



Genetic diversity and structure of Ugandan tea (*Camellia sinensis* (L.) O. Kuntze) germplasm and its implication in breeding

Kaweesi Tadeo · Kawooya Ronald · Turyahebwa Vereriano · Tumuhimbise Robooni

Received: 4 April 2023 / Accepted: 17 June 2023
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

Abstract Genetic improvement of tea for both health and economic benefit requires proper genetic characterization of sequence variation present in tea accession especially those conserved in the *ex-situ* genebanks. Here, we studied the genetic variation and population structure of 376 tea germplasm at Rwebitaba Tea Research Centre (RTRC), a tea research station with the largest collections of tea germplasm in East Africa. The study was conducted using 8480 (22.7%) diversity array technology (DArT) SNP markers. These markers had a high call rate of 77% with mean polymorphic information content (P.I.C) of 0.12 and minor allelic frequency of 0.08. The expected heterozygosity (H_e) varied from 0.05 to 0.15 while the observed heterozygosity (H_o) varied from 0.01 to 0.26. The overall inbreeding coefficient (FIS) was 0.45, with some populations showing a high level of outcrossing ($F_{IS} = -0.035$). Analysis of molecular variance revealed a high within-population variance (88.9%) showing weak genetic differentiation between populations

($\Phi_{PT} = 0.111$). Structure analysis showed the presence of eight clusters, with all sub-populations highly admixed except for a few tea clones such as Ch-06 and Ch-07. Phylogenetic analysis confirmed the clustering, showing interrelatedness between different tea genotypes, which will facilitate the selection of parents for increased genetic gain and breeding efficiency in tea breeding. In general, the genetic diversity of tea genotypes is relatively low, with high variation within the existing tea populations. These results showed that the existing variation can be exploited for the development of unique tea genotypes in Uganda.

Keywords Tea · *Camellia sinensis* · Genetic diversity · DArTseq · Population structure · Uganda

Introduction

Tea (*Camellia sinensis* (L.) O. Kuntze) is one of the most consumed beverages in the world due to its favorable taste, aroma, and health benefits (Wang et al. 2022). It is mainly consumed in different forms, either as “non-fermented tea”, “fermented tea” or “semi-fermented tea”, mainly referred to as green, black, and oolong tea respectively. Differences in consumer preferences are due to the unique attributes and benefits each type of tea offers, which are directly dependent on the nature of processing. For black tea, fermentation transforms the polyphenol

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10722-023-01641-5>) contains supplementary material, which is available to authorized users.

K. Tadeo (✉) · K. Ronald · T. Vereriano · T. Robooni
Rwebitaba Zonal Agricultural Research and Development
Institute, P.O Box 96, Kabarole District, Fort portal,
Uganda
e-mail: karln1844@yahoo.com

in tea leaves into theaflavin, thearubigin, and theabrownin, which influences the strength, colour, briskness, aroma, taste, and pungency associated with black tea quality (Rahman et al. 2020). On the other hand, the quality of green tea is influenced by the proportion of theanine (α -glutamylethylamide) relative to the total polyphenol, which influences the mellowness and sweet saltiness (Umani taste) (Kaneko et al. 2006). The health benefits of green tea are majorly associated with catechins (3,3,4,5,7-pentahydroxyflavan) (Sharangi 2009), and these include, offering protection to the cell by scavenging oxygen free radicals, anti-viral, anti-inflammatory, anti-allergic and anti-cancer effects (Shi et al. 2016; Bae et al. 2020). Another form of non-fermented tea is purple tea, a variety of *Camellia sinensis* with a stronger antioxidant capacity (Joshi et al. 2017). It contains a high level of anthocyanins (135-fold) compared to normal green tea (Kerio et al. 2013; Sun et al. 2016). It also contains 1, 2-di-O-galloyl-4, 6-O-(S)-hexahydroxydiphenoyl- β -D-glucose (GHG), a hydrolysable tannin, only found in purple tea (Yagi et al. 2009). Anthocyanins and GHG has been shown to demonstrate excellent anti-obesity, anti-hypertensive, anti-carcinogenic, and anti-ageing effects (Lee et al. 2009; Shimoda et al. 2015; Blesso 2019). To exploit the potential of tea in its improvement, there is a need to characterize the available genetic resource to identify the best tea germplasm for breeding and product development.

Tea was introduced in Uganda by the British Government, through the Royal Botanical Gardens of England and first planted at the Botanical Gardens in Entebbe in 1900. The planting materials were brought from Assam India. By 1914, tea growing reached Toro, Bunyoro and Ankole regions of Uganda. The first successful tea plantations in Uganda were established by the Europeans and Asians in Central region of Uganda in 1924. The first tea exports were made to the United Kingdom in 1945. The increased demand for processed tea on the export market opened opportunities for the Ugandan smallholder farmers to grow tea on out-grower contract arrangements in the late 1950's and early 1960's. By 1955, the Agricultural Enterprises Limited (AEL), a private company owned by the Uganda Development Cooperation (UDC) was formed to develop 6 tea estates and 6 factories namely, Bugambe in Hoima, Muzizi in Kibale,

Mwenge in Kyenjojo, Kiko in Kabarole, Ankole in Bushenyi and Salama in Mukono (MAAIF 2017). As more Ugandan smallholder farmers joined tea growing, a parastatal body called the Uganda Tea Growers Cooperation (UTGC) was established in 1966 with a mandate to encourage associations of tea growers to form cooperative societies which would acquire processing factories. The first smallholder tea factories established by UTGC in 1969 were Igara (Bushenyi), Kayonza (Kanungu), Mabale (Kyenjojo), Mpanga (Kabarole) and other factory sites in Mityana, Buhweju and Kyenjojo, making tea Uganda's third foreign exchange earner after coffee and cotton. However, from 1971 to 1985, the political upheavals that resulted in the expulsion of Asians negatively impacted both the tea industry and tea research in Uganda.

Tea Research and development in Uganda started in the 1960s at Rwebitaba Tea Research Centre (TRC) in Kyenjojo district. Rwebitaba Tea Research Centre was established in 1961 as the one of tea breeding station of the Tea Research Institute of East Africa (TRIEA). Research efforts focused on (i) collection and assembling of tea genotypes at Rwebitaba TRC, (ii) evaluation of yield in different tea genotypes and (iii) determining the effective fertilizer rate for optimal performance of different tea accessions. From these efforts, four tea clones were selected based on their relative yield. These included (i) UTR100/5, (ii) UTR108/82, (iii) UTR6/8 and (iv) UTR31/8. These tea clones were pre-leased and are currently grown as commercial tea varieties in Uganda. The 5th commercial tea variety (UTR303/577) was introduced to Uganda from Kenya as a high-yielding commercial variety. The political upheavals in the 1970s led to an abrupt ending of the breeding activities affecting the tea improvement program in Uganda. Subsequently, in 2003, National Agricultural Research Organization rehabilitated and re-established Rwebitaba Tea Research Station to resume tea research and produce high-quality and high-yielding tea clones. Currently, the institute has a large collection of tea genetic material constituting 148 tea genotypes in the *ex-situ* genebank, 228 tea clones in experimental field plots, and seed barie constituting over 14,005 tea genotypes. The tea genetic resource was left uncharacterized. It is therefore important that tea genotypes at RTRC are characterized genetically and physiochemically to generate profiles that can facilitate proper tea

germplasm conservation, germplasm exploitation for tea genetic improvement, and downstream commercialization in Uganda.

