Chapter 2

Review of Literature
CHAPTER 2

REVIEW OF LITERATURE

2.1. TEA

2.1.1. Tea taxonomy

Tea is a woody evergreen plant of the genus *Camellia* in the Theaceae family. Cultivated tea consists of three species each with specific plant type viz. *Camellia sinensis* (China type), *Camellia assamica* (Assam type) and *Camellia assamica* sub sp. *lasiocalyx* (Cambod type) (Wight 1962). The classifications of these cultivars was initially proposed by Sealy (1958) based on leaf characteristics. This classification was revised by Wight (1962) on the basis of morphological characters such as leaf size, leaf shape, length of pistil and flower sizes (Takeda and Toyao 1980). Based on this, a bush with small leaves, resistant to cold are characterized as China type, while Assam type are tall with large leaves (Sealy 1958) and Cambod type are intermediate between the China and Assam type (Kingdom-Ward 1950; Robert *et al.* 1958). However, differences in caffeine, flavanols and amino acids contents in leaves of cultivated tea were also used for the classification (Nagata and Sakai 1984).

2.1.2. Tea origin

Southeast Asia is the original home for tea. According to Wight (1959), the primary centre of origin of tea is considered to be around the point of intersection of latitude 29°N and longitude 98°E near the source of the river of Irrawaddy, the point of
confluence where lands of Assam, North Myanmar, southwest China and Tibet meet. Secondary centres of origin are considered to be located in southeast China, Mizoram and Meghalaya (Kingdon-Ward 1950). The above areas are, therefore, considered to be the zone of origin and dispersion of the genus *Camellia* as a whole (Sealy 1958). However, presently tea cultivation is spread within the latitudinal range of 45°N–34°S.

### 2.1.3. Genome Size and Diversity

The genome size in terms of 4C DNA amount for *Camellia sinensis* is 15.61±1.06 pg where 1C DNA is equal to 3824 mega base pair (Mbp) and 1 pg = 980 Mbp (Hanson *et al.* 2001). Generally, tea chromosomes are small and tend to clump together due to ‘stickiness’. Tea is diploid \(2n = 2x = 30\); where basic chromosome number \(x = 15\) (Morinaga *et al.* 1929). Chromosome lengths range between 1.28μm to 3.44μm (Bezbaruah 1971). The \(r\) value (ratio of long arm to short arm) for all the 15 pairs of chromosomes ranges from 1.00 to 1.91. This consistency in diploid chromosome number suggests a monophyletic origin for all *Camellia* species. However, few higher ploidy levels, such as triploids, for example, TV-29, HS-10 A, UPASI-3, UPASI-20 \(2n = 3x = 45\), tetraploids \(2n = 4x = 60\), pentaploids \(2n = 5x = 75\) and aneuploids \(2n = 2x±1\) to 29) have also been identified (Singh 1980; Zhan *et al.* 1987).

Eighty two species of the genus *Camellia* had been described till 1958 (Sealy 1958). Currently more than 325 species have been described (Mondal 2002a). This may indicate genetic instability and high out-breeding nature of the genus. Presently, world-wide over 600 cultivated varieties are available, of which many have unique
traits such as high caffeine content, blister blight disease tolerant, etc. (Mondal 2009).

2.1.4. Tea clones and bi-clonal stocks

The first scientific attempt to select improved tea in North-East India was made by Stiefelhagen brothers in 1860 by establishing standard sources of tea seeds. Indigenous Assam tea was improved by following the technique of mass selection. The yield increased considerably, because of line breeding for desirable morphological features that are genetically linked with the characteristics of Assam tea. However, the seed grown plants were not uniform and unpredictable as their characters were governed by the genotype of their parents. Therefore, it was felt necessary to develop clonal cultivars in tea (Mondal 2009).

Secondly, with the increase of the region specific need of the industry, almost all tea producing countries have developed their clones or seed stocks. In India, Tea Research Association (TRA), Tocklai, Jorhat released the first three clones i.e. TV1, TV2 and TV3 in 1949. Development of bi-clonal seed stocks was also initiated in the late 1970s and early 1980s. Over the years, 32 clones and 14 bi-clonal seed stocks had been developed in TRA. In South India, the breeding work at UPASI (The United Planters' Association of Southern India), Tamil Nadu has resulted in the release of 28 clones and 5 bi-clonal seed stocks.

