Construction of a DArT-seq marker–based genetic linkage map and identification of QTLs for yield in tea (*Camellia sinensis* (L.) O. Kuntze)

M. P. Malebe[1,*], R. K. Koech[1,2], E. G. N. Mbanjo[1,3], S. M. Kamunya[2], A. A. Myburg[1], Z. 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] International Institute of Tropical Agriculture (IITA), PMB 5320 Ibadan, Oyo, Nigeria

*Correspondence to: M. P. Malebe. Email: pelly@tuks.co.za

Abstract

As the second most consumed non-alcoholic beverage, the tea plant (*Camellia sinensis*) has high economic value. Tea improvement efforts that largely target economic traits such as yield have traditionally relied on conventional breeding approaches. The tea plant's perennial nature and its long generation time make conventional approaches time-consuming and labour-intensive. Biotechnology provides a complementary tool for accelerating tea improvement programmes through marker-assisted selection (MAS). Quantitative trait loci (QTLs) identified on linkage maps are an essential prerequisite to the implementation of MAS. QTL analysis was performed on yield data over 3 years (2010–2012) across two sites (Timbilil and Kangaita, in Kenya), based on two parental framework linkage maps arising from a population of 261 F_1 progeny, derived from a reciprocal cross between GW Ejulu and TRFK 303/577. The maps contain 15 linkage groups each, this corresponds to the haploid chromosome number of tea ($2n=2x=30$). The total length of the parental maps was 1028.1 cM for GW Ejulu and 1026.6 cM for TRFK 303/577 with an average locus spacing of 5.5 cM and 5.4 cM, respectively. A total of 13 QTLs were identified over the three measurement years. The 13 QTLs had LOD values ranging from 1.98 to 7.24 and explained 3.4% to 12% of the phenotypic variation. The two sites had seven mutually detected QTLs.

Keywords: Yield . QTL . NGS marker . Linkage map . *Camellia sinensis*

Introduction

Camellia sinensis, commonly known as the tea plant, has many cultivars that are selected for different regions of the world. Tea is the second most widely consumed beverage, only second to water. This makes it an economically important crop. The dry young leaves of *Camellia sinensis* are used to produce black, green, oolong, white and purple tea. Grown in over 50 countries, tea has significant economic and social importance. The top three tea-growing countries are China, India and Kenya (FAOSTAT 2018). Globally, Kenya is ranked third in terms of annual tea volume produced, and tea is Kenya's top export commodity among agricultural products (Kenya National Bureau of Statistics 2012; FAOSTAT 2018).

Therefore, it is an important source of income and employment (Pettigrew and Richardson 2014).

Besides it being a major cash crop, tea has documented medicinal properties and health benefits (Khan and Mukhtar 2007; Taylerson 2012). The healing properties of tea have been accredited to the antioxidant properties of the tea flavonoids and phenolic compounds such as epigallocatechin gallate (EGCG) (Mondal et al. 2004).

Tea global consumption has increased by 60% between 1993 and 2000. The human population is predicted to increase by 20% by 2050; significant increase in tea global consumption is also forecasted (Pettigrew and Richardson 2014; OECD/FAO 2020). Therefore, increasing crop yield is of prime importance as yield limitation threatens food security and is detrimental to the economy.

Conventional breeding has been used to boost crop yield. However, the perennial nature of tea and its long generation time make conventional approaches time-consuming and labour-intensive. Biotechnology has the potential to speed up the development of high yield crops through marker-assisted selection (MAS). MAS applicability and effectiveness depend on identifying markers that are tightly linked to genes or quantitative trait loci (QTL) that reliably predict a trait phenotype (Collard and Mackill 2008).

The development of a genetic linkage map is a prerequisite for marker-trait association and trait dissection. Meiotic maps are important assets to deciphering genome structure, organization and evolution. The first linkage map of the tea plant was reported by Hackett et al. (2000) using random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) DNA. The linkage map developed was from 208 markers present in the female parent only; the map covered 1349.7 cM, with an average distance of 11.7 cM between loci (Hackett et al. 2000). Several other maps have subsequently been generated. Taniguchi et al. (2012) constructed a high-density reference linkage map of tea using 54 F_1 clones. This core map contained 1124 markers including 441 simple sequence repeats (SSRs), seven cleaved amplified polymorphic sequence (CAPS), two sequence-tagged sites (STS) and 674 RAPDs and had a total map length of 1218 cM with an average distance between markers of 4.35 cM. Ma et al. (2014) constructed a moderately saturated genetic map in tea plant using 406 SSR markers and 183 individuals. The map had a total length of 1143.5 cM, with an average locus spacing of 2.9 cM. Bali et al. (2015) have developed the first genetic linkage map in Indian beveragial tea using 234 DNA markers (AFLP and RAPD) and 87 F_1 individuals. Using the regression algorithm, the authors successfully positioned 73.5% of the total markers with total map length of 2051.7 cM and 14.96 cM average distance between each (Bali et al. 2015). Tan et al. (2016) used 483 SSR markers to develop a genetic linkage map of 1226.2 cM with an average marker distance of 2.5 cM. RNA sequencing has been used to develop 29 new SSR markers, these were used along with 649 other markers (AFLPs, public SSRs and RAPDs) to generate a genetic linkage map that was 1441.6 cM in length and had an average of 4.7 cM average spacing (Chang et al. 2017). Recently, Xu et al. (2018) used 2b-restriction site-associated DNA (2b-RAD) sequencing to obtain 4463 markers for constructing a 1678.52-cM high-density map with an average interval of 0.40 cM.

The data in this current study was analysed as a pseudo-testcross based on DArT-seq dominant marker systems; separate maps were constructed for each parent. Koech et al. (2019) merged these resulting parental maps using the “hkxhk” locus that was heterozygous

in both parents as anchor markers. Koech et al. (2018) successfully used the integrated linkage map to identify QTLs for drought tolerance in tea and black tea quality traits. The functional annotation of the identified putative QTLs that associated with caffeine, catechin biosynthesis and drought tolerance was performed (Koech et al. 2019; Koech et al. 2020). While several linkage maps for tea exist, there is only one report on QTL analysis for yield in tea to date. Kamunya et al. (2010) developed the linkage map using 42 F_1 clonal progeny on a map containing 19 maternal and 11 paternal linkage groups that covered 1411.5 cM with mean interval of 14.1 cM between loci. The use of a small population may have led to overestimated and spurious QTLs; a larger population should be used (Kamunya et al. 2010).

The tea genome is estimated to be 3.2 Gb in size (Xia et al. 2017; Wei et al. 2018). The first available tea draft genome sequences had large numbers of scaffolds (14,051–37,618) (Xia et al. 2017; Wei et al. 2018). Recently, a chromosome-scale genome assembly of tea was released (Chen et al. 2020; Xia et al. 2020) obtained a final assembly of 2.94 Gb, accounting for 91.9% of the estimated genome size, 86.7% of which was anchored into 15 pseudo-chromosomes. Chen et al. (2020) developed a chromosome-scale assembly of 2.98 Gb, accounting for a larger percentage (94.7%) of the estimated genome size.

Using next-generation sequencing (NGS) markers for genome characterization is cost-effective and time-saving and allows for high-throughput development of markers. Diversity Arrays Technology sequences (DArT-seq) stand as an appropriate and cost-effective system to discover hundreds of polymorphic genomic loci, scoring thousands of unique genomic-wide DNA fragments in one single experiment, without requiring existing DNA sequence information. The DArT complexity reduction approach in combination with Illumina short read sequencing is applied in crop breeding and genetic studies (Sansaloni et al. 2011; Dracatos et al. 2019; Malebe et al. 2019; Nadeem et al. 2020).

Here, we report the construction of framework maps of tea using DArT-seq markers and we show the usefulness of the map for complex trait dissection and identification of novel QTLs for yield.

Materials and methods

Plant material

The reciprocal cross, St 504 (TRFK 303/577 × GW Ejulu) and St 524 (GW Ejulu × TRFK 303/577), involving two heterozygous commercial tea clones GW Ejulu × TRFK 303/577 and its 261 F_1 clonal progeny, was established in 2007 by the Tea Research Institute (TRI) in Kenya. The cross was chosen on the basis of the differing parental attributes. Clone GW Ejulu is a Kenyan China type (*Camellia sinensis* var. *sinensis*) that produces high-quality black tea but is a low yielder, whereas clone TRFK 303/577 is an Assam type (*Camellia sinensis* var. *assamica*) that is high-yielding, and tolerant to water stress. Cuttings were collected from individual seedling bushes rooted and raised in the nursery for 1 year prior to field transplanting. The trial was set up in a completely randomized block design with three replicates in clonal plots spaced at 0.61 m within rows and 1.22 m between rows (i.e. 13,448 plants/ha) in two sites. A guard row of clone TRFK 303/1199 surrounded each replicate. The two sites were in Timbilil (0° 22′ S, 35° 21′ E, 2180 m asl), Kericho County, and Kangaita (0° 30′ S, 37° 18′ E, 2100 m asl), Kirinyaga County. The recommended management practices were followed: Fertilizer was applied at a rate of 150 kg N per hectare per year in the form of NPKS 25:5:5:5 compound fertilizer (Anon 2002). The first formative pruning

was carried out in November 2008 at 45.72 cm at both the Kangaita and Timbilil sites. Pruning was carried out again in January 2012 at 50.8 cm at both sites.

Yield data analysis

Yield data collection took place in the form of plucked two leaves and a bud each year during 2010, 2011 and 2012. Depending on availability of crop, harvesting was carried out at intervals of 7 to 10 days each year. The yield from individual plots was extrapolated to green leaf yield (kg) which was converted to made tea per hectare (mt/ha) by a conversion factor of 0.225 prior to statistical analyses. QTL analysis was performed on yield data for 2010, 2011, 2012 and long-term annual yield means.

Isolation of genomic DNA

Fresh leaves were collected from the TRI and frozen at − 20 °C overnight prior to DNA extraction. DNA extraction from the frozen leaves was carried out using a modified protocol of the CTAB method (Gawel and Jarret 1991). DNA concentration and quality were determined using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific Inc., MA, USA).

Sequence-based DArT genotyping

Sequence-based genotyping was performed on the Illumina HiSeq 2500 at Diversity Arrays Technology Pty Ltd (Canberra, Australia) as described by Sansaloni et al. (2011) using *Pst*I and *Mse*I restriction enzymes. Prior to shipment, DNA quality was verified by testing the digestibility of the DNA using restriction enzyme *Pst*I (Fermentas, Burlington, Canada) and *Eco*RI (Promega, Madison, USA). Each restriction digestion reaction was done as described by the manufacturer. The digested product was then resolved on a 1% agarose gel.

Sequence-based DArT marker filtering

The scored data was received in binary format. Presence or absence of each marker in each individual was scored as 1 or 0, respectively. The markers generated by Diversity Arrays Technology Pty Ltd underwent quality control. Non-informative markers (non-polymorphic markers in the progeny) were discarded. Markers that had conflicting results in the duplicated parents were removed as well as the markers with missing proportion of more than 10%. Only markers that were scored differently (0 and 1) in the parent, resulting in polymorphism within the progeny, were retained.

Linkage analysis

Linkage analysis was performed using JoinMap4.1 (Van Ooijen 2006). The population type code was CP (cross pollination) and the two-way pseudo-testcross strategy for heterozygous outbreeding species was used (Grattapaglia and Sederoff1994). The coding was as follows: firstly, lmxll represented locus that is heterozygous in TRFK 303/577 and homozygous in GW Ejulu. Secondly, nnxnp represented locus that is heterozygous in GW Ejulu and homozygous in TRFK 303/577. Finally, hkxhk represented locus that was heterozygous in both parents. The JoinMap “create maternal and paternal population nodes” option was selected and segregation data was split into maternal and paternal datasets. Separate parental linkage maps were constructed for both St 504 and St 524 sub-populations. The segregation

deviation (SD) from the expected Mendelian ratios was assessed by the calculation of chi-square statistics for each marker. Distorted markers that had a chi-square value of more than 9.9, that disturbed the neighbouring markers, were excluded. Linkage groups were established using a minimum logarithm of odds ratio (LOD) for linkage ranging from 3 to 12. Preferential selection of nodes that showed a stable number of markers at a high LOD was applied. Calculation of recombination frequencies for all the pairs of markers that belong to a certain linkage group determined the map order. The parameters for the recombination frequencies calculation were as follows: maximum recombination threshold value of 0.4; minimum LOD score for calculating map distance of 1; goodness-of-fit jump threshold for removal of loci of 5 and number of added loci after which to perform a ripple of 1. The Kosambi genetic mapping function was used to convert the recombination frequency into map distance. The linkage groups for St 504 and St 524 sub-populations were combined for parental map integration using the JoinMap function combined groups for map integration. MapChart 2.3 (Voorrips 2002) was used to draw the final map.

QTL mapping

The data for each parental meiosis was analysed separately. The <lmxll> and <nnxnp> datasets were translated to a doubled haploid population (DH). QTL mapping was achieved using MapQTL 6.0 software (Van Ooijen and Kyazma 2009). A permutation test (1000 iterations) was performed at a genome-wide level of 5% to determine the LOD score threshold for QTL declaration. Interval mapping was performed to detect putative QTLs for yield. This was followed by multiple QTL model (MQM) to refine QTL positions. QTL intervals were indicated on groups using MapChart 2.3 using the LOD-1 confidence region. MAPQTL 6.0 was used to calculate the percentage of phenotypic variance explained by each QTL.

Results

Phenotypic data

Variation of mean yield across Kangaita and Timbilil experimental sites was evaluated (Fig. 1 and Suppl. Fig. 1 and 2). Annual mean yield in Kangaita among the progeny had a range of 696–1981 kg mt/ha (Table 1). The annual mean yield in Timbilil ranged from 342 to 2596 kg mt/ha. The F_1 means for Kangaita and Timbilil were 1041and 1662 mt/ha, respectively. The mid-parent value, calculated as the average of the yield of both parents, for Kangaita and Timbilil was 1625 and 1343 mt/ha, respectively. The mid-parent heterosis (occurs when the hybrid displays high yield which is greater than the average of both parents) for Kangaita and Timbilil was calculated at − 36% and 24%, respectively. The best performing progeny at Kangaita gave 18% higher yield over the mid-parent value. The best performing progeny at Timbilil outperformed the mid-parent value by 48%. The correlation calculated for long-term annual yield means between the two sites was 0.18 ($p < 0.001$) representing a weak positive linear relationship. Overall yield assessed in the two sites did not behave the same: F_1 means were lower than the parental means for Kangaita and higher than the parental means for Timbilil. The normality of distribution for yield as measured in both Kangaita and Timbilil displayed a continuous distribution.

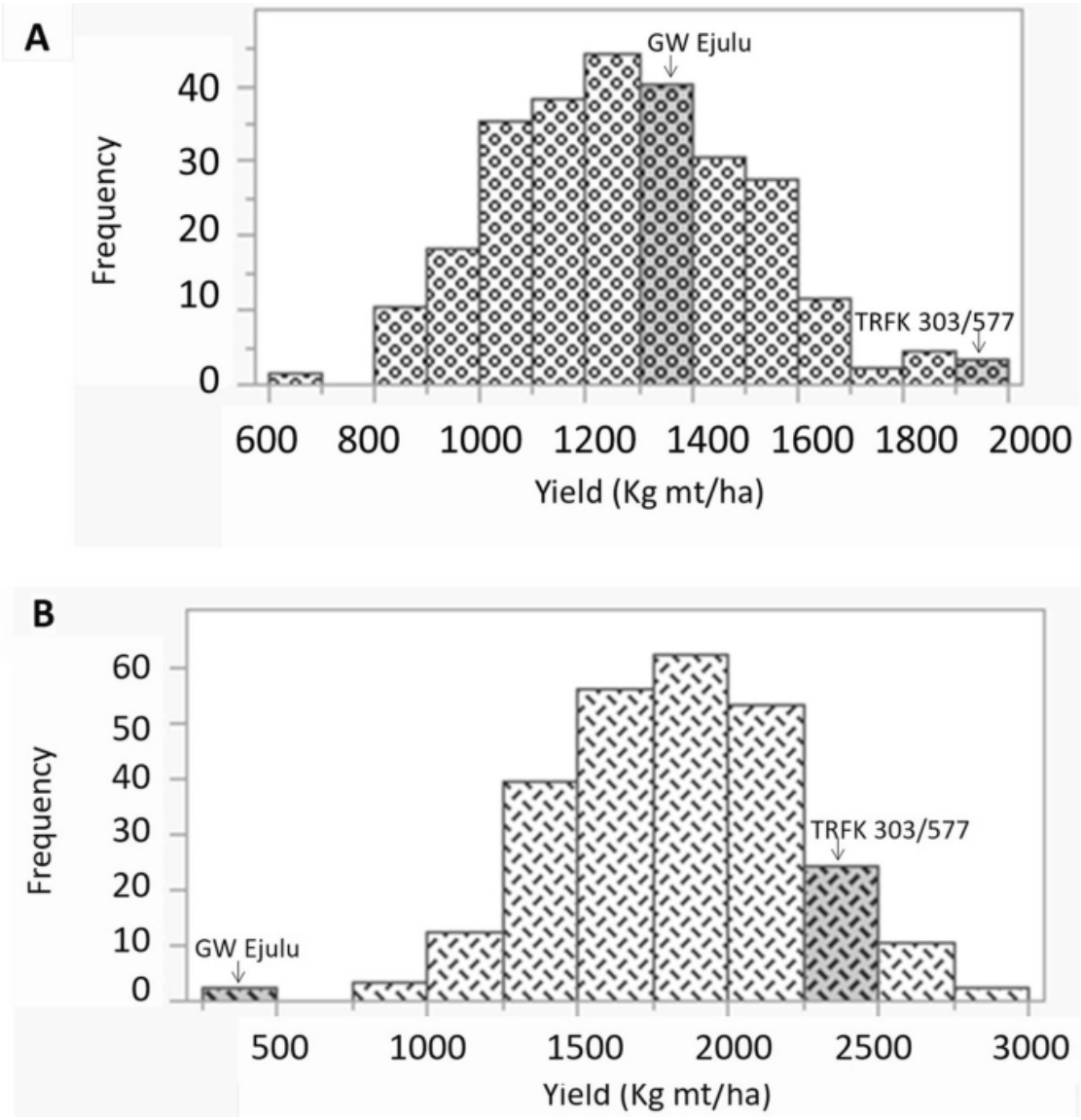

Fig. 1. Frequency distribution of annual means for St 504 and St 524 progeny at the two sites: **a** Kangaita and **b** Timbilil. Parental values are indicated with shaded bar graphs; transparent shading indicates first filial (F_1) generation

Table 1. Phenotypic data analysis at the two sites, Kangaita and Timbilil

	Kangaita	Timbilil
Annual mean (kg mt/ha)	696–1981	342–2596
F_1 means (kg mt/ha)	1041	1662
Mid-parent value (kg mt/ha)	1625	1343
Mid-parent heterosis	−36%	24%

Marker filtering

Following the quality control of the DNA, 261 offspring (109 from St 504 and 152 from St 524) and the duplicated parents underwent DArT-seq. A total of 17,474 DArT-seq markers were developed and the data underwent quality control (Table 2). Out of the 17,474 DArT markers assessed, 1187 (6.8%) were non-informative. We identified 2370 (14.5%) markers with conflicting genotypes (i.e. coded as 0 and 1 or missing code in the NGS data) in TRFK 303/577 and GW Ejulu, respectively. We reported 3018 (21%) markers that had a missing proportion of more than 10%. Markers that were duplicated within the offspring were deleted, 124 (1.1%). The initial nearest neighbour fit was calculated and the 3605 (33.5%) markers with high nearest neighbour fit values were removed. In total, 10,304 markers were filtered out. This resulted in 7170 coded markers, of which 2207 were GW Ejulu informative, 2929 were TRFK 303/577 informative, while 2034 were non-informative.

Table 2. DArT-seq marker filtering for quality control

Quality control	Number of markers remaining
Total	17,474
Polymorphic markers (6.8% of non-polymorphic markers were removed)	16,287
Non-conflicting markers	13,917
Missing proportion <10%	10,899
Remove duplicated markers	10,775
Initial nearest neighbour fit	7170

Construction of linkage maps

The GW Ejulu framework map consisted of 187 informative segregating markers spread over the 15 linkage groups covering 1028.1 cM, with an average locus spacing of 5.5 cM (Fig. 2). The number of mapped markers per linkage group ranged from 4 (linkage group 2) to 21 (linkage group 15) with an average of 13 markers per linkage group (Table 3). The linkage group size ranged from 12.2 cM (linkage group 2) to 95.1 cM (linkage group 1); the average size was 68.5 cM. Distorted segregation was observed in 58 markers (Suppl. Table 1).

Fig. 2. GW Ejulu DArT-seq marker–based genetic map of the tea plant, total map length of 1028.1 cM. Map positions are indicated on the left of the bar in cM and names of loci are indicated of the right of the bar. Linkage groups 5, 6, 7, 8, 10 and 15 showing positions of the multiple yield QTLs. Significant QTL for yield is shown on the right of the bar. The bars and lines indicate 1-LOD and 2-LOD support intervals

Table 3 GW Ejulu linkage map marker distribution among the linkage groups

Linkage group	Number of markers	Total length covered (cM)	Average distance between markers (cM)
LG1	12	95.1	7.9
LG2	4	12.2	3.1
LG3	8	72.8	9.1
LG4	16	91.8	5.7
LG5	14	65.7	4.7
LG6	16	69.1	4.3
LG7	16	79.4	5.0
LG8	12	47.2	3.9
LG9	8	68.3	8.5
LG10	12	55.2	4.6
LG11	8	38.7	4.8
LG12	19	79.2	4.2
LG13	12	94.6	7.9
LG14	9	75.1	8.3
LG15	21	83.7	4.0
Total	187	1028.1	5.5

The TRFK 303/577 framework map consisted of 190 informative segregating markers spread over the 15 linkage groups covering 1026.6 cM, with an average locus spacing of 5.4 cM (Fig. 3.). The number of mapped markers per linkage group ranged from 8 (linkage group 7) to 18 (linkage group 2) with an average of 13 markers per linkage group (Table 4). The linkage group size ranged from 12.4 cM (linkage group 7) to 109.1 cM (linkage group 9); the average size was 68.4 cM. Distorted segregation was observed in 54 markers in the TRFK 303/577 map. The 15 linkage groups generated corresponded to the haploid number of chromosomes found in tea ($2n=2x=30$). Both parental maps had similar total map length (1028.1 cM and 1026.6 cM).

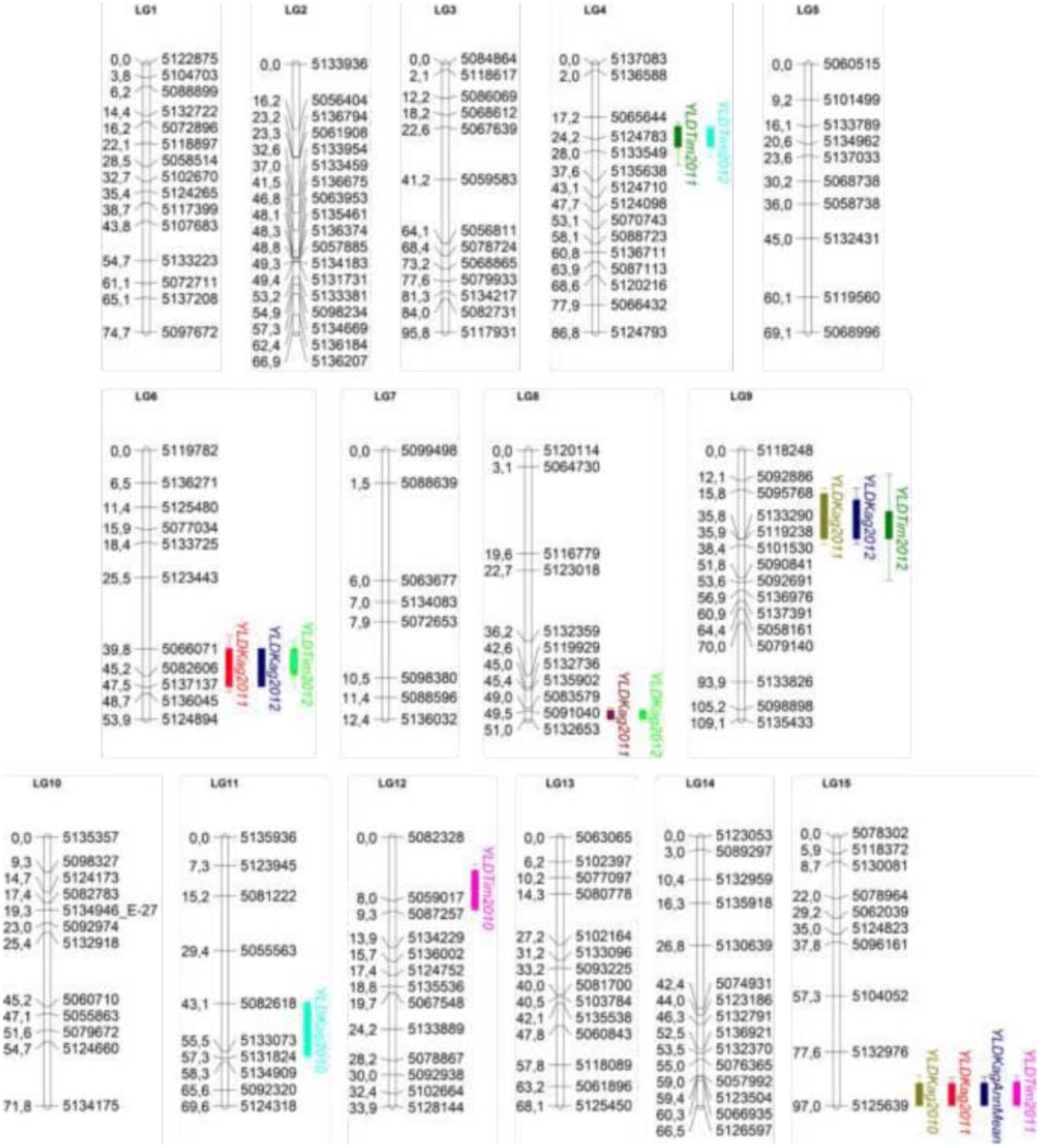

Fig. 3. TRFK 303/577 DArT-seq marker–based genetic map of the tea plant, total map length 1026.6 cM. Map positions are indicated on the left of the bar in cM and names of loci are indicated of the right of the bar. Linkage groups 4, 6, 8, 9, 11, 12 and 15 showing positions of the multiple yield QTLs. Significant QTL for yield is shown on the right of the bar. The bars and lines indicate 1-LOD and 2-LOD support interval

Table 4. TRFK 303/577 linkage map marker distribution among the linkage groups

Linkage group	Number of markers	Total length covered (cM)	Average distance between markers (cM)
LG1	15	74.7	5.0
LG2	18	66.9	3.7
LG3	13	95.8	7.4
LG4	15	86.8	5.8
LG5	10	69.1	6.9
LG6	11	53.9	4.9
LG7	8	12.4	1.5
LG8	11	51.0	4.6
LG9	15	109.1	7.2
LG10	12	71.8	5.9
LG11	10	69.6	6.7
LG12	13	33.9	2.6
LG13	14	68.1	4.9
LG14	15	66.5	4.4
LG15	10	97.0	9.7
Total	190	1026.6	5.4

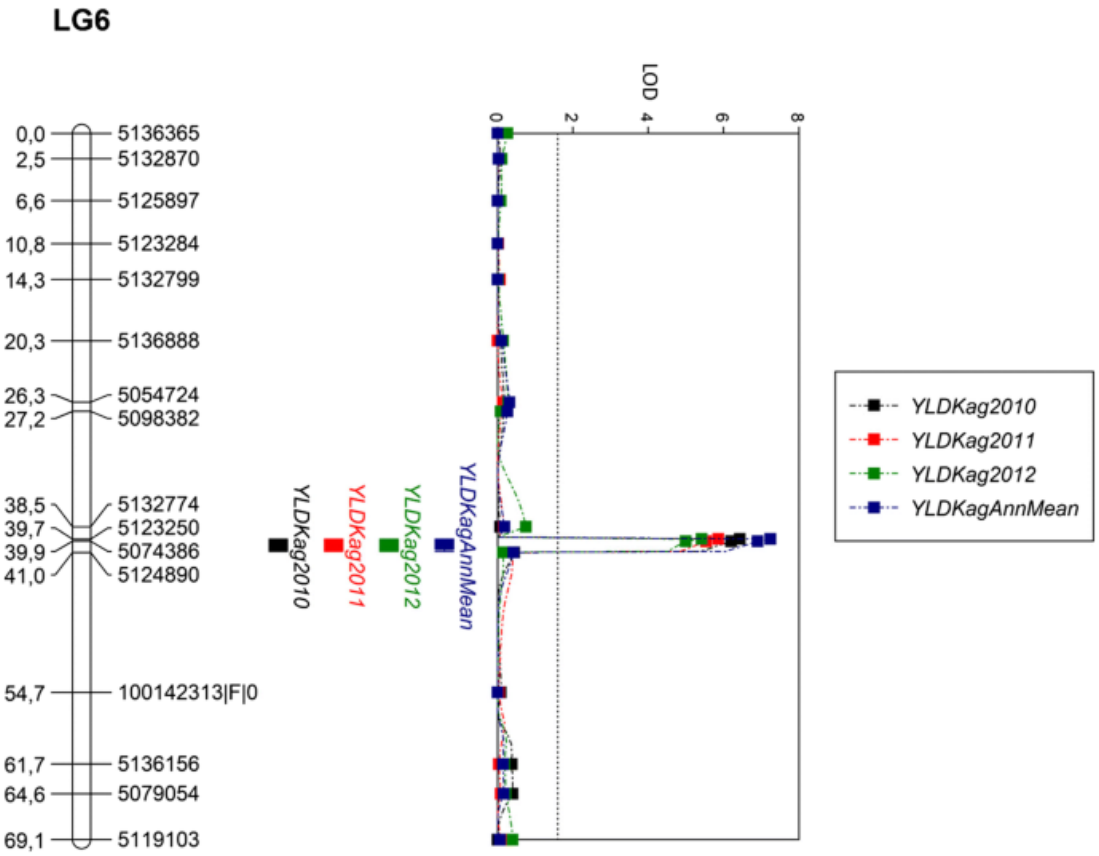

Fig. 4. Locations of QTL for yield in GW Ejulu linkage group 6. LOD score is plotted against marker location. QTL identified across the 3 years (2010–2012) at Kangaita site

QTL analysis

A total of 13 QTLs associated with yield of tea were revealed by both interval mapping and MQM analyses, across the two sites over the 3 years. The 13 QTLs were located on 13 linkage groups, namely, GW Ejulu linkage groups 5, 6, 7, 8, 10 and 15 as well as TRFK 303/577 linkage groups 4, 6, 8, 9, 11, 12 and 15 (Suppl. Fig. 3 and Suppl. Fig. 4). A QTL located on GW Ejulu linkage group 6 was validated across the 3 years (Fig. 4). Over the period, seven QTLs were mutually detected at the two sites (Table 5). The Kangaita site had three QTLs that were detected, while the remaining three QTLs were detected in the Timbilil site. The phenotypic variance explained by each QTL ranged from 3.4 to 12%. Out of the 13 QTLs, seven were inherited directly from the high-yield parent, TRFK 303/577. The remaining six QTLs were inherited directly from GW Ejulu. The nucleotide sequences of the DArT-seq markers are presented in Suppl. Table 2. We performed homology search against cultivar Yunkang 10's genome (Xia et al. 2017). Putative annotation of the QTL markers was carried out, 8 of 13 QTL markers were annotated (Suppl. Table 3). The sequences of the remaining five QTLs did not align significantly to Yunkang 10's genome.

Table 5. QTLs for yield detected in tea plant at Kangaita and Timbilil sites

Site	Linkage group	Locus	LOD threshold	LOD score	Inheritance	Phenotypic variance explained (%)
Kangaita & Timbilil	5	5132285	1.8	2.99	GW Ejulu	5.1
Kangaita & Timbilil	5	5118812	1.8	2.30	GW Ejulu	4.0
Kangaita & Timbilil	5	5135127	1.8	2.32	GW Ejulu	3.8
Kangaita	6	5123250	1.8	7.24	GW Ejulu	12.0
Kangaita	6	5074386	1.8	6.90	GW Ejulu	11.5
Timbilil	7	100013614|F|0	1.7	4.09	GW Ejulu	6.8
Timbilil	7	5136366	1.7	3.78	GW Ejulu	6.4
Kangaita & Timbilil	8	5085451	1.6	4.19	GW Ejulu	7.1
Kangaita & Timbilil	8	100127514|F|0	1.6	3.78	GW Ejulu	6.4
Kangaita & Timbilil	8	5117807	1.6	3.41	GW Ejulu	5.8
Kangaita & Timbilil	10	5133598	1.7	2.87	GW Ejulu	4.9
Kangaita & Timbilil	10	5056921	1.7	2.47	GW Ejulu	4.3
Kangaita & Timbilil	15	5122834	1.9	3.55	GW Ejulu	6.0
Kangaita & Timbilil	15	5095072	1.9	3.42	GW Ejulu	5.8
Kangaita & Timbilil	15	5111164	1.9	2.15	GW Ejulu	3.7
Timbilil	4	5124783	1.7	3.65	TRFK 303/577	6.2
Timbilil	4	5133549	1.7	2.01	TRFK 303/577	3.5
Kangaita & Timbilil	6	5066071	1.6	4.09	TRFK 303/577	7.0
Kangaita & Timbilil	6	5082606	1.6	3.54	TRFK 303/577	6.1
Kangaita	8	5091040	1.5	3.26	TRFK 303/577	5.6
Kangaita	8	5132653	1.5	3.03	TRFK 303/577	5.2
Kangaita & Timbilil	9	5133290	1.9	2.47	TRFK 303/577	4.3
Kangaita	11	5133073	1.4	2.01	TRFK 303/577	3.5
Kangaita	11	5082618	1.4	1.98	TRFK 303/577	3.4
Timbilil	12	5059017	1.7	2.31	TRFK 303/577	4.0
Kangaita & Timbilil	15	5125639	1.8	2.85	TRFK 303/578	4.9

Discussion

Yield and environment

Quantitative traits are governed by multiple genes under environmental influence. Yield recording was done more or less simultaneously in the two sites. F_1 means were lower than the parental means for Kangaita and higher than the parental means for Timbilil. This could be attributed to genotype by environment interactions. There were seven stable QTLs that influenced the yield trait as observed at the two sites, demonstrating their potential as candidate markers for marker-assisted selection. GW Ejulu linkage groups 5, 8, 10 and 15 as well as TRFK 303/577 linkage groups 6, 9 and 15 had QTLs that were detected in both sites. This is the first report of QTLs that influenced yield in tea at different two sites.

The remaining three QTLs that were only detected in Kangaita site and the other three QTLs that were only detected in the Timbilil site was due to gene-environment interactions (Kamunya et al. 2010). The complex environmental factors in the two sites may have had predominant effects on QTLs governing yield, implying that multiple QTLs with small to moderate effects exist for each yield trait and it only requires significant environmental change to trigger a QTL (Kamunya et al. 2010). The genetic variance among a collection of genotypes may change with the environment, the effects of given allele substitutions may be different in one environment than in another (Kearsey and Pooni 1996). This is illustrated by the correlation calculated for long-term annual yield means between the two sites of 0.18 ($p < 0.001$). The QTL on GW Ejulu linkage group 6 that consistently associated with yield over the 3 years within Kangaita indicated potential as a candidate marker for site-specific marker-assisted selection. This highlights the need for selection and evaluation of yield to be done at the particular target site of exploitation for optimal performance of a variety (Kearsey and Pooni 1996). The tea improvement programme in Kenya has incorporated a crucial phase, namely clonal adaptability trial ahead of cultivar release for commercial utilization (TRI 2017). The two parental maps are available for genetic studies and QTL mapping of traits of agronomical importance such as yield in the tea plant.

DArT-seq markers

DArT mapping studies in plants have suggested that DArT markers have a reasonably uniform genomic distribution (Kullan et al. 2012). A marker density of less than 10 cM is recommended for genome-wide QTL mapping (Doerge 2002). The parental maps had an average distance between two markers of 5.5 cM and 5.4 cM, so the constructed genetic maps in the present study are considered to be suitable for identification of QTLs. The previously generated maps used AFLP, CAPS, RAPD, SSR and STR markers proving to be informative for genetic analysis but were still limited in throughput for rapid genome-wide genetic analysis. Therefore, more genetic maps were constructed using larger mapping populations and increasing the number of informative markers (Koech et al. 2018; Xu et al. 2018). The DArT-seq markers used in the framework maps provide anchor points for map integration. As research heads towards the full-genome sequencing of tea, these parental framework maps will be useful in establishing the alignment of DNA sequences of tea.

Genetic linkage map

The first linkage map of the tea plant was first reported by Hackett et al. (2000). More recently, Xu et al. (2018) constructed a highly saturated genetic map in tea plant using 4463

SNPs and SSRs markers and 327 F_1 individuals; QTLs related to caffeine content and flavonoids were mapped.

The parental framework maps developed in this study spanned a total length of 1028.1 and 1026.6 cM which is close to the 1143.5 cM tea genome obtained by Ma et al. (2014). Variations in the number of recombination events in the two parents as well as variations in the locations and number of mapped loci may result in differences in the total map length. The 15 linkage groups of each constructed parental map correspond to tea's haploid chromosome number ($n = 15$); this indicates that the markers were spread across the genome. Future investigation may establish the relationship of the current parental framework maps with the tea reference map by utilizing anchor markers from the reference map (Taniguchi et al. 2012). There is also potential to locate the mapped markers on the newly established chromosome-scale genome of tea (Chen et al. 2020; Xia et al. 2020).

Segregation distortion is observed when genotypic frequencies of a locus deviate from the expected Mendelian ratios and has been described in tea (Hackett et al. 2000; Huang et al. 2005; Kamunya et al. 2010; Hu et al. 2012; Ma et al. 2014). In this study, 12% of the loci exhibited slight segregation distortion ($0.01<P\leq0.05$) and 18% were severely distorted ($P\leq0.01$). This range is comparable to previously observed segregation distortion in tea (12–32.9%) (Hackett et al. 2000; Kamunya et al. 2010; Hu et al. 2012; Ma et al. 2014). The distorted markers were distributed across the linkage groups, except for GW Ejulu linkage group 12 and TRFK 303/577 linkage groups 1, 7, 8, 9, 11 and 15. There are several factors that contribute to segregation distortion, including sample size, genotyping errors, gametic selection and zygotic selection. An in-depth study may be carried out to investigate the reason for the observed segregation distortion.

QTL mapping of tea yield

This study is the first report of the identification of QTLs for yield in tea plant using a saturated linkage map. The phenotypic variance explained by the QTLs for yield ranged between 3.4 and 12%. The low level of phenotypic variance explained by these QTLs suggests that yield may be controlled by a larger number of critical genes.

Conclusion

In this study, we have developed moderate coverage genetic linkage maps of tea using 261 F_1 individuals derived from a reciprocal cross. These maps were used to investigate the genetic basis of yield. The maps could be used to refine the newly released tea reference genome sequence; they also constitute an important asset for further genomic studies in the tea plant. The results of this study provide a foundation for further characterization of QTLs for utilization in improvement programmes targeting economic traits in tea. As phenotypic scores become available, the maps developed in this study will be used to identify associations with other traits of economic importance.

Acknowledgements

The authors thank the staff of Crop Improvement and Management, Tea Research Institute, who assisted in phenotypic data collection and extraction of DNA from the mapping population. We acknowledge Mr. Richard Mose for his critical reading and feedback on this manuscript.

The first author acknowledges the financial assistance of the Carnegie Regional Initiative in Science and Education (Carnegie-RISE) through the Southern African Biochemistry and Informatics for Natural Products (SABINA) network. The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged by the first author. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF, Carnegie-RISE or SABINA.

Data archiving

The DArT sequences have been submitted to NCBI: BioProject PRJNA398959 (Suppl. Table 4)
https://www.ncbi.nlm.nih.gov/biosample?LinkName=bioproject_biosample_all&from_uid=39895.

References

Anon (2002) Tea growers handbook, 5th edn. Tea Research Foundation of Kenya, Kericho, p 261

Bali S, Mamgain A, Raina SN, Yadava SK, Bhat V, Das S, Pradhan AK, Goel S (2015) Construction of a genetic linkage map and mapping of drought tolerance trait in Indian beveragial tea. Mol Breed 35(5):1–20

Chang Y, Oh EU, Lee MS, Kim HB, Moon DG, Song KJ (2017) Construction of a genetic linkage map based on RAPD, AFLP, and SSR markers for tea plant (*Camellia sinensis*). Euphytica 213(8):190

Chen JD, Zheng C, Ma JQ, Jiang CK, Ercisli S, Yao MZ, Chen L (2020) The chromosome-scale genome reveals the evolution and diversification after the recent tetraploidization event in tea plant. Hort Res 7(1):1–1

Collard BC, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos Trans R Soc Lond B Biol 363(1491):557–572

Doerge RW (2002) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev Genet 3(1):43–52

Dracatos PM, Haghdoust R, Singh RP, Huerta Espino J, Barnes CW, Forrest K, Hayden M, Niks RE, Park RF, Singh D (2019) High-density mapping of triple rust resistance in barley using DArT-seq markers. Front Plant Sci 10:467

OECD/FAO (2020) OECD-FAO Agricultural Outlook 2020-2029, FAO, Rome/OECD Publishing, Paris. https://doi.org/10.1787/1112c23b-en

FAOSTAT (2018) Food and Agriculture Organisation of the United Nations. Statistical database. FAO, Rome. http://www.fao.org

Gawel NJ, Jarret RL (1991) A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*. Plant Mol Biol Rep 9(3):262–6

George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156(4):2081–2092

Grattapaglia D, Sederoff R (1994) Genetic linkage maps of Eucalyptus grandis and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137(4):1121–1137

Hackett CA, Wachira FN, Paul S, Powell W, Waugh R (2000) Construction of a genetic linkage map for *Camellia sinensis* (tea). Heredity 85:346–355

Hu Z, Reecy JM, Wu X (2012) Design database for quantitative trait loci (QTL) data warehouse, data mining, and meta-analysis. In: Rifkin SA (ed) Quantitative trait loci (QTL) methods and protocols. Humana Press, San Diego, pp 121–144

Huang J, Li J, Huang Y, Luo J, Gong Z, Liu Z (2005) Construction of AFLP molecular markers linkage map in tea plant. J Tea Sci 25:7–15

Kamunya SM, Wachira FN, Pathak RS, Korir R, Sharma V, Kumar R, Bhardwaj P, Muoki RC, Ahuja PS, Sharma RK (2010) Genomic mapping and testing for quantitative trait loci in tea (*Camellia sinensis* (L.) O. Kuntze). Tree Genet Genomes 6(6):915–929

Kearsey MJ, Pooni HS (1996) The genetical analysis of quantitative traits. Stanley Thorne, Cheltenham, p 381

Kenya National Bureau of Statistics (2012) Kenya facts and figures 2012

Khan N, Mukhtar H (2007) Tea polyphenols for health promotion. Life Sci 81(7):519–533

Koech RK, Malebe PM, Nyarukowa C, Mose R, Kamunya SM, Apostolides Z (2018) Identification of novel QTL for black tea quality traits and drought tolerance in tea plants (*Camellia sinensis*). Tree Genet Genomes 14(1):9

Koech RK, Malebe PM, Nyarukowa C, Mose R, Kamunya SM, Joubert F, Apostolides Z (2019) Functional annotation of putative QTL associated with black tea quality and drought tolerance traits. Sci Rep 9(1):1465

Koech RK, Malebe PM, Nyarukowa C, Mose R, Kamunya SM, Loots T, Apostolides Z (2020) Genome-enabled prediction models for black tea (*Camellia sinensis*) quality and drought tolerance traits. Plant Breed 00:1–13

Kullan ARK, Van Dyk MM, Jones N, Kanzler A, Bayley A, Myburg AA (2012) High-density genetic linkage maps with over 2,400 sequence-anchored DArT markers for genetic dissection in an F2 pseudo-backcross of *Eucalyptus grandis*× *E. urophylla*. Tree Genet Genomes 8:163–175

Ma JQ, Yao MZ, Ma CL, Wang XC, Jin JQ, Wang XM, Chen L (2014) Construction of a SSR-based genetic map and identification of QTLs for catechin content in tea plant (*Camellia sinensis*). PLoS One 9(3):e93131. https://doi.org/10.1371/journal.pone.0093131

Malebe MP, Mphangwe NI, Myburg AA, Apostolides Z (2019) Assessment of genome-wide DArT-seq markers for tea *Camellia sinensis* (L.) O. Kuntze germplasm analysis. Tree Genet Genomes 15(4):1–9

Mondal TK, Bhattacharya A, Laxmikumaran M, Ahuja PS (2004) Recent advances of tea (*Camellia sinensis*) biotechnology. Plant Cell Tiss Org 76:195–254

Nadeem MA, Karaköy T, Yeken MZ, Habyarimana E, Hatipoğlu R, Çiftçi V, Nawaz MA, Sönmez F, Shahid MQ, Yang SH, Chung G (2020) Phenotypic characterization of 183 Turkish common bean accessions for agronomic, trading, and consumer-preferred plant characteristics for breeding purposes. Agron 10(2):272

Pettigrew J, Richardson B (2014) A social history of tea: tea's influence on commerce, culture & community. Benjamin Press, New Port Richey

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011) Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. BMC Proc 5(7):P54

Tan LQ, Wang LY, Xu LY, Wu LY, Peng M, Zhang CC, Wei K, Bai PX, Li HL, Cheng H, Qi GN (2016) SSR-based genetic mapping and QTL analysis for timing of spring bud flush, young shoot color, and mature leaf size in tea plant (*Camellia sinensis*). Tree Genet Genomes 12(3):52

Taniguchi F, Furukawa K, Ota-Metoku S, Yamaguchi N, Ujihara T, Kono I, Fukuoka H, Tanaka J (2012) Construction of a high-density reference linkage map of tea (*Camellia sinensis*). Breed Sci 62:263–273

Taylerson K (2012) The health benefits of tea varieties from *Camellia sinensis*. The Plymouth Stud Sci 5(1):304–312

Tea Research Institute (TRI) (2017). Annual technical report. Kenya agricultural & livestock research organization (KALRO) pp 1–50

Van OoiJen JW (2006) JoinMap ® 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands

Van Ooijen J, Kyazma B (2009) MapQTL 6. Software for the mapping of quantitative trait loci in experimental populations of diploid species. Kyazma BV, Wageningen

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wei C, Yang H, Wang S, Zhao J, Liu C, Gao L, Xia E, Lu Y, Tai Y, She G, Sun J (2018) Draft genome sequence of Camellia sinensis var. sinensis provides insights into the evolution of the tea genome and tea quality. PNAS 1;115(18): E4151–8

Xia E, Zhang H, Sheng J, Li K, Zhang Q, Kim C, Zhang Y, Liu Y, Zhu T, Li W (2017) The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. Mol Plant 10(5):866–877

Xia E, Tong W, Hou Y, An Y, Chen L, Wu Q, Liu Y, Yu J, Li F, Li R, Li P (2020) The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into genome evolution and adaptation of tea plants. Mol Plant 13:1013–1026

Xu L, Wang L, Wei K, Tan L, Su J, Cheng H (2018) High-density SNP linkage map construction and QTL mapping for flavonoid-related traits in a tea plant (*Camellia sinensis*) using 2b-RAD sequencing. BMC Genomics 19:955