2.1. POLYPLOIDY

Polyploidy is the process of genome doubling producing organisms with multiple sets of chromosomes (Moshgabadi and Adams, 2010). Polyploid organisms contain a multiple or combination of the chromosome sets found in the same or a closely related diploid species. Polyploidy can arise from spontaneous somatic chromosome duplication or as a result of non-disjunction of the homologous chromosomes during meiosis resulting in diploid gametes (Ramsey and Schemske, 2002). It can also be artificially induced by treatment with drugs, such as colchicine, which inhibits cell division.

The haploid number (n) is the number of chromosomes that occurs in gametes. This represents half the chromosome number in somatic cells, which is designated 2n. A diploid species has 2n = 2x in its somatic cells, and n = x in its gametes. Some species have a higher ploidy, for example an autotetraploid (four basic sets of chromosomes) which has somatic cells with 2n = 4x and gametes with n = 2x. Within the five kingdom classification system, polyploids have naturally risen in three kingdoms: plantae, protista and animalia (Baatout, 1999; Storchova and Pellman, 2004). In the plant kingdom polyploids are common in angiosperms and ferns, (Schuettpelz et al., 2008) but are rare in gymnosperms (Briggs and Walters, 1997). Furthermore, genome duplication or polyploidy is believed to have occurred during the evolution of 30 – 70 percent of angiosperms (Masterson, 1994; Nadler 2009).

2.2. POLYPLOID BREEDING

Initially polyploidy acquired attraction due to their unique cytogenetics and their reproductive isolation (Blakeslee, 1921; Jørgensen, 1928). Soon it was recognized that polyploids also show distinctive phenotypic traits and hybrid vigor useful for agriculture (Muentzing, 1951; Randolph and Hand, 1940;
Many crop species, including wheat, maize, sugar cane, coffee, cotton and tobacco, are polyploid, either through intentional hybridization and selective breeding (e.g. some blueberry cultivars) or as a result of a more ancient polyploidization event (e.g. maize) (Ramsey and Schemske, 2002). The importance of polyploidy the fields of cytogenetics, physiology, breeding, cytotaxonomy and biogeography in conjunction with new possibilities put forth by various molecular techniques has all spurred a resurgence of interest in issues of origin and establishment of lineages (Mable, 2003). Polyploidy seems to be favored in long lived/perennial plants possessing various vegetative means of propagation and in those with frequent occurrences of natural inter-specific hybridizations (Hilu, 1993). The polyploid species formed independently from heterozygous diploid progenitors may provide important source of genetic variation (Soltis and Soltis, 2000). Polyploids often have improved horticultural or agronomic traits such as larger fruit, thicker leaves and robust stems (Kehr, 1996), so plant breeders sometimes favor polyploids in breeding programs (Nadler, 2009).

Vitamin A activity in tetraploid maize is increased as compared to the diploid (Randolf and Hand, 1940) is about 40% more than diploid. Similarly, the vitamin C content of vegetables and fruits has been known to increase following chromosome doubling. The nicotine content of tetraploid tobacco is about 18-33% higher than in diploid species (Briggs and Knowles, 1967).

### 2.3. CYTOLOGY OF GENUS CAMELLIA

Cytology of the genus *Camellia* was studied since the early 1970s and reveals many interesting features. Chromosome number has been established for the most available taxa of Camellia including tea (Bezbaruah, 1971; Kondo, 1975, 1986). From extensive cytological studies, it is suggested that tea is a diploid plant \((2n = 2X = 30; \text{ basic chromosome number } X=15)\) (Bezbaruah, 1971). However few higher ploidy levels such as triploids, tetraploids, pentaploids and aneuploids have also been identified (Singh, 1980). In karyotype analysis, grouping by chromosome size was difficult in the *Camellia* taxa since the chromosome vary continuously from the largest to the smallest. Furthermore,
even in the best preparation, homologous chromosomes did not appear identical (Kondo, 1977). Relatively little intraspecific karyotypic variation had been observed in the cultivated species of *Camellia* studied (Kondo, 1979). Satellite chromosomes in karyotype within many accessions of certain *Camellia* species are morphologically and quantitatively variable. Thus karyotypes including characteristics of sat-chromosomes are not of taxonomic significance for *Camellia*. It was shown by Kondo and Parks (1979) that the C-banding method can be applied to the somatic mid-metaphase chromosomes in *Camellia* taxa. This differentially stains bands in somatic mid-metaphase chromosomes and permits the identification of individual chromosomes. Thus it is possible to identify the homologous pairs of chromosomes more precisely and perhaps even to measure chromosome divergence between different clones within the same species with same or similar karyotypes (Kondo and Parks, 1980).

