Chapter 5
Discussion
5.1. PRIMER COMBINATION SELECTION

For revealing polymorphism for an organism with small genome size (10^6-10^7 bp), AFLP primers with one or two selective nucleotides are necessary. According to Vos et al. (1995) the number of amplification products generated by the AFLP is related to the size of the genome and the number of selective nucleotides added to the 3' end of the primers. Species with a large genome size are expected to generate a large number of AFLP bands making analyses more cumbersome, therefore, additional selective nucleotides are required for those organisms with larger genomes (10^8-10^10 bp) (Blears et al. 1998). Introduction of one more restriction enzyme limits the number of bands. Theoretically, using the TE-AFLP technique the number of fragments which are potentially amplified can be easily reduced 20 fold compared to traditional AFLP (van der Wurff et al. 2000). Considering all the nucleotides are randomly distributed in the genome the expected frequency of the restriction sites of two rare cutters and one frequent cutter in tea genome (4.0x 10^9 bp) is estimated to be approximately 1.758x10^7. Due to the addition of a second rare cutter in TE-AFLP, extra sites have been added to increase the discriminatory power of the technique.

Like RAPD, the standard TE-AFLP with only one step also has high probability of generating artifactual bands associated with misannealing of primers and resulting in poor reproducibility. In the present study, to avoid this problem, a pre-amplification
step was added. Addition of the pre-amplification step also gave more of the amplified products for subsequent selective amplification.

Five primer combinations were selected in the final screening. An average of 36 prominent fragments within 1Kbp were obtained per reaction, indicating that silver stained polyacrylamide gel retained the resolution power of the radio-labelling technique. High resolution silver stained polyacrylamide gel without radioactive isotopes proved to be a more efficient technique to visualize TE-AFLP markers in tea. Silver-staining reduced the time and cost as well as eliminated hazard of working with radioisotopes (Vantoai et al. 1996, Chalhoub et al. 1997). In addition, the high resolution and recovery of fragments from the dried gels made silver staining a more useful and versatile detection method for AFLP amplification products.

In the biclonal population taken for the present study, the selected primer combinations could resolve differences at the level of intra-species. Mueller and Wolfenbarger (1999) also reported that AFLP markers have the potential to resolve genetic differences at the level of ‘DNA fingerprints’ for individual identification and parentage analysis. Several studies had identified inter- and intraspecific hybrid individuals with AFLP (Liu et al. 1998; Congiu et al. 2001; Bensch et al. 2002; Chauhan et al. 2004). However, the usefulness of AFLP markers for systematics rests more on the rapid grouping of closely related lineages. Phylogenetic inferences based on similarities of AFLP profiles become problematic for higher taxonomic levels, because the high variability of AFLP markers reduces similarities between distant taxa to the level of chance (Mueller and Wolfenbarger 1999).
A ‘narrow-down’ strategy for the primer combination and increasing number of samples from 1 to 5, and finally to 17 samples in three different steps was followed. This strategy was beneficial in choosing the optimised TE-AFLP primer combinations. Four marker indices were used to examine the overall efficiency of the primer combinations selected. PIC has been used extensively in diversity and marker studies (Vos et al. 1995; 245 Milbourne et al. 1997; Rosales et al. 2005; Tatikonda et al. 2009). PIC values for the TE-AFLP primer combinations in the study showed that it is an efficient parameter to reveal the diverse nature of the genotypes but it is independent to the number of fragments per primer combination. In the study, the PIC value ranged from 0.33 (E-AAC / P-ATG) to 0.40 (E-AAT / P-CGA) with an average of 0.36 which is higher than the TE-AFLP primer combinations used in *Pongamia pinnata* diversity study (Sharma et al. 2011).