Several studies have used different molecular markers to determine the genetic variability and relationship among tea accessions. Molecular markers such as Random amplified polymorphic DNA (RAPD), Inter-simple sequence repeat (ISSR), Restriction fragment polymorphism (RFLP), Simple sequence repeats (SSR), Amplified fragment length polymorphism (AFLP) and Single nucleotide polymorphism (SNP) have been applied in cultivar identification (Wachira et al. 1995; Matsumoto et al. 2002), genetic diversity (Kaundun and Matsumoto 2002; Fang et al. 2012) and genetic mapping (Malebe et al. 2021). SNP markers offer greater opportunities to study genetic resources, genotype-phenotype association, and application of molecular breeding techniques like genomic selection. Genome-wide SNP discovery has been done using Genotyping-by-sequencing (GBS) (Elshire et al. 2011), RAD-Seq (Baird et al. 2008), and Diversity Array Technology (DArT) (Sansaloni et al. 2011). Of these, DArTseq offers a low-cost genotyping method that scores thousands of unique genome-wide markers without prior sequence information, using an effective complexity reduction method that targets low-copy sequences (active genes in the genome). This study was initiated to determine tea germplasm's structure, genetic and physicochemical diversity at Rwebitaba TRC for effective conservation and utilization in genetic improvement and/or commercialization using mid-density DArTseq markers.

Materials and methods

Plant materials and sample collection

Leaf samples were collected from a total of 376 germplasm at Rwebitaba TRC (0°38'58.5"N 30°25'42.6"E (Location Code 6GGGJCXG+HQ)). Tea germplasm at Rwebitaba TRC is grouped into different populations, which may give a hint on the history of tea improvement at the tea station. Therefore, tea samples were taken from tea genotypes in the *ex-situ* tea germplasm (comprising two sections), experimental fields, chinary tea collection and assembly of purple tea genotypes. A total of 147

tea samples were collected from the tea *ex-situ* tea genebank, with 82 samples taken from the genebank clonal section (Genebank_A) and 66 tea samples taken from the seedling section (Genebank_B). A third batch of samples (164 tea samples) were collected from the tea genotypes established in experimental fields (Rweb_Field_A), while 32 tea samples were collected from field with tea genotypes with Chinary tea characteristics (Rweb_Field_C). The last batch comprising 33 tea samples were taken from the field with tea genotypes with purple leaves (Rweb_Field_P). For each sample, 8 leaf discs were obtained using a leaf-cutting punch and were placed in an appropriate well on 96-well tube storage rack according to the plate design. The leaf samples were dried using silica gel in a desiccator before they were sent to SEQART AFRICA (Nairobi, Kenya) for DArTseq genotyping.

DArTseq genotyping

Tea leaf samples were sent to SEQART AFRICA located at International Livestock Research Institute in Nairobi for Genotyping. DNA extraction was done using a Nucleomag Plant DNA extraction kit (Macherey-Nagel inc, Bethlehem, PA. USA). The genomic DNA extracted was in the range of 50–100ng/ul. DNA quality and quantity were checked on 0.8% agarose. Libraries were constructed according to (Kilian et al. 2012) DArTseq complexity reduction method through digestion of genomic DNA using a combination of PstI and MseI enzymes and ligation of barcoded adapters and common adapter followed by PCR amplification of adapter-ligated fragments. Libraries were sequenced using Single Read sequencing runs for 77 bases. Next-generation sequencing was carried out using Hiseq2500.

SEQART AFRICA uses a genotyping by Sequencing (GBS) DArTseq™ technology, which provides rapid, high quality, and affordable genome profiling, even from the most complex polyploid genomes. DArTseq markers scoring was achieved using DArTsoft14, an in-house marker scoring pipeline based on algorithms. Two types of DArTseq markers were scored: SilicoDArT markers and SNP markers, both scored as binary for presence /absence (1 and 0, respectively) of the restriction fragment with the marker sequence in genomic representation

of the sample. Both SilicoDArT markers and SNP markers were aligned to Tea reference genome, *Teaplant_CSS_Genome* (Xia et al. 2019, 2020) to identify chromosome positions.

Data analysis

Quality analysis of SNP and SilicoDArT markers

The quality of both DArTSeq SNP and SilicoDArT markers was analyzed based on minor allelic frequency, call rate (%), polymorphic information content (PIC) and reproducibility (%) using DArTView software (<https://software.kddart.com/kdxdplore/dartview/>). Genetic richness, diversity and evenness parameters such as observed heterozygosity (H_o), expected heterozygosity (H_e), pairwise differentiation between the populations (F_{ST}), average inbreeding coefficient (F_{IS}), Nei's unbiased gene diversity, Shannon index, evenness and index of association were calculated using DArTR (Gruber et al. 2018) and Poppr package version 2.9.3 (Kamvar et al. 2014) for R software (The R Foundation for Statistical Computing, Vienna: <https://www.R-project.org/>). Population differentiation and gene flow were assessed using analysis of molecular variance (AMOVA) and coefficient of genetic differentiation among population (Φ_{IPT}) using *poppr.amova* and *randtest* functions of *poppr* package respectively.

Population structure of tea germplasm at Rwebitaba TRC

Population structure was assessed using LEA version 3.8.0, an R package for estimating ancestry coefficient and the number of ancestry population (Frichot and François 2015). Inference of admixture coefficient from the genotypic matrix was generated based on sparse non-negative matrix factorization (sNMF) using R function *snmf*. Results from this analysis are closely comparable to those generated using Bayesian clustering programs such as STRU CTURE (Pritchard et al. 2000). Runs of sNMF were performed for values of the number of clusters set to $K=2-10$. The function *snmf* was also used to compute and plot an entropy criterion curve. The minimum cross-entropy value was selected as the possible number of ancestral populations or clusters that best explain the genotypic data (Alexander and

Lange 2011; Frichot et al. 2014). A discriminant analysis of principal components (DAPC) was done to confirm the results using *find.clusters* and *dapc* functions of *DartR* R package version 2.0.4. The lowest Bayesian Information Criterion (BIC) values were used to infer the optimal number of genetic clusters (K) to describe the genetic data. Each tea genotype was assigned to a given cluster based on membership probabilities.

Genetic relationship among tea accessions

To estimate pairwise relatedness, the genetic distances among different tea accession were analyzed based on Prevosti's absolute genetic distance (Prevosti et al. 1975) using *prevosti.dist* function implemented in *poppr* R package. The genetic relationship between the 376 tea accessions was visualized using the Neighbour Joining method, implemented in Molecular Evolutionary Genetics Analysis software (MEGA) version 11 (Tamura et al. 2021). Briefly, the DArT trimmed sequences for each genotype was concatenated using *gl2fasta* function in *dartR* to generate a single composite haplotype in fasta format. The heterogeneous state was resolved by randomly assigning one or another SNP variant to an individual. The composite haplotype fasta file was re-aligned and used in MEGA11 to generate a neighbour-joining tree using default parameters.

Functional diversity of tea germplasm at Rwebitaba TRC

To determine the functional diversity of different tea genotypes at Rwebitaba TRC, physicochemical analysis was conducted to determine the relative quantity of different physicochemical components within each tea genotype. A total of 278 tea genotypes were analyzed for total polyphenol, catechins, theanine, caffeine, and rate of fermentation as described in (Csupor et al. 2014; Nalugo et al. 2022). A ratio of theanine to total polyphenol in the studied tea genotype was used to categorize different tea genotypes into different functional groups. The scale used included > 10 for high-quality green tea genotypes, 5–10 for both

green and black tea, and < 5 for high-quality black tea genotypes.

Results

Analysis of DArTSeq-based SNP and silicodart markers

DArTSeq genotyping of 376 tea germplasm at RTRC yielded a total of 37,391 DArTSeq-based SNP markers. The generated markers had a high call rate ranging from 38 to 100%, with mean call rate of 77%. Polymorphic information content (PIC) of DArTSeq-based SNPs ranged from 0.03 to 0.49 with a mean PIC of 0.12 (Table S1). Minor allelic frequency of the called SNPs ranged between 0.01 and 0.50, with a mean of 0.08 while the heterozygosity ranged from 0 to 0.69, with a mean of 0.06. Filtering of the generated SNP markers with a call rate of more than 90% and reproducibility of 95% resulted in a total of 8480 SNPs markers, which were used in subsequent analysis. For some analyses, such as the estimation of genetic distance, structure, and phylogenetic analysis, all SNPs with callrate and reproducibility above 70% were used to capture the effects of rare alleles in the population.