### 2.4. SELECTION CRITERIA OF PROMISING GENOTYPES

Traditionally, yield and morpho-chemical characteristics are the basis of selection in tea breeding (Mandal, 2008). The area of the plucking surface of a plucked bush, leaf size, area of the mature leaves in the maintenance canopy, plucking point density and pruning weight are considered as important criteria for the assessment of yield capacity. To screen the genotypes during its early age, 2-3 year old seed generated plants were used to study the leaf area, height and diameter of the stem (Barua, 1989).

For selection of genotypes with promising organoliptic characteristics, morphological or chemical parameters showing positive correlation with the quality of made tea are used (Das et al., 2012a) One of such characteristics that have been used for a long time is the hairiness of the young tea leaves which shows a positive correlation with quality of orthodox tea (Wight, 1958; Wight and Barua 1954; Wight and Gilchrist 1963). Colour of the green leaf is another characteristic taken as an index for quality of genotypes at the field. In broad leaved *C. assamica*, genotypes with light green leaves are considered as good source for CTC manufacturing (Venkataramani and Padmanabhan, 1964). Tea
bushes with yellow-green shoots show significant correlation with quality of made tea (Harler, 1964). The characteristic flavonoids of tea which play an important role in determining organoliptic and pharmaceutical properties of tea is also important for the pigmentation of the tea leaves (Nijveldt et al., 2001). Therefore, such a correlation may provide interesting relationship with flavonoid content and leaf colour of the genotypes.

When young tea shoots are exposed to chloroform, the green color of the leaf changes to brown. The rapidity of changing colour indicates the fermentation ability of the enzyme polyphenols oxidase (PPO) and is an important factor determining the quality of black tea. Therefore, this criterion is given significant weightage for selection of promising genotypes for black tea production (Toyao et al., 1971). Recently, extensive biochemical characterization of made tea revealed several important biochemical checkpoints, presence of which in the green tea leaves can be taken as important selection criteria of promising tea cultivars. It is found that there is a positive and significant correlation between total quality score (TQS) and green tea leaf chemical components like caffeine, nitrogen, amino acids and catechins (polyphenols) (Liang et al., 2003). These characters are influenced by environmental factors and may show a continuous variation with a high degree of plasticity. Yet, they can serve as good markers for quality, provided the selection procedure is repeated for a number of times in different seasonal variations. Molecular marker techniques are being applied by several institutes for identifying DNA regions that may directly or indirectly affect the quality of tea genotypes (Mondal, 2008). Such approaches will provide the breeders with DNA based selection tools for developing potential genotypes.

2.5. QUALITY OF BLACK TEA

Black tea is the most important form of made tea consumed across the world. Black tea quality depends mainly on the components and colour of the tea infusions. Tea prices vary greatly, depending on the quality which has traditionally been assessed by a tea taster who has developed a language of his own to describe various quality attributes of a tea infusion. Black tea quality is
a complex quantitative multigenic character determined by quantitative variation of several secondary metabolites (Mahanta and Barua, 1992). There are several parameters that are used for assessing quality of black tea (Owuor, 1992; Owuor and Obanda, 2001). Biochemical tests for predicting black tea quality from green leaf composition include carotenoid and chlorophyll content (Taylor et al., 1992), caffeine content (Millin et al., 1969) and catechin (flavanol) composition (Hilton and Palmer-Jones, 1973; Obanda et al., 1997).