The AFLP generates a large number of fragments per primer combination. This leads to a higher EMR because the number of polymorphic bands is directly proportional to EMR. The multilocus nature of AFLP markers is well suited for detecting polymorphism and distinguishing genotypes using few primer combinations. In this study, EMR for E-ACG/P-AAA (kk) primer combination was the highest (15.93) where the polymorphic bands were highest. Since MI is a product of two parameters: EMR and PIC, it is used to calculate the overall utility of a maker system. MI together with PIC value has been used to assess the discriminatory power of AFLP primer combinations used in several cross pollinated plants including Radish (PIC = 0.24, MI = 5.14) (Muminovic et al. 2005); *Jatropha curcus* L. (PIC = 0.26, MI =
Chapter 5 – Discussion

25.13) (Tatikonda et al. 2009) and Pongamia pinnata L. (PIC = 0.31, MI = 30.75) (Kesari et al. 2010).

The discriminatory power of the TE-AFLP primer combinations was also evaluated using the two parameters i.e. MI and PIC in Pongamia pinnata accessions (MI = 7.95, PIC = 0.30) (Sharma et al. 2011). In the present investigation, MI values were reported in the range of 1.04 to 5.73 (average 2.75) and PIC in the range of 0.33 to 0.40 (average 0.36). Discriminating as many accessions as possible would be the most important feature of a given primer combination. Various studies reported a high MI in AFLP markers compared with other marker system (Bohn et al. 1999; Hongtrakul et al. 1997; Powell et al. 1996; Russell et al. 1997). There is a reduction in the number of bands generated in TE-AFLP. Consequently the MI value was much lower than the standard AFLP.

Rp is well suited for comparing primers, primer-enzyme combinations or probe-enzyme combinations generated by RFLP, RAPD, ISSR-PCR or AFLP analyses (Prevost and Wilkinson 1999). Since Rp is the sum of all the informativeness of each band, higher the number of polymorphic bands of a primer combination, higher is the Rp. In the present work, Rp value was highest for E-ACG/P-AAA (kk) primer combination and MI value was also the highest for the same. The combined use of all these parameters clearly showed the ability of the primer combinations to distinguish the different genotypes. The primers with higher Rp values have a greater capacity to separate genotypes (Prevost and Wilkinson 1999) hence they are useful for bigger applications.
Laurentin and Karlovsky (2007) reported that there was a lack of correlation among PIC, MI and Rp indicating that a single parameter was not sufficient to assess the informativeness of a primer combination. We too found no specific relation among the parameters used by us. In this study, the PIC was highest in E-AAT/P-CGA (fh) whereas EMR, MI, Rp were highest in E-ACG/P-AAA (kk). However, the values of the latter three parameters are directly proportionate to the number of polymorphic fragments generated while PIC is independent of it. There is not much difference of PIC values (an average of 0.36) in the primer combinations undertaken which showed that all the five primer combinations could be used for variation studies on a wider scale. However, EMR, MI and Rp were markedly higher in E-ACG/P-AAA (kk) showing that it was the most informative among the five primer combinations. Thus, TE-AFLP is better suited for analysis of larger genomes. The addition of the pre-amplification step helped improve the accuracy and reproducibility of amplification, and produced much cleaner bands. The number and sequence of the selective bases of the primers in the selective amplification further enhanced the efficacy of this protocol. Silver staining was found to be more advantageous compared to radio-labelling and was used for visualizing the DNA fragments of the selective PCR amplification in the sequencing gel. Therefore, we recommend use of silver stained TE-AFLP for analysing a larger number of tea samples.

5.2. SEGREGATION STUDIES

5.2.1. Segregation analysis

Five TE-AFLP primer combinations which were selected in the final screening were used in mapping analysis of 117 genotypes of the bi-clonal population of TS 463 i.e.
F<sub>1</sub> generation of a cross between two tea clones TV1 (assam-china) and TV19 (comob). Two kinds of segregating AFLP markers could be identified. Type 1:1 markers, where one parent was heterozygous (presence of band, \( A/a \)) and the other was homozygous (absence of bands, \( a/a \)) and the \( F_1 \) offspring were expected to segregate in a 1:1 (presence of band, \( A/a \): absence of band, \( a/a \)) ratio. The other type was a 3:1 marker where both parents were heterozygous (presence of bands in both cases; \( A/a \)) and the offspring were expected to segregate in a 3:1 (presence of bands, \( A/a \) or \( A/A \): absence of bands, \( a/a \)) ratio. All segregating markers were tested for goodness of fit to the 1:1 and 3:1 Mendelian ratio using chi-square analysis and a significance level of \( P = 0.1 \). A total of 153 polymorphic markers were generated, out of which 100 markers comprised of the first type (type 1:1 marker) and 53 markers comprised of the second type (type 3:1 marker). The number of 1:1 markers that segregated in the TV1 parent was more than TV19 parent (59 and 41 respectively).