Clonal selection is an important and widely adopted method of tea plant improvement because of wide heterogeneity in the existing seedling population (Barua 1963; Shanmugarajah 1994). Clones are genetically uniform and give uniform yield and quality. Genetic stability of clones is vital for tea germplasm
preservation, breeding and production. However, clonal degeneration, a gradual loss of vigour and yield with age of a variety is also a well known phenomenon in vegetatively propagated crops (Forbes and Watson 1992). The increase in clonal plantation and continuous crossbreeding with small selection of superior trees can lead to a reduction in the gene pool of tea. So it is necessary to study the worldwide distributed genetic diversities of tea to overcome future problems associated with narrowness of genetic base of the modern tea cultivars.

2.1.5. Tea as beverage

The custom of brewing leaves of the tea plant for a beverage has its origin in China, with numerous records dating back more than 2,000 years (Tanaka and Taniguchi 2007). Based on the fermentation process, tea are divided broadly into three types: deeply fermented tea (black tea), semi-fermented tea (oolong tea) and non-fermented tea (green tea). Approximately 76-78% of the tea produced and consumed worldwide is black tea, 20-22% is green tea, and <2% is oolong tea (Mukhtar et al. 1994; Stoner et al. 1995; Bushman 1998; Cabrera et al. 2003). Green tea is mainly produced and consumed in East Asia and recently has gained attention as a healthy beverage (Higdon and Frei 2003; Crespy and Williamson 2004; Harada et al. 2005) in regions such as the United States and European countries, where green tea was not previously popular (Ujihara et al. 2011). Planted area totals about 2.3 million hectares, with China, India, Sri Lanka, Kenya, and Indonesia being the major producers. Major importers include the UK, Russia, Pakistan, the USA, Egypt, and Japan. Black tea accounts for about 70% of the world production, and green tea for most of the remainder (Tanaka and Taniguchi 2007).
2.1.6. Importance of Tea and uses

The economic importance of the genus *Camellia* is attributed primarily to tea. Tea was used initially as medicine, later as beverage and has a proven future potential of becoming an important industrial and pharmaceutical raw material. Many studies have authenticated many beneficial claims of tea, majority of which are attributed to its polyphenolic constituents. There are more than 700 constituent chemicals in tea leaves (Chen 1999). Among them, tea polyphenols and flavonoids have been reported to have strong antioxidant activity (Wiseman *et al.* 1977; Allemain 1999) which is responsible for most of the beneficial effects of tea. Among tea the polyphenols, and in particular catechins, have received immense attention (Cabrera *et al.* 2003). The major green tea catechins are epicatechin (EC), epigallocatechin (EGC), epicatechin gallate (ECG), and epigallocatechin gallate (EGCG). EGCG makes up about 40% of the total catechin content and is widely accepted as the major antioxidant ingredient in green tea and commercial green tea dietary supplements (GTDS) (Stoner *et al.* 1995; Cabrera *et al.* 2003).

Tea polyphenols inhibit the absorption of dietary fats and cholesterols (Chen *et al.* 2000). Several experimental evidences point to the potential of tea to protect against cancer. Especially green tea reduces the incidence of cancers of the stomach, small intestine, pancreas, lung, breast, skin, urinary bladder, prostrate, oesophagus and mouth (Vasisht *et al.* 2003). Green tea drinking also has been shown to possess anti-diabetic activity (Gomes *et al.* 1995), anti-arthritis activity (Tapiero *et al.* 2002), anti-plaque activity (Yu *et al.* 1995), antiviral activity (Okubo *et al.* 1995), anti-HIV (Human Immunodeficiency Virus) activity (Hashimoto *et al.* 1996),...
effect (Kwanashie et al. 1989) and anti-microbial activity (Hamilton-Miller 1995). It has also been reported that green tea polyphenols exhibit neuromuscular, anti-angiogenic, anti-hepatotoxic, anti-proliferative/apoptotic and immunomodulatory effects (Sueoka et al. 2001). Recent studies also showed consumption of green tea had inhibitory effects on cancerous cells of breast (Wu and Butler 2011), oesophagus and lungs (Yuan 2011) and ovaries (Lee et al. 2012).

