CHAPTER 1: INTRODUCTION

Seaweeds are a subset of a larger group known as the algae that live in marine or brackish water environment. They do not have true roots, stem or leaves and the body of the plant consists of holdfast, stipe and blade. Thus, algae are defined as non-flowering photosynthetic organisms ranging from the microscopic phytoplankton to the macroscopic marine algae or seaweeds (Phang, 2006).

Seaweeds are important components of aquatic ecosystems and are valuable sources of food, ingredients in cosmetics, fertilizers and hydrocolloids (e.g. agar and alginate). Some seaweeds have the capacity to remove heavy metals from water which is known as bioremediation and can potentially be used in biomonitoring of such pollutants (Chan et al., 2006). The taxonomy and systematics, in addition to physiological aspects related to mass cultivation and the production of useful products especially of commercially important species have been the focus of research (Chan et al., 2006). However, seaweeds are well-known for their high morphological plasticity. The classification of seaweeds based on morphological criteria can be problematic due to the lack of distinctive reproductive structures and changing in morphological structures in response to environmental conditions (de Senerpont Domis et al., 2003). Despite being chosen from the same clone and cultivated under identical conditions, these species may still vary in terms of yields and morphological features. Therefore, molecular tools are useful to confirm and assess the limits of morpho-species and to delineate boundaries between species (Manhart & McCourt, 1992; John & Maggs, 1997) especially for commercial exploitation, such as with species of *Kappaphycus*, *Gracilaria*, *Sargassum*, and *Caulerpa*.

Over the years, a number of genetic marker systems have been developed for genetic diversity studies, species and strain identification, genome mapping and marker-

trait association studies. These include random amplified polymorphic DNA (RAPDs), restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs), microsatellites or simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs) (Liu & Cordes, 2004). They are among the popular markers compared to allozyme and mtDNA markers that have very low polymorphism and required prior molecular information for the target species. RFLPs and RAPDs are the first few molecular markers that are widely used. However, the RFLP technique is laborious as they required large amount of good quality DNA for restriction digestion and Southern blotting, which is expensive, hazardous and time-consuming. RAPD marker has poor reproducibility, dominant inheritance and difficulty in scoring bands, that leads to inappropriate inferences (Joshi et al., 1999; Sharma et al., 1998). Thus, better alternatives have been developed namely AFLPs, SSRs, and SNPs. These three markers are the most popular markers nowadays.

Microsatellite markers or simple sequence repeats (SSRs), are tandem repeats of 1-6 nucleotides that are distributed throughout the genomes of most eukaryotic species (Zane et al., 2002). SSRs markers are useful for a variety of applications in plant genetics and breeding because of the uniqueness and value of microsatellites arises from their multiallelic nature, co-dominant transmission, relative abundances, ease of detection by PCR, good genome coverage and require only a small amount of starting DNA (Qureshi et al., 2004). Besides, microsatellites have been revealed to be more polymorphic in comparison to other markers, shown by the amount of diversity detected during evaluation of microsatellites. The high discriminatory power of microsatellites is also significant for analyzing variation in the gene pools of crops (Powell et al., 1996).

The conventional approaches in obtaining microsatellite markers involved PCR amplification and sequencing of the genome. However, with the large number establishment of expressed sequence tags (ESTs) and genome sequences in public

database, development of SSRs markers is possible from potentially useful resources (Rudd, 2003). The chloroplast microsatellite markers (cpSSRs) were the first developed as genetic markers by Powell et al. in 1995 (Powell et al., 1995a,b) and nowadays a rapid increase in chloroplast genome sequences enhanced the development and identification of cpSSRs in various plants, such as in *Chiloglottis* orchids (Ebert & Peakall, 2008) and torrey pine (*Pinus torreyana*) (Provan et al., 1999). In 1993, the identification of SSRs in ESTs of plant species was carried out by Morgante and Olivieri but the volume of available sequence data at that time was very limited (Morgante & Olivieri, 1993). To date, abundance of EST sequences for different organisms are available in GenBank and these EST-SSRs are transferable among distantly related species than SSRs obtained from genomic sequences (Varshney et al., 2005). This means the identification of microsatellite markers can be achieved by simple database searches and the developed SSRs can be used across different species without the complete sequence of the genome.

Perhaps, cultivar identification was the first application of microsatellites in plants, where microsatellites have been used to genotype diverse materials such as grapevine (*Vitis vinifera*) (Thomas, 1993) and soybean (Rongwen, 1995). Later on, microsatellites were used to determine the hybridity of intraspecific somatic hybrids of potato (*Solanum tuberosum*) (Provan et al., 1996) and also for analyzing genetic variation in rice (*Oryza sativa*). Due to the high discriminatory power of SSRs, study has found that 28% of the allelic variability was lost during the process of cultivar development of rice (Yang et al., 1994). This showed that microsatellites play an important role in securing plant variety rights. Hence, this is the first attempt to use the bioinformatics tools that are available, to identify microsatellite markers from GenBank and to evaluate their applicability on *Gracilaria* species from Malaysia and other countries.

1.1 Importance of Seaweed Taxonomy Studies

Taxonomy is the branch of biological systematic which includes naming, identifying, and classifying of organisms (Winston, 1999). There are at least four major approaches to reconstruct the evolutionary history/phylogeny of a group of species: classical evolutionary taxonomy, phonetics (numerical taxonomy), cladistics and molecular taxonomy. DNA genomes have thus become valuable reservoirs of taxonomy characters potentially important for phylogenetic analysis (Coleman & Goff, 1991). However, molecular phylogenies are not fully congruent with morphological taxonomy (Fama et al., 2002) and cryptic or sibling species have sometimes been detected for several ‘‘species’’ complexes that were previously identified exclusively by morphology (Lewis & Flechtner, 2004; Zuccarello & West, 2003). Therefore, the combination of both molecular and morphological techniques is a promising approach for delineating species boundaries (de Senerpont Domis et al., 2003; Kawai, 2004; Nam et al., 2000; Yoshida et al., 2000).

The systematic and phylogenetic studies of seaweeds were initiated in the early 1990s, when the extensive research and development in modern biotechnology starting to increase steadily. The molecular biotechnology of seaweeds which included the biotechnology on identification, modification, production and utilization of marine algal molecules are important, not only in the strain improvement and development of algae as bioreactors, but also in the determination of taxonomic status among algal species of interest (Qin et al., 2004). However, the morphological species concept remains as a basis for species-level and intraspecific studies (Wattier & Maggs, 2001) especially in *Gracilaria*. For example, the taxonomic and systematic problems within the members of Gracilariaceae have not been resolved due to the lack of distinct morphological diagnostic characters and unreliable developmental characters of the female reproductive system (Wattier et al., 1997; Iyer et al., 2005). Small subunit rDNA and

RuBisCO spacer sequences were employed to determine the phylogenetic relationships on Gracilariaceae from South Africa. Results showed that the markers were able to resolve the southern African gracilariod complex into three species: *Gracilaria gracilis*, *Gracilariopsis longissima* and *Gracilariopsis funicularis* (Iyer et al., 2005). This is important as Gracilariaceae (Rhodophyta) is one of the families that possess economic potential for agar production and as feed for abalone.

Another example of species with uncertain taxonomy is *Gracilaria salicornia* (C. Agardh) due to the limitation of distinct morphological characteristics (Yamamoto, 1978). Several features were chosen by Xia (1986) in order to distinguish *G. salicornia* (C. Ag.) Dawson, *G. cacalia* (J. Ag.) Dawson, *G. minor* (Sond.) Durairatnam and *G. canaliculata* (Kützing) Sonder (= *G. crassa* Harvey ex J. Agardh) but the features found were not consistent. Thus, the following related species: *Corallopsis opuntia* J. Agardh; *C. dichotoma* Ruprecht; *C. minor* (Sonder) J. Agardh; *C. cacalia* J. Agardh; *Gracilaria cacalia* (J. Agardh) Dawson; *C. salicornia* (C. Agardh) Greville; *G. salicornia* var. minor Sonder; *G. minor* (Sonder) Chang et Xia; *G. crassa* Harvey ex J. Agardh; *G. canaliculata* (Kützing) Sonder; and *Sphaerococcus canaliculatus* (Kützing) were classified as synonyms of *G. salicornia* (Abbott, 1988). Furthermore, the mode of branching is the only distinct feature used to distinguish *G. salicornia* and its closely related species as many of the related species have similar reproductive structures. Correct identification is important as different species have different biochemical composition and uses. Besides, genetic analysis can reveal the extent of population structuring and give an indication of subgroup distinctiveness (Beebee & Rowe, 2008). Therefore, taxonomy studies are important to ensure that seaweeds and other organisms are correctly identified and recognized (Bird & van der Meer, 1993).

1.2 Objectives of Research

Research Questions:

1. Using the bioinformatics tools, is it possible to mine microsatellite markers from the chloroplast genome and EST database of seaweeds downloaded from GenBank?

2. Are the microsatellite markers mined from chloroplast genome and EST database of seaweeds suitable for use to differentiate the species of *Gracilaria* according to populations, strains and taxa?

3. What is the difference between SSRs that are mined from chloroplast genome and EST database of *Gracilaria tenuistipitata* in terms of polymorphism, number and pattern of the SSRs?

4. What is the difference between SSR markers mined from GenBank and the commonly used marker, *cox*1 gene in terms of variability on the species tested?

The objectives of this research study are:

i) To mine microsatellite markers or simple sequence repeats (SSRs) from chloroplast genome and EST database of *Gracilaria* species using available bioinformatics tools.

ii) To assess the suitability of the genomic SSR markers for differentiating between different populations and species of *Gracilaria*.

iii) To assess the suitability of the EST-SSR markers for differentiating between different populations and species of *Gracilaria.*

iv) To compare the polymorphism, number and pattern of SSRs that are mined from chloroplast genome and EST database of *Gracilaria tenuistipitata.*

v) To compare the variability of the SSR markers mined from GenBank and the commonly used marker, *cox*1 gene on the species tested.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction to Seaweeds

The study of algae is called phycology. The word phycology is derived from the Greek word *phykos*, which means "seaweed" (Lee, 1999). The algae do not arise from a single common ancestor like plants and animals. Instead, they are a group composed of many lineages (they are "polyphyletic", Graham & Wilcox, 2000). Algae are important photosynthetic producers where they grow not only in both freshwater and marine environments but also found in damp terrestrial environments (Sze, 1998).

Algae can be either autotrophic or heterotrophic but they are generally considered to be autotrophs. If they are autotrophic, they use inorganic nutrients, such as carbon dioxide, water, phosphates, and inorganic nitrogen as a source of carbon (Sze, 1998). Autotrophs can be photoautotrophic, using light as a source of energy, or chemoautotrophic, oxidizing inorganic compounds for energy. If they are heterotrophic, the algae use organic compounds for growth. Heterotrophs can be further divided to photoheterotrophs which use light as a source of energy, or chemoheterotrophs where they oxidizing organic compounds for energy (Lee, 1999).

Algae have an extremely confusing array of cell cycles, cell morphologies and live in a multitude of habitats (Bhattacharya & Medlin, 1998). Their cell arrangements to form multicellular bodies or thalli and their pigments for photosynthesis also vary greatly. Algae lack plants' complex reproductive structure but both algae and land plants have photosynthesis system based on chlorophyll *a* (Sze, 1998). The algae include many bacteria-sized (1-5 μm) coccoid taxa (e.g. *Micromonas* spp. and *Chlorella* spp. in the Chlorophyta and *Pelagomonas* spp. in the Heterokonta) and also the world's largest protists, the kelps such as *Macrocystis* spp. in the Geterokonta, which is up to 30 m in length (Bhattacharya & Medlin, 1998).

Seaweeds, or marine macroalgae are usually found in the sub-tidal region up to a depth where 0.01% photosynthetic light is available. Many seaweeds are plant-like in their appearances; having stem, root and leaf analogs in the form of anchoring holdfasts, stipes, and blades. Nevertheless, they are only remotely related to land plants. The similarities in macroscopic structure between seaweeds and land plants are due to parallel evolution (Graham & Wilcox, 2000). The distribution and variety among seaweeds are determined by the plant pigments, light, exposure, temperature, depth, tides and the shore characteristic that combined to create different environment for seaweeds. Based on the recent biochemical, physiological and electron microscopic studies, the criteria used to distinguish the different algal group are based on the: photosynthetic pigments; storage food products; cell wall component and fine structure of the cell and flagella (Dhargalkar & Kavlekar, 2004). Thus, seaweeds are classified into brown (Phaeophyta), red (Rhodophyta), and green (Chlorophyta) seaweeds (Phang, 2006).

In 2010, only 31 countries and territories were recorded with algae farming production and 99.6 percent of global cultivated algae production came from eight countries: China (58.4 percent, 11.1 million tons), Indonesia (20.6 percent, 3.9 million tons), the Philippines (9.5 percent, 1.8 million tons), the Republic of Korea (4.7 percent, 901 700 tons), Democratic People's Republic of Korea (2.3 percent, 444 300 tons), Japan (2.3 percent, 432 800 tons), Malaysia (1.1 percent, 207 900 tons) and the United Republic of Tanzania (0.7 percent, 132 000 tons) (FAO, 2012). In Malaysia, the tally of marine algae now stands at 373 specific and intraspecific taxa (17 taxa of Cyanophyta, 102 Chlorophyta, 182 Rhodophyta and 72 Phaeophyta) (Phang, 2006).

Brown seaweeds (Phaeophyta) are exclusively marine and the coloration is due to the presence of fucoxanthin and beta-carotene pigments. They can be ranged from microscopic filaments like *Ectocarpus siliculosus* to the largest size *Macrocystis*

pyrifera (kelp) (Dhargalkar & Kavlekar, 2004). In Malaysia, the most frequently found species are *Sargassum*, *Dictyota* and *Padina*. They inhabit sandy areas, mudflats, coral reefs, rocky shores and mangrove areas (Phang, 2006). Phang and Yoshida described a new species *Sargassum stolonifolium* from Penang Island, west coast of Peninsular Malaysia that exhibit the phenomena of new plantlets derived from the first leaves (Phang & Yoshida, 1997). Brown seaweeds are important as food, hydrocolloids production, fertilizer and soil amendment, animal fodders and biomonitors. The large brown algae such as *Macrocystis* and *Ascophyllum* are the main species used for manure while *Ecklonia* and *Fucus* are also sold as soil additives and functioned as both fertilizer and soil amendment. The environment is contaminated with heavy metals released by many industries through their wastewaters (Brauckmann, 1990). Milled, dried species of the brown seaweeds such as *Macrocystis*, *Ecklonia* and *Laminaria* were able to adsorb copper, zinc and cadmium ions from the wastewater. Extractions of alginate from brown seaweeds are found to adsorb heavy metals just as effectively as the seaweed from which it is derived. In natural or effluent waters, seaweeds can be used as biomonitors in the detection of algal nutrients or substances that are toxic to algae (Graham & Wilcox, 2000).

Green seaweeds (Chlorophyta) are photosynthetic eukaryotes bearing double membrane-bound plastids containing chlorophyll *a* and *b* and a unique stellate structure linking nine pairs of microtubules in the flagellar base (Bremer et al., 1987; Kenrick & Crane; 1997). Starch is usually stored inside the plastid when it is present and cell wall is composed of cellulose (Graham & Wilcox, 2000). Most of the green algae are microscopic and rarely more than a meter in greatest dimension. Their body (thallus) size and habit range from microscopic swimming or non-motile forms, such as nanoplankton, benthos, or lichen phycobionts, to macroscopic. The thallus structure of green algae varied from swimming and non-motile unicells, to filaments, colonies,

various levels of tissue organization, and branching morphologies. Cells may be uninucleate or coenocytic. An example of algae which has many nuclei dispersed throughout the cytoplasm of giant cell is *Caulerpa* (Lewis & McCourt, 2004). In Malaysian waters, Chlorophyta consists of the second highest number of taxa. There are thirteen species and seven varieties of *Caulerpa* have been recorded and most of them are in coral reefs. Eight of these, i.e., *C. lentillifera*, *C. serrulata*, *C. taxifolia*, *C. scalpelliformis*, *C. peltata*, *C. sertulariodes*, *C. racemosa* and *C. verticillata* are commonly found in coral reefs, indicated by recent collections data (Phang, 2006).

Taxonomy, physiology and biochemistry have been the main research focus since 1940s, when the potential of agar production from seaweeds was recognized. Physiological aspects such as mass cultivation of seaweeds and the production of hydrocolloids and pigments have been of particular interest. Our basic understanding of seaweed biology depends on these areas of research. Seaweeds had received little attention worldwide and were not included in efforts to elucidate gene functions, despite the progress and development of functional genomics in terrestrial plants (Chan et al., 2006). When the development of genetic transformation techniques on seaweeds (Cheney & Kurtzman, 1992) and the characterization of genes involved in carbohydrate synthesis (Zhou & Ragan, 1995) were established, studies on the molecular genetics of seaweeds were initiated in 1990s. After the first use of the expressed sequence tag (EST), which is a relatively inexpensive and quick approach to study seaweed genomics (Lluisma & Ragan, 1997), genetic research in seaweeds has entered a new phase. The move to have a genetic model for each of these seaweed groups appears desirable as there are differences in physiology and developmental biology among the red, brown and green seaweeds (Chan et al., 2006).