Linkage maps of tea had been generated using different tea plant sources and various types of molecular markers. Hackett et al. (2000) generated 420 (RAPD and AFLP) markers of which 304 markers comprised of 1:1 marker type while 116 markers comprised of 3:1 marker type in tea (\emph{Camellia sinensis}). Kamunya et al. (2010) generated 260 markers (RAPD,SSR,AFLP) of which 149 markers showed 1:1 segregation ratio and 118 displayed 3:1 segregation ratio and Taniguchi et al. (2012) generated 1124 markers (SSR,RAPD,CAPS,STS). The number of markers generated in our study was much lower when compared due to lower number of primer combinations used in the study. However, primer combinations used in this study
generated on average 20 polymorphic markers between the parents (type 1:1 markers) which was higher than AFLP markers generated by Hackett et al. (2000) (an average of 10.5 polymorphic bands per primer combination). The high polymorphic rate observed in our study may be due to the fact that the samples taken for the study were from artificially selected populations (Moore et al. 1999; Li et al. 2006) since tea is a plantation crop. In addition, the F1 population was obtained by crossing two highly heterozygous parents belonging to different species both of which were open pollinated.

Using the same primer combinations, 59 and 41 markers segregated through TV1 and TV19 parent respectively. This showed that TV1 parent was heterozygous for a larger number of loci than TV19 parent. But in case of informativeness for linkage analysis, markers segregating through TV19 were more informative as 30 out of 41 markers (73.17%) were mapped in the TV19 map while 36 out of 59 markers (61.02%) were mapped in TV1 map. Three markers which were highly distorted ($P < 0.01$) from Mendelian ratio were excluded from linkage analysis from each parent. Therefore, eight TV19 markers and twenty TV1 markers remained unassigned to linkage groups. Among the five primer combinations, primer combination E-ACG/P-AAA (kk) was most informative, generating the highest number (40) of markers. If the markers segregating in either parent (1:1 marker type) were considered, markers generated by E-ACG/P-AAA (kk) and E-AAC/P-ATG (hi) primer combinations were highest (11 and 9) in TV1 parent and TV19 parent respectively.
5.2.2. Segregation distortion

All segregation markers were checked with $\chi^2$ test ($P < 0.1$) to identify deviations from Mendelian 1:1 ratio. Distorted markers were suffixed with a minus (-) for homozygote deficiency and a plus (+) for homozygote excess. The markers from the TV1 parent had a higher proportion of distortions compared to TV19 parent (i.e. 11 markers in TV1 parent and 6 markers in TV19). The proportion of distorted segregating markers in this study was 17% for markers heterozygous in either parent.

The percentage of significantly distorted markers detected (17%) was similar to that observed in linkage analyses of pine (14–15%; Kubisiak et al. 1995), eucalyptus (15%; Marques et al. 1998), Quercus (18%; Barreneche et al. 1998), willow (18%; Hanley et al. 2002), and artichoke (14%; Lanteri et al. 2006).

The markers with highly distorted segregation ratios at the 1% level ($P<0.01$) were excluded from linkage analysis. However, we included the markers showing distorted segregation ratios ($0.05 < P < 0.1$) in the linkage analysis. In TV1 parent and TV19 parent data, 8 and 3 distorted markers ($0.05 < P < 0.1$) were included respectively in the mapping data sets.

Markers with distorted segregation ratios ($0.05 < P < 0.1$) were included in the linkage analysis to see if these markers could help in connecting between markers to obtain a longer linkage group. In addition to this, it would also help in identifying possible regions of distortion by investigating regions of clustering of markers. However, most of the distorted markers from both the parents were found to be unlinked and could not be placed on the map. No clustering of the distorted markers was also observed thereby regions of segregation distortion could not be identified in
this study. This was due to less distorted markers ($0.05 < P < 0.1$) (to avoid false linkages) were linked to the linkage groups in the study, i.e. only 4 distorted markers were linked to the TV1 map while only one distorted marker was linked to TV19 map.

Several reasons for distortion of segregation ratios in many plants have been reported that range from biological to non biological factors. The biological factors include chromosome loss (Kasha and Kao 1970), genetic isolation mechanisms (Zamir and Tadmor 1986), and the presence of viability genes (e.g., Hendrick and Muona 1990; Beavis and Grant 1991; Liedl and Anderson 1993; Bradshaw and Stettler 1994). Nonbiological factors include scoring errors (Devey et al. 1994; Xu et al. 1997; Nikaido et al. 1999), marker sampling error and low population size (Lu et al. 1998) and sampling errors (Plomion et al. 1995; Echt and Nelson 1997). The nonbiological factor for the distortion may also be the superimposition on the gel of non-allelic amplified products corresponding to different loci (Virk et al. 1998). Transmission distortion between genetically different genomes (Fishman et al. 2001); zygotic selection (Rick 1969) or using parents from different populations containing a high genetic load (Fishman et al. 2001) can also lead to distortion in segregation ratios.

Another significant issue that hampers all aspects of map development is experimental error such as wrongly scored individuals or contamination.

Segregation distortion may also be caused by a shortage in identical-by descent homozygotes due to a high genetic load (Launey and Hedgecock 2001). This is due to harmful recessive alleles being filtered out of the family line in outcross species. In our study, homozygote deficiency accounted for more of the segregation distortion
(76.47%) compared to homozygous excess (3.92%). Genetic load is likely a contributing factor to the segregation distortion observed. Similar results have also been reported in other outbreeding tree species such as populus (Bradshaw and Stettler 1994) and eucalyptus (Gion et al. 2000). Another contributing factor for segregation distortion in our study could be scoring error as AFLP has been reported to have and error rate of 0.6-2% per band (Arens et al. 1998; Janssen et al. 1997).

5.3. LINKAGE MAPPING

5.3.1. Linkage map

Two separate genetic linkage maps for TV1 and TV19 were constructed using the two-way pseudo-testcross strategy (Grattapaglia and Sederoff 1994) based on the two separate data sets obtained for heterozygous segregating genetic markers. The pseudo-testcross approach has been widely used for construction of linkage maps in tea (Tanaka 1996; Hackett et al. 2000; Kamunya et al. 2010; Taniguchi et al. 2012) and in many other outbreeding species.

In our study, TV1 map consisted of 36 loci, spread over 10 linkage groups while in TV19 map, 30 loci were distributed in 11 linkage groups. The number of linkage groups obtained in this study was therefore less than the haploid chromosome number (n=15) for this species. Kamunya et al. (2010) also could obtain only 19 maternal and 11 paternal linkage groups in tea.

Estimates of total genome map size have been calculated on tea in various studies including a female map that covered 1349.7 cM with an average distance of 11.7 cM between loci (Hackett et al. 2000); a maternal map spanning 1.012 cM, while the paternal map covered total length of 399.5cM, with mean distance between markers
Chapter 5 – Discussion

being 14.7 and 12.9 cM, respectively (Kamunya et al. 2010). Recently, in a high reference map developed by Tanaguchi et al. (2012) the core map of tea covered a total length of 1218 cM. In our study, 10 linkage groups of TV1 map covered a total map length of 388.9 cM at an average distance of 38.89 cM between the markers while in TV19 map, 11 linkage groups covered a total map length of 410.7 cM with an average distance of 37.34 cM between the markers.

Therefore, linkage maps generated in this study had comparatively low genome length due to less number of markers. The linkage maps were also incomplete as indicated by the lower number of linkage groups when compared to the haploid chromosome number. Large intervals (> 20cM) were also observed in both the maps generated including all the markers. The existence of these minor linkage groups and unlinked markers (20 markers in TV1 parent and 8 markers in TV19 parent) also indicated that there were many large gaps with few markers (Kesseli et al. 1994). All these shortcomings were attributed to the less number of markers used for linkage map generation.

The density and resolution of the linkage map will be increased if more molecular markers were added. The addition of markers to these maps should reduce the larger intervals and thus the gaps (Lallias et al. 2007). Furthermore, increasing the marker density may result in the linkage groups containing only two or three markers to coalesce into larger linkage groups (Baranski et al. 2006).

Few distorted markers (0.05 < P < 0.1, to avoid false linkages) were included in the linkage analysis and only 5 markers were linked in the parental maps (4 marker in TV1 parent and 1 marker in TV19 parent). Since only 5 markers were linked to
linkage groups and no clustering was observed, regions of segregation distortion could not be identified in this study.

Another factor contributing to the low map length could be the presence of genotyping errors which can overestimate the proportion of double crossovers and thus severely inflating map lengths (Hackett and Broadfoot 2003). Therefore, the rigorous data checking and exclusion of numerous scored markers with an apparent excess of double crossovers may have had the effect of keeping map length to a minimum. The experiences gained in this study shall be used in subsequent studies to overcome these limitations.

5.3.2. Map comparison

The heteroduplex markers which were descended from both the parents (3:1 marker type) were essential for the integration of the separate parental linkage maps into one consensus map. However, in our study bridging markers were clustered among themselves and not linked with markers specific to either parent (1:1 markers) (Fig. 4.3.3.3). This uneven distribution throughout the parental maps may be due to the low density of markers generated. So, these bridging markers could not be used to detect homology between linkage groups with markers specific to either parent. Since there is no information on homology of the linkage groups, integration of linkage groups to make an integrated/consensus map was not possible.

The addition of more markers (both marker types) to the parental maps is required for the establishment of a consensus map. The addition of more informative co-dominant markers such as microsatellites, SNP, CAPS (Omura et al. 2000, Yamamoto et al. 2002, 2005, 2007), will further increase the accuracy of the
consensus map, as they will serve as anchor loci between the two parental maps (Powell et al. 1996). Recently a consensus map based on a combination of different markers (AFLP, SSR and CAPS) has been established in tea by Taniguchi et al. (2012).

5.3.3. Marker distribution

The AFLP marker distribution was analyzed by calculating the Pearson correlation coefficient between the number of AFLP markers in the linkage groups and the size of the linkage groups (Yu and Guo 2003), *t*-test was applied to test the significance of correlation coefficient at *P* = 0.1 level.

The Pearson correlation coefficient (*r* = 0.91 and *r* = 0.83) indicated that there was a highly significant positive correlation between the number of markers and linkage group size. This supports our contention that the distribution of AFLP markers was relatively even in linkage groups of TV1 and TV19 maps.

The distribution of the mapped AFLP markers was also classified according to the markers generated by 5 primer combinations used for analysis by visual examination (Fig. 4.3.4.1 and fig 4.3.4.2). In this study, we found the accumulation of markers (1:1 markers) on a particular linkage group amplified by the same primer combination and the distribution was relatively even in both the parental maps. The number of AFLP markers in the linkage groups that had high positive correlation with the size of the linkage groups also supports this conclusion.
Chapter 5 – Discussion

The AFLP markers segregating at 3:1 ratio were not distributed along with the 1:1 markers, and therefore remained uninformative in finding homology between linkage groups.

5.3.4. Genome coverage

The estimated genome lengths determined for both the parental maps in this study were 766.1 cM and 894.5 cM for TV1 and TV19 parent respectively. The difference seen between the two maps is not an uncommon occurrence as it had also been observed in maps of other tea plants (Kamunya et al. 2010; Tanaguchi et al. 2012). Possible reasons for the variations between the estimates for both the parents are:

(i) the distribution of markers along the chromosome differs in both parents. (ii) recombination frequencies occurring in both gametes may have differed, and (iii) errors within the data, such as statistical and experimental errors (Wu et al. 2004).

The map coverage of both the maps was low i.e. 50.76% and 45.91 % for TV1 and TV19 respectively.

The addition of genetic markers, especially different genetic markers, such as additional microsatellites, EST and SNPs (Wang et al. 2004) are needed to improve genome coverage. Although low, it has potential to serve as a basis for the development of a denser and more saturated linkage map.

5.4. SEQUENCE ANALYSIS

The current results demonstrate the ability of combining the advantages of PCR and restriction pattern fingerprinting methods in AFLP analysis method that could detect single nucleotide polymorphisms in addition to minor insertions, deletions, transpositions, duplications or inversions. Those sequences that turned up in our
analysis indicated that the particular sequences were subjected to relatively frequent sequence variations. Most of the genes for which map positions are reported in this linkage map were polymorphic regions that reveal putative functions relating to housekeeping genes, functioning in general plant metabolism. Thus it showed that a number of multilocus housekeeping genes accumulate neutral variation/mutation providing more discrimination in the genome. Sometimes housekeeping genes may be of considerable value in determining evolutionary mechanisms and evaluating diversity. Although the similarity of housekeeping genes might be an indication of genome similarity, there are many examples, particularly in large genomes, where dramatic differences in gene content have been found in organisms (e.g. Roca et al., 2003). Even distantly related plant species have in common these genes with similar functions and a high degree of sequence conservation. In our study, 12 mapped markers each of both TVl and TV19 showed homology with the sequences from different database of NCBI GenBank. The homologous sequences came from several species; unigenes that showed homology with TV1 markers were from Solanum lycopersicum, Ricinus communis and majority were Vitis vinifera. All the 12 sequences also showed homology with sequences from the whole genome shotgun sequence database, 7 mapped markers and one unlinked marker showed homology with whole genome sequences of Camellia sinensis var sinensis (Table 4.4.1.1. and 4.4.1.3.). In case of TV19 markers, homologous unigenes were from various plant species e.g. Populus trichocarpa, Fragaria vesca supsp. vesca, Betula pendula, Solanum lycopersicum, Cicer arietenum. 9 out the 11 TV19 mapped markers and one unlinked marker showed homology with whole genome sequences of Camellia sinensis var sinensis (Table 4.4.1.2. and 4.4.1.3).
Chapter 5 - Discussion

Those sequences that did not identify any similarities of the DNA fragment with any known genes or anonymous sequences after a BLASTn search at NCBI were subjected to find ORF region by ORF finder (NCBI). The sequences were translated for all possible reading frames but did not result in any known protein (BLASTp). The absence of homology could be caused by several factors. A large proportion of sequences were too short to allow statistically meaningful matches. However, for some sequences, the absence of homologous sequences in the public databases may indicate specific roles for them in *C. sinensis* (Shi et al. 2011).

Unigenes of various plant species showed homology with some of the sequences in our study. In TV 1 map, out of 12 sequence (mapped) markers, three markers showed homology with unigenes in the database (Table 4.4.1.1.). The details of the unigenes are as follows:

1. Peptide/nitrate transporter

Two of the sequences aligned with the putative mRNA for nitrate transporter of *Vitis vitifera* and *Fragaria vesca* subsp. *vesca*. This unigene is involved in primary assimilation of inorganic nitrogen and amino acid metabolism. So, this would be helpful in understanding physiological processes important for tea cultivation and quality.

2. Chromodomain helicase DNA binding protein

The CHD (Chromodomain Helicase DNA Binding Protein) family of proteins is known to be involved in the regulation of gene expression, recombination and chromatin remodelling via their chromatin specific interactions and activities, e.g. CHD2 is a multifunctional protein that is involved in the regulation of the vital cellular processes of pre-mRNA splicing, DNA repair and DNA damage induced
transcriptional activation of stress response genes at the molecular and cellular levels
(Woodage et al. 1997)

3. Wall-associated receptor kinase 2-like
So far, WAKs have been identified only in *Arabidopsis* and they play important roles in cell expansion, heavy-metal stress tolerance and pathogenic bacteria resistance in the plants (Lally et al. 2001; He et al. 1998; Hou et al. 2005).

In TV 19 map, 8 mapped markers showed homology with unigenes in the database (Table 4.4.1.2). The details of the unigenes are as follows:

1. Chromatin-remodeling factor PICKLE-like
   It is related to the function of inhibition of some late embryogenesis genes after imbibition to promote germination (Perruc et al. 2007).

2. Phylocalpain
   Calpain, a calcium-dependent cysteine, plays an essential role in basic cellular cells. Phytocalpain plays a key role in the determination of the proliferation and differentiation fates of cells during organ development of dicotyledonous plants (Ahn et al. 2004).

3. Protein decapping 5-like
   A study in *Arabidopsis thaliana* found that Decapping 5 (DCP5) is required for mRNA decapping, P-body formation and translational repression during postembryonic development (Xu and Chua 2009).
4. PREDICTED: histidine kinase 4-like

In *Arabidopsis thaliana*, His-kinase was recently suggested as a sensor for cytokinins which have important function in cell division and differentiation (Nishimura *et al.* 2004).

5. DNA/RNA-binding protein KIN17-like

Kin17 is believed to be involved in the cellular response to DNA damage, gene expression, and DNA replication.

6. PREDICTED: cytokinin receptor 1 (CRE1) mRNA

Elevated cytokinin levels maintain high cellular redox potentials during drought and resistance to biotic and abiotic stresses in tobacco (*Nicotiana tabacum*) (Rivero *et al.* 2007). However, recent findings have shed light on a distinct role of cytokinins in plant immune responses (Choi *et al.* 2010; Sano *et al.* 1994; Sano *et al.* 1996). For these reasons, cytokinin receptor might be an important application target for generating drought tolerant and disease resistant in tea.

7. PREDICTED: glutamate-rich WD repeat-containing protein 1-like

These proteins form a very large family that is both diverse in function and structure. They play key roles in the formation of protein-protein complexes in nearly all the major cellular pathways.

8. Putative NADPH-dependent oxidoreductase

NADPH-DOs include essential enzymes for a variety of cell functions in plants, such as NADPH oxidases which control defence mechanisms in plants (Mehlby *et al.* 1996) and also play role in the control of plant growth by zinc and auxin (*Ogawa* 2004).
One unlinked TV1 marker also showed homology with the unigenes in the GenBank.

1. Cysteine-rich receptor-like protein kinase 29

Cysteine-rich repeat RLKs (CRKs) functions during oxidative stress, pathogen attack and the application of salicylic acid (SA) (Du and Chen 2000; Wrzaczek et al. 2010).

Using the above searches and analyzing the homology, 4 sequences were selected which showed homology with the existing tea EST database. EST sequences (regulatory sequences + exons) could be considered as a good representation of coding sequences. The sequences were translated in all possible six reading frames (3 forward and three reverse) and a similarity search was done against the protein databanks. This approach was found to be successful to predict gene specific products. ORFs of two sequences i.e. gka11 and hga1 showed homology with unigenes in the NCBI database.

AFLP markers were sequenced for purposes of physical mapping and, at least in some cases, to generate codominant marker systems. However no informative codominant markers (intercross marker) from the sequences could be achieved that could be linked to any of the linkage groups.

Current studies have shown that functional genes appear to be relatively less compared to the non coding regions. There have been reports that AFLPs cluster around the centromeres and telomeres where functional genes are rare. Moreover, the noncoding nature of AFLPs may provide a rational explanation for their high variability.
However, some sequences showed homology with unigenes which were related to nitrogen assimilation, drought, pathogen resistance, heavy-metal stress tolerance, resistance to biotic and abiotic stresses. The identification of such markers is an important resource for research aimed at understanding physiological processes important for tea cultivation and quality. The identification of markers closely linked to desirable traits is a prerequisite for the application of MAS breeding programmes, as the markers will be used as identifiers to locate the presence of desirable traits in breeding families (Wang et al. 2004).

These findings will lead to the genetic enhancement of this species, which is 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.