Besides these properties, it is a dietary source of various important constituents among which flavanoids, amino acids, Vitamins (C, E and K), folic acid, Manganese, Potassium, Fluorides and polysaccharides which are important to human health (Chen 1999). Tea is gaining popularity as an important health drink in view of the above properties. The production of high quality tea with regional characteristics has nevertheless remained a highly profitable business. Tea plays a pivotal role in the national economy, a source of revenue and the job opportunities for almost all the producing countries.

2.2. DNA MARKERS AND THEIR APPLICATIONS IN TEA

In the recent years, with the remarkable advances in DNA based molecular markers techniques, several molecular markers, such as RFLP (Restriction Fragment of Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism), CAPS (Cleaved Amplified Polymorphic Sequence), ISSR (Inter Simple Sequence Repeat), SSR (Simple Sequence Repeat), EST-SSR (Expressed Sequence Tags based SSR), and ALPs (Amplicon Length Polymorphisms), etc. have been developed and widely applied in the tea plant.
Tea is highly heterogenous, highly cross-pollinated as well as freely out breeding with wild relatives in nature. Consequently a broad genetic variation exists in the cultivated tea gene pool. Some of which have valuable potential for tea industry in future. Therefore, many studies have been focused on estimation of genetic relationship and determination of genetic diversity of the tea germplasms using different molecular markers.

2.2.1. Restriction Fragment Length Polymorphism (RFLP)

Matsumoto et al. (1994) employed an RFLP analysis that used Phenylalanine ammonia-lyase sequence (PAL) cDNA as a DNA probe to determine the existence of PAL as a single gene and to study difference between Chinese and Japanese tea. Later, Matsumoto et al. (2002) used PAL marker as a tool for the classification and evaluation of tea resource. Matsumoto et al. (2004) were the first to assess genetic diversity with RFLP using phenylalanine ammonia-lyase sequence.

2.2.2. Random Amplified Polymorphic DNA (RAPD)

Since the first report on RAPD markers in Kenyan tea germplasm (Wachira et al. 1995), they have been used to analyze phylogenetic relationships amongst tea genotypes in different regions, including India (Mondal 2000), Japan (Tanaka and Yamaguchi 1996; Chen and Yamaguchi 2002, 2005), China (Chen 1998, Chen et al. 1998, 1999, 2002a, b), South Africa (Wright et al. 1996), Taiwan (Lai et al. 2001), South Korea (Kaundun et al. 2000; Kaundun and Park 2002), Portugal (Jorge et al. 2003) and Pakistan (Gul et al. 2007). Roy and Chakravorty (2010) revealed genetic diversity and relationships among tea (Camellia sinensis) cultivar by using RAPD
and ISSR based markers. Recently, Afridi et al. (2011) estimated genetic diversity in tea genotypes cultivated in Pakistan using RAPD markers.

2.2.3. Cleaved Amplified Polymorphic Sequence (CAPS)

Based on the sequence information of previously characterized tea genes, such as phenylalanine ammonia-lyase (PAL), chalcone synthase and dihydroflavonol 4-reductase, Kaundun and Matsumoto (2003) reported CAPS markers for the analysis of 52 tea samples of diverse origin. CAPS markers based on the above three genes were also studied for comparative analysis of Korean and Japanese green tea trees (Cho et al 2010). Expressed Sequence Tag (EST)-based Cleaved Amplified Polymorphic Sequence (CAPS) markers were also developed in tea plant for cultivar identification (Ujihara et al. 2011). Recently, genetic diversity among 30 tea cultivars in Sichuan province of China was investigated by PCR-RFLP analysis of cpDNA using 7 sets of chloroplast primers (Chen et al. 2012). CAPS marker for the two genes, phenylalanine ammonia-lyase (PAL) and chalcone synthase 2 (CHS2) were also developed which were related to the catechin content in tea (Elangbam 2012).