2.2 Uses of Seaweeds

To date, more than 14.7 million tons of seaweeds are commercially produced worldwide (Buchholz et al., 2012) and seaweeds have been used by humans for 14,000 years (Dillehay et al., 2008). According to the Wamyosho (the oldest Chinese-Japanese dictionary in Japan), in an earlier stage of the Heiam era (794-1185), twenty-one species of marine algae including green, brown and red algae were already used as food among people (Nisizawa et al., 1987). Japanese people are the main consumers with an average of 1.6 kg (dry weight) per year per capita (Fujiwara et al., 1984).

Porphyra (Nori) cultured entirely for the production of edible seaweed consumed in Japan and is marketed in thin, uniformly sized, sheets of dried and pressed weed which has many uses in a variety of traditional dishes, including sushi. Walker et al. showed that 30% of *Porphyra* spp. provides a suitable fish meal replacement in the diet of juvenile Atlantic cod (Walker et al., 2009). The popular food named "Kombu" (*Saccharina japonica*) with 4.9 million tons produced annually, is widely used as soup stock, vegetable and rice dishes after soaking in soy sauce. The other popular seaweed product "Wakame" (*Undaria pinnatifida*), produced 1.7 million tons per year and is marketed mainly in a desalted form and eaten chiefly with soybean soup (FAO 2011a; Buchholz et al., 2012).

In Malaysia, species of *Gracilaria* and *Gelidium* are consumed as salads and species of the green seaweed, *Caulerpa*, are eaten raw or cooked, as salads or desserts. A number of brown seaweeds, *Sargassum* and *Turbinaria*, are consumed fresh, cooked in coconut milk or smoked dried (Chapman, 1970). A soup made up of *Sargassum* and *Turbinaria* can be a source of iodine that "cools" the body system while crushed *Corallina* (Amphiroa) can be consumed by children to eject worms (Phang, 2006). Usage of seaweeds as antibiotics can be obtained from *Halimeda opuntia*,

Acanthophora spicifera, *Eucheuma spinosum*, *Gracilaria* spp., *Hypnea musciformis*, *Dictyopteris* spp., and *Sargassum* spp. (Phang, 2006).

The great variety of chemical compositions and therefore bioactive properties in different species of seaweeds make them a good source for medical and pharmaceutical uses. Some sulphated polysaccharides from red algae exhibit antiviral activities towards viruses responsible for human diseases, for example *Aghardhiella tenera* and *Nothogenia fastigiata* (Smit, 2004). A galactan sulphate from *Aghardhiella tenera* was tested by Witvrouw et al. (1994) while Damonte et al. (1994) and Kolender et al. (1995) discovered a xylomannan sulphate from *Nothogenia fastigiata* which will act against human immunodeficiency virus (HIV), Herpes simplex virus (HSV) type 1 and 2 and also respiratory syncytial virus (RSV). Characteristic of most of the algal polysaccharides, particularly those of *Aghardhiella tenera* and *Nothogenia fastigiata*, meet the requirement of antiviral polysaccharides that is having very low cytotoxic activities towards mammalian cells (De Clercq, 1996).

Dried seaweeds that have been milled to a fine powder can be used as animal fodder. *Ascophyllum* is the main raw material for seaweed meal as it is so accessible. It contains useful amounts of minerals such as potassium, phosphorus, magnesium, calcium, sodium, chlorine and sulphur, trace elements and also vitamins. *Ascophyllum nodosum* are harvested annually and processed to seaweed meal which is used as a supplementary animal fodder (Moen et al., 1997). *Hypnea* and *Acanthophora* are used as feed for abalone (Phang, 2006). Besides, the addition of 4% *Sargassum* to the feed of shrimp cultures can reduce the cholesterol by 29% which will be beneficial for human consumption (Casas-Valdez et al., 2006).

The hydrocolloids: agar, carrageenan, and alginates were still the focus of the seaweed industry and these substances have thickening, emulsifying and gelifying

properties which can be used in diverse fields (Buchholz et al., 2012). Alginates have been extracted from brown algae since the 19-century (Phaeophyceae) particularly from the species of *Laminaria, Aschophyllum, Macrocystis, Sargassum, Durvillaea, Ecklonia*, and *Lessonia* (Ohno & Largo, 1998; McHugh, 2003) as they containing D-mannuronic and L-guluronic acids that are useful in providing viscosity to aqueous solutions (Renn, 1997). The latest innovation is the inclusion of seaweeds *Ascophyllum nodosum* in the manufacture of textile fiber that nurtures the skin and protects against free radicals (Smartfiber, 2010). The hydrocolloids and other seaweed components from *Ulva lactuca, Ascophyllum nodosum, Laminaria longicruris, Saccharina latissima, Laminaria digitata, Alaria esculenta,* various *Porphyra* species, *Chondrus crispus*, and *Mastocarpus stellatus*, together made up the ingredients of cosmetics (Cosmetic Ingredient Dictionary 2002–2011; Buchholz et al., 2012).

Carrageenan is a complex mixture of several polysaccharides where the amount and composition of polysaccharides vary with the types of seaweed used (Towle & Whistler, 1973). Like agar, carrageenan is also used to solidify and emulsify solutions. The raw materials used in the extraction of carrageenan are *Chondrus crispus* and to a lesser extent, *Gigartina stellata*. Approximately half of all carrageenan is derived from *Eucheuma*. The major producing countries for *Eucheuma* are Philippines and Indonesia with smaller amounts from Tanzania (Bixler & Porse, 2010). Carrageenans are composed of alternating units of modified galactose. There are three types of carrageenan, which are kappa, lambda, and iota carrageenan but the two main groups of carrageenan are assigned as k- and i-types (De Ruiter & Rudolph, 1997). Carrageenan is used by food industry especially in the baking, confectionery and dairy product markets. Gelling properties of carrageenan for structural stability are important in fillings for tarts while carrageenan added with cooked ham and poultry products can enhance salting characteristics and improve sliceability. Carrageenan is also used to prevent chocolate

milk from sedimenting and to allow solidification of custards, ice-creams and milk puddings. Toothpastes, canned and frozen pet foods are incorporated with carrageenan during the production process to assist in gel formation and solidification (Therkelsen, 1993).

2.3 Rhodophyta (Red Algae)

Red algae (Rhodophyta) are the most diverse in the tropics compared to the green and brown algae. They are eukaryotes, and majority of the species are marine, photosynthetic, and macroscopic (Maggs et al., 2007). The marine algae are generally found in intertidal and sub-littoral regions. The red algae *Bangiomorpha pubescens* Butterfield, is believed to be one of the oldest taxonomically resolved eukaryotic fossils (Butterfield, 2000).

There are two layers in the cell wall of red algae: (i) the inner wall layer consisting of microfibrils of cellulose and (ii) the outer wall layer of pectic materials. In Bangiophyceae, species such as *Porphyra* do not have cellulose in the cell wall; instead they have microfibrils made up of xylan (Awasthi, 2009). Floridean starch is a reserve polysaccharide in the red algae and is always located in the cytoplasm of the cell whereas green algae and plants store starch within the chloroplasts. The cells of red algae exhibit typical eukaryotic organization where they are characterized by the absence of flagella, having unstacked thylakoid membranes and lacking an encircling endoplasmic reticulum membrane in the red algal plastids (Maggs et al., 2007). Red algae possess only chlorophyll *a* with accessory red/blue phycobilin pigments, predominantly the red-colored phycoerythrin, occur in stalked phycobilisomes on thylakoids (van den Hoek et al., 1995).

The red algae are divided into two groups: Bangiophyceae and Florideophyceae based on morphological, anatomical, and life-history differences (Dixon, 1973). The unicellular form of Bangiophyceae has simpler morphology structure and sexual reproduction system. Their life histories are poorly known but appear to be diverse (Brodie & Irvine, 2003). *Porphyra* has a sheet-like thallus but most of the part only one cell thick and displays a heteromorphic life history, which means they are morphologically different in both phases of the life cycles (Drew, 1954; Bell & Hemsley, 2002). Besides Bangiales, most of the Bangiophyceae are asexual (Brodie & Irvine, 2003). The Florideophyceae are more complex and filamentous in construction. For example, *Chondrus crispus* Stackhouse has much diverse morphological structures and complicated haploid-diploid life histories. The product of fertilization is not the diploid sporophyte, but a hemi-parasitic diploid tissue (the "gonimoblast") surrounded by female nutritive tissue, called the "cystocarp" (or carposporophyte). In this stage, the cystocarp will release numerous genetically identical diploid spores that give rise to tetrasporophytes. When these tetrasporophytes matured, they will produce haploid tetraspores via meiosis and grow into male and female haploid gametophytes (Hawkes, 1990; Maggs et al., 2007).

In Malaysia, red seaweeds comprise the highest number of taxa with 182 taxa out of the 373 specific and intraspecific taxa. *Eucheuma* and *Kappaphycus* are found from lower intertidal to upper sub-tidal areas in Sabah and around islands in Peninsular Malaysia. The large foliose species of *Halymenia* dominate the subtidal bedrock areas, while the proliferous branching thalli of *Laurencia* and *Hypnea* species inhabit the bedrocks at the intertidal regions. Erect coralline (*Amphiroa*, *Jania*) as well as crustose coralline (*Lithothamnion, Peyssonnelia*) are Rhodophytes commonly found in the coral reefs especially in the cleaner deep waters around the islands. Thirty-eight new records,

including one new species, were reported from the Malaysian-Japanese collaboration from 1995 (Phang, 2006).

2.4 Gracilariaceae

The family Gracilariaceae Nägeli belongs to Gracilariales, the order of marine red algae described by Fredericq and Hommersand (1989). The genus *Gracilaria* was established by Greville since 1830 and he assigned 14 species to this genus. In 1852, Agardh redefined it based on the differences in external form and internal structure. Although the structure of the reproductive organs has been considered in the classification of species since 1926, it was not prominent until 1949, when Dawson separated the genus *Gracilariopsis* from *Gracilaria* based on the absence of nutritive filaments and the small size of the gonimoblast cells. However in 1967, Papenfuss found that the nutritive filament was not always present in British material of *Gracilaria verrucosa* and thus the two genera were merged. In 1989, the study on spermatangial development and post-fertilization events in *Gracilariopsis lemaneiformis* (Bory) Dawson by Fredericq and Hommersand restored the genus *Gracilariopsis* as they distinguished from the genus *Gracilaria* by the lack of tubular nutritive cells, the superficial position of gametangial mother cells and a few other characters.

In 1963, Chang and Xia established the genus *Polycavernosa* Chang and Xia when they found gracilarioid algae with multiple spermatangial cavities and darkly staining, irregularly shaped conspicuous basal nutritive filaments, emerging from the bottom of the gonimoblasts. In 1987, Xia and Abbott found that the spermatangial conceptacles are clustered rather than randomly distributed over the surface of the thallus and showed that *Hydropuntia urvillei* Montagne (1842) is actually similar to Polycavernosa. Wynne (1989) transferred 14 species of *Polycavernosa* and two species

of *Gracilaria* to *Hydropuntia*. However, in 1991, Abbott, Zhang and Xia re-examined the western Pacific specimens of *Polycavernosa* as they found the mixture of *Verrucosa*-type and *Polycavernosa*-type spermatangia on the same branchlets and thalli when describing the new species, *Gracilaria mixta*. There are some species of *Gracilaria* maintaining these *Verrucosa*-type spermatangia throughout the plant. Hence, 16 species of *Hydropuntia* were transferred to *Gracilaria*.

Yamamoto (1975) divided the genus into three subgenera based on three different types of male organs from the 15 Japanese *Gracilaria*: *Gracilariella* – with superficial spermatangia, scattered continuously over the frond surface; *Textoriella* – with spermatangia conceptacles shallowly depressed; and *Gracilaria* – with deep pot-shaped conceptacles and spermatangia covering the entire inner surface of conceptacles. In 1999, Tseng and Xia examined hundreds of male plants of *Gracilaria* and agree with Yamamoto in his subgenera disposition. They have placed the *Hydropuntia* as the fourth subgenus of *Gracilaria* and *Gracilariopsis* in the first subgenus *Gracilariella*.

In Malaysia, twenty species of *Gracilaria* have been reported, mainly inhabiting mangroves, sandy mudflats and rocky shores (Lim & Phang, 2004). The common species are, *Gracilaria changii* (Xia and Abbott), Abbott, Zhang and Xia; *Gracilaria edulis* (Gmelin) Silva; *Gracilaria salicornia* (C. Agardh) Dawson and *Gracilaria tenuistipitata* Chang and Xia (Phang, 2006). Nowadays, the genus *Gracilaria* has replaced *Gelidium* as the most important source of agar in the world.

2.4.1 Economic Importance of *Gracilaria*

Gracilaria species are commercially important in the hydrocolloids production (Armisen, 1995). In 2009, 86,100 tons of hydrocolloids were traded in comprising of 58% of carrageen, ~31% alginates, and ~11% agar (Bixler & Porse, 2010). Although direct

consumption is most common in the Asia-Pacific region, algal hydrocolloids are used worldwide in a great variety of food items as emulsifying, gelling, or water retention agents (Indergaard & Østgaard, 1991; Murata & Nakazoe, 2001; Bartsch et al., 2008). In Malaysia, several species of *Gracilaria* (e.g. *Gracilaria changii* and *Gracilaria tenuistipitata*) and *Gelidium* are used as salads and for agar extraction. In Sabah, *Eucheuma* is consumed as food and for the extraction of carrageenan. Besides that, *Gracilaria* species are recognized for their antibiotic properties (Phang, 2006).

The word "agar" itself is of Malayan origin. In Japan, they are known as "kanten" where the agar is marketed in the form of sheets, bars, sticks and other shapes. Majority of the processed agarophytes is collected from the natural beds. The high demand for agar has caused the shortage of *Gelidium* species and nowadays *Gracilaria* species is replacing *Gelidium* for agar production. Agar is used as a stabilizer for emulsion, and as a constituent of skin creams, ointments and lotions in the cosmetic industry (Levring et al., 1969). It is also indispensable as a solid culture medium in medical bacteriology and microbiological research (Buchholz et al., 2012). In the pharmaceutical industry, agar has been used for many years as a smooth laxative. Other applications of agar are found in the photographic industry, in paper manufacture, and as a finishing and sizing agent in the textile industry.

Chapman (1970) discovered the uses of agar particularly in the production of linoleum and artificial leather, as a sound and heat insulator, as an ingredient for paints and in the manufacture of storage batteries. Agars are also applied in the production of intricate casts which is used in dentistry, criminology and dye making. Agars are commonly used as ingredients for food in South-East Asia and elsewhere as they are non-digestible (Fleurence, 1999).

2.4.2 *Gracilaria changii* (Xia et Abbott) Abbott, Zhang et Xia

Gracilaria changii was first described by Abbott, Zhang and Xia (1991) resulted from the collections made during the Siboga Expedition in the Indonesian region, 1899-1900. The type locality of this species is in Malaysia situated about 200 m northeast of beacon opposite Glugor Marine Police Jetty, Pulau Pinang, Malaysia (Xia & Abbott 1987). *G. changii* mainly inhabits mangrove areas that fringe the west coast of Peninsular Malaysia. They have irregular, alternate, or secund in two to four orders branching with an abrupt constriction at the base, forming a slender stipe. The distal end of stipe is slightly swollen and tapering towards the tip (Lim & Phang, 2004). Experiment has been done on the agar yield and gel strength of *G. changii* and results showed that the agarose yield ranged between 13.38% and 16.33% whereas the gel strength ranged between 737 and 950 g cm^{-2} (Santos & Doty, 1983). Later on, Phang et al. monitored the variation in agar content and gel strength of a wild population of *G. changii* from Morib, west coast Peninsular Malaysia over a 15-month period and found that the agar yield and gel strength were ranged from 12% to 25% dry weight and 294 to 563 g cm^{-2} respectively (Phang et al., 1996). This showed that *G. changii* has potential to produce good quality agar that will be beneficial in various industries.

In Malaysia, the molecular study on *G. changii* started since 2003, when Gan et al. developed the genetic transformation and gene expression system for *G. changii*. In 2007, Sim et al. identified the sex-specific markers for *G. changii* using random amplified polymorphic DNA (RAPD). There is no further report until 2011, Yow et al. studied the genetic diversity of this species using the mitochondrial *cox*1 gene and recently, they have published another paper on *Gracilaria* species using *cox*2-3 spacer as a comparative marker to *cox*1 gene (Yow et al., 2012).

2.4.3 *Gracilaria tenuistipitata* var. *liui* Chang et Xia

Gracilaria tenuistipitata was first described by Chang and Xia in 1976 and they are distributed not only throughout the tropical and subtropical regions in the Western Pacific (Chang & Xia, 1988) but also in the Indian Ocean (Terada et al., 2000). This seaweed is intensively cultured in China for the production of agar and fodder for abalone (Lin & Liao, 1999). This species' wide tolerance to cultivation environment, high growth rate and high agar yield make it suitable for cultivation (Collén et al., 2003). A new variety, *G. tenuistipitata* var. *liui* was reported in 1988 (Chang & Xia, 1988) and this species can be found in Vietnam (Nguyen, 1992; Tseng & Xia, 1999; Tsutsui et al., 2005), Thailand (Lewmanomont, 1994), and Philippines (Abbott, 1994). *Gracilaria tenuistipitata* Chang and Xia was reported for the first time from Kuah, Pulau Langkawi (Terada et al., 2000) followed by Batu Laut, Morib in Selangor, Malaysia (Lim & Phang, 2004; Phang, 2006).

In 2004, Hagopian et al. constructed the chloroplast genome of *Gracilaria tenuistipitata* var. *liui*, which was the first plastid genome being successfully completed from the subclass Florideophycidae (Rhodophyta) (Hagopian et al., 2004). The life history of *Gracilaria tenuistipitata* var. *liui* was completed *in vitro* by Barufi et al. (2010) and the mitochondria genome was successfully sequenced in 2010 (Takahashi, 2010). Recently, the expressed sequence tags (ESTs) of *Gracilaria tenuistipitata* var. *liui* were established and deposited in a public database, the National Center for Biotechnology Information (NCBI) (Collén et al., 2012). This finding shows that there is increasing interest in *G. tenuistipitata* as a model organism for functional genomics which may aid in phylogenetic studies.

2.5 DNA Isolation and Purification

The isolation of DNA from algae especially red algae has been proven difficult (Chesnick & Cattolico, 1993; Wattier et al., 2000), due to the cell walls and hydrocolloids (agars and carrageenans) that may inhibit endonuclease and DNA polymerase activities (Jin et al., 1997).

Plants and macroalgae can yield large quantities of nucleic acids due to the large genome sizes. In macroalgae, nuclear DNA per cell varies over four orders of magnitude (200 - 0.2 pg); while in algal species with smaller genomes, there can be a 1000-fold difference between the sizes of the nuclear genome than that of the plastid genome (Coleman & Goff, 1991). Plants and macroalgae contain two other genomes, which are mitochondrial and plastid. The plastid DNA genome is a double-stranded DNA circle, which contains the genes for plastid rRNA, tRNA and other proteins in both plants and macroalgae. In contrast to macroalgae, which range from 73 to over 400 kilobase (kb), plant genomes can range from 120 - 217 kb (Palmer, 1991). Macroalgae have only a single copy of the ribosomal gene (Jorgensen et al., 1987) and small mitochondrial genomes ranging from 100 - 500 kb pairs. They are found in both circular and linear forms, similar to fungi and protozoa. Uni-parental inheritance is observed in macroalgae which is similar to metazoans (Harris, 1989).

The extraction of DNA from seaweed cells that are heavily embedded in sulfated polysaccharides (cell walls and intracellular matrix) involving ultracentrifugation which is time-consuming, labour intensive and expensive (Fain et al., 1988; Goff & Coleman, 1988; Parson et al., 1990; Shivji, 1991). It is vital that the method used for DNA purification is rapid, inexpensive, requires little cellular material, yielding DNA that can be used directly for PCR, and is effective across a broad range of algal lineages.

Furthermore, the use of phenol should be avoided as it is a highly toxic material (Fawley & Fawley, 2004).

Most of the published techniques for DNA extraction from green algae (Meusnier et al., 2004), brown algae (Philips et al., 2001) and red algae (Hong et al., 1997; Wattier et al., 2000) required grinding tissues in liquid nitrogen. Grinding of algal material in liquid nitrogen will released viscous soluble polysaccharides (Brasch et al., 1981) that are hardly to separate from DNA. Thus, cesium chloride (CsCl)-gradient ultracentrifugation (La Claire et al., 1997; Phillips et al., 2001), cetyl trimethyl ammonium bromide (CTAB) treatments (e.g. Fawley & Fawley, 2004), or lithium chloride (LiCl) (Hong et al., 1992, 1997) methods have been applied during DNA isolation in order to obtain high quality of DNA.

Recently, Varela-Alvarez et al. (2006) have presented a new method to extract DNA in recalcitrant algae and it is highly recommended for seaweed DNA extractions. This method involves a combination of two altered procedures developed previously by Triboush et al. (1998) for nuclei isolation, and by Doyle and Doyle (1990) for DNA isolation in land plants. The protocol is rapid, inexpensive, reproducible and effective in isolating genomic DNA from *Caulerpa* species. Moreover, this procedure provides genomic DNA of high quality with no degradation and with high yields from a small amount of raw material. Varela-Alvarez et al. (2006) also stated that this DNA isolation technique has been successfully used in PCR, cloning, hybridization, and in the construction of genomic libraries. Wattier et al. have also reported a DNA isolation method for red seaweed and the protocol is a modification of Dellaporta et al. (1983) technique for land plants. They have successfully minimised polysaccharide co-isolation and the DNA isolated from all red seaweeds was of good quality (Wattier et al., 2000).

2.6 Molecular Approaches

Since the early 1990s, the numbers of both phylogenetic studies and markers used have increased steadily. Phylogenetic studies have been carried out at all taxonomic levels across most of the spectrum of red algal biodiversity (Maggs et al., 2007). The genome of seaweeds can be subdivided into three components: the nuclear genome and the organellar genomes, which are mitochondrial (mtDNA) and chloroplastic (cpDNA). Although many of the markers have been employed for studies at various taxonomic levels, mitochondrial genes and plastid are mostly exploited to resolve the relationships between species belonging to one genus or one family; plastid or nuclear ribosomal DNA sequences are mostly used at higher taxonomic levels; and spacer sequences are almost exclusively used to infer haplotype trees or networks for closely related species and populations (Maggs et al., 2007). Due to the increasing availability of faster evolved organellar markers (*rbc*L, *rbc*L-S spacer, *cox*2-3 spacer), the use of the relatively slowly evolving SSU ribosomal DNA marker were gradually decreased. The most recently developed markers are highly variable markers mainly used for differentiation or study of phylogenetic relationship at the genus and species level, for example, the mitochondrial *cox*1 gene (Yang et al., 2007; Kim et al., 2012) and *cox*2-3spacer (Zuccarello et al., 2006; Vidal, 2008), the plastid genes *psb*A (Kim et al., 2006) and *psa*A (Yang & Boo, 2004), and the universal rice primer (URP) markers (Kang et al., 2002; Aggarwal et al., 2008).

The other genetic marker types that are used in this field include fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellite and single nucleotide polymorphism (SNP). These markers can be divided into two major classes: single-locus/co-dominant and multilocus/dominant markers. Single locus/co-dominant markers usually target a unique site within the nuclear genome and able to distinguish between

homozygous and heterozygous genotypes. Multilocus/dominant markers detect many genomic sites at a time but not able to distinguish homozygotes from heterozygotes (Maggs et al., 2007).

2.6.1 Random Amplified Polymorphic DNA (RAPD)

In 1990, random amplified polymorphic DNA (RAPD) technique was first developed (Welsh & McClelland, 1990; Williams et al., 1990) to randomly amplify template DNA using PCR without prior knowledge of the target sequence with a single arbitrary oligonucleotide primer of 8 – 10 base pair (bp) in length. Low annealing temperatures (34℃ – 40℃) are used for the amplification to produce a multilocus fingerprint pattern and scored as present and absent. Two criteria must be met in choosing the arbitrary RAPD primers: minimum of 40% GC content and the absence of palindromic sequence (a base sequence is exactly the same when one strand is read from left to right and the other strand is read from right to left) (Williams et al., 1990). The PCR products that are resolved in 1.5% - 2.0% agarose gels and stained with ethidium bromide (EtBr), can be viewed under UV light or resolved in polyacrylamide gels in combination with either $AgNO_3$ staining (Huff et al., 1993; Vejl, 1997; Hollingsworth et al., 1998), radioactivity (Pammi et al., 1994), or fluorescently labeled primers or nucleotides (CorleySmith et al., 1997; Weller & Reddy, 1997).

The advantages of RAPD marker are that no previous knowledge of the genome is required; large number of loci and individuals can be screened using low cost of agarose gel electrophoresis and low amount of DNA is required. However, they are subject to low reproducibility due to low annealing temperature used in the PCR amplification and difficulty in scoring bands, which leads to inappropriate inferences (Liu & Cordes, 2004). The presence of artifactual bands (noninformative) produced by

nested primer binding site (Schierwater et al., 1996; Rabouam et al., 1999) and intrastrand annealing during PCR amplification (Hunt & Page, 1992; Caetano-Anolles et al., 1992) were reported to influence the reliability of RAPD data for genetic diversity and genome mapping studies. Another concern is the lack of informativeness of RAPDs at the population level (Isabel et al., 1999), for example there is little or no differentiation in the genotypic information among the populations of white pines (*Pinus strobus* L.) when small population sample sizes are used. The other limitation of RAPD markers is that they are dominantly inherited and thus they are not able to distinguish the homozygous dominant from heterozygous individuals. There is no fragment produced from homozygous recessive alleles as the amplification is disrupted in both alleles (Semagn et al., 2008).

In Malaysia, molecular research using RAPD has focused on *Gracilaria* species for studying the two variants of *Gracilaria salicornia* collected at different localities. Two primers (OPA 01 and OPK 07) were able to show that these two morphological variants were due to genetic difference but not plasticity (Lim et al., 2001). The identification of sex-specific markers for *G. changii* were also carried out by Sim et al. using the same technique and results showed that RAPD was able to differentiate tetrasporophytes, females and males plants (Sim et al., 2007). In China, Wang et al. has showed the feasibility of the RAPD method in the genetic analysis and germplasm identification of the gametophytes of the *Undaria pinnafitida* that is important to provide molecular basis for their breeding (Wang et al., 2006). RAPD is also prominent in plants where RAPD was first used to differentiate the species of *Panax*, a type of ginseng that has a long history of usage in traditional medicine. The two main types of ginseng are Asian ginseng and American ginseng. American ginseng is believed to have cooling effects (yin) as opposed to Asian ginseng that has heating effects (yang). The differences between these main types of ginseng are the biological activity caused by

different bioactive compounds (ginsenosides) (Popovich et al., 2012). The identification and differentiation of *Panax* species was achieved by using RAPD method and *P. notoginseng* can be identified by the absence of ginsenoside Rc in the extract of *P. notoginseng* with Eastern blotting (Tanaka et al., 2006).

This technique has also been applied in diverse fields, for example the identification of eight species of fishes (Partis & Wells, 1996) and mollusks (Klinbunga et al., 2000; Crossland et al.,1993), genetic variation study in wild black tiger shrimp, *Penaeus monodon* (Tassanakajon et al., 1998) and genetic diversity study on the eastern oyster (*Crassostrea virginica*) (Hirschfeld et al., 1999).

2.6.2 Restriction Fragment Length Polymorphism (RFLP)

In 1975, restriction fragment length polymorphisms (RFLPs) were first used to identify DNA sequence polymorphisms for genetic mapping of a temperature-sensitive mutation of adeno-virus serotypes (Grodzicker et al., 1975). Later on, it was used for human genome mapping by Botstein et al. (1980) and then adopted for plant genomes (Helentjaris et al., 1986; Weber & Helentjaris, 1989). This technique is based on the principal that two individuals of the same species (with almost identical genomes) will always differ at a few nucleotides due to insertion/deletion, point mutation, translocation, inversion and duplication that can result in the gain, loss or relocation of a restriction site. Thus, digestion of DNA with restriction enzymes reveals a pattern difference between DNA fragment sizes among individuals, populations and species.

Restriction enzymes are enzymes isolated from bacteria that recognize specific four, six or eight base pair (bp) sequences in DNA and cut double-stranded DNA whenever these sequences are encountered. Most of the researchers used six-cutter restriction enzymes as they are readily available, cheaper and produced fragments in the

size range of 200 to 20,000 bp, which can be easily resolved by agarose gel electrophoresis (Potter & Jones, 1991; Semagn et al., 2006). Hence, the procedures of RFLP markers can be summarised as: i) digestion of double-stranded DNA with restriction enzymes; ii) separation of the restriction fragments in agarose gel; iii) transfer of separated fragments from agarose gel to a filter by Southern blotting; iv) detection of individual fragments by nucleic acid hybridization with labeled probes and v) autoradiogrphy (Perez de la Vega, 1993; Terachi, 1993; Landry, 1994).

RFLPs are co-dominant markers, i.e., both alleles in an individual are observed in the analysis, in linkage analysis and breeding. They can be easily determined in homozygous or heterozygous state of an individual and due to the large size difference, scoring is relatively easy (Liu & Cordes, 2004; Sharma et al., 2008). Besides, the size difference of the fragments is often large and thus the scoring is relatively easy. However, RFLP markers have low level of polymorphism with few loci are detected per assay. Large amount of high quality DNA are required for restriction digestion and Southern blotting, making it laborious, time-consuming and expensive to develop markers in species that are lacking molecular information (Semagn et al., 2006; Sharma et al., 2008).

There are very few studies of RFLP analysis on seaweeds and the early molecular study using this technique was in 1988 by Goff and Coleman, for a geographical study of *Gracilariopsis andersonii* (Grunow) Dawson (as *Gracilaria sjoestedtii*), showing populations over a 2000 km range to be remarkably genetically homogeneous. In 1996, RFLP was used to solve the taxonomic problems of *Gracilaria* (Gracilariales, Rhodophyta) from Chile (González et al., 1996).

2.6.3 Amplified Fragment Length Polymorphism (AFLP)

Amplified Fragment Length Polymorphisms (AFLPs) are a PCR-based, multi-locus fingerprinting technique that using a limited set of generic primers for amplification without any prior knowledge of the target genome. This technique was introduced in 1995 by Vos et al. based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA (Vos et al., 1995). AFLPs combine the strengths and overcome the weakness of the RFLP and RAPD methods by ligating primer recognition sequences (adaptors) to the restricted DNA (Lynch & Walsh, 1998).

Three steps are involved in this technique: i) restriction digestion of genomic DNA and ligation of oligonucleotide adaptors of known sequence; ii) two rounds of selective PCR amplification of sets of restriction fragments; and iii) gel analysis of the amplified fragments (Vos et al., 1995). The first PCR amplification occurred when the primers are able to anneal to fragments that have adaptor and the second amplification is achieved by the use of complementary base pairs that matched the additional nucleotides flanking the restriction sites. The additional nucleotide to the PCR primers could reduce the amplified fragments and thus enable to resolve on gel electrophoresis. It is possible for the combinations of many primers to amplify different subsets. AFLPs allow the amplification and analysis of many loci simultaneously.

The major advantages of AFLP method include large numbers of revealed polymorphisms, high reproducibility due to stringent PCR conditions and are extremely useful as tools for DNA fingerprinting, cloning and mapping of variety-specific genomic DNA sequences (Sharma et al., 2008). This technique is more expensive than RAPDs, but large number of loci can be analyzed from a single run and thus the cost per marker is reduced (Liu & Cordes, 2004). Its major weakness includes the need of special equipment such as automated gene sequencers for electrophoretic analysis of

fluorescent labels (Liu & Cordes, 2004), requires more number of steps to produce the result and most of the AFLP loci are dominant, which does not differentiate dominant homozygotes from heterozygotes (Semagn et al., 2006). The first application of this technique tested on the red alga *Chondrus cripus* and they identified both conservative and variable markers within and between populations while some markers were unique to individuals. They concluded that this method is suitable for genome mapping and population genetic investigations as AFLP is sensitive, reproducible and efficient for algal studies (Donaldson et al., 1998).

2.6.4 Microsatellite Markers (SSRs)

Microsatellites, alternatively known as simple sequence repeats (SSRs, Jacob et al., 1991), short tandem repeats (STRs, Craig et al., 1988), or simple sequence length polymorphisms (SSLPs, Rassmann et al., 1991), are tandem repeats of 1 – 6 nucleotides that are arranged throughout the genomes of most eukaryotic species (Powell et al., 1996; Zane et al., 2002). Some authors defined microsatellite as 2 – 8 bp repeats (Armour et al., 1999) and some defined as 1 – 5 bp repeats (Schlotterer, 1998). Chambers and MacAvoy (2000) followed the original authors and suggested 2 – 6 bp as the size of repeat units in microsatellites. They are the markers of choice for a variety of applications in plant genetics and breeding due to their multiallelic nature, higher levels of polymorphism, co-dominant inheritance and relatively abundant (Sun et al., 2006). Microsatellite loci have been identified for a number of seaweeds, such as the red *Gracilariopsis lemaneiformis* (Pang et al., 2010), the green *Ulva intestinalis* (Kostamo et al., 2008) and the brown *Saccharina (Laminaria) japonica* (Liu et al., 2010).

The predominant mutation mechanism in microsatellite is believed to be caused by a process called “slipped-strand mispairing” (Levinson & Gutman, 1987). During

DNA repair and replication, slipped-strand mispairing occurs whenever unpaired loops form. This process involved local denaturation and displacement of the double-stranded DNA, followed by mispairing of complementary bases at the site of an existing short tandem repeat. The consequences of this mispairing, when followed by replication or repair, will lead to the insertions or deletions of one or several of the short repeat units. This process has bias toward the duplication of shorter repeat units (Levinson & Gutman, 1987).

Microsatellite loci can be classified on the basis of the repeat motif length (i.e. dinucleotide, trinucleotide, tetranucleotide, etc.) and motif contiguity (e.g. perfect or interrupted). Characteristics of the repeat motif (type, length and contiguity) appear to affect the rate of mutation and levels of allelic variation. Interruptions within the core sequence stabilize the arrays of repeats and lowering levels of polymorphism, making microsatellites with interrupted repeats less variable than loci with pure repeats (Richards & Sutherland, 1994; Pepin et al., 1995; Petes et al., 1997). Levels of allelic diversity are also correlated with repeat length (Weber, 1990), as loci with longer repeats are generally more polymorphic than loci composed of short motif (Beckmann & Weber, 1992).

Primers of microsatellites for PCR amplification are designed from sequences flanking specific microsatellite loci and it is necessary to create and screen a genomic DNA library for the presence of microsatellites, sequence the regions flanking the loci, then design and test PCR primers for these loci (Ashley & Dow, 1994). The conventional method of developing SSR markers are labour intensive, time-consuming and costly (Kantety et al., 2002; Varshney et al., 2002; Yu et al., 2004), but now an abundance of expressed sequence tags (ESTs) and genomic sequences are available in online databases. To date, the potential usage of microsatellite markers in seaweeds has been tested in *Gracilaria gracilis* (Wattier et al., 1997) and *Porphyra haitanensis* (Xie

et al., 2009) for assessing genetic variability, *Undaria pinnatifida* for exploring genetic structure (Voisin et al., 2005), *Porphyra* sp. for germplasm identification (Sun et al., 2006), *Asparagopsis taxiformis* for determining ploidy level and sexual reproduction (Andreakis et al., 2007), and *Gracilariopsis lemaneiformis* (Pang et al., 2010) and *Chondrus crispus* for performing genetic diversity studies (Provan & Maggs, 2012).

The method of assaying and analyzing microsatellite loci has several advantages over other molecular techniques. For example, in the minisatellite "DNA fingerprinting" approach, variation occurs as the result of varying numbers of repeat units. However, the core units in microsatellites are generally 1 – 6 bp in length rather than the minisatellites which is 10 – 60 bp. The standard multilocus DNA fingerprinting examines variability of many minisatellite loci at one time, therefore allelic relationships among bands are generally unknown, and genotypes at specific loci cannot be scored. The presence of many bands of unknown loci specificity makes it difficult to evaluate relationships among individuals. However, co-dominant alleles at single microsatellite loci can be unambiguously scored by size, making gel to gel comparisons straightforward (Ashley & Dow, 1994). Microsatellite analysis relies on polymerase chain reaction amplification of DNA (PCR; Saiki et al., 1988) rather than Southern blotting and hybridization with labeled probes. It is therefore much faster and requires much less DNA, which can be of low quality.

For visualization, PCR primers may be end-labeled with ^{32}P-labeled nucleotides or ^{35}S-labeled nucleotides. Another method is silver staining which does not require the use of radioactivity (Ashley & Dow, 1994). The co-dominant inheritance of microsatellite loci should yield one (for homozygotes) or two (for heterozygotes) bands in the appropriate size range. However, there are few fainter bands present as a result of PCR artifacts during amplification. Longer repeats (tri- and tetranucleotide) have the advantage of being less prone to such artifacts but they are rarer than dinucleotide

repeats (Rassmann et al., 1991; Hughes & Queller, 1993). Recently, automated DNA sequences and fragment analyzers have become more affordable and products can be visualized with fluorescent labeling. The use of primers from one species on another ("cross-amplification") is very useful in plants (Ritland & Ritland, 1994). These highly polymorphic SSR markers are useful in germplasm characterization, marker-assisted selection, cultivar identification, and genetic diversity and phylogenetic relationship studies (Wang et al., 2009).

All marker types have some pitfalls and microsatellite markers also have several challenges despite many advantages. However, the versatility of microsatellites to address the ecological questions outweighs their drawbacks for many applications and the pitfalls can be avoided by careful selection of loci. For example, the possibility of null alleles in which the SSRs fail to amplify due to primer site variation, and thus do not produce a visible amplicon. This problem occurred either because the PCR conditions are not ideal or the primer-binding region contains mutations that inhibit binding (Selkoe & Toonen, 2006). Redesigning primers to bind to a different region of the flanking sequence, or adjusting PCR conditions can help to improve null allele problems (Pemberton et al., 1995). The other phenomenon called "homoplasy" occurs because alleles are scored as size variants, so it is possible for alleles of identical size to be the products of different evolutionary lineages. This can result in false inference, where allele of identical sizes (identity of state) is taken to imply the same structure and common genetic history (identity by descent) (Selkoe & Toonen, 2006). This always happens when compound and/or interrupted loci are used (Chambers et al., 1997; Adams et al., 2004). Table 2.1 summarized the characteristics and potential applications of the markers discussed.

Table 2.1: Comparison of different DNA markers that are widely used in plants (adapted from Semagn et al., 2006).

Features	RFLP	SSR	AFLP	RAPD
Amount of DNA required	High	Low	Medium	Low
Inheritance	Co-dominant	Co-dominant	Dominant	Dominant
Type of Polymorphism	Single base changes, Insertion, Deletion	Changes in length of repeats	Single base changes, Insertion, Deletion	Single base changes, Insertion, Deletion
Level of Polymorphism	Medium	High	High	High
Reproducibility (Reliability)	High	High	High	Intermediate
Type of probes/ primers	Low copy genomic DNA or cDNA clones	Specific repeat DNA sequence	Specific sequence	Usually 10 bp random nucleotides
Ease of Use	Labour intensive	Easy	Difficult Initially	Easy
Detection of Alleles	Yes	Yes	No	No
Major Applications	Linkage mapping	Linkage mapping, Population studies, Paternity analysis	Linkage mapping, Population studies	Fingerprinting for population studies, Hybrid identification

RFLP: Restriction Fragment Length Polymorphism; SSR: Simple Sequence Repeat or Microsatellite; AFLP: Amplified Fragment Length Polymorphism; RAPD: Random Amplified Polymorphic DNA

2.6.4.1 Comparison of Genic and Genomic Microsatellite Markers

The conventional method of developing SSR markers involving generating of a genomic library and it is generally made using small fragments (300 - 600 bp) of DNA (Rassmann et al., 1991; Hughes & Queller, 1993). High molecular weight DNA is digested to completion using one or more restriction enzymes having four base pair recognitions sites. The use of several restriction enzymes has the effect on increasing the proportion of fragments containing microsatellites by frequently restricting the DNA in nonrepetitive regions, but not in microsatellite regions. There are two reasons that size selection of DNA is important. First, the entire insert fragment can be sequenced, guaranteeing that the microsatellite region of any positive clone will be found. Second, clones of single insert fragments can be verified by length to avoid identifying PCR primers from concatemers of two or more insert fragments from different regions of the genome (Ashley & Dow, 1994). M13 is a good choice for constructing the microsatellite libraries as single-stranded DNA can be easily harvested for sequencing once M13 plaques are obtained. Double-stranded DNA can be sequenced by using a denaturing process, but the reliability of single-stranded sequence is higher, particularly for microsatellite regions which may have different annealing properties than non-repeat regions (Ashley & Dow, 1994).

Once clones have been obtained, recombinant DNA from plaques or colonies is transferred to a membrane and hybridized to a radioactive probe using standard protocols. Dinucleotide repeats such as $(CA/GT)_n$ are often used as probes as they are highly abundance in eukaryotic genomes. Autoradiographs of membranes are used to identify positive plaque, which mean plaques that hybridize to the dinucleotide probes. DNA is then extracted from positive clones for use in sequencing. Perfect repeats (with no substitutions interrupting the core motif) with a length of 15 or more repeat units should be used to design PCR primers. Screening several positive clones for

identification of useful microsatellite loci is necessary for several reasons. The number of repeats may be too low, repeats may be imperfect, the microsatellite may located too close to the flanking sequences, or the flanking sequences may be inappropriate for primer design (e.g. having very low GC content or composed of other repeat motifs). When an appropriate microsatellite locus has been identified, PCR primers are designed using computer programs (Ashley & Dow, 1994).

Nowadays, many expressed sequence tags (ESTs) sequencing projects have been performed for seaweeds that can be obtained in the online database. They are short DNA sequences produced from cDNAs and expressed in a cell at a particular given time. cDNA is a stable compound and it represents only expressed DNA sequence as it is generated from mRNA that represents exons by excising introns. Originally, ESTs were produced to identify gene transcripts, but are now been used in gene discovery, for obtaining data on gene expression and regulation, sequence determination, and for developing highly valuable molecular markers, such as EST-based RFLPs, SSRs, and SNPs (Semagn et al., 2006). ESTs have also been used for designing probes for DNA microarrays that is used to determine gene expression. However, EST data have some limitations. A few particular tissues sampled may not be sequenced at all as the transcripts are in low abundance. Besides, ESTs only represent partial sequences of the original transcripts, and give no information about genomic position, gene order, introns or regulatory motifs (Bouck & Vision, 2007). Despite these limitations, ESTs are still a valuable resource for molecular ecology.

The development of SSRs from ESTs databases has been reported in a few seaweeds such as *Laminaria digitata* (Billot et al., 1998), *Cladophoropsis membranacea* (van der Strate et al., 2000), *Fucus serratus* and *Fucus evanescens* (Coyer et al., 2002), *Enteromorpha intestinalis* (Alström-rapaport & Leskinen, 2002), *Postelsia palmaeformis* (Whitmer et al., 2002), *Ascophyllum nodusum* (Olsen et al.,

2002), *Gracilaria chilensis* (Guillemin et al., 2005), *Laminaria japonica* (Shi et al., 2007) and *Chondrus crispus* (Provan & Maggs, 2012). There are also genomic libraries that have been constructed for a few seaweeds and can be found in online database for example, plastid genome of *G. tenuistipitata* (Hagopian et al., 2004) and *Porphyra purpurea* (Reith & Munholland, 1995); mitochondrial genome of *Chondrus crispus* (Leblanc et al., 1995) and *Porphyra purpurea* (Burger et al., 1999).

Complete sequences of the chloroplast genome led to the unexpected identification of chloroplast microsatellites (Powell et al., 1995a,b; Vendramin et al., 1996). There are a few introns in the tRNA regions contain repeats. Powell et al. (1995a,b) first identified mononucleotide repeats of microsatellites in chloroplast sequences, finding mostly A or T repeats. Chloroplast DNA (cpDNA) varies in terms of their haploid nature, uniparental inheritance and lack of recombination compared with the nuclear genome (Zuccarello et al., 1999; Wattier & Maggs, 2001). Chloroplast SSRs (cpSSRs) have been used mostly in plant studies such as the design of cpSSRs primers to amplify targeted regions in a diverse array of plant species (Chung & Staub, 2003), the development of universal primers to amplify SSRs in grasses by Provan et al. (2004a) and the development of a set of universal cpSSRs primers to explore cpDNA diversity among sub-tropical and tropical fruit crops (Cheng et al., 2006). However, few studies of cpSSRs have been conducted in algae, such as the development of universal primers for the amplification of chloroplast coding and non-coding regions in Chlorophyta and Rhodophyta (Provan et al., 2004b).

Genic microsatellites (EST-SSRs) have a number of advantages over genomic SSRs in which they tend to be more widely transferable between species and genera as the primers were designed from the more conserved coding regions of the genome. EST-SSRs are located in gene rich regions of the genome and that the sequence can be compared to protein sequence databases, possibly knowing the functional identity of a

particular marker locus (Bouck & Vision, 2007). The disadvantage of EST-SSRs is that they are less polymorphic in comparison to genomic SSRs as they are situated in more conserved regions of the genome (Varshney et al., 2002).

2.6.5 Single Nucleotide Polymorphism (SNP)

Single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide in the genome differs between members of a species (or between paired chromosomes in an individual). SNPs may occur in the coding, non-coding and intergenic regions of the genome, thus enabling the discovery of genes as a result of the differences in the nucleotide sequences (Ayeh, 2008). SNP is the new generation molecular makers that can be used for mapping, marker-assisted breeding and map-based cloning (Batley et al., 2003). A large amount of SNP data is available in higher plants, but very limited data are available in seaweeds. In plants, study of SNPs in *Arabidopsis* was initiated where over 37,000 SNPs have been identified (Jander et al., 2002). It has been reported in maize that there occurs a frequency of one non-coding SNP per 31 bp and 1 coding SNP per 124 bp in 18 maize genes assayed in 36 inbred lines (Ching et al., 2002). In seaweeds, a set of 35 SNP markers has been developed for the brown alga *Fucus vesiculosus* and has successfully discriminate populations of this species at different spatial scales (Canovas, 2011).

SNPs are excellent markers in modern phylogeography and evolutionary ecology (Garvin et al., 2010). They are less polymorphic than microsatellite markers but they are more abundant throughout the genome. Due to the low mutations rates compare to microsatellites, SNPs are more evolutionarily stable (Canovas, 2011). Their discovery with insertions/deletions has formed the basis of most differences between alleles and thus they can be explained as any polymorphism between two genomes that

is based on a single nucleotide exchange. In recent years, rapid identification of SNP markers using EST data sources has led to high-volume, cost-effective SNP discovery in plants that can be used for estimating functional genetic diversity (Picoult-Newberg et al., 1999 and Varshney et al., 2007).

2.7 Molecular Studies in *Gracilaria*

In mid-1980s, the first molecular phylogenetic study that included red algae was carried out. The most ancient eukaryote lineage was found to be the red algae based on the analyses using 5S ribosomal sequences of representative prokaryotic and eukaryotic organisms (Hori et al., 1985; Lim et al., 1986; Hori & Osawa, 1987). Other molecular study includes restriction fragment length polymorphism (RFLP) analysis of plastid DNA for a geographical study of *Gracilariopsis andersonii* (Grunow) Dawson (as *Gracilaria sjoestedtii*) by Goff and Coleman, 1988. In 1990, the phylogenetic position of *Gracilariopsis* was determined from 18S nrDNA sequences (Bhattacharya et al., 1990). In red algae, the small (*rbc*S) and large (*rbc*L) subunits of the ribulose biphosphate carboxylase/oxygenase (RuBisCO) gene are organised as a co-transcribed operon including a short highly variable intergenic spacer (the rubisco spacer) (Kostrzewa et al., 1990). This spacer region was employed for the study amongst member of the *Gracilaria verrucosa* species complex (Destombe & Douglas, 1991). In 1992, the first molecular study at family to species level on red algae was carried out using the SSU nrDNA to determine the phylogenetic relationships between members of the Gracilariaceae (Bird et al., 1992). In 1994, they used sequence comparison to determine the relationships of unidentified species among the Gracilariaceae.

The phylogenetic study on red seaweeds has gradually increased since the early 1990s. In 1996, DNA analyses were conducted using RFLP and RAPD, tested on

morphotypes of *Gracilaria* collected from different geographical locations (González, 1996). Later on, single locus microsatellites were used in the taxonomic study of *Gracilaria gracilis* (Wattier, 1997). Although the number of molecular phylogenetic markers used has increased, plastid and mitochondrial genes are still mostly used by the researchers for studies at various taxonomic levels.

Studies on the structure of their reproductive organs and the phylogenetic relationships among species using *rbc*L sequence analyses have produced three clades at the genus level, namely *Gracilaria sensu stricto*, *Gracilariopsis*, and *Hydropuntia* (Gurgel & Fredericq, 2004). Nowadays, DNA sequence analysis is the most widely used molecular technique for inferring phylogenetic relationships at the species level within the Gracilariaceae (Iyer et al., 2005; Gargiulo et al., 2006; Bellorin et al., 2008; Kim et al., 2008*a*). *Gracilaria* and *Gracilariopsis* have been confirmed to be monophyletic groups based on the study from global phylogeny and taxonomic biogeography of Gracilariaceae conducted (Bellorin et al., 2002; Gurgel & Fredericq, 2004; Guillemin et al., 2008) whereas *Hydropuntia* has received less support as a monophyletic group (Gurgel & Fredericq, 2004). Recently, DNA barcoding has become an important tool for the identification of gracilarioid species (Saunders, 2009; Kim et al., 2010).

CHAPTER 3: MATERIALS AND METHODS

3.1 Sample Collection and Specimen Processing

Most of the *Gracilaria* specimens were collected at various localities in Malaysia and some of the samples were provided by the co-authors from other countries. For example, *Gracilaria changii* were collected from Morib, Selangor, Malaysia (Figure 3.1) ii) Middle Banks, Penang, Malaysia iii) Sungai (Sg.) Merbok, Kedah, Malaysia iv) Teluk Pelandok (Figure 3.2) and Pantai Dickson, Negeri Sembilan, Malaysia v) Gua Tanah, Batu Besar (Figure 3.3) and Batu Tengah, Malacca, Malaysia vi) Sg. Pulai and Sg. Kong-Kong, Johor, Malaysia vii) Sandakan, Sabah, Malaysia and viii) Bintulu, Sarawak, Malaysia. *Gracilaria tenuistipitata* were collected from i) Batu Laut, Selangor, Malaysia (Figure 3.4) ii) Middle Banks, Penang, Malaysia (Figure 3.5 and 3.6) iii) Kuah, Pulau Langkawi, Malaysia (Figure 3.7) iv) Kelantan, Malaysia (Figure 3.8). Specimens from other countries such as *Gracilaria manilaensis* from Cebu, Philippines (Figure 3.9), *Gracilaria fisheri* from Pattani, Thailand (Figure 3.10) and *Gracilaria tenuistipitata* from Quy Kim, Hai Phong, Vietnam (Figure 3.11) and Pattani, Thailand (Figure 3.12). Table 3.1, 3.2 and 3.3 shows the species name, location and number of samples that have been used for different studies.

All specimens were identified based on the morphological descriptions from Xia & Abbott (1987) and Lim & Phang (2004). Female gametophytes (haploid) and tetrasporophytes (diploid) were identified by their reproductive organs. Due to the low number of female gametophytes that we obtained, only tetrasporophytes were used in this study. The samples were kept in ice chest prior transported back to laboratory. In the laboratory, the samples were washed thoroughly with seawater, followed by distilled water and UHQ water. The samples were blotted dry with a clean towel paper and kept in the sealed plastic bags with silica gel. The samples were kept at -20 ℃ in refrigerator.

Figure 3.1: Fresh specimen of *Gracilaria changii* from Morib, Selangor, Malaysia.

Figure 3.2: Herbarium specimen of *Gracilaria changii* from Teluk Pelandok, Negeri Sembilan, Malaysia.

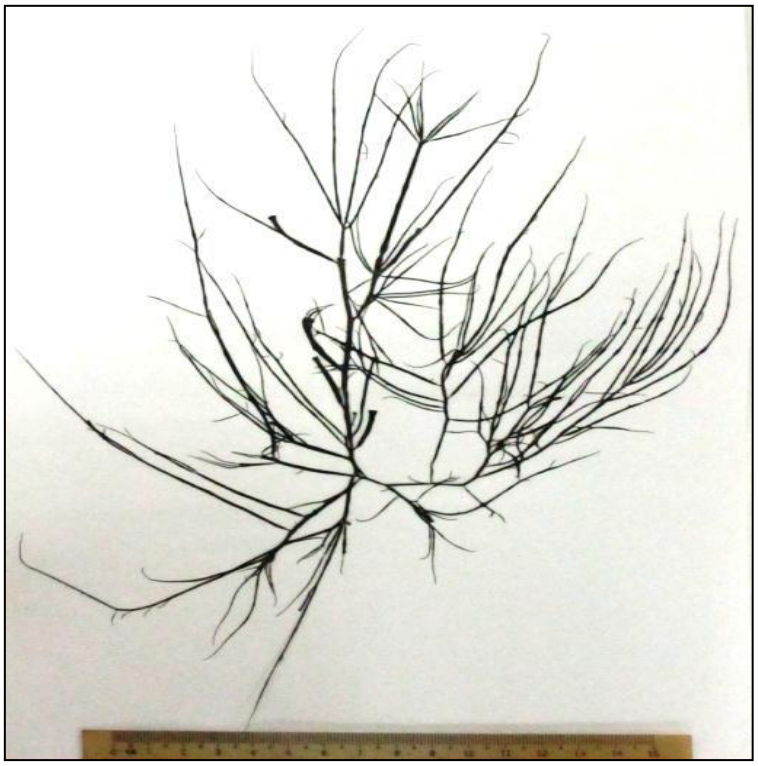


Figure 3.3: Herbarium specimen of *Gracilaria changii* from Batu Besar, Malacca, Malaysia.

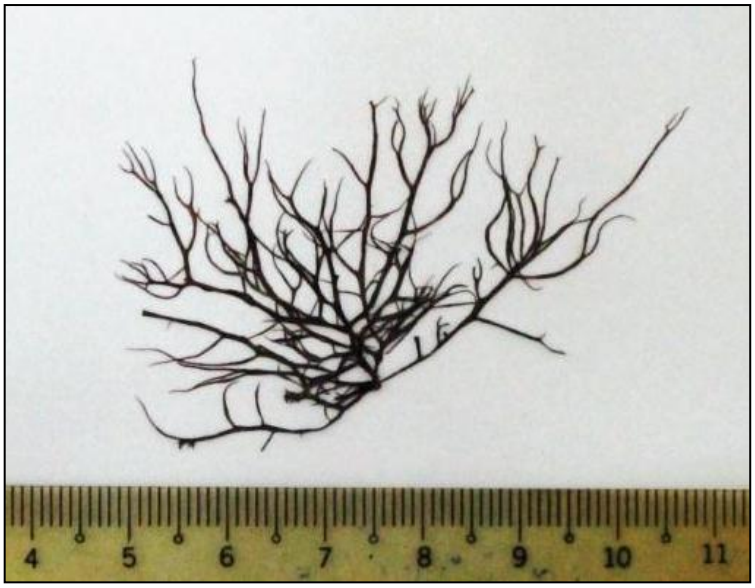


Figure 3.4: Herbarium specimen of *Gracilaria tenuistipitata* from Batu Laut, Selangor, Malaysia.

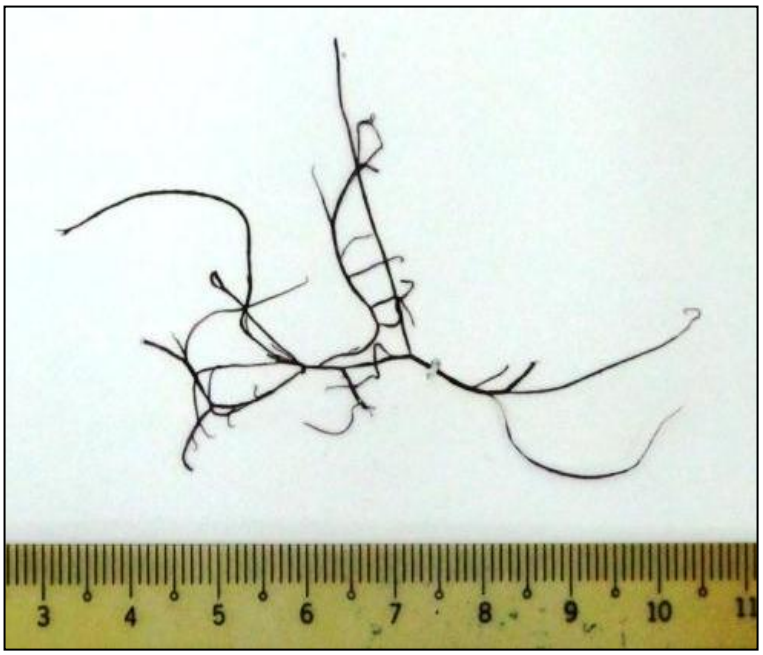


Figure 3.5: Herbarium specimen of *Gracilaria tenuistipitata* from Middle Banks, Penang, Malaysia.

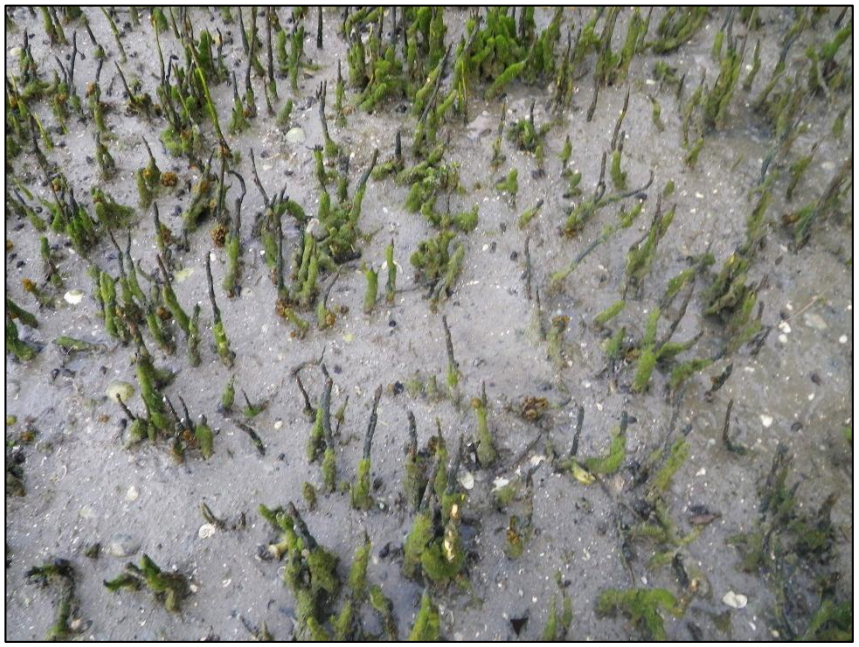


Figure 3.6: The collection site of *Gracilaria tenuistipitata* in a mangrove area at Middle Banks, Penang, Malaysia.

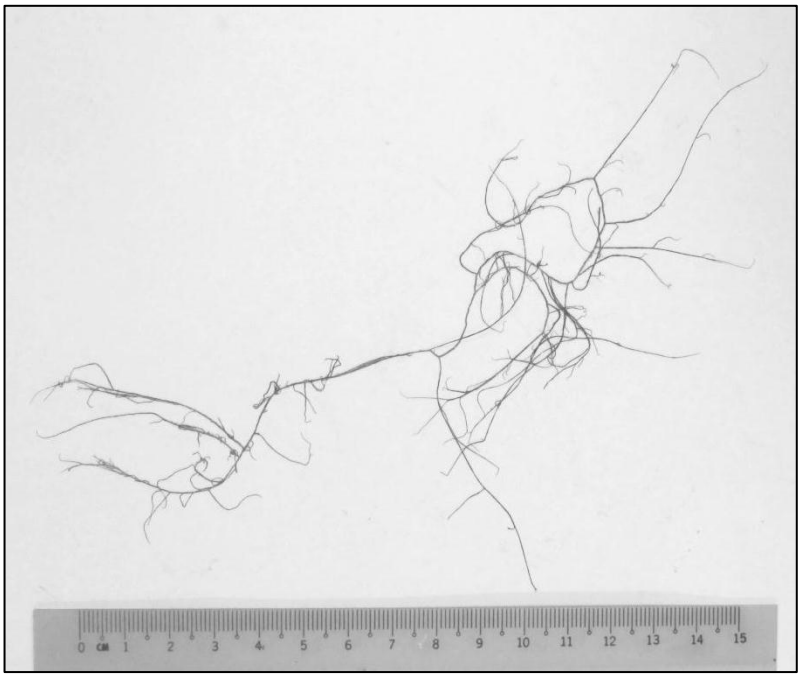


Figure 3.7: Herbarium specimen of *Gracilaria tenuistipitata* from Kuah, Pulau Langkawi, Malaysia.

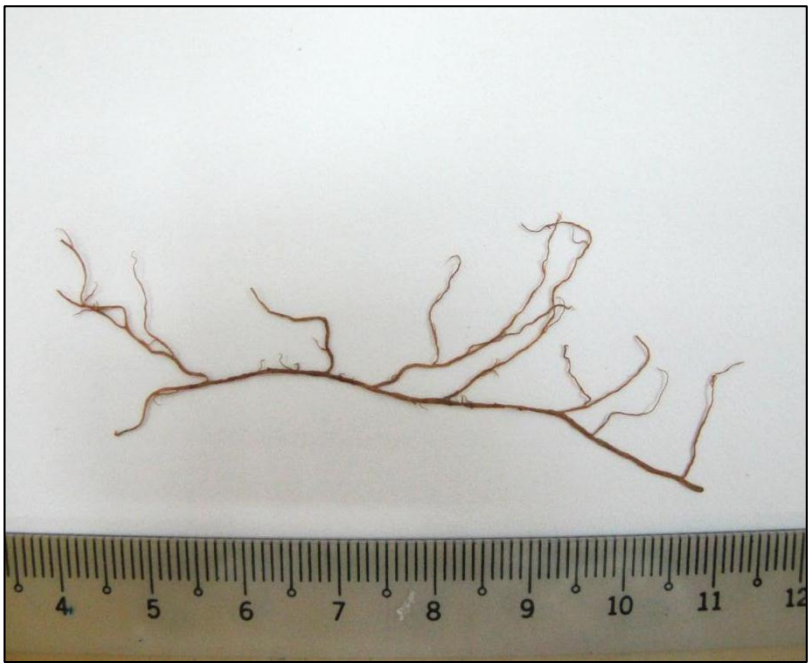


Figure 3.8: Herbarium specimen of *Gracilaria tenuistipitata* from Kelantan, Malaysia.

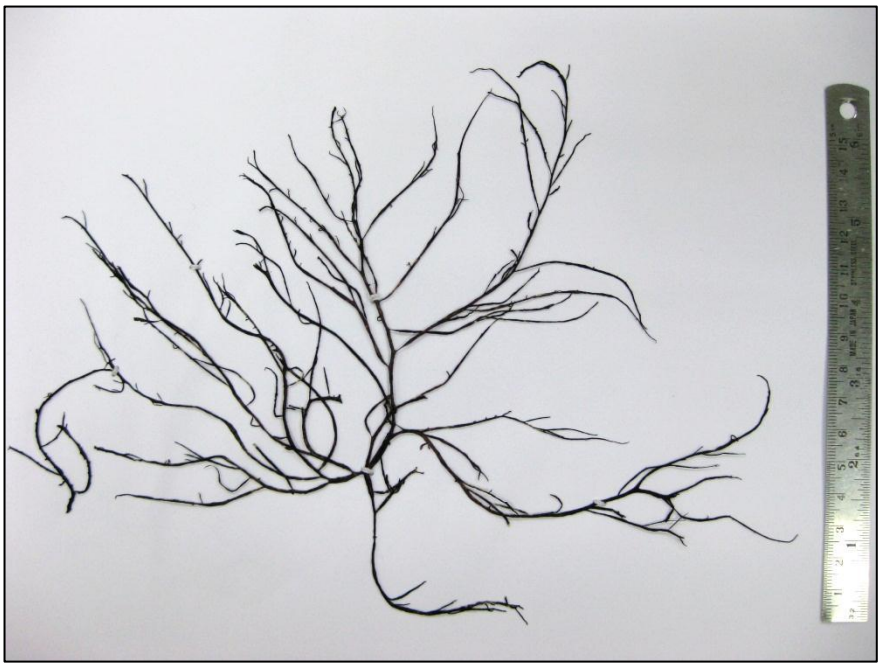


Figure 3.9: Herbarium specimen of *Gracilaria manilaensis* from Cebu, Philippines.

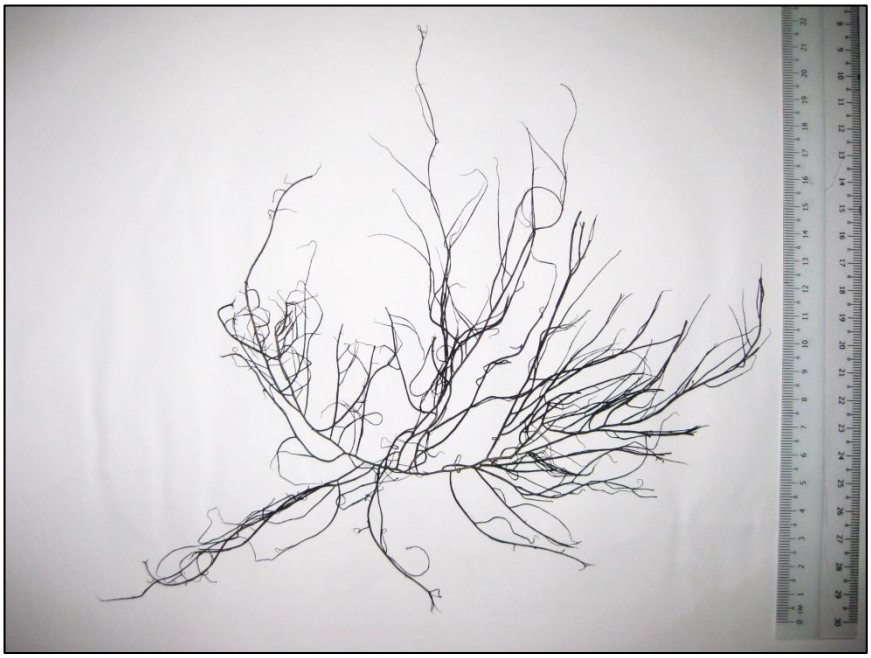


Figure 3.10: Herbarium specimen of *Gracilaria fisheri* from Pattani, Thailand.

Figure 3.11: Herbarium specimen of *Gracilaria tenuistipitata* from Quy Kim, Hai Phong, Vietnam.

Figure 3.12: Herbarium specimen of *Gracilaria tenuistipitata* from Pattani, Thailand.

Table 3.1: Species name, locality and number of samples used for the analysis of primer-pairs designed from chloroplast genome and ESTs of *Gracilaria tenuistipitata*.

Species Name	Locality		Number of Samples
	Pulau Korea Besar, Penang	Malaysia	15
	Batu Laut, Selangor	Malaysia	15
G. tenuistipitata	Pulau Langkawi, Kedah	Malaysia	10
	Kelantan	Malaysia	10
	Quy Kim, Hai Phong	Vietnam	15
	Pattani	Thailand	15
G. changii	Morib, Selangor	Malaysia	5
G. canaliculata = G. crassa		Malaysia	4
G. salicornia		Malaysia	5
G. cacalia		Malaysia	5

Table 3.2: Species name, locality and number of samples used for the analysis of primer-pairs designed from ESTs of *Gracilaria* species (*G. changii*, *G. gracilis* and *Gp. lemaneiformis*).

Species Name	Locality		Number of Samples
	Morib, Selangor		10
	Batu Besar, Malacca		10
	Batu Tengah, Malacca		10
	Gua Tanah, Malacca		10
	Middle Banks, Penang		10
G. changii	Sg. Merbok, Kedah	Malaysia	8
	Teluk Pelandok, Negeri Sembilan		10
	Pantai Dickson, Negeri Sembilan		10
	Sg. Kong-Kong, Johor		9
	Sg. Pulai, Johor		10
	Sandakan, Sabah		10
	Bintulu, Sarawak		10
G. edulis (Figure 3.13)		Malaysia	15
G. salicornia (Figure 3.14)		Malaysia	7
G. manilaensis	Cebu	Philippines	10
G. fisheri	Pattani	Thailand	10

Table 3.3: *Gracilaria changii* from various localities in Malaysia with number of specimens and haplotype identified using *cox*1 gene marker.

Species Name	Locality	Number of Samples	Haplotype (Yow et al., 2011)
G. changii	Sungai Merbok, Kedah	7	C1
	Middle Banks, Penang	11	C1
	Morib, Selangor	10	C1, C4 and C5
	Gua Tanah, Malacca	9	C1
	Batu Besar, Malacca	8	C1 and C2
	Batu Tengah, Malacca	7	C1
	Sungai Pulai, Johor	6	C1
	Teluk Pelandok, Negeri Sembilan	4	C3 and C6

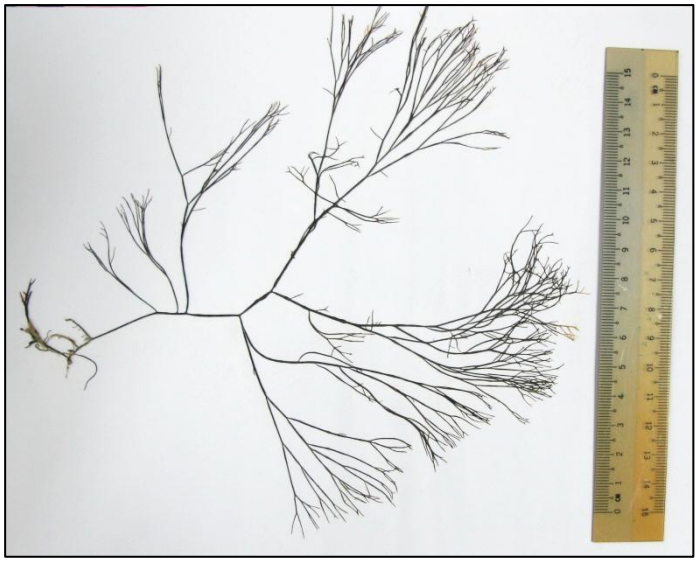


Figure 3.13: Herbarium specimen of *Gracilaria edulis* from Batu Besar, Malacca.

Figure 3.14: Herbarium specimen of *Gracilaria salicornia* from Morib, Selangor.

3.2 Molecular Analyses

3.2.1 Data Mining of Microsatellite Markers

Seaweed ESTs and genome DNA sequences were downloaded from GenBank (http://www.ncbi.nlm/nih.gov) and saved in FASTA format. Perl scripts were developed to search for microsatellite tandem repeats in the genome and EST sequences data using the MIcroSAtellite (MISA) search module (http://pgrc.Ipk-gatersleben.de/misa/). CAP3 software (Huang & Madan, 1999) was used to assemble the sequences into contigs as the EST sequences were in redundant form. To avoid homoplasy, only perfect SSRs (with no substitutions interrupting the core motif) and nucleotides with repeated patterns of length two or more were used in this study.

3.2.2 Designing and Synthesizing of Primers for SSRs

Primers were designed using Primer3 (Rozen & Skaletsky, 2000). The parameters used for the design of SSR primers were defined in such a way that the primer melting temperatures varied from 48℃ to 55℃ with primer length within 20 – 24 bp; GC content between 50% - 70% and the expected product size between 280 bp and 350 bp. If the numbers of the generated primers are too large, further selection to minimize the number of primers will be conducted. The selection will be based on several criteria that have been shown from literature review to have high success rate such as the number of tandem repeats and the pattern of the repeat sequences (Ashley & Dow, 1994).

3.2.3 Genomic DNA Extraction from Plant Tissue using DNeasy Plant Mini Kit

DNAs were extracted from *Gracilaria* samples using the DNeasy Plant Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. Before genomic DNA extraction, *Gracilaria* were ground to a fine powder by using a mortar and pestle with liquid nitrogen. The starting material used was weight approximately 0.1 g of dried samples. The grounded sample was transferred to a 1.5 mL microcentrifuge tube.

Initially, 400 µL of lysis Buffer AP1 and 4 µL of RNase A Stock Solution (100 mg/mL) was added. The mixture was incubated for 10 minutes at 65℃ in a water bath. During the incubation, the mixture was mixed 2 to 3 times by inverting the tube gently.

Then, 130 µL of buffer AP1 was added into the lysate. It was mixed and incubated for 5 minutes on ice. The lysate was centrifuged for 5 minutes at the speed of 12,000 rpm. After centrifugation, the lysate was applied to the QIAshredder Mini Spin Column (lilac) sitting in a 2 mL collection tube and was centrifuged for 2 minutes at the speed of 12,000 rpm. Approximately 450 µL of lysate was recovered. Then, 1.5 volumes of binding Buffer AP3/E were added to the cleared lysate and the mixture was mixed by inverting the microcentrifuge tube.

Next, 650 µL of the mixture including any precipitate which may have formed was applied to the DNeasy Mini Spin Column sitting in a 2 mL collection tube. The mixture was centrifuged for 1 minute at 8,000 rpm. The DNA was bounded to the membrane of the DNeasy Mini Spin Column and the flow-through was discarded.

After that, the DNeasy column was transferred to a new 2 mL collection tube and 500 µL of Buffer AW was added to the DNeasy column and was centrifuged for 1 minute at 8,000 rpm. The flow-through was discarded and this was repeated once.

Subsequently, the DNeasy column was centrifuged for 2 minutes at 12,000 rpm to dry the membrane. It is important to dry the membrane of the DNeasy column since residual ethanol may interfere with subsequent reactions. This spin ensures that no residual ethanol will be carried over during elution. The flow-through and collection tube were then discarded. Following the spin, the DNeasy column was removed from the collection tube carefully so that the column would not contact with the flow-through as this would have resulted in carrying over of ethanol. After washing with Buffer AW, it was observed that the DNeasy Mini Spin Column membrane is slightly coloured.

Finally, for the elution process, the DNeasy column was transferred to a new 1.5 mL microcentrifuge tube. 50 µL of preheated (65 ℃) Buffer AE was pipetted directly onto the DNeasy membrane. It was incubated for 5 minutes at room temperature and then centrifuged for 1 minute at 8,000 rpm to elute the DNA.

3.2.4 Optical Density (OD) Reading

Optical density measurement was carried out to quantify the DNA yield and purity. 1.0 µL of our sample was added with 49 µL of AE buffer in the cuvette and measured using spectrophotometer. The absorbance of UV light at 260 nm wavelength by nucleic acids gives an estimate of concentration and the ratio of readings taken at 260 nm and 280 nm wavelength gives an indication of the purity of nucleic acid. For pure DNA, ratio of OD 260/280 should be in the range of 1.8 – 2.0.

3.3 Microsatellite Markers (SSRs)

3.3.1 PCR Parameters

The designed primer-pairs were used to test for successful amplification in all the specimens collected. Polymerase chain reactions (PCRs) were performed in a total volume of 15 µL containing 1.5 µL 10x PCR buffer (Takara Biotechnology, Dalian, China), 0.2 mM dNTP mix (Takara), 0.5 U *Taq* polymerase (Takara), 0.3 mM of each primer-pairs, and 25 - 50 ng genomic DNA with UHQ water added to a total volume of 15 µL. In order to obtain the optimum annealing temperature (*Ta*) for PCR reactions, a range of annealing temperature (46 ℃, 50 ℃ and 54 ℃) were tested with 14 *Gracilaria* specimens. We found that *Ta* with 50 ℃ and 54 ℃ gave higher number of amplicons compared to 46 ℃. Therefore, we tested again using *Ta* = 52 ℃ and this annealing temperature showed better results in terms of quantity and quality of amplicons. Thus, *Ta* was set at 52 ℃ for all the specimens tested. The amplification reaction conditions were as follows: 5 minutes denaturation step at 94 ℃, followed by thirty-five cycles of 1 minute at 94 ℃, 1 minute at 52 ℃, 2 minutes elongation at 72 ℃, and a final extension at 72 ℃ for 5 minutes (Table 3.4). To ensure that those specimens with no amplification were not linked with standard PCR problems (low quality of DNA or experimental errors), each failed PCR was repeated using different amount of DNA or re-extract the DNA of a particular specimen.

Table 3.4: Parameters for PCR amplification.

Phase	Temperature	Duration
Denaturation	94 ℃	5 min
Denaturation	94 ℃	1 min
Annealing	52 ℃	1 min
Extension	72 ℃	2 min
		35 cycles
Final cycle	72 ℃	5 min
Storage	4 ℃	∞

3.3.2 SSRs Analysis

The amplification products were separated on a 3.0% MetaPhor® agarose gel (FMC or Cambrex Corporation, USA) at 75 V for 75 minutes and visualized by SYBR SAFE staining (Invitrogen, USA) under UV light.

3.4 Data Analysis

3.4.1 PCR Amplifications Using Labeled Primer and DNA Fragment Analysis (FA)

Primer-pairs that showed good amplification in all the *Gracilaria* species were considered usable. The forward primer of each primer-pair was fluorescently [69-carboxy-fluorecine (FAM)] labeled and PCR amplification was conducted. The optimized annealing temperature for PCR amplifications was 52 ℃, and this optimum temperature was used to test on all the specimens. Amplification products were then sent for fragment analysis to detect alleles using an automated DNA sequencer (First Base Sdn. Bhd.).

3.4.2 Data Analysis Using GelQuest, ClusterVis and TreeMe Software

SSR products amplified by each primer-pair were scored using a binary matrix method with "1" (presence) and "0" (absence). These DNA fingerprints of *Gracilaria* specimens from different locations were constructed using GelQuest software (Hepperle, 2002) based on the size standard template ABI GeneScan 50-500 and followed with similarity matrix cluster analysis based on the UPGMA (unweighted pair group method using arithmetic means) using ClusterVis software version 1.4.2 (Hepperle, 2002).

When selecting the allele size for the cluster analysis, in order to avoid background noise, only the peak which was 50% or higher than the highest peak, was selected for each sample. The allele size range was set to include all the selected alleles. Cluster analysis was performed and the dendrogram generated based on the presence and absence of the band. The generated UPGMA dendrogram was visualized and edited using the TreeMe software (Hepperle, 2004).

3.4.3 DNA Purification and DNA Sequencing

To verify that the polymorphisms were not because of indels in the regions flanking the cpSSRs, three PCR products of *G. tenuistipitata* from each locality for every primer-pair (i.e. GT1, GT2, GT3, GT4, GT5, GT6, GT7 and GT8) were randomly selected, purified and sequenced. The purification process was carried out using LaboPass^{TM} PCR Purification Kit (Seoul, Korea) according to the manufacturer's instructions.

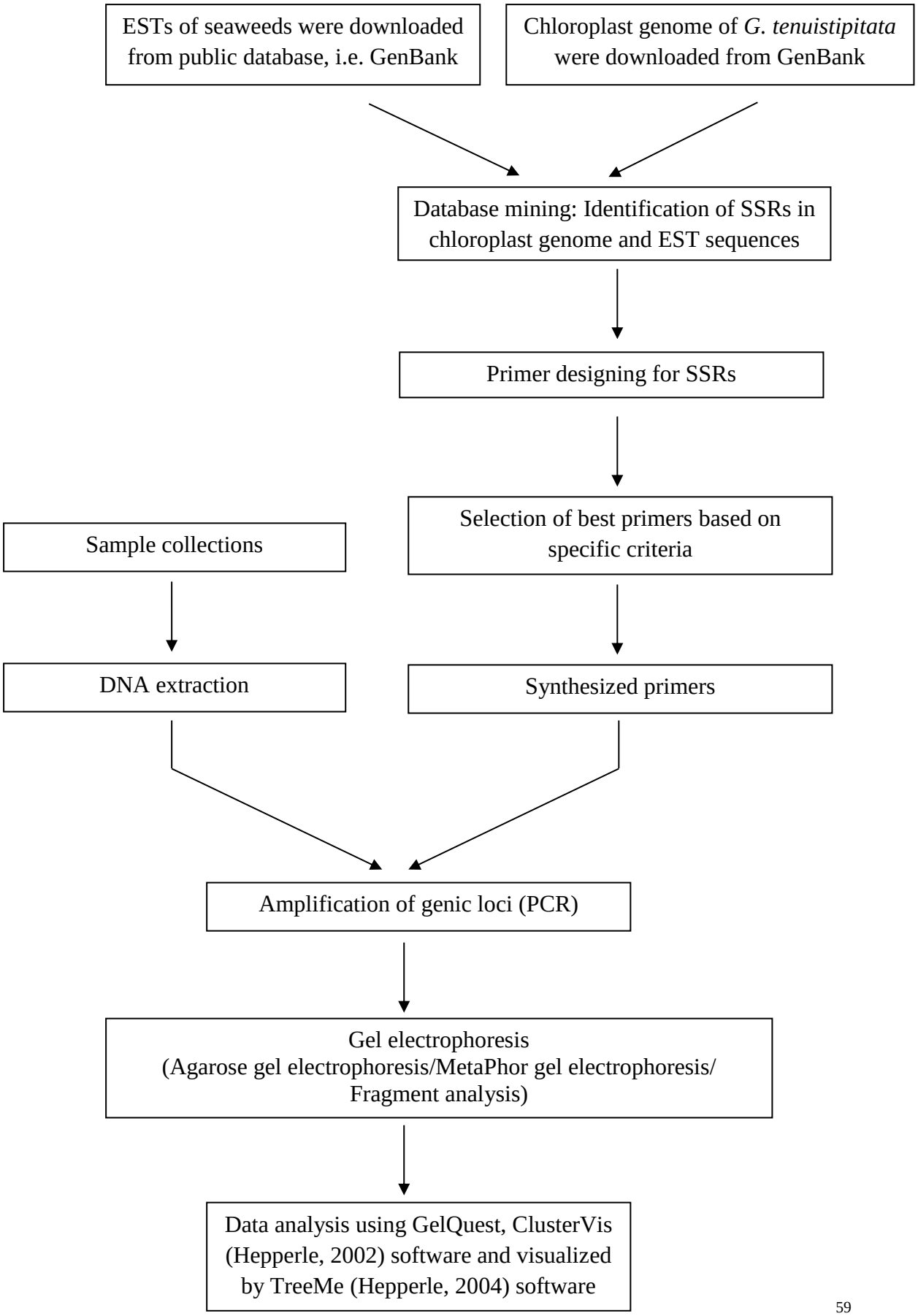
Firstly, 5 volumes of Buffer CB were added to 1 volume of the sample and mix. The mixture was then transferred to a spin column and centrifuged for 1 minute at

13,000 rpm. The spin column was removed from the tube and the flow-through was discarded from the collection tube. Spin column was reinserted into the collection tube.

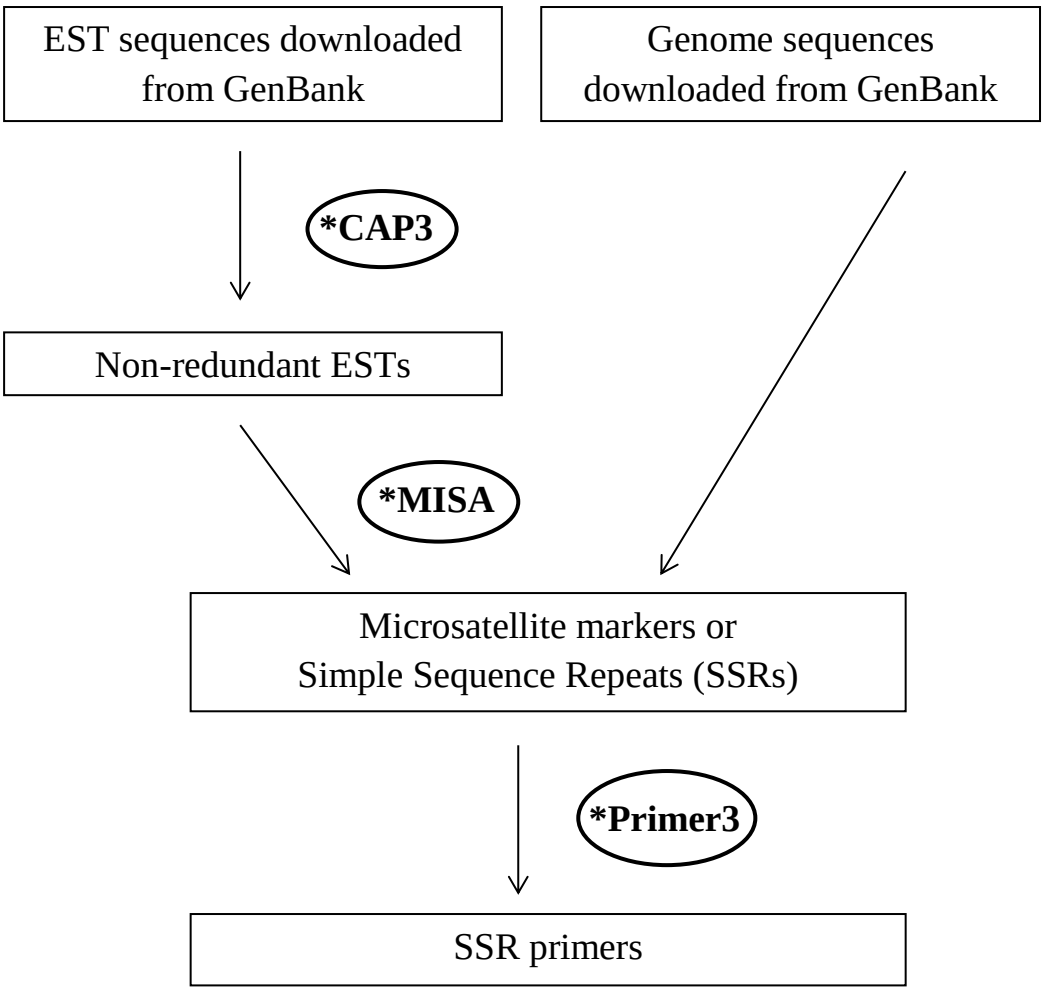
Next, 750 µL of Buffer NW was added to the spin column and centrifuged for 1 minute at 13,000 rpm. Then, the spin column was transferred to a new, sterile 1.5 mL tube. Finally, 50 µL of Buffer EB was added to the center of the membrane in the spin column, incubated for 5 minutes at room temperature and then centrifuged for 1 minute at 13,000 rpm to elute the purified DNA.

The purified DNA samples were then sent for automated DNA sequencing service (First Base Sdn. Bhd.). The sequencing data were then analyzed and edited using Chromas 1.45 (Technelysium Pty Ltd., Australia) and BioEdit 7.0.9.0 (Hall, 1999) software. Edited sequences were aligned by the CLUSTAL X program (Thompson et al., 1997).

The research methodology of this study can be summarized as shown in the flow chart below:



An overview of SSRPrimer pipeline



* **MISA**: SSR repeat finder, (http://pgrc.ipk-gatersleben.de/misa/)

*** Primer3**: PCR primer design program, (Rozen & Skaletsky, 2000) (http://www-genome.wi.mit.edu/genome_software/other/primer3.html)

*** CAP3:** DNA sequence assembly program (Huang & Madan, 1999)

How to execute Perl script in command prompt

1. cap3 <EST.file> -p95 > GC.cap3

This command will trace the similarity of the bases in EST sequences and assemble them in the same contig. The stringency level of 95% similarity is set and GC.cap3 is the name given for the output file. Two output file will be obtained: a) GC.contigs and b) GC.singleton. GC. contigs contained the EST sequences that have been assembled while GC.singlets contained those EST sequences that couldn't be grouped into contig.

Example:

```
>gi|327365314|gb|HS979443.1|HS979443 GT01017F05f Gracilaria
tenuistipitata tetrasporophyte EST Gracilaria tenuistipitata var. liui
cDNA clone Gt01017F05, mRNA sequence
GTCGACCACCATTGGGCAAGGACGTTGTTTCTTCCTATTTGCTCTTGGTCAGTAGAAACAATCGCAGGGA
AGATGCTGCGCACTGCAGCAAGAAGCGCTATGCGCTTCGGCGCTTCGAGATCGTTGCGCGGACTTGCTAC
GAAGACCGATCTTCTCCCTCCTCTCAAGGTTGCCGTCACAGGAGCTGCCGGACAAATCGGGTATGCTCTT
GTTATGCGCATCGCGAGTGGGGAAATGTTAGGTCCAGATCAGCTCTTAGAGTTAAGCCTCATCGAGACCG
AAGCGGGAATGTCTCCTTTGCGTGGGGTTATGATGGAACTGGAGGATGGGGCTTTCCCCTTGCTCTCTAA
GGTCACCGCGACTACGGATCTTCGCGAAGGCTTTGGAGATGCCTCATTTGCCATGCTTATCGGAGCCCAG
CCACGTACCAAAGGTATGGAACGCTCCGACTTGATGGCAGCGAATGCAAAGATCTTCGCGGAGCAAGGAA
AGGCCTTGAACGAATCTGCTGCGGAAAATGCCAAGGTGCTTGTGGTCGGTAATCCTGCCAACACGAACGC
GCTCATTGCATCCGAGAATGCTCCGGATATGTGGCCAGAACAGTTCATTGCAATGACTCGTCTCGACCAG
TCTCGAGCAAAAGCTTTATTGGCCGAAAAAGCCGACGTTCCTGTCAAGGACGTTGACCGCGTGATCATCT
GGGGTAATCATTCGGCCACCCAATATCCAGATATTTCCCATGCGCCCCATTGGGGGGAAATGGGCAAAGT
CAATTGTCACTGATGAAAAAGGGATTAAAGAGGGACTTCATTCCGCGGGGCCAAAAAACCAGGGGCCCAG
GGTTATCG

>gi|327365312|gb|HS979441.1|HS979441 GT01017F03f Gracilaria
tenuistipitata tetrasporophyte EST Gracilaria tenuistipitata var. liui
cDNA clone Gt01017F03, mRNA sequence
ATGGACCCCCATTGCGAAATGCTGGGAGGAAAGCCTGACTGCGTCATGCGTGCGGAGCCGTTAGAAGACT
TTGGAGGCGGTCACCCGGATCCCAACCTCACTTATGCCGCTCAATTGGTGGCCAAATGCGATCCAAAACA
AAACCACGACGCACCGGACATGGGCGCTGCTTCTGATGGAGATGGCGATCGAAACATGATTCTTGGAAGA
GGATTTTTTGTCACCCCGTCTGACTCGGTGGCTGTCATCGCTGCAAAGGCTGTTGATGCCATCCCGTACT
TCTGCAAAGGCTTGAAGGGAGTGGCACGAAGTATGCCCACTGCTGGCGCTCTTGACCGCGTTGCGAACGC
GATGGGAATCGAGTTGCACGAAGTACCTACTGGTTGGAAGTACTTTGGAAACCTCATGGATGCCGGTCGC
GCTCAAATTTGCGGTGAGGAGTCATTCGGAACTGGATCCGATCACGTCAGGGAAAAGGATGGAATCTTTG
CGATTTTGGCGTGGCTTTCTATCATTGCACATGAGAGTGAAGGTAAGGATAAGGTTGTGAGCATTGAGGA
TATTGTCACCTCTCATTGGAAGACATATGGCAGAAATTATTTCTCTAGATACGATTATGAGGAAGTTGAA
AGCGAGGCTGCCAACGCTATGATGGGCCATCTTGATGAATTACAGGGCAAAGATGAATGCCGAGAGAAGT
ACAGAGGATGAGCCAATGACTCTTGACCCTGAATTTGACACCCAGGGTGCTGGTGGGGGATAATTTTTCG
TACGTCGACCCGGTGGATGGGTCGGAGGCTATAGGGTCAGGCCCGTCAATTTGGGTTTCCGGATGGGAA
```

Some examples of *G. tenuistipitata* EST sequences downloaded from GenBank. The important information obtained including the name of organism (*G. tenuistipitata* var. *liui*), the accession number (e.g. HS979443) and also the DNA type of the organism (mRNA sequence).

2. cat GC.contigs GC.singlets > TUG_GC.fasta

This command will combine both sequences in GC.contigs and GC.singlets into another file named TUG_GC.fasta. TUG = Tandem Unique Gene (or more commonly known as unigenes) are unique sequences obtained after assembly of all the reads from one single library. After the combination from both sources, the dataset are ready to run.

Example:

```
>Contig1
TTTTTTTTTTTTTTTTTTGACCTTTCTAAATTTTTTAAATTACAATATATCTATTATCT
TTTCTTAATTACAATACATCCATCATCTAAGACTAAAAAAAAGTCTTGCGCCAACCACTA
TTCCTCAAAACATTCAAAATTTTACCTATACTCTTACAATATTAACAATTTTTTCCAATC
AAAAACCAATCCAGGATATACACAAATCCAGTGCATTTTTGTAGGACTTCTCAAACTTGT
CTATCAGCAGGAAGTTTGTAGTCAGTAGCAGAGGGAGGTCTATGCACCAAAAAGCCAGAA
ATCTAAACTTATCTATTGGCATCAAAACAAACCAATCTGATGTAGGCATCAGGAAAAGCT
TTTGTGCAAGCTTAGATCTCACTCAACACTTGAGAACCGTCCGTGCATCCAAACATAGGT
AACTTCCACATTGTCCAATATCTGTTATCTTAATAAGTAGAAGCTACTGGTCCCATTCTA
ATGCAGTTTTCATTACCAGCATAAGCTTGGTCCGAAGCTGCAAATTCAAGACAGGGAGTC
CAAGAGTTGGTCAAGATGTAGTCTACTTGTTTACTAATTTGTTCATCTGTCAAGGGAGGC
AAAAAGGAGAAAGTTTCAAACATTTTGTTATTGAAAGGCTACCAAACCATCATCTCTCTG
TTTTTGTTAGCTGTGCTAATTGCAAATCCTCTTTTGTGAGAAGAAATTTTGGGAGTTGCA
GCAGGTTTGGAACACCCCTTGCTCAAAACAACACTTTTGTTCATCATTGTTGCAGCCATG
ATTGTCTATATTAATATAGATTGAGTTAAAGGGAAC
>Contig2
TTTTTTTTTTTTTTTTTTTTAAAATATTACCATTTTTATTTACACAAAATCTCTCACTG
CATCTAACTTCTACGCATCATAACATACTCATAATCTTTACAAATAAATACTACATAAGT
ATATCTACCATCTAACAAGCTAAATAAAGTTTCTACACCAAGCATAATCTGAACACATTT
TTAATTTTTCTGCATTCAATTTCAAGTTTTCAGTTAGTTCAAATACCTGCAAATTTAAAA
ATTTTGATCAATTTTCTTGCTCGAATGTTCCAGTACCTTCATTAATAGCAGCAAAGAGGG
TAAAGGCAATGAAAGCCAAAAGGAGTGGTTCAGTATCACTGAGGGGTAAACCAGTTTCGA
TATCAAACTGTGCAAGCGGCCCTTTTCCCGTAACAGCCTCACCGATTAAAGCAAAAGCAA
AACCTAATTGTGCAACACGGCCAACAAACAACTCATTAGCTTTGGTGAATCCAAATCCGC
TAATTCCAAAAAAATTTGAGAGGCTGCAAAATGGAGATAGAAGGATCTTACAAAGAACCTT
TGGGTCTTTCTTCCTCCTCCTCATCGCTGGGCACAAACACCCCCTTTGCTGGCAATAAAG
CGGCCACCAAATTGAAAATTACGAGGCCGATGACAAGAGGCTCGGTGTCCGTGATGGGGA
TGCCAGTTTCTAGACCGAATTGAGCCAATGCACCTTTGCCCGTCAAAAGTTCTCCGATGA
```

Some of the contigs obtained from the previous commands. Bases in the EST sequences that have 95% similarity will be grouped in one contig.

```
>AC_Contig1
TGGATCCAAAGAATTCGGCACGAGGTTTTCCTACGCTCAATTTACAACAGTTCTTTCTCT
CCAACATGTCTTTGTGCAACTCTGCAAACTTTCAAGCTAACAGGACGTTTGCTCCCAAAC
AAGCAAACTCCTTGTTTCTACGCCAGGTTCGCGCCAAAACGGTGAACCCTAAAACAAGTT
TTAGCAGAACTGTTTACGCCACTGTGGCAGAGCCTGGAGTAGCTTCTGCGCCTGTGCCTG
AACCTGTGACTACTGGTCCTGCAACTATACCCCTTGTGTCTTAAGCTGCCACCAGGAAAA
TCAAATAAAAGAATGTTGCAGTTATCACTGGAGCTTCTTCTGGCTTGGGATTGAATGCTG
CCAAGGCACTAAGTATGAGAGGTGATTGGCATGTGGTTCTAGCTTGCAGGGATTTCTTGA
AAGCCGAACGCGCCGCTCAACAAATGGGCATGCTGAAAGACTCCTACACCACCATGCACC
TGGACCTATCCTCCCTCGAAAGCGTCCGTTAATTCGTGGACAACTTCCGACGAGCCAATC
TACCCCTAGACTGCTTCGTTGCAAACGCAGCTGTGTACCTCCCCACCGCAAAAGAACCCA
GATTTACCGCAGACGGATTCGAGCTAAGCGTCGGAACCAACCATTTGGGACATTTCTTGC
TATGCAACTTGTTGTTGGAAGATATGATTAAAACCAAGAAAGAAAAAAATTCGCAACCCA
GAATGATTATTGTTGGTTCCATCACGGGAAATACCAATACTGTTGCTGGAAATGTGCCTC
CCAAGGCTAACCTTGGAGACTTGCGGGGGTTGGCTGCAGGCTTGAACGGAAGCACTTCTG
TAATGATCGATGGCGGAGAGTTTGATGGTGCTAAGGCGTATAAGGATAGCAAAGTTTGTA
ATATGTTGACTATGCGATAATTCCACAAACGCTTCAGTGAAGCTACTGGCATCTCGTTCT
CTTCATTGTATCCAGGTTGCATTGCTACCACGGGTCTATTCAGAGAACATTACCAATTGT
TCCGCACATTGTTCCCTCCATTCCAAAAGTACATCACCAAAGGGTACGTGTCCGAGGAAG
ATGCCGGTACAAGGTTGGCTCAGGTGATGAGCGACCCCGAATTGAGCGCATCCGGTGTGT
ACTGGAGTTGGGATAATGCTACTGGTGCATTTGTAAATGAACTCTCTGAGGAAGCTAGCG
ATGATGCAAAGGGAGAGAAGTTGTGGGAGTTGAGTGAGGTCGCTGTCGGAATGAAATGAA
ACAAATTTATTTCAAAAAAGGCCGCACTTTTTTATCAATTATTATTGCAGAAACAATCAA
TAAAATTTAGAGTTTGTCGGTAAAATTTATATTGTTTACTAGCTAAATATTTCTTTAATT
TTGGCTATGCTGGTTAAATATTAAATAATAATGGCAAATATTTTTTTTGTGAGAAAGGGG
TCTCTGCTTATTGGCAGAGACTATGTTTTTGTTCTATTTGGTGCATGAAATCGAAAATTT
GTACAAAATTTTCAATACTTTTTTCTTGAATGTAAAAAGTACATTTTGGTTATCTTTGTT
```

The remaining of the EST sequences that couldn't be assembled in contig will be grouped as singlets.

3. myMICROSAT.pl <TUG_GC.fasta>

This command will execute the software MISA and Primer3. MISA will trace out the simple sequence repeats (SSRs) needed and exclude those SSRs with only one repeat and also Poly (A) and Poly (T) tails. The output file from MISA, <FileName>.fasta.p3in will be used by Primer3 for SSRs primers designed. A range of parameters, for example the primer melting temperature, primer length, GC content and the expected product size have set for the designed of primers. The output files from this execution will be <FileName>.fasta.p3out after Primer3 has done the job. Both the output files from MISA and Primer3 will be combined and desired SSRs with designed primers can be obtained in <FileName>.fasta.results.

Example:

	ID	SSR_nr	SSR_type	SSR	size	start	end	FORWARD_PRIMER1(5'-3')	Tm	size	REVERSE_PRIMER1(5'-3')	Tm	size	PRODU
2	Contig57	1	p3	(GGC)5	15	329	343	ACTGATATACCCTTCGGATT	52.784	20	AAGATACCCGGACTGTTG	53.3	18	329
3	Contig76	1	p2	(AG)8	16	622	637	TCTAGAGGCTGAATGAAAGA	52.754	20	ATGAATCTAATCTGGTGTGC	52.5	20	319
4	Contig78	1	p2	(GT)9	18	477	494	ACTTTAAGATGGTGACATGG	52.476	20	TGGAGAATAATTGAGCAAAC	53	20	340
5	Contig99	1	p3	(CAT)5	15	1003	1017	AAACATGTCATTCAAGTGGT	52.745	20	GGAAAAATAGTTGAGGAAGG	53.2	20	308
6	Contig133	1	p2	(AT)6	12	61	72	CATACGTTCACAGGGCTAC	54.062	19	ATTACGTACTCAGCTTCGTG	53.2	20	348
7	Contig134	1	p3	(TGC)5	15	47	61	CTCCACTAACACTTCAGTCG	53.367	20	TTGAGAAGAGAGTTCTGTCG	53.1	20	297
8	Contig140	1	p2	(CT)6	12	657	668	CACGGACTTATGGTAATTTG	53.254	20	CCATATGGTCCAGTACCTTT	54	20	341
9	Contig143	1	p2	(TA)8	16	887	902	CAGTCTTCTGAGTATCACGG	53.259	20	AACTGTTGAAATGCTACCTG	52.9	20	321
10	Contig205	1	p3	(CCG)5	15	253	267	GTTCCCAGCTAGACTTCAG	52.403	19	GAAGCGGAGAGGAATTAG	52.5	18	303
11	Contig315	1	p3	(CAA)9	27	846	872	AGGAATTTAACTTTTCCGAC	52.804	20	AACAAGTCTATCACAGCGAG	53	20	282
12	HS979395	1	p2	(AT)11	22	539	560	CGAAGTTAAGGACAAGTTCA	53.103	20	AAAGGTTTCATCAAAGCAC	52.8	19	327
13	HS979333	1	p3	(GAC)17	51	46	96	GGCATTTAATTCTAAACAGC	52.002	20	GGCACACATGGAGAATATC	53.6	19	345
14	HS979254	1	p3	(AAG)7	21	297	317	TTTCAAAATCTGTTTCAACC	52.846	20	AGCGTCGACAGTATCAGTAG	53.1	20	281
15	HS979254	2	p3	(AAG)6	18	477	494	GTCCCTTCCTTTGACTTATC	52.889	20	TGAGTTCCGAATAGGTGTAA	53.3	20	349
16	HS979215	1	p3	(GGC)5	15	322	336	ACTGATATACCCTTCGGATT	52.784	20	AAGATACCCGGACTGTTG	53.3	18	329
17	HS979174	1	p3	(CAA)12	36	83	118	CAATAACAACACCAACAACA	52.766	20	TGCCAATCTTCTTACTGACT	52.5	20	293
18	HS979170	1	p2	(CT)8	16	678	693	AGAATAGGTCGATCCTTCAT	52.826	20	TTTGAAATTGTTACGGTAGG	53	20	330
19	HS979102	1	p3	(TGC)5	15	47	61	TGCATTTCACTTACCTGG	52.797	18	TTGAGAAGAGAGTTCTGTCG	53.1	20	296
20	HS979073	1	p3	(CAA)12	36	408	443	CAATAACAACACCAACAACA	52.766	20	TGCCAATCTTCTTACTGACT	52.5	20	293
21	HS979028	1	p3	(CAA)12	36	396	431	CAATAACAACACCAACAACA	52.766	20	GGTGCAATTCTTCTTACTGA	53	20	295
22	HS979007	1	p3	(CCA)5	15	84	98	CTATACCATGTCTGGACCAC	52.7	20	CGACGATTACCTCTTGATTA	53	20	332
23	HS978982	1	p3	(ACC)5	15	85	99	TATTCCTAGATCCAAAAGCA	52.701	20	TCTTCGCTTTCAGTTTCTTA	53.1	20	285
24	HS978982	2	p3	(AGG)5	15	258	272	TATTCCTAGATCCAAAAGCA	52.701	20	TCTTCGCTTTCAGTTTCTTA	53.1	20	285
25	HS977954	1	p3	(GAC)5	15	480	494	GTGACAAGTCCAAGATTCTG	53.004	20	AGCCATTTGTTTACTTGTTG	53	20	305

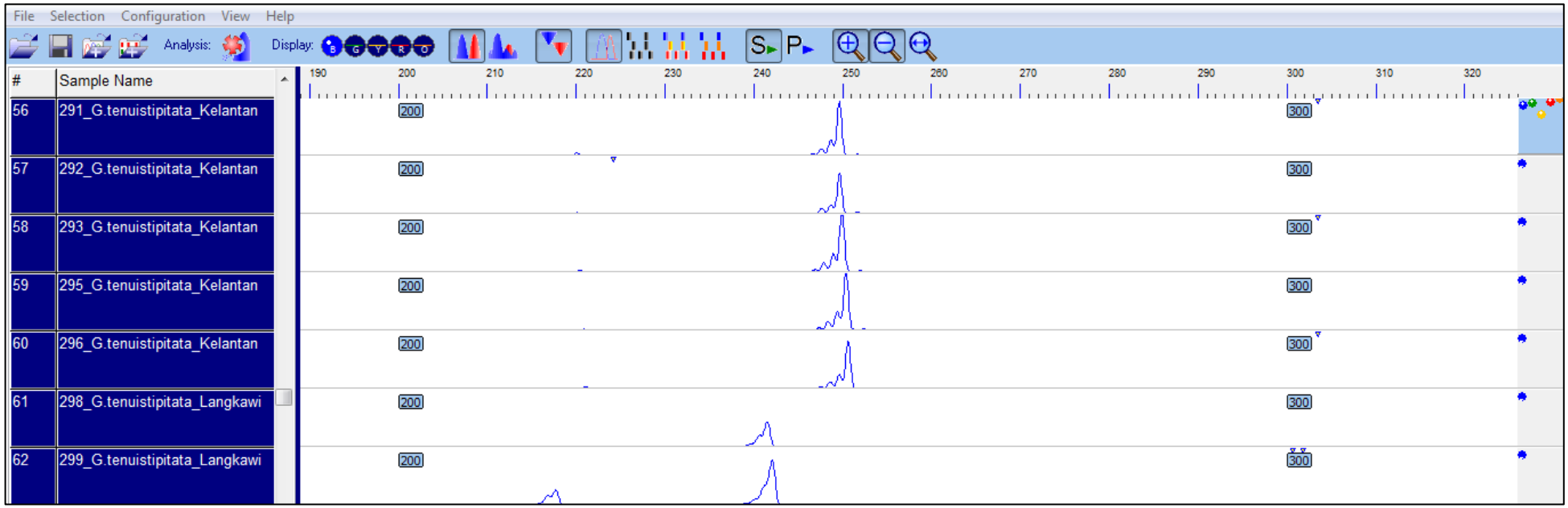
The results file obtained include the ID name, type of SSR (e.g. p3 means tri-nucleotide), the simple sequence repeats (SSRs), forward and reverse primers and the expected product size.

How to analyze SSR products using GelQuest, ClusterVis and TreeMe software

1. Analyzing SSR products using GelQuest software

SSR products from different localities were analyzed using GelQuest software (Hepperle, 2002) based on the size standard template ABI GeneScan 50-500. When selecting the allele size for the cluster analysis, in order to avoid background noise, only the peak which was 50% or higher than the highest peak, was selected for each sample. The allele size range was set to include all the selected alleles.

Example:

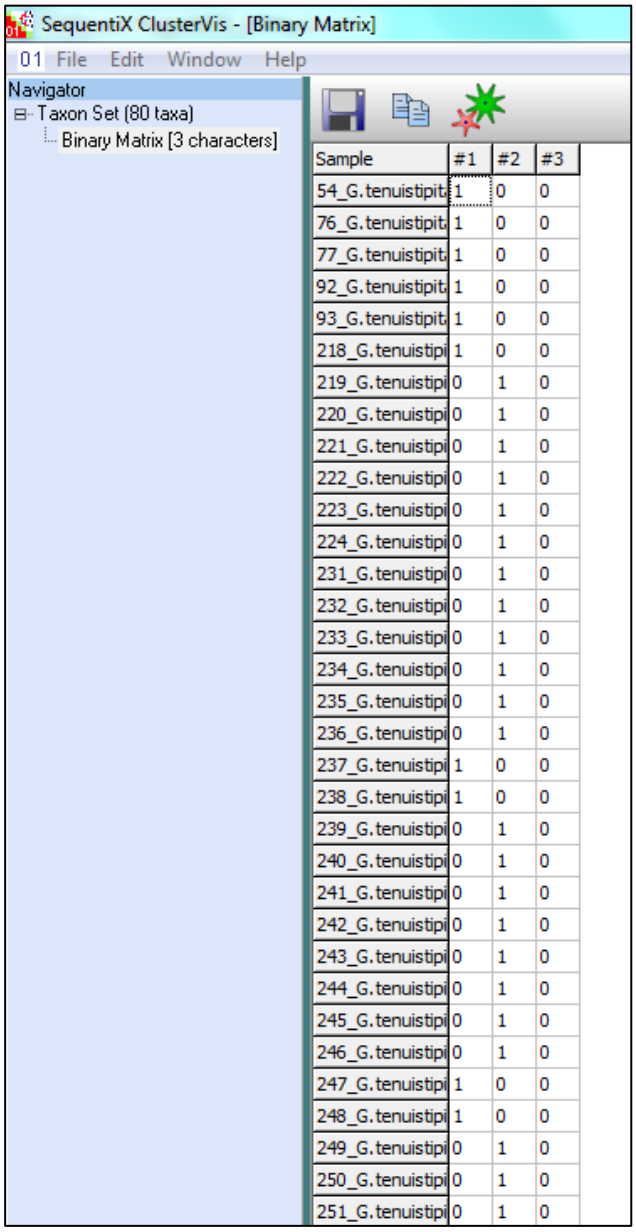


In GelQuest, the sample name (*G. tenuistipitata*) is listed on the left hand side while the respective product size is scored on the right hand size. We chose blue as the display color because we have labeled our forward primer with FAM for PCR amplification and fragment analysis.

2. SSR products scored by ClusterVis software using binary matrix method

SSR products amplified by each primer-pair were scored using a binary matrix method with “1” (presence) and “0” (absence) using ClusterVis software (Hepperle, 2002). Cluster analysis was performed and the dendrogram generated based on the presence and absence of the band.

Example:



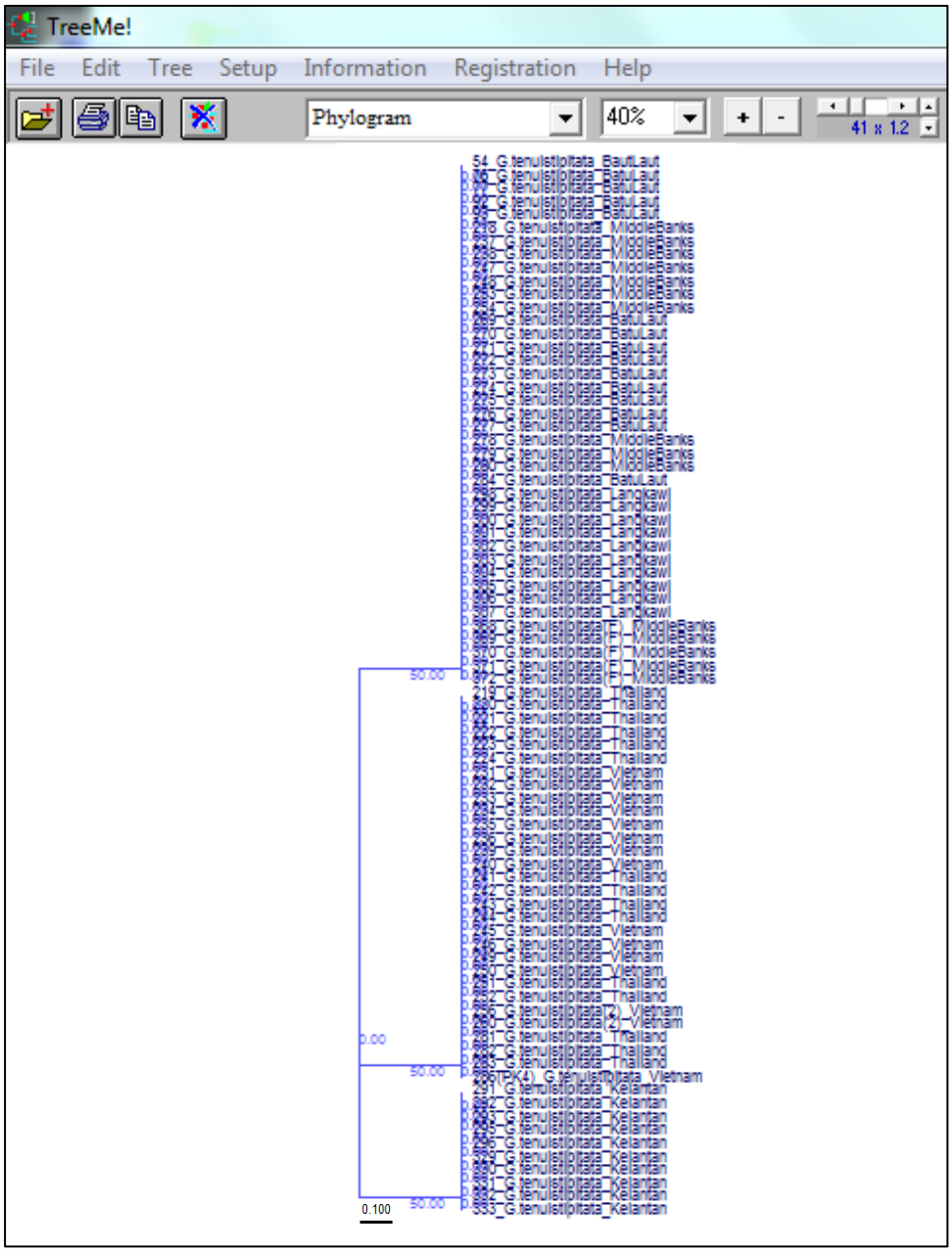
Sample	#1	#2	#3
54_G.tenuistipit	1	0	0
76_G.tenuistipit	1	0	0
77_G.tenuistipit	1	0	0
92_G.tenuistipit	1	0	0
93_G.tenuistipit	1	0	0
218_G.tenuistipi	1	0	0
219_G.tenuistipi	0	1	0
220_G.tenuistipi	0	1	0
221_G.tenuistipi	0	1	0
222_G.tenuistipi	0	1	0
223_G.tenuistipi	0	1	0
224_G.tenuistipi	0	1	0
231_G.tenuistipi	0	1	0
232_G.tenuistipi	0	1	0
233_G.tenuistipi	0	1	0
234_G.tenuistipi	0	1	0
235_G.tenuistipi	0	1	0
236_G.tenuistipi	0	1	0
237_G.tenuistipi	1	0	0
238_G.tenuistipi	1	0	0
239_G.tenuistipi	0	1	0
240_G.tenuistipi	0	1	0
241_G.tenuistipi	0	1	0
242_G.tenuistipi	0	1	0
243_G.tenuistipi	0	1	0
244_G.tenuistipi	0	1	0
245_G.tenuistipi	0	1	0
246_G.tenuistipi	0	1	0
247_G.tenuistipi	1	0	0
248_G.tenuistipi	1	0	0
249_G.tenuistipi	0	1	0
250_G.tenuistipi	0	1	0
251_G.tenuistipi	0	1	0

In ClusterVis, one can know the number of samples used in the analysis (E.g. 80 taxa) and different number of characters scored using the binary matrix method.

3. Editing of UPMGA dendrogram using TreeMe software

Example of similarity matrix cluster analysis based on the UPGMA (unweighted pair group method using arithmetic means) using ClusterVis software version 1.4.2 (Hepperle, 2002). The generated UPGMA dendrogram was visualized and edited using the TreeMe software (Hepperle, 2004).

Example:



In TreeMe, different types of topology such as Phylogram and Cladogram can be chose to view the generated dendrogram.