Mainly, four flavanol derivatives are found in tea: (−)-epicatechin (EC), (−)-epigallocatechin (EGC), EC gallate (ECG), and EGC gallate (EGCG) (Punyasiri et al., 2004; Vallsa et al., 2009). During the black tea fermentation, an enzymatic oxidation of tea polyphenols takes place, leading to formation of a series of coloured chemical compounds, such as theaflavins (TFs) and thearubigins (TRs), which are responsible for the characteristic taste of the black tea liquor (Liang et al., 2003). Regression analysis of tasters’ preferences for black teas against green tea leaf chemical components showed positive and significant correlations for epicatechin gallate (ECG), epigallocatechin gallate (EGCG) and caffeine (Obanda et al., 1997). It was showed that fresh leaf epicatechin (EC) and ECG content together correlated well with the total score of black tea and the total theaflavin content of black tea correlated significantly with the value of the tea (Liang et al., 2003).

Caffeine is a purine alkaloid and biosynthetically derived from xanthosine (Ashihara et al., 2008). Purine alkaloids contain N-alkyl derivatives such as caffeine, theophylline and theobromine that have been widely used due to their pharmacological properties. It is observed that caffeine contributes towards the briskness of black tea (Deb and Ullah, 1968). Caffeine complexes with the polyphenols in tea, mainly theaflavins (Roberts and Smith, 1961; Roberts, 1962; Collier et al., 1972) and modifies the taste characteristics of both caffeine and thea-flavins (Millin et al., 1969) making the tea taste brisker (Sanderson et al., 1976). Besides quality, pharmaceutical benefits of black tea are also dependent on the components of the tea infusions. The biological benefits of tea are due to the strong antioxidant and anti-angiogenic activity of catechins.
They can modulate carcinogen metabolism and together with caffeine are potential inhibitor of cell proliferation (Demeule et al., 2002; Kazi et al., 2002). Researchers have endeavoured to develop chemical and physical methods for identifying tea quality. Capillary electrophoresis, electronic tongue and lipid membrane taste sensor have been applied to tea quality estimation (Horie and Kohata, 1998; Wright, 2005). Unfortunately, tea breeding was never done in India considering all of these traits, neither for diploids, nor for polyploids.

2.6. BIOSYNTHETIC PATHWAYS RELATED TO BLACK TEA QUALITY

Catechin and caffeine biosynthesis are considered to be the most important pathways related to the quality of *Camellia* genotypes (Wright, 2005). Both the pathways are regulated in a coordinated way by a series of important genes (Gohain et al., 2005)

2.6.1. CATECHIN PATHWAY

Vascular plants can divert a large amount of carbon from aromatic amino acid metabolism into biosynthesis of natural products (Winkel and Burbulis, 1999). Diversion of carbon flow towards flavonoid biosynthesis takes place either by suppression of lignin biosynthesis or by activating key enzymes of the phenylpropanoid metabolism and flavonoid metabolism (Gohain et al., 2005).

A. PHENYLPROPANOID METABOLISM

Catechins are synthesized via the flavonoid biosynthetic pathway which starts with phenylalanine as the precursor (Figure 1.1). The reactions of phenylpropanoid metabolism involve three enzymes, phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumarate coenzyme (4CL) (Hahlbrock and Scheel, 1989). As metabolic flux for the operation of the flavonoid pathway is maintained through the activities of PAL and C4H, catechin biosynthesis in tea is critically dependent on the products of these enzymes (Singh et al., 2008).
Phenylalanine ammonia-lyase (PAL)

The enzyme phenylalanine ammonia lyase (PAL) catalyses the deamination of phenylalanine to cinnamic acid, which is further, modified by hydroxylase and O-methyltransferases (Dixon et al., 1999). Since PAL resides at a metabolically important position, linking the phenylpropanoid secondary pathway to primary metabolism, the regulation of overall flux into phenylpropanoid metabolism has been suggested to be modulated by PAL as a rate-limiting enzyme (Ro and Douglas, 2004).

Cinnamate-4-hydroxylase (C4H)

Transcinnamic acid is hydroxylated at the para-position by a cytochrome P450 monooxygenase enzyme, cinnamate-4-hydroxylase (C4H), in conjunction with NADPH:cytochrome P450 reductase (CPR) (Ro and Douglas, 2004). C4H catalyzes the hydroxylation of trans-cinnamate to trans-4-coumarate in the initial oxygenation step of phenylpropanoid biosynthesis, which introduces the 4’-hydroxyl that is common to most flavonoids (Davies and Schwinn, 2005).

4-Coumarate:CoA ligase (4CL)

4-Coumarate: CoA ligase (4CL; EC 6.2.1.12) catalyzes the activation of 4-coumarate and a few related substrates to the respective CoA esters and thus channels the common, phenylalanine-derived building block into the otherwise widely distinct branches of general phenylpropanoid metabolism (Hamberger and Hahlbrock, 2004). This results various classes of secondary compounds with essential functions like production of lignin for structural support, flavones and flavonols for UV protection, anthocyanins as well as chalcones as pigments for the attraction of pollinators and seed distributors. Therefore, it is considered as a key enzyme for flavonoid biosynthesis (Davies and Schwinn, 2005).
Figure 2.1: (A) Generalized view of phenylpropanoid pathway and the enzymes involved in formation of 4-oumaryl – Co-A, from which diverse group of secondary metabolites are synthesized.

(B) Possible biosynthetic pathways of flavan-3-ols in *Camellia sinensis* leaves.
B. FLAVONOID METABOLISM

*Chalcone Synthase (CHS)*

The flavonoid metabolism starts with the formation of a carbon 15 backbone by CHS. Chalcones are then converted to a range of other flavonoids in a pathway of intersecting branches, with intermediate compounds involved in the formation of more than one type of end product. CHS carries out a series of sequential decarboxylation and condensation reactions, using 4-coumaroyl-CoA and three molecules of malonyl-CoA, to produce a polyketide intermediate that undergoes cyclization and aromatization reactions to form the resultant chalcone structure naringenin chalcone (Davies and Schwinn, 2005).

*Chalcone Isomerase (CHI)*

Subsequently, chalcone isomerase (EC 5.5.1.6) catalyzes the stereo-specific cyclization of chalcones into naringenin (5, 7, 4′-trihydroxyflavanone) via an acid base catalysis mechanism (Jez et al., 2000). CHI catalyzes the intramolecular cyclization of bicyclic chalcones into tricyclic (S)-flavanones. The activity of CHI is essential for the biosynthesis of flavanone precursors and phenylpropanoid plant defense compounds (Jez and Noel, 2002).

*Flavanone – 3 – Hydroxylase (F3H)*

Naringenin can be converted by flavonoid 3′-hydroxylase (EC 1.14.13.21) and flavonoid 3′,5′-hydroxylase (EC 1.14.13.88) to produce eriodictyol (5,7,3′,4′-tetrahydroxyflavanone) and Dihydrotricetin (5,7,3′,4′,5′-pentahydroxyflavanone) respectively. Naringenin, eriodictyol and dihydrotricetin are the flavanones. Almost all flavonoid compounds can be derived from these flavanones (Shi et al., 2011).

*Flavone synthase (FLS, EC 1.14.11.2)*

FLS catalyzes the conversion of flavanones to flavones, but flavanone 3-hydroxylase (EC 1.14.11.9) can convert these flavanones to dihydroflavonols.

*Dihydroflavonol 4-reductase (DFR, EC 1.1.1.219)*

Following this reaction, the divergent conversions of dihydroflavonols include production of flavan-3, 4-diols (leucoanthocyanidin) catalyzed by DFR. DFR
catalyzes the stereo specific reduction of dihydroflavonols to leucoanthocyanidins (flavan-3, 4-diol) using NADPH as a cofactor. DFR belongs to the single-domain-reductase/epimerase/dehydrogenase (RED) protein family, which has also been termed the short chain dehydrogenase/reductase (SDR) superfamily. This contains other flavonoid biosynthetic enzymes, in particular the anthocyanidin reductase (ANR), leucoanthocyanidin reductase (LAR) and isoflavone reductase (IFR) (Labesse et al., 1994; Johnson et al., 2001).

**Anthocyanidin Synthase (ANS, EC 1.14.11.19)**

The role of ANS in the biosynthetic pathway is to catalyze reduction of the leucoanthocyanidins to the corresponding anthocyanidins. Formation of (-)-epiflavon 3-ols (e.g. (-)-epicatechin, (-)-epigallocatechin) can be achieved by a two-step reaction on leucoanthocyanidin by anthocyanidin synthase and anthocyanidin reductase.

**Anthocyanidin Reductase (ANR, EC 1.3.1.77)**

ANR uses anthocyanidins such as cyanidin to synthesize epicatechin in the presence of NADH or NADPH (Xie et al., 2003). The first evidence for the role of ANR in epicatechin biosynthesis was provided by the analysis of the *BANYULS* genes of *Arabidopsis thaliana* and Medicago (Xie et al., 2004). It was found that tea genotypes with higher concentrations of epicatechins had higher expression of *ANR* than genotypes with lower concentrations of epicatechin.

### 2.6.2. CAFFEINE PATHWAY

Caffeine is a purine alkaloid and its biosynthesis starts by the conversion of “Phosphoribosyl pyrophosphate” (PRPP) to “Inosine monophosphate” (IMP). PRPP is formed from ribose 5-phosphate by the enzyme ribose-phosphate diphosphokinase.

**Inosine Monophosphate Dehydrogenase (IMPDH)**

IMPDH is very important for oxidation of IMP to xanthosine monophosphate (XMP). Li et al. (2008) indicate that caffeine synthesis should be reduced and
that free purine nucleotides, including IMP, will accumulate if IMPDH activity is blocked.

**Guanine Monophosphate Synthetase (GMPS)**

GMPS is involved in the de novo synthesis of guanine nucleotides and catalyzes the amination of XMP to Guanine Monophosphate (GMP). In the de novo synthesis of purine nucleotides, IMP is the branch point metabolite at which the pathway diverges to the synthesis of either guanine or adenine nucleotides.

**Guanine Monophosphate Reductase (GMPR)**

This gene encodes an enzyme that catalyzes the irreversible and NADPH-dependent reductive deamination of GMP to IMP. The protein also functions in the re-utilization of free intracellular bases and purine nucleosides.

**5'-nucleotidase (5'-Nase)**

The enzyme encoded by 5'-Nase converts Xanthosine 5'-phosphate (XMP) to xanthosine by releasing an inorganic phosphate molecule.

**7-Methylxanthosine Synthase (7-XMT)**

The first step in the biosynthetic pathway of caffeine from xanthosine is the conversion of xanthosine to 7-methylxanthosine. This reaction is catalysed by 7-methylxanthosine synthase and the genes encoding 7-methylxanthosine synthase were isolated from *C. arabica* (Mizuno et al., 2003, 2005).

**N-Methyl Nucleosidase (N-Me Nase)**

The second step of caffeine biosynthesis involves a nucleosidase which catalyses the hydrolysis of 7-methylxanthosine. Although N-methylnucleosidase (EC 3.2.2.25) was partially purified from tea leaves, isolation of the native enzyme as well as DNA encoding the enzyme has not yet been achieved (Ashihara et al., 2008).
Theobromine Synthase (Monomethylxanthine N-Methyltransferase)/Caffeine Synthase (dimethylxanthine N-methyltransferase)

The last two steps of caffeine synthesis are catalysed by another SAM-dependent N-methyltransferase called caffeine synthase (EC 2.1.1.160) which catalyzes conversion of 7- methylxanthine to caffeine via theobromine (Ashihara and Suzuki, 2004; Ashihara et al., 2010). The gene encoding caffeine synthase was cloned from young tea leaves by Kato et al. (2000).

2.6. REAL TIME PCR ANALYSIS OF mRNA

Among the many methods currently available for quantifying mRNA transcript abundance, reverse transcription-polymerase chain reaction (RT-PCR) has proved to be the most sensitive. Recently, several protocols for real-time relative RT-PCR using the reporter dye SYBR Green I have appeared in the literature. In these methods, sample and control mRNA abundance is quantified relative to an internal reference RNA whose abundance is known not to change under the differing experimental conditions (Marino et al., 2003). Therefore, RT-PCR has proven to be a powerful method to quantify gene expression. Two different methods of analyzing data from real-time, quantitative PCR experiments exist: absolute quantification and relative quantification. Absolute quantification determines the input copy number of the transcript of interest, usually by relating the PCR signal to a standard curve. Relative quantification describes the change in expression of the target gene relative to some reference group such as an untreated control or a sample at time zero in a time-course study. Quantifying the relative changes in gene expression using real-time PCR requires certain equations, assumptions, and the testing of these assumptions to properly analyze the data. The $2^{\Delta\Delta CT}$ method is used to calculate relative changes in gene expression (Livak and Schmittgen, 2001).

2.7. WHOLE GENOME TRANSCRIPT PROFILING

Several methods are available to profile gene expression in response to different growth condition or different physiological status of an organism. The
expression profile for each sample may be estimated separately and then compared by methods that depend on specific hybridization of probes to DNA microarrays (Lipshutz et al., 1999) or on the counting of tags or signatures of DNA fragments (Brenner et al., 2000; Velculescu et al., 1995). Differences in gene expression between two samples can be directly compared by methods such as differential display (Liang and Pardee, 1992), two-colour microarray hybridization (Brown and Botstein, 1999), subtractive cloning techniques (Sagerstrom et al., 1997), and combinations of these (Pardinas et al., 1998; Yang et al., 1999). These approaches have been successfully used to identify genes differentially expressed in two populations that exhibit large changes in expression levels, or genes that are expressed at high concentrations in terms of number of copies per cell. Closed systems such as DNA microarrays require genomic sequence information in order to identify differentially expressed transcripts. Open systems have the flexibility of identifying uncatalogued sequences. However, many techniques have a low efficiency of identifying rare genes that are differentially expressed (Martin and Pardee, 2000). This problem becomes worse when the change in expression level of rare transcripts is small.

2.8. SUPPRESSION SUBTRACTION HYBRIDIZATION METHOD OF TRANSCRIPT PROFILING

Genes expressed can play important role in establishing differentiated phenotypes; therefore, their identification is essential for a complete understanding of cellular changes. Suppression subtractive hybridization (SSH) is a technology that allows for PCR-based amplification of only cDNA fragments that differ between a control and experimental transcriptome (Al-Taweel and Fernando, 2011). The singular advantage of SSH lies in the ability to identify differentially expressed genes, irrespective of the level of expression, in the absence of sequence information (Buzdin and Lukyanov, 2007). Differences in relative abundance of transcripts are highlighted, as are genetic differences between species. The technique relies on the removal of dsDNA formed by hybridization between a control and test sample, thus eliminating cDNAs or gDNAs of similar abundance, and retaining
differentially expressed, or variable in sequence, transcripts or genomic sequences. It is a molecular transcriptomic approach for enrichment and detection of secondary metabolite related transcripts, involving a normalization step, whereby pools of housekeeping (e.g. primary metabolite) cDNA fragments from an organism of interest (“tester”) are depleted by combined hybridization and PCR amplification of sequences that are also in a reference strain (“driver”) (Chauhan and Varma, 2009). The SSH method is based on a suppression PCR effect and combines normalization and subtraction in a single procedure. The normalization step equalizes the abundance of cDNA fragments within the target population, and the subtraction step excludes sequences that are common to the populations being compared. This dramatically increases the probability of obtaining low-abundance differentially expressed cDNA and simplifies analysis of the subtracted library.

The remaining cDNA fragments, highly enriched for tester-specific (secondary metabolite) sequences, are then cloned for further analysis (Chauhan and Varma, 2009). Differences in mRNA abundance are alleviated by the hybridization step that normalizes sequence abundance during the course of subtraction by standard hybridization kinetics.