2.2.4. Inter Simple Sequence Repeats (ISSR)

Twenty-five diverse Indian tea were analyzed using ISSR markers. A dendogram was constructed using the unweighted pair group method analysis (UPGMA) method and revealed three distinct clusters of Cambod, Assam and China type, which concur with the known taxonomical classification of tea (Mondal 2002b). ISSR markers have also been used to analyze the genetic diversity of 27 Taiwanese tea cultivars (Lai et al. 2001), tea plants from Yunnan province of China (Liu et al. 2010) and
molecular identification of tea cultivars (Yao et al. 2005). The discrimination of tea germplasm at the inter-specific level was analyzed using ISSR markers (Liu B et al. 2012).

2.2.5 Simple Sequence Repeats (SSR)

Freeman et al. (2004) developed 15 SSR primers which revealed a great variability across a wide range of tea clones. Isolation and characterisation of 11 microsatellite loci from Camellia sinensis in Taiwan had been done using PCR- based isolation of microsatellite arrays (PIMA) (Hung et al. 2008). Recently, genetic diversity and relationship of clonal tea (Camellia sinensis) cultivars in China was revealed using SSR markers (Fang et al. 2012).

2.2.6. Expressed Sequence Tags based SSR (EST-SSR)

Zhao et al. (2007) generated 24 novel EST derived microsatellites from tea plant (Camellia sinensis). Microsatellite markers developed from genomic libraries can belong to the transcribed region or the non transcribed region of the genome. Over the past few years, various EST projects and studies (Sharma and Kumar 2005; Park et al. 2004; Chen et al. 2006) have generated publicly available EST sequence data in tea. Sharma et al. (2009) predicted 1,223 unigenes from 2,181 expressed sequence tags (ESTs) of tea (Camellia sinensis L.) which were available in publicly available sequence database. Ninety six primer pairs could be designed from 83.5% of SSR containing unigenes. Of these, 61 (63.5%) primer pairs were experimentally validated and used to investigate the genetic diversity among the 34 accessions of different Camellia spp. Seventy four novel polymorphic EST-SSR markers in tea plant were also identified and characterised by Ma et al. (2010). Recently, genetic
distribution of China tea germplasm was revealed by EST-SSR markers (Yao et al. 2012).

2.2.7. Amplified Fragment Length Polymorphism (AFLP)

Paul and co-workers (1997) were the first to employ AFLP markers to detect genetic diversity and differentiation of different Indian and Kenyan tea clones. Later, AFLP markers were also employed to study genetic variation amongst 49 different south Indian (Balasaravanan et al. 2003) and 27 Darjeeling tea cultivars (Mishra and Sen-Mandi 2001), and the phylogenetic relationship among tea cultivars from South Korea (Lee et al. 2003) and Japan (Wachira et al. 2001). Polymorphism and genetic relations among tea genotypes from turkey were also revealed by AFLP markers (Kafkas et al. 2009). Recently, Sharma et al. (2010) studied genetic diversity of commercially important tea germplasm in India using seven AFLP primer combinations.

2.3. GENETIC LINKAGE MAP

A mapping project is often started with the objective to detect linkage between one or several markers and a trait of interest. To achieve the ultimate goal of genetic improvement of tea, further efforts are required to construct a high density map using informative DNA markers and to locate the quantitative trait loci (QTL) of important agronomic traits, quality and resistance, showing bright prospects in tea breeding through marker assisted selection to integrate economically important traits onto the linkage map. Although the density of integrated genetic maps can be extremely high, it may still not be sufficient to tag a specific gene.
Many software packages were developed for linkage maps e.g. MAPMAKER (Lander et al. 1987; Lincoln et al. 1993), JOINMAP (Stam 1993; van Ooijen 2006), CRIMAP MAP (Donis-Keller et al. 1987; Green et al. 1989), MapManager QTX (Manly et al. 2001), GMENDEL (Liu and Knapp 1990), ANTMAP (Iwata and Ninomiya 2006), etc. Genetic maps have been constructed for several species using various markers like RFLP, RAPD, STS, microsatellites, proteins and recently AFLP markers. Over the years, the number of linkage maps based on AFLP has strongly increased. AFLPs are widely used owing to the large amount of detected polymorphism and the fact that sequence information is not required which is further facilitated by low cost per marker (Schlotterer 2004).

Advantages of AFLP

Since AFLP can generate many polymorphic bands without prior sequence knowledge, it is powerful technique for generating linkage maps. The technique has the following advantages:

- The markers produced are reliable and reproducible within and between the laboratories.
- The AFLP technique can be used for DNA samples of any origin or complexity. Small sequence variations can be detected using only small quantities of genomic DNA (0.05–0.5 mg).
- The capacity to reveal many polymorphic bands in an experiment (high multiplex ratio) is a major advantage of AFLP markers.
- They segregate in a Mendelian fashion and can be used for population genetic and QTL analyses.
The error level is very low as AFLP amplifications are performed under conditions of high selectivity (at high stringency) (Vos et al. 1995; Mueller and Wolfenbarger 1999).

Modified AFLP methods have also been developed for other species in recent years (Suazo and Hall 1999; Lindstedt et al. 2000; Ranamukhaarachchi et al. 2000; James et al. 2003; Gaafar et al. 2003; Kazachkova et al. 2004; Masumu et al. 2006; Giammanco et al. 2007; Esteve-Zarzoso et al. 2010; Sharma et al. 2011). Seonha et al. (2003) used a modified AFLP technique to study the phylogenetic relationship among 37 accessions of the genus *Camellia* consisting of Japanese tea, Korean tea and some *Camellia* species closely related to tea.

Although, AFLP is an important marker that detects high polymorphism but in large genomes ($10^8$-$10^{10}$bp) (Blears et al. 1998) due to generation of large number of bands the analyses were more cumbersome. A modification of conventional AFLP, TE-AFLP (Three Endonuclease-AFLP) method (van der Wurff et al. 2000) provides high discriminatory power and reduction in the number of bands. So, it would be suitable for tea genome which was estimated to be $4.0 \times 10^9$ bp.

Negi et al. (2005) did a comparative study of three different AFLP based methods and they concluded that TE-AFLP was the best technique for fingerprinting of tea. Recently, TE-AFLP has been successfully used for assessment of genetic diversity of biodiesel species *Pongamia pinnata* accessions (Sharma et al. 2011). It had also concluded that the easy scorability of TE-AFLP profiles is desired in studies requiring genotyping of large number of individuals across many gels.
2.3.1. Linkage maps in tea

A first linkage map for tea plant was constructed with RAPD markers by Tanaka (1996) and the markers related with theanine content, date of bud sprouting, resistance to anthracnose and tolerance to cold were detected (Tanaka 1996).

Another linkage map from the female parent, SFS150, was established with RAPD and AFLP markers (Hackett et al. 2000). There were 126 markers, covering 1349.7 cM, with an average distance of 11.7 cM between loci on the map.

An AFLP linkage map for tea plant was also constructed in China. The map of a female parent included 17 linkage groups and contained 208 markers, covering a total length of 2457.7 cM. The average distance between markers was 11.9 cM. A map from male parent included 16 linkage groups and located 200 markers, covering a total length of 2545.3 cM, and the average distance between markers was 12.8 cM (Huang et al. 2005).

A partial genetic map of backcross F1 generation between ‘Zhenong 129 (selected from the open pollination of ‘Fuding Dabaicha’×‘Yunnna Dayecha’) and ‘Fuding Dabaicha’ was also generated using RAPD and ISSR markers (Huang et al. 2006).

However, in previous studies, the number of individuals for mapping was limited and the density is not high enough to meet the demand of precise mapping.

Kamunya et al. (2010) used forty seven primers (21 RAPD primers, 20 AFLP and six SSR primer pairs) for complete genotyping of the *Camellia sinensis*. 260 informative markers were generated, out of which 100 markers that showed 1:1 segregation were used to construct a linkage map. The map contained 30 (19
maternal and 11 paternal) linkage groups that spanned 1,411.5 cM with mean interval of 14.1 cM between loci.

Recently, a high reference combined map was developed in a population of 54 F₁ clones derived from reciprocal crosses between ‘Sayamakaori’ and ‘Kana-Ck17’. The parental maps contain 441 SSRs, 7 CAPS, 2 STS and 674 RAPDs. The core map contains 15 linkage groups that covered a total length of 1218 cM (Taniguchi et al. 2012).

These constructed maps were still limited to locate QTLs linked with some important traits due to their low distribution of molecular markers.

2.4. SEQUENCING IN TEA

In plants, markers were sequenced for purposes of physical mapping. So, BLAST searches were performed to identify homologous sequences from the public databases. Identification of important markers would facilitate the gene mapping and marker aided selection in tea. The identification and validation of 61 new Unigene derived microsatellite (UGMS) markers from publicly available sequence database and 1,223 unigenes were predicted from 2,181 expressed sequence tags (ESTs) of tea. Out of 61 UGMS markers identified and validated, 36 of these UGMS markers correspond to the Arabidopsis protein sequence data with known functions (Sharma et al. 2009). This will have a major impact on genetic analysis, gene mapping and marker assisted breeding.

Recent advances in large-scale RNA sequencing (2.59 gigabase pairs of the transcriptome from poly(A)⁺ RNA of C. sinensis) was analysed using high
throughput Illumina RNA-seq approach by Shi et al. (2011) to generate large expression datasets for functional genomics. This analysis obtained 127,094 unigenes, which consisted of 788 contig clusters and 126,306 singletons. The category of secondary metabolism related genes covered 2.7% (427 of the functional genes) of the total genes identified, out of which thirteen unigenes related to theanine and flavonoid synthesis and were validated. Whole genome sequencing of tea had not been done but four Camellia cDNA libraries were available in Camellia ESTs in GenBank, including the EST sequences from the young root cDNA library of the tea plant (Shi and Wan 2009) (GenBank accession: GE652554.1-FE861258.1), two reported C. sinensis cDNA library respectively named subtractive cDNA library special for young leaves of the tea plant (Park et al. 2004) (GenBank accession: CV699876.1-CV699527.1) and the young leaf cDNA library of the tea plant (Chen et al. 2005) (GenBank accession: CV067174.1-CV013548.1), and another drought-stressed root SSH cDNA library of C. sinensis var. assamica (GenBank accession: GW316945.1-GT969202.1).

Therefore, BLAST searches were needed to perform alignments with whole genome shotgun contigs or non reductant nucleotides sequences of Arabidopsis and other dicots or alignment with the four Camellia cDNA libraries from the public databases. If markers, which are closely linked to desirable traits, are identified then it will facilitate the marker assisted early selection and shorten breeding procedures in tea. Thus, it will speed up the cultivar improvement programs in tea, a perennial crop which has a long juvenile period and is highly heterozygous.
Chapter 2 – Review of Literature

Tea plantation is often affected by many factors ranging from abiotic (drought, cold, etc.) to biotic factors (pest, fungal diseases, bacterial diseases). Understanding and identifying of not only a particular gene but those responsive genes that affect a particular trait are also needed to enhance breeding in tea. For example, drought tolerance is a polygenic trait which affects morphological, physiological, biochemical and molecular processes of plants (Zhu 2002). A set of drought responsive genes and their pattern of expression were identified under controlled condition in tea (Mishra and Sen-Mandi 2001).

So far, one transgenic tea plant obtained by the Agrobacterium-mediated transformation of somatic embryos was reported (Mondal et al. 2001). Meanwhile, several research projects using both particle bombardment and Agrobacterium-mediated or combined transformation methods were also reported (Luo and Liang 2000; Zhao et al. 2001; Wu et al. 2003, 2005). With advances in technology, if novel genes responsible for drought tolerance or disease resistance are identified, it could be introduced to other drought stress or disease susceptible plants. By this way, the chances of survival of important plant species against biotic and abiotic stresses would be increased and thus the yield of this important cash crop would be increased.

Thus, to achieve the ultimate goal of genetic improvement of tea, further efforts are required to construct a high density map and to locate the quantitative trait loci (QTL) and other important agronomic traits. Although the density of integrated genetic maps can be extremely high, it may still not be sufficient to tag a specific gene. Maps keep improving but to obtain a “complete map” that includes the
sequences and location of all genes of an organism, the work is quite vast and there is still a long way to go.