Chapter 4
Results & Discussion

If I have seen further than others, it is by standing upon the shoulder of giants.

-Isaac Newton

4.1. Morphological documentation

The tea clones or planting material for northern regions has been certified by Tocklai Research Association (TRA) based on standard, yield, quality, suitability, etc. Some of the TRA certified clones for Darjeeling based on standard are AV2, B 7/1A/76, B 9/3/76, BB157, Lingia 12, TV 14, T78, T135, T383, P1258, RR17/14, SKM 1, TV 1, Thurbo 3, Thurbo 9, Badamtam 15/263, P1404, T246. The TRA-approved Darjeeling clones with high yield are Happy Valley 39, RR4/5, TV 19, CP1, K1/1, Sundaram, B/5/63, T 253, and the quality clones are B688, B777, and T145.

Fig.4.1. Leaves of *Camellia sinensis* showing alternate and serrate morphology

The clones of Darjeeling and Dooars chosen for our study comprised of all the three main types of tea i.e the China type (small-sized leaf), Assam type (large-sized leaf), and the Cambod type (intermediated sized leaf). Generally, tea (*Camellia sinensis*) is an evergreen shrub or tree with simple, alternate, and serrate leaf (Figure 4.1). It is a diploid plant (2n=30) with 15 chromosomes. It is usually cross-pollinated since it is a self-incompatible plant. It has bisexual and regular flowers usually with five petals and sepals. It has multiple two-celled stamens with anthers. The ovary is superior, 2-4 locular, axile, and rarely solitary. It bears a capsule-like fruit with recalcitrant seeds as shown in Figure 4.2 (the seeds which losses viability on drying).

Tea or *Camellia sinensis* usually share more or less a common morphology among the different types and only differs in leaf morphology (size, shape, pubescence, etc), shoot (large, small, or medium). The leaf and shoot morphology of the studied tea clones is provided in Table 4.1. Different parameters like shoot size, pubescence, yield, flavor, drought resistance, suitability, resistance against pests, etc. were evaluated for the clones under investigation and the detailed information is given in Table 4.1. Among the 33 studied tea clones, the standard clones included clones like T 78, T 246, T 383, AV2, P312, RR 17/144, P 1404, P 1258, TTV 1, SKM 1, BT 15/263, BS 7/1A/ 76, BS 9/3/76, TS 378. The clones with high yield included T 253, RR 4/5, CP 1, HV 39, Sundaram, TV 19, and the quality clones included clones like T 145, K 1/1, B 777, and B 688 (Singh, 2005).

Fig.4.2. Capsule like fruit of *C. sinensis* with recalcitrant seeds

Clones like AV2, B 668, T 383, and K 1/1 comprised of the clones with very good flavor, and the rest are categorized into average (T 145, T 253, T 246, B 777, RR 4/5, CP 1, P1404, P 1258, Sundaram, HV 39) or good flavor (B 157, T 78, T 383, RR 17/144, TV 19, TS 378, TS379, B 777, P 312, SKM 1, BT 15/ 263, BS 7/1A/76, BS 9/3/76). Likewise, the clones are categorized as per their yield into clones with very high yield (RR 17/144, CP 1, P1404, Sundaram, HV 39, TV 19), high yield (P 312, AV 2, T 253, RR 4/5, T 78, T 383, P 1258, TTV 1, BT 15/263, BS 7/1A/76, BS 9/3/76, SKM 1, TS 379), average yield (T 145, B 777, B 688, K1/1, TS 378) and

Table 4.1. Data collected for clones under investigation considering various parameters (Singh, 2005)

	CATEGORY
Standard	T 78, T 246, T383, AV2, P 312, RR 17/144, P 1404, P 1258, TTV 1, SKM 1, BT 15/263, BS 7/1A/76, BS 9/3/76, TB 3, TB 9, TB 3, TS 378, TS 379, T 157
Yield	T 253, RR 4/5, CP 1, HV 39, Sundaram, TV 19
Quality	T 145, K 1/1, B777, B688
	SHOOT SIZE
Small	RR 4/5,TB 3, TS 378
Medium	T 78, AV2, P312, RR 17/144, P1258, TTV 1, SKM 1, BT 15/263, BS 9/3/76,TB 9,T 145, B 777, B 688, K1/1
Large	T 253, T 246, T 383, CP 1, P 1404, Sundaram, BS 7/1A/76, TV 19
	PUBESCENCE
Low	B 157, T 145, TV 19, Sundaram
High	P 312, AV2, RR 4/5, T 78, T 383, RR 17/144, P 1404, K 1/1, TTV 1, HV 39, BS 7/1A/76, BT 15/263,TB 3, TB 9, TS 378, P 1258, BS 9/3/76
Very high	T 253, T 246, B 777, B 688, SKM 1, CP 1
	YIELD
Average	T 145, B 777, B 688, K 1/1, TS 378
Above average	B 157, T 246, TB 3
High	P 312, AV 2, T 253, RR 4/5, T 78, T 383, P 1258, TTV 1, BT 15/263, BS 7/1A/76, BS 9/3/76, SKM 1, TB 9
Very high	RR 17/144, CP 1, P 1404, Sundaram, HV 39, TV 19
	FLAVOUR
Average	T 145, T 253, T 246, B 777, RR 4/5, CP 1, P 1404, P 1258, Sundaram, HV 39
Good	B 157, T 78, T 383, RR 17/144, TV 19, TS 378, B 777, P 312, SKM 1, BT 15/263, BS 7/1A/ 76, BS 9/3/76, TB 3, TB 9
Very good	AV2, B 688, T 383, K 1/1
	DROUGHT RESISTANCE
Poor	T 246
Fair	B 777, K 1/1, TB 9
Good	P 312, B 157, T 145, AV 2, T 253, RR 4/5, B 688, T 78, T 383, P 1404, P 1258, TTV 1, Sundaram, HV 39, SKM 1, BT 15/263, TV 19, BS 7/1A/76, BS 9/3/76, TB 3
Very good	RR 17/144, CP 1, TS 378
	SUITABILITY
Sub-marginal/ droughty conditions	P 312, B 157, T 145, T 253, T 78, T 383, AV 2, RR 4/5, P 1258, TTV 1, TS 378, BS 7/1A/76, BS 9/3/76, TB 3, TB 9
Infilling and Interplanting	T 78, RR 4/5, T 253, CP 1, HV 39, P 1404, Sundaram, SKM 1, BT 15/263, TS 378, TV 19

above-average yield (B 157, T 246). Some abiotic stress like drought were also considered for evaluation and the clones are categorized following their ability to resist drought with clones having poor (T 246), fair (B 777, K1/1), good (P312, B 157, T 145, AV2, T 253, RR4/5, B 688, T 78, T 383, P 1404, P 1258, TTV 1, Sundaram, HV 39, SKM 1, BT 15/263,

Table 4.2. Data collected for resistance or susceptibility against different diseases and pest (Singh, 2005)

RESISTANT TO MITE	AV2, B/5/63[***], Sundaram, TS-378[***], HV39[***], TV 19
RESISTANT TO BLISTER BLIGHT	AV2, P312[****], TS 378[*], T 383, TS 378[***], Nandadevi, TB 9[***], T 145[***], T 78[***]
RESISTANT TO RED SPIDER	HV 39[*], T 383, B/5/63[***], T 246[**], B 157[****]
SUSCEPTIBLE TO BLISTER BLIGHT	HV 39[****], T 253, TTV 1, K1/1, B 15/263, B5/1A/76[*], B 777, T 135, BS 9/3/76[**], CP 1, P 1404[*], P 1258[**], RR 4/5[**], B 157
SUSCEPTIBLE TO RED SPIDER	T 253, TTV 1, B 15/263, B5/1A/76[*], BS 9/3/76[**], P 1258[**], RR 4/5[**], TB 3, TB 9, T 78
SUSCEPTIBLE TO MITES	K 1/1, B5/1A/76[*], B 777, BS 9/3/76[**], P 1404[*], P 1258[**], RR 4/5[**]

[*]Slightly resistant or susceptible; [**]Moderately resistant or susceptible; [***] Fairly resistant or susceptible; [****]Highly resistant or susceptible

TV19, BS 7/1A/76, BS 9/3/76)) and very good (RR 17/144, CP 1, TS 378) resistance towards drought.

The studied clones also exhibit resistance against various pests and diseases (Table 4.2.). The tea clones, resistant to mites include AV2, B/5/63, Sundaram, TS 38, HV 39, TV 19 and, the clones resistant to red spiders are HV39, T383, B/5/63, B668, T 246, B 157. Some clones are also resistant to the havoc-causing blister blight disease. The resistant clones towards the blister blight are AV2, P312, TS 378, T 383, Thurbo 9, T 145, T78. However, some clones are susceptible to the mite, red spider, and blister blight, the details of which are provided in Table 4.2. Morphological descriptors do not give us a complete idea about the genetic variation or diversity due to excess cross-breeding. However, prior knowledge about the phenotype is crucial for further breeding programs.

4.2. Molecular Documentation

Crude DNA was isolated from 33 tea samples of clones using the protocol given by Doyle and Doyle, (1987). After the successful CTAB extraction of DNA, agarose gel electrophoresis was performed to visualize the bands and any form of impurities. Further, protein and other contaminants were removed employing the isolated DNA for downstream purification. The Agarose gel electrophoresis revealed distinct and clear bands. The purity or quality of DNA was further checked using DNA quantifying methods where

the DNA was quantified using UV spectrophotometer at 260 nm and 280 nm filters. The ratio of 260/280 equivalent to 1.8 was considered of good quality. DNA of good quality will have an A_{260} /A_{280} ratio of 1.7–2. However, a value of 1.6 does not render the unsuitability of DNA for further analysis or application, but a lower ratio below 1.7 indicates the presence of more contaminants in the DNA. Hence, combining the three steps mentioned above i.e., DNA isolation, DNA purification, and DNA quantification allowed a good amount of pure DNA from the leaves of different clones of tea (*Camellia sinensis*) that can be utilized for further PCR amplification.

Tea is an out-crossing species, where the selected elite tea genotypes are vegetatively propagated and later released in the gardens as clonal varieties. Identification of the clones has been mostly done in a traditional way using morphological descriptors like size and shape of the leaf, fruit, etc. However, tea is an outcrossing and heterozygous crop, the different descriptors like the morphological, biochemical, and physiological descriptors show continuous variation with increased plasticity (Lai *et al.,*2001). Morphological traits are influenced by environmental factors like age of the plant and phenology which makes use of such descriptors difficult in identification and discernment of genetic diversity. Therefore, molecular markers such as RAPD, ISSR, AFLP, etc. can be utilized to study the genome directly since the molecular markers are least affected by environmental influence thus eliminating the limitations of a phenotype observation (Lai *et al.,*2001).

4.2.1. RAPD - Random Amplified Polymorphic DNA

RAPD marker developed by Williams *et al.*(1990) has been used extensively to identify and characterize cultivars and is also used to determine the genetic diversity within and between cultivated tea and other related species of *Camellia* (Lai *et al.,*2001). RAPD is a technique used routinely to evaluate the genetic relationship of variation among the species, varieties, and cultivars. Evaluation of genetic diversity using DNA molecular markers unlike the morpho-anatomical is the easiest when it comes to studying the genetic variation or relatedness among the species or cultivars (Roy and Chakraborty, 2009). Previous reports suggest the use of molecular markers like AFLP (Paul *et al.,*1997), RFLP (Devarumath *et al.,*2002), RAPD (Wachira *et al.,*1995; Kaundun *et al*., 2002; Roy and Chakraborty,2007) and

Table 4.3. Total number of amplified bands, number of polymorphic and monomorphic bands and percentage of polymorphism generated by the RAPD primers

SL no.	PRIMER	TB	PB	MB	POL%
1.	OPA01	24	24	0	100%
2.	OPA02	16	16	0	100%
3.	OPA03	22	22	0	100%
4.	OPA04	18	18	0	100%
5.	OPA05	18	18	0	100%
6.	OPA06	13	13	0	100%
7.	OPA07	21	21	0	100%
8.	OPA08	22	22	0	100%
9.	OPA09	20	20	0	100%
10.	OPA10	22	22	0	100%
11.	OPA11	15	15	0	100%
12.	OPA12	16	16	0	100%
13.	OPA13	15	15	0	100%
14.	OPA14	14	14	0	100%
15.	OPA15	14	13	1	92.90%
16.	OPA16	9	9	0	100%
17.	OPA17	13	11	2	84.61%
18.	OPA18	17	17	0	100%
19.	OPA19	24	24	0	100%
20.	OPA20	21	21	0	100%
21.	OPB01	23	23	0	100%
22.	OPB02	14	14	0	100%
23.	OPB03	15	15	0	100%
24.	OPB04	15	15	0	100%
25.	OPB05	25	25	0	100%
26.	OPB06	13	13	0	100%
27.	OPB07	23	23	0	100%
28.	OPB08	16	16	0	100%
29.	OPB09	22	22	0	100%
30.	OPB10	26	26	0	100%
31.	OPB11	11	11	0	100%
32.	OPB12	16	16	0	100%
33.	OPB13	21	21	0	100%
34.	OPB14	20	20	0	100%
35.	OPB15	23	23	0	100%
36.	OPB17	18	18	0	100%
37.	OPB18	17	17	0	100%
38.	OPB19	19	19	0	100%
39.	OPB20	22	22	0	100%
40.	OPF09	18	18	0	100%
41.	OPG19	14	14	0	100%
42.	OPN13	14	14	0	100%
43.	OPH04	18	18	0	100%
44.	OPN04	17	16	1	94.11%
45.	OPN19	13	13	0	100%
	Total	**807**	**803**	**4**	**99.50%**

TB= total bands, PB = polymorphic bands, MB = monomorphic bands, POL% = polymorphism percentage

ISSR (Mingzhe *et al.,* 2010) to study the genetic variation within the tea germplasm. Random Amplified Polymorphic DNA (RAPD) markers have been efficiently utilized earlier to detect genetic variation or establish a genetic relationship in 25 tea genotypes comprising of garden clones, TRA garden series, Darjeeling clones, and some novel clones (Baruah *et al*., 2010).

In our study, a total of 46 different decamer primers have been employed to study the genetic diversity of 33 tea accessions. (Table 4.3). Out of the 46

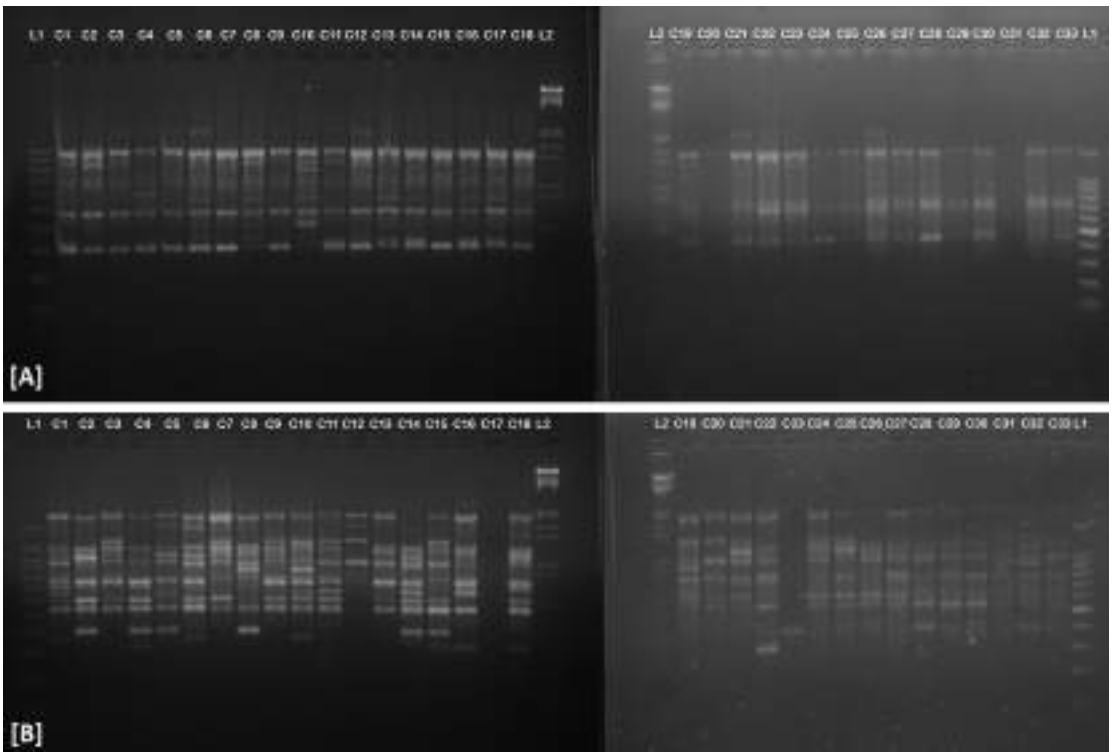

Fig.4.3. RAPD screening: L1- 100 bp DNA ladder, L2- : λ DNA/EcoRI/HindIII double digest DNA ladder, 1-33-tea clones; (A) amplification with OPB06 (B) amplification with OPB15

RAPD primers screened, 45 primers resulted in distinct and scorable bands .

RAPD screening rendered a total of 807 bands out of which 4 bands were monomorphic and the remaining 803 bands showed 100% polymorphism (Table 4.3). The genetic diversity of 27 superior tea germplasms from three different regions i.e., Japan, Korea, and Taiwan were examined using 50 RAPD primers where only 17 primers generated 58 polymorphic bands (Kaundun *et al.,* 2000). The 45 random primers showed a polymorphism percentage of 99.50% and an average PIC value of 0.41. Earlier, 10 random primers used to screen genetic diversity among tea genotypes showed 51.47% polymorphism and a PIC value of 0.356 (Baruah *et al*., 2010). The number of polymorphic bands ranged between 9(OPA16) to 26 (OPB10) (Table 4.3). The RAPD profile of the tea clones generated using primers OPB06 (A) and OPB15 (B) is given in Figure 4.3.

4.2.2. Intersimple Sequence Repeat (ISSR) analysis

ISSR markers developed by Zietkiewicz *et al.* (1994) can also be employed for the same purpose i.e., to study genetic diversity (Lai *et al.,*2001). A total of 15 ISSR primers were further employed to study the genetic variation among the 33 tea clones. All the primers amplified

Table 4.4. Total number of amplified bands, number of polymorphic and monomorphic bands and percentage of polymorphism generated by the ISSR primers

Sl No.	PRIMER	TB	PB	MB
1	UBC807	16	16	0
2	UBC808	14	14	0
3	UBC810	20	20	0
4	UBC811	17	17	0
5	UBC818	15	15	0
6	UBC822	17	17	0
7	UBC824	17	17	0
8	UBC825	28	28	0
9	UBC834	23	23	0
10	UBC836	19	19	0
11	UBC841	16	16	0
12	UBC856	26	26	0
13	UBC873	27	27	0
14	UBC813	21	21	0
15	UBC815	22	22	0
		298	**298**	**0**

TB= total bands, PB = polymorphic bands, MB = monomorphic bands, *100% polymorphism

distinct and scorable bands. Previously twenty 3'-anchored ISSR primers were screened among tea, out of which six primers produced clear and reproducible bands (Lai *et al.,*2001). A total of 298 bands were generated out of which all the bands were polymorphic showing 100% polymorphism (Table 4.4) with an average PIC value of 0.42. The polymorphic bands ranged between 14 (UBC808) to 28(UBC825). The ISSR profile of the studied clones amplified using the primer UBC818 (A), UBC10 (B) is given in Figure 4.4.

4.2.3. DNA fingerprinting analysis

The genetic relatedness is depicted in the dendrogram given (Figure 4.5) where we can see the tea samples clustered into two main groups and many other subgroups showing genetic diversity. The highest similarity was observed between C22 (P-1258) and C23 (RR4/5) with a nodal value of 0.84. The same nodal value of similarity was also shared by the clones C28(T-246) and C29(TV-19). The dendrogram revealed genetic variation within the studied sample. The results were further supported or validated by the 2D (Supplementary Figure 4.1) and 3D (Supplementary Figure 4.2) plot which revealed two major clusters thus pointing towards genetic diversity within the studied clones. A similarity matrix was further drawn using the Dice coefficient of similarity (Nei and Li, 1979) ranging from 0.581 to 0.844 (Figure 4.6).

The ISSR primers also revealed similar kinds of results where the clones studied clustered into two major groups and other subgroups as depicted by the dendrogram (Figure 4.7).

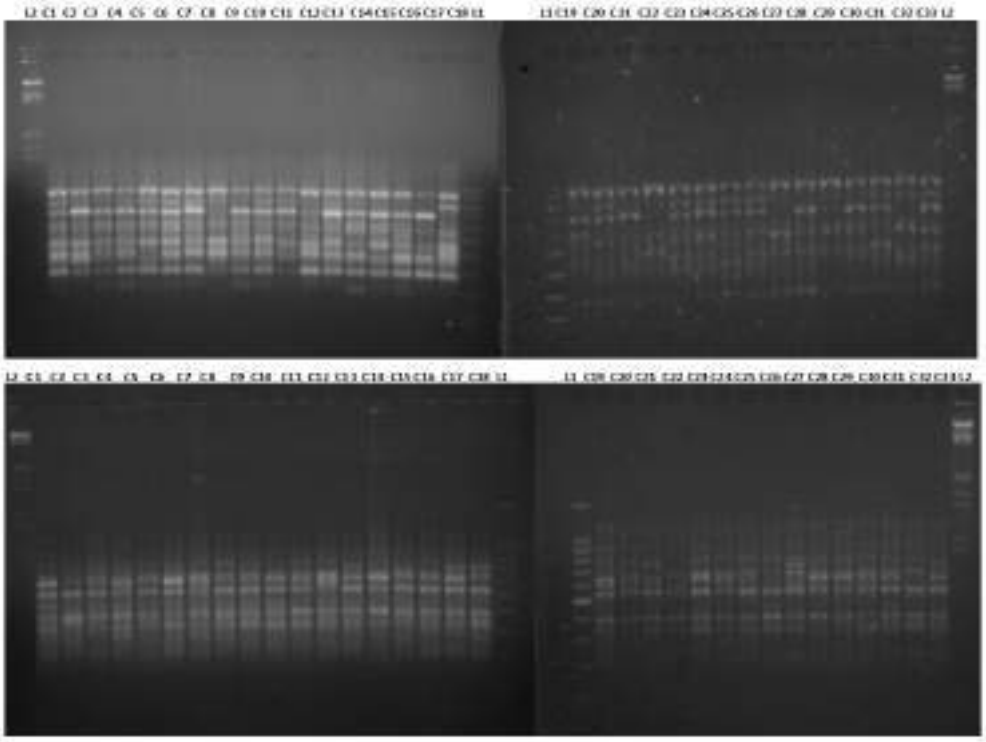

Fig.4.4. ISSR screening: L1- 100 bp DNA ladder, L2- : λ DNA/EcoRI/HindIII double digest DNA ladder, 1-33-tea variety; (A) amplification with UBC818 (B) amplification with UBC10

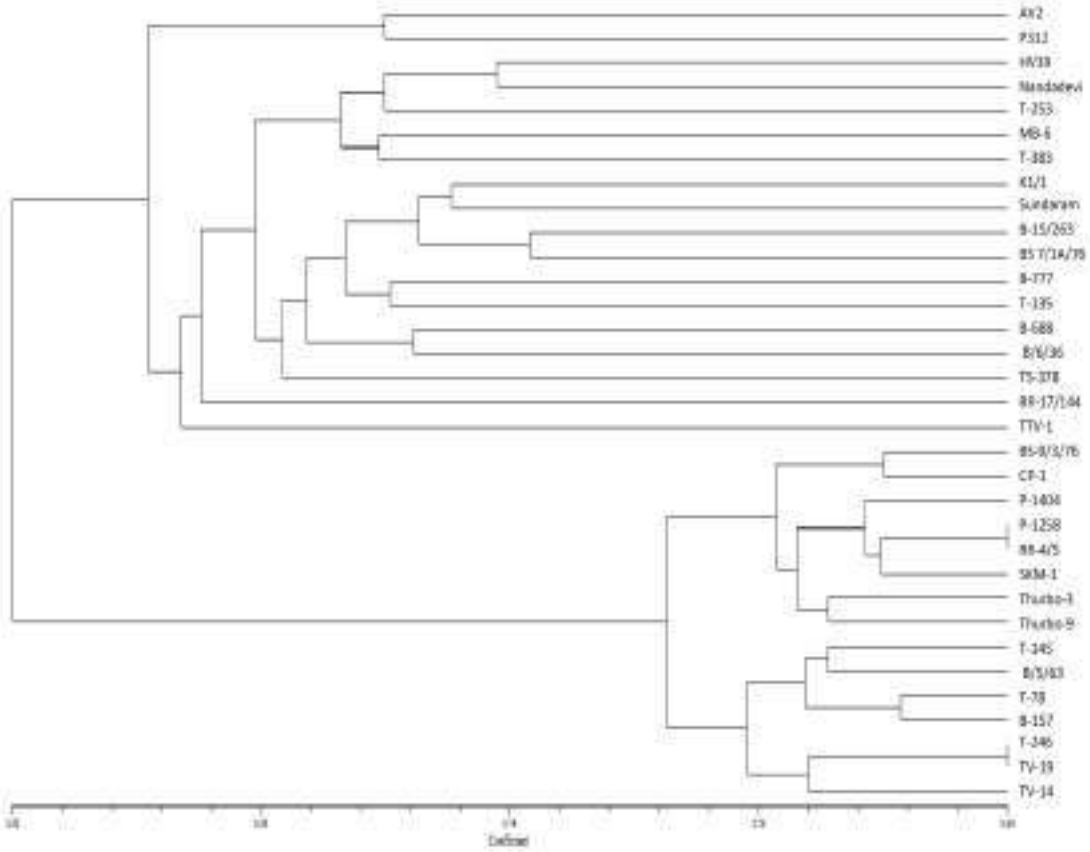

Fig.4.5. Dendrogram obtained from UPGMA cluster analysis of RAPD markers illustrating the genetic relationships among the 33 tea clones

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33
C1	1.00																																
C2	0.71	1.00																															
C3	0.69	0.69	1.00																														
C4	0.68	0.69	0.71	1.00																													
C5	0.66	0.70	0.73	0.70	1.00																												
C6	0.67	0.69	0.70	0.69	0.71	1.00																											
C7	0.64	0.68	0.69	0.65	0.66	0.68	1.00																										
C8	0.63	0.65	0.70	0.67	0.72	0.71	0.65	1.00																									
C9	0.67	0.67	0.72	0.66	0.69	0.67	0.68	0.71	1.00																								
C10	0.63	0.67	0.70	0.67	0.66	0.70	0.67	0.66	0.71	1.00																							
C11	0.64	0.66	0.71	0.70	0.71	0.70	0.68	0.71	0.72	0.74	1.00																						
C12	0.64	0.66	0.67	0.67	0.67	0.68	0.66	0.68	0.71	0.68	0.70	1.00																					
C13	0.64	0.64	0.66	0.67	0.66	0.68	0.68	0.68	0.72	0.72	0.71	0.71	1.00																				
C14	0.65	0.66	0.69	0.69	0.67	0.68	0.66	0.69	0.68	0.70	0.69	0.71	0.71	1.00																			
C15	0.63	0.65	0.65	0.66	0.69	0.67	0.64	0.65	0.69	0.67	0.69	0.68	0.70	0.69	1.00																		
C16	0.62	0.65	0.69	0.68	0.66	0.68	0.66	0.68	0.67	0.70	0.67	0.69	0.70	0.68	0.68	1.00																	
C17	0.65	0.66	0.69	0.69	0.65	0.68	0.65	0.65	0.69	0.68	0.71	0.69	0.72	0.70	0.68	0.71	1.00																
C18	0.61	0.63	0.65	0.66	0.66	0.65	0.63	0.65	0.65	0.68	0.66	0.68	0.69	0.68	0.68	0.66	0.70	1.00															
C19	0.61	0.64	0.66	0.60	0.67	0.62	0.64	0.61	0.63	0.64	0.66	0.62	0.61	0.62	0.60	0.62	0.63	0.64	1.00														
C20	0.64	0.65	0.66	0.64	0.66	0.63	0.63	0.61	0.64	0.66	0.65	0.62	0.61	0.62	0.61	0.61	0.62	0.63	0.82	1.00													
C21	0.63	0.62	0.64	0.63	0.65	0.63	0.63	0.61	0.63	0.64	0.65	0.64	0.63	0.63	0.60	0.63	0.63	0.63	0.79	0.82	1.00												
C22	0.62	0.64	0.64	0.65	0.66	0.64	0.64	0.61	0.64	0.67	0.66	0.64	0.62	0.64	0.63	0.62	0.64	0.65	0.80	0.82	0.82	1.00											
C23	0.61	0.61	0.62	0.61	0.65	0.61	0.60	0.61	0.61	0.62	0.63	0.62	0.59	0.59	0.58	0.61	0.60	0.61	0.79	0.81	0.81	0.84	1.00										
C24	0.63	0.63	0.64	0.62	0.64	0.63	0.64	0.62	0.62	0.64	0.65	0.63	0.60	0.62	0.61	0.61	0.61	0.63	0.78	0.78	0.81	0.82	0.82	1.00									
C25	0.60	0.63	0.63	0.62	0.66	0.63	0.62	0.62	0.61	0.65	0.65	0.64	0.63	0.63	0.61	0.62	0.61	0.63	0.78	0.81	0.80	0.80	0.80	0.80	1.00								
C26	0.61	0.60	0.62	0.61	0.63	0.62	0.59	0.61	0.61	0.62	0.63	0.62	0.61	0.62	0.59	0.58	0.61	0.61	0.76	0.80	0.78	0.79	0.79	0.81	0.80	1.00							
C27	0.62	0.60	0.64	0.62	0.66	0.65	0.60	0.62	0.63	0.65	0.65	0.63	0.62	0.62	0.61	0.61	0.63	0.62	0.74	0.78	0.76	0.76	0.79	0.79	0.79	0.80	1.00						
C28	0.62	0.62	0.64	0.63	0.62	0.60	0.61	0.60	0.62	0.64	0.65	0.63	0.63	0.63	0.59	0.61	0.62	0.62	0.76	0.76	0.77	0.78	0.78	0.75	0.78	0.78	0.79	1.00					
C29	0.64	0.63	0.65	0.63	0.66	0.64	0.62	0.64	0.63	0.64	0.65	0.64	0.62	0.63	0.60	0.61	0.63	0.63	0.76	0.78	0.78	0.79	0.80	0.79	0.80	0.77	0.81	0.84	1.00				
C30	0.61	0.61	0.67	0.64	0.66	0.63	0.63	0.62	0.61	0.63	0.64	0.62	0.60	0.64	0.58	0.62	0.62	0.61	0.74	0.77	0.75	0.77	0.80	0.76	0.77	0.74	0.79	0.78	0.82	1.00			
C31	0.61	0.63	0.64	0.62	0.65	0.62	0.61	0.60	0.61	0.66	0.63	0.63	0.61	0.62	0.59	0.62	0.62	0.62	0.74	0.75	0.76	0.77	0.77	0.79	0.80	0.76	0.79	0.76	0.80	0.79	1.00		
C32	0.62	0.64	0.66	0.64	0.67	0.62	0.62	0.64	0.63	0.64	0.65	0.64	0.62	0.65	0.61	0.64	0.64	0.63	0.73	0.78	0.75	0.75	0.77	0.77	0.79	0.77	0.80	0.78	0.77	0.78	0.82	1.00	
C33	0.62	0.62	0.65	0.64	0.66	0.63	0.63	0.60	0.62	0.62	0.64	0.64	0.63	0.62	0.61	0.63	0.63	0.61	0.74	0.75	0.76	0.76	0.77	0.77	0.79	0.75	0.80	0.77	0.81	0.80	0.80	0.80	1.00

Fig.4.6. Similarity matrix of RAPD analysis calculated using Dice coefficient (Nei and Li, 1979) based on band-sharing from the binary data

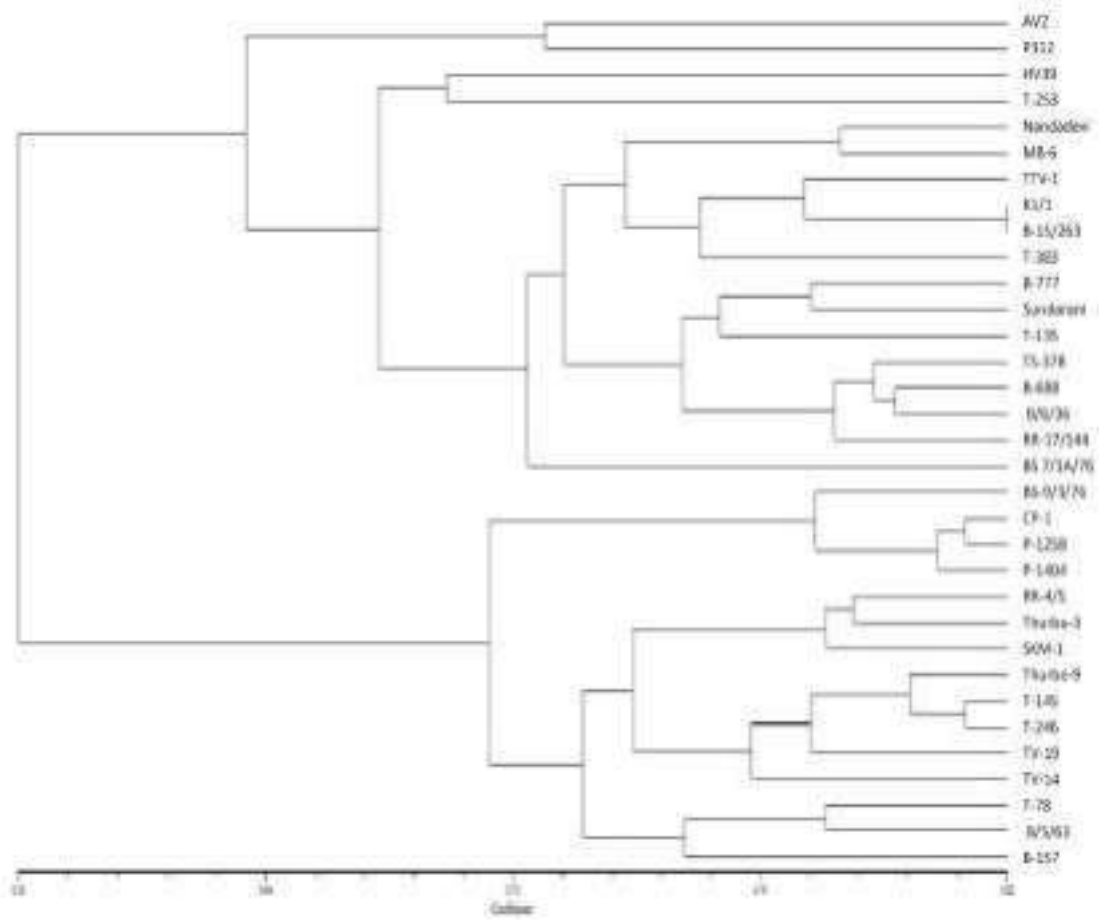

Fig.4.7. Dendrogram obtained from UPGMA cluster analysis of ISSR markers illustrating the genetic relationships among the 33 tea clones

The highest similarity was observed between C9 (K1/1) and C10 (B-15/263) with a nodal value of 0.82. Our findings were corroborated by 2D (Supplementary Figure 4.3) and 3D plots (Supplementary Figure 4.4). A similarity value ranging from 0.500 to 0.817 was observed from the similarity matrix drawn (Figure 4.8).

The results of RAPD and ISSR profiling more or less corroborated by clustering into broad two groups and many other subgroups. However, the pattern of sub-grouping does not reveal as such a clear pattern of grouping irrespective of the sample source, sampling area, or the place of selection or collection. Our findings thus suggest an extensive exchange of genetic resources between the sampling regions. A minimum genetic diversity was recorded previously among the tea populations irrespective of the place of selection or collection (Baruah *et al*., 2010). In our study ISSR fingerprinting revealed more polymorphism than RAPD fingerprinting. A similar kind of result was reported earlier (Devarumath *et al.,* 2002) where ISSR profiling detected more polymorphic loci than RAPD profiling.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32	C33
C1	1.00																																
C2	0.71	1.00																															
C3	0.66	0.69	1.00																														
C4	0.67	0.63	0.68	1.00																													
C5	0.63	0.67	0.70	0.71	1.00																												
C6	0.66	0.66	0.66	0.74	0.78	1.00																											
C7	0.59	0.68	0.68	0.68	0.75	0.78	1.00																										
C8	0.60	0.66	0.69	0.70	0.70	0.70	0.74	1.00																									
C9	0.58	0.66	0.67	0.70	0.70	0.75	0.80	0.76	1.00																								
C10	0.61	0.67	0.67	0.73	0.71	0.72	0.74	0.73	0.82	1.00																							
C11	0.58	0.62	0.66	0.68	0.66	0.66	0.70	0.72	0.72	0.75	1.00																						
C12	0.60	0.63	0.67	0.66	0.69	0.73	0.73	0.70	0.77	0.70	0.73	1.00																					
C13	0.61	0.67	0.62	0.63	0.67	0.68	0.71	0.72	0.74	0.76	0.68	0.77	1.00																				
C14	0.63	0.65	0.70	0.63	0.69	0.70	0.71	0.70	0.74	0.72	0.72	0.74	0.76	1.00																			
C15	0.59	0.65	0.63	0.62	0.71	0.68	0.71	0.68	0.74	0.73	0.70	0.72	0.75	0.76	1.00																		
C16	0.58	0.65	0.66	0.60	0.68	0.68	0.73	0.66	0.70	0.71	0.70	0.76	0.72	0.71	0.79	1.00																	
C17	0.62	0.66	0.65	0.62	0.68	0.69	0.71	0.71	0.72	0.76	0.69	0.73	0.77	0.75	0.78	0.79	1.00																
C18	0.61	0.66	0.67	0.66	0.70	0.70	0.73	0.70	0.72	0.75	0.71	0.73	0.74	0.73	0.78	0.77	0.78	1.00															
C19	0.58	0.57	0.61	0.57	0.60	0.56	0.60	0.56	0.61	0.61	0.59	0.59	0.61	0.63	0.62	0.60	0.61	0.58	1.00														
C20	0.58	0.57	0.58	0.60	0.59	0.58	0.58	0.55	0.58	0.59	0.60	0.57	0.60	0.61	0.58	0.60	0.61	0.58	0.80	1.00													
C21	0.58	0.58	0.60	0.59	0.60	0.55	0.59	0.54	0.59	0.60	0.59	0.60	0.63	0.61	0.60	0.62	0.59	0.61	0.77	0.80	1.00												
C22	0.56	0.57	0.54	0.55	0.57	0.59	0.57	0.54	0.56	0.57	0.57	0.57	0.61	0.61	0.58	0.59	0.58	0.60	0.74	0.81	0.80	1.00											
C23	0.55	0.58	0.59	0.58	0.58	0.59	0.58	0.56	0.59	0.59	0.58	0.57	0.59	0.58	0.60	0.62	0.60	0.60	0.71	0.76	0.73	0.78	1.00										
C24	0.61	0.63	0.61	0.56	0.58	0.61	0.58	0.57	0.60	0.61	0.58	0.57	0.60	0.61	0.60	0.59	0.58	0.60	0.68	0.69	0.71	0.77	0.78	1.00									
C25	0.55	0.56	0.55	0.56	0.56	0.56	0.55	0.55	0.56	0.56	0.57	0.54	0.56	0.58	0.58	0.58	0.56	0.54	0.69	0.71	0.68	0.72	0.78	0.77	1.00								
C26	0.56	0.58	0.50	0.56	0.54	0.58	0.56	0.53	0.57	0.60	0.61	0.57	0.60	0.59	0.58	0.61	0.57	0.57	0.67	0.71	0.70	0.72	0.75	0.75	0.77	1.00							
C27	0.56	0.58	0.54	0.54	0.57	0.57	0.56	0.54	0.57	0.56	0.54	0.55	0.60	0.60	0.58	0.58	0.56	0.56	0.67	0.70	0.71	0.75	0.77	0.75	0.77	0.79	1.00						
C28	0.56	0.59	0.56	0.60	0.57	0.58	0.56	0.54	0.55	0.59	0.58	0.53	0.58	0.59	0.56	0.59	0.56	0.58	0.65	0.70	0.71	0.72	0.69	0.72	0.74	0.79	0.81	1.00					
C29	0.57	0.58	0.57	0.59	0.61	0.60	0.59	0.56	0.59	0.63	0.58	0.58	0.59	0.58	0.61	0.61	0.60	0.58	0.64	0.69	0.70	0.69	0.68	0.68	0.68	0.74	0.77	0.80	1.00				
C30	0.58	0.63	0.61	0.61	0.62	0.62	0.63	0.60	0.64	0.61	0.64	0.61	0.64	0.59	0.63	0.60	0.57	0.60	0.65	0.69	0.69	0.69	0.71	0.74	0.72	0.74	0.76	0.76	0.76	1.00			
C31	0.53	0.60	0.57	0.59	0.61	0.60	0.57	0.56	0.58	0.62	0.60	0.55	0.59	0.58	0.60	0.60	0.62	0.58	0.65	0.70	0.66	0.70	0.72	0.74	0.73	0.76	0.78	0.76	0.70	0.76	1.00		
C32	0.53	0.52	0.53	0.54	0.57	0.55	0.56	0.54	0.59	0.60	0.59	0.55	0.58	0.56	0.58	0.61	0.59	0.57	0.61	0.68	0.65	0.67	0.65	0.66	0.69	0.70	0.72	0.73	0.72	0.72	0.74	1.00	
C33	0.53	0.53	0.56	0.59	0.59	0.59	0.57	0.55	0.60	0.61	0.57	0.56	0.61	0.58	0.57	0.59	0.60	0.58	0.65	0.70	0.68	0.71	0.68	0.69	0.70	0.70	0.71	0.73	0.70	0.73	0.77	0.74	1.00

Fig.4.8. Similarity matrix of ISSR analysis calculated using Dice coefficient (Nei and Li, 1979) based on band-sharing from the binary data

4.3. DNA barcode

The advancement of sequencing technology has uplifted molecular research, and even a minor difference whether being intraspecific or interspecific can be detected which would otherwise not be easily accomplished using other morphological or robust molecular techniques. DNA barcoding utilizes numerous plastid and nuclear regions to identify species. Seven different plastid regions have been explored in land plants where they have suggested the combination of rbcL+matK to be most precise and appropriate for plant barcode analysis (Group *et al.*, 2009). Successful utilization of a combination of matK+ITS and rbcL+trnH-psbA to study 100% differences between *Cassia* species has already been reported in previous studies (Purushothaman *et al.*, 2014). The matK region solely has been used efficiently to differentiate *Vachellia* species from other *Acacia* species (Steven *et al.*, 2009), where the study highlighted the possibility of utilizing matK for separating taxa at the genus level. Some previous studies have highlighted the successful amplification and utilization of the matK region to explore the phylogeny in both monocots and dicots such as Zingiberaceae (Selvaraj *et al.*, 2008), Erythronium (Allen *et al.*, 2003), *Myristica fragrans* (Tallei and Kolondam, 2015), local tomato (Caprar *et al.*, 2017) and oil-bearing roses (Wang *et al.*, 2012). Apart from species recognition or variation study, nowadays the barcode technique is also used to detect other contaminants. For instance, one study reports the origin of bamboo leaves as well as the presence of adulterants in manufactured bamboo tea products. The adulterants present in bamboo tea have counter-indications for pregnant women in products (Horn and Haser, 2016). The DNA barcodes like rbcL, matK, ITS2, and psbA-trnH have been used previously to differentiate between the commercial non-*Camellia* tea and the adulterants present in it just to ensure its safety. However, a limited number of original plant sequences in GenBank limited the findings of the study (Long *et al.*, 2014).

4.3.1. matK Amplification and Sequencing

The primer successfully amplified the matK (Figure 4.9) region of all the 33 tea clones under investigation. The size of the PCR product amplified approximately ranged between 900 bp –1000 bp. However, the sequencing result deduced the size of the final amplified product ranging from 644 bp to 876 bp. The accession number for the submitted sequences is provided in Table 4.5 & Appendix- C.

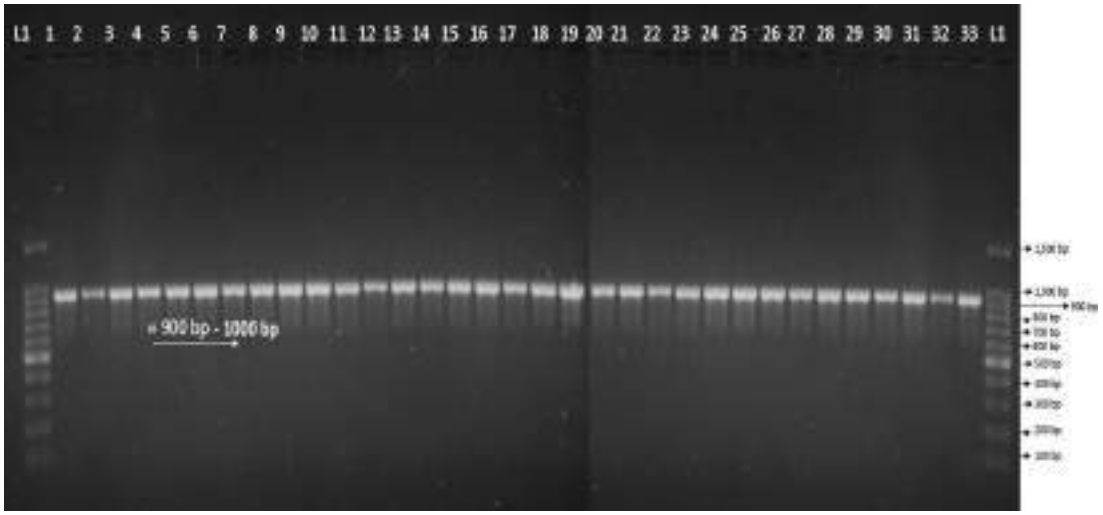

Fig.4.9. Amplification of the matK region. Lane L1- 100 bp DNA ladder; Lane 1-33: 33 tea clones.

Table 4.5. Accession number and details of the submitted matK sequence retrieved from NCBI

Sl. no	Full name	Abbreviation	NCBI accession	Unique Id	Base pairs
1.	Ambari Vallai 2	AV2	MH649284	1111	757bp
2.	Phoobsering 312	P312	MK393394	1112	871bp
3.	Happy valley 39	HV39	MH791417	1113	864bp
4.	Tukdah 253	T-253	MH920315	1114	876bp
5.	Nanda Devi	TS 378	MH920316	1115	816bp
6.	Makaibari-6	MB-6	MH920317	1116	758bp
7.	Teesta Valley 1	TTV-1	MH920318	1117	861bp
8.	Kopati 1/1	K1/1	MH920319	1119	774bp
9.	Balasun 7/1A/76	BS 7/1A/76	MK393393	11111	833bp
10.	Bunnockburn 777	B-777	MK393395	11112	757bp
11.	Sundaram	B/5/63	MK393396	11113	644bp
12.	Tukdah-135	T-135	MK393397	11114	763bp
13.	Bunnockburn 688	B-688	MK393398	11116	833bp
14.	Golconda	Golconda	MK393399	11117	644bp
15.	RungliRungiliot 17/144	RR-17/144	MK393400	11118	826bp
16.	Balasun 9/3/76	BS-9/3/76	MK393401	11119	756bp
17.	Chiradew Parbat1	CP-1	MK393402	11120	761bp
18.	Phoobsering 1404	P-1404	MK393403	11121	763bp
19.	Phoobsering 1258	P-1258	MK393404	11122	751bp
20.	RungliRungiliot 4/5	RR-4/5	MK393405	11123	746bp
21.	Sikkim 1	SKM-1	MK424865	11124	867bp
22.	Thurbo 3	Thurbo-3	MN480321	11125	761bp
23.	Thurbo 9	Thurbo-9	MN480322	11126	707bp
24.	Tukdah 145	T-145	MK424866	11127	761bp
25.	Tukdah 246	T-246	MK424867	11128	756bp
26.	Tocklai variety 19	TV-19	MK424868	11129	750bp
27.	Tocklai variety 14	TV-14	MK424869	11130	735bp
28.	Tukdah 78	T-78	MK424870	11131	755bp
29.	Bannockburn 157	B-157	MK424871	11132	735bp

4.3.2. Blast Result

As per the blast results, a total of 24 clones out of 29 clones was found to be 100% identical with *Camellia sinensis*. The percentage of similarity with *Camellia sinensis* as shown by the remaining five clones were 99.29% (Thurbo 9), 99.61% (Thurbo 3), 99.64% (RR-4/5), 99.88% (SKM-1), and 99.89% (P312). However, the clones like Thurbo 3 and Thurbo 9 showed a relatively higher percentage of similarity with other species of *Camellia*, i.e., 99.87% (Thurbo 3) and 99.57% (Thurbo 9) similarity with *Camellia mairei* (KJ197933.1). The clones were therefore placed under different species of *Camellia* with a percentage similarity value below 99.64%. The matK locus placed two genera *Myristica* and *Knena* differently at a sequence similarity of 99.43% while genus *Virola* differed with 99.25% (Tallei and Kolondam, 2015). With a similarity percentage of 99.64%, tomatoes were placed within the same species contradicting the assumption of required percent identity as 99.74 –100% to place the organisms within the same species (Lawodi, 2013). Our study has differentiated two species at percent identity below 99.64% where the clones like Thurbo 3 showed 99.61% and Thurbo 9 showed 99.29% identity with *C. sinensis*. However, our study showed 99.87% (Thurbo 3) and 99.57% (Thurbo 9) identity with *C. mairei* which is comparatively higher than the identity shared with *C. sinensis*.

4.3.3. Sequence Alignment and Phylogenetic Tree Construction

The phylogenetic tree constructed using neighbor-joining (Figure 4.10) and UPGMA (Supplementary Figure 4.5) method depicted variation among the sequences. The combined nucleotide sequences of almost all the clones clustered together except for clones like Thurbo 3 (11125), Thurbo 9 (11126), P312 (1112), and RR-17/144 (11118). The clones like Thurbo 3 and Thurbo 9 clustered together whereas the clones P312 and RR-17/144 diverged from the main group. To validate our results, a phylogenetic tree was further constructed considering a reference matK sequence of different *Camellia* species taken from NCBI (KJ197933.1). The clones like Thurbo 3 (11125) and Thurbo 9 (11126) are now clustered with the reference sequence of *C. mairei* (KJ197933.1) as shown in the neighbor-joining (Figure 4.11) and UPGMA tree (Supplementary Figure 4.6).

4.3.4. Sequence Analysis

The genetic distances for the matK sequence as calculated using Nucleotide: Maximum composite likelihood method and Nucleotide

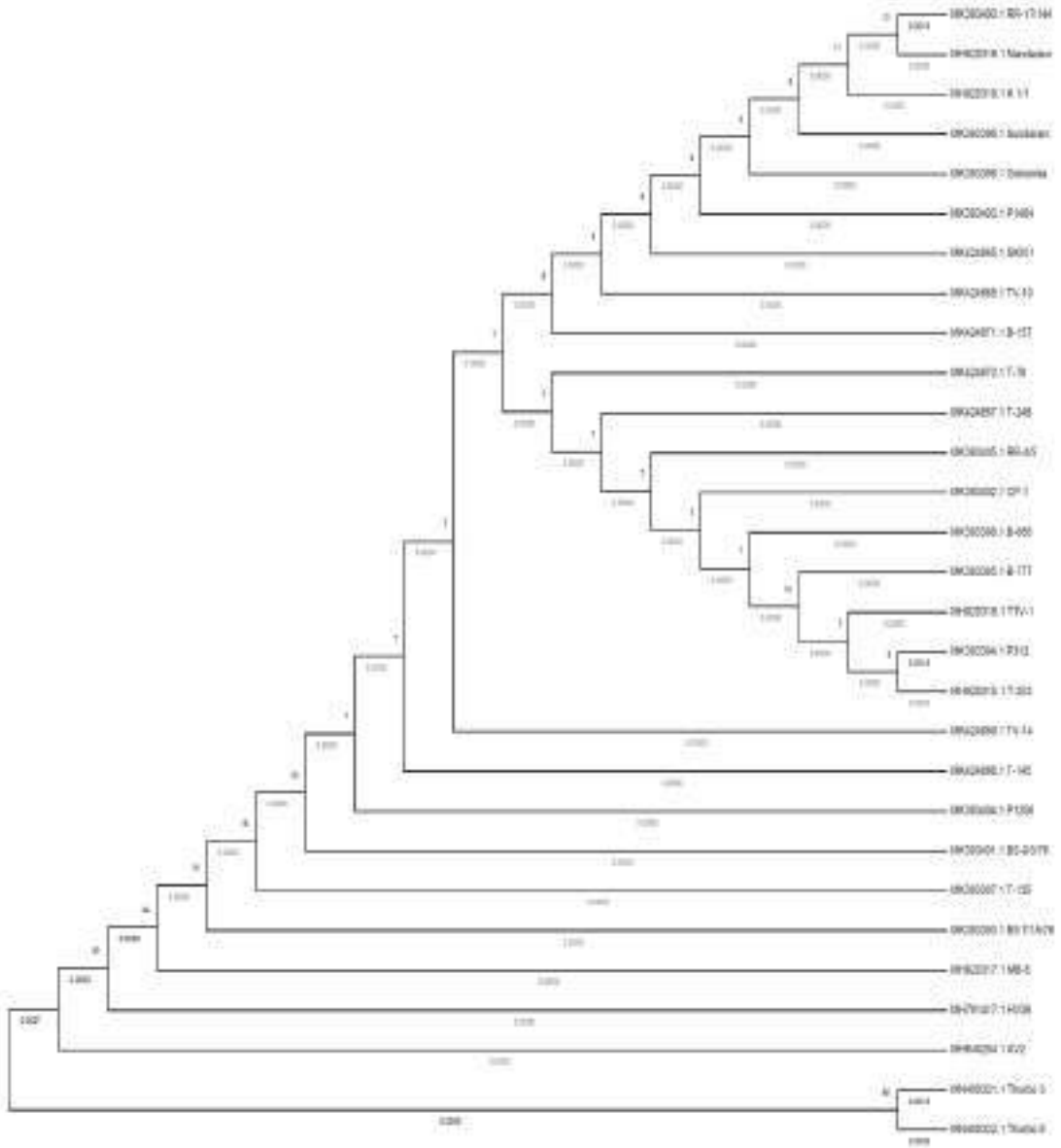

Fig.4.10. Neighbor joining tree method showing the genetic relationship of matK region between 29 tea clones

Kimura 2-parameter method ranged between 0 to 0.0090 (Figure 4.12) and 0 to 0.0089 (Supplementary Figure 4.7) respectively. The overall calculated mean distance was 0.0013. The results show the number of base substitutions per site and are based on an analysis of a total of 29 sequences with all codon positions included and with a total of 563 positions in the final dataset excluding the eliminated positions containing gaps and missing data. The matK sequence of Thurbo 3, Thurbo 9, and *C. mairei* shared two unique variable sites that differed from the rest of the sequences. It was further validated by analyzing a high consensus sequence of 563 bp prepared using Multalin software. The high consensus region showed multiple

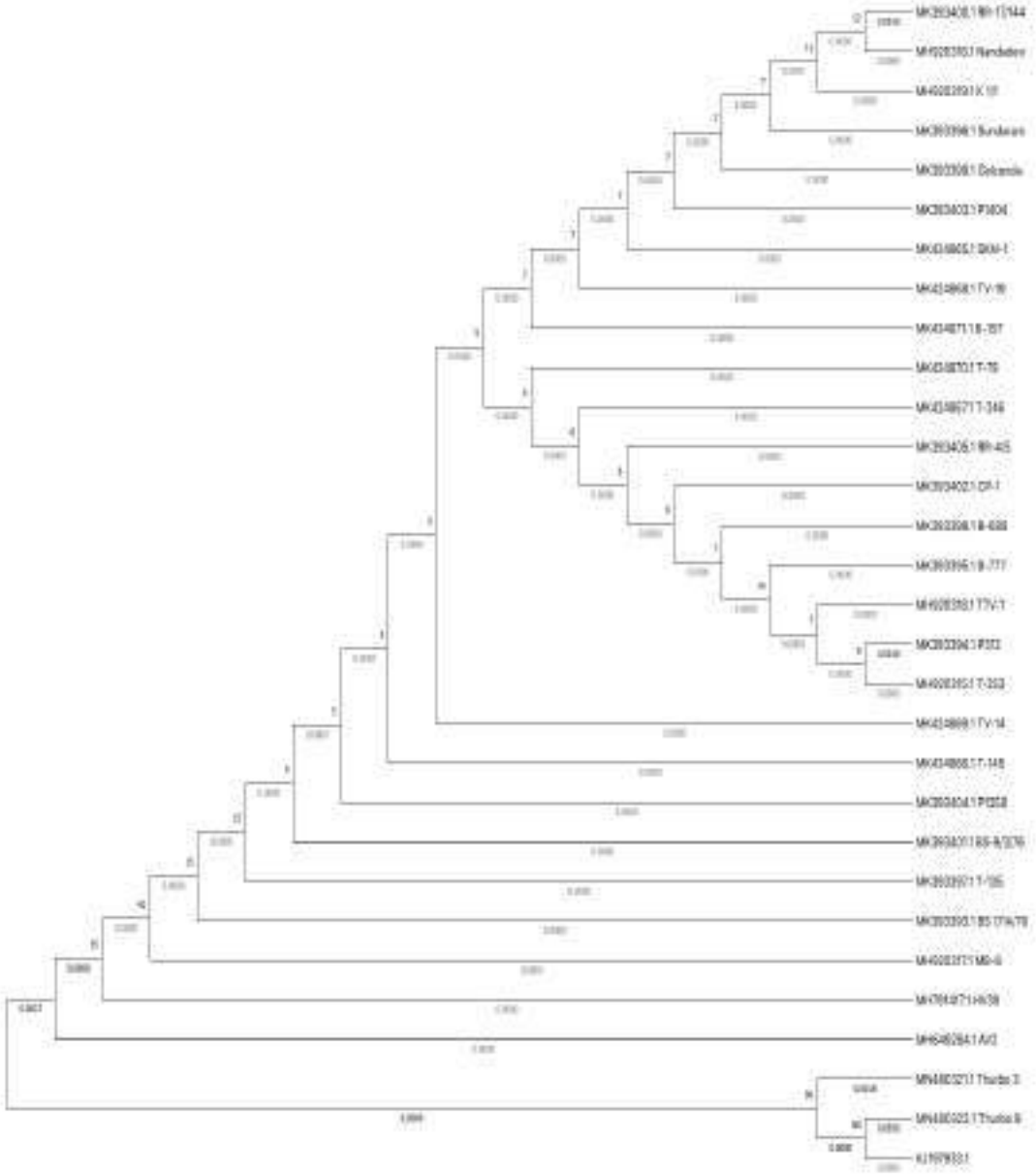

Fig.4.11. Neighbour joining tree method showing the genetic relationship of matK region between 29 tea clones along with sequence of *Camellia mairei* (KJ197933.1) taken from NCBI

variable sites, where Thurbo 3 (11125) showed 3 substitutions, Thurbo 9 (11126) showed 4 substitutions, and some single substitutions were found in P312 (1112) and RR-17/144 (11118) as shown in Figure 4.13 and Supplementary material 1 (SM1). Similarly, the low consensus region also revealed a total of five nucleotide substitutions or variations in clones like SKM-1 (3 substitutions), Thurbo 9 (1 substitution), and P1258 (1 substitution) as shown in Supplementary material 2 (SM2). Earlier, two species *T. cope* and

T. wightii (Ragupathy *et al.*, 2009) did not show variation at rbcl locus but

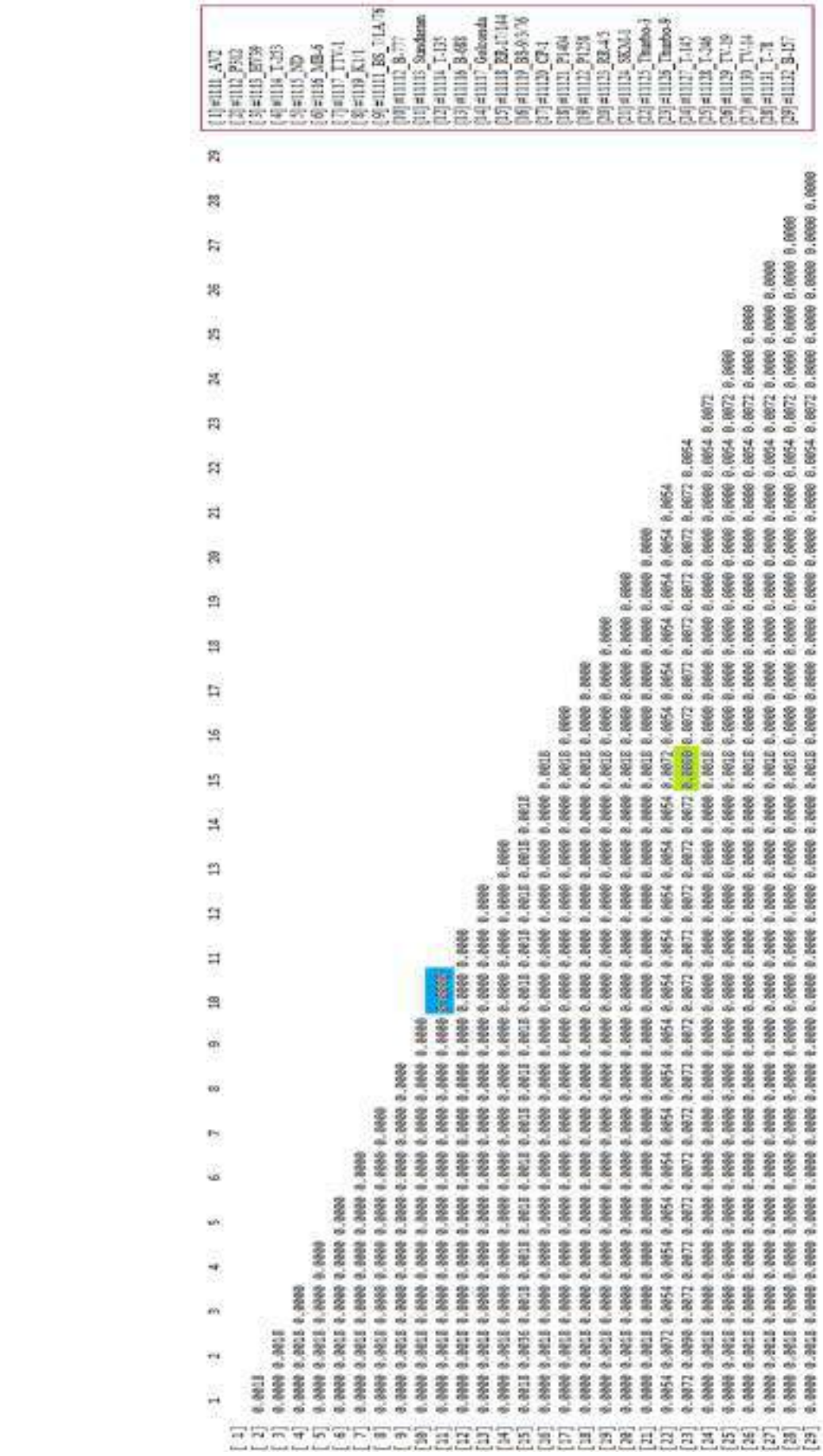

Fig.4.12. Genetic distances of the matK sequence calculated using Nucleotide: Maximum Composite Likelihood method.

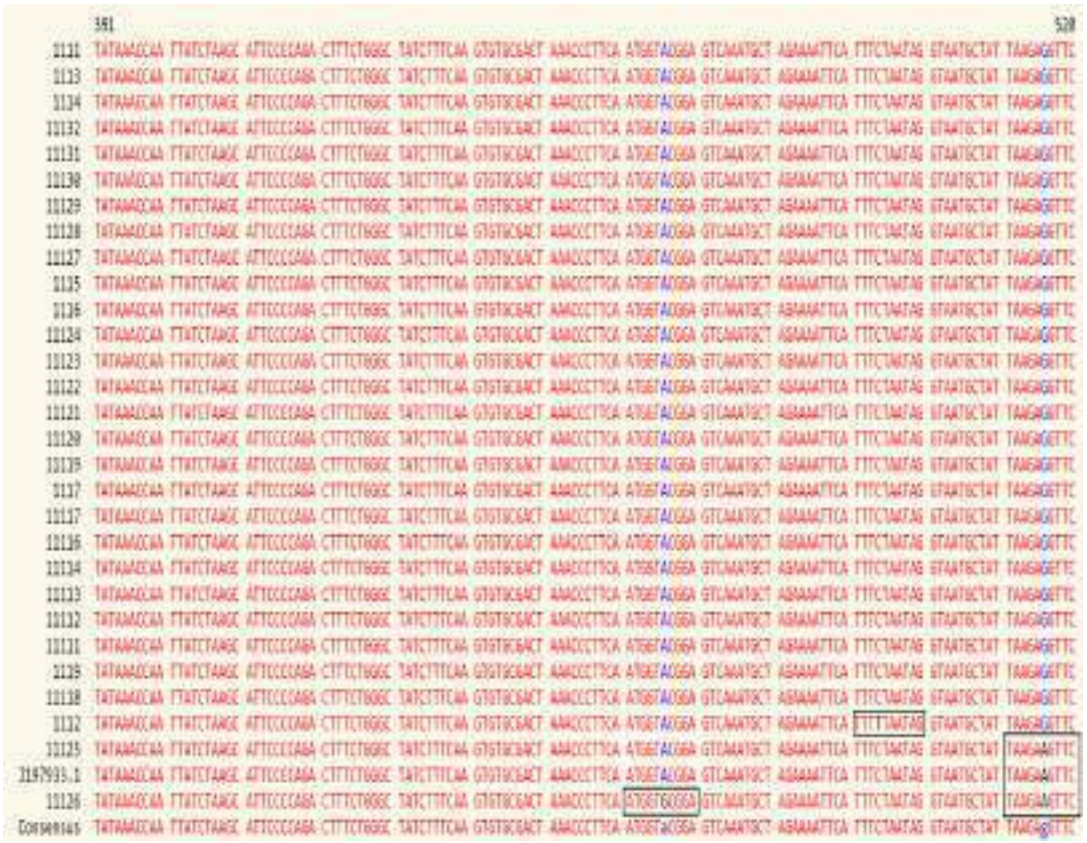

Fig.4.13. Consensus region of the aligned sequences showing nucleotide substitution as highlighted by black box

showed a variable site in the matK region where it showed 2 nucleotide variation and trnHpsbA showed 1 nucleotide variation. The interspecific variation could be broadened if the two loci are considered as a two-gene approach and therefore they report interspecific variation at p-distance 0.002– 0.003 but with the p-distance value of 0.00, they report the absence of intraspecific variation. A total of three variable sites was reported in trnH-psbA sequences of seven tomato varieties with the genetic distance ranging from 0 to 0.004 (Caprar *et al.*, 2017). On the other hand, the study reported no variable sites in the rbcL, rpoC1, and rpoB sequences thus suggesting it to be 100% conserved within the species. The ability of matK locus to differentiate only at the genus level within the family of Myristicaceae has been reported earlier where the matK locus failed to differentiate *Myristica* at species level showing 100% similarity in blast analysis with other three species of *Myristica* sharing three nucleotide differences with *Rivola sebifera* and four nucleotide differences with *Knema laurina*. However, our study reports a total of nine variable sites in the high consensus region and a total of five variable sites in the low consensus

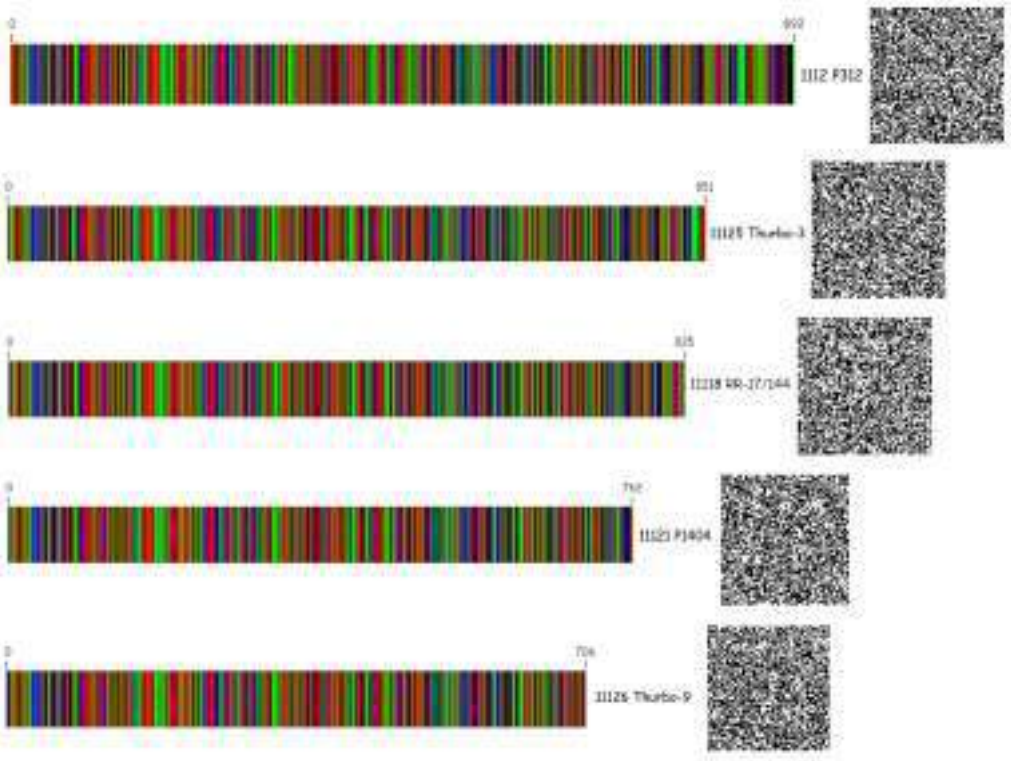

Fig.4.14. Illustrative representation of matK sequences as barcode and QR code

region of matK sequences within the same species of *Camellia sinensis*. Therefore, we report intraspecific variation and conclude with a fact of matK sequence not being 100% conserved within the same species of *Camellia*.

The matK sequence under investigation is represented illustratively as barcode and QR code [Figure 4.14 and Supplementary material 3 (SM3)]. The four different color code in the barcode depicts the different nucleotide composition (A, T, G or C) of the matk sequence. The QR code generated can be decoded as DNA sequences which make data storage and retrieval easy.

Fig 4.15. Herbarium of Thurbo 3 (*C. sinensis*) with the partial sequence information of its matk region enscrypted as QR code

The sequence information can also be enscrypted as QR code along with the herbarium of the specimen as shown in Figure 4.15 .

4.4. Phytochemical screening

4.4.1. Qualititative tests

The three polar solvents i.e., acetone, methanol, and ethanol persistently gave better results in almost all the qualitative tests like tests for cardiac glycoside, steroid, flavonoid, diterpenes, terpenoids, saponin, tannin, and reducing sugar as shown in Figure 4.16. These three solvents showed the highest potency to extract important phytochemicals from tea leaves. Extracting solvent and its nature (polar/

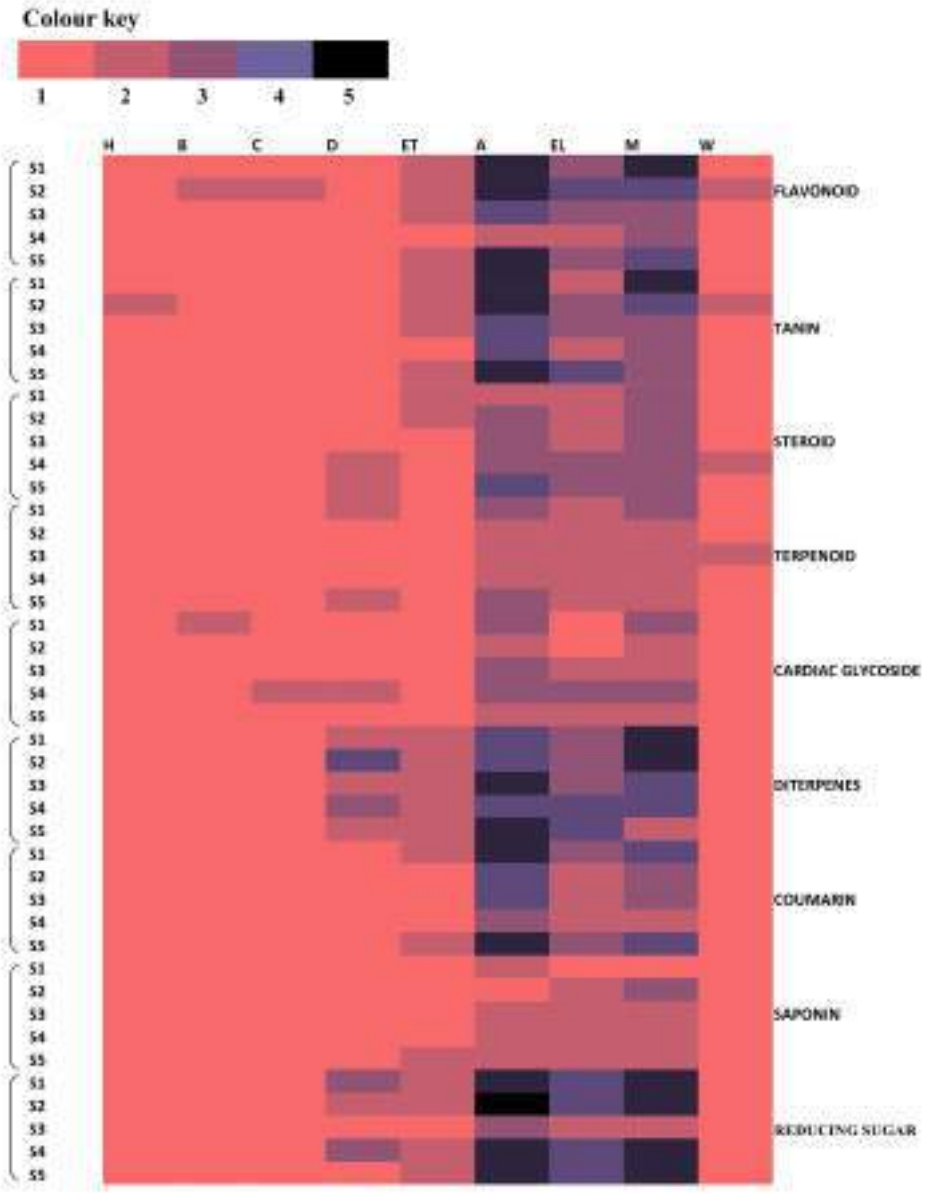

Fig.4.16. Heatmap representing the qualitative phytochemical profiling of different extracts of selected tea varieties. Colour key provided from pink to black represents the intensity in increasing order

non-polar) is very crucial in the extraction of phytochemicals since the compounds have different polarities, chemical characteristics, and solubilities (Ozarkar, 2005). An abundant number of phytochemicals was present in the extracts prepared using acetone, methanol, and ethanol, and traces of phytochemicals like flavonoid, tannin, and cardiac glycoside was found in the extracts prepared using less polar or non-polar solvents like hexane, benzene, and chloroform. High polar solvents like water, ethanol, and acetone extracted major phytochemicals groups like flavonoids, alkaloids, terpenoids, saponins, and phenols from plant extracts of *C. sinensis* (purple tea) when compared with non-polar ethyl acetate and chloroform (Geoffrey *et al.,* 2014). One of the reports suggests methanol as the most suitable solvent than acetone and ethyl acetate to extract phytochemicals from black packaged tea. Methanol showed major extraction properties for extracting phytochemicals like flavonoid, reducing sugar triterpenes, lipid and tannin while other solvents showed minimum activity (Patil *et al.,* 2016). Along with the use of varying solvents for extraction, the time of extraction, and the procedure followed (cold extraction / hot extraction), the state of the sample also plays a crucial role during phytochemical extraction since the phytochemicals slowly degenerate from the time of plucking up to manufacturing. Qualitative screening of phytochemicals very much plays an important role since the presence of a phytochemical of interest or importance may lead to further isolation, purification, and characterization of desired phytochemical of pharmaceutical importance (Ugochukwu *et al.,* 2013).

Since the antioxidant ability of a particular sample or extract cannot be determined using only qualitative method, so methods like total phenol and flavonoid estimation, DPPH free radical scavenging assay, and FRP assay provided some insight into the potent capability of extracting solvents to extract crucial phytochemical compounds with the antioxidant or bioactive property.

4.4.2. DPPH assay

At a single concentration of 200 µg/ml, the percentage of inhibition or the highest radical scavenging percentage in the DPPH assay was observed in acetone extracts followed by methanol extract and ethanol extract (Figure 4.17). The lowest radical scavenging activity was shown by benzene extracts.

The antioxidant potential of the two

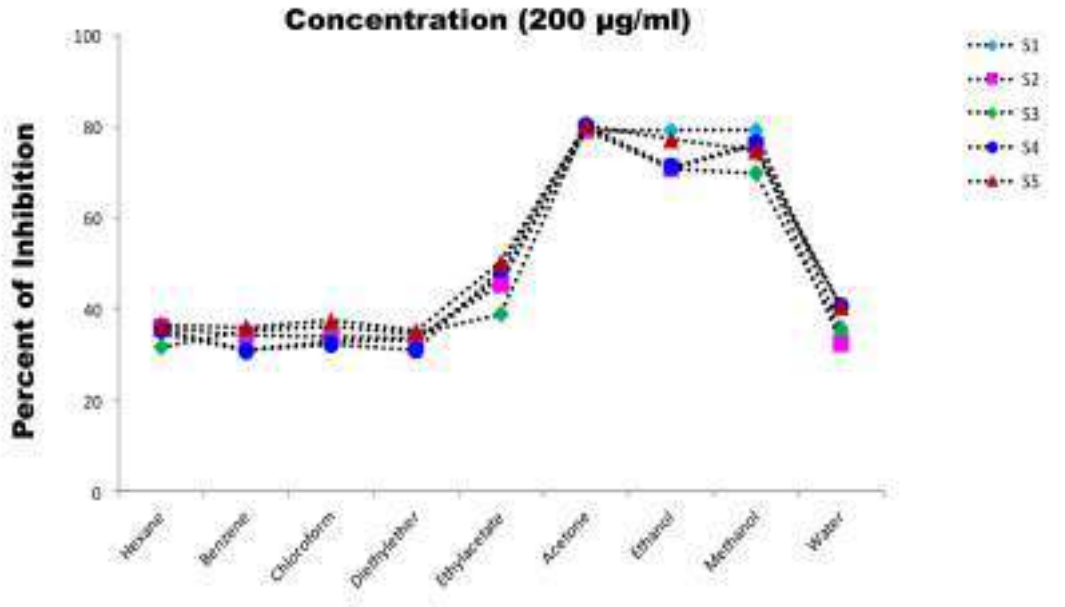

Fig.4.17. DPPH scavenging activity of different solvent leaf extracts of tea clones at single concentration (200µg/ml)

best solvent extracts i.e., the acetone extracts and the methanol extracts were further compared alongside the standard ascorbic acid using various concentrations ranging from 1 mg/ml up to 5 mg/ml. Acetone (Figure 4.18.a) and methanol extracts (Figure 4.18.b) showed antioxidant activity or free radical scavenging property almost similar to the standard used. When compared among all the five clones, acetone gave promising results than methanol. The sample concentration providing 50% of inhibition is the IC_{50} value. The lower the IC_{50} value the higher the antioxidant activity. It was determined by plotting a curve of inhibition using various concentrations. Most of the time the IC_{50} value increases with increasing concentration

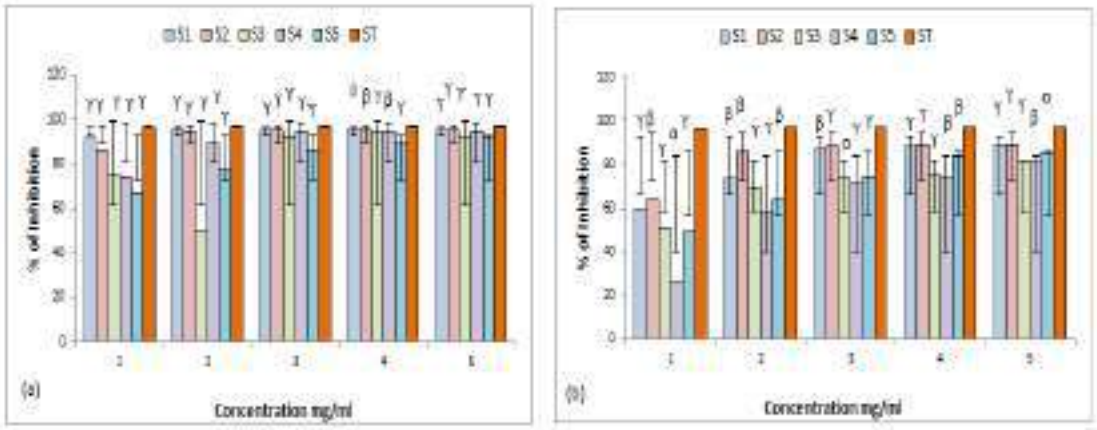

Fig.4.18.DPPH scavenging activity of (a) acetone extracts; methanol extracts; p<0.05; β p<0.01; ϒp<0.001; € non-significant

Table 4.6. IC_{50} value of acetone and methanol extracts prepared in different concentrations (mg/ml) during DPPH assay

Accession	IC_{50} (mean ± sd) of acetone extracts(mg/ml)	IC_{50} (mean ± sd) of methanol extracts(mg/ml)
S1 (TS569)	0.111±0.001	0.635±0.028
S2 (China)	0.159±0.002	0.478±0.028
S3 (AV2)	0.600±0.002	1.030±0.061
S4 (P312)	0.294±0.008	1.595±0.297
S5 (Assam)	0.515±0.002	1.016±0.039
ST (Ascorbic acid)	0.057±0.000	0.056±0.000

Data expressed as means of triplicates ±sd

(Labar *et al*., 2019). Here, the lowest IC_{50} value was recorded as 0.111 ± 0.001 mg/ml (S1 acetone extracts), and 0.478 ± 0.028 mg/ml (S2 methanol extracts) when compared against the IC50 value of the standard ascorbic acid i.e., 0.057 ± 0.000 mg/ml (Table 4.6).

A strong antioxidant property of methanol (Turkmen *et al.,* 2006), ethanol, and acetone extract of black tea has been previously reported. Varying percent concentration of the solvent resulted in varying antioxidant activity with 50% ethanol and 50% acetone showing maximum antioxidant activity in a mate and black tea where they inferred that the solvent potential in phytochemical extraction can be enhanced or reduced just by altering the concentration percentage of the solvent. The result also highlighted the importance of hot water in extraction where the extracts showed moderate (black tea) and higher (mate tea) antioxidant potential.

4.4.3. Ferric reducing power assay

Acetone extracts (Figure 4.19.a) comparatively showed higher ferric reducing power than methanol (Figure 4.19.b) and the results were almost at par with the standard ascorbic acid. Previously methanol extract of fresh tea leaf has shown the highest reducing power based on the sample state. The reducing power also varied among the sample type i.e., greater ferric reducing power was showed by shoot extract which was followed by young leaves and mature leaves (Chan *et al.,* 2007).

4.4.4. Total Phenol

The highest value of total phenol was recorded in acetone extracts of seed clone (S2) as 37.77 mg GAE/g followed by methanol and ethanol as mentioned in Table 4.7. On the other hand, the lowest value was recorded for chloroform extracts of S2 (0.25 mg GAE/g of total phenols). The minimum

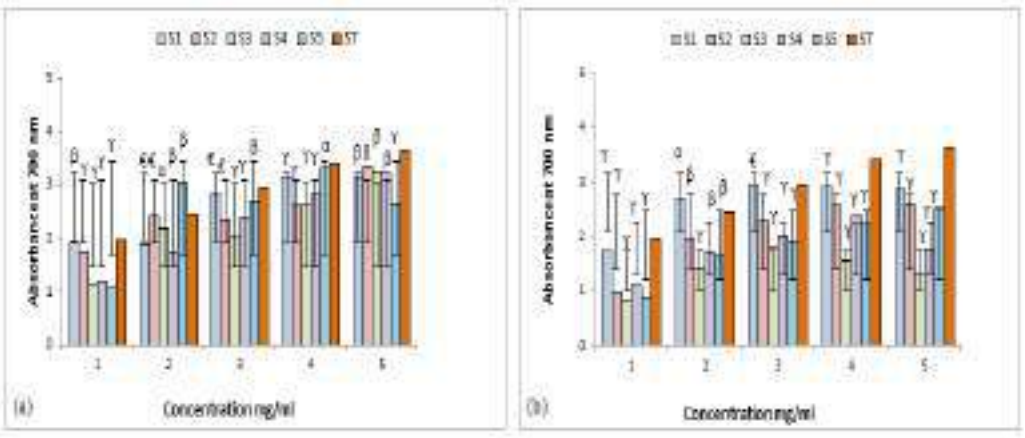

Fig.4.19. Ferric reducing power of acetone extracts (a); methanol extracts (b); p<0.05; β p<0.01; ϒp<0.001; € non-significant

level of phenol was recorded in extracts of non-polar solvents like hexane, benzene, and chloroform. However, this result did not just limit to the varying range of polarity. It varied among the different samples along with varying solvents. Previous works suggest green tea to be a richer source of phenolics than white tea where they reported the potential of 40% aqueous ethanol to extract catechin (Rusak *et al.,* 2008). The estimation of total phenol content using the Folin-Ciocalteu reagent method showed solvents like acetone (50%), N, N-dimethylformamide (DMF) (50%), ethanol (50%), and methanol (50%), to be suitable for extracting total phenols (Turkmen *et al.*, 2006).

4.4.5. Total flavonoid

The highest total flavonoid content

Table 4.7. Determination of total phenol content (TPC) expressed as mg/g GAE

Solvents	S1 (mgGAE/g)	S2 (mgGAE/g)	S3 (mgGAE/g)	S4 (mgGAE/g)	S5 (mgGAE/g)
Hexane	1.61	0.60	3.66	1.00	1.08
Benzene	2.14	0.38	2.86	1.21	1.50
Chloroform	1.77	0.25	1.50	1.45	1.85
Diehylether	3.42	7.18	2.49	2.14	3.34
Ethylacetate	6.41	5.40	3.77	3.69	5.42
Acetone	32.44	37.77	32.06	14.57	18.65
Ethanol	18.25	16.92	17.88	14.25	13.72
Methanol	32.89	31.50	23.82	10.73	11.50
Water	4.04	1.53	2.36	2.52	2.25

Table 4.8. Determination of total flavonoid content (TFC) expressed as mg QE/g

Solvents	S1 (mgQE/g)	S2 (mgQE/g)	S3 (mgQE/g)	S4 (mgQE/g)	S5 (mgQE/g)
Hexane	412.94	446.74	399.42	418.73	537.52
Benzene	406.18	441.91	488.27	401.35	463.16
Chloroform	428.39	482.47	412.94	467.02	477.64
Di ethyl ether	467.99	473.78	433.22	448.67	561.66
Ethyl acetate	470.88	507.58	433.22	424.53	469.92
Acetone	513.76	675.62	617.67	579.04	669.82
Ethanol	528.83	557.80	570.35	506.62	611.88
Methanol	722.94	581.94	545.24	519.17	574.22
Water	424.53	406.18	432.25	423.56	431.29

(Table 4.8) was found in the methanol extracts of S1(722.94 mg QE/g) and the lowest value was recorded that for hexane extracts of S3 (399.42 mg QE/ g).

4.5. Chemical Characterization and bioactivity study

4.5.1. Gas chromatography-mass spectrometry

GC-MS identified crucial bioactive compounds (Appendix D), which could impart antimicrobial properties. The compounds under scrutiny were Phenol, 3,5bis (1,1-dimethyl ethyl), caffeine, and Vitamin E, which could impart antimicrobial properties. GC-MS analysis detected these compounds in the acetone extracts of S3 (Figure 4.20) and methanol extracts of S1 (Supplementary Figure 4.8). The GC-MS data is provided in Appendix D (Table 1– 4). A previous study reports the caffeine-independent antimicrobial activity of coffee extracts against pathogenic strains like *Staphylococcus epidermidis* and *Enterococcus faecalis* (Runti *et al.,* 2015). Contrary to this result, caffeine was also found to enhance or increase the antimicrobial activity of amoxicillin when used against the gram-positive strain i.e., *Staphylococcus aureus* (Esimone *et al.,* 2008). Vitamin E, a known compound with powerful antioxidant activity and the anti-inflammatory property was found to inhibit the growth of mostly gram-positive bacteria rather than gram -negative bacteria due to the presence of lipopolysaccharides in their outer

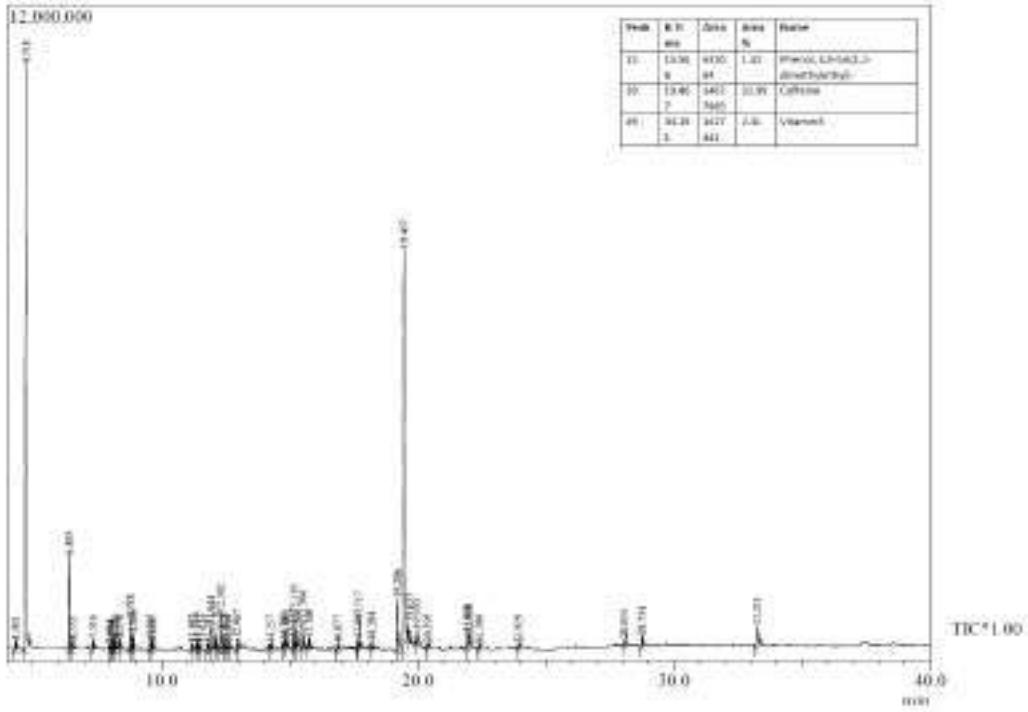

Fig.4.20. Chromatogram of the acetone extracts of AV2 (S3)

membrane (Jehad *et al.,* 2013). The presence of phenol, 3, 5-bis (1, 1-dimethyl ethyl) in GCMS data of *Ulva reticulata* indicated its probable role for imparting anti-microbial effect against *Staphylococcus aureus* and *Escherichia coli* (Dhanya *et al.,* 2016).

4.5.2. Antimicrobial screening

Out of the four bacterial strains studied, acetone and methanol extracts were more effective or bactericidal against the gram-positive *S. aureus* (Figure 4.21). Both the extracts showed minimum or no activity towards the other bacterial strains employed. A broad-spectrum antibiotic i.e., Streptomycin sulfate was used as a standard and the plant extracts were pipetted into each well with volumetric difference. The volumetric increase in the sample ultimately increased the inhibition, which was at par with 100 µl of the standard (Supplementary Figure 4.9). The minimum Inhibition concentration or the MIC value was found to be 4mg/ml for acetone extracts (Supplementary Figure 4.10) and 8mg/ml for the methanol extracts (Figure 4.22).

Previous reports suggest tea extract to be effective against various bacterial strains with bactericidal activity against *Staphylococcus aureus* and *Yersinia enterocolitica* (Yam *et al.,* 1997). However, our study reported acetone and methanol extract to show potent antimicrobial and bactericidal activity against the studied strain, *S. aureus*.

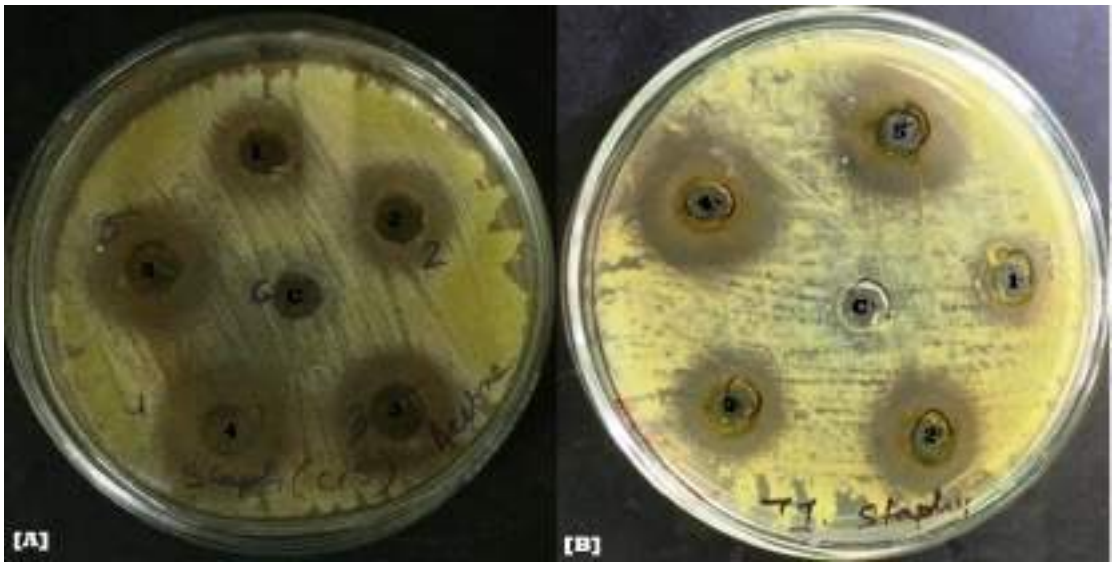

Fig.4.21. Antimicrobial activity of tea extracts against *S. aureus* [A] Acetone [B] Methanol; 1-5 different tea clones; C sterility control (DMSO)

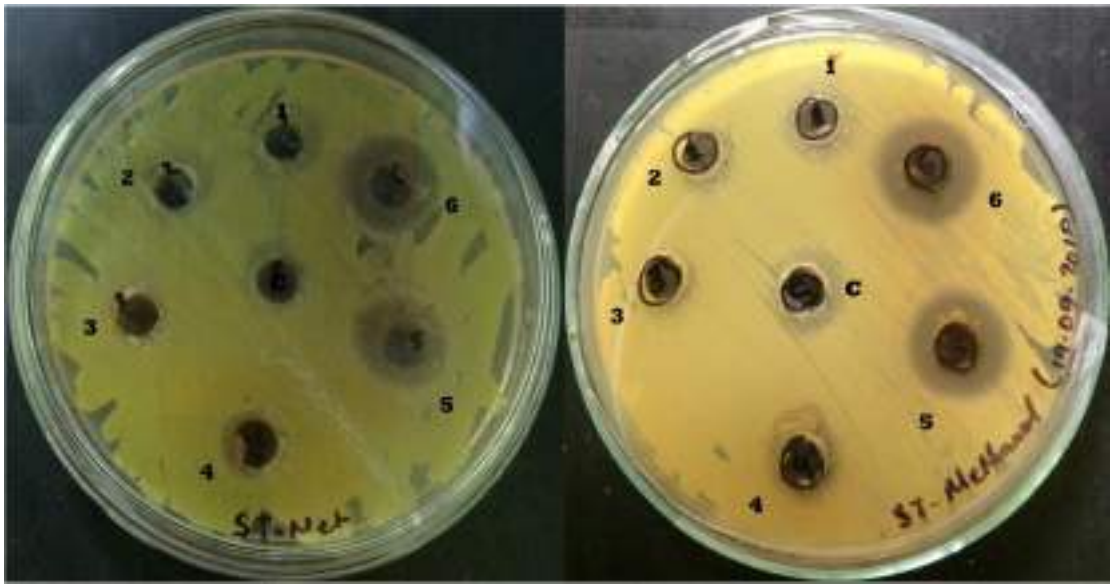

Fig.4.22.MIC value determination from antimicrobial activity of tea extracts against *S. aureus* using agar well diffusion method (C) Methanol extracts of S4(D) Methanol extracts of S7; 1-5 different concentrations 1-0.25mg/ml, 2– 05.mg/ml, 3– 1mg/ml, 4– 4mg/ml, 5– 8 mg/m,6– 10mg/ml; C sterility control DMSO

Both the extracts showed minimum or negligible activity against other strains examined.

4.5.3. In silico analysis

Compounds like caffeine and Vitamin E did not show effective binding with the gyrase protein. The compound Phenol, 3,5bis (1,1-dimethylethyl) showed significant binding to 5CDN with the binding affinity of -7.2 kcal/ mol (Figure 4.23). This result showed the significant effect of phenol with the gyrase protein.

Phenol is a well-known antimicrobial

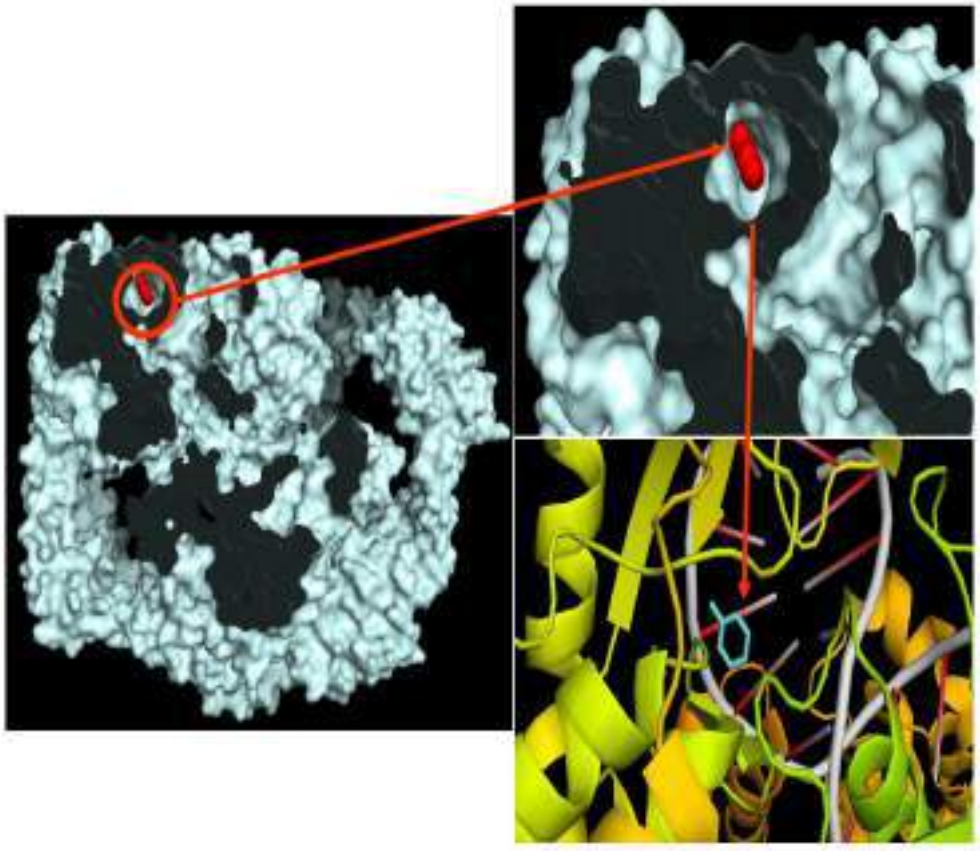

Fig.4.23. *In silico* docking of *S.aureus* DNA gyrase (pdb id: 5CDN) showing significant binding affinity -7.2 kcal/mol with phenol

agent with both bacteriostatic and bactericidal effects which targets the bacterial cell wall, cytoplasmic membrane, organelles, cytosol, and spores (Sabbinenj, 2016). The high penetrating power of the phenol into any organic matter may render its easy access or entry inside the cytoplasmic membrane of bacteria via passive diffusion and active transport mechanisms. Along with its easy permeability, phenol is also a strong oxidizing agent which converts akin to oxide and gas, eventually leading to destabilization of the spores (Sabbinenj, 2016). Therefore, we can infer that the binding of phenol to gyrase A protein may exert a strong antimicrobial activity.

4.6. Purple Tea and its utility in green synthesis of nanoparticles

4.6.1. Silver nanoparticles (AgNPs)

A green route for the rapid synthesis of silver nanoparticles employing extract of purple tea (*C. sinensis*) has been established in our study. The phytochemicals present in the purple tea acts as both reducing as well as a capping agent since earlier reports suggest that the plant extracts contain different phytochemicals like

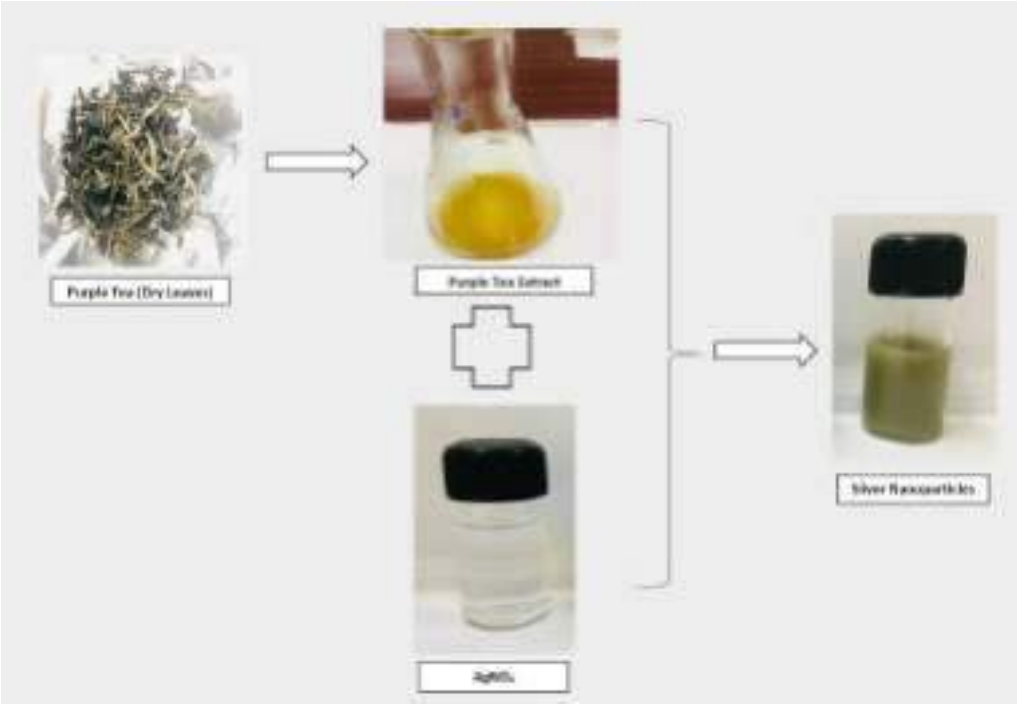

Fig.4.24. Schematic diagram of synthesis of silver nanoparticles from purple tea extract

Fig.4.25. Synthesis of silver nanoparticles characterized by change in color

flavonoid, terpenoid, phenol derivatives, etc. which shows the bioreduction property in the presence of metal salts (Kumar *et al.*, 2017). The formation of silver nanoparticles synthesized from purple tea extract was primarily screened by the change in color of the reaction mixture (Figure 4.24). The color of the solution changed from yellow to green after 5 minutes of incubation (Figure 4.25). A previous study on the synthesis of silver nanoparticles using pu-erh tea extract also reports the formation of silver nanoparticles by a change in color of the solution to dark brown

(Loo *et al.*, 2012). The formation or synthesis of AgNPs was further confirmed by its characterization using UV-visible spectroscopy (UV-vis), X-ray diffraction (XRD), and Scanning electron microscopy (SEM).

4.6.1.1. Characterization of silver nanoparticles

The reaction parameters like reaction temperature, the concentration of reactants used in the solution, time of reaction, pH, etc. have to be controlled and optimized to obtain the desired monodispersed nanoparticles with tunable size and morphology (Tripathy *et al.,* 2010).

4.6.1.1.1. Effect of varying molar concentration of $AgNO_3$

Firstly, the effect of different molar concentrations was studied to explore the optimum synthesis of silver nanoparticles. A varying molar concentration of silver nitrate was screened in the range of 1 mM to 10mM (Figure 4.26). The reaction mixture was exposed to bright sunlight for 15 minutes. The reaction mixtures containing different metal ion ($AgNO_3$) concentrations; 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM and 10 mM produced (Surface Plasmon Resonance) SPR band at 414 nm, 417 nm, 418 nm, 431 nm, 430 nm, 428 nm, 437 nm, 436 nm, 439 nm and 442 nm respectively. It was observed that the SPR band of 10mM was the least intense as compared to others (Figure 4.26). The steady increase in SPR band intensity up to 3mM revealed that the synthesis of silver

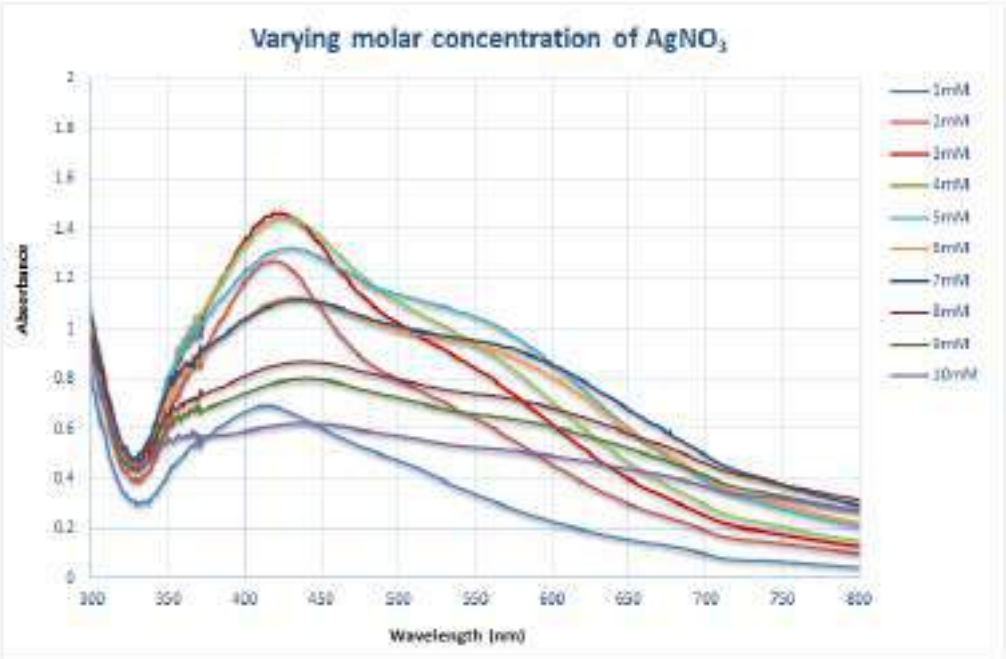

Fig.4.26.A varying molar concentration of silver nitrate was screened for AgNPs synthesis in the range of 1 mM -10mM

nanoparticles increased with the increasing concentration of $AgNO_3$. Synthesis of AgNPs increased while increasing the $AgNO_3$ concentration up to 2 mmol L^{-1} that was evident from a regular increase in SPR band intensity (Philip, 2010; Bar *et al.*, 2009). A redshift in the SPR band from 453 nm to 459 nm is directed towards the increase in the size of the nanoparticle (Jagtap and Bapat, 2013). However, the intensity of the SPR band (431nm) of 4mM decreased when compared against 3mM (418nm). The intensity of the SPR band of 3mM increased but following that the intensity of the band decreased.

A redshift towards a higher wavelength was observed from 418 to 442 nm with an increase in the concentration of $AgNO_3$. The redshift towards a higher wavelength with an increase in the concentration of $AgNO_3$ directly relates to an increase in the size of the silver nanoparticles (Mock *et al.*, 2002). Thus, 3mM was chosen as the optimal molar concentration of $AgNO_3$ for this current study. The obtained AgNPs were found to be stable only for 24 hours and after that settled at the bottom due to agglomeration. The previous study reports such agglomeration during green synthesis of nanoparticles, which was also a case, reported in the previous study (Kumar *et al.,* 2017).

4.6.1.1.2. Effect of sunlight exposure

When the reaction mixture containing $AgNO_3$ and plant extract was exposed to bright sunlight, the color of the solution changed from colorless to brown rapidly within 10 minutes of incubation, and as time advanced the color became darker. However, the

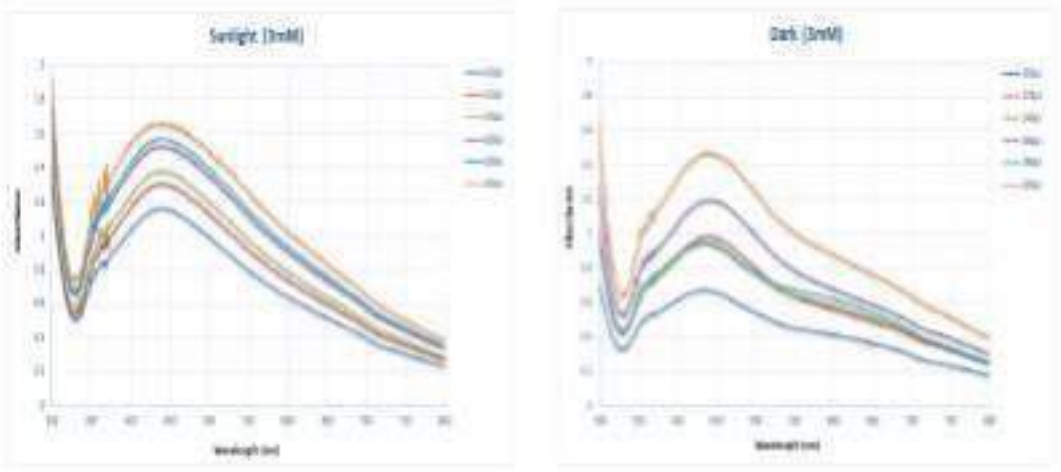

Fig.4.27.UV-vis absorption spectrum of purple tea extract mediated AgNPs synthesized using varying conditions of light

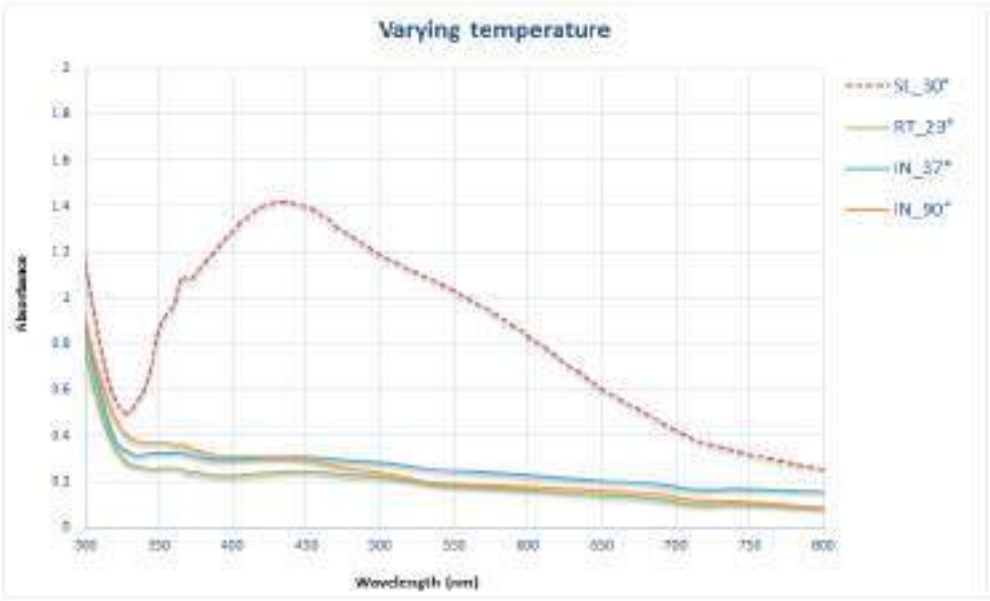

Fig.4.28.UV-vis absorption spectrum of purple tea extract mediated AgNPs synthesized using varying conditions of temperature

reaction mixture kept in the dark also rendered the same color but the change was not rapid as compared to the sample exposed to the bright sunlight. This result highlights the importance of the photocatalytic action of sunlight on the synthesis of AgNPs (Kumar *et al.,* 2017). UV–vis spectroscopy further monitored the reaction mixture kept at varying conditions of light and darkness (Figure 4.27).

4.6.1.1.3. Effect of varying temperature

A different range of temperature was used to study the effect of temperature on AgNPs synthesis. The sample or the reaction mixture was exposed to bright sunlight (30°C), room temperature of 23°C and incubator temperatures of 37°C and 90°C. The reaction mixture exposed to varying temperature conditions for 1 hour was monitored using UV Vis spectroscopy. The sample kept at 30°C (under sunlight) produced the sharper SPR band at 430 nm whereas the other reaction mixture showed a blue shift towards a lower wavelength (304nm) with a decrease in the SPR band intensity (Figure 4.28).

4.6.1.2. Structural Analysis

The structural properties of the sample were studied by carrying out the (X-Ray Diffraction) XRD analysis employing Rigaku Ultima IV. The crystalline profile of our sample clearly shows the peaks of Silver, the major peaks (Figure 4.29) occurring at the diffraction angle 38.03°, 44.15°, and 64.40° correspond to the Ag [111], Ag [002], and Ag [220] respectively. These values of the diffraction angles for the Ag peaks closely match with the PDF card number 00-004-0783

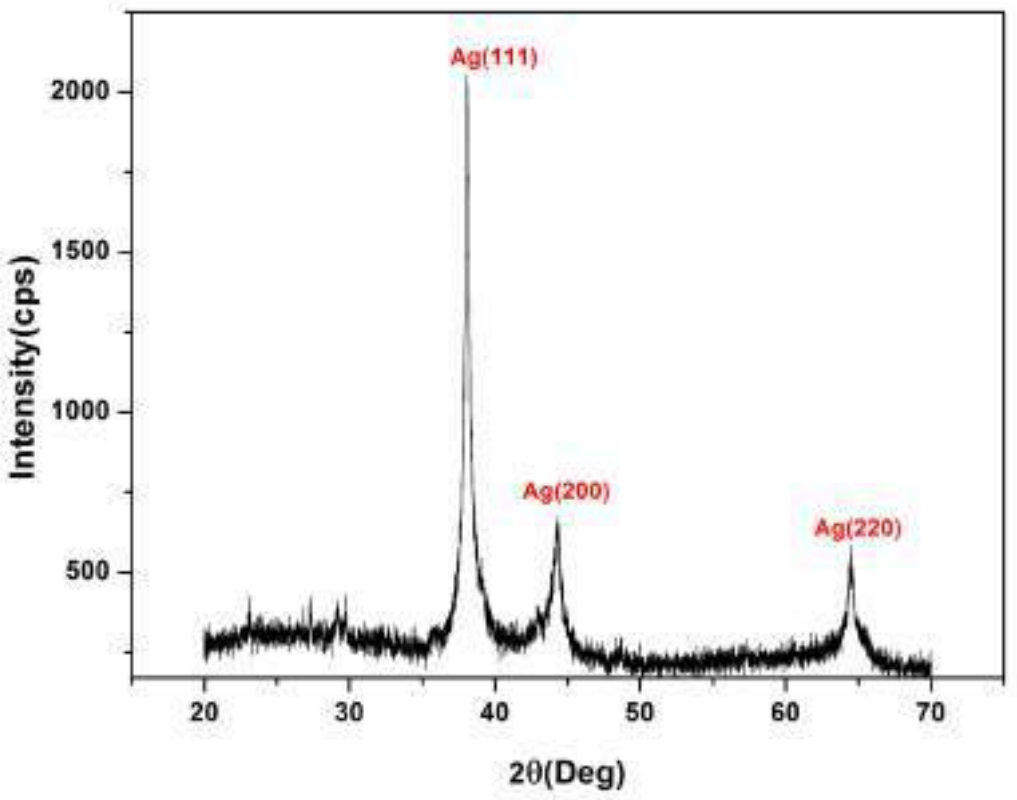

Fig.4.29.XRD pattern of synthesized silver nanoparticle. The crystalline profile of the sample clearly shows the peaks of Silver, the major peaks occurring at the diffraction angle 38.03°, 44.15° and 64.40° corresponding to the Ag [111], Ag [002] and Ag [220]

indicating the formation of silver nanoparticles. The size of the nanoparticle was determined using the Debye Scherrer's formula given as,

$$D = \frac{K\lambda}{\beta\cos(\theta)}$$

Where,

K- Boltzmann Constant

λ – Wavelength of radiation

β – FWHM value of the corresponding peak

θ- Diffraction angle

The size of the nanoparticle hence determined from our analysis was of the order of 38.4 nm. Thus, we can infer from the XRD study, that pure Ag nanoparticles are formed with the particle size of the order of 38 nm.

4.6.1.3. Scanning electron microscopy (SEM)

The SEM imaging revealed varying morphology with the majority of silver nanoparticles to be non-spherical or either irregular in shape with a rough surface, and few were rod-shaped (Supplementary Figure 4.11). The

particle size ranged from 10nm to 40 nm. The particle size of a few exceeded the range, which is probably due to the overlapping of the synthesized nanoparticles. The particles were agglomerated due to crosslinking (Shankar *et al.*, 2017) or solvent evaporation of solvent during the preparation of the sample (Jagtap and Bapat 2013).

4.6.1.4. Antimicrobial property

The synthesized AgNPs were found to show antimicrobial properties against *Staphylococcus aureus* and *Bacillus subtilis* (Figure 4.30) out of the four strains studied namely *Staphylococcus aureus*, *Bacillus subtilis, Escherichia coli*, and *Klebsiella pneumonia.* The plant extract itself showed antimicrobial activity against the gram-positive strains and showed minimum or no activity against the gram-negative ones. However, the antimicrobial activity enhanced with the formation of silver nanoparticles (AgNPs) using the purple tea extract.

4.6.2. Zinc oxide nanoparticles (ZnO NPs)

The schematic diagram of ZnO nanoparticle synthesis is given in Figure 4.31. The brown color of the ZnO NPs arises due to the capping action of purple tea biomolecules on the surface of the nanoparticles. The formation of ZnO nanoparticles synthesized using leaf extract of purple tea was initially monitored by the change in color occurring during the reaction period (Figure 4.32). Upon addition of tea extract to the reaction mixture, the color slowly changes to pale white which then progresses to yellowish-green, and finally, the color of the end product was brown. The

Fig.4.30. Antimicrobial activity of purple tea extract mediated AgNPs:
(a) *S. aureus*, (b) *E. coli*

Fig.4.31. Schematic diagram of synthesis of Zinc Oxide nanoparticles (ZnO NPs) from purple tea extract

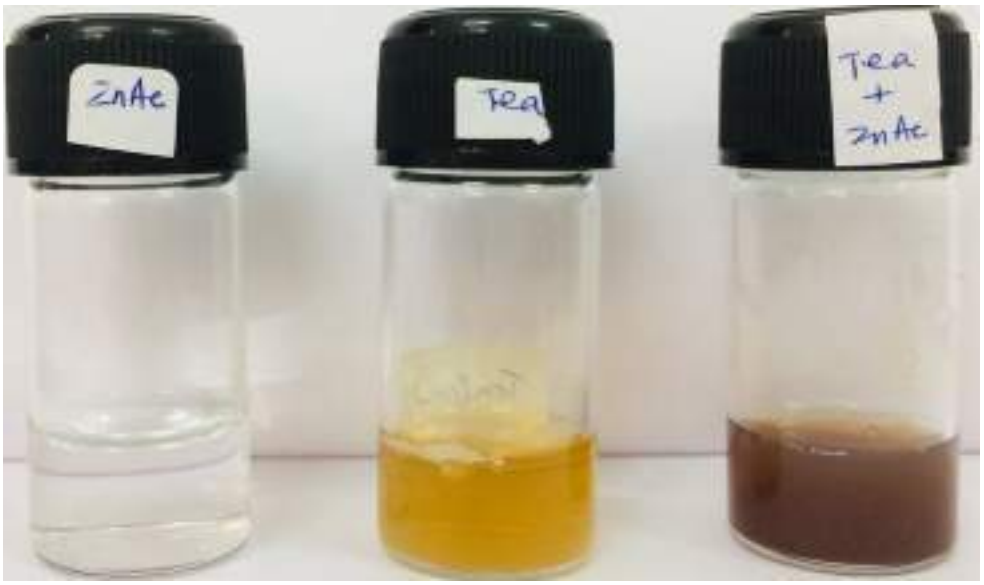

Fig.4.32. Synthesis of Zinc oxide nanoparticles characterized by change in color

color change arises due to the excitation of surface plasmon resonance in the metal nanoparticles, indicating the formation of ZnO nanoparticles.

4.6.2.1. ZnONPs characterization

4.6.2.1.1.UV–characterization

The UV-Vis spectrum of the ZnO nanoparticle is provided in Figure 4.33. The synthesis of ZnO product in nanoscale was confirmed by the highly blue-shifted, maximum absorption peak occurring around 273 nm. The zinc oxide synthesized using leaf extract of *Sesbania grandiflora* (Rajendran and Sengodan, 2017) exhibited UV-visible absorption peaks at 235 nm which is nearly similar to our results. Another work related to

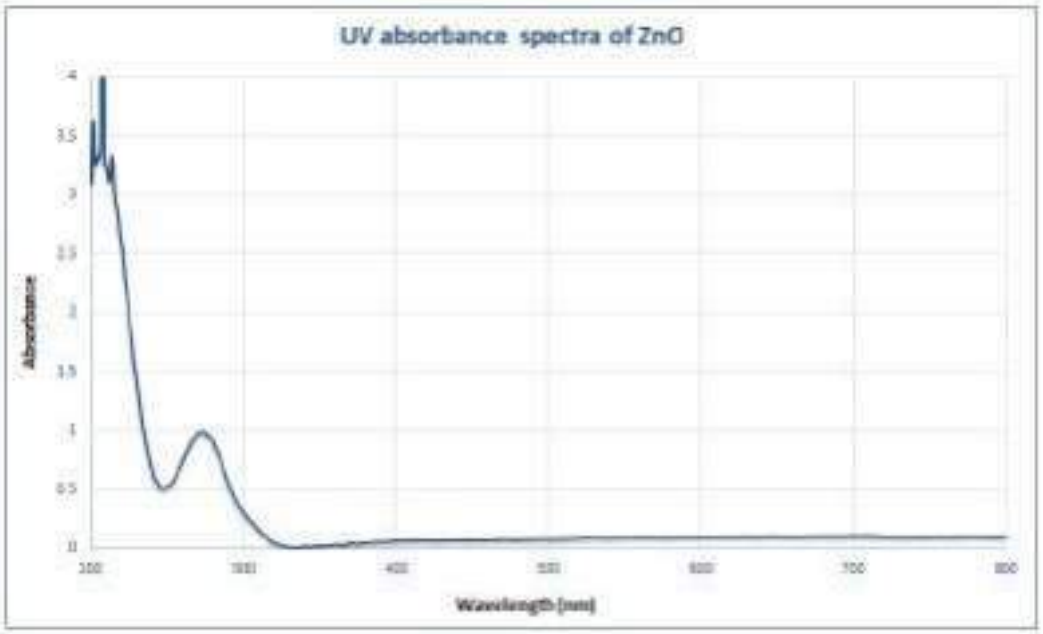

Fig.4.33.UV-vis absorption spectrum of Zinc oxide nanoparticle synthesized from purple tea extract

ZnO synthesis using the microemulsion method also reports a similar kind of result (Kumar and Rani, 2013). Usually, maximum absorption for bulk ZnO approximately occurs around 385 nm. The decrease in absorption may result from agglomeration and settling down of nanoparticles in the cuvette. However, we conclude from our analysis that the synthesized zinc nanoparticles were photosensitive in the UV region since they showed an intense absorbance peak of ~270 (Savi *et al.*, 2014). When the concentration of the plant extract was increased, the absorbance value also increased where the peak intensified at 400nm. The increased concentration of the phytochemicals in the extract increases the efficacy of the extract to reduce the precursor present in the solution rapidly which eventually also enhances the synthesis of nanoparticles as well as increases the absorbance value (Isaac *et al.*, 2013).

4.6.2.1.2. Structural Analysis using XRD (X-ray diffraction)

To study the crystalline properties of the sample, we performed the (X-Ray Diffraction) XRD analysis of the sample, employing Rigaku Ultima IV. The X-ray diffraction profile of the sample shows a crystalline nature. The detailed analysis of the peaks indicates the presence of ZnO, the occurrence of which is visibly clear in the sample annealed at temperature 400°C (Figure 4.34). The major peaks [100], [002], and [101] of ZnO occur at the diffraction angles 31.69º and 34.77º and 36.1º respectively. These values of

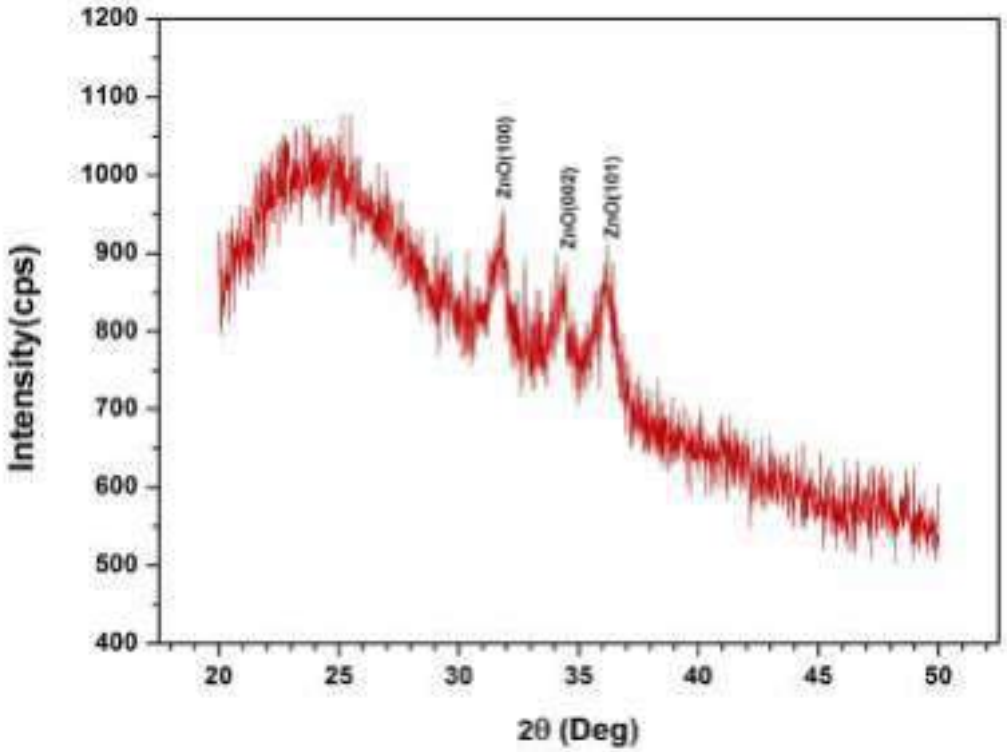

Fig.4.34.XRD pattern of synthesized Zno NPs (Calcination temperature - 400° C). The major peaks [100], [002] and [101] of ZnO occur at the diffraction angles

the diffraction angles for the ZnO peaks closely match with the PDF card number 01-080-0074 indicating the formation of pure ZnO nanoparticles. The size of the nanoparticle was determined using the Debye Scherrer's formula given as,

$$D = \frac{K\lambda}{\beta\cos(\theta)}$$

Where,

K- Boltzmann Constant

λ – Wavelength of radiation

β – FWHM value of the corresponding peak

θ- Diffraction angle

The size of the nanoparticle hence determined from our analysis was of the order of 22.4 nm. Though the samples annealed at 400°C show a clear formation of ZnO nanoparticles, the samples annealed at 100°C (Figure 4.35) shows the occurrence of only the [002] peak with a minimum intensity. The profile also shows a broad background, somewhat like a hump, along with the crystalline peaks, the possible reason for which could be due to the presence of the organic compound in the material. The broad background is minimized slightly at higher temperatures probably due to the removal of the organic compound at a higher temperature.

The occurrence of this arch-like peak

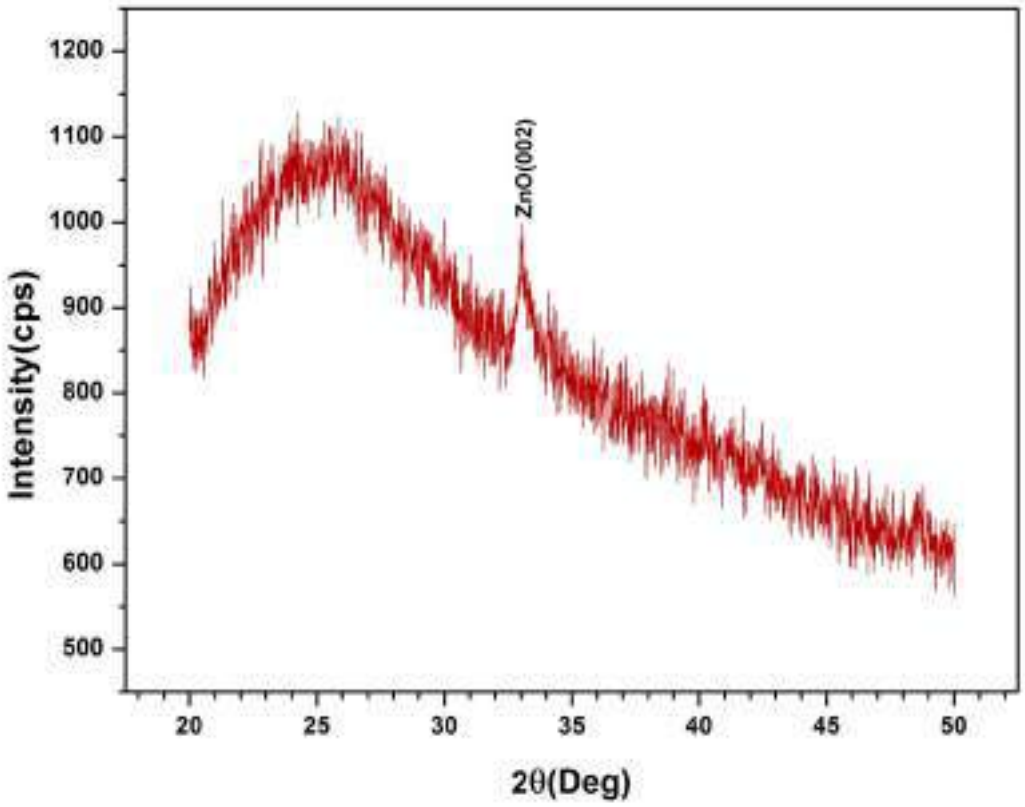

Fig.4.35.XRD pattern of synthesized Zno NPs (Calcination temperature - 100° C). Single peak [002] of ZnO occur at the diffraction angle 34.77º

in the profile also indicates the presence of tea in the sample as reported in various works.

4.6.2.1.3. Scanning electron microscopy (SEM)

SEM was carried out to determine the surface morphology of the synthesized ZnO. The micrographs (Figure 4.36) showed network formation or cross-linking during the synthesis of ZnO. This indicates agglomeration (Shankar *et al.*, 2017) during the process of synthesis. Moreover, the synthesized ZnO was found to be spherical with an approximate size of around 15-25nm.

4.6.2.2. Antimicrobial property

The antimicrobial activity of the plant extract (purple tea extract), zinc acetate, and zinc nanoparticle (ZnO) was screened against bacterial strains including both the gram-positive (*Staphylococcus aureus*, *Bacillus subtilis*) and gram-negative strains (*Escherichia coli*, *Klebsiella pneumoniae*). We found that zinc acetate itself has an antimicrobial property towards all the investigated strains while our plant extract was effective against the strains like *Staphylococcus aureus* and *Bacillus subtilis* and we solely aimed to see the

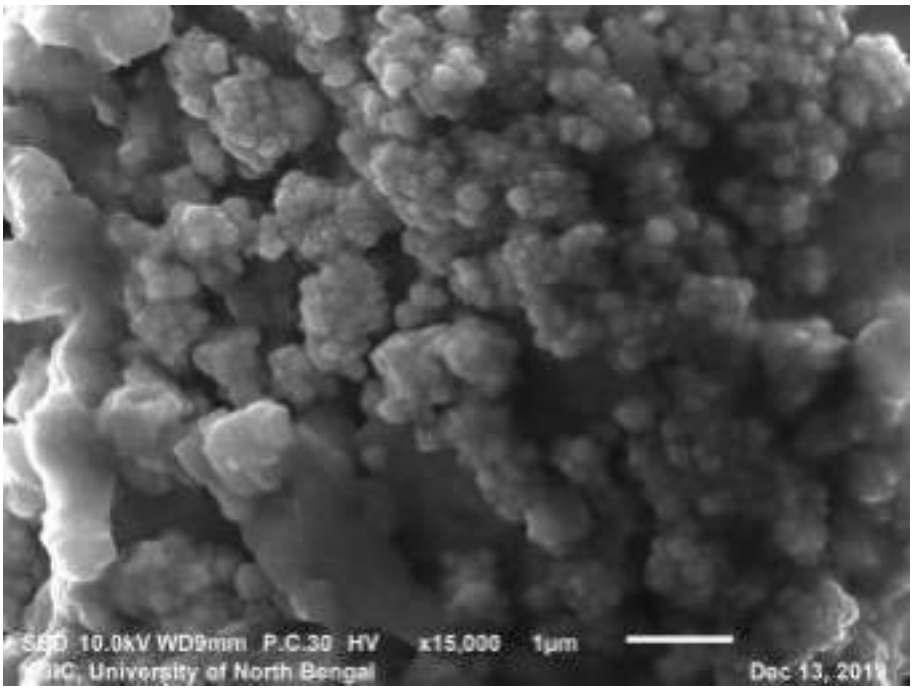

Fig.4.36.SEM image of ZnONPs synthesized using aqueous purple tea extract

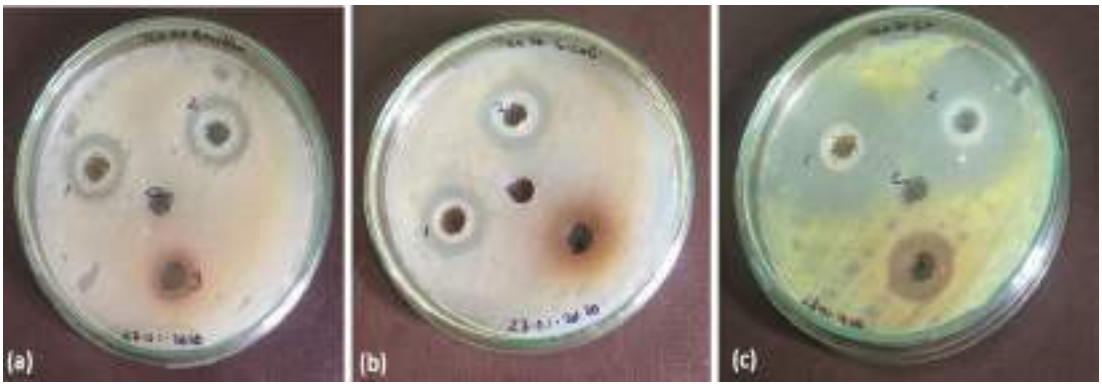

Fig.4.37. Antimicrobial activity of purple tea extract mediated ZnONPs: (a) *Bacillus subtilis*, (b) *Escherichia coli*, (c) *Staphylococcus aureus*

antimicrobial activity of the synthesized ZnO. The inhibition zone of the ZnO was bigger when compared solely to Zinc acetate and plant extract. Thus, we found that the antimicrobial property was enhanced in the synthesized zinc nanoparticles (Figure 4.37). The synthesized ZnO showed antimicrobial activity against all the strains but the inhibition property towards *Escherichia coli* was due to the antimicrobial property of zinc acetate alone since the plant extract did not show significant antimicrobial property towards *E. coli*.

4.7. Tea Bioinformatics

4.7.1. Factors governing codon usage behavior

Availability of whole-genome sequence of the reference genome of

Camellia sinensis var. *sinensis* cv. Shuchazao (Xia *et al*., 2020) allowed a detailed study of its codon usage and amino acid usage patterns. *Camellia sinensis* was found to be AT-rich (AT 55% and GC 45%). The tea genome preferred AT-rich codons over the GC ones. The overall codon usage pattern revealed that 15 amino acids were coded by AT-rich codons with RSCU > 1 as marked in bold (Table 4.9). Out of 27 optimal codons (RSCU>1), 19 optimal codons were AT-rich. Out of the 61 codons, 29 codons were AT-rich out of which 22 codons were found to end with either Adenine or Thymine. Usage of AT-rich codons like AAG, GAA, GAT, AAT, AAA, GTT, GCT, TTG, CTT, TCT, and ATG was quite perceptible from the given Table 4.9. Preference for AT-rich codons was evident from codon usage and amino acid usage analysis. Axis 1 of Relative synonymous codon usage (RSCU) was found to exhibit a significantly high and positive correlation with A3 and T3 indices (Table 4.10) which pointed towards a mighty effect of the AT nucleotide compositional constraint in shaping the codon usage pattern of *Camellia sinensis*. Various important correlation patterns between the codon usage parameters pointed out the importance of nucleotide compositional constraint in directing codon usage variation. Cases of nucleotide composition governing the codon usage have been reported in unicellular as well as multicellular eukaryotes like *Entamoeba histolytica* (Romero *et al*., 2000), different human genes (Karlin and Mrazek, 1996), and nuclear genes of *Oryza sativa*, *Zea mays*, and *Tuber aestivum* (Liu and Xue, 2005).

GC3 versus Nc plots were plotted where we considered total protein-coding genes (TPC) along with potentially highly expressed genes (PHX) and potentially lowly expressed genes (PLX) as shown in Figure 4.38. We also plotted a GC3 versus Nc plot (Figure 4.39) considering total protein-coding genes along with genes of different metabolic pathways rendering synthesis of important metabolites like flavonoid, phenylpropanoid, flavone and flavonol, caffeine biosynthesis, and streptomycin biosynthesis. GC3 versus Nc plot has been suggested to be effective in the proper assessment of probable factors that govern codon usage patterns among genes and genomes (Wright 1990). Previous studies have stated that all the genes would fall on a continuous Nc plot if the codon usage is solely governed by nucleotide compositional constraint (Wright 1990). The fact that nucleotide compositional constraint was having a pronounced effect on the codon usage behavior of *C. sinensis* was evident

Table 4.9. Overall codon usage table of *C. sinensis*

AA	CODON	N	RSCU	AA	CODON	N	RSCU
Phe	**UUU**	367773	**1.13**	**Ser**	**UCU**	343826	**1.51**
	UUC	284965	0.87		UCC	190052	0.84
Leu	UUA	181203	0.72		**UCA**	308178	**1.36**
	UUG	405410	**1.6**		UCG	119393	0.53
	CUU	351724	**1.39**		**AGU**	228392	**1.01**
	CUC	241117	0.95		AGC	172260	0.76
	CUA	158489	0.63	**Thr**	**ACU**	259137	**1.38**
	CUG	181791	0.72		ACC	173905	0.93
Ile	**AUU**	393437	**1.43**		**ACA**	245262	**1.31**
	AUC	229411	0.83		ACG	73037	0.39
	AUA	202753	0.74	Ala	GCU	374391	1.53
Val	**GUU**	385936	**1.54**		GCC	191336	0.78
	GUC	174705	0.7		GCA	319405	1.31
	GUA	153571	0.61		GCG	92867	0.38
	GUG	290487	1.16	**Arg**	CGU	92103	0.71
Pro	CCU	256840	1.38		CGC	69350	0.53
	CCC	124824	0.67		CGA	97952	0.75
	CCA	271364	1.46		CGG	8314	0.6
	CCG	88950	0.48		**AGA**	237445	**1.83**
Gly	GGU	308462	1.22		AGG	4585	1.57
	GGC	178355	0.7	**His**	**CAU**	225773	**1.22**
	GGA	309165	1.22		CAC	144199	0.78
	GGG	16733	0.86	**Gln**	**CAA**	319649	**1.16**
Asn	**AAU**	424258	**1.22**		CAG	232013	0.84
	AAC	271869	0.78	**Asp**	**GAU**	537901	**1.37**
Lys	AAA	425374	0.95		GAC	249058	0.63
	AAG	66503	**1.05**	**Glu**	**GAA**	511574	**1.06**
Tyr	**UAU**	239990	**1.15**		GAG	451716	0.94
	UAC	176468	0.85				
Cys	**UGU**	165325	**1.12**				
	UGC	130754	0.88				

Bold letters represent the amino acids preferentially coded by AT rich codons
RSCU cumulative RSCU values by summation of RSCU of all genes, *N* number of codons, *AA* amino acids

from the GC3 versus Nc plots. However, many genes were found to position well below the continuous Nc curve suggesting the influence of factors other than compositional bias. Apart from the role of nucleotide compositional constraint, the impact of translational selection on codon usage patterns was apparent from the RSCU based scatter plots and GC3 versus NC plots. It has already been reported that if translational selection acts on a gene, the aforesaid gene lies distantly below the continuous plot (Nair *et al*., 2012). Efficient translation, under the deep influence of natural selection, was most prevalent on PHX genes. Codon usage variation due to the influence of translational selection has been observed previously in organisms like *Caenorhabditis* (Moriyama and Powell, 1997).

Multivariate statistical analysis (correspondence analysis) of RSCU

Table 4.10. Correlation analysis between different parameters of codon and amino acid usage obtained from SPSS

	T3s	C3s	A3s	G3s	CAI	CBI	Fop	GC3s	GC	L_sym	L_aa	Gravy	Aromo	MeanCost	Axis1_RSCU	Axis2_RSCU	Axis3_RSCU	Axis4_RSCU	Axis1_RAAU	Axis2_RAAU	Axis3_RAAU	Axis4_RAAU
T3s	1.000	-.758**	.293**	-.327**	.814**	-.139**	-.061**	-.832**	-.665**	.212**	.212**	-.039**	.034**	.037**	.789**	-.061**	.329**	.108**	.084**	.138**	-0.009	-.190**
C3s	-.758**	1.000	-.523**	.033**	-.702**	.353**	.332**	.825**	.608**	-.217**	-.217**	.077**	.150**	-.052**	-.867**	-.336**	-.198**	.014**	-.214**	-.103**	.114**	.121**
A3s	.293**	-.523**	1.000	-.461**	.567**	-.199**	-.104**	-.730**	-.675**	.151**	.152**	-.192**	-.048**	.178**	.685**	-.117**	-.218**	-.342**	.234**	.064**	.157**	-.138**
G3s	-.327**	.033**	-.461**	1.000	-.286**	-.220**	-.211**	.561**	.353**	-.095**	-.094**	-.209**	-.132**	.184**	-.344**	.696**	.018**	.226**	.384**	.102**	-.190**	.043**
CAI	.814**	-.702**	.567**	-.286**	1.000	-.032**	.107**	-.809**	-.837**	.183**	.183**	-.171**	.058**	.271**	.846**	-.080**	.013*	.015**	.269**	.228**	.130**	-.262**
CBI	-.139**	.353**	-.199**	-.220**	-.032**	1.000	.943**	.180**	.146**	-.093**	-.094**	.080**	0.003	-.102**	-.213**	-.320**	-.197**	-0.010	-.139**	-.118**	.061**	.071**
Fop	-.061**	.332**	-.104**	-.211**	.107**	.943**	1.000	.120**	-.036**	-.089**	-.090**	.047**	.066**	.037**	-.138**	-.331**	-.238**	-.012*	-.039**	.019**	.106**	.055**
GC3s	-.832**	.825**	-.730**	.561**	-.809**	.180**	.120**	1.000	.780**	-.232**	-.232**	0.003	0.008	-.036**	-.923**	.125**	-.125**	.134**	-.038**	-.095**	-.053**	.163**
GC	-.665**	.608**	-.675**	.353**	-.837**	.146**	-.036**	.780**	1.000	-.205**	-.207**	-.071**	-.184**	-.485**	-.792**	.113**	.075**	.102**	-.154**	-.452**	-.369**	.234**
L_sym	.212**	-.217**	.151**	-.095**	.183**	-.093**	-.089**	-.232**	-.205**	1.000	1.000**	-.016**	-.053**	0.010	.217**	0.001	.081**	-.040**	.035**	.027**	.087**	-.131**
L_aa	.212**	-.217**	.152**	-.094**	.183**	-.094**	-.090**	-.232**	-.207**	1.000**	1.000	-.012*	-.047**	.015**	.218**	0.002	.080**	-.039**	.033**	.033**	.086**	-.128**
Gravy	-.039**	.077**	-.192**	-.209**	-.171**	.080**	.047**	0.003	-.071**	-.016**	-.012*	1.000	.379**	.136**	-.028**	-.134**	.023**	-.024**	-.680**	.536**	.059**	.278**
Aromo	.034**	.150**	-.048**	-.132**	.058**	0.003	.066**	0.008	-.184**	-.053**	-.047**	.379**	1.000	.417**	-.016**	-.109**	-.043**	-.014**	-.505**	.358**	.097**	.133**
MeanCost	.037**	-.052**	.178**	.184**	.271**	-.102**	.037**	-.036**	-.485**	0.010	.015**	.136**	.417**	1.000	.130**	-.024**	-.084**	-0.009	.279**	.683**	.355**	.179**
Axis1_RSCU	.789**	-.867**	.685**	-.344**	.846**	-.213**	-.138**	-.923**	-.792**	.217**	.218**	-.028**	-.016**	.130**	1.000	0.000	0.000	0.000	.134**	.185**	.021**	-.179**
Axis2_RSCU	-.061**	-.336**	-.117**	.696**	-.080**	-.320**	-.331**	.125**	.113**	0.001	0.002	-.134**	-.109**	-.024**	0.000	1.000	0.000	0.000	.181**	-0.009	-.188**	-.045**
Axis3_RSCU	.329**	-.198**	-.218**	.018**	.013*	-.197**	-.238**	-.125**	.075**	.081**	.080**	.023**	-.043**	-.084**	0.000	0.000	1.000	0.000	-.019**	-.051**	-.089**	.079**
Axis4_RSCU	.108**	.014**	-.342**	.226**	.015**	-0.010	-.012*	.134**	.102**	-.040**	-.039**	-.024**	-.014**	-0.009	0.000	0.000	0.000	1.000	.046**	.012*	-.105**	.038**
Axis1_RAAU	.084**	-.214**	.234**	.384**	.269**	-.139**	-.039**	-.038**	-.154**	.035**	.033**	-.680**	-.505**	.279**	.134**	.181**	-.019**	.046**	1.000	.023**	-.035**	-.042**
Axis2_RAAU	.138**	-.103**	.064**	.102**	.228**	-.118**	.019**	-.095**	-.452**	.027**	.033**	.536**	.358**	.683**	.185**	-0.009	-.051**	.012*	.023**	1.000	-.048**	-.028**
Axis3_RAAU	-0.009	.114**	.157**	-.190**	.130**	.061**	.106**	-.053**	-.369**	.087**	.086**	.059**	.097**	.355**	.021**	-.188**	-.089**	-.105**	-.035**	-.048**	1.000	-.012*
Axis4_RAAU	-.190**	.121**	-.138**	.043**	-.262**	.071**	.055**	.163**	.234**	-.131**	-.128**	.278**	.133**	.179**	-.179**	-.045**	.079**	.038**	-.042**	-.028**	-.012*	1.000

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

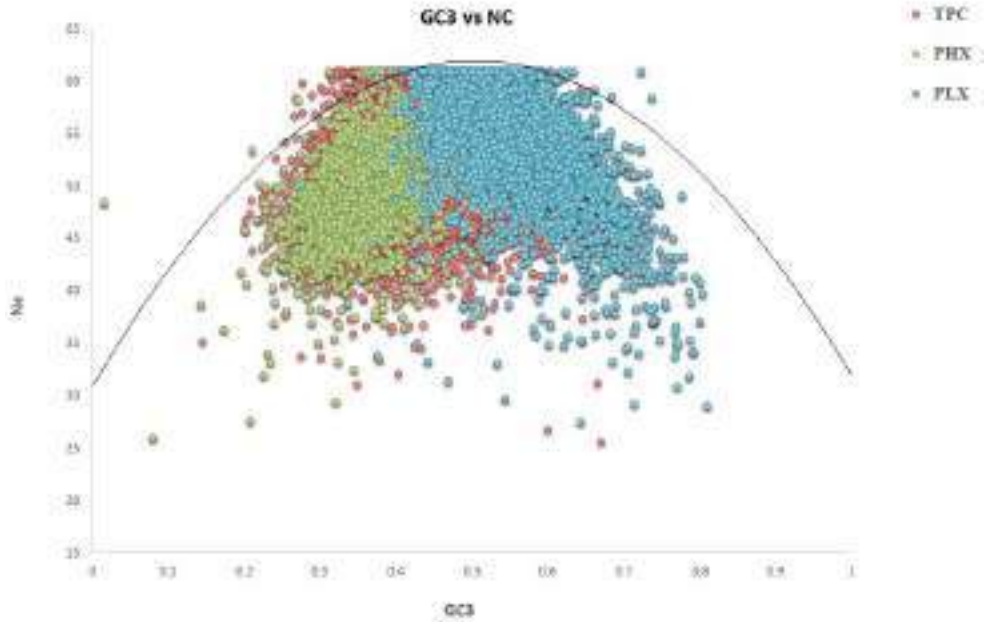

Fig.4.38.GC3 versus Nc plot for *Camellia sinensis* genome. Orange colored circles represent total protein coding genes (TPC), blue colored circles represent potentially lowly expressed genes (PLX), greencolored circles represent potentially highly expressed genes (PHX)

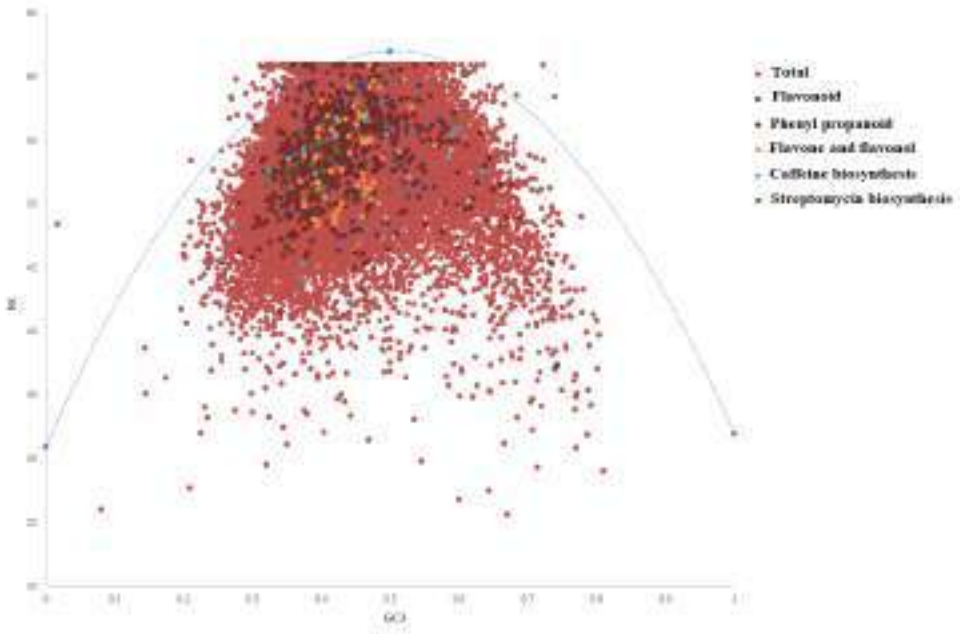

Fig.4.39.GC3 versus Nc plot for *Camellia sinensis* genome highlighting some important metabolic genes

data was executed to gain a better insight into the tentative factors dictating codon usage signatures in the tea genome. Correspondence analysis (CoA) based on RSCU data revealed that Axes 1 mainly contributed to total inertia (3.78) in the 59- dimensional hyperspace. Such an observation signified that Axis 1 has been the principal axis governing the variation

among genes. RSCU based scatter plots revealed an interesting fact that PHX and PLX genes clustered separately at the extreme opposite ends of the major axis of separation of genes (Figure 4.40). Such differential trends of codon usage among the potential highly expressed and lowly expressed genes signified a marked impact of translational selection to be influential on the *C. sinensis* genome.

Gene expression level was found to be another important determinant influencing codon usage pattern in genome of *C. sinensis* as obvious from the CoA plots. Such cases of gene expressivity affecting codon usage have been reported in *Caenorhabditis elegans*, *Drosophilla melanogaster*, and *Arabidopsis thaliana* (Duret and Mouchiroud, 1999), *Corynebacterium glutamicum* (Liu *et al*. 2010), etc. It was interesting to note that there was a strong significant and positive correlation between Axis 1 of RSCU and CAI (Table 4.10). Meaningful correlation analysis and RSCU based scatter plots clearly emphasized the influence of gene expression level to be a major determinant in framing the codon usage pattern of the *C. sinensis* genome.

The PHX genes were the ones to encode for proteins like disease resistance family protein, nucleic acid-binding protein, plant calmodulin-binding protein, SAP domain-containing protein, galactose oxidase, leucine-rich repeat, thioesterase superfamily protein, etc. On the other

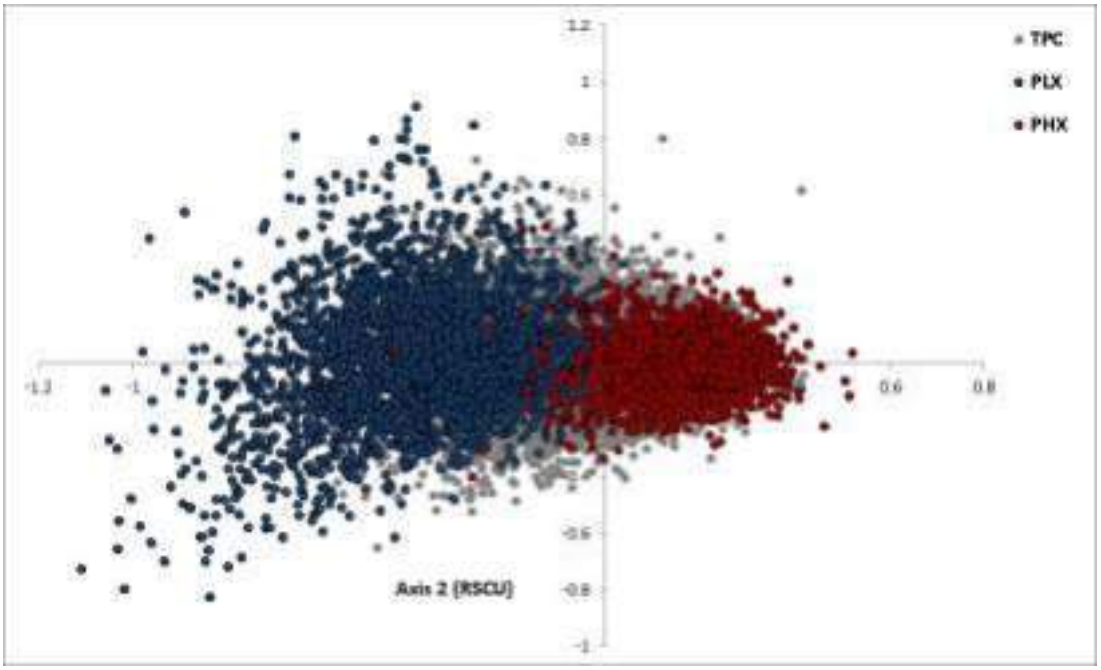

Fig.4.40.RSCU scatter plot for *Camellia sinensis*. Gray colored circles represent total protein coding genes (TPC), blue colored circles represent potentially lowly expressed genes (PLX), red colored circles represent potentially highly expressed genes (PHX)

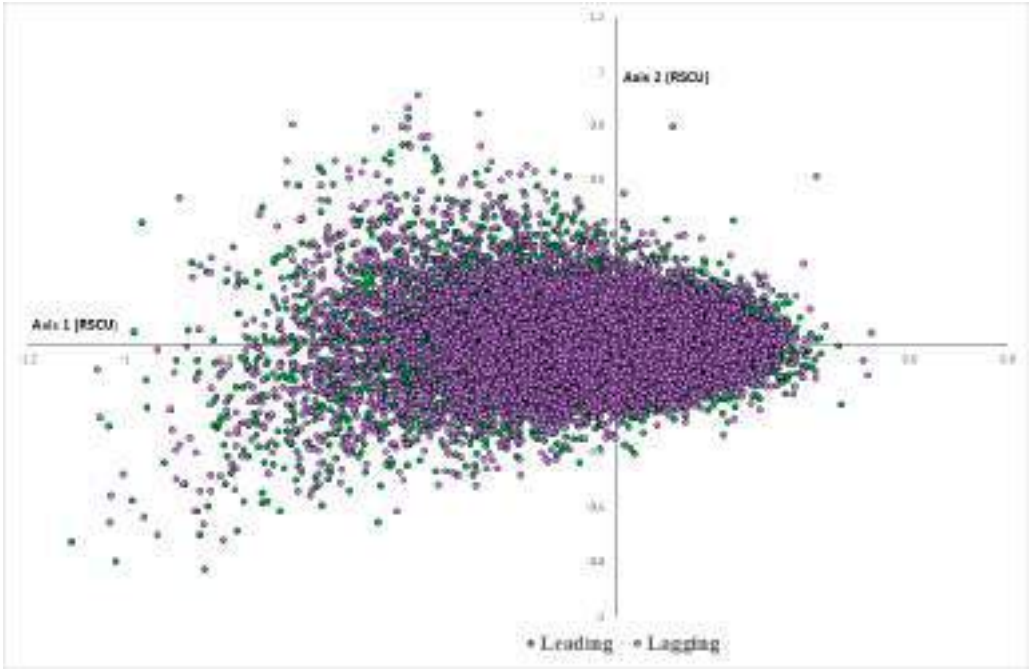

Fig.4.41. RSCU scatter plot for the leading and lagging strand-specific genes of *C. sinensis* showing absence of replicational selection. The green colored circle represents genes on the leading strand and the purple colored represents genes on the lagging strand

hand, the PLX genes were the ones to encode proteins like calcium-binding EF-hand family protein, cupredoxin superfamily, carbohydrate-binding X8 domain superfamily, etc. The list of PHX and PLX genes is given in Supplementary material 4 (SM4) .

The length of the coding sequence was found to correlate strongly with Axis 1 of RSCU (Table 4.10) suggesting the role of length of coding sequences in codon usage variation. The fact that replicational and transcriptional selection dictates the codon usage signatures in organisms, experiencing strand (leading/lagging strand of replication) asymmetry, has been well established (Das *et al*., 2006; Guo and Yu, 2007; Guo and Yuan, 2009; Lafay *et al*., 1999; McInerney, 1998; Wei and Guo, 2010). RSCU based scatter plots of leading and lagging strand-associated genes of the tea genome revealed that the complimentary strand -associated genes clustered together signifying the absence of replication-associated mutational pressure (Figure 4.41). Correlation studies of codon usage parameters directed towards the role of length of coding sequences in influencing codon usage behavior.

A significant positive correlation between gene expression level and gene length in the tea genome pointed towards the tendency of the potential highly expressed genes to be longer in stretch than the lowly expressed genes. However, *C. sinensis* being a

eukaryotic organism has deviated from the trends reported earlier in eukaryotes like *D. melanogaster* (Miyasaka, 2002), *C. elegans* (Marais and Duret, 2001) where the length of coding sequences was found to correlate negatively with the level of gene expression. Though PHX genes in eukaryotes may be shorter in length to minimize energy cost (Moriyama and Powell, 1998), the behavioral pattern of the genome prefers a longer stretch of highly expressed genes over shorter ones. Such an observation of the positive correlation between gene expression levels and gene length to minimize misincorporation and missense errors to increase translational accuracy has been formerly reported in prokaryotic organisms like *E. coli* (Eyre-Walker, 1996), *Pseudomonas aeruginosa* (Gupta and Ghosh, 2001), *Streptococcus pneumonia* (Hou and Yang, 2001), etc.

4.7.2. Exploring the sources of amino acid usage heterogeneity

Amino acids like Leucine (L), Serine (S) were used in the highest frequencies as evident from the amino acid usage (Figure 4.42). The other amino acids used in higher frequencies were Alanine (A), Lysine (K), Glutamic acid (E), Valine (V), and Glycine (G). The first and the second principal axes i.e., Axis 1 and Axis 2 were mainly found to govern the amino acid usage variation in the tea genome). The two major axes of RAAU data were found to exhibit significant correlations with hydropathicity index GRAVY and aromaticity of the gene products of *C. sinensis* genome (Table 4.10).

It was perceptible from RAAU data-based scatter plot that the protein-coding genes separated based on higher or lower values of Gravy and Aromo value. The genes with high GRAVY (high values of hydrophobicity) represented by red color and the genes with low GRAVY (low values of hydrophobicity) marked by blue color (Figure 4.43.a) clustered separately along the axis of separation. Similarly, the genes with higher aromo (purple color) and lower aromo (green color) clustered separately in the RAAU data-based scatter plot (Figure 4.43.b). Detailed analysis of amino acid usage-based parameters revealed the possible sources of amino acid usage variation in *C. sinensis*. Hydrophobicity and aromaticity were found to govern amino acid usage signatures. The role of gene expression level in amino acid usage in all the *C. sinensis* was obvious from the high correlation of CAI and the two major axes of separation of genes based on RAAU i.e., Axis 1 and Axis 2. Axis 1 and Axis 2 of RAAU

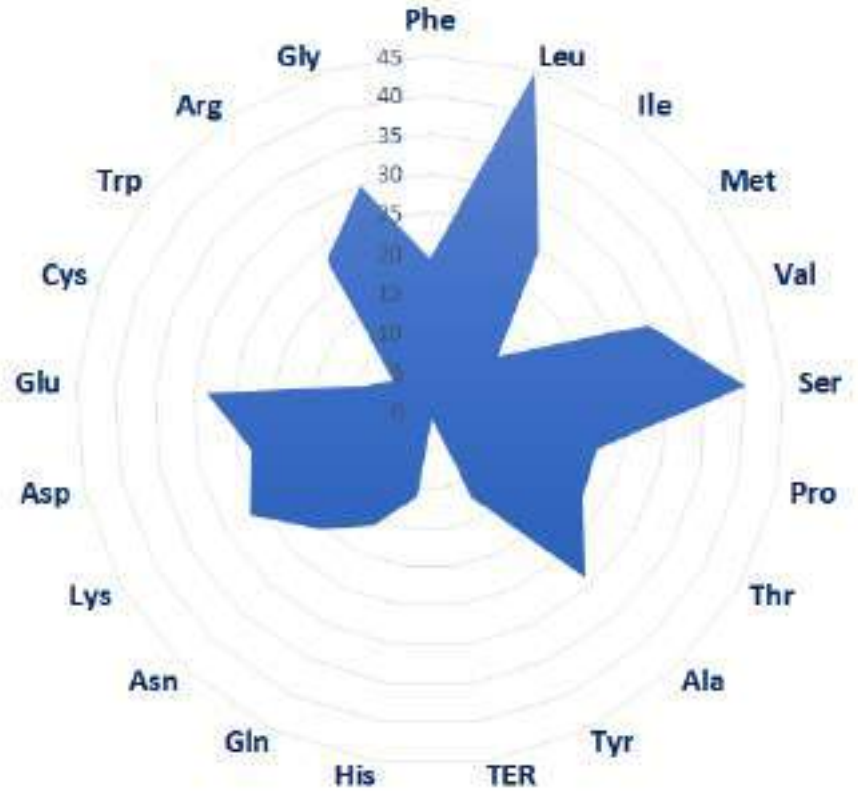

Fig.4.42.A radar plot showing amino acid usage in *C. sinensis* genome

were found to correlate significantly with gene expressivity, as estimated by CAI (Table 4.10). Thus, gene expression level not only influenced the codon usage signatures but also contributed substantially to producing the observed amino acid usage variations in the genome.

4.7.3. Protein-energy cost

It has been reported in prokaryotic organisms that the highly expressed gene products tend to be less expensive in terms of biosynthetic energy (protein -energy cost) and judiciously abide by the policy of cost-minimization (Roy *et al*., 2015; Seligmann, 2003). A significant positive correlation (Table 4.10) was noted between expression level (CAI) and energy cost in the tea genome ($r = 0.271$, $p<0.01$). Significant positive correlation of gene expression level with protein-energy cost indicated towards the utilization of less energetically costly amino acids in highly expressed genes which have also been reported in *Saccharomyces cerevisiae* (Kahali *et al*., 2007; Raiford *et al*., 2008). Increased utilization of inexpensive amino acids in highly expressed genes to lessen the use of expensive amino acids is true for prokaryotes (Akashi and Gojobori, 2002; Heizer *et al*., 2006).

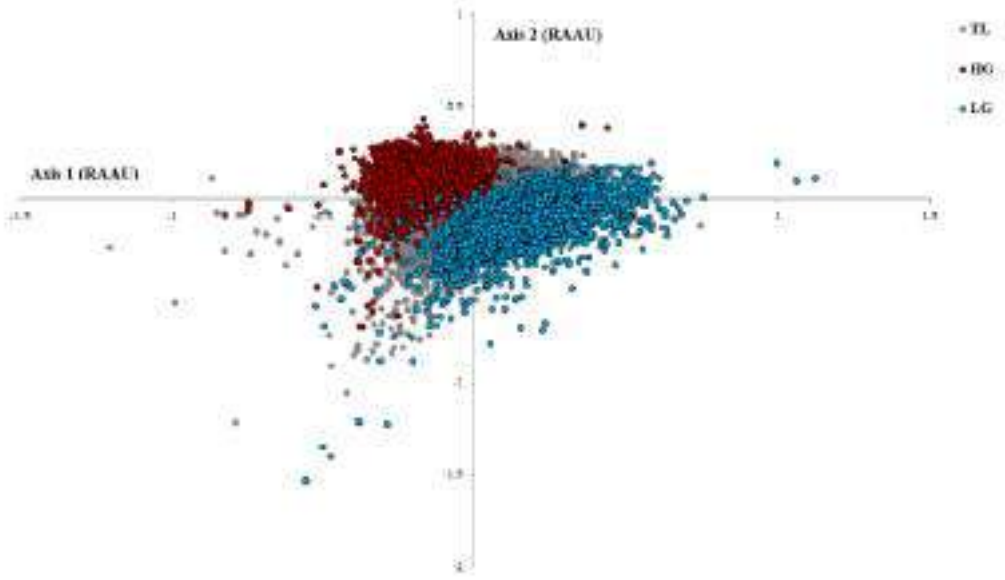

Fig.4.43.a.RAAU scatter plot showing genes with high hdrophobicity (red circle) and genes with low hydrophobicity (blue circle) of *C. sinensis* cluster separately . Grey coloured regions represents total genes

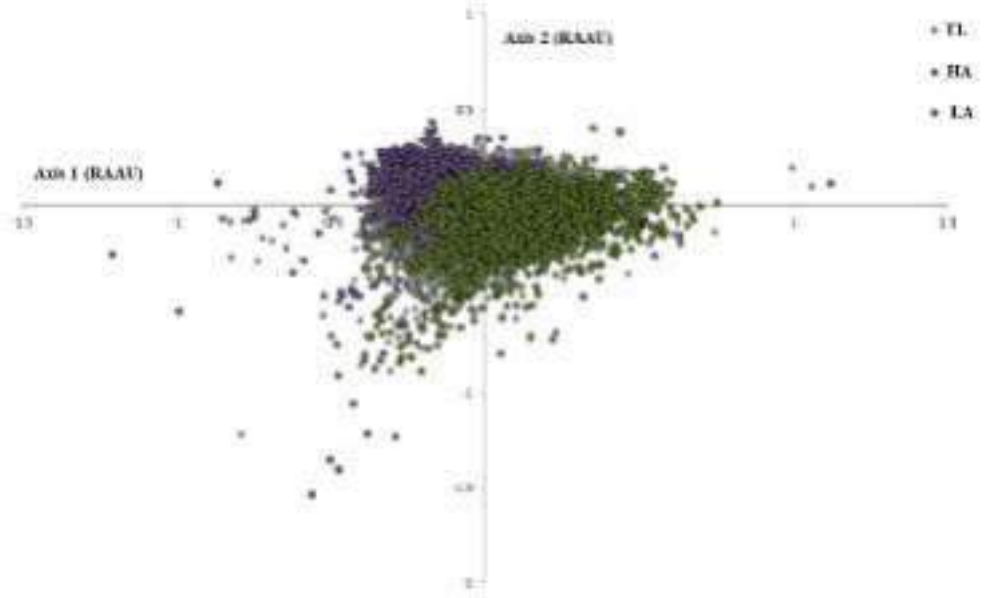

Fig.4.43.b.RAAU scatter plot showing genes with high aromo (green circle) and genes with low aromo (purple circle) of *C. sinensis* cluster separately .Grey coloured regions represents total genes

4.7.4. Eukaryotic Orthologous Groups (KOG)

Extensive profiling of Eukaryotic Orthologous Groups (KOG) revealed that the tea genome was mostly enriched with the proteins representing the KOG categories T (Signal transduction mechanisms), O (Posttranslational modification, protein turnover, chaperons), Q (Secondary metabolites biosynthesis, transport, and catabolism), G (Carbohydrate transport and metabolism), K (Transcription), C (Energy production and conversion),

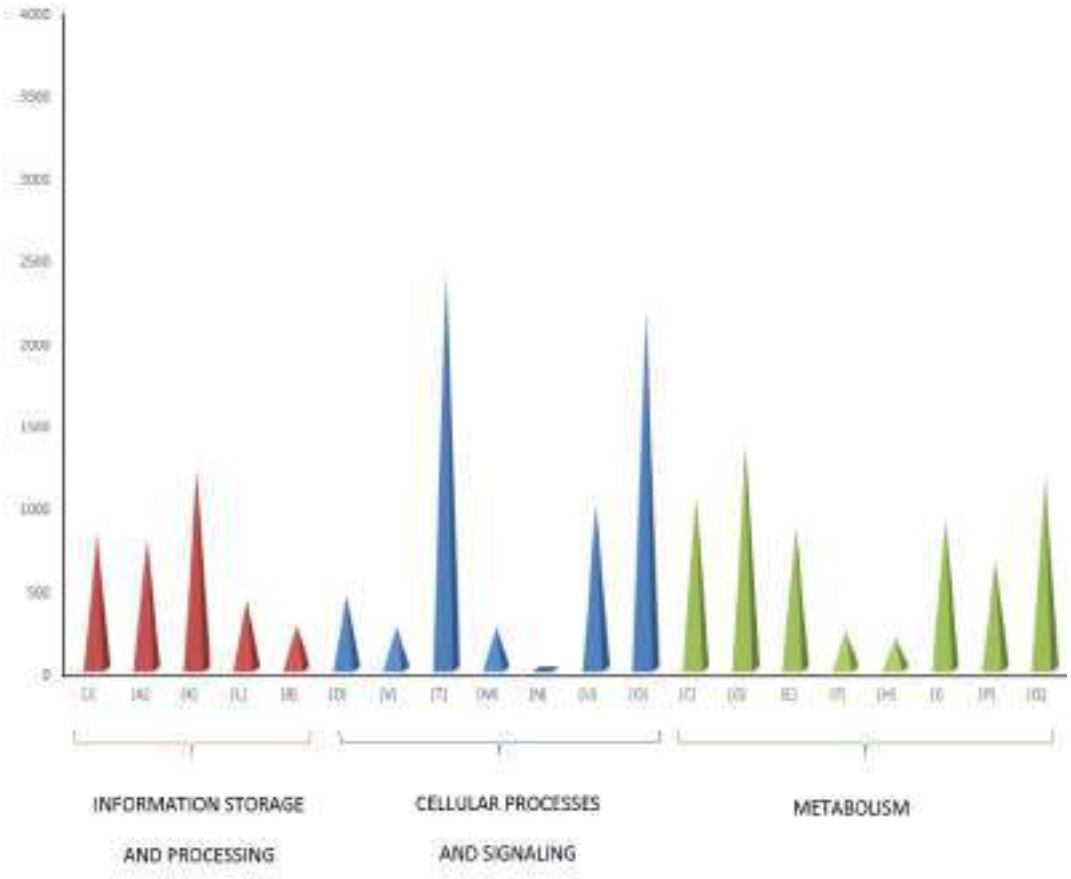

Fig.4.44. Characterization of Eukaryotic orthologous groups (KOGs) of the *C. sinensis* genome

U(Intracellular trafficking, secretion, and vesicular transport (Figure 4.44). However we did not consider the poorly characterized categories R (general function prediction) and S (unknown function).

Thus, the KOG categories in the majority were under the group like metabolism-related proteins (T, O, Q, G, and C), cellular processes and signaling (T, U, O), and information storage and processing (J, A, K).

4.8. *In silico* Polypharmacology of tea

4.8.1. Polypharmacological studies of Purple tea

4.8.1.1. Phytocompounds and their respective target proteins

GC/MS analysis of Purple tea (PT) extract yielded a list of 29 component chemicals (Table 5- Apendix D). Four phytocompounds, namely Hexadecanoic acid, 1,2,3-benzenetriol, Caffeine, and γ -Sitosterol, have more than 80% interaction probability with 19 target proteins, according to target prediction.

4.8.1.2. Phytocompounds and their Drug likeliness assessment

The details of this analysis are provided in Supplementary material 5 (SM5) and the results are discussed below briefly. Hexadecanoic acid and 1,2,3-benzenetriol have permeability across the blood-brain barrier (BBB), whereas caffeine and γ -Sitosterol do not. All of the substances, except for γ -Sitosterol, have a high GI absorption rate. None of the chemicals are permeability glycoprotein substrates (P -gp). One of the most important properties that influence drug absorption is solubility. Three topological approaches are employed in SwissADME to predict water solubility (Gfeller *et al.,* 2014). Hexadecanoic acid was found to be 'moderately soluble', 1,2,3-benzenetriol and Caffeine are 'very soluble,' while γ -Sitosterol is 'poorly soluble,'. The Lipinski rule-of-five is one of the filters used in this study to determine a compound's 'drug-likeliness.' The rule is followed by 1,2,3-benzenetriol and Caffeine, while Hexadecanoic acid and γ -Sitosterol each have one violation. However, Hexadecanoic acid has the greatest Bioavailability score (chance of a chemical having at least 10% oral bioavailability in rats or detectable Caco-2 permeability) among the substances (0.85).

4.8.1.3. Interaction with human proteins

The purple tea-derived compounds interacted with a total of 19 human proteins (Table 4.11, Figure 4.45). The network showing interaction of target proteins of Purple tea (PT) derived compounds and other proteins of *Homo sapiens* is shown in Figure 4.46. The compound 1, 2, 3-Benzenetriol interacted with epidermal growth factor (EFGR), with proto-oncogene tyrosine-protein kinase Fyn, and Carbonic anhydrase (CA) 1, 2, and 6. The EFGR activates numerous signaling pathways to transform extracellular inputs into suitable cellular responses. Fyn regulates cell growth and survival, cell adhesion, integrin-mediated signaling, cytoskeletal remodeling, cell motility, immunological response, and axon guidance, among other biological activities. The regulation of fluid secretion into the anterior chamber of the eye is linked to carbonic anhydrase (CA) proteins. These CA proteins contribute to intracellular pH regulation in the upper villous epithelium of the duodenum during proton-coupled peptide absorption and boost chloride-bicarbonate exchange activity, which aids CO2 transport into the blood.

Table 4.11. Target proteins associated with PT and their probability of interaction

Compound	Target	Common Name	Uniport ID	Probability
1,2,3-Benzenetriol	Tyrosine-protein kinase FYN	FYN	P06241	0.91
	Epidermal growth factor receptor erbB1	EGFR	P00533	0.91
	Carbonic anhydrase II	CA2	P00918	0.91
	Carbonic anhydrase I	CA1	P00915	0.91
	Carbonic anhydrase VI	CA6	P23280	0.91
Coffeine	Acetylcholinesterase	ACHE	P22303	0.85
	HERG	KCNH2	Q12809	0.85
	Adenosine A2a receptor	ADORA2A	P29274	0.85
γ-Sitosterol	Nuclear receptor ROR-gamma	RORC	P51449	0.92
	Niemann-Pick C1-like protein 1	NPC1L1	Q9UHC9	0.92
	LXR-alpha	NR1H3	Q13133	0.92
	Sterol regulatory element-binding protein 2	SREBF2	Q12772	0.83
	HMG-CoA reductase	HMGCR	P04035	0.83
Hexadecanoic Acid	Fatty acid-binding protein adipocyte	FABP4	P15090	0.94
	Peroxisome proliferator-activated receptor alpha	PPARA	Q07869	0.94
	Fatty acid-binding protein muscle	FABP3	P05413	0.94
	Fatty acid-binding protein epidermal	FABP5	Q01469	0.94
	Peroxisome proliferator-activated receptor delta	PPARD	Q03181	0.94
	Fatty acid-binding protein intestinal	FABP2	P12104	0.94

Acetylcholinesterase (ACE), potassium voltage-gated channel subfamily H member 2 (KCNH2), and adenosine A2a receptor (ADORA2A) are all associated with caffeine . Through fast hydrolysis of acetylcholine delivered into the synaptic cleft, ACE plays a key role in signal transduction at the neuromuscular junction. Additionally, this protein was linked to neuronal death.

Proteins like Fatty acid-binding protein adipocyte (FABP4), Peroxisome proliferator-activated receptor alpha (PPARA), Fatty acid-binding protein muscle (FABP3), Fatty acid-binding protein epidermal (FABP5), Peroxisome proliferator-activated receptor delta (PPARD), and Fatty acid -binding protein intestinal (FABP2) were found to form a direct network with Hexadecanoic Acid. In adipocytes, these proteins are linked to the Lipid Transport Protein. PPARD

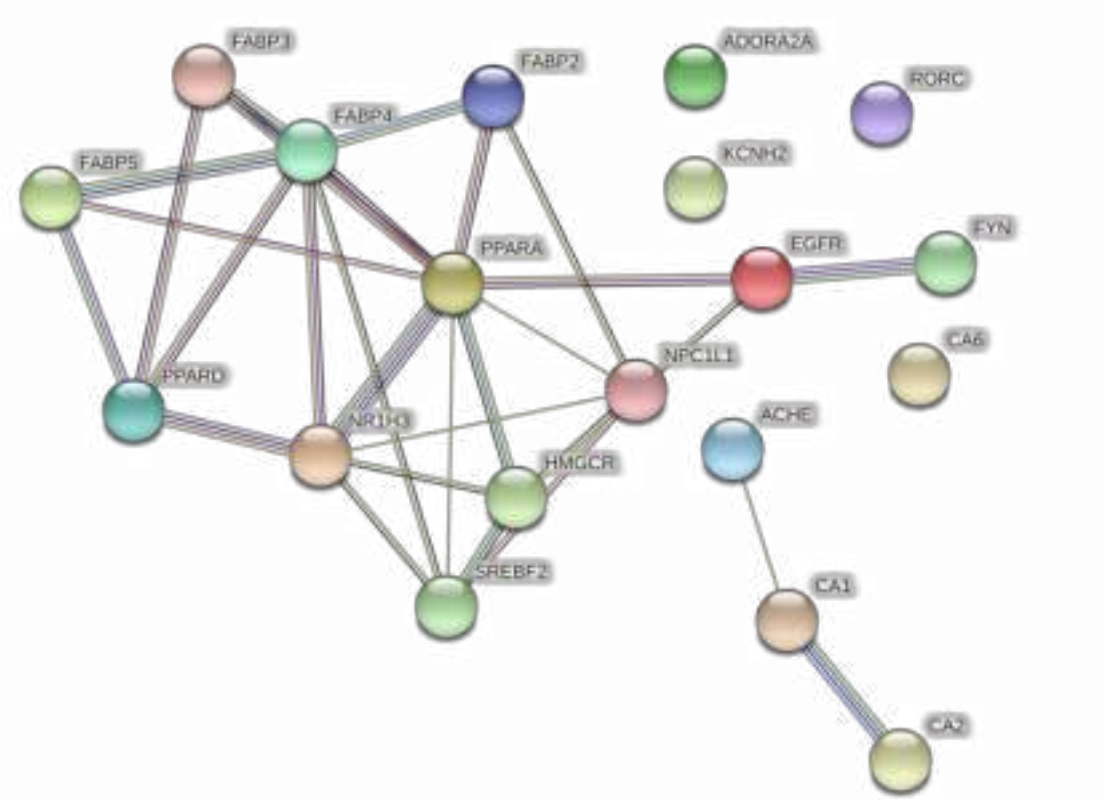

Fig 4.45. Interaction between the target proteins of Purple tea (PT) derived compounds

also functions as a Ligand-activated transcription factor that prefers polyunsaturated fatty acids like gamma -linoleic acid and eicosapentaenoic acid. They regulate the peroxisomal beta-oxidation pathway of fatty acids after activation.

γ-Sitosterol was found to interact with nuclear receptor ROR-gamma (RORC), Niemann-Pick C1-like protein 1 (NPC1L1), LXR-alpha (NR1H3), Sterol regulatory element-binding protein 2 (SREBF2), and HMG -CoA reductase (HMGCR). These interacting proteins are essential for the body's homeostasis to be maintained. RORC controls cellular differentiation, immunity, peripheral circadian rhythm, lipid, steroid, xenobiotic, and glucose metabolism, among other things. Through cholesterol uptake across the plasma membrane of the intestinal enterocyte, NPC1L1 plays a critical function in cholesterol homeostasis. Multiple lipid transport abnormalities result from a loss-of-function mutation in this protein.

With a p-value of less than 0.05, GO analysis of the target genes reveals that they are related with 152 BP, 62 MF, and 32 CC (Supplementary Material 6: ST 2, 3 and 4) The top 15 GO results based on gene count are shown in Figure 4.47. The majority of the genes

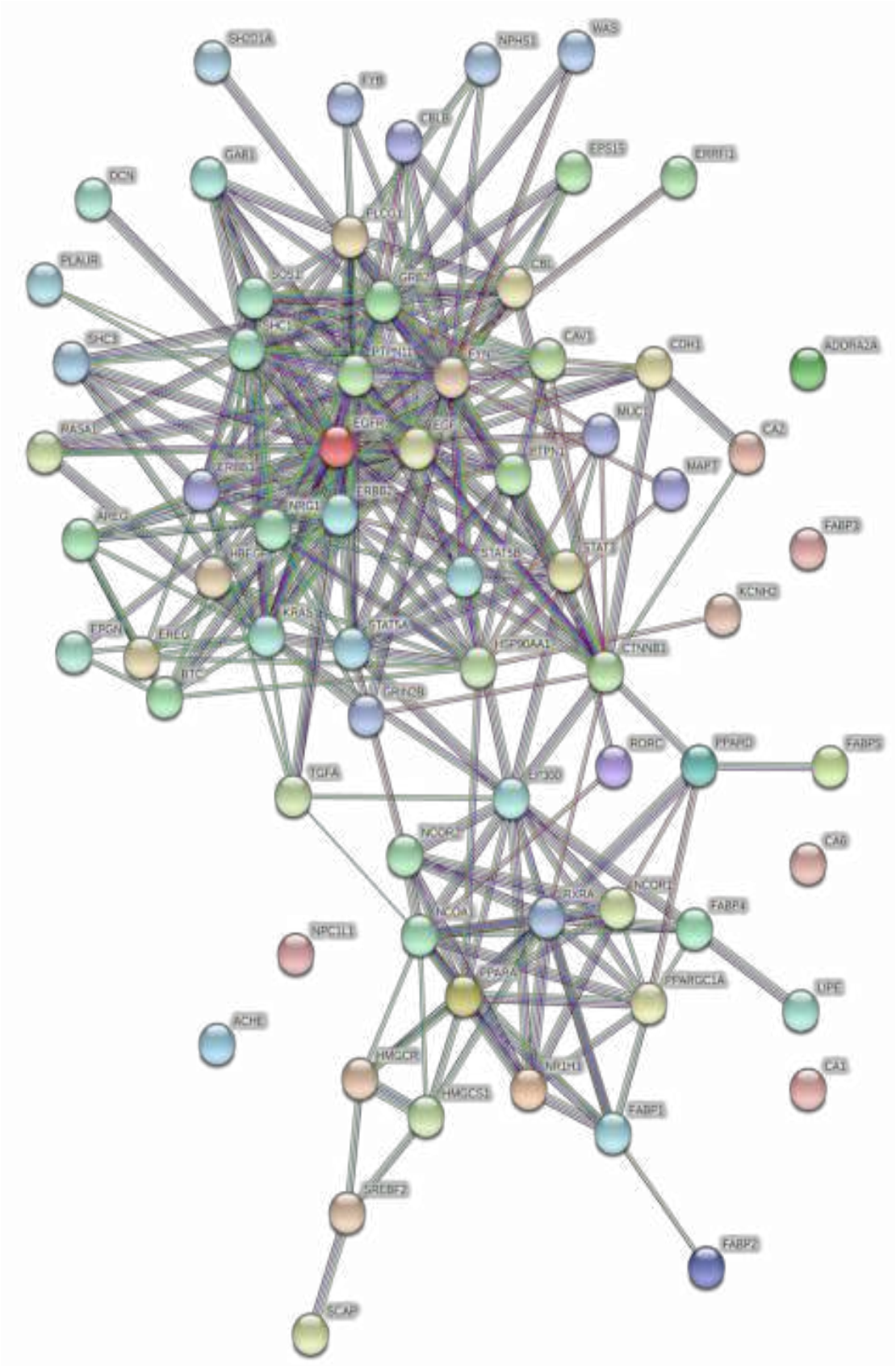

Fig 4.46. PPI network showing interaction of target proteins of Purple tea (PT) derived compounds and other proteins of *Homo sapiens*

(58) are involved in the molecular function of protein binding. The biggest number of genes (32), according to cell component analysis, are found in the cytosol.

As a result of this investigation, it was discovered that human proteins that interact directly with the selected compounds are functionally important and are primarily linked to cell growth and survival, axon guidance, neural transport, immune response, cell motility, lipid metabolism, and cholesterol metabolism. Furthermore, KEGG pathway enrichment analysis shows that PT targets are associated with pathways like ErbB signaling (21 proteins), PPAR signaling (8 proteins), Insulin signaling (10 proteins), Ras signaling (12 proteins), Natural killer cell-mediated cytotoxicity (9 proteins), Neurotrophin signaling (8 proteins), and others (Supplementary Material 7 : ST 5, SF 1,2,3,4,5 and 6).

4.8.1.4. Role in Diseases

4.8.1.4.1. Gut-Brain Axis

Compounds produced from PT were discovered to interact with many genes linked to various illnesses like Hypercholesterolemia (increased levels of HDL cholesterol in the blood), Type 2 diabetes, edema, cardiovascular illnesses, colorectal cancer, Alzheimer's disease, atherosclerosis, insulin resistance, obesity, hyperlipidemias, and several other metabolic syndromes (Supplementary Material 6: ST 1, Figure 4.48). As a result, these proteins were divided into two categories: (a) gut-related disorders and (b) neurological diseases, indicating the impact of PT on the Gut-Brain Axis (GBA). The GBA is a two-way communication system that connects the enteric and neurological systems. This is the link between the brain's emotional and cognitive centers and the functioning of the peripheral intestine.

We discovered 34 proteins that interacted with PT-derived chemicals were linked to Type 2 Diabetes and Edema while evaluating the data (Supplementary Material 6: ST 1 TGFA, EGFR, ERBB3, PLCG1, STAT5B, EGF, GAB1, NRG1, GRB2, SOS1, and HBEGF were discovered to be linked to the ErbB signaling pathway among the 34 proteins studied (Supplementary Material 7 :ST 5, SF 2). This pathway is important for directing intracellular signaling cascades, either directly or indirectly, in a variety of hereditary diseases (Holbro and Hynes, 2004). Peptides linked to epidermal growth factor (EGF) connect to ErbB receptors, activating intrinsic kinases (Fiske *et al.,* 2009). These Fyn-kinases phosphorylate or dephosphorylate a

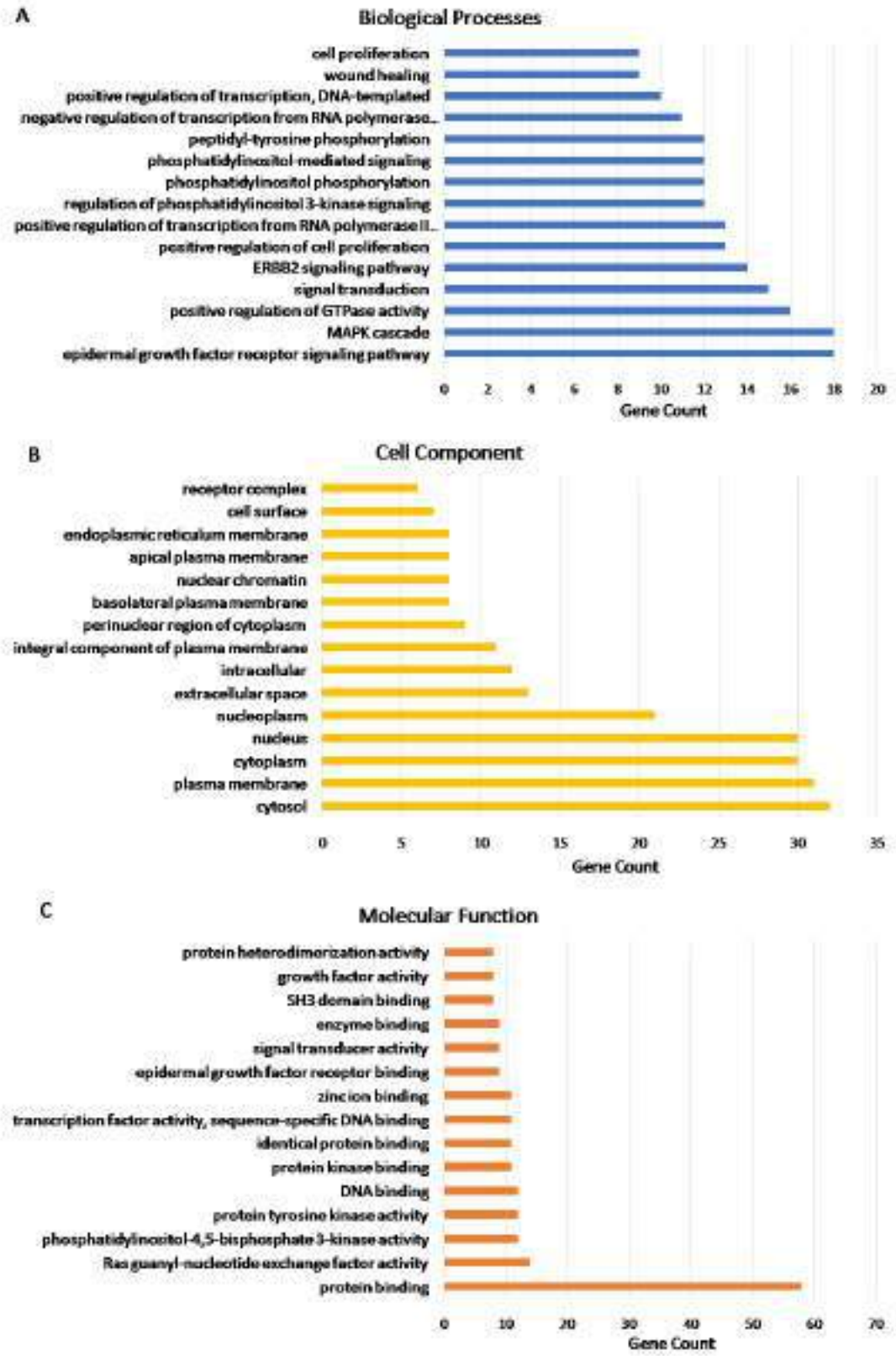

Fig 4.47. Gene Ontology Enrichment Analysis. The top 15 results (with p-value < 0.05) of Gene ontology enrichment analysis of the target genes show them to be associated with different (A) Biological Processes, (B) Cell Components and (C) Molecular Functions

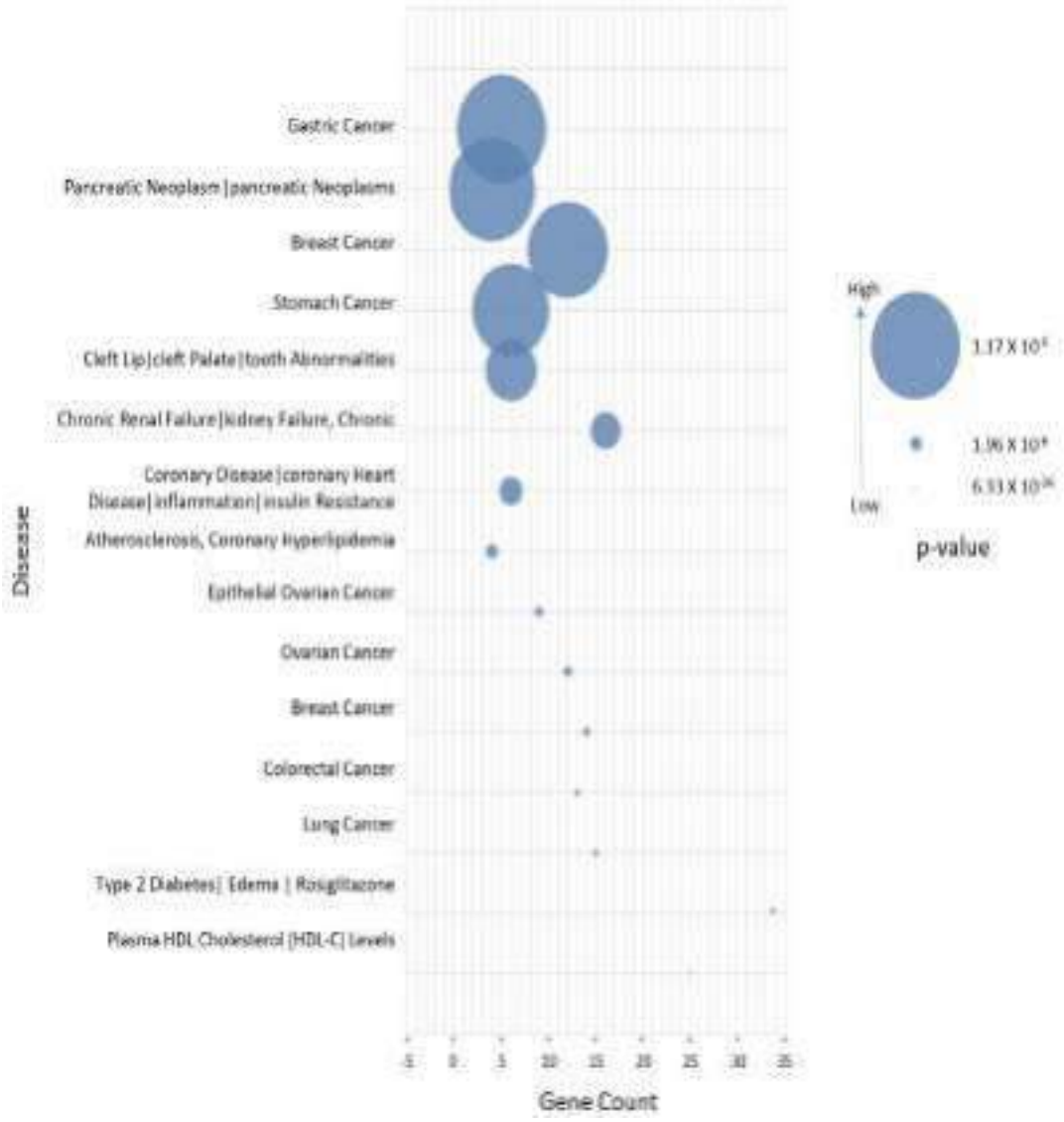

Fig 4.48. Functional Annotation (Disease). The top 15 results (with p-value < 0.05) of Functional annotation of target genes show their association with diseases such as plasma HDL cholesterol (Hypercholesterolemia), Type 2 diabetes, edema, coronary diseases, etc. Size of the bubbles indicate the p-value

variety of target proteins, regulating the cell-cell contact cascade (Schreier *et al.,* 2021). Thus, the intrinsic signaling cascades broadly linked with the Gut-Brain axis were revealed to be modulated by both EGFR related to ErbB signaling and Fyn kinase, which was discovered to be the principal target of PT-derived components.

4.8.1.4.2. Immune Booster

A total of 9 targets namely SHC3, SHC1, SH2D1A, KRAS, PTPN11, GRB2, FYN, PLCG1, and SOS1 were found to be related with 'Natural killer cell-mediated cytotoxicity in our body' in a KEGG pathway enrichment study of the PT targets (Supplementary Maetrial 7: ST 5, SF 6). Some of the

key immune cells linked with the innate immune system include natural killer (NK) cells, macrophages, and other phagocytes. These cells serve a key part in cell-mediated cytotoxicity that is mediated by antibodies (ADCC). Furthermore, NK cells can kill tumor cells without the need for immunization or immune activation (Waldhauer and Steinle, 2008). In the presence of increased reactive oxygen species (ROS), these NK cells are especially vulnerable to and can lose their ability to destroy antigens (Nakamura and Matsunaga, 1998). PT can also protect NK cells from ROS-induced damage, enhancing NK cell-mediated immunity. Furthermore, anti-oxidant properties are invariably linked to apoptosis induction and cancer inhibition (Harris and De Nicola, 2020). The PT interacts with proteins involved in the PI-3 kinase and MAP kinase pathways (Downward, 2004). These pathways play an important role in cell growth, survival, aging, and apoptosis regulation. As a result, PT chemicals aid in the general preservation of the normal cell cycle, avoiding the start of uncontrolled cellular proliferation, which leads to cancer.

4.8.1.4.3. Alzheimer's disease management

Alzheimer's disease (AD) is a progressive neurological illness that causes brain atrophy and cell death. This is the most prevalent form of dementia, and it causes patients' intellectual abilities to deteriorate over time, as well as causes significant damage to their behavioral and social skills. All of these things have an impact on a person's ability to think and function independently in the long run. The tau protein has been linked to the development of Alzheimer's disease. This protein is mostly found in brain cells and is linked to the production of microtubules (Wenk, 2003). These microtubules aid in the transportation of vital elements such as nutrients from one section of the nerve cell to another. Tau can diffuse throughout the brain as an oligomer, allowing chemical and electrical messages to pass via synapse (Fiske *et al.,* 2009). When tau accumulates, however, clinical consequences similar to those seen in Alzheimer's disease develop. In Alzheimer's disease, two primary accumulations have been identified: (a) β -amyloid (A) plaques and (b) tau neurofibrillary tangles (NFTs). These disrupt the brain's and neurons' homeostasis, resulting in Alzheimer's disease (Wenk, 2003). Tyrosine-protein kinases have been discovered in several studies on this disease. Fyn is crucial in the treatment of Alzheimer's disease. They regulate

cell growth and survival, cell adhesion, integrin-mediated signaling, cytoskeletal remodeling, cell motility, immunological response, and axon guidance, among other biological activities (Shirazi and Wood, 1993). Furthermore, Fyn regulates the nervous system's functionality by phosphorylating proteins involved in synaptic signal transmission. Recent research has also demonstrated the role of Fyn in preventing the onset of Alzheimer's disease via regulating tau aggregation (Briner *et al.,* 2020). Fyn knockout mice also lacked normal myelination, according to previous research. The tau protein interacts with the microtubules, and Fyn interacts with it. Both Fyn and tau are found in oligodendrocyte rafts, generating a Fyn -tau complex. Tau is attached to the SH3 domain of Fyn (Fyn has both SH2 and SH3 domains) on one side and the microtubules on the other in this complex. The interaction of Fyn-tau microtubules permits oligodendrocytes to link with nearby neurons throughout their extension process (Rojo *et al.,* 2006).

The phytocompounds in PT have been discovered to interact with Fyn kinase through the ErbB signaling pathway. Furthermore, Functional Annotation reveals that five of PT drugs' key targets (ACHE, FYN, HMGCR, PPARA, PPARD) are linked to Alzheimer's disease (Supplementary Material 6: ST 1). These chemicals can also control axonal outgrowth, neurotransmitter retrograde transmission, and the neuro-myelination process. As a result, we might speculate that PT can maintain neuronal homeostasis by interacting with the Fyn kinase protein, preventing or at least delaying the onset of AD and associated neurological diseases.

4.8.1.4.4. Hyperglycemia

People with diabetes might be affected by high blood sugar (hyperglycemia). Hyperglycemia is caused by several causes that might aggravate Type 2 Diabetes (T2D). T2D is a chronic condition that affects glucose metabolism and renders the body insulin resistant. This means that the cells will not employ insulin to utilize glucose. The insulin signaling pathway is linked to Fyn-kinase and other Src-kinase family proteins. They work with the help of the plasma membrane's lipid raft microdomains (PM). Post-transcriptional changes shift Fyn kinase from the cytoplasm to the plasma membrane (Dai *et al.,* 2017). They are specifically found within the lipid rafts. Fyn can phosphorylate caveolin protein in lipid rafts in response to insulin stimulation. Fyn also interacts with flotillin and other proteins found in lipid rafts (Dai *et al.,* 2017). Fyn aids in

fatty acid absorption and oxidation, which is directly linked to insulin signaling, through these interactions (Dai *et al.,* 2017). Furthermore, constitutively active Src family proteins can block pyruvate kinase, the enzyme that catalyzes the final step of glycolysis (phosphor-enol pyruvate →). As a result, the use of glucose as a source of energy is inhibited. Furthermore, inhibiting Fyn and other Src kinases can stop adipocyte differentiation from happening. These findings confirmed Fyn kinase's role in insulin signaling. As previously established, PT drugs interact with the Fyn kinase via the ErbB signaling pathway to keep it active in adipocyte development and insulin signaling. Accurate insulin signaling will keep glucose uptake and utilization in check, reducing the risk of insulin resistance and type 2 diabetes. LIPE, PTPN1, SHC3, SHC1, CBLB, KRAS, GRB2, CBL, SOS1, and PPARGC1A are among the Insulin signaling pathways ten targets (Supplementary Material 7 , -ST 5, SF 4.Furthermore, PT drugs interact with eight proteins of the peroxisome-proliferator-activated-receptors (PPAR) signaling pathway (Supplementary Material 7: ST 5, SF 3) that play an essential role in adaptive thermogenesis (FABP1, FABP2, RXRA, FABP4, FABP5, NR1H3, PPARA, and PPARD). This is the production of body heat in stressful situations (Polvani *et al.,* 2016). When people are stressed, their rate of gluconeogenesis increases, which boosts blood glucose levels. Insulin is secreted to generate heat (or energy) from glucose. In the absence of stress, the entire situation is 'cooled down.' In the presence and absence of stress, adaptive thermogenesis mediated by PPAR aids in maintaining insulin homeostasis (Polvani *et al.,* 2016). PPAR interacts with the RAS signaling system and is involved in insulin signaling, as well as glucose and lipid metabolism (Polvani *et al.,* 2016). As a result, regular PT consumption can help to maintain a healthy blood glucose level in the body, reducing the risk of type 2 diabetes and insulin resistance.

4.8.1.4.5. Hypercholesterolemia, obesity, and coronary heart disease management

The prevalence of hypercholesterolemia, obesity, and coronary heart disease is quickly rising over the world. Excess fat buildup in tissues other than adipose tissues is sometimes referred to as obesity. Fat is stored mostly in the liver and skeletal muscles, which can result in obesity-related inflammation, high blood cholesterol, and coronary heart disease. Obesity and related pathophysiological

problems are linked to several biological pathways. One of them is PPAR, which is still linked to lipid metabolism, adipocyte development, fatty acid breakdown, and bile acid production (Polvani *et al.,* 2016). As a result, PPAR is linked to maintaining cholesterol homeostasis and fat metabolism (Polvani *et al.,* 2016). Furthermore, Fyn-significance kinases in obesity and associated illnesses have been confirmed in recent investigations (Jung et al., 2008). Furthermore, functional annotation and GO enrichment analysis revealed that PT compounds influenced 'plasma HDL cholesterol (HDL-C) levels and 'coronary heart disease' linked genes (25 genes and 6 genes, respectively) (Supplementary Material 6: ST 1). As a result, we can speculate that PT drugs that interact with PPAR and Fyn play a direct role in reducing obesity and coronary heart disease.

4.8.2. Polypharmacological studies of the mature leaf of *C. sinensis*

4.8.2.1. Identification of target proteins and physiochemical analysis of the derived Phyto-compounds

Analysis of the phytocompounds from mature leaf (ML) extract of *C. sinensis* (Table 1-4: Appendix D), revealed five compounds with interaction probability of more than 60% i.e., Caffeine, Hexadecanoic acid, Octadecanoic acid, Olean-12-en- 3-one, and Stigmasta-7,22-dien-3-ol (Table 4.12). Out of the five compounds (Supplementary Material 8), only hexadecenoic acid showed blood-brain barrier (BBB) permeability. Gastro-intestinal (GI) absorption was higher in compounds like Caffeine, Hexadecanoic acid, and Octadecanoic acid. No compounds among five serve as the substrate of permeability glycoprotein (P-gp). Here, only Caffeine abides by the rule of drug likeliness implemented using Lipinski rule-of-five. Whereas, the other compounds showed one violation each. The compounds like hexadecenoic acid and Octadecanoic acid showed the highest Bioavailability score (0.85).

4.8.2.2. Phyto compounds and their interaction with human proteins

A total of 12 human proteins was found to directly interact with the phytocompounds found in mature tea leaf (Figure 4.49). The Protein-Protein interaction data obtained from STRING Database with minimum required interaction score (0.400) and maximum number of interactors set to none is provided in Supplementary Material 9. The network showing interaction of target proteins of Mature Tea Leaf (ML) derived compounds and other proteins of *Homo sapiens* with minimum required interaction score

Table 4.12: Target proteins associated with ML and their proability of interaction

Compound	Target	Common name	Uniprot ID	Probability*
Caffeine	Acetylcholinesterase	ACHE	P22303	0.85
	HERG	KCNH2	Q12809	0.85
	Adenosine A2a receptor	ADORA2A	P29274	0.85
Hexadecanoic acid	Fatty acid binding protein adipocyte	FABP4	P15090	0.94
	Peroxisome proliferator-activated receptor alpha	PPARA	Q07869	0.94
	Fatty acid binding protein muscle	FABP3	P05413	0.94
	Fatty acid binding protein epidermal	FABP5	Q01469	0.94
	Peroxisome proliferator-activated receptor delta	PPARD	Q03181	0.94
	Fatty acid binding protein intestinal	FABP2	P12104	0.94
Octadecanoic acid	Peroxisome proliferator-activated receptor alpha	PPARA	Q07869	0.93
	Peroxisome proliferator-activated receptor delta	PPARD	Q03181	0.93
	Fatty acid binding protein adipocyte	FABP4	P15090	0.71
	Fatty acid binding protein epidermal	FABP5	Q01469	0.71
Olean-12-en- 3-one	Cytochrome P450 19A1	CYP19A1	P11511	0.61
Stigmasta-7,22-dien-3-ol	Androgen Receptor	AR	P10275	0.68

(0.900) and maximum number of interactors set to 50 is shown in Figure 4.50. The interaction of compounds like caffeine and hexadecenoic acid is already mentioned in the above section (4.8.1.3). Octadecanoic acid was also found to interact with Peroxisome proliferator-activated receptor alpha (PPARA), Peroxisome proliferator-activated receptor delta (PPARD), Fatty acid-binding protein adipocyte (FABP4), Fatty acid-binding protein, and epidermal (FABP5). Both hexadecenoic acid and Octadecanoic acid showed a direct network with proteins directly associated with Lipid transport protein in adipocytes.

Olean-12-en- 3-one was directly

related to Cytochrome P450 family 19 (CYP19A1) which catalyzes the formation of aromatic C18 estrogens from C19 androgens.

Stigmasta-7,22-dien-3-ol on the other hand showed direct relation with Androgen Receptor (AR). These receptors are ligand-activated transcription factors that control cellular proliferation and differentiation in target tissues by regulating eukaryotic gene expression. The activity of transcription factors is influenced by the presence of binding coactivator and corepressor proteins. NR0B2 inhibits the activation of transcription. HIPK3 and ZIPK/ DAPK3 activate it but do not phosphorylate it.

The Gene Ontology results based on gene count showed a majority of the genes to be involved in the molecular function (Supplementary Materia10) of lipid binding and Biological processes (Supplementary Materia 11) like the triglyceride catabolic process (Figure 4.51).

4.8.2.3. Compounds derived from the mature leaf

The compounds were found to interact with several genes linked to various diseases with (Supplementary Materia 13) like Type 2 Diabetes, Obesity, Polycystic ovary syndrome, higher level of plasma HDL cholesterol (Hypercholesterolemia), high level of triglycerides, Alzheimer's disease,

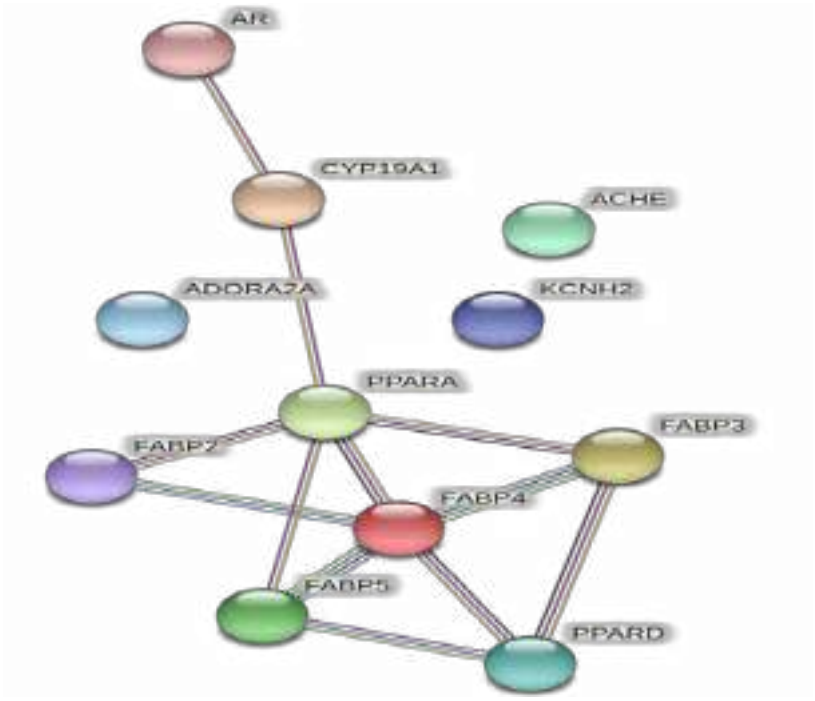

Fig 4.49. Phytocompound-target interaction and PPI network of *C. sinensis* Mature leaf (ML) associated proteins

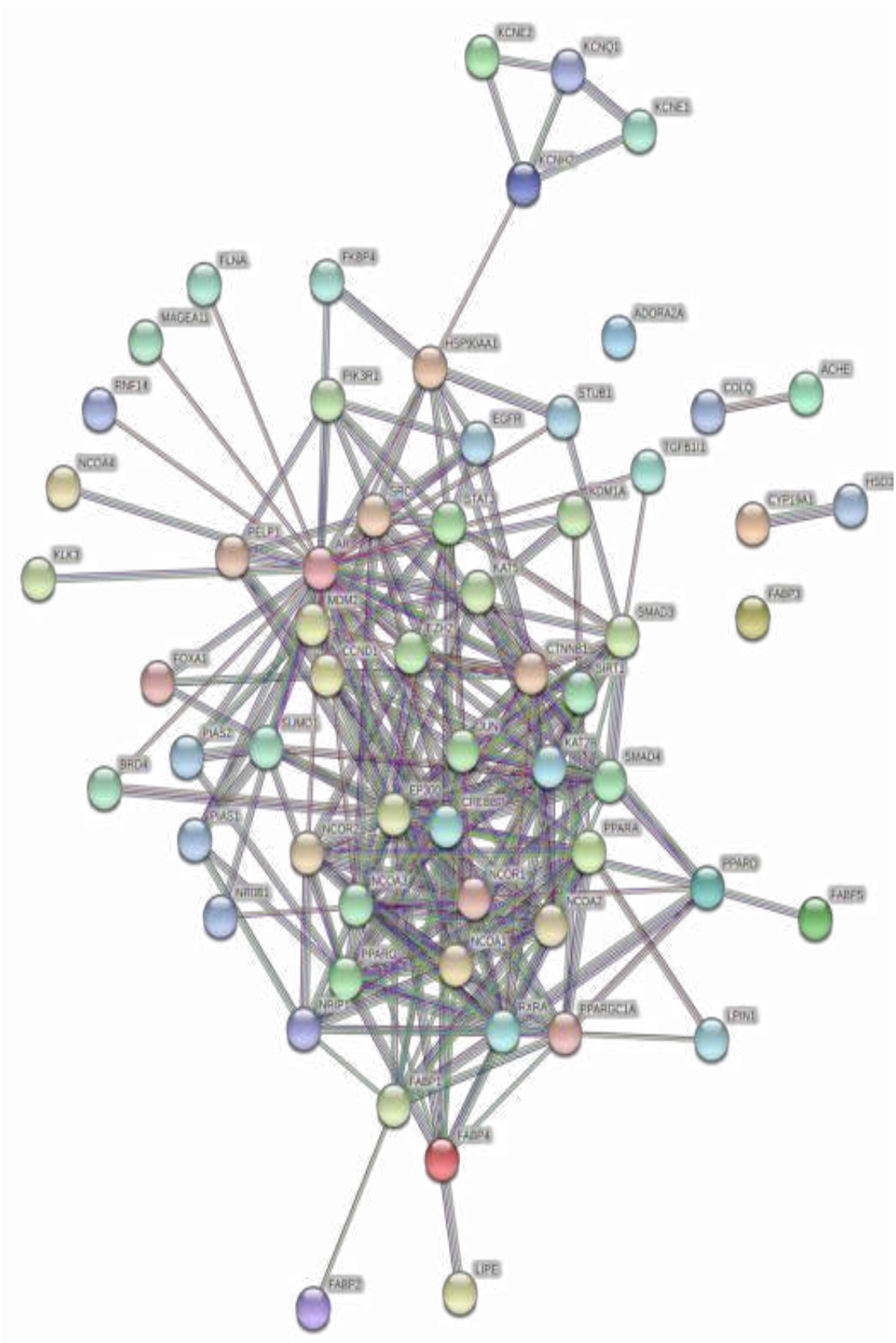

Fig 4.50. PPI network showing interaction of target proteins of Mature Tea Leaf (ML) derived compounds and other proteins of *Homo sapiens*

atherosclerosis, insulin resistance, metabolic syndrome, cardiovascular and coronary heart diseases, ovarian cancer, fatty liver, anxiety disorder, hypertension, and several other diseases. The top 15 results with p-value < 0.05 of Functional annotation of target genes is provided in Figure 4.52 which shows their association with diseases such as Type 2 diabetes, edema, Polycystic ovary syndrome, plasma HDL cholesterol (Hypercholesterolemia), coronary diseases, etc. Decreasing the size of the bubbles as well as the color gradiation from light pink to deep red indicates decrease in the the p-Value.

Among the 12 studied proteins, 11 proteins interacted with compounds associated with Type 2 diabetes and edema, 7 proteins were linked with Obesity, 4 proteins were linked with polycystic ovary syndrome, 5 proteins were linked with plasma HDL cholesterol, Alzheimer's disease, and hypertension (Supplementary Material 12).

Diabetes, cardiovascular disease, and cancer are the most frequent and complex diseases that are currently having a significant impact on global health issues that lead to death. Diabetic individuals are at a greater risk of heart problems, and cardiovascular complications are the primary cause of mortality in diabetic patients (Mirza *et al.,* 2019). Furthermore, type 2 diabetes can lead to an increase in cholesterol issues, blood pressure, and obesity, all of which can lead to cardiovascular difficulties. Aside from the increased risk of heart disease in diabetic people, researchers are looking into a possible link between diabetes and cancer. Millions of people have died as a result of cancer, which has now become a major global threat. Diabetes is thought to increase the incidence of certain malignancies, which may be attributed to hyperinsulinemia (Volkers, 2000). Diabetic people are at a much-increased risk for several forms of cancer. Type 2 diabetes has been linked to a variety of malignancies, and it has been claimed that treatments for type 2 diabetes may directly or indirectly affect cancer cells (Giovannucci *et al.,* 2010).

PCOS (polycystic ovarian syndrome) is a prevalent and complex endocrine disorder in women that is regarded as one of the leading causes of infertility (Cheshmeh *et al.,* 2021). PCOS is linked to hyperandrogenism, hyperinsulinemia, and hypothalamic–pituitary–adrenal axis dysfunction. Changes in the pituitary–ovarian axis, ovulation difficulties, and irregular menstruation affect these women, and they also experience mood swings,

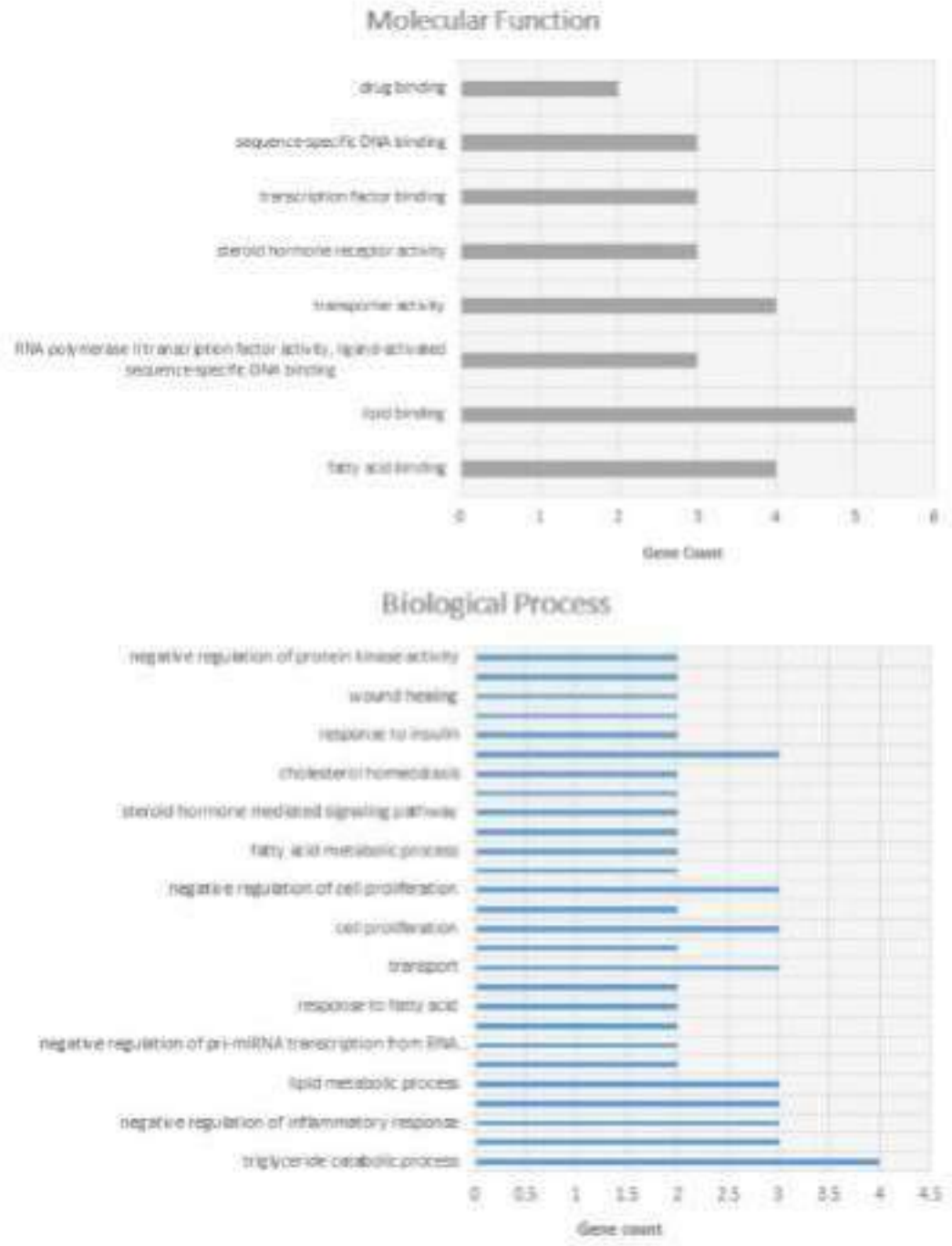

Fig 4.51. Gene Ontology Enrichment Analysis. The top 15 results (with p-value < 0.05) of Gene ontology enrichment analysis of the target genes show them to be associated with different Molecular Functions and Biological Processes

anxiety, despair, and mood changes (Machado *et al.,* 2020). Inter-linked metabolic disorders, such as weight gain and obesity, Insulin resistance, type 2 diabetes, and cardiovascular disease (Zhang *et al.,* 2020) are common in patients with PCOS. According to recent studies, genetic factors have a major role in the development of obesity and insulin

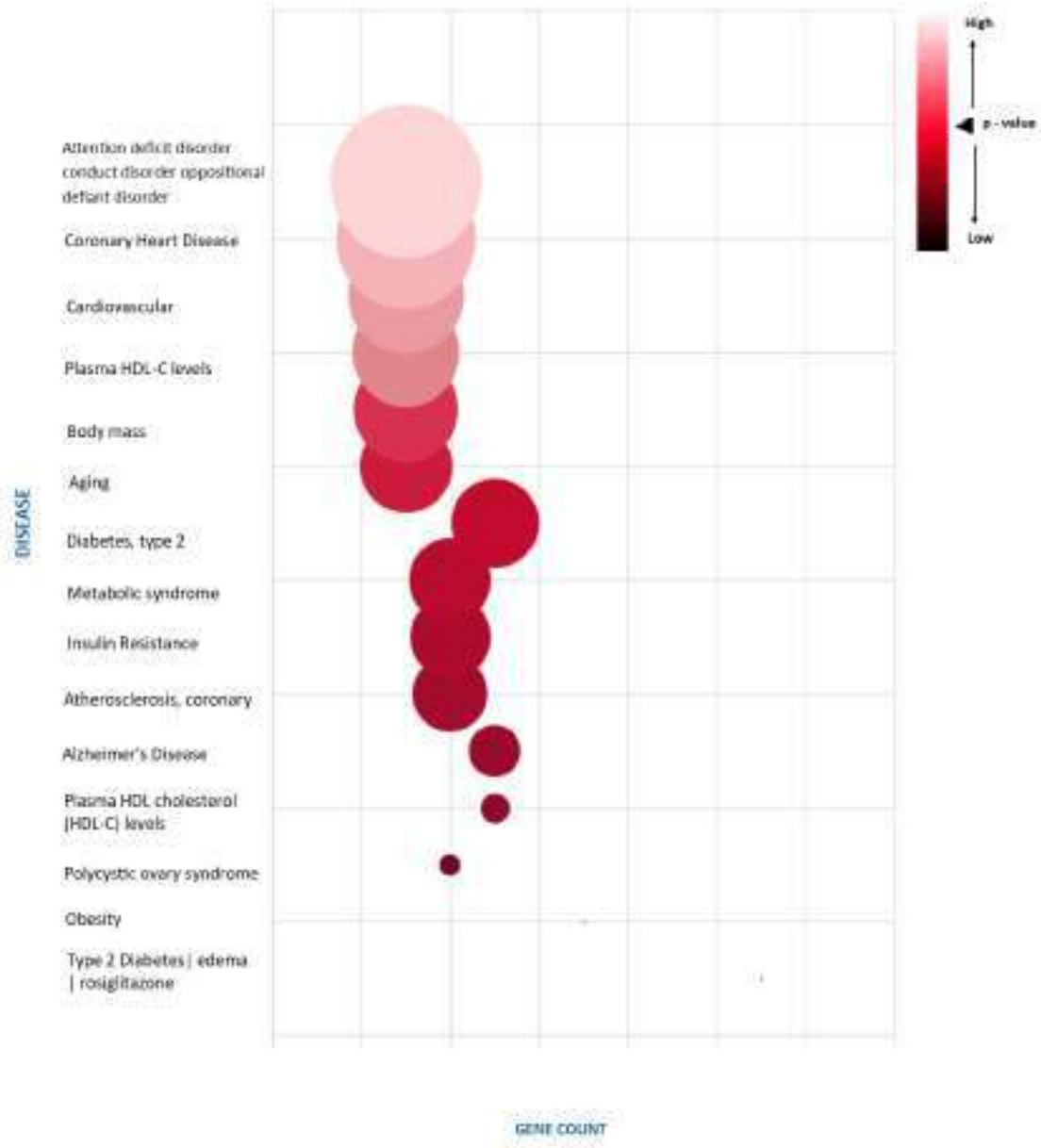

Fig 4.52. Functional Annotation (Disease). The top 15 results (with p-value < 0.05) of Functional annotation of target genes show their association with diseases such as plasma HDL cholesterol (Hypercholesterolemia), Type 2 diabetes, edema, Polycystic ovary syndrome, Obesity etc. Decreasing size of the bubbles as well as the color gradiation from light pink to deep red indicates decrease in the the p-value

resistance in PCOS patients (Khorshidi *et al.,* 2018). Furthermore, oxidative stress and an increase in inflammatory cytokines have been linked to the development of PCOS (Machado *et al.,* 2020).

4.8.2.4. Proteins related with PPAR signaling pathway and their role in several diseases

A total of 6 proteins (FABP2, FABP3, FABP4, FABP5, PPARA, PPARD), among the, studied 12 proteins were

found to be connected with PPAR (peroxisome proliferator-activated receptors) signaling pathway (Supplementary Material 13). The nuclear receptor superfamily includes the peroxisome proliferator-activated receptors (PPARs) which are activated by fatty acids and their derivatives. Invertebrates, PPAR has three subtypes namely PPAR alpha (PPARA), beta/ delta (PPARB/PPARD), and gamma (PPARG) with distinct expression patterns. Each of them binds fatty acids and eicosanoids and is encoded by a different gene. PPARalpha regulates gene expression associated with lipid metabolism in the liver and skeletal muscle, which aids in the clearance of circulating or cellular lipids. PPARbeta/delta is implicated in cell growth and lipid oxidation. PPARgamma enhances blood glucose absorption by promoting adipocyte differentiation (https:// www.genome.jp/entry/hsa03320).

The peroxisome proliferator-activated receptors (PPARA, PPARB/D, and PPARG), serve as ligand-inducible transcription factors and play important roles in glucose and lipid metabolism. These are well-known diabetes treatment receptor that not only affects the cardiovascular system but is also found in many human solid tumors. Peroxisome proliferator-activated receptors (PPARs) have a well-established role in a variety of chronic illnesses, including diabetes, cancer, inflammation, and atherosclerosis (Mirza *et al.,* 2019). It has a crucial function in maintaining lipid, glucose, and energy homeostasis (Towfighi and Ovbiagele, 2008). The PPARs are key therapeutic targets for atherosclerosis, inflammation, and hypertension and the careful and advanced design of a partial agonists of PPAR may also increase the therapeutic importance for several diseases like cardiovascular, cancer, and inflammation (Mirza *et al.,* 2019).

PPARA and other members of this family of receptors, such as PPARB/D and PPARG, are found in the brain and other organs and have a role in oxidative stress, energy balance, mitochondrial fatty acid metabolism, and inflammation. In the brain, PPAR- plays a role in the regulation of genes that code for proteins involved in glutamate homeostasis and cholinergic/ dopaminergic transmission. Downregulation of PPARA may reduce anti-oxidative and anti-inflammatory activities, and may be responsible for changes in fatty acid transport, lipid metabolism, and mitochondrial function in Alzheimer's disease (AD) patients' brains. In neurodegenerative and neurodevelopmental illnesses, specific PPAR-activators may help improve

brain cell metabolism and cognitive performance (Wojtowicz *et al.,* 2020).

PPARG is extensively expressed in adipose tissue and has a variety of activities, including adipocyte development, lipid storage, mitochondrial uncoupling protein expression, leptin downregulation, and insulin sensitivity modulation (Towfighi and Ovbiagele, 2008). They play a role in inflammatory processes and are involved in cell cycle regulation and insulin sensitivity development (Lehrke and Lazar, 2005). PPAR-agonists are thought to serve as negative regulators in T cell development, activating inflammatory responses and hence having a role in adaptive immunity (Sun *et al.,* 2017). PPARG regulates glucose homeostasis, lipid metabolism, and is a key therapeutic target for the treatment of type 2 diabetes as well as metabolic syndrome (Mirza *et al.,* 2019). It increases endothelial cell function by reducing inflammation in diabetes and atherosclerosis (Hsueh and Law, 2003). PPARG also plays an important function in cardiovascular disease. PPARG is abundantly expressed in atherosclerotic lesions, and its activation increases inflammatory effects in cardiovascular cells (Hamblin *et al.,* 2009). PPARG, in addition to its role in cardiovascular systems, has a strong affinity for tumor -related issues and plays an important role in cancer cell growth regulation. Many cancer cells have been found to contain it. Potent ligands that activate PPARG activation can inhibit cancer cell proliferation and differentiation. It plays a role in lipid metabolism as well as cancer cell growth. As a result, it could be used as a therapeutic target for cancer treatment (Mirza *et al.,* 2019). Activators of the PPARA and PPARG receptors can lower atherogenic triglycerides, enhance cardio-protective HDL levels, and improve insulin resistance simultaneously. Since PPARA and PPARD have significant fat-burning activity (Kersten *et al.,*2000) dual, selective, or triple agonists for PPAR-A/D/G are being investigated, since they can overcome the negative effects of PPARG agonists (Towfighi and Ovbiagele, 2008).■