On the other hand, a total of 25,636 SilicoDart markers were generated from genotyping 147 tea accessions. Their call rate ranged from 81 to 100% with a mean call rate of 98% and polymorphic information content ranging from 0.01 to 0.50, with a mean PIC of 0.07. The SilicoDart markers

generated were very highly reproducible, ranging from 95 to 100%.

Genetic diversity of Ugandan tea germplasm at Rwebitaba TRC

Analysis of the genetic diversity of 376 Ugandan tea genotypes revealed an overall gene diversity of 0.150 and observed heterozygosity of 0.07. To estimate the genetic diversity between the five tea populations studied, three indexes were calculated. These included the Shannon-wiener index (H), Stoddard and Taylor index (G) and Simpson's index (lambda). Based on the three indexes, Rweb_Field A population had the highest genetic diversity with H and G index of 5.10 and 164 respectively. This was followed by Genebank A and Genebank B populations respectively (Table 1). All five populations did not show a significant difference in the Simpson index. Furthermore, all populations showed an even distribution of genetic abundances within their respective populations, with evenness (E.5) of 1, showing that the different tea populations at Rwebitaba TRC had an equal abundance of genotypes.

Both, index of association (Ia) and the standardized index of association (rbarD) were calculated to estimate the mode of reproduction and linkage equilibrium. Based on the index of association, Genebank A population had the highest value of Ia followed by Rweb_Field P and Rweb_Field C with 464.0, 244.4 and 215.3 respectively. Rweb_Field A and Genebank B showed low values of Ia of 50.0 and 97.7 showing that the rate of recombination within these populations is relatively high. Comparing the values of standardized index of association, only

Table 1 Genetic richness, diversity and evenness of Ugandan tea genotypes at Rwebitaba TRC

Population	N	MLG	eMLG	SE	H	G	Lambda	E.5	Hexp	Ia	rbarD
Genebank A	82	82	32	6.24e-06	4.41	82	0.988	1	0.074	464.0	0.146
Rweb_Field_A	164	164	32	0.00e-00	5.10	164	0.994	1	0.084	50.0	0.009
Genebank B	65	65	32	2.93e-06	4.17	65	0.985	1	0.081	97.7	0.020
Rweb_Field_C	32	32	32	0.00e-00	3.47	32	0.969	1	0.179	215.3	0.036
Rweb_Field_P	33	33	32	0.00e-00	3.50	33	0.970	1	0.142	244.4	0.049
Total	376	376	32	2.33e-05	5.93	376	0.997	1	0.103	284.4	0.041

N – Number of individuals observed, MLG – Number of multilocus genotypes, eMLG – the number of MLG at the smallest sample size ≥ 10 based on rarefaction, SE – Standard error based on eMLG, H – Shannon-wiener index of MLG diversity, G – Stoddard and Taylor's index of MLG diversity, Lambda – Simpson's index, E.5 – Evenness, Hexp – Nei's unbiased gene diversity, Ia – Index of association and rbarD – The standardized index of association

genebank A had non-significant values of observed r_{barD} of 0.146 showing that no linkage exists within this population, confirming that these tea genotypes were introduced and established as clones. The other four populations showed significant observed r_{barD} values showing linkage and confirming recombination between the tea genotypes within these populations.

Molecular variance and genetic composition of tea populations at Rwebitaba TRC

Analysis of molecular variance to detect divergence within and among the five tea populations partitioned 11.06% of the variation between population and 88.94% among individuals within population, detected at $P < 0.01$ and low genetic differentiation among tea genotypes population (PhiPT) of 0.111 (Table 2). The total population differentiation due to the proportion of genetic variation that is within the population relative to that between populations (Fixation index, F_{ST}), was 0.149 ($G_{ST} = 0.18$). A pairwise comparison of population differentiation between the studied tea populations at Rwebitaba TRC revealed that the genetic differentiation ranged from 0.009 (Genebank B and Rweb_Field A) to 0.221 (Rweb_Field P and Genebank A). Generally, genetic differentiation of tea genotypes was moderate with the highest recorded between Rweb_Field P and

Genebank A (Table 3). Tea genotypes within Rweb_field P populations had unique alleles exhibited by relatively higher genetic differentiation compared to other populations. It had 19.8% and 18.6% genetic differentiation with Rweb_Field A and Genebank B respectively. On contrary, it had a very low genetic differentiation (0.099) with Rweb_Field C, showing that the two populations shared a large number of similar alleles.

The observed heterozygosity between the studied tea population was low, varying from 0.051 to 0.154. The highest observed heterozygosity was observed in Rweb_Field C and Rweb_Field P with 0.154 and 0.147 respectively. The observed heterozygosity for different tea genotypes at the individual level ranged between 0.015 and 0.260 (Fig. 1). The tea genotype with the highest observed heterozygosity was Ch-15, a Chinari tea genotype. Other tea genotypes with high observed heterozygosity (above 0.200) included Ch-03, Ch-11, L4-04, Ch-04, Ch-14, Ch-09, and Ch-19, all tea genotypes with morphological characteristics of China tea, except one purple tea clone L4-04. The inbreeding coefficient (F_{IS}) was analyzed to determine the extent of inbreeding in the studied tea population. The overall inbreeding coefficient was 0.448. Of the five populations, only one (Rweb_Field_P) indicated a relatively higher level of outcrossing with a negative F_{IS} value of -0.035 (Fig. 2). Genebank A and Genebank B showed a

Table 2 Analysis of molecular variance for genetic differentiation among and within clusters of Ugandan tea genotypes at Rwebitaba TRC

	Degrees of freedom	Sum of square	Mean square	Sigma	(%)	<i>P</i> value	<i>F</i> -statistics
Between population	4	2.63	0.66	0.01	11.06	0.01	$F_{ST} = 0.123$
Within population	371	26.01	0.07	0.07	88.94		$F_{IS} = 0.448$
Total	375	28.64	0.07	0.08	100		

PhiPT – Genetic differentiation among accession populations = 0.111

Table 3 Pairwise F_{ST} values between tea populations at Rwebitaba TRC

	Genebank A	Rweb_Field_A	Genebank B	Rweb_Field_C	Rweb_Field_P
Genebank A	0.000				
Rweb_Field_A	0.092	0.000			
Genebank B	0.064	0.009	0.000		
Rweb_Field_C	0.143	0.139	0.124	0.000	
Rweb_Field_P	0.221	0.198	0.186	0.099	0.000

Fig. 1 Observed heterozygosity of individual tea genotypes from Rwebitaba TRC

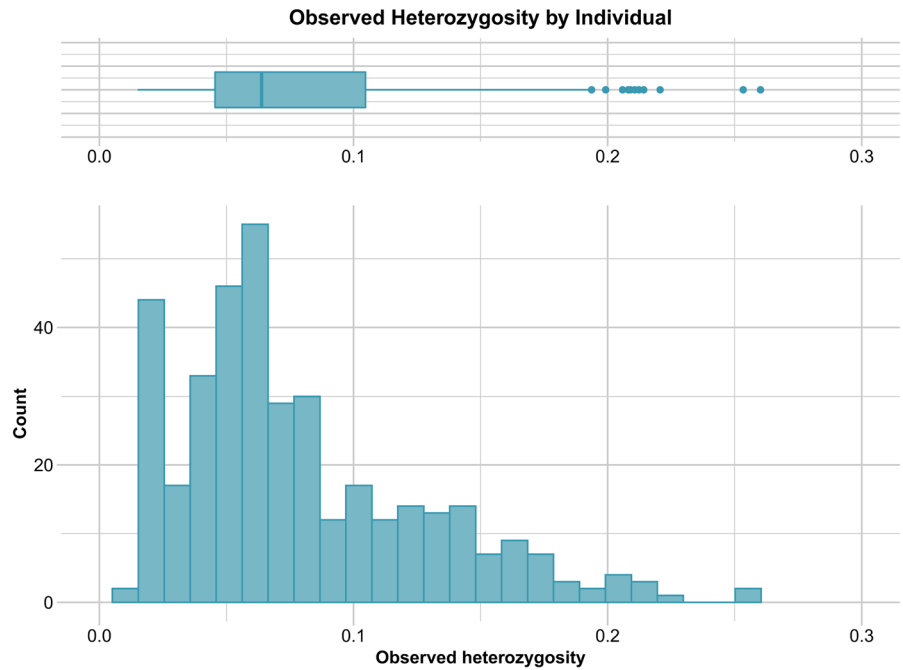
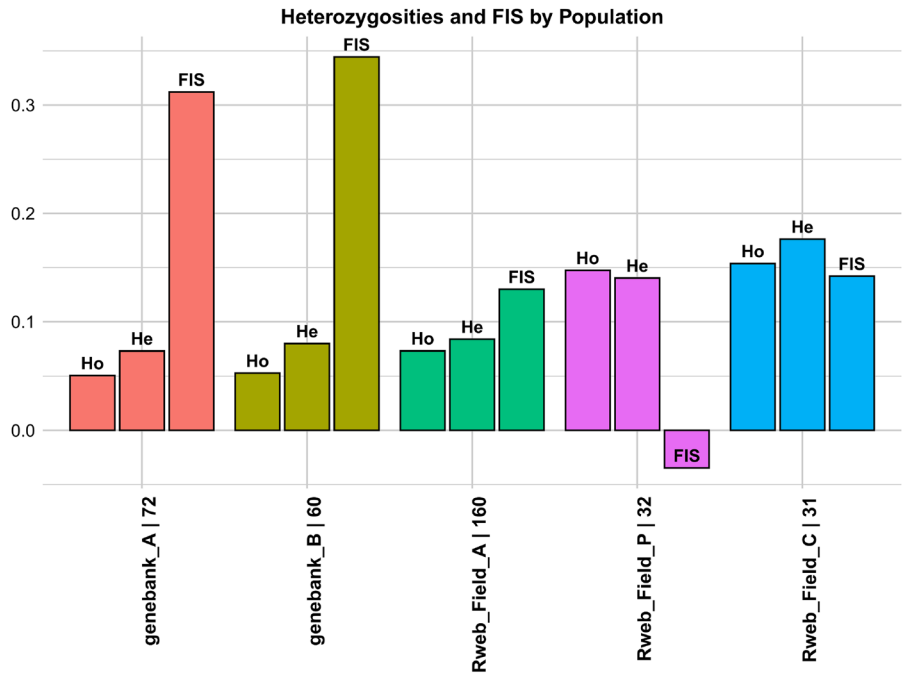


Fig. 2 Heterozygosities (observed and expected heterozygosity) and inbreeding coefficient of different tea populations at Rwebitaba TRC



relatively high level of inbreeding among the tea genotypes within these populations, with F_{IS} values of 0.312 and 0.344 respectively.

Population structure of tea genotypes at Rwebitaba TRC

The population structure of the 376 tea genotypes as determined using Landscape and Ecological

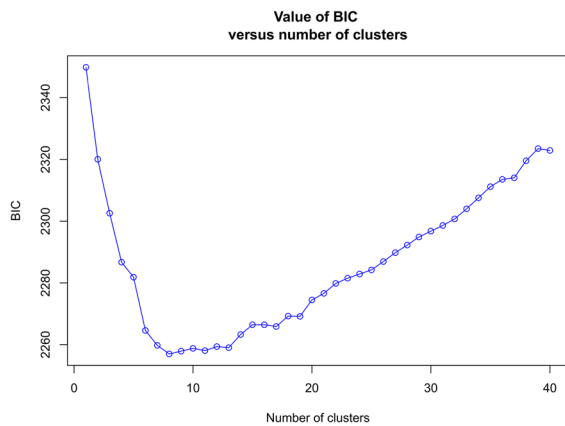


Fig. 3 Number of clusters within the studied tea population as determined by Bayesian Information criterion (BIC). The lowest BIC value relates to the appropriate number of clusters within the population

Association Studies software (LEA) revealed eight ancestral populations based on the entropy criterion and Tracy-Widom distribution (Figure S1 and Figure S2). The number of ancestral populations was confirmed by Discriminant Analysis of Principal Components (DAPC), which revealed a total of eight clusters based on the lowest BIC values (Fig. 3). The revealed clusters had varied numbers of genotypes ranging from 10 to 151 genotypes. Cluster 7 had the highest number of genotypes (151) distributed in all the tested populations. A barplot based on the Q-ancestry matrix (Fig. 4) showed that three of the eight sub-population had a relatively high number of tea genotypes. All tea sub-populations were majorly admixed, except for a few genotypes within particular sub-population such as 246 (Ch-06) and 258 (Ch-07).

The clustering of tea genotypes based on DAPC showed a varying number of tea genotypes per cluster. Both China tea and purple tea genotypes formed unique clusters, cluster 3 and cluster 6

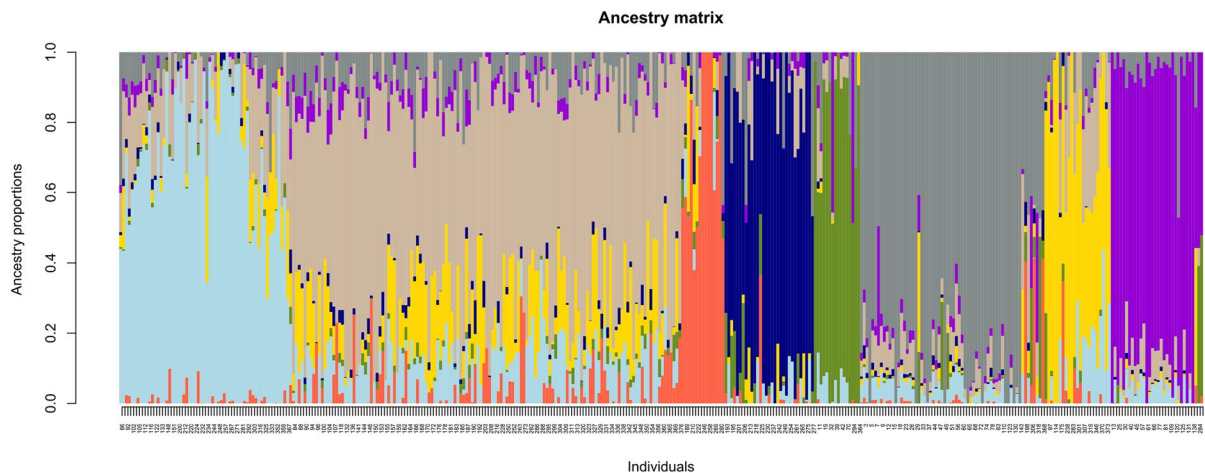


Fig. 4 Structure of tea population at Rwebitaba TRC. The ancestry matrix shows eight sub-population with individuals with high admixture at different ancestry proportions

Table 4 Number of tea genotypes per cluster within the defined tea populations at Rwebitaba TRC

Population	1	2	3	4	5	6	7	8
Genebank A	0	46	0	20	13	0	3	0
Genebank B	18	9	0	8	0	0	24	6
Rweb_Field_A	33	1	0	0	0	0	121	9
Rweb_Field_C	0	2	9	3	3	0	2	13
Rweb_Field_P	0	0	1	0	0	31	1	0
Total	51	58	10	31	16	31	151	28

respectively, with cluster 6 comprising 93.9% of the evaluated purple tea genotypes (Table 4). These two clusters were distinct and did not cluster with any tea accession in either Genebank A or Genebank B. Cluster 1 comprised only tea genotypes from genebank B and Rweb_Field_A while cluster 5 comprised tea genotypes from Genebank_A and Rweb_Field_C (China tea clones, characterized by small leaves and high shoot density). Clusters with tea genotypes from more than three populations include cluster 2, cluster 7, and cluster 8, showing the relatedness and contribution of tea genotypes in Genebank A and Genebank B to the development of tea genotypes in various experimental fields.

To visualize the distribution of these clusters, a DAPC plot was made, which confirmed the distinctness of both cluster 3 and cluster 6 (Fig. 5). Phylogenetic analysis using the Neighbour-joining method clustered all 31 purple tea genotypes (L1-XX to L4-XX, yellow colour) for cluster 6 and the 9 Chinary clones (Ch-03, Ch-01, UTR101, Ch-13, Ch-15, Ch-06, Ch-09, Ch-07 and Ch-14, Blue colour) for cluster 3 (Fig. 6), confirming the clustering obtained using DAPC analysis.

To estimate the amount of variation among the tested tea genotypes, a Principal Coordinate Analysis (PCOA) was conducted. Based on PCOA, the first

two dimensions in the ordination explained only 19.9% of the variation present in the data, with the first two principal coordinate accounting for 12.3% and 7.6% of the variance, respectively (Fig. 7). The pairwise genetic distances based on Prevosti's absolute genetic distance were estimated at the individual level. Pairwise genetic distances among the studied tea genotypes ranged from 0.00 to 0.59. The highest pairwise genetic distances were mainly recorded among Chinary tea clones such as Ch-11, Ch-15, UTR101, GB_S54, GB_S61, Ch-06, Ch-04, Ch-27, Ch-19 and Ch-24. Other tea genotypes with high pairwise genetic distances include GB_C17, GB_C45, GB_C40, GB_C44, GB_C58, GB_C68 and GB_C82 from Genebank A and a few tea clones such as UTR108/66, UTR103/2, UTR115/46, UTR1/25, UTR30/18, UTR141/191, UTR6/95, UTR141/189 and UTR1/14 from Rweb_Field_A (Table S2). These are potential genetic stocks for producing desirable genetic recombination and highly segregating tea populations for Uganda's high-quality and high-yielding tea genotypes.

Fig. 5 Discriminant analysis of principal components (DAPC) plot showing the distribution of different clusters of tea genotypes at Rwebitaba TRC

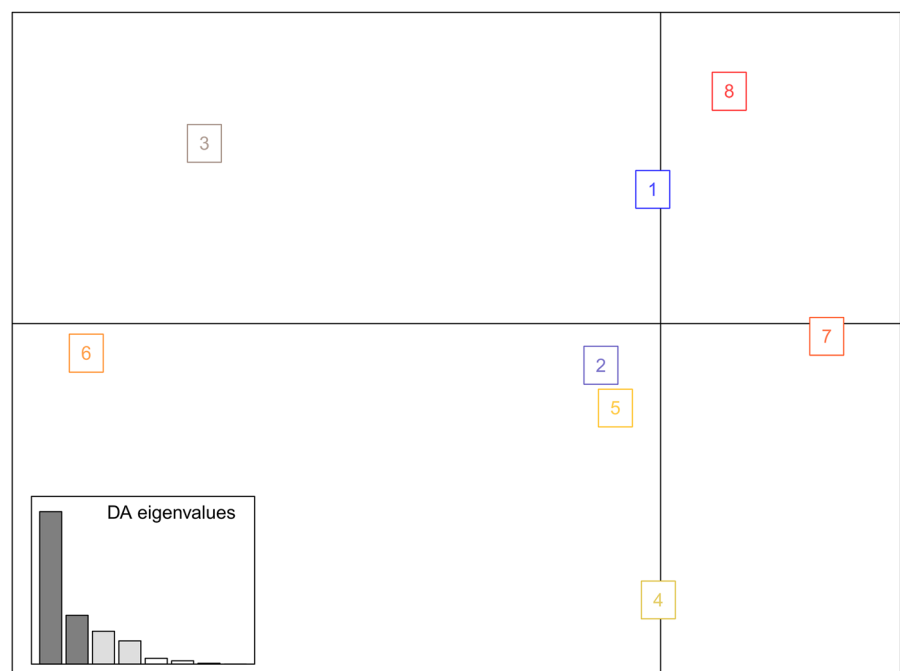


Fig. 6 Phylogenetic analysis based on Neighbour-joining methods showing genetic relatedness of different tea genotypes at Rwebitaba TRC. Different colours represent the different clusters that form the population structure of tea at Rwebitaba TRC. These include Cluster 1 (magenta-pink), Cluster 2 (green), Cluster 3 (blue, majorly china type with small leaves and high shoot density), Cluster 4 (lily white), Cluster 5 (dark gray, china type/hybrids with moderately sized leaves), Cluster 6 (yellow, purple tea genotypes with purple leaves and high anthocyanin level), Cluster 7 (light violet, assam type) and Cluster 8 (Maverick, purplish)

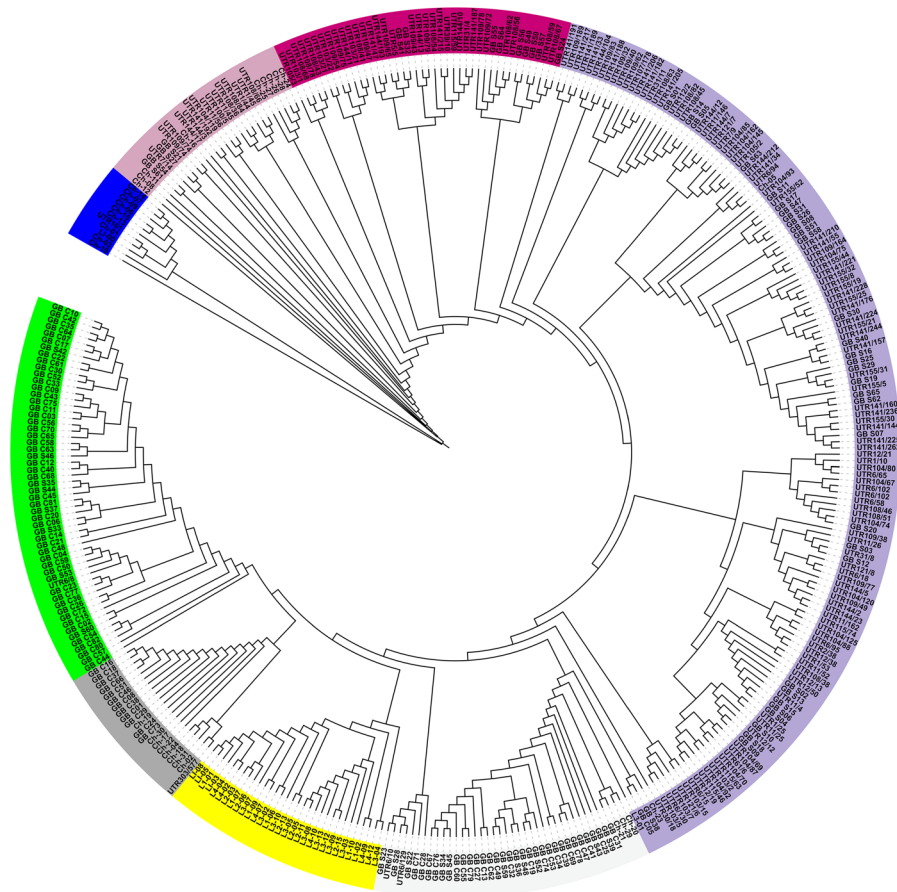


Fig. 7 Principal Coordinate Analysis showing the amount of variation within the tea populations from Rwebitaba TRC. Populations of tea genotypes at Rwebitaba were distinguished by different colours. Genotypes in Genebank_A showed in bright orange, Genebank_B in gold, Rweb_Field_A in light red, Rweb_Field_C in Jordy blue, and Rweb_Field_P magenta pink

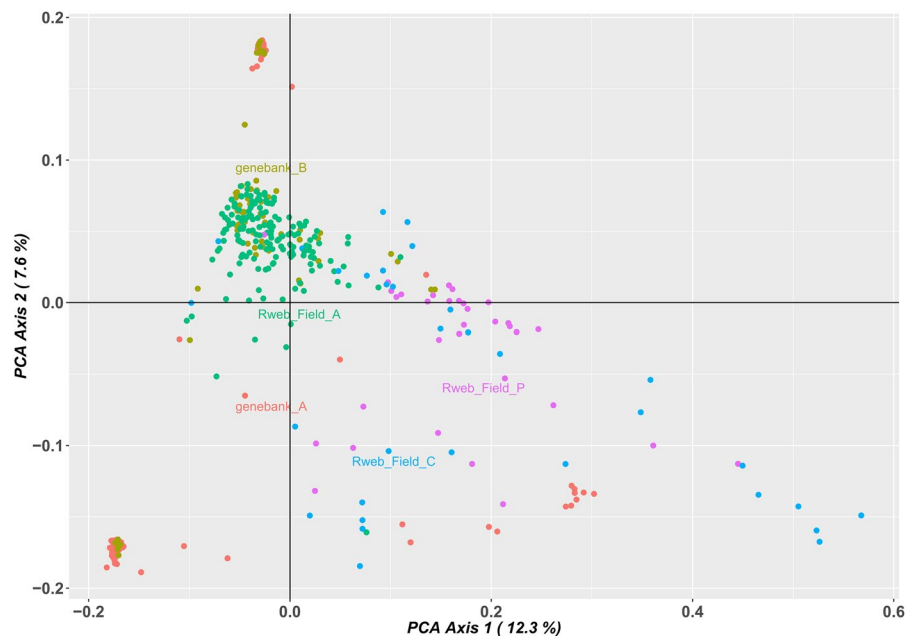


Table 5 Number of tea clones per functional category

Ratio	Number of tea genotypes	Category
> 10	2	Green tea
5–10	90	Both green and black tea
< 5	186	Black tea

Functional diversity of tea genotypes at Rwebitaba TRC

A total of 278 tea genotypes, mostly genebank accession were characterized based on the ratio of total theanine and total polyphenol to identify the best tea genotypes for black tea and green tea production. Analysis revealed a variation in both the amount of theanine and total polyphenol among the screened tea genotypes. The amount of theanine ranged from 1.57 to 856.52 mg/ml while total polyphenol ranged from 52.50 to 244.39 mg/g FW. Characterization based on the ratio of theanine to total polyphenol revealed only two tea genotypes with a ratio greater than 10, while 90 tea genotypes had a ratio ranging from 5 to 10 and 186 tea genotypes had a ratio lower than 5 (Table 5). Based on these ratios, two tea genotypes, GB_C20 and GB_C21 are the best tea genotypes for green tea production while 186 tea genotypes are good for black tea production. The high number of tea genotypes characterized as black tea genotypes shows that previous breeding efforts focused on the selection of high-quality black tea genotypes. A total of 8 tea genotypes with the lowest ratio among the 186 black tea genotypes were selected as the best black

tea genotypes. Among the 8 selected tea genotypes, three (UTR303/577, 31/8, and UTR6/8) are high-quality black tea genotypes, currently used as the best commercial black tea genotypes in Uganda (Table 6).

In addition to the selection index (ratio of theanine to total polyphenol), other parameters such as the amount of theanine, polyphenol, catechins, caffeine, and rate of fermentation were considered. A total of 30 tea genotypes were selected from each category (low selection index with high catechin content for black tea and high selection index with high theanine content for green tea) and clustered using hierarchical cluster analysis to form a cluster dendrogram (Fig. 8). The cluster dendrogram revealed two major clusters which, majorly show the best tea genotypes for black and green tea production. A high number of tea genotypes clustered as black tea are tea genotypes from Rweb_Field_A while many genotypes characterized as green tea genotypes are from Genebank A (Fig. 8). Among the tea genotypes clustered as black tea, four Genebank_B accessions GB_S01, GB_S03, GB_S04 and GB_S06 outperformed all the selected black tea genotypes, therefore can be selected as genetic stocks or candidate tea clones for multi-location evaluation and release as high-quality black tea genotypes. These results show that breeding efforts at Rwebitaba TRC majorly focused on the selection of black tea genotypes while the green tea accessions in Genebank A have not yet been exploited.

Table 6 Tea genotypes selected per functional category

Category	List of selected tea genotypes
Green tea	GB_C20, GB_C21
Dual purpose but best for green tea	GB_C22, GB_S35, GB_S30, GB_C23, GB_C26, UTR141/183, GB_C19
Dual purpose but best for black tea	UTR104/85, GB_S61, GB_S60, GB_S37, UTR12/12, GB_S18, GB_S44, UTR1/44
Black tea	UTR141/192, UTR144/3, UTR141/4, UTR6/8, UTR303/577, UTR31/8, UTR144/13, UTR31/11

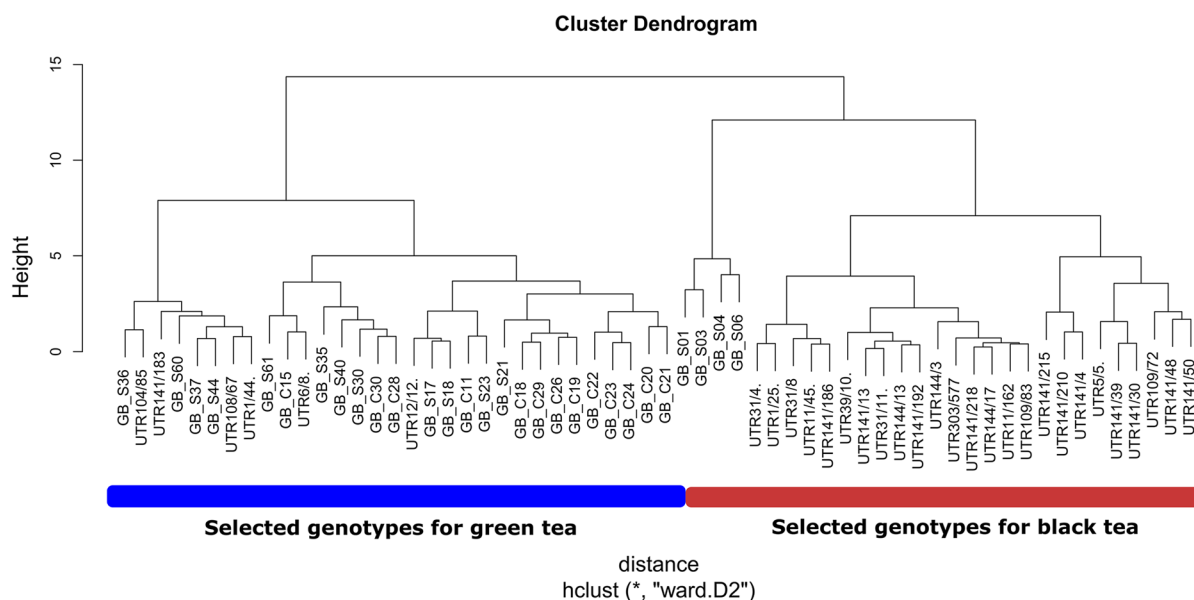


Fig. 8 Functional clustering of tea genotypes based the five parameters; amount of theanine, polyphenol, catechins, caffeine, and rate of fermentation. Based on these parameters,

theselected tea genotypes were clusters into two functional categories, high quality black or high quality green tea

Discussion

Understanding the genetic diversity of germplasm is critical for the proper utilization and conservation of germplasm (Fang et al. 2014; Lopes et al. 2015). This study was initiated to genetically characterize tea germplasm at Rwebitaba Tea Research Centre. As one of the sub-stations of TRIEA, many breeding activities such as the introduction of tea genotypes, evaluation, and selection (Seurei 1997), were carried out until the late 1970s, leaving a large resource of uncharacterized germplasm. In order to utilize this genetic resource for tea improvement in Uganda, DArTseq SNP markers were used to assess genetic diversity, population structure, and hierarchical clustering of 376 tea genotypes. Results from this study revealed that the generated SNP markers were highly reproducible and moderately informative based on the polymorphic information content (Botstein et al. 1980). We used only SNP markers in this study because dominant markers such as SilicoDArT add little information to population studies (Gapare et al. 2021). The overall genetic diversity (0.15) and population differentiation (G_{ST} , 0.18) of tea genotypes at RTRC was low, consistent with the RAPD and ISSR analysis of the Kenyan tea cultivars (Wachira

et al. 2001; Yao et al. 2008). In addition, the Analysis of Molecular Variance also confirmed that the highest variation (88.94%) was within populations than between populations (11.06%), showing the impact of directional selection and focus on particular traits during tea breeding in the East African region.

Tea breeding efforts in East Africa have focused on the generation of high-quality and high-yielding tea genotypes for black tea processing. As a result, the characterization of tea genotypes at Rwebitaba TRC revealed that many tea genotypes were black tea genotypes (Table 6). In addition, most tea genotypes characterized as the best black tea genotypes (UTRXXX) as revealed in Fig. 8 belonged to Rweb_Field A, while a majority of the green tea genotypes belonged to Genebank A. These insights show that limited work has been done in the breeding or selection of green tea genotypes for specialty tea production. In the early stages of tea breeding in TRIEA, crosses were majorly done using *C. assamica* tea clone 6/8 as a parent (Seurei 1996). In this study, this clone has been characterized as a high-quality black tea clone (Table 6) and clustered with Genebank A genotypes (GB_CXX, Fig. 6), showing that the majority of the tea genotypes in Genebank A belong to *C. assamica* group. Another commercial

tea variety, UTR 303/577 clustered with Chinary tea genotypes such as Ch-02, Ch-31, Ch-18 and Ch-04 showing that it is a chinary tea genotype. However, the morphological characteristics of UTR303/577 are different from the typical China tea clone UTR101, which has small leaves and high shoot density. Therefore, it can be concluded that UTR303/577 is a hybrid, belonging to a unique cluster distinct to that of china tea UTR101. On the other hand, purple clone L1-00 clustered with china tea clone UTR101, away from other purple genotypes (LX-XX, yellow cluster, Fig. 6) showing that it could also be a hybrid combining the attributes of china tea and high anthocyanin level making it a good tea clone for production of nutraceuticals or healthy products. The other purple tea genotypes (LX-XX) formed a unique cluster suggesting that (i) they have a common origin and (ii) they have not been exploited as parents in tea breeding in Uganda.

Structure analysis using both DAPC and LEA revealed eight populations, with high levels of genetic admixture. Earlier studies of Kenya tea germplasm using nuclear microsatellites revealed three population conforming to the tea taxon affinity; Assam tea (*Camellia sinensis* var *assamica*), China tea (*Camellia sinensis* var *sinensis*), and Cambod tea (*Camellia assamica* subsp *lasiocalyx*) (Wambulwa et al. 2016). Our study did not conform to this taxon affinity. This could be due to the differences in the markers used in assessing the population structure. Recently, another study conducted using genome-wide markers (DArTseq) revealed five populations, which combined tea type (Cambod and China type) and origin of tea genotypes (TRFCA or Tea Research Institute) (Malebe et al. 2019). The eight populations revealed in this study may (i) relate to the multiple origins of East Africa's tea breeding stocks; China, Japan, India and other locations (Wambulwa et al. 2016), and (ii) hybrid nature of tea genotypes arising from crossing different tea types contributing to the high genetic admixture within the studied tea germplasm (Yang et al. 2016).

The field layout of tea trials at RTRC shows the history of tea improvement activities that were conducted at the station. Tea genotypes distributed within the experimental field (Rweb_Field A) had the highest genetic diversity based on both the Shannon-wiener index and Stoddard and Taylor index. The highest number of tea genotypes in Rweb_Field

A belonged to cluster 7 and cluster 1, with high association or relatedness to tea genotypes in Genebank B and limited association or relatedness to tea genotypes in Genebank A (Table 4). These results show that tea genotypes in Genebank A may not be the progenitors of tea genotypes in experimental fields (Rweb_Field A). This also means that tea genotypes in the experimental field were most probably introduced as seeds in different families, evaluated and a few selected tea genotypes were planted in the *ex-situ* genebank section B (Genebank B). Similarly, Wachira et al. (1995) reported that Kenyan tea introductions were based on the progeny from crosses undertaken in one region (Assam). Furthermore, phylogenetic analysis (Fig. 6, light violet cluster) clustered Rweb_Field A genotypes (UTRXXX) more with Genebank B genotypes (GB_SXXX) than Genebank A genotypes (GB_CXXX). These results show that tea improvement at Rwebitaba TRC between 1960 and late 1970s concentrated more on introduction and evaluation and less on controlled hybridization.

To investigate this, we used both index of association (I_a) and standardized index of association ($rbarD$) (Table 1) to estimate the mode of reproduction and linkage equilibrium. Only Genebank A had non-significant values of observed $rbarD$ showing that no linkage exists within this population, confirming that tea genotypes within Genebank A were introduced and established as clones. The high value of the index of association of genotypes within Genebank A shows a very low recombination rate, confirming that limited hybridization has been conducted to generate Genebank A tea genotypes. These results show that the tea genotypes in the *ex-situ* genebank have not yet been fully utilized or exploited. Other populations with low rate of recombination include Rweb_Field P and Rweb_Field C, comprising of Chinary and purple tea genotypes and forming distinct clusters (cluster 3 and 6 respectively, Table 4). The low recombination among these populations contributes to the low genetic diversity observed at Rwebitaba TRC. In addition, the low genetic diversity could be attributed to a high rate of clonal propagation. Balloux et al. (2003) reported that the high rate of clonal propagation, especially in vegetatively propagated crops increases heterozygosity and variation within a population while reducing genetic diversity and genetic differentiation among populations. Therefore,

the low genetic diversity in Ugandan tea germplasm necessitates deliberate efforts towards broadening the Ugandan tea germplasm base by collecting and conserving new tea individuals and/or hybridization for successful future breeding initiatives of tea in Uganda.

Furthermore, the closer the observed heterozygosity to the expected heterozygosity, the higher the genetic diversity of the population (Hansson and Westerberg 2002; Slate et al. 2004). This was observed in two populations (Rweb_Field A and Rweb_Field P, Fig. 2) showing that tea germplasm within these populations had relatively high genetic richness. In addition, Rweb_Field P and Rweb_Field C had high levels of observed heterozygosity, genetic differentiation (F_{ST}) and genetic distance (Table 5). The uniqueness of these genotypes show that they have unique alleles that (i) can be exploited in controlled hybridization program for heterosis and (ii) can contribute to the diversification of tea products (green tea, purple tea and other health products).

Conclusion

Based on these results, the amount of genetic diversity within the tea germplasm at Rwebitaba TRC is low. The variation within the existing populations such as Genebank A, Rweb_Field_P and Rweb_Field_C has not been well exploited. The presence of genetically diverse tea genotypes especially the China tea type offers an opportunity to produce highly segregating progenies with high heterosis and desirable gene recombination. This, in addition to the introduction of new tea genotypes, will maximize genetic diversity, and increase breeding efficiency and genetic gain for increased productivity and profitability from tea production.

Acknowledgements The authors thank the Government of Uganda and Rwebitaba ZARDI for the financial through CGS grant No. CCGS/05/08/18, MoSTI/NRIP/2020–2021/79 and organizational support, respectively.

Author contributions KT: conceptualization, design, data analysis, interpretation, and manuscript writing and review; RT: conceptualization and manuscript review; KR & TV: data collection & analysis, manuscript review. All authors read and approved the final manuscript.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

References

- Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:1–6. <https://doi.org/10.1186/1471-2105-12-246/FIGURES/3>
- Bae J, Kim N, Shin Y, Kim S-Y, Kim Y-J (2020) Activity of catechins and their applications. Biomed Dermatol 4:1–10. <https://doi.org/10.1186/S41702-020-0057-8>
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA et al (2008) Rapid SNP Discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:e3376 <https://doi.org/10.1371/JOURNAL.PONE.0003376>
- Balloux F, Lehmann L (2003) de Meeûs T (2003) The Population Genetics of Clonal and Partially Clonal Diploids. Genetics 164:1635–1644. <https://doi.org/10.1093/genetics/164.4.1635>
- Blesso CN (2019) Dietary anthocyanins and Human Health. Nutrients. <https://doi.org/10.3390/NU11092107>
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314
- Csupor D, Boros K, Jedlinszki N, Hohmann J (2014) A validated RP-HPLC-DAD method for the determination of l-Theanine in tea. Food Anal Methods 7:591–596. <https://doi.org/10.1007/S12161-013-9659-8/METRICS>
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES et al (2011) A robust, simple genotyping-by-sequencing (GBS) Approach for High Diversity Species. PLoS ONE 6:e19379. <https://doi.org/10.1371/JOURNAL.PONE.0019379>
- Fang W, Cheng H, Duan Y, Jiang X, Li X (2012) Genetic diversity and relationship of clonal tea (*Camellia sinensis*) cultivars in China as revealed by SSR markers. Plant Syst Evol 298:469–483. <https://doi.org/10.1007/S00606-011-0559-3/TABLES/4>
- Fang WP, Meinhardt LW, Tan HW, Zhou L, Mischke S, Zhang D (2014) Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP) markers. Hortic Res 2014 1(1 1):1–8. <https://doi.org/10.1038/hortres.2014.35>
- Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929. <https://doi.org/10.1111/2041-210X.12382>
- Frichot E, Mathieu F, Trouillon T, Bouchard G, François O (2014) Fast and efficient estimation of individual ancestry

- coefficients. *Genetics* 196:973–983. <https://doi.org/10.1534/GENETICS.113.160572>
- Gapare WJ, Kilian A, Stewart AV, Smith KF, Culvenor RA (2021) Genetic diversity among wild and cultivated germplasm of the perennial pasture grass *Phalaris aquatica*, using DArTseq SNP marker analysis. *Crop Pasture Sci* 72(10):823–840. <https://doi.org/10.1071/CP21112>
- Gruber B, Unmack PJ, Berry OF, Georges A (2018) Dartr: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour* 18:691–699. <https://doi.org/10.1111/1755-0998.12745>
- Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. *Mol Ecol* 11:2467–2474. <https://doi.org/10.1046/J.1365-294X.2002.01644.X>
- Joshi R, Rana A, Kumar V, Kumar D, Padwad YS, Yadav SK et al (2017) Anthocyanins enriched purple tea exhibits antioxidant, immunostimulatory and anticancer activities. *J Food Sci Technol* 54:1953. <https://doi.org/10.1007/S13197-017-2631-7>
- Kamvar ZN, Tabima JF, Grünwald NJ, (2014) Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *Peer J*. <https://doi.org/10.7717/PEERJ.281/TABLE-6>
- Kaneko S, Kumazawa K, Masuda H, Henze A, Hofmann T (2006) Molecular and sensory studies on the umami taste of Japanese green tea. *J Agric Food Chem* 54:2688–2694. <https://doi.org/10.1021/JF0525232>
- Kaundun SS, Matsumoto S (2002) Heterologous nuclear and chloroplast microsatellite amplification and variation in tea, *Camellia sinensis*. *Genome* 45:1041–1048. <https://doi.org/10.1139/G02-070>
- Kerio LC, Wachira FN, Wanyoko JK, Rotich MK (2013) Total polyphenols, catechin profiles and antioxidant activity of tea products from purple leaf coloured tea cultivars. *Food Chem* 136:1405–1413. <https://doi.org/10.1016/J.FOODC.2012.09.066>
- Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H et al (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol* 888:67–89. https://doi.org/10.1007/978-1-61779-870-2_5
- Lee SH, Park SM, Park JH, Shin DY, Kim GY, Ryu CH et al (2009) Induction of apoptosis in human leukemia U937 cells by anthocyanins through down-regulation of Bcl-2 and activation of caspases. *Int J Oncol* 34:1077–1083. <https://doi.org/10.3892/IJO-00000234>
- Lopes MS, El-Basyoni I, Baenziger PS, Singh S, Royo C, Ozbek K et al (2015) Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J Exp Bot* 66:3477–3486. <https://doi.org/10.1093/JXB/ERV122>
- Malebe M, Mphangwe N, Myburg A, Apostolides Z (2019) Assessment of genome-wide DArT-seq markers for tea *Camellia sinensis* (L.) O. Kuntze germplasm analysis. *Tree Genet Genomes* 15:1–9. <https://doi.org/10.1007/S11295-019-1358-1/METRICS>
- Malebe MP, Koech RK, Mbanjo EGN, Kamunya SM, Myburg AA, Apostolides Z (2021) Construction of a DArT-seq marker-based genetic linkage map and identification of QTLs for yield in tea (*Camellia sinensis* (L.) O. Kuntze). *Tree Genet Genomes* 17:1–11. <https://doi.org/10.1007/S11295-021-01491-1/METRICS>
- Matsumoto S, Kiriiwa Y, Takeda Y (2002) Differentiation of Japanese green tea cultivars as revealed by RFLP analysis of phenylalanine ammonia-lyase DNA. *Theor Appl Genet* 104:998–1002. <https://doi.org/10.1007/S00122-001-0806-Z>
- Nalugo RG, Kaweesi T, Kawooya R, Nuwamanya E, Mugisa C, Namutebi V et al (2022) Physicochemical analysis of Ugandan tea (*Camellia sinensis*) germplasm reveals potential commercial green and black tea varieties. *J Crop Improv*. <https://doi.org/10.1080/15427528.2022.2095317>
- Prevosti A, Ocaña J, Alonso G (1975) Distances between populations of *Drosophila subobscura*, based on chromosome arrangement frequencies. *Theor Appl Genet* 45:231–241. <https://doi.org/10.1007/BF00831894>
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959. <https://doi.org/10.1093/GENETICS/155.2.945>
- Rahman MM, Hossain MM, Das R, Ahmad I (2020) Changes in phytochemicals and determination of optimum fermentation time during black tea manufacturing. *J Sci Res* 12:657–664. <https://doi.org/10.3329/JSR.V12I4.45452>
- Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D et al (2011) Diversity arrays technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. *BMC Proceedings* 5:1–2. <https://doi.org/10.1186/1753-6561-5-S7-P54>
- Seurei P (1996) Tea improvement in Kenya: a review. *Tea Board of Kenya* 17:76–90
- Seurei P (1997) Tea improvement in Kenya: progress, constraints and perspectives. *Tea - Tea Board of Kenya (Kenya)* 18:87–100. 10.3/JQUERY-UIJS
- Sharangi AB (2009) Medicinal and therapeutic potentialities of tea (*Camellia sinensis* L.) – A review. *Food Res Int* 42:529–535. <https://doi.org/10.1016/J.FOODRES.2009.01.007>
- Shi M, Nie Y, Zheng XQ, Lu JL, Liang YR, Ye JH (2016) Ultraviolet B (UVB) photosensitivities of tea catechins and the relevant chemical conversions. *Molecules*. <https://doi.org/10.3390/MOLECULES21101345>
- Shimoda H, Hito S, Nakamura S, Matsuda H (2015) Purple tea and its extract suppress diet-induced fat accumulation in mice and human subjects by inhibiting fat absorption and enhancing hepatic carnitine palmitoyltransferase expression. *Int J Biomed Sci* 11:67
- Slate J, David P, Dodds KG, Veenvliet BA, Glass BC, Broad TE et al (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* 93:255–265. <https://doi.org/10.1038/sj.hdy.6800485>
- Sun B, Zhu Z, Cao P, Chen H, Chen C, Zhou X et al (2016) Purple foliage coloration in tea (*Camellia sinensis* L.) arises from activation of the R2R3-MYB transcription factor CsAN1. *Sci Rep* 6:1–15. <https://doi.org/10.1038/srep32534>

- Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* 38:3022–3027. <https://doi.org/10.1093/MOLBEV/MSAB120>
- Wachira FN, Waugh R, Hackett CA, Powell W (1995) Detection of genetic diversity in tea (*Camellia sinensis*) using RAPD markers. *Genome* 38:201–210. <https://doi.org/10.1139/G95-025>
- Wachira F, Tanaka J, Takeda Y (2001) Genetic variation and differentiation in tea (*Camellia sinensis*) germplasm revealed by RAPD and AFLP variation. *J Hort Sci Biotechnol* 76:557–563
- Wambulwa MC, Meegahakumbura MK, Chalo R, Kamunya S, Muchugi A, Xu JC et al (2016) Nuclear microsatellites reveal the genetic architecture and breeding history of tea germplasm of East Africa. *Tree Genