SUMMARY

An extensive study was carried out on the quality potential of triploid genotypes of two polyploid breeding populations in tea. The study was extended to molecular level to find the functional behaviour of the genotypes with contrasting biochemical profiles of caffeine and catechin. The whole study involved five major areas:

1. Two potential polyploid breeding populations were identified on the basis of the flavonoid content,geographical originand population sizeof the parents. Using squash preparations of metaphase plates, consistency in chromosome number was investigated in 130 progenies of both the populations and triploid progenies with consistent chromosome count of 45 were identified.

2. Quantitative estimation of catechins and caffeine in the green leaves of the identified triploid progenies and their parents was carried out using HPLC. Genotypes accumulating highest percentage of total catechin were identified and were marked as heterotic for flavonoid content. Genotypes with lower catechin and caffeine content were recorded as non-heterotic genotypes.

3. The molecular diversity of the genotypes of both the populations was analyzed using microsatellite based SSR marker analysis. Invesigation on markerallele and trait correlation was done which might be useful for further molecular analysis of trait specific marker identification.

4. The affect of polyploidy on the known pathway genes of flavonoid and caffeine biosynthesis were identified using real time PCR quantification.

5. The quality of tea depends on many factors and hundreds of biochemicals are involved in determining quality of black tea. To understand the affect of polyploidization on different other metabolic pathways/regulatory units a SSH approach was followed. Using this technique different EST's were identified that have differentially enriched in the heterotic transcriptome of a successful genotype – T-30. Using RT-PCR quantification, candidate genes were

identified which are differentially expressed in the heterotic genotypes of both the breeding populations.

To find the effect of polyploidization on the genotypes of tea, the first and basic requirement was to find a polyploid population with the genetic backgroundfor quality attributes and a procedure for selection of the parents and progenies. Catechins (flavan-3-ol) are polyphenolic compounds of tea leaves and they contribute to the flavour and taste of black tea. Therefore catechin content was used as one of the criteria for selecting genotypes with quality characters. Based on the diverse geographic origin, size of the breeding population and concentration of total catechins, two breeding populations were identified, where the diploid progenitor had a higher level of catechin content.

A consistent ploidy level was necessary to find the effect of polyploidy on the genotypes. Using metaphase chromosome plates of the progenies, 97 triploid genotypes with a consistent chromosome number of 45 were identified.

Based on the estimation of major catechins and caffeine content of the identified triploids, it was clear that polyploid progenies have the capacity for producing metabolites that have the modulating affect of the black tea quality. However, we could not detect any additive affect of polyploidy on the accumulation of catechins and caffeine. We found that both positive and negative heterosis for catechin and caffeine content occurs within the progenies. It was also observedthat, progenies producing high catechins also produce high caffeine and the triploidization does show effect on the metabolite content. Whether this effect is positive or negative may be determined by the genotype of the progenies.The identification of distinct groups of metabolite accumulation in comparison to their parents indicated unique genetic makeup of the hybrids. The difference in metabolite content was understood to be a function of genetic differences between the progenies.

To understand the genetic behaviour of the inherited triploid genome of the progenies we used SSR based marker analysis. In majority of the loci, both the parents and progenies amplify only two alleles. Moreover, for these loci, either the diploid or the tetraploid parent is heterozygous. For the tetraploid parent,

this condition was found to be confusing for determining the exact copy number of the alleles. Within the progenies, a complex inheritance pattern could be observed. Due to heterozygous nature of the parents at different loci, even in the F1 progenies we could obtain a segregating hybrid population. The ratio of segregating genotypes varies for each of the locus. Moreover, we also found evidence for presence of null alleles within both the populations. This makes a very complex mode of segregating pattern with identification of progenies with unusual genotypes and private alleles. Hence we concluded that a unique genotypic contribution from the parents is possible in polyploid breeding programs involving tetraploid and diploid parents. We believe that there is a relation of chromosome behaviour at meiosis and genetic segregation ratios in progenies.

To find any evidence of relationships between the unique genotypes and metabolite contentof the progenies, a statistical approach was followed. A few alleles were determined which show statistical association with caffeine and catechin level of the genotypes. However, to establish such a relationship, more loci with differentiating alleles for the tetraploid and diploid parents need to be enriched. Then only, unique genotypes of the progenies could be studied in a more reliable manner. In our study we identified four such loci where some of the progenies amplify three alleles per locus. A basic problem with analysis of polyploid genome is the unavailability of statistical simulations for efficient calculation of frequency based and distance based relationship studies. Therefore more focused research in this area is required in future. From this work,it was clear that some of the triploid progenies with their unique genotypes contribute to the enhancement of the secondary metabolic profiles.This justifies polyploid breeding for quality.

By extending our research to the functional behaviour of the progenies, we looked into the expression pattern of the genes/ESTs in the progenies. For initial studies, we selected 11 genes from the pathway of flavonoid and caffeine biosynthesis and tried to find if any detectable difference in expression occurs between the genotypes. We found that like secondary metabolites, at

expression level also effect of polyploidy is non additive. Hence it was clear that increase of genomic copy number is not always reflected at the transcriptomic or metabolite level. We identified five genes from flavonoid pathway (PAL, CHS, CHI, DFR and ANR) and one gene from caffeine pathway (CS), that show characteristic higher gene expression pattern in the heterotic genotypes. This result is based on analysis of transcripts at one time point only. Therefore, an extensive expression study at different stages of biosynthesis should be followed up to gather more information in this regard. In this study, the affect of genotypic contribution of the diploid parent on the functional behaviour of the triploid progenies was found to be prominent. Since the tetraploid parent was common to both the parents, the difference of expression patterns in both the populations was definitely a function of the genetic contribution of the diploid parent. Therefore proper selection of the diploid parent is necessary for the success of polyploid breeding program. Production of new tetraploids may be an option to extend the diversity of tetraploid parent. Since the polyploids of tea are productive, hardy and adaptive to environmental stress conditions, identification of quality potential of the polyploids can be considered as a major breakthrough in the area of tea breeding.

In this study, we used progenies that were derived from different allelic conformations of the same parental contributions within a breeding population. Therefore, it was obvious that allelic variations might be causing beneficial effects in the secondary metabolic pathways in some of the genotypes. So, there must be some regulatory sites at which the genetic variations of the progenies are regulating the secondary metabolite levels. We were interested in the regulatory sites, so that further insight to the breeding of successful polyploids can be achieved. We also wanted to find the other transcripts or domains that were retained preferentially within the heterotic transcriptome so that candidate genes for flavonoid content could be identified. We could identify at least three level of regulation at which the transcriptome of heterotic progeny T-30 is regulated. We found MYB-bHLH-WD40 transcription factor

cascade to be over-expressed in the subtracted cDNA pool of T-30. Similarly, due to identification of differentially accumulated domains, flavonoid modifying enzymes and sub-cellular localization of flavonoids were recognized as another level at which heterotic transcriptome is being regulated. Identification of other over-represented transcription factors, protein modifying enzymes, chaperons, domains of proteins involved in signal transduction and transcriptional regulation in the heterotic transcriptome also suggests that these genes are necessary for the overall success of secondary metabolite biosynthetic pathways. The carbon backbone of each of the secondary metabolite biosynthetic pathways is derived either from the carbohydrate or from citric acid cycle compounds. Differential expression of the genes from these pathways suggests an enriched source of carbon backbone for flavonoid biosynthesis. Identification of transcripts from sikkimate and polyketide pathways suggests an efficient conversion of carbon backbones from the primary metabolic pathways to the secondary metabolite biosynthetic pathways.

A comprehensive transcriptome dataset has been obtained from the subtracted cDNA pool of the heterotic genotype with an ample poolofEST's involved in several major metabolic pathways. Apart from "flavonoid biosynthesis", "carbohydrate metabolism" and "energy metabolism" we identified five other pathways to which the representative EST's belong. These pathways are – "metabolism of terpenoids and polyketides", "lipid metabolism", "amino acid metabolism" and "phenylalanine metabolism" and "tryptophan metabolism". If we consider EST representations from these pathways as an indicator of probable up-regulation of the pathways, then the T-30 genotype can be considered as promising genotype with quality characteristics. Based on the sequencing of differentially expressed transcriptome of the most heterotic genotype, we selected 18 EST's as candidate genes for validating the subtracted library. We were interested to find whether these EST's are capable of differentiating the two heterotic and non-heterotic groups of genotypes with extreme range of total catechin content. We found that, expression pattern of GST, CoNMT, A3OG, ERF and UCR complex may serve as important

markers in differentiating these groups. These EST's were found to have a characteristic mode of expression in the heterotic and non-heterotic groups of genotypes of both the populations. Moreover, apart from these, some other EST's were also identified in both Stock 615 and Stock 616. This indicates that for each population due to different genetic recombination or due to different parental combinations different transcriptional regulation may occur. Although we cannot control the recombination event, by selecting a proper parent, an attempt can be made to organize the transcriptome for transcripts involved in the better adaptability and quality related secondary metabolite biosynthesis. Therefore, a proper selection strategy for parents based on reliable quantification methodologies should be practiced, so that, complex traits like quality can also be expected in polyploid progenies.

The metabolic, structural and functional diversities of the triploid genome demonstrated quantitative enhancement of biochemical profile of triploid *Camellia* germplasms with identification of positively modulated quality related genes. This transcriptome dataset can serve as an important platform for public information on genomics, gene expression, and functional genomics studies in polyploid breeding of tea. These studies provided a comprehensive survey of the gene expression and genome changes accompanying polyploid formation and the authenticity as well as applicability of polyploid breeding for quality germplasms in tea. The present work represents a platform on which complicated problems related to polyploid mapping can be solved within our framework, integrating statistics, genetics and computational science.