Combined linkage and association mapping of putative QTLs controlling black tea quality and drought tolerance traits

Robert. K. Koech[1], [2], Richard Mose[3], Samson M. Kamunya[2] and Zeno Apostolides[1]*

[1]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria 0002, South Africa

[2]Kenya Agriculture and Livestock Research Organization, Tea Research Institute, P.O. Box 820, Kericho 20200, Kenya

[3]James Finlay (Kenya) Limited, P.O. Box 223, Kericho 20200, Kenya.

*Corresponding author: za@up.ac.za

Abstract

The advancements in genotyping have opened new approaches for identification and precise mapping of quantitative trait loci (QTL) in plants, particularly by combining linkage and association mapping (AM) analysis. In this study, a combination of linkage and AM approach was used to precisely identify and authenticate putative QTLs associated with black tea quality traits and %RWC. The population structure analysis clustered two parents and their respective 261 F_1 progenies from the two reciprocal crosses into two clusters with 141 tea accessions in cluster one and 122 tea accessions in cluster two. The two clusters were of mixed origin with tea accessions in population TRFK St. 504 clustering together with tea accessions in population TRFK St. 524. A total of 70 putative QTLs linked to black tea quality traits and percent relative water contents (%RWC) were detected in interval mapping (IM) method. Cofactors obtained in IM were used in multiple QTL model (MQM) mapping where 40 putative QTLs were detected. The phenotypic variance for each QTL ranged from 2.8-23.3% in IM and 4.1-23% in MQM mapping. Using Q-model and Q+K-model in AM, a total of 49 DArTseq markers were

associated with 16 phenotypic traits. Significant marker-trait association results obtained using AM were similar to those obtained in IM, and MQM mapping with six more QTLs detected in AM. The functional annotations of six additional putative QTLs detected in AM were involved in purine/thiamine biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, carbon fixation and abiotic stress. The application of a combination of linkage and AM approach appears to have great potential and could reduce the cost and improve the efficacy and accuracy of selection of desirable traits in tea breeding.

Introduction

Linkage mapping has been a key tool for identifying the genetic basis of quantitative traits in plants. However, linkage mapping is usually limited by low polymorphism or small population size, which is required for suitable crosses. Also, the mapping resolution is limited by few recombination events, which are considered to estimate the genetic distances between marker loci causative genomic regions for QTLs. AM also known as LD mapping has extensively been used to circumvent the limitations posed by linkage mapping in plants in the last few years (1). AM which exploits broader genetic diversity offers more advantages over linkage mapping in terms of mapping resolution, allele number, saves time in establishing a marker-trait association and its application in a breeding program (2). Although linkage mapping and AM each offer more advantages over the other, they are often applied in conjunction to validate the QTL identified, thus reducing spurious associations.

The integrated approach of linkage-AM has been used in other plants such as *Arabidopsis* (3), wheat (4) and maize (5) to dissect quantitatively inherited traits. In tree species, an integrated method of linkage mapping and AM to decipher the nature of genetic architecture of potential QTLs for growth traits has been reported in poplar hybrids (6) and maritime pine (7). In grapes, (8) combined linkage mapping and AM to study the genetic patterns of anthocyanin content, a

49 determinant of berry colour, in grapes. In a study to investigate genetic relationship between
50 tea caffeine synthase gene (TCS1) and the caffeine content in tea plant and its related species,
51 AM was used (9). In this current study, we combined the high QTL detection power of the
52 linkage mapping with the high-resolution power of AM. The linkage-AM approach will not
53 only accelerate the pace of QTL mapping in tea breeding improvement but will also precisely
54 identify reliable QTLs linked to black tea quality and drought tolerance in tea. The objective
55 of this study was to combine linkage mapping and AM to precisely identify and authenticate
56 putative QTLs linked to caffeine content, catechins fractions, theaflavins fractions, tea tasters'
57 scores, and %RWC.

58 # Material and methods

59 ## Population type

60 Two pseudo-testcross populations used in this current study consisted of 109 F_1 progeny from
61 TRFK St. 504 (TRFK 303/577 x GW Ejulu) and 152 F_1 progeny from TRFK St. 524 (GW
62 Ejulu x TRFK 303/577) as described previously (10).

63 ## Phenotypic data

64 A total of 16 phenotypic traits were used as described previously (10).

65 ## Population structure

66 The population structure was inferred from the 1,421 DArT markers data previously used to
67 construct genetic linkage map using the STRUCTURE software version 2.3.4 (11). Twenty
68 independent runs were carried out using the following parameters; number of populations (K)
69 from 1 to 10, burn-in time and Markov Chain Monte Carlo (MCMC) replication number were
70 both set to 100,000 for model of admixture with correlated allele frequencies. The natural
71 logarithms of the probability data (LnP(K)) and the *ad hoc* delta K statistical were calculated

using STRUCTURE Harvester (12), and the optimal K according to the delta K value was then selected (13). Finally, the population structure matrix (Q) was obtained by integrating 20 independent replicate runs and applying CLUMPP software (14). STRUCTURE bar plots were plotted using STRUCTURE Plot v 2.0 (15).

Linkage mapping, QTL analyses and allelic effects

Six thousand five hundred and eighty-eight DArTseq markers derived from *C. sinensis* genomic DNA of the two parents and 261 F_1 progeny were subjected to linkage mapping analyses using JoinMap 4.0 (16) and QTL mapping analyses using MapQTL 6.0 software (17) as described previously (Koech *et al.,* 2018).

The allelic effects were estimated as $A_f=[(\mu ac+\mu ad)-(\mu bc+\mu bd)]/4$ for female additivity, $A_m=[(\mu ac+\mu bc)-(\mu ad+\mu bd)]/4$ for male additivity and $D=[(\mu ac+\mu bd)-(\mu ad+\mu bc)]/4$ for dominance where μac, μad, μbc and μbd are estimated phenotypic means associated to each of the four possible genotypic classes ac, bc, ad and bd, derived from ab×cd cross (18).

Association mapping

Using the Q and K matrices data files from STRUCTURE software as a covariates, 1,421 DArT markers of the 261 F_1 progeny were tested for association with each phenotype using a general linear model (GLM), mixed linear model (MLM) that included kinship, phylogenetic tree in TASSEL (trait analysis by association, evolution and linkage) Version 5.2.43 software (19). Principal component analysis (PCA) was performed using JMP Pro 14 to visualise the dispersion of the association panel in a graph. The Q and kinship (K) matrices were used to correct the effects of population substructure in the association panel which can cause false-positive associations. The p-value and R^2 were used to determine whether a QTL is associated with the marker and the magnitude of QTL effects, respectively.

Results

Frequency distribution

The frequency distribution results for 16 phenotypic traits were described previously (10).

Population structure

A total of 261 tea accessions, 109 for population TRFK St. 504 and 152 for population TRFK St.524 were combined to construct a panel to study their genetic structure. The two tea populations, TRFK St. 504 and TRFK St. 524 are a reciprocal cross of two parents TRFK 303/577 and GW Ejulu. A population structure analysis showed that the value of Evanno's ΔK presented a sharp spike at K = 2, which suggested that this population panel was clustered into two groups (Fig 1). The two groups or clusters are represented by two clades with parental clone TRFK 303/577 clustering with 141 progenies in cluster one (blue) and parental clone GW Ejulu (not shown) clustering with 122 progenies in cluster two (black), respectively (Fig 2). The two clusters were accession with mixed origin with tea accessions in TRFK St. 504 population clustering together with tea accessions in TRFK St. 524 population. Furthermore, the results from PCA (Fig 3) and a phylogenetic tree-based on Nei's genetic distance was in agreement with the structure analysis results (Fig 4). The population structure of the two tea populations was examined with the neighbour-joining algorithm using Euclidian distance on the DArTseq marker band intensities.

Fig 1. Delta K values plotted from 1 to 10 for 261 tea accessions panel.

Fig 2. Population structure of 261 tea accessions panel (K = 2)

Fig 3. Principal component analysis of 261 tea accessions. represents cluster 1, represents cluster 2

Fig 4. Phylogenetic tree-based on Nei's genetic distance. Blue clade represents cluster 1, black clade represents cluster 2.

Linkage mapping and association mapping

Interval mapping and allelic effects

The genetic linkage map was constructed using a total of 1,421 DArT markers as described previously (10). A total of 70 putative QTLs were identified for the 16 phenotypic traits measured using the interval mapping (Table 1). Of all the putative QTLs identified, most QTLs were found in almost all linkage groups, except LG03, LG05 and LG11. For black tea quality traits, 67 putative QTLs were detected and were located on almost all linkage groups, except on LG03, LG05 and LG11. The remaining three putative QTLs which are associated with drought tolerance were identified in LG02, LG06 and LG09. Eight putative QTLs were identified on LG01, seven in LG02 and LG12, 10 in LG04, five in LG06 and LG15, three in LG07, LG08, LG09, two in LG10, 15 in LG13, four in LG15, respectively. The highest number of putative QTLs were associated with catechin (13), EGC (12), tea tasters' score (12), caffeine (10) and EGC (9), respectively. The phenotypic variance for all the identified putative QTLs ranged from 2.8% for EGC to 23.3% for qECG, respectively. The phenotypic variance for %RWC ranged from 5.7 to 7.3%, while for the tea tasters' score, it ranged from 5.8 to 9.1%. Half of the allelic effects identified to be associated with black tea quality and drought tolerance traits had a positive additive effect. There were eight allelic effects associated with female additivity, and 18 were associated with male additive effects, respectively. The highest positive female and male additive effects were qEGC and qECG, respectively. Although, some putative QTLs also exhibited high negative allelic effects, which were from either female or male additive effects or both, such as qEGCG, qCAT, qEC and qCaffeine (Table 1). All the putative QTLs identified in IM were used as cofactors in MQM mapping.

141 Table 1. Linkage mapping results for putative QTLs in black tea quality and drought tolerance traits in St. 504 and St. 524 using interval
142 mapping in MapQTL 6.0

Nr	[a]Trait	Locus	LG	Position (cM)	LOD	PEV	Af	Am	D	Allelic effect
1	AST	5115441_E-26	1	96.6	4.5	9.1	0.224	-0.061	-0.015	Af
2	AST	5135087	1	97.3	4.5	9.1	0.234	-0.057	-0.014	Af
3	BRK	5115441_E-26	1	96.6	3.6	7.3	0.196	-0.082	-0.020	Af
4	BRK	5135087	1	97.3	3.5	7.2	0.204	-0.078	-0.019	Af
5	ECG	5128890	1	96.4	6.8	11.7	0.326	0.046	0.011	Af
6	ECG	5135087	1	97.3	6.6	11.3	0.347	0.053	0.013	Af
7	ECG	5115441_E-26	1	96.6	6.5	11.1	0.331	0.051	0.013	Af
8	EGC	5133866	1	87.1	3.3	5.8	0.109	0.218	0.054	Am
9	CAFF	5064585	2	50.9	3.3	5.8	5.849	-9.702	-2.425	Af; Am
10	CAT	5135436	2	0.0	3.6	6.1	-0.838	-11.195	-2.799	Am
11	EC	5072338	2	2.2	4.3	7.2	-14.648	-12.555	-3.139	Af; Am
12	EGC	5124128	2	7.7	3.3	5.6	-36.447	1.544	0.386	Af
13	RWC	5136794	2	60.9	4.3	7.3	0.482	1.130	0.283	Am
14	TF1	5084595	2	4.5	3.2	5.7	-0.952	8.158	2.039	Am
15	TF2	5084595	2	4.5	4.1	7.3	2.290	9.140	2.285	Am
16	AR	100158044|F|0	4	70.1	2.9	6	-0.268	0.090	0.023	Af
17	CAFF	5112599	4	68.6	3.8	6.7	2.413	11.852	2.963	Am
18	CAT	5063001	4	30.3	5.3	4.1	-17.604	-53.467	-13.367	Am
19	EC	5123475	4	27.2	14.0	21.9	3.485	44.620	11.155	Am
20	EC	5123257	4	26.7	13.4	21.1	1.703	43.813	10.953	Am
21	EC	5134490	4	26.4	12.4	19.6	0.260	41.569	7.005	Am
22	ECG	5087113	4	17.7	3.8	5.8	-2891.857	3715.230	928.808	Af; Am
23	EGC	5123475	4	27.2	3.7	2.8	7071.043	-6846.394	-1711.598	Af; Am
24	EGCG	5087017	4	37.2	4.9	6	-320.975	-200.380	-50.095	Af; Am
25	ECG	5098382	6	57.0	5.2	8.7	35.650	-32.760	-8.190	Af; Am
26	EGC	5073424	6	66.4	3.9	6.7	-13.905	14.749	3.687	Af; Am
27	EGC	5124993	6	66.6	3.9	6.6	-12.321	15.635	3.909	Af; Am
28	RWC	5082606	6	66.2	3.3	5.7	-2.094	-0.814	-0.203	Af
29	TF1	5136045	6	69.6	4.3	7.7	2.795	3.782	0.946	Af; Am
30	CAFF	5064391	7	48.1	3.4	6	-12.264	-3.397	-0.849	Af
31	CAFF	5056614	7	48.3	3.3	5.9	-12.029	-3.985	-0.996	Af
32	CL	5132432	7	72.7	3.2	6.5	-0.138	-0.118	-0.029	Af; Am
33	CAFF	5134558	8	18.8	4.3	7.5	17.375	-12.332	-3.083	Af; Am
34	CAFF	5111497	8	18.8	4.1	7.1	11.542	-12.927	-3.232	Af; Am
35	CAT	5130194	8	12.5	3.9	6.6	-29.556	-9.720	-2.430	Af

36	AST	5123950	9	87.6	3.2	6.6	-0.036	0.178	0.045	Am
37	BRK	5123950	9	87.6	3.6	7.4	-0.012	0.190	0.047	Am
38	RWC	5130531	9	6.7	4.1	7	1.933	2.876	0.719	Am
39	AR	5075627	10	26.9	3.4	6.9	-0.064	0.110	0.027	Am
40	ECG	5136108	10	20.6	3.3	5.7	4.297	34.819	8.705	Am
41	CAT	5123751	12	43.0	4.2	6.5	-36.478	-3.162	-0.791	Af
42	CAT	5127224	12	42.9	4.2	6.5	-36.467	-3.045	-0.761	Af
43	ECG	5136790	12	50.4	15.0	23.3	-22.365	60.821	15.205	Am
44	ECG	5135536	12	50.4	15.0	23.2	-22.316	60.808	15.202	Am
45	EGC	5123751	12	43.0	5.0	8.5	50.465	-30.542	-7.635	Af; Am
46	EGC	5127224	12	42.9	5.0	8.4	50.241	-30.563	-7.641	Af; Am
47	EGCG	5104630	12	48.2	4.2	7.2	18.115	-52.080	-13.020	Am
48	BRT	5088162	13	29.8	3.0	6.1	0.006	-0.142	-0.035	Am
49	BRT	5135810	13	26.0	2.8	5.8	0.007	-0.137	-0.034	Am
50	CAFF	5088162	13	29.8	3.4	5.4	7.183	-29.136	-7.284	Am
51	CAFF	5080631	13	35.6	2.9	3.1	-14.651	23.420	5.855	Af; Am
52	CAT	5122819	13	10.2	4.3	7.3	-27.341	-14.521	-3.630	Af; Am
53	CAT	5111268	13	58.1	4.6	7.2	14.200	17.804	4.451	Af; Am
54	CAT	5129729	13	9.8	4.2	7.2	-26.752	-14.467	-3.617	Af; Am
55	CAT	5115793	13	58.7	4.3	6.8	12.635	17.248	4.312	Af; Am
56	CAT	5136782	13	58.5	4.2	6.6	12.029	17.237	4.309	Af; Am
57	CAT	5103784	13	50.6	3.8	6.1	4.188	21.966	5.491	Am
58	CAT	5122899	13	50.5	3.8	6	3.656	21.921	5.480	Am
59	ECG	5123761	13	60.7	3.3	5.8	-0.469	-0.112	-0.028	Af
60	ECG	5072523	13	60.3	3.0	5.2	-0.261	-0.231	-0.058	Af; Am
61	ECG	5114985	13	61.6	2.8	5	-0.439	-0.088	-0.022	Af
62	EGC	5136623	13	50.6	3.7	6.4	8.946	-29.764	-7.441	Am
63	BRT	5125626	14	71.1	3.3	6.8	0.000	0.143	0.036	Am
64	CAFF	5123053	14	5.5	4.0	7.1	-7.622	-11.572	-2.893	Af; Am
65	CAFF	5054639	14	6.4	3.9	6.9	-7.260	-11.533	-2.883	Af; Am
66	CAT	5132370	14	60.7	6.5	10.8	-19.945	35.937	8.984	Af; Am
67	EGC	5132370	14	60.7	3.7	6.4	46.492	-7.514	-1.878	Af
68	EC	5085963	15	25.4	4.4	7.5	11.674	-24.472	-6.118	Am
69	EC	5099958_E-25	15	25.1	4.4	7.4	11.686	-24.101	-6.025	Am
70	ECG	5111164	15	75.1	4.2	7.2	22.426	34.202	8.550	Af; Am
71	EGCG	5114089	15	32.1	4.0	6.8	-1.178	56.567	14.142	Am

[a] Putative QTL identified in Interval Mapping using MapQTL 6.0. LOD- Logarithm of odds for putative QTL in IM (LOD>2); PEV- percentage phenotypic variation explained by markers in IM; Af- Female additivity; Am- Male additivity; D- Dominance

Multiple QTL model mapping and allelic effects

A total of 40 putative QTLs (LOD > 3.0) were identified in 16 phenotypic traits using MQM mapping method from the selected cofactors in IM. Similarly, no putative QTLs in LG03, LG05 and LG11, respectively were associated with any of the phenotypic traits. Most of the putative QTLs previously identified in IM were also identified MQM mapping. Of these 40 putative QTLs identified, 37 putative QTLs that were associated with black tea quality traits while three were associated with qRWC (Table 2). The three putative QTLs associated with qRWC were similar to those identified in IM. Four putative QTLs were identified in LG01 and LG12, seven in LG02, five in LG04, three in LG06, LG09, LG13, LG14 and LG15, two in LG07 and LG08 and one in LG01, respectively. A similar trend was observed in the number of putative QTLs for each trait as in IM with a high putative QTL number for catechin, ECG, EGC and caffeine, respectively. The phenotypic variance for all the identified putative QTL ranged from 4.1% for qCAT to 23% for qECG, respectively. The phenotypic variance for qRWC was similar to IM results, and it ranged from 5.7 to 7.3%, while for the tea tasters' score, it ranged from 6.0 to 9.1%. For allelic effects, similar results to IM with qEGC and qECG showing high positive female additive and male additive effects, respectively, this trend was also observed in MQM mapping. The putative QTLs identified using both IM, and MQM mapping were almost similar although the efficiency and the accuracy of detecting QTL is achieved by employing MQM instead of the single QTL model used in IM.

Table 2. Linkage mapping for putative QTLs in black tea quality and drought tolerance traits in St. 504 and St. 524 using multiple QTL model mapping in MapQTL 6.0

Nr	[b]Trait	Locus	LG	Position (cM)	LOD	PEV	Af	Am	D	Allelic effect
1	AST	5135087	1	97.3	4.5	9.1	0.234	-0.057	-0.014	Af
2	BRK	5115441_E-26	1	96.6	3.6	7.3	0.196	-0.082	-0.020	Af
3	ECG	5128890	1	96.4	6.8	11.7	0.326	0.046	0.011	Af
4	EGC	5133866	1	87.1	3.3	5.8	0.109	0.218	0.054	Am
5	CAFF	5064585	2	50.9	3.3	5.8	5.849	-9.702	-2.425	Af; Am
6	CAT	5135436	2	0.0	3.6	6.1	-0.838	-11.195	-2.799	Am
7	EC	5072338	2	2.2	4.3	7.2	-14.648	-12.555	-3.139	Af; Am
8	EGC	5124128	2	7.7	3.3	5.6	-36.447	1.544	0.386	Af
9	RWC	5136794	2	60.9	4.3	7.3	0.482	1.130	0.283	Am
10	TF1	5084595	2	4.5	3.2	5.7	-0.952	8.158	2.039	Am
11	TF2	5084595	2	4.5	4.1	7.3	2.290	9.140	2.285	Am
12	AR	100158044|F|0	4	70.1	2.9	6	-0.268	0.090	0.023	Af
13	CAFF	5112599	4	68.6	3.8	6.7	2.413	11.852	2.963	Am
14	CAT	5063001	4	30.3	5.3	4.1	-17.604	-53.467	-13.367	Am
15	EC	5123475	4	27.2	14.0	21.9	3.485	44.620	11.155	Am
16	EGCG	5087017	4	37.2	4.9	6	-320.975	-200.380	-50.095	Af; Am
17	ECG	5098382	6	57.0	5.2	8.7	35.650	-32.760	-8.190	Af; Am
18	EGC	5073424	6	66.4	3.9	6.7	-13.905	14.749	3.687	Af; Am
19	RWC	5082606	6	66.2	3.3	5.7	-2.094	-0.814	-0.203	Af
20	CAFF	5064391	7	48.1	3.4	6	-12.264	-3.397	-0.849	Af
21	CL	5132432	7	72.7	3.2	6.5	-0.138	-0.118	-0.029	Af; Am
22	CAFF	5134558	8	18.8	4.3	7.5	17.375	-12.332	-3.083	Af; Am
23	CAT	5130194	8	12.5	3.9	6.6	-29.556	-9.720	-2.430	Af
24	AST	5123950	9	87.6	3.2	6.6	-0.036	0.178	0.045	Am
25	BRK	5123950	9	87.6	3.6	7.4	-0.012	0.190	0.047	Am
26	RWC	5130531	9	6.7	4.1	7	1.933	2.876	0.719	Am
27	ECG	5136108	10	20.6	3.3	5.7	4.297	34.819	8.705	Am
28	CAT	5123751	12	43.0	4.2	6.5	-36.478	-3.162	-0.791	Af
29	ECG	5136790	12	50.4	15.0	23.3	-22.365	60.821	15.205	Am
30	EGC	5123751	12	43.0	5.0	8.5	50.465	-30.542	-7.635	Af; Am
31	EGCG	5104630	12	48.2	4.2	7.2	18.115	-52.080	-13.020	Am
32	CAFF	5088162	13	29.8	3.4	5.4	7.183	-29.136	-7.284	Am

33	CAT	5111268	13	58.1	4.6	7.2	14.200	17.804	4.451	Af; Am
34	EGC	5136623	13	50.6	3.7	6.4	8.946	-29.764	-7.441	Am
35	BRT	5125626	14	71.1	3.3	6.8	0.000	0.143	0.036	Am
36	CAFF	5123053	14	5.5	4.0	7.1	-7.622	-11.572	-2.893	Af; m
37	CAT	5132370	14	60.7	6.5	10.8	-19.945	35.937	8.984	Af; Am
38	EC	5085963	15	25.4	4.4	7.5	11.674	-24.472	-6.118	Am
39	ECG	5111164	15	75.1	4.2	7.2	22.426	34.202	8.550	Af; Am
40	EGCG	5114089	15	32.1	4.0	6.8	-1.178	56.567	14.142	Am

[b] Putative QTL identified in Multiple QTL Model Mapping using MapQTL 6.0

LOD- Logarithm of odds for putative QTL in MQM (LOD>3); PEV- percentage phenotypic variation explained by markers in MQM; Af- Female additivity; Am- Male additivity; D- Dominance

CAFF- Caffeine; CAT- Catechin; EC- Epicatechin; ECG- Epicatechin gallate; EGC- Epigallocatechin gallate; EGCG- Epigallocatechin gallate; TF1- Theaflavin; TF2- Theaflavin-3-gallate; TF3- Theaflavin-3'-gallate; TF4- Theaflavin-3,3'-digallate; CL- Colour; BRT- Brightness; AST- Astringency; BRK- Briskness; AR- Aroma; RWC- percent relative water content

Association mapping

Four models GLM, GLM (Q), mixed linear model with kinship matrix (MLM) (K) and mixed linear model with kinship matix and population structure matrix (MLM) (Q+K) in TASSEL software v5.2.43 were used to determine AM and the effects of population structure on the AM to reduce the inflation of false-positive associations. The p-values were plotted cumulatively for each model, and the distribution examined. However, no further analysis was done using GLM model since marker-trait association results were characterised by excess small p-values, which indicated an abundance of spurious associations (S1 Fig). The association analysis using the Q-model and Q+K-model detected a total of 49 DArT markers associated with 15 different black tea quality traits and one drought-tolerant trait. The GLM (Q) model showed associations between DArT markers and traits ($p < 0.01$) and was confirmed using the MLM (Q+K) model (Table 3, S2-S4 Figs). The MLM Q+K-model (MLM with Q-matrix and K-matrix as a correction for population structure) showed a good fit for the p-values ($p < 0.01$), as compared to GLM (Q) models, which were characterised by a few associations with excess of small p-values, which indicates abundance of spurious associations (Table 3, S2 and S4 Figs). Also, the MLM Q+K-model also showed better small p-values than the K-model (MLM with K-matrix as a correction for population structure) (S3 and S4 Figs). The GLM (Q) model may not have accounted for the heterogeneity of the genetic background in some tea cultivars under study, which may have resulted in false-positive associations. Ideally, the distribution of p-values should follow a uniform distribution with less deviation from the expected p-values. However, the two models were used to compare the effect of population matrix and kinship matrix on GLM and MLM, respectively, in associating mapping. Therefore, taking into account the performance of the two different models, the results from the MLM (Q+K) model appeared to have controlled better population structure and kinship relationships than GLM (Q) model. While it might be tempting to consider p-values that remain extreme after MLM (Q+K) model

correction as true associations, we need to keep in mind that minor allele frequency also influences p-values.

A total of 49 DArT markers were associated with 16 different phenotypic traits in six out of 15 linkage groups (Table 3). Since, all 1,421 DArT markers in the genotypic dataset had a known map location, the loci associated with particular traits could be allocated to a linkage group. The numbers of marker loci associated with traits together with their location and positions in different linkage groups at a test level of $p< 0.01$ in the two test models are shown in Table 3. Choosing a significance level of $p< 0.01$ which involves multiple testing corrections in both GLM (Q) and MLM (Q+K) models, 32 DArT marker loci were detected for all individual catechins (C, EC, ECG, EGC, EGCG), scattered over five linkage groups namely LG01, LG02, LG04, LG12 and LG15 (Table 3). For caffeine trait, three DArT markers were only located in LG02 while 11 DArT markers for all individual theaflavins traits (TF1, TF2, TF3, TF4) were detected in LG04 and LG15 (Table 3). The trait for tea liquor colour which is an indicator for black tea quality in terms of tea tasters' score appeared to be associated with only one DArT marker locus at the significance level of $p< 0.01$ in LG07 (Table 3). Drought tolerance trait (%RWC) was also associated with only one DArT marker locus detected in LG02 (Table 3). The majority of the DArT markers loci associated with the traits were found in LG04, LG12, LG15, LG02, and LG01. The three DArT markers associated with catechin trait in LG04, namely, 5136945, 5064764 and 5134490 had the highest proportion of the percent explained phenotypic variation of 41%, 38%, and 25%, respectively. The DArT markers that also showed high proportion of percent explained phenotypic variation were also observed in LG04 for qEGCG, qEGC, and qTF1 traits, respectively. In this study, a pleiotropic locus which is associated with a single locus affecting two or more distinct phenotypic traits was observed. Several pleiotropic loci were identified including DArT markers 5136945, 5112438, 5136058, 5123257, 5106022 and 5064764 which were associated with qC, qEC, qEGC, qTF1, qTF2 and

225 qTF4 phenotypic traits (Table 3). All the pleiotropic loci detected and associated with the
226 respective phenotypic traits were located in LG04.

227

228 Table 3. Association mapping results for black tea quality and drought tolerance traits in St. 504 and St.524 using the GLM (Q) and MLM (Q+K) method,
229 respectively ($p < 0.01$)

Nr	[c]Trait	Marker	LG	Position (cM)	p-Value	Marker R^2 GLM (Q)	p-Value	Marker R^2 MLM (Q+K)
1	ECG	5128890	1	96.4	1.18E-03	0.042	5.31E-03	0.031
2	EGC	5086476	1	87.5	8.20E-03	0.031	8.54E-04	0.043
3	EGC	5122985	1	87.5	1.47E-04	0.056	2.41E-03	0.038
4	EGC	5133866	1	87.1	8.94E-03	0.030	6.62E-04	0.046
5	CAFF	5134416	2	44.4	2.46E-04	0.057	5.58E-04	0.050
6	CAFF	5136551	2	43.3	9.60E-04	0.048	7.51E-03	0.030
7	CAFF	5137282	2	50.7	2.27E-03	0.042	9.72E-03	0.026
8	EC	5132500	2	7.9	2.91E-03	0.036	1.37E-04	0.052
9	RWC	5136794	2	60.9	8.58E-03	0.032	3.02E-03	0.033
10	CAT	5063001	4	30.3	2.34E-09	0.098	6.67E-08	0.109
11	CAT	5064764	4	28.8	6.42E-37	0.326	7.80E-22	0.379
12	CAT	5136945	4	28.1	6.39E-40	0.345	1.81E-21	0.406
13	EC	5112438	4	27.4	4.90E-07	0.087	1.50E-05	0.067
14	EC	5123257	4	26.7	3.89E-08	0.101	7.16E-07	0.089
15	EC	5134490	4	26.4	5.15E-19	0.228	6.21E-15	0.246
16	EC	5136058	4	27.6	1.00E-06	0.083	5.31E-06	0.073
17	EC	5136410	4	29.9	2.39E-05	0.065	1.77E-05	0.065
18	ECG	5087113	4	17.7	3.74E-03	0.035	2.92E-04	0.048
19	EGC	5112438	4	27.4	1.50E-11	0.148	1.95E-08	0.134
20	EGC	5123257	4	26.7	9.59E-13	0.162	6.95E-10	0.167
21	EGC	5136058	4	27.6	3.43E-11	0.143	3.60E-09	0.146
22	EGCG	5074553	4	39.5	3.26E-14	0.210	1.86E-12	0.221
23	EGCG	5123463	4	28.9	3.38E-17	0.250	1.09E-13	0.266
24	EGCG	5134853	4	37.6	7.03E-14	0.205	2.00E-09	0.170
25	EGCG	5136554	4	38.7	1.46E-16	0.242	2.08E-11	0.222
26	EGCG	100011901|F|0	4	38.7	1.24E-04	0.066	3.13E-03	0.039
27	TF1	5064764	4	28.8	2.35E-10	0.149	5.78E-06	0.088
28	TF1	5106352	4	26.0	1.95E-04	0.061	8.45E-03	0.032
29	TF1	5123463	4	28.9	1.33E-10	0.153	2.66E-06	0.097
30	TF1	5136945	4	28.1	1.69E-11	0.166	8.20E-07	0.107
31	TF2	5063001	4	30.3	9.76E-04	0.050	4.22E-03	0.035
32	TF2	5064764	4	28.8	1.85E-08	0.122	7.36E-05	0.063

33	TF2	5106022	4	29.1	2.99E-09	0.134	3.06E-04	0.052
34	TF4	5064764	4	28.8	8.64E-06	0.082	1.33E-03	0.042
35	TF4	5106022	4	29.1	1.38E-06	0.094	1.57E-03	0.040
36	TF4	5136945	4	28.1	6.26E-07	0.099	1.59E-04	0.060
37	CL	5132432	7	72.7	1.31E-03	0.047	7.30E-04	0.049
38	ECG	5059563	12	49.0	8.91E-09	0.111	2.29E-05	0.068
39	ECG	5136077	12	42.8	1.91E-07	0.094	6.34E-03	0.027
40	ECG	5136790	12	50.4	7.87E-11	0.137	1.58E-04	0.051
41	EGC	5135536	12	50.4	2.25E-04	0.053	2.82E-06	0.086
42	EGCG	5087581	12	48.3	1.51E-04	0.065	1.75E-03	0.040
43	EGCG	5088456	12	47.9	7.70E-04	0.053	2.46E-04	0.060
44	EGCG	5133629	12	47.7	7.17E-04	0.054	1.95E-03	0.041
45	EC	5085963	15	25.4	3.52E-04	0.049	9.09E-03	0.029
46	EC	5132852	15	24.8	1.09E-04	0.056	2.50E-03	0.036
47	EC	5099958_E-25	15	25.1	1.00E-04	0.056	6.37E-03	0.031
48	TF3	5083854	15	15.2	3.52E-03	0.043	2.77E-03	0.035
49	TF3	5136247	15	16.3	1.57E-03	0.049	5.40E-04	0.046

[c] Putative QTL identified in General Linear Model and Mixed Linear Model methods using TASSEL software v5.2.43

LG- Linkage group

Marker R^2- percentage phenotypic variation explained by markers in both GLM (Q) and MLM (Q+K) methods

Q- Population structure matrix

K- Kinship matrix

CAFF- Caffeine; CAT- Catechin; EC- Epicatechin; ECG- Epicatechin gallate; EGC- Epigallocatechin gallate; EGCG- Epigallocatechin gallate; TF1- Theaflavin; TF2- Theaflavin-3-gallate; TF3- Theaflavin-3'-gallate; TF4- Theaflavin-3,3'-digallate; CL- Colour; BRT- Brightness; AST- Astringency; BRK- Briskness; AR- Aroma; RWC- percent relative water content

241 ## A comparison of linkage mapping and association mapping

242 The putative QTLs detected using IM, and MQM mapping in MapQTL 6.0 were compared
243 with marker-trait association results obtained using GLM (Q) and MLM (Q+K) models in AM.
244 The significant marker-trait association results obtained using AM were similar to with the
245 results obtained using MapQTL 6.0 (Tables 4 and 5). The IM analysis produced 14 significant
246 QTLs (LOD > 3.0) that were also found to be associated with the phenotypic traits in AM
247 analysis (Table 4). From the 14 QTLs, a total of 11 markers were associated with four
248 individual catechins (qCAT, qEC, qECG, and qEGC) and one marker each was associated with
249 qTF2, qCL, and qRWC, respectively. On the other hand, the MQM mapping analysis produced
250 eight significant QTLs (LOD > 3.0) which were also found to be significantly associated with
251 the phenotypic traits in AM analysis (Table 5). Of these, five markers were associated with
252 four individual catechins (qCAT, qEC, qECG, and qEGC) and one marker each was associated
253 with qTF2, qCL, and qRWC, respectively. Furthermore, the four mapping models were
254 compared with each other to find markers that could associate with phenotypic traits in all four
255 mapping models (Table 6). A total of 13 markers were found to be associated with four
256 individual catechins (qCAT, qEC, qECG, and qEGC) and one marker each was associated with
257 qTF2, qCL, and qRWC, respectively. Most of the markers that were associated with different
258 phenotypic traits were only located in LG01, LG02, LG04, LG12 and LG15 with a majority of
259 markers showing both positive and negative male additive effects.

260

Table 4. Association mapping and linkage mapping results for putative QTLs in black tea quality and drought tolerance traits in St.504 and St.524 using GLM (Q), MLM (Q + K) methods ($p < 0.01$) in TASSEL software v5.2.43 and interval mapping in MapQTL 6.0

Nr	[ac]Trait	Marker	LG	Position (cM)	p-Value	Marker R^2 GLM (Q)	p-Value	Marker R^2 MLM (Q+K)
1	ECG	5128890	1	96.4	1.18E-03	0.042	5.31E-03	0.031
2	EGC	5133866	1	87.1	8.94E-03	0.030	6.62E-04	0.046
3	RWC	5136794	2	60.9	8.58E-03	0.032	3.02E-03	0.033
4	CAT	5063001	4	30.3	2.34E-09	0.098	6.67E-08	0.109
5	EC	5123257	4	26.7	3.89E-08	0.101	7.16E-07	0.089
6	EC	5134490	4	26.4	5.15E-19	0.228	6.21E-15	0.246
7	ECG	5087113	4	17.7	3.74E-03	0.035	2.92E-04	0.048
8	EGC	5123257	4	26.7	9.59E-13	0.162	6.95E-10	0.167
9	TF2	5063001	4	30.3	9.76E-04	0.050	4.22E-03	0.035
10	CL	5132432	7	72.7	1.31E-03	0.047	7.30E-04	0.049
11	ECG	5136790	12	50.4	7.87E-11	0.137	1.58E-04	0.051
12	EGC	5135536	12	50.4	2.25E-04	0.053	2.82E-06	0.086
13	EC	5085963	15	25.4	3.52E-04	0.049	9.09E-03	0.029
14	EC	5099958_E-25	15	25.1	1.00E-04	0.056	6.37E-03	0.031

[ac] Putative QTL identified in both Interval Mapping (MapQTL 6.0), General Linear Model and Mixed Linear Model methods (TASSEL software v5.2.43)
Marker R^2- percentage phenotypic variation explained by markers in GLM and MLM

CAT- Catechin; EC- Epicatechin; ECG- Epicatechin gallate; EGC- Epigallocatechin gallate; TF2- Theaflavin-3-gallate; CL- Colour; RWC- Percent relative water content

Table 5. Association mapping and linkage mapping results for putative QTLs in black tea quality and drought tolerance traits in St.504 and St.524 using GLM (Q), MLM (Q + K) methods ($p < 0.01$) in TASSEL software v5.2.43 and multiple QTL model mapping in MapQTL 6.0

Nr	[bc]Trait	Marker	LG	Position (cM)	p-Value	Marker R2 GLM (Q)	p-Value	Marker R2 MLM (Q+K)
1	ECG	5128890	1	96.4	1.18E-03	0.042	5.31E-03	0.031
2	EGC	5133866	1	87.1	8.94E-03	0.030	6.62E-04	0.046
3	RWC	5136794	2	60.9	8.58E-03	0.032	3.02E-03	0.033
4	CAT	5063001	4	30.3	2.34E-09	0.098	6.67E-08	0.109
5	TF2	5063001	4	30.3	9.76E-04	0.050	4.22E-03	0.035
6	CL	5132432	7	72.7	1.31E-03	0.047	7.30E-04	0.049
7	ECG	5136790	12	50.4	7.87E-11	0.137	1.58E-04	0.051
8	EC	5085963	15	25.4	3.52E-04	0.049	9.09E-03	0.029

[bc] Putative QTL identified in both multiple QTL model mapping (MapQTL 6.0), General Linear Model and Mixed Linear Model methods (TASSEL software v5.2.43)

Marker R^2- percentage phenotypic variation explained by markers in GLM and MLM

CAT- Catechin; EC- Epicatechin; ECG- Epicatechin gallate; EGC- Epigallocatechin gallate; TF2- Theaflavin-3-gallate; CL- Colour; RWC- Percent relative water content

Table 6. Association mapping and linkage mapping results for putative QTLs in black tea quality and drought tolerance traits in St.504 and St.524 using GLM (Q), MLM (Q + K) methods ($p < 0.01$) in TASSEL software v5.2.43 and interval mapping and multiple QTL model mapping in MapQTL 6.0

Nr	[abc] Trait	Marker	LG	Position (cM)	p-Value	Marker R2 GLM (Q)	p-Value	Marker R2 MLM (Q+K)	LOD	PEV	Af	Am	D	Allelic effects
1	ECG	5128890	1	96.4	1.18E-03	0.042	5.31E-03	0.031	6.8	11.7	0.326	0.046	0.011	Af
2	EGC	5133866	1	87.1	8.94E-03	0.030	6.62E-04	0.046	3.3	5.8	0.109	0.218	0.054	Am
3	RWC	5136794	2	60.9	8.58E-03	0.032	3.02E-03	0.033	4.3	7.3	0.482	1.13	0.283	Am
4	CAT	5063001	4	30.3	2.34E-09	0.098	6.67E-08	0.109	5.3	4.1	-17.6	-53.45	-13.37	Am
5	EC	5123257	4	26.7	3.89E-08	0.101	7.16E-07	0.089	13.4	21.1	1.703	43.81	10.95	Am
6	ECG	5087113	4	17.7	3.74E-03	0.035	2.92E-04	0.048	3.8	5.8	-2891.86	3715.23	928.81	Af; Am
7	EGC	5123257	4	26.7	9.59E-13	0.162	6.95E-10	0.167	13.4	21.1	1.703	43.81	10.95	Am
8	TF2	5063001	4	30.3	9.76E-04	0.050	4.22E-03	0.035	5.3	4.1	-17.604	-53.47	-13.37	Am
9	CL	5132432	7	72.7	1.31E-03	0.047	7.30E-04	0.049	3.2	6.5	-0.138	-0.118	-0.029	Af; Am
10	ECG	5136790	12	50.4	7.87E-11	0.137	1.58E-04	0.051	15	23.3	-22.37	60.82	15.21	Am
11	EGC	5135536	12	50.4	2.25E-04	0.053	2.82E-06	0.086	15	23.2	-22.32	60.81	15.20	Am
12	EC	5085963	15	25.4	3.52E-04	0.049	9.09E-03	0.029	4.4	7.5	11.67	-24.47	-6.12	Am
13	EC	5099958_E-25	15	25.1	1.00E-04	0.056	6.37E-03	0.031	4.4	7.4	11.69	-24.1	-6.03	Am

[abc] Putative QTL identified using both Interval Mapping and Multiple QTL Model Mapping (MapQTL 6.0) and General Linear Model and Mixed Linear Model method (TASSEL software v5.2.43)

Marker R^2- percentage phenotypic variation explained by markers in GLM and MLM

LG- Linkage group

LOD- Logarithm of odds ratio for putative QTL

PEV- percentage phenotypic variation explained by markers in IM and MQM

Af- Female additivity

Am- Male additivity

D- Dominance

CAT- Catechin; EC- Epicatechin; ECG- Epicatechin gallate; EGC- Epigallocatechin gallate; TF2- Theaflavin-3-gallate; CL- Colour; RWC- Percent relative water content

297 **Functional annotation of putative QTLs in linkage and association mapping**

298 The putative QTLs detected in both IM, and MQM mapping were identified by the BLAST
299 algorithm, searched on the reference tea genome (NCBI) and assigned functions on Blast2GO
300 database. A total of 53 QTLs were detected in both interval and MQM mapping methods, of
301 which 45 proteins were assigned functions. An additional six markers in LG02, LG04, and
302 LG12 were found to be associated with caffeine, EC, ECG and EGC traits in AM (Table 7).
303 The six additional QTLs detected in AM were also functionally annotated using the above
304 databases. The putative candidate genes identified were involved in purine or thiamine
305 biosynthesis, phenylalanine biosynthesis, tyrosine and tryptophan biosynthesis, carbon fixation
306 (photosynthesis) and abiotic stress. The putative QTL, qCaffeine in LG02 was putatively
307 annotated N-(5'phosphoribosyl) anthranilate (PRA) isomerase enzyme. The two other putative
308 QTLs identified for qEC and qEGC located in LG04 were annotated as phosphoribulokinase
309 or uridine kinase family proteins. Two putative QTLs, for qEC in LG04, were annotated as
310 autophagy-related protein 11 and protein tyrosine kinase, respectively.

311

Table 7. Association mapping, linkage mapping and functional annotation protein of putative QTLs in black tea quality traits in St.504 and St.524 based on IM and MLM (Q + K) method ($p < 0.01$) in MapQTL 6.0 and TASSEL software v5.2.43, respectively.

Nr	Trait	Marker	LG	Pos (cM)	p-Value	A	p-Value	B	p-Value	Annotated protein	Function
1	[abc]ECG	5128890	1	96.4	1.2E-03	0.042	5.3E-03	0.031	2.0E-25	['Actin']	Cell signalling (response to cold stress, dehydration)
2	[c]CAFF	5137282	2	50.7	2.3E-03	0.042	9.7E-03	0.026	8.0E-09	["N-(5'phosphoribosyl) anthranilate (PRA) isomerase"]	Involved in the first and intermediate pathway in the purine/thiAmine biosynthesis; It also participates in the phenylalanine, tyrosine and tryptophan biosynthesis pathway
3	[abc]ECG	5087113	4	17.7	3.7E-03	0.035	2.9E-04	0.048	6.0E-22	['impB/mucB/sAmB fAmily']	Involved in UV protection through DNA repair
4	[ac]TF1	5106352	4	26	2.0E-04	0.061	8.5E-03	0.032	2.0E-06	['Thiolase, C-terminal domain']	Benzoic acid biosynthesis (volatile compounds)
5	[c]EC	5112438	4	27.4	4.9E-07	0.087	1.5E-05	0.067	9.0E-21	['Phosphoribulokinase / Uridine kinase fAmily']	Carbon fixation (photosynthesis) which are products for shikimate pathway (phenolics)
6	[c]EGC	5112438	4	27.4	1.5E-11	0.148	2.0E-08	0.134	9.0E-21	['Phosphoribulokinase / Uridine kinase fAmily']	Carbon fixation (photosynthesis) which are productsfor shikimate pathway (phenolics)
7	[ac]EC	5134490	4	26.4	5.2E-19	0.228	6.2E-15	0.246	3.0E-08	['Aminotransferase class I and II']	Phenylalanine, tyrosine and tryptophan biosynthesis
8	[ac]EGCG	5134853	4	37.6	7.0E-14	0.205	2.0E-09	0.17	2.0E-06	['Diacylglycerol kinase catalytic domain']	Water, salt, ROS, cold and freezing stress response
9	[c]EC	5136058	4	27.6	1.0E-06	0.083	5.3E-06	0.073	8.0E-09	['Autophagy-related protein 11']	Abiotic stress response (stomatal closure)
10	[ac]EGC	5136058	4	27.6	3.4E-11	0.143	3.6E-09	0.146	8.0E-09	['Autophagy-related protein 11']	Abiotic stress response (stomatal closure)
11	[c]EC	5136410	4	29.9	2.4E-05	0.065	1.8E-05	0.065	7.0E-06	['Protein tyrosine kinase']	ABA signalling
12	[ac]EGCG	5088456	12	47.9	7.7E-04	0.053	2.5E-04	0.06	4.0E-23	['Protein kinase domain']	Abiotic stress response
13	[c]ECG	5136077	12	42.8	1.9E-07	0.094	6.3E-03	0.027	5.0E-13	CSA026168	-

[abc] Putative QTL identified using both Interval Mapping and Multiple QTL Model Mapping (MapQTL 6.0), General Linear Model and Mixed Linear Model (TASSEL software v5.2.43), respectively.

[c] Putative QTL identified in General Linear Model and Mixed Linear Model (MLM) using TASSEL software v5.2.43; A- Marker R^2 GLM (Q); B- Marker R^2 MLM (Q+K)

Discussion

In this current study, a combination of linkage mapping and AM was used to provide a mutual and precise authentication of QTL, which will enable more reliable results to be obtained. Linkage mapping method has a relatively low genome resolution while AM method is affected by population structure and individual relationships. The combination of linkage mapping and AM has successfully been applied to many plant studies (20-24).

The population structure analysis in discovery mapping panel used in the current study identified two groups or clusters. The dendrogram demonstrated two subspecies of the two parental clones, TRFK 303/577 and GW Ejulu, which are of *C. sinensis* var. *assamica* and *C. sinensis* var. *sinensis* variety, respectively. The neighbour-joining tree for the 261 tea accessions grouped them into two major clusters on the base of Nei's genetic distance. The phylogenetic relationships for the parental cultivars and their progenies clustering into their respective groups in separate clades based on their parental genotypes were consistent with their genetic backgrounds (25). The parental clone TRFK 303/577, which is an open-pollinated progeny of clone TRFK 6/8, is high-yielding, drought-tolerant, medium in black tea quality, caffeine, and individual catechins. On the other hand, parental clone GW Ejulu is a low-yielding, high black tea quality and moderate levels of caffeine, but high in total catechins and individual catechin contents (26). The bulk of tea cultivars used for green and black tea production have been derived through individual selection, hybridisation and molecular breeding of *C. sinensis* var. *assamica* and *C. sinensis* var. *sinensis* varieties to produce tea with desirable characteristics (27). Therefore, developing a few markers that have the capability of discriminating the two subspecies and their cultivars is of paramount interest. The results on the linkage mapping for all the phenotypic traits and the QTL positions in different linkage groups in this study were similar as reported in (10).

In this study, AM was conducted with three different models, Q, K, and Q+K. The observed –log10 (P) values for QTL deviated from the expected –log10 (P) values in the Q method (GLM), indicated that there might be a few false-positive. However, the addition of genetic relatedness K (kinship) used in a mixed linear model has proven to be more powerful and reduces the number of false-positive associations (28). This is because a lot more QTLs were identified in this current study using Q-model as compared with K and Q + K-model. Therefore, there could be higher chances of false-positive or negative errors in AM due to complex population structure (29, 30). In this study, a mixed linear model approach using the K-matrix or a combination (Q+K) performed better than a GLM (Q). However, the K or the Q + K methods did show to be too strict and resulted in the missing of some possibly useful QTLs. Nearly all the significant QTLs identified in AM using the Q, and Q+K models were in line with those identified in MapQTL 6.0 (10). The percentages of phenotypic variation explained (R^2) by various markers analysed using AM were significant as those analysed using MapQTL 6.0. This is in agreement with previous studies reported in plants on efficiency and robustness of combining linkage mapping and AM for precise identification of QTL (5, 31).

The six additional QTLs identified in AM were found to be associated with purine or thiamine biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, carbon fixation (photosynthesis) and abiotic stress. The putative QTL, qCaffeine located in LG02 was putatively annotated N-(5'phosphoribosyl) anthranilate (PRA) isomerase enzyme. The enzyme is involved in the first and intermediate pathway in the purine or thiamine biosynthesis and tryptophan biosynthesis pathway (32). Caffeine (1,3,7-trimethylxanthine) and other methylxanthines such as theobromine (3,7-dimethylxanthine) and methyluric acids present in tea are classified as purine alkaloids (33). Therefore, the putative QTL, qCaffeine could be associated with the biosynthesis of caffeine, theobromine and methyluric acids in tea. Also,

365 phenylalanine, tyrosine, and tryptophan are products of shikimate pathway which is involved
366 in the biosynthesis of plant flavonoids, including catechins (34).
367 The putative QTLs, qEC, and qEGC in LG04 were annotated as phosphoribulokinase/uridine
368 kinase family proteins which are involved in carbon fixation or photosynthesis in plants (35).
369 Photosynthesis is an important process in plants for provision of carbon skeletons to the
370 shikimate pathway (36). Catechins are synthesised in the leaves of the tea plant through the
371 acetic-malonic acid and shikimic-cinnamic acid metabolic pathways (37). The chalcone and
372 gallic acid are produced from the shikimic acid pathway, which then produce the different
373 catechins (38). The two putative QTLs, qEC annotated putatively as autophagy-related protein
374 11 and protein tyrosine kinase are involved in abiotic stress response through a process of
375 stomatal closure and ABA signalling (39, 40). Stress conditions can interfere with
376 photosynthetic energy production in plants, which leads to stomatal closure, which inhibits
377 CO_2 intake and thereby reduces photosynthetic activity in leaves (41). Abscisic acid has been
378 reported to confer abiotic stress-tolerance in crop plants (39). In stress conditions like drought,
379 extreme temperature, and high salinity, content in plants increases considerably, inspiring
380 stress-tolerance effects that help plants, adapt, and survive under these stressful situations (42).
381 Also, ABA is also required for plant growth and development under non-stress conditions (43).

382 # Conclusion

383 The tea plant is a woody perennial crop with long generation intervals, which usually hinders
384 its genetic improvement by conventional breeding methods. The approach of combining the
385 high QTL detection power of genetic linkage mapping with the high-resolution power of AM
386 allowed identification and precise authentication of putative QTL controlling black tea quality
387 and drought tolerance traits. Based on the two mapping approaches, the putative QTLs
388 associated with caffeine content, catechins fractions, theaflavins fractions, tea tasters' scores,

389 and %RWC detected in linkage mapping and AM were not significantly different except that
390 an additional six more QTLs were detected using AM method. Therefore, the combination of
391 linkage and association strategies can be used to identify closely linked molecular markers for
392 robust MAS and trait introgression. This will accelerate tea breeding progress in terms of
393 reduced breeding time, reduced cost, increased efficiency, and precision of selection.

394 Acknowledgments

395 The authors are grateful to KALRO-Tea Research Institute (Kenya) and the University of
396 Pretoria (South Africa) for providing space for field experiments, facilitating data collection
397 and analysis and laboratory work.

398 References

399 1. Kraakman AT, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA. Linkage
400 disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics.
401 2004;168(1):435-46.
402 2. Soto-Cerda BJ, Cloutier S. Association mapping in plant genomes. Genetic diversity
403 in plants: InTech; 2012.
404 3. Sterken R, Kiekens R, Boruc J, Zhang F, Vercauteren A, Vercauteren I, et al. Combined
405 linkage and association mapping reveals CYCD5; 1 as a quantitative trait gene for
406 endoreduplication in Arabidopsis. Proceedings of the National Academy of Sciences.
407 2012;109(12):4678-83.
408 4. Shi W, Hao C, Zhang Y, Cheng J, Zhang Z, Liu J, et al. A Combined Association
409 Mapping and Linkage Analysis of Kernel Number Per Spike in Common Wheat (Triticum
410 aestivum L.). Frontiers in Plant Science. 2017;8:1412.

411 5. Li X, Zhou Z, Ding J, Wu Y, Zhou B, Wang R, et al. Combined linkage and association
412 mapping reveals QTL and candidate genes for plant and ear height in maize. Frontiers in plant
413 science. 2016;7:833.
414 6. Du Q, Gong C, Wang Q, Zhou D, Yang H, Pan W, et al. Genetic architecture of growth
415 traits in Populus revealed by integrated quantitative trait locus (QTL) analysis and association
416 studies. New Phytologist. 2016;209(3):1067-82.
417 7. Bartholomé J, Bink MC, van Heerwaarden J, Chancerel E, Boury C, Lesur I, et al.
418 Linkage and association mapping for two major traits used in the maritime pine breeding
419 program: height growth and stem straightness. PloS one. 2016;11(11):e0165323.
420 8. Fournier-Level A, Lacombe T, Le Cunff L, Boursiquot J, This P. Evolution of the
421 VvMybA gene family, the major determinant of berry colour in cultivated grapevine (Vitis
422 vinifera L.). Heredity. 2010;104(4):351.
423 9. Jin J-Q, Yao M-Z, Ma C-L, Ma J-Q, Chen L. Association mapping of caffeine content
424 with TCS1 in tea plant and its related species. Plant Physiology and Biochemistry.
425 2016;105:251-9.
426 10. Koech RK, Malebe PM, Nyarukowa C, Mose R, Kamunya SM, Apostolides Z.
427 Identification of novel QTL for black tea quality traits and drought tolerance in tea plants
428 (Camellia sinensis). Tree Genetics & Genomes. 2018;14(1):9.
429 11. Pritchard J, Wen X, Falush D. Documentation for STRUCTURE
430 software.[Documentation file]. Available with the program at http://pritch bsd uchicago
431 edu/structure html. 2010.
432 12. Earl DA. STRUCTURE HARVESTER: a website and program for visualizing
433 STRUCTURE output and implementing the Evanno method. Conservation genetics resources.
434 2012;4(2):359-61.

435 13. Mezmouk S, Dubreuil P, Bosio M, Décousset L, Charcosset A, Praud S, et al. Effect of
436 population structure corrections on the results of association mapping tests in complex maize
437 diversity panels. Theoretical and applied genetics. 2011;122(6):1149-60.
438 14. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program
439 for dealing with label switching and multimodality in analysis of population structure.
440 Bioinformatics. 2007;23(14):1801-6.
441 15. Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG. STRUCTURE PLOT: a program
442 for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus.
443 2014;3(1):431.
444 16. Van Ooijen J. JoinMap® 4, Software for the calculation of genetic linkage maps in
445 experimental populations. Kyazma BV, Wageningen. 2006;33:10.1371.
446 17. Van Ooijen JW, Boer M, Jansen R, Maliepaard C. MapQTL 4.0: software for the
447 calculation of QTL positions on genetic maps (user manual). Plant Research International;
448 2000.
449 18. Sadok IB, Celton J-M, Essalouh L, El Aabidine AZ, Garcia G, Martinez S, et al. QTL
450 mapping of flowering and fruiting traits in olive. PLoS One. 2013;8(5):e62831.
451 19. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL:
452 software for association mapping of complex traits in diverse samples. Bioinformatics.
453 2007;23(19):2633-5.
454 20. Fan R, Xiong M. Combined high resolution linkage and association mapping of
455 quantitative trait loci. European Journal of Human Genetics. 2003;11(2):125.
456 21. Jung J, Fan R, Jin L. Combined linkage and association mapping of quantitative trait
457 loci by multiple markers. Genetics. 2005;170(2):881-98.

458 22. Lu Y, Zhang S, Shah T, Xie C, Hao Z, Li X, et al. Joint linkage–linkage disequilibrium
459 mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance
460 in maize. Proceedings of the National Academy of Sciences. 2010;107(45):19585-90.
461 23. Lou Q, Chen L, Mei H, Wei H, Feng F, Wang P, et al. Quantitative trait locus mapping
462 of deep rooting by linkage and association analysis in rice. Journal of experimental botany.
463 2015;66(15):4749-57.
464 24. Cao Y, Li S, Wang Z, Chang F, Kong J, Gai J, et al. Identification of major quantitative
465 trait loci for seed oil content in soybeans by combining linkage and genome-wide association
466 mapping. Frontiers in plant science. 2017;8:1222.
467 25. Kamunya S, Wachira F, Pathak R, Korir R, Sharma V, Kumar R, et al. Genomic
468 mapping and testing for quantitative trait loci in tea (Camellia sinensis (L.) O. Kuntze). Tree
469 genetics & genomes. 2010;6(6):915-29.
470 26. TRFK. TRFK clonal catalogue. In: TRFK, editor. 2012.
471 27. Yao M, Chen L, Liang Y. Genetic diversity among tea cultivars from China, Japan and
472 Kenya revealed by ISSR markers and its implication for parental selection in tea breeding
473 programmes. Plant Breeding. 2008;127(2):166-72.
474 28. Finno CJ, Aleman M, Higgins RJ, Madigan JE, Bannasch DL. Risk of false positive
475 genetic associations in complex traits with underlying population structure: a case study. The
476 Veterinary Journal. 2014;202(3):543-9.
477 29. Müller B, Stich B, Piepho H. A general method for controlling the genome-wide type I
478 error rate in linkage and association mapping experiments in plants. Heredity.
479 2011;106(5):825.
480 30. George A. Controlling type 1 error rates in genome-wide association studies in plants.
481 Heredity. 2013;111(1):86.

482 31. Mammadov J, Sun X, Gao Y, Ochsenfeld C, Bakker E, Ren R, et al. Combining powers
483 of linkage and association mapping for precise dissection of QTL controlling resistance to gray
484 leaf spot disease in maize (Zea mays L.). BMC genomics. 2015;16(1):916.
485 32. Li J, Zhao J, Rose AB, Schmidt R, Last RL. Arabidopsis phosphoribosylanthranilate
486 isomerase: molecular genetic analysis of triplicate tryptophan pathway genes. The Plant Cell.
487 1995;7(4):447-61.
488 33. Ashihara H, Sano H, Crozier A. Caffeine and related purine alkaloids: biosynthesis,
489 catabolism, function and genetic engineering. Phytochemistry. 2008;69(4):841-56.
490 34. Ghasemzadeh A, Ghasemzadeh N. Flavonoids and phenolic acids: Role and
491 biochemical activity in plants and human. Journal of medicinal plants research.
492 2011;5(31):6697-703.
493 35. Miziorko HM. Phosphoribulokinase: current perspectives on the structure/function
494 basis for regulation and catalysis. Advances in Enzymology and Related Areas of Molecular
495 Biology, Volume 74. 2000:95-127.
496 36. Janacek SH, Trenkamp S, Palmer B, Brown NJ, Parsley K, Stanley S, et al.
497 Photosynthesis in cells around veins of the C3 plant Arabidopsis thaliana is important for both
498 the shikimate pathway and leaf senescence as well as contributing to plant fitness. The Plant
499 Journal. 2009;59(2):329-43.
500 37. Vuong QV, Stathopoulos CE, Nguyen MH, Golding JB, Roach PD. Isolation of green
501 tea catechins and their utilization in the food industry. Food Reviews International.
502 2011;27(3):227-47.
503 38. Chu D, Juneja L. General chemical composition of green tea and its infusion. Chemistry
504 and applications of green tea. 1997:13-22.
505 39. Sah SK, Reddy KR, Li J. Abscisic acid and abiotic stress tolerance in crop plants.
506 Frontiers in plant science. 2016;7:571.

507 40. Batoko H, Dagdas Y, Baluska F, Sirko A. Understanding and exploiting autophagy
508 signaling in plants. Essays in biochemistry. 2017;61(6):675-85.
509 41. Chaves MM, Flexas J, Pinheiro C. Photosynthesis under drought and salt stress:
510 regulation mechanisms from whole plant to cell. Annals of botany. 2009;103(4):551-60.
511 42. Ng LM, Melcher K, Teh BT, Xu HE. Abscisic acid perception and signaling: structural
512 mechanisms and applications. Acta Pharmacologica Sinica. 2014;35(5):567.
513 43. Nakashima K, Yamaguchi-Shinozaki K. ABA signaling in stress-response and seed
514 development. Plant cell reports. 2013;32(7):959-70.

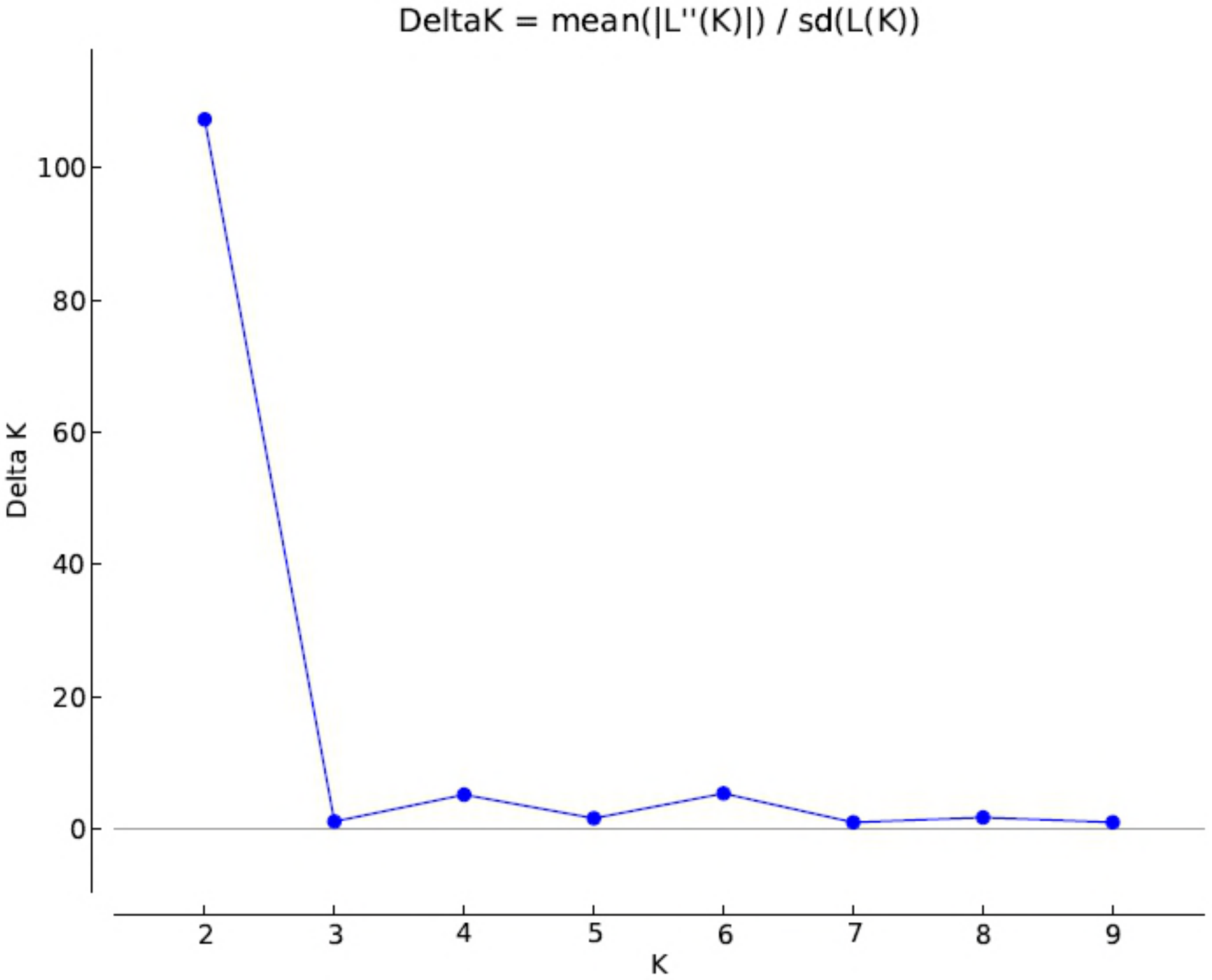
515 Supporting information

516 **S1 Fig. Trait-wise association Q-Q plots for GLM.** AR- Aroma; AST- Astringency; BRK-
517 Briskness; BRT- Brightness; C- Catechin; CAFF- Caffeine; CL- Colour; EC- Epicatechin;
518 ECG- Epicatechin gallate; EGC- Epicatechin gallate; EGCG- Epigallocatechin gallate; RWC-
519 Percent relative water content; TF1- Theaflavin; TF2- Theaflavin-3-gallate; TF3- Theaflavin-
520 3'-gallate; TF4- Theaflavin-3,3'-digallate

521 **S2 Fig. Trait-wise association Q-Q plots for GLM (Q).** AR- Aroma; AST- Astringency;
522 BRK- Briskness; BRT- Brightness; C- Catechin; CAFF- Caffeine; CL- Colour; EC-
523 Epicatechin; ECG- Epicatechin gallate; EGC- Epicatechin gallate; EGCG- Epigallocatechin
524 gallate; RWC- Percent relative water content; TF1- Theaflavin; TF2- Theaflavin-3-gallate;
525 TF3- Theaflavin-3'-gallate; TF4- Theaflavin-3,3'-digallate

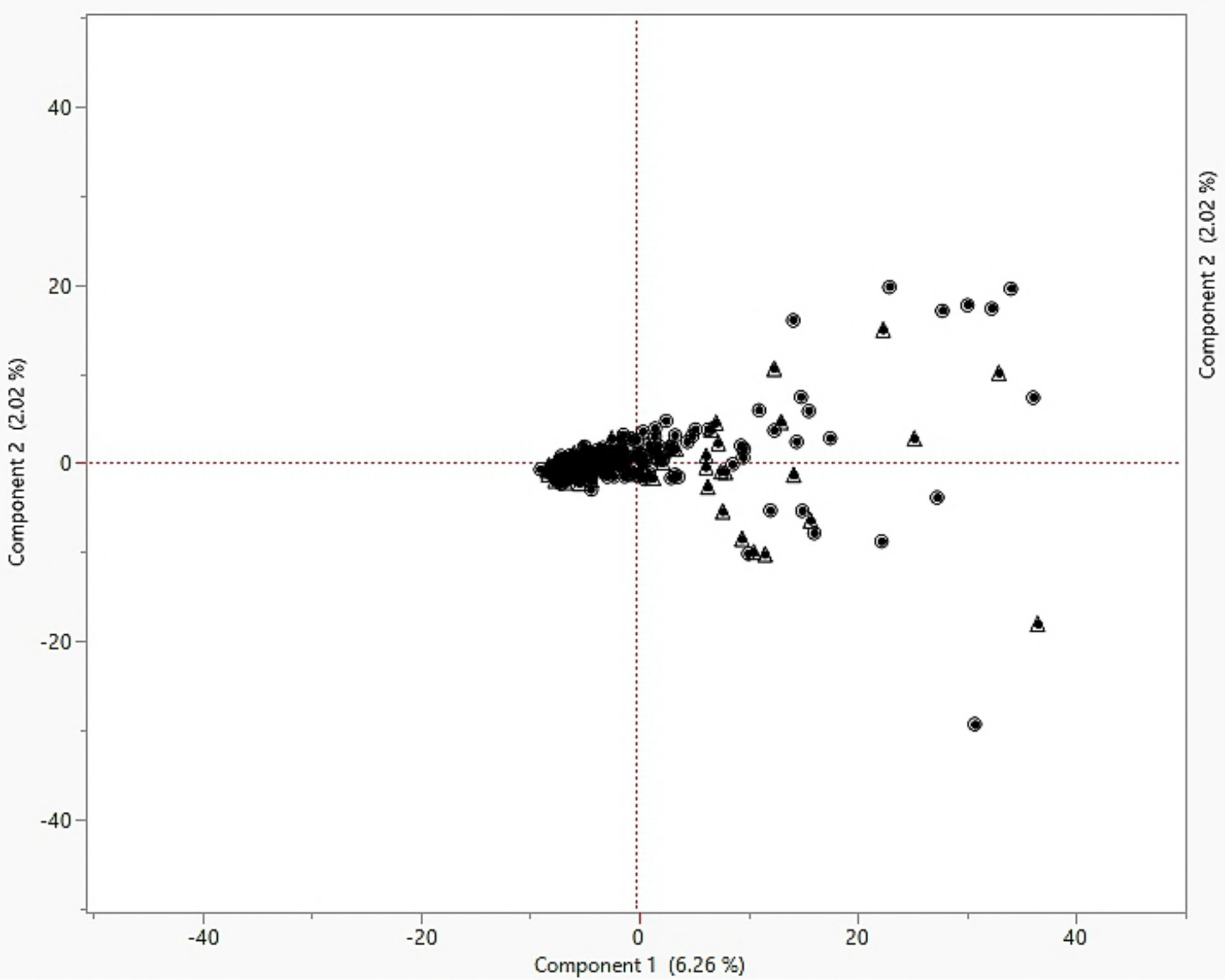
526 **S3 Fig. Trait-wise Q-Q association plots for mixed linear model MLM (K).** AR- Aroma;
527 AST- Astringency; BRK- Briskness; BRT- Brightness; C- Catechin; CAFF- Caffeine; CL-
528 Colour; EC- Epicatechin; ECG- Epicatechin gallate; EGC- Epicatechin gallate; EGCG-
529 Epigallocatechin gallate; RWC- Percent relative water content; TF1- Theaflavin; TF2-
530 Theaflavin-3-gallate; TF3- Theaflavin-3'-gallate; TF4- Theaflavin-3,3'-digallate

S4 Fig. Trait-wise Q-Q association plots for mixed linear model MLM (Q+K). AR- Aroma; AST- Astringency; BRK- Briskness; BRT- Brightness; C- Catechin; CAFF- Caffeine; CL-Colour; EC- Epicatechin; ECG- Epicatechin gallate; EGC- Epicatechin gallate; EGCG-Epigallocatechin gallate; RWC- Percent relative water content; TF1- Theaflavin; TF2-Theaflavin-3-gallate; TF3- Theaflavin-3'-gallate; TF4- Theaflavin-3,3'-digallate

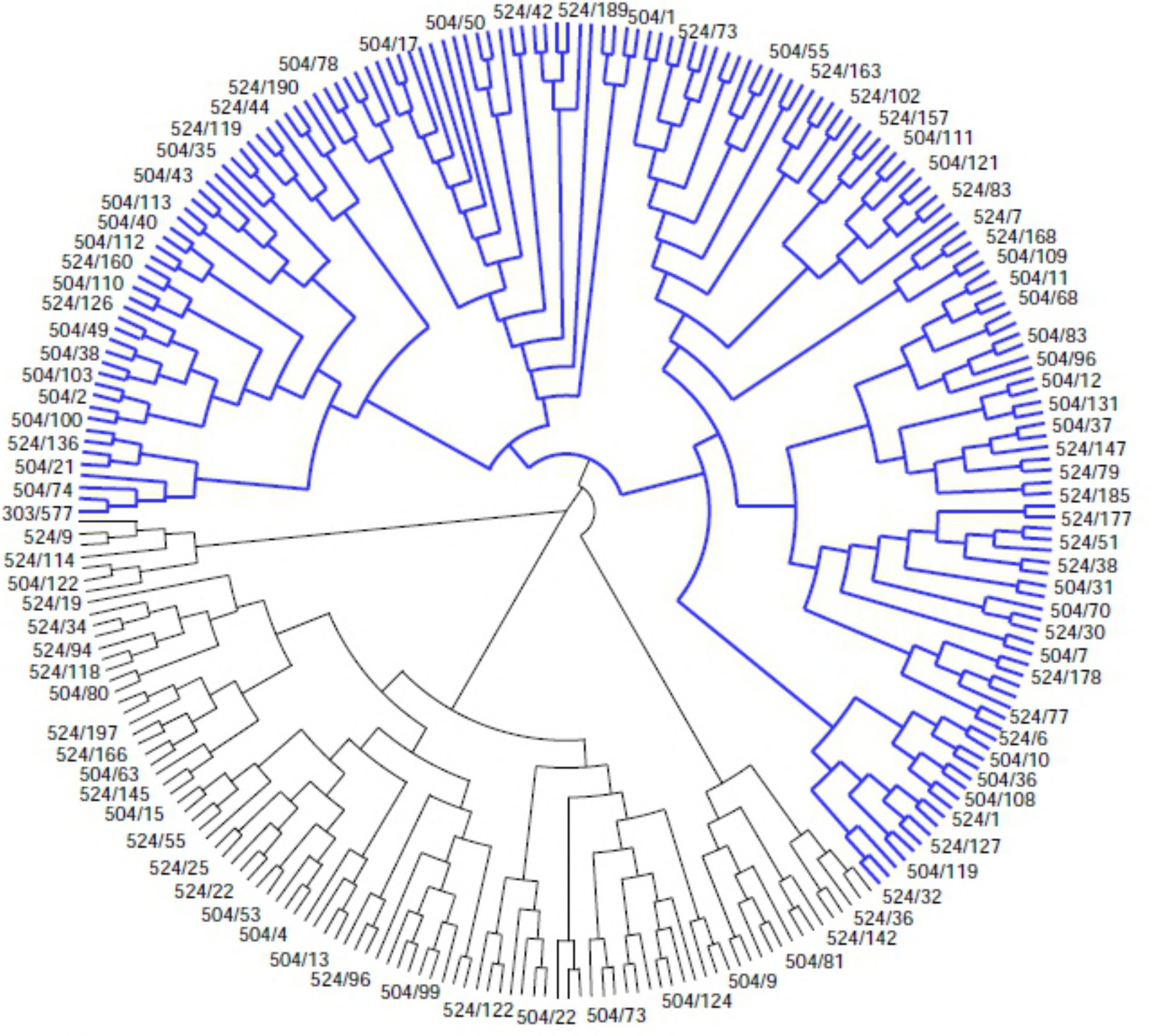


Figure

Figure


40
20
0
-20
-40
Component 2 (2.02 %)
Component 2 (2.02 %)
-40
-20
0
20
40
Component 1 (6.26 %)

Figure


504/50
524/42
524/189
504/1
504/17
524/73
504/78
504/55
524/190
524/163
524/44
524/102
524/119
524/157
504/35
504/111
504/43
504/121
504/113
524/83
504/40
524/7
504/112
524/168
524/160
504/109
504/110
504/11
524/126
504/68
504/49
504/83
504/38
504/96
504/103
504/12
504/2
504/131
504/100
504/37
524/136
524/147
504/21
524/79
504/74
524/185
303/577
524/177
524/9
524/51
524/114
524/38
504/122
504/31
524/19
504/70
524/34
524/30
524/94
504/7
524/118
524/178
504/80
524/77
524/197
524/6
524/166
504/10
504/63
504/36
524/145
504/108
504/15
524/1
524/55
524/127
524/25
504/119
524/22
524/32
504/53
524/36
504/4
524/142
504/13
504/81
524/96
504/9
504/99
504/124
524/122
504/22
504/73

Figure