*. Academic Press.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden, and I. G. Young. 1981. 'Sequence and Organization of the Human Mitochondrial Genome'. *Nature* 290(5806):457–65.

Ariyaratne, Pramila Nuwantha, and Wing-Kin Sung. 2011. 'PE-Assembler: De Novo Assembler Using Short Paired-End Reads'. *Bioinformatics* 27(2):167–74.

Arun, Alok, Susana M. Coelho, Akira F. Peters, Simon Bourdareau, Laurent Pérès, Delphine Scornet, Martina Strittmatter, Agnieszka P. Lipinska, Haiqin Yao, Olivier Godfroy, Gabriel J. Montecinos, Komlan Avia, Nicolas Macaisne, Christelle Troadec, Abdelhafid Bendahmane, and J. Mark Cock. 2019. 'Convergent Recruitment of TALE Homeodomain Life Cycle Regulators to Direct Sporophyte Development in Land Plants and Brown Algae' edited by S. McCormick and C. S. Hardtke. *ELife* 8:e43101.

Auwera, Geraldine A. Van der, Mauricio O. Carneiro, Christopher Hartl, Ryan Poplin, Guillermo del Angel, Ami Levy-Moonshine, Tadeusz Jordan, Khalid Shakir, David Roazen, Joel Thibault, Eric Banks, Kiran V. Garimella, David Altshuler, Stacey Gabriel, and Mark A. DePristo. 2013. 'From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline'. *Current Protocols in Bioinformatics* 43(1):11.10.1-11.10.33.

Avia, Komlan, Susana M. Coelho, Gabriel J. Montecinos, Alexandre Cormier, Fiona Lerck, Stéphane Mauger, Sylvain Faugeron, Myriam Valero, J. Mark Cock, and Pierre Boudry. 2017. 'High-Density Genetic Map and Identification of QTLs for Responses to Temperature and Salinity Stresses in the Model Brown Alga Ectocarpus'. *Scientific Reports* 7(1):43241.

Bahassi, El Mustapha, and Peter J. Stambrook. 2014. 'Next-Generation Sequencing Technologies: Breaking the Sound Barrier of Human Genetics'. *Mutagenesis* 29(5):303–10.

Barnes, Harold. 1988. *Oceanography and Marine Biology, An Annual Review*. CRC Press.

Barraclough, Timothy G., and Vincent Savolainen. 2001. 'Evolutionary Rates and Species Diversity in Flowering Plants'. *Evolution* 55(4):677–83.

Barsanti, Laura, and Paolo Gualtieri. 2014. *Algae: Anatomy, Biochemistry, and Biotechnology, Second Edition*. CRC Press.

Bartholomäus, Anika, Daniel Wibberg, Anika Winkler, Alfred Pühler, Andreas Schlüter, and Mark Varrelmann. 2016. 'Deep Sequencing Analysis Reveals the Mycoviral Diversity of the Virome of an Avirulent Isolate of Rhizoctonia Solani AG-2-2 IV' edited by V. Gupta. *PLOS ONE* 11(11):e0165965.

Bartsch, Inka, Johanna Vogt, Constanze Pehlke, and Dieter Hanelt. 2013. 'Prevailing Sea Surface Temperatures Inhibit Summer Reproduction of the Kelp Laminaria Digitata at Helgoland (North Sea)'. *Journal of Phycology* 49(6):1061–73.

Berglund, Eva C., Anna Kiialainen, and Ann-Christine Syvänen. 2011. 'Next-Generation Sequencing Technologies and Applications for Human Genetic History and Forensics'. *Investigative Genetics* 2(1):23.

Birky, C. William. 2001. 'The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models'. *Annual Review of Genetics* 35(1):125–48.

Boetzer, Marten, and Walter Pirovano. 2012. 'Toward Almost Closed Genomes with GapFiller'. *Genome Biology* 13(6):R56.

Bolton, John J. 2010. 'The Biogeography of Kelps (Laminariales, Phaeophyceae): A Global Analysis with New Insights from Recent Advances in Molecular Phylogenetics'. *Helgoland Marine Research* 64(4):263–79.

Borowitzka, Michael Armin, and Navid Reza Moheimani. 2013. 'Sustainable Biofuels from Algae'. *Mitigation and Adaptation Strategies for Global Change* 18(1):13–25.

Brennan, Liam, and Philip Owende. 2010. 'Biofuels from Microalgae—A Review of Technologies for Production, Processing, and Extractions of Biofuels and Co-Products'. *Renewable and Sustainable Energy Reviews* 14(2):557–77.

Bruhn, Annette, Gunhild Brynning, Aase Johansen, Mads Sarauw Lindegaard, Helle Hagen Sveigaard, Betina Aarup, Lene Fonager, Lene Lykke Andersen, Michael Bo Rasmussen, Martin Mørk Larsen, Dieter Elsser-Gravesen, and Margrethe Elise Børsting. 2019. 'Fermentation of Sugar Kelp (Saccharina Latissima)—Effects on Sensory Properties, and Content of Minerals and Metals'. *Journal of Applied Phycology* 31(5):3175–87.

Burtin, Patricia. 2003. 'Nutritional Value of Seaweeds'. *Electron. J. Environ. Agric. Food Chem* 2.

Cameron, Frank K. 1913. 'Kelp and Other Sources of Potash'. *Journal of the Franklin Institute* 176(4):347–83.

Carver, Tim, Ulrike Böhme, Thomas D. Otto, Julian Parkhill, and Matthew Berriman. 2010. 'BamView: Viewing Mapped Read Alignment Data in the Context of the Reference Sequence'. *Bioinformatics* 26(5):676–77.

Chapman, A. R. O., and J. E. Lindley. 1980. 'Seasonal Growth of Laminaria Solidungula in the Canadian High Arctic in Relation to Irradiance and Dissolved Nutrient Concentrations'. *Marine Biology* 57(1):1–5.

Charoensiddhi, Suvimol, Andrew J. Lorbeer, Christopher M. M. Franco, Peng Su, Michael A. Conlon, and Wei Zhang. 2018. 'Process and Economic Feasibility for the Production of Functional Food from the Brown Alga Ecklonia Radiata'. *Algal Research* 29:80–91.

Charrier, Bénédicte, Susana M. Coelho, Aude Le Bail, Thierry Tonon, Gurvan Michel, Philippe Potin, Bernard Kloareg, Catherine Boyen, Akira F. Peters, and J. Mark Cock. 2008. 'Development and Physiology of the Brown Alga Ectocarpus Siliculosus: Two Centuries of Research'. *New Phytologist* 177(2):319–32.

Chaudhury, Nandini Ray, Dimpal Sanghvi, and Bhanukumar Jain. 2018. 'Macroalgae Species as Zonal Indicators of Coral Reef: A Case Study from Bet Shankhodhar Reef, India'. *Wetlands Management - Assessing Risk and Sustainable Solutions*.

Cherry, Paul, Cathal O'Hara, Pamela J. Magee, Emeir M. McSorley, and Philip J. Allsopp. 2019. 'Risks and Benefits of Consuming Edible Seaweeds'. *Nutrition Reviews* 77(5):307–29.

Chisti, Yusuf. 2007. 'Biodiesel from Microalgae'. *Biotechnology Advances* 25(3):294–306.

Christiansen, Irene Kraus, Geir Kjetil Sandve, Martina Schmitz, Matthias Dürst, and Eivind Hovig. 2015. 'Transcriptionally Active Regions Are the Preferred Targets for Chromosomal HPV Integration in Cervical Carcinogenesis' edited by M. L. Tornesello. *PLOS ONE* 10(3):e0119566.

Cole, Kathleen M., and Robert G. Sheath. 1990. *Biology of the Red Algae*. Cambridge University Press.

Cole, Logan W., Wenhu Guo, Jeffrey P. Mower, and Jeffrey D. Palmer. 2018. 'High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants'. *Molecular Biology and Evolution* 35(11):2773–85.

Dankworth, Marie, Sandra Heinrich, Stein Fredriksen, and Inka Bartsch. 2020. 'DNA Barcoding and Mucilage Ducts in the Stipe Reveal the Presence of *Hedophyllum Nigripes* (Laminariales, Phaeophyceae) in Kongsfjorden (Spitsbergen)' edited by M. Edwards. *Journal of Phycology* jpy.13012.

Dantas, Gisele Pires De Mendonça, Raquel Godinho, João Stenghel Morgante, and Nuno Ferrand. 2009. 'Development of New Nuclear Markers and Characterization of Single Nucleotide Polymorphisms in Kelp Gull (Larus Dominicanus)'. *Molecular Ecology Resources* 9(4):1159–61.

Das, Surajit, and Neelam Mangwani. 2015. 'Ocean Acidification and Marine Microorganisms: Responses and Consequences'. *Oceanologia* 57(4):349–61.

Day, Anil, and Michel Goldschmidt-Clermont. 2011. 'The Chloroplast Transformation Toolbox: Selectable Markers and Marker Removal'. *Plant Biotechnology Journal* 9(5):540–53.

Dayton, P. K. 1985. 'Ecology of Kelp Communities'. *Annual Review of Ecology and Systematics* 16(1):215–45.

Dhargalkar, V. K., and X. N. Verlecar. 2009. 'Southern Ocean Seaweeds: A Resource for Exploration in Food and Drugs'. *Aquaculture* 287(3):229–42.

Dieck (Bartsch), I. tom. 1992. 'North Pacific and North Atlantic Digitate Laminaria Species (Phaeophyta): Hybridization Experiments and Temperature Responses'. *Phycologia* 31(2):147–63.

Doyle, Jeffrey. 1991. 'DNA Protocols for Plants'. Pp. 283–93 in *Molecular Techniques in Taxonomy*, *NATO ASI Series*, edited by G. M. Hewitt, A. W. B. Johnston, and J. P. W. Young. Berlin, Heidelberg: Springer.

Draget, Kurt I., and Catherine Taylor. 2011. 'Chemical, Physical and Biological Properties of Alginates and Their Biomedical Implications'. *Food Hydrocolloids* 25(2):251–56.

Dühring, Ulf, Klaus-Dieter Irrgang, Katja Lünser, Julia Kehr, and Annegret Wilde. 2006. 'Analysis of Photosynthetic Complexes from a Cyanobacterial Ycf37 Mutant'. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1757(1):3–11.

Dyall, Sabrina D., Mark T. Brown, and Patricia J. Johnson. 2004. 'Ancient Invasions: From Endosymbionts to Organelles'. *Science* 304(5668):253–57.

Edgar, Robert C. 2004. 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput'. *Nucleic Acids Research* 32(5):1792–97.

El Arroussi, Hicham, Redouane Benhima, Najib El Mernissi, Rachid Bouhfid, Chakib Tilsaghani, Iman Bennis, and Imane Wahby. 2017. 'Screening of Marine Microalgae Strains from Moroccan Coasts for Biodiesel Production'. *Renewable Energy* 113:1515–22.

Ercegović, A. 1960. 'Značajne crte vegetacije alga Jadranskog mora'. *Acta Botanica Croatica* 18–19(1):17–36.

Finnegan, P. M., and G. G. Brown. 1990. 'Transcriptional and Post-Transcriptional Regulation of RNA Levels in Maize Mitochondria.' *The Plant Cell* 2(1):71–83.

Fleurence, J., M. Morançais, and J. Dumay. 2018. '9 - Seaweed Proteins'. Pp. 245–62 in *Proteins in Food Processing (Second Edition)*, *Woodhead Publishing Series in Food Science, Technology and Nutrition*, edited by R. Y. Yada. Woodhead Publishing.

Fleurence, Joël. 1999. 'Seaweed Proteins: Biochemical, Nutritional Aspects and Potential Uses'. *Trends in Food Science & Technology* 10(1):25–28.

Fleurence, Joël, Michèle Morançais, Justine Dumay, Priscilla Decottignies, Vincent Turpin, Mathilde Munier, Nuria Garcia-Bueno, and Pascal Jaouen. 2012. 'What Are the Prospects for Using Seaweed in Human Nutrition and for Marine Animals Raised through Aquaculture?' *Trends in Food Science & Technology* 27(1):57–61.

Frampton, Matthew, and Richard Houlston. 2012. 'Generation of Artificial FASTQ Files to Evaluate the Performance of Next-Generation Sequencing Pipelines'. *PLOS ONE* 7(11):e49110.

Francisco, Érika C., Débora B. Neves, Eduardo Jacob-Lopes, and Telma T. Franco. 2010. 'Microalgae as Feedstock for Biodiesel Production: Carbon Dioxide Sequestration, Lipid Production and Biofuel Quality'. *Journal of Chemical Technology & Biotechnology* 85(3):395–403.

Fu, Gang, Nana Kinoshita, Chikako Nagasato, and Taizo Motomura. 2014. 'Fertilization of Brown Algae: Flagellar Function in Phototaxis and Chemotaxis'. Pp. 359–67 in *Sexual Reproduction in Animals and Plants*, edited by H. Sawada, N. Inoue, and M. Iwano. Tokyo: Springer Japan.

Gall, Erwan Ar, Frithjof C. Küpper, and Bernard Kloareg. 2004. 'A Survey of Iodine Content in Laminaria Digitata'. *Botanica Marina* 47(1):30–37.

Gan, Han Ming, Mark B. Schultz, and Christopher M. Austin. 2014. 'Integrated Shotgun Sequencing and Bioinformatics Pipeline Allows Ultra-Fast Mitogenome Recovery and Confirms Substantial Gene Rearrangements in Australian Freshwater Crayfishes'. *BMC Evolutionary Biology* 14(1):19.

Gerard, V. A. 1982. 'In Situ Water Motion and Nutrient Uptake by the Giant Kelp Macrocystis Pyrifera'. *Marine Biology* 69(1):51–54.

Giegé, Philippe, Michaela Hoffmann, Stefan Binder, and Axel Brennicke. 2000. 'RNA Degradation Buffers Asymmetries of Transcription in Arabidopsis Mitochondria'. *EMBO Reports* 1(2):164–70.

Glöckner, Gernot, André Rosenthal, and Klaus Valentin. 2000. 'The Structure and Gene Repertoire of an Ancient Red Algal Plastid Genome'. *Journal of Molecular Evolution* 51(4):382–90.

Gómez-Díaz, Diego, and José M. Navaza. 2003. 'Rheology of Aqueous Solutions of Food Additives: Effect of Concentration, Temperature and Blending'. *Journal of Food Engineering* 56(4):387–92.

Gomez-Zavaglia, Andrea, Miguel A. Prieto Lage, Cecilia Jimenez-Lopez, Juan C. Mejuto, and Jesus Simal-Gandara. 2019. 'The Potential of Seaweeds as a Source of Functional Ingredients of Prebiotic and Antioxidant Value'. *Antioxidants* 8(9).

Gould, Sven B., Ross F. Waller, and Geoffrey I. McFadden. 2008. 'Plastid Evolution'. *Annual Review of Plant Biology* 59(1):491–517.

Great Barrier Reef Marine Park Authority. 2003. *State of the Great Barrier Reef Report.* Townsville, Qld.: Great Barrier Reef Marine Park Authority.

Greiner, Stephan, Pascal Lehwark, and Ralph Bock. 2019. 'OrganellarGenomeDRAW (OGDRAW) Version 1.3.1: Expanded Toolkit for the Graphical Visualization of Organellar Genomes'. *Nucleic Acids Research* 47(W1):W59–64.

Guiry, Michael D. 2012. 'How Many Species of Algae Are There?' *Journal of Phycology* 48(5):1057–63.

Gustafsson, Claes M., Maria Falkenberg, and Nils-Göran Larsson. 2016. 'Maintenance and Expression of Mammalian Mitochondrial DNA'. *Annual Review of Biochemistry* 85(1):133–60.

Hasselström, Linus, Jean-Baptiste Thomas, Jonas Nordström, Gunnar Cervin, Göran M. Nylund, Henrik Pavia, and Fredrik Gröndahl. 2020. 'Socioeconomic Prospects of a Seaweed Bioeconomy in Sweden'. *Scientific Reports* 10(1):1610.

Hillman, Bradley I., and Guohong Cai. 2013. 'Chapter Six - The Family Narnaviridae: Simplest of RNA Viruses'. Pp. 149–76 in *Advances in Virus Research*. Vol. 86, *Mycoviruses*, edited by S. A. Ghabrial. Academic Press.

Jackson, G. A. 1987. 'Modelling the Growth and Harvest Yield of the Giant Kelp Macrocystis Pyrifera'. *Marine Biology* 95(4):611–24.

Jakob, Sabine S., and Frank R. Blattner. 2006. 'A Chloroplast Genealogy of Hordeum (Poaceae): Long-Term Persisting Haplotypes, Incomplete Lineage Sorting, Regional Extinction, and the Consequences for Phylogenetic Inference'. *Molecular Biology and Evolution* 23(8):1602–12.

Javaux, Emmanuelle J., and Kevin Lepot. 2018. 'The Paleoproterozoic Fossil Record: Implications for the Evolution of the Biosphere during Earth's Middle-Age'. *Earth-Science Reviews* 176:68–86.

Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, and S. Hunter. 2014. 'InterProScan 5: Genome-Scale Protein Function Classification'. *Bioinformatics* 30(9):1236–40.

Katoh, Kazutaka, and Daron M. Standley. 2013. 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability'. *Molecular Biology and Evolution* 30(4):772–80.

Kawai, Hiroshi, and Eric C. Henry. 2017. 'Phaeophyta'. Pp. 267–304 in *Handbook of the Protists*, edited by J. M. Archibald, A. G. B. Simpson, and C. H. Slamovits. Cham: Springer International Publishing.

Keeling, Patrick J. 2010. 'The Endosymbiotic Origin, Diversification and Fate of Plastids'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1541):729–48.

Kellogg, Elizabeth A. 2009. 'The Evolutionary History of Ehrhartoideae, Oryzeae, and Oryza'. *Rice* 2(1):1–14.

Kolb, Nada, Luciana Vallorani, Nada Milanović, and Vilberto Stocchi. 2004. 'Evaluation of Marine Algae Wakame (Undaria Pinnatifida) and Kombu (Laminaria Digitata Japonica) as Food Supplements'. *Food Technology and Biotechnology* 42(1):57–61.

Koonin, Eugene V. 1991. 'The Phylogeny of RNA-Dependent RNA Polymerases of Positive-Strand RNA Viruses'. *Journal of General Virology,* 72(9):2197–2206.

Koonin, Eugene V. 2010. 'The Origin and Early Evolution of Eukaryotes in the Light of Phylogenomics'. *Genome Biology* 11(5):209.

Krumhansl, Kira A., Daniel K. Okamoto, Andrew Rassweiler, Mark Novak, John J. Bolton, Kyle C. Cavanaugh, Sean D. Connell, Craig R. Johnson, Brenda Konar, Scott D. Ling, Fiorenza Micheli, Kjell M. Norderhaug, Alejandro Pérez-Matus, Isabel Sousa-Pinto, Daniel C. Reed, Anne K. Salomon, Nick T. Shears, Thomas Wernberg, Robert J. Anderson, Nevell S. Barrett, Alejandro H. Buschmann, Mark H. Carr, Jennifer E. Caselle, Sandrine Derrien-Courtel, Graham J. Edgar, Matt Edwards, James A. Estes, Claire Goodwin, Michael C. Kenner, David J. Kushner, Frithjof E. Moy, Julia Nunn, Robert S. Steneck, Julio Vásquez, Jane Watson, Jon D. Witman, and Jarrett E. K. Byrnes. 2016. 'Global Patterns of Kelp Forest Change over the Past Half-Century'. *Proceedings of the National Academy of Sciences* 113(48):13785–90.

Ku, Chuan, Shijulal Nelson-Sathi, Mayo Roettger, Filipa L. Sousa, Peter J. Lockhart, David Bryant, Einat Hazkani-Covo, James O. McInerney, Giddy Landan, and William F. Martin. 2015. 'Endosymbiotic Origin and Differential Loss of Eukaryotic Genes'. *Nature* 524(7566):427–32.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms' edited by F. U. Battistuzzi. *Molecular Biology and Evolution* 35(6):1547–49.

Kurtz, Stefan, Adam Phillippy, Arthur L. Delcher, Michael Smoot, Martin Shumway, Corina Antonescu, and Steven L. Salzberg. 2004. 'Versatile and Open Software for Comparing Large Genomes'. *Genome Biology* 5(2):R12.

Langmead, Ben, and Steven L. Salzberg. 2012. 'Fast Gapped-Read Alignment with Bowtie 2'. *Nature Methods* 9(4):357–59.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. 'Clustal W and Clustal X Version 2.0'. *Bioinformatics* 23(21):2947–48.

Le Quéré, Corinne, Robbie M. Andrew, Pierre Friedlingstein, Stephen Sitch, Julia Pongratz, Andrew C. Manning, Jan Ivar Korsbakken, Glen P. Peters, Josep G. Canadell, Robert B. Jackson, Thomas A. Boden, Pieter P. Tans, Oliver D. Andrews, Vivek K. Arora, Dorothee C. E. Bakker, Leticia Barbero, Meike Becker, Richard A. Betts, Laurent Bopp, Frédéric Chevallier, Louise P. Chini, Philippe Ciais, Catherine E. Cosca, Jessica Cross, Kim Currie, Thomas Gasser, Ian Harris, Judith Hauck, Vanessa Haverd, Richard A. Houghton, Christopher W. Hunt, George Hurtt, Tatiana Ilyina, Atul K. Jain, Etsushi Kato, Markus Kautz, Ralph F. Keeling, Kees Klein Goldewijk, Arne Körtzinger, Peter Landschützer, Nathalie Lefèvre, Andrew Lenton, Sebastian Lienert, Ivan Lima, Danica Lombardozzi, Nicolas Metzl, Frank Millero, Pedro M. S. Monteiro, David R. Munro, Julia E. M. S. Nabel, Shin-ichiro Nakaoka, Yukihiro Nojiri, X. Antoni Padín, Anna Peregon, Benjamin Pfeil, Denis Pierrot, Benjamin Poulter, Gregor Rehder, Janet Reimer, Christian Rödenbeck, Jörg Schwinger, Roland Séférian, Ingunn Skjelvan, Benjamin D. Stocker, Hanqin Tian, Bronte Tilbrook, Ingrid T. van der Laan-Luijkx, Guido R. van der Werf, Steven van Heuven, Nicolas Viovy, Nicolas Vuichard, Anthony P. Walker, Andrew J. Watson, Andrew J. Wiltshire, Sönke Zaehle, and Dan Zhu. 2017. 'Global Carbon Budget 2017'. *Earth System Science Data Discussions* 1–79.

Levasseur, Wendie, Patrick Perré, and Victor Pozzobon. 2020. 'A Review of High Value-Added Molecules Production by Microalgae in Light of the Classification'. *Biotechnology Advances* 41:107545.

Li, Fay-Wei, Li-Yaung Kuo, Kathleen M. Pryer, and Carl J. Rothfels. 2016. 'Genes Translocated into the Plastid Inverted Repeat Show Decelerated Substitution Rates and Elevated GC Content'. *Genome Biology and Evolution* 8(8):2452–58.

Li, Heng, Jue Ruan, and Richard Durbin. 2008. 'Mapping Short DNA Sequencing Reads and Calling Variants Using Mapping Quality Scores'. *Genome Research* 18(11):1851–58.

Littler, Mark M., and Diane S. Littler. 2011. 'Algae-Macro'. Pp. 30–38 in *Encyclopedia of Modern Coral Reefs: Structure, Form and Process*, edited by D. Hopley. Dordrecht: Springer Netherlands.

Liu, Xiaojie, Kenny Bogaert, Aschwin H. Engelen, Frederik Leliaert, Michael Y. Roleda, and Olivier De Clerck. 2017. 'Seaweed Reproductive Biology: Environmental and Genetic Controls'. *Botanica Marina* 60(2):89–108.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. 'Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2'. *Genome Biology* 15(12):550.

Lowe, Todd M., and Sean R. Eddy. 1997. 'TRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence'. *Nucleic Acids Research* 25(5):955–64.

Lüning, K., and I. tom Dieck. 1989. 'Environmental Triggers in Algal Seasonality'. *Botanica Marina* 32(5).

Lüning, Klaus. 1990. *Seaweeds: Their Environment, Biogeography, and Ecophysiology*. John Wiley & Sons.

Macaya, E. C., and G. C. Zuccarello. 2010. 'Genetic Structure of the Giant Kelp Macrocystis Pyrifera along the Southeastern Pacific'. *Marine Ecology Progress Series* 420:103–12.

Mamatha, B. S., K. K. Namitha, Amudha Senthil, J. Smitha, and G. A. Ravishankar. 2007. 'Studies on Use of Enteromorpha in Snack Food'. *Food Chemistry* 101(4):1707–13.

Marguerat, Samuel, Brian T. Wilhelm, and Jürg Bähler. 2008. 'Next-Generation Sequencing: Applications beyond Genomes'. *Biochemical Society Transactions* 36(5):1091–96.

McWilliam, Hamish, Weizhong Li, Mahmut Uludag, Silvano Squizzato, Young Mi Park, Nicola Buso, Andrew Peter Cowley, and Rodrigo Lopez. 2013. 'Analysis Tool Web Services from the EMBL-EBI'. *Nucleic Acids Research* 41(W1):W597–600.

Medvedev, Paul, Monica Stanciu, and Michael Brudno. 2009. 'Computational Methods for Discovering Structural Variation with Next-Generation Sequencing'. *Nature Methods* 6(11):S13–20.

Mendes, Ana, Teresa Lopes da Silva, and Alberto Reis. 2007. 'DHA Concentration and Purification from the Marine Heterotrophic Microalga Crypthecodinium Cohnii CCMP 316 by Winterization and Urea Complexation'. *Food Technology and Biotechnology* 45(1):38–44.

Mundt, Florian, Sandra Heinrich, and Dieter Hanelt. 2019. 'RNA Isolation from Taxonomically Diverse Photosynthetic Protists'. *Limnology and Oceanography: Methods* 17(3):190–99.

Muñoz, Raul, and Benoit Guieysse. 2006. 'Algal–Bacterial Processes for the Treatment of Hazardous Contaminants: A Review'. *Water Research* 40(15):2799–2815.

Nagle, Nick, and Pete Lemke. 1990. 'Production of Methyl Ester Fuel from Microalgae'. *Applied Biochemistry and Biotechnology* 24(1):355–61.

Nei, Masatoshi, and Sudhir Kumar. 2000. *Molecular Evolution and Phylogenetics*. Oxford University Press.

Neori, Amir, Thierry Chopin, Max Troell, Alejandro H. Buschmann, George P. Kraemer, Christina Halling, Muki Shpigel, and Charles Yarish. 2004. 'Integrated Aquaculture: Rationale, Evolution and State of the Art Emphasizing Seaweed Biofiltration in Modern Mariculture'. *Aquaculture* 231(1):361–91.

Neori, Amir, Michael D. Krom, Steve P. Ellner, Claude E. Boyd, Dan Popper, Ruth Rabinovitch, Patrick J. Davison, Orit Dvir, Daniel Zuber, Michal Ucko, Dror Angel, and Hillel Gordin. 1996. 'Seaweed Biofilters as Regulators of Water Quality in Integrated Fish-Seaweed Culture Units'. *Aquaculture* 141(3):183–99.

Nibert, Max L., Minh Vong, Karen K. Fugate, and Humberto J. Debat. 2018. 'Evidence for Contemporary Plant Mitoviruses'. *Virology* 518:14–24.

Nisizawa, Kazutosi, Hiroyuki Noda, Ryo Kikuchi, and Tadaharu Watanabe. 1987. 'The Main Seaweed Foods in Japan'. *Hydrobiologia* 151(1):5–29.

Norall, T. L., A. C. Mathieson, and J. A. Kilar. 1981. 'Reproductive Ecology of Four Subtidal Red Algae'. *Journal of Experimental Marine Biology and Ecology* 54(2):119–36.

O'Brien, Sofie A., Kyoungho Lee, Hsu-Yuan Fu, Zion Lee, Tung S. Le, Christopher S. Stach, Meghan G. McCann, Alicia Q. Zhang, Michael J. Smanski, Nikunj V. Somia, and Wei-Shou Hu. 2018. 'Single Copy Transgene Integration in a Transcriptionally Active Site for Recombinant Protein Synthesis'. *Biotechnology Journal* 13(10):1800226.

Oppliger, Luz Valeria, Peter von Dassow, Sarah Bouchemousse, Marine Robuchon, Myriam Valero, Juan A. Correa, Stéphane Mauger, and Christophe Destombe. 2014. 'Alteration of Sexual Reproduction and Genetic Diversity in the Kelp Species Laminaria Digitata at the Southern Limit of Its Range'. *PLoS ONE* 9(7).

Palmer, Jeffrey D., Bernardita Osorio, Jane Aldrich, and William F. Thompson. 1987. 'Chloroplast DNA Evolution among Legumes: Loss of a Large Inverted Repeat Occurred Prior to Other Sequence Rearrangements'. *Current Genetics* 11(4):275–86.

Patarra, Rita Ferreira, Lisete Paiva, Ana Isabel Neto, Elisabete Lima, and José Baptista. 2011. 'Nutritional Value of Selected Macroalgae'. *Journal of Applied Phycology* 23(2):205–8.

Pavlov, Alexey K., Eva Leu, Dieter Hanelt, Inka Bartsch, Ulf Karsten, Stephen R. Hudson, Jean-Charles Gallet, Finlo Cottier, Jonathan H. Cohen, Jørgen Berge, Geir Johnsen, Marion Maturilli, Piotr Kowalczuk, Sławomir Sagan, Justyna Meler, and Mats A. Granskog. 2019. 'The Underwater Light Climate in Kongsfjorden and Its Ecological Implications'. Pp. 137–70 in *The Ecosystem of Kongsfjorden, Svalbard*, *Advances in Polar Ecology*, edited by H. Hop and C. Wiencke. Cham: Springer International Publishing.

Podkorytova, A. V., L. H. Vafina, E. A. Kovaleva, and V. I. Mikhailov. 2007. 'Production of Algal Gels from the Brown Alga, Laminaria Japonica Aresch., and Their Biotechnological Applications'. *Journal of Applied Phycology* 19(6):827–30.

Prabhudas, Sudheesh K., Sowjanya Prayaga, Parani Madasamy, and Purushothaman Natarajan. 2016. 'Shallow Whole Genome Sequencing for the Assembly of Complete Chloroplast Genome Sequence of Arachis Hypogaea L.' *Frontiers in Plant Science* 7.

Prendergast, James G. D., Carys Pugh, Sarah E. Harris, David A. Hume, Ian J. Deary, and Allan Beveridge. 2019. 'Linked Mutations at Adjacent Nucleotides Have Shaped Human Population Differentiation and Protein Evolution'. *Genome Biology and Evolution* 11(3):759–75.

Provan, Jim, Keith Glendinning, Ruth Kelly, and Christine A. Maggs. 2013. 'Levels and Patterns of Population Genetic Diversity in the Red Seaweed Chondrus Crispus (Florideophyceae): A Direct Comparison of Single Nucleotide Polymorphisms and Microsatellites'. *Biological Journal of the Linnean Society* 108(2):251–62.

Raji, Hayatu, and Edgar Hartsuiker. 2006. 'Double-Strand Break Repair and Homologous Recombination in Schizosaccharomyces Pombe'. *Yeast* 23(13):963–76.

Rana, Shivani, Klaus Valentin, Inka Bartsch, and Gernot Glöckner. 2019. 'Loss of a Chloroplast Encoded Function Could Influence Species Range in Kelp'. *Ecology and Evolution* 9(15):8759–70.

Rasala, Beth A., and Stephen P. Mayfield. 2015. 'Photosynthetic Biomanufacturing in Green Algae; Production of Recombinant Proteins for Industrial, Nutritional, and Medical Uses'. *Photosynthesis Research* 123(3):227–39.

Roleda, Michael Y. 2016. 'Stress Physiology and Reproductive Phenology of Arctic Endemic Kelp Laminaria Solidungula J. Agardh'. *Polar Biology* 39(11):1967–77.

Ronquist, Fredrik, Maxim Teslenko, Paul van der Mark, Daniel L. Ayres, Aaron Darling, Sebastian Höhna, Bret Larget, Liang Liu, Marc A. Suchard, and John P. Huelsenbeck. 2012. 'MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space'. *Systematic Biology* 61(3):539–42.

Roure, Béatrice, Naiara Rodriguez-Ezpeleta, and Hervé Philippe. 2007. 'SCaFoS: A Tool for Selection, Concatenation and Fusion of Sequences for Phylogenomics'. *BMC Evolutionary Biology* 7(1):S2.

Sabir, Jamal S. M., Dhivya Arasappan, Ahmed Bahieldin, Salah Abo-Aba, Sameera Bafeel, Talal A. Zari, Sherif Edris, Ahmed M. Shokry, Nour O. Gadalla, Ahmed M. Ramadan, Ahmed Atef, Magdy A. Al-Kordy, Fotoh M. El-Domyati, and Robert K. Jansen. 2014. 'Whole Mitochondrial and Plastid Genome SNP Analysis of Nine Date Palm Cultivars Reveals Plastid Heteroplasmy and Close Phylogenetic Relationships among Cultivars'. *PLOS ONE* 9(4):e94158.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. 'DNA Sequencing with Chain-Terminating Inhibitors'. *Proceedings of the National Academy of Sciences* 74(12):5463–67.

Schaffelke, B., and K. Lüning. 1994. 'A Circannual Rhythm Controls Seasonal Growth in the Kelps Laminaria Hyperborea and L. Digitata from Helgoland (North Sea)'. *European Journal of Phycology* 29(1):49–56.

Schiel, David R., and Michael S. Foster. 2006. 'The Population Biology of Large Brown Seaweeds: Ecological Consequences of Multiphase Life Histories in Dynamic Coastal Environments'. *Annual Review of Ecology, Evolution, and Systematics* 37(1):343–72.

Schmid, Matthias, and Dagmar B. Stengel. 2015. 'Intra-Thallus Differentiation of Fatty Acid and Pigment Profiles in Some Temperate Fucales and Laminariales'. *Journal of Phycology* 51(1):25–36.

Schrider, Daniel R., Jonathan N. Hourmozdi, and Matthew W. Hahn. 2011. 'Pervasive Multinucleotide Mutational Events in Eukaryotes'. *Current Biology* 21(12):1051–54.

Schuster, Stephan C. 2008. 'Next-Generation Sequencing Transforms Today's Biology'. *Nature Methods* 5(1):16–18.

Ścieszka, Sylwia, and Elżbieta Klewicka. 2019. 'Algae in Food: A General Review'. *Critical Reviews in Food Science and Nutrition* 59(21):3538–47.

Secq, Marie-Pierre Oudot-Le, Susan Loiseaux-de Goër, Wytze T. Stam, and Jeanine L. Olsen. 2006. 'Complete Mitochondrial Genomes of the Three Brown Algae (Heterokonta: Phaeophyceae) Dictyota Dichotoma, Fucus Vesiculosus and Desmarestia Viridis'. *Current Genetics* 49(1):47–58.

Secq, Marie-Pierre Oudot-Le, Bernard Kloareg, and Susan Loiseaux-De Goër. 2002. 'The Mitochondrial Genome of the Brown Alga Laminaria Digitata: A Comparative Analysis'. *European Journal of Phycology* 37(2):163–72.

Sekimoto, Hiroyuki. 2017. 'Sexual Reproduction and Sex Determination in Green Algae'. *Journal of Plant Research* 130(3):423–31.

Sheath, Robert G., and Morgan L. Vis. 2015. 'Chapter 5 - Red Algae'. Pp. 237–64 in *Freshwater Algae of North America (Second Edition)*, *Aquatic Ecology*, edited by J. D. Wehr, R. G. Sheath, and J. P. Kociolek. Boston: Academic Press.

Shendure, Jay, and Hanlee Ji. 2008. 'Next-Generation DNA Sequencing'. *Nature Biotechnology* 26(10):1135–45.

Shibata, Yasuyuki, Kazuo Jin, and Masatoshi Morita. 1990. 'Arsenic Compounds in the Edible Red Alga, Porphyra Tenera, and in Nori and Yakinori, Food Items Produced from Red Algae'. *Applied Organometallic Chemistry* 4(3):255–60.

Sievers, Fabian, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D. Thompson, and Desmond G. Higgins. 2011. 'Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega'. *Molecular Systems Biology* 7(1):539.

Simpson, Jared T., Kim Wong, Shaun D. Jackman, Jacqueline E. Schein, Steven J. M. Jones, and İnanç Birol. 2009. 'ABySS: A Parallel Assembler for Short Read Sequence Data'. *Genome Research* 19(6):1117–23.

Singh, R. N., and Shaishav Sharma. 2012. 'Development of Suitable Photobioreactor for Algae Production – A Review'. *Renewable and Sustainable Energy Reviews* 16(4):2347–53.

Smale, Dan A., Michael T. Burrows, Pippa Moore, Nessa O'Connor, and Stephen J. Hawkins. 2013. 'Threats and Knowledge Gaps for Ecosystem Services Provided by Kelp Forests: A Northeast Atlantic Perspective'. *Ecology and Evolution* 3(11):4016–38.

Smetacek, Victor, and Adriana Zingone. 2013. 'Green and Golden Seaweed Tides on the Rise'. *Nature* 504(7478):84–88.

Starko, Samuel, Marybel Soto Gomez, Hayley Darby, Kyle W. Demes, Hiroshi Kawai, Norishige Yotsukura, Sandra C. Lindstrom, Patrick J. Keeling, Sean W. Graham, and Patrick T. Martone. 2019. 'A Comprehensive Kelp Phylogeny Sheds Light on the Evolution of an Ecosystem'. *Molecular Phylogenetics and Evolution* 136:138–50.

STENECK, ROBERT S., MICHAEL H. GRAHAM, BRUCE J. BOURQUE, DEBBIE CORBETT, JON M. ERLANDSON, JAMES A. ESTES, and MIA J. TEGNER. 2002. 'Kelp Forest Ecosystems: Biodiversity, Stability, Resilience and Future'. *Environmental Conservation* 29(4):436–59.

Stiller, John W., and Benjamin D. Hall. 1997. 'The Origin of Red Algae: Implications for Plastid Evolution'. *Proceedings of the National Academy of Sciences* 94(9):4520–25.

Su, Huei-Jiun, Todd Barkman, Weilong Hao, Samuel Jones, Julia Naumann, Elizabeth Skippington, Eric Wafula, Jer-Ming Hu, Jeffrey Palmer, and Claude dePamphilis. 2018. 'Novel Genetic Code and Record-Setting AT-Richness in the Highly Reduced Plastid Genome of the Holoparasitic Plant Balanophora'. *Proceedings of the National Academy of Sciences* 116:201816822.

Tamura, Koichiro, Glen Stecher, Daniel Peterson, Alan Filipski, and Sudhir Kumar. 2013. 'MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0'. *Molecular Biology and Evolution* 30(12):2725–29.

Tatewaki, Masakazu. 1966. 'Formation of a Crustaceous Sporophyte with Unilocular Sporangia in Scytosiphon Lomentaria'. *Phycologia* 6(1):62–66.

Teagle, Harry, Stephen J. Hawkins, Pippa J. Moore, and Dan A. Smale. 2017. 'The Role of Kelp Species as Biogenic Habitat Formers in Coastal Marine Ecosystems'. *Journal of Experimental Marine Biology and Ecology* 492:81–98.

Tineo, Daniel, Karol B. Rubio, Jegnes B. Melendez, Jani E. Mendoza, Jhonsy O. Silva, Jhordy Perez, Eggleantina E. Esquerre, Melissa Perez-Alania, Samia L. Fernandez, Smith E. Aguilar, Fernando Chuquizuta, Yadira M. Olano, Renzo P. Hoyos, Jaris E. Veneros, Ligia M. Garcia, Natalia Arakaki, Enrique Garcia-Candela, Manuel Oliva, Andres Mansilla, Martha S. Calderon, Jeffery R. Hughey, and Danilo E. Bustamante. 2019. 'Analysis of the Complete Organellar Genomes of the Economically Valuable Kelp Lessonia Spicata (Lessoniaceae, Phaeophyceae) from Chile'. *Mitochondrial DNA Part B* 4(2):2581–82.

Vauchel, Peggy, Raymond Kaas, Abdellah Arhaliass, Régis Baron, and Jack Legrand. 2008. 'A New Process for Extracting Alginates from Laminaria Digitata: Reactive Extrusion'. *Food and Bioprocess Technology* 1(3):297–300.

Verhaeghe, Elodie Françoise, Aurélien Fraysse, Jean-Luc Guerquin-Kern, Ting-Di Wu, Guillaume Devès, Charles Mioskowski, Catherine Leblanc, Richard Ortega, Yves Ambroise, and Philippe Potin. 2008. 'Microchemical Imaging of Iodine Distribution in the Brown Alga Laminaria Digitata Suggests a New Mechanism for Its Accumulation'. *JBIC Journal of Biological Inorganic Chemistry* 13(2):257–69.

Wambugu, Peterson W., Marta Brozynska, Agnelo Furtado, Daniel L. Waters, and Robert J. Henry. 2015. 'Relationships of Wild and Domesticated Rices (Oryza AA Genome Species) Based upon Whole Chloroplast Genome Sequences'. *Scientific Reports* 5(1):13957.

Wang, Xiuliang, Zhanru Shao, Wandong Fu, Jianting Yao, Qiuping Hu, and Delin Duan. 2013. 'Chloroplast Genome of One Brown Seaweed, Saccharina Japonica (Laminariales, Phaeophyta): Its Structural Features and Phylogenetic Analyses with Other Photosynthetic Plastids'. *Marine Genomics* 10:1–9.

Wehr, John D., Robert G. Sheath, and J. Patrick Kociolek. 2015. *Freshwater Algae of North America: Ecology and Classification*. Elsevier.

Wells, Mark L., Philippe Potin, James S. Craigie, John A. Raven, Sabeeha S. Merchant, Katherine E. Helliwell, Alison G. Smith, Mary Ellen Camire, and Susan H. Brawley. 2017. 'Algae as Nutritional and Functional Food Sources: Revisiting Our Understanding'. *Journal of Applied Phycology* 29(2):949–82.

Wendel, Jonathan F., Curt L. Brubaker, and Tosak Seelanan. 2010. 'The Origin and Evolution of Gossypium'. Pp. 1–18 in *Physiology of Cotton*, edited by J. McD. Stewart, D. M. Oosterhuis, J. J. Heitholt, and J. R. Mauney. Dordrecht: Springer Netherlands.

Wernberg, Thomas, Gary A. Kendrick, and Benjamin D. Toohey. 2005. 'Modification of the Physical Environment by an Ecklonia Radiata (Laminariales) Canopy and Implications for Associated Foliose Algae'. *Aquatic Ecology* 39(4):419–30.

Wilde, Annegret, Katja Lünser, Friedrich Ossenbühl, Jörg Nickelsen, and Thomas Börner. 2001. 'Characterization of the Cyanobacterial Ycf37: Mutation Decreases the Photosystem I Content'. *Biochemical Journal* 357(1):211–16.

Wolstenholme, David R. 1992. ‘Animal Mitochondrial DNA: Structure and Evolution’. Pp. 173–216 in *International Review of Cytology*. Vol. 141, edited by D. R. Wolstenholme and K. W. Jeon. Academic Press.

Wong, K. H., and Peter C. K. Cheung. 2000. ‘Nutritional Evaluation of Some Subtropical Red and Green Seaweeds: Part I — Proximate Composition, Amino Acid Profiles and Some Physico-Chemical Properties’. *Food Chemistry* 71(4):475–82.

Wray, Gregory A. 2007. ‘The Evolutionary Significance of Cis -Regulatory Mutations’. *Nature Reviews Genetics* 8(3):206–16.

Ye, Naihao, Xiaowen Zhang, Miao Miao, Xiao Fan, Yi Zheng, Dong Xu, Jinfeng Wang, Lin Zhou, Dongsheng Wang, Yuan Gao, Yitao Wang, Wenyu Shi, Peifeng Ji, Demao Li, Zheng Guan, Changwei Shao, Zhimeng Zhuang, Zhengquan Gao, Ji Qi, and Fangqing Zhao. 2015. ‘Saccharina Genomes Provide Novel Insight into Kelp Biology’. *Nature Communications* 6(1):1–11.

Yoshihisa, Tohru. 2014. ‘Handling TRNA Introns, Archaeal Way and Eukaryotic Way’. *Frontiers in Genetics* 5.

Zhang, Lei, Xumin Wang, Tao Liu, Haiyang Wang, Guoliang Wang, Shan Chi, and Cui Liu. 2015. ‘Complete Plastid Genome of the Brown Alga Costaria Costata (Laminariales, Phaeophyceae)’. *PLOS ONE* 10(10):e0140144.

Zheng, Zhou, Hao Chen, Huaiping Wang, Weiming Jiang, Qian Cao, and Ning Du. 2019. ‘Characterization of the Complete Mitochondrial Genome of Bull Kelp, Nereocystis Luetkeana’. *Mitochondrial DNA Part B* 4(1):630–31.

Zhu, Andan, Wenhu Guo, Sakshi Gupta, Weishu Fan, and Jeffrey P. Mower. 2016. ‘Evolutionary Dynamics of the Plastid Inverted Repeat: The Effects of Expansion, Contraction, and Loss on Substitution Rates’. *New Phytologist* 209(4):1747–56.

Žuljević, Ante, Akira F. Peters, Vedran Nikolić, Boris Antolić, Marija Despalatović, Ivan Cvitković, Igor Isajlović, Hrvoje Mihanović, Slavica Matijević, Dawn M. Shewring, Simonepietro Canese, Christos Katsaros, and Frithjof C. Küpper. 2016. ‘The Mediterranean Deep-Water Kelp Laminaria Rodriguezii Is an Endangered Species in the Adriatic Sea’. *Marine Biology* 163(4):69.

6 Abbreviation

bp	Base-pairs
°C	Celsius
Ca	Circa
cm	Centimeter
cpDNA	Chloroplast deoxyribonucleic acid
CTAB	Cetyl trimethyl ammonium bromide
DNA	Deoxyribonucleic acid
dNTP	Deoxy ribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
g	Gram
g	Relative centrifugation force
HTG	Horizontal gene transfer
Indel	Insertion-deletion
IR	Inverted repeat
kb	Kilo base
kbp	Kilo basepair
LD	Light Dark
MAFFT	Multiple alignment using fast fourier transform
mg	Milligram
ml	Millilitre
mM	Millimolar
M	Molar (mol/l)
mRNA	Messenger ribonucleic acid
MUMmer	Maximal Unique Matches
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
numer	Nucleotide Maximal Unique Matches
ORFs	Open reading frames
PES	Provasoli Enriched Seawater medium
PCR	Polymerase chain reaction
Poly a	Polyadenylic acid
rRNA	Ribosomal ribonucleic acid
RNase	Ribonuclease

SBS	Sequencing by synthesis
SNPs	Single nucleotide polymerase
TE	Tris-EDTA
tRNA	Transfer ribonucleic acid
v/v	Volume per volume
w/v	Weight per volume
μg	Micro gram
μm	Micro molar

7 List of figures

8 List of tables

9 Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Rana, S., Valentine, K., Blanfuné, A., Reynes, L., Thibaut, T., Bartsch, I., Glöckner, G. 2020. Kelp evolution inferred from complete mitochondrial and plastid genomes- detection of an independent introduction of foreign sequences into the mitochondrial genome in brown algae. *Journal of Phycology*. Submitted.

Rana, S., Valentine, K., Bartsch, I., Glöckner, G. 2019. Loss of chloroplast encoded function could influence species range in kelp. *Ecology and Evolution*.2019;9:8759-8770.

Köln, den 10.08.2020

(Shivani Rana)

10 Curriculum Vitae

Personal information

Name:	Shivani Rana
Address:	Glück-Auf-Str. 5, 50169, Kerpen
Date of birth:	19.02.1989
Place of birth:	Meerut, India
Nationality:	Indian

Education-Work

2017-2020	PhD. in Bioinformatics at the Institute of Biochemistry I, Medical Faculty, University of Cologne, Germany
2016-2017	Research assistant, at the Institute of Biochemistry I, Medical Faculty, University of Cologne, Germany
2014-2015	Research assistant, at the University of Delhi, South Campus, New Delhi, India
2011-2013	Master of Technology in Biotechnology Uttar Pradesh Technical University, Lucknow, India
2009-2011	Master of Science in Biotechnology C.C.S. University, Meerut, India
2006-2009	Bachelor of Science in Biotechnology C.C.S. University, Meerut, India
2004-2006	Secondary Education Meerut, India Focus Subjects: Biology, Chemistry, Physics

11 Lebenslauf

Persönliche Daten

Name:	Shivani Rana
Momentane Anschrift:	Glück-Auf-Str. 5, 50169, Kerpen
Geburtstag:	19.02.1989
Geburtsort:	Meerut, Indien
Staatsangehörigkeit:	Indisch

Ausbildung

2017-2020	Promotion in Bioinformatik (biologische Richtung) am Institut für Biochemie I, Medizinische Fakultät, Universität zu Köln
2016-2017	Wissenschaftliche Mitarbeiter, am Institut für Biochemie I, Medizinische Fakultät, Universität zu Köln
2014-2015	Wissenschaftlicher Mitarbeiter, am Universität zu Delhi, Indien
2011-2013	Master of Technology, in Biotechnologie, am Uttar Pradesh Technical Universität, Lucknow, Indien
2009-2011	Master of Science, in Biotechnologie, am C.C.S. Universität, Meerut, Indien
2006-2009	Bachelor of Science, in Biotechnologie, am C.C.S. Universität, Meerut, Indien
2004-2006	Hochschule Meerut, Indien Schwerpunkte: Biologie, Chemie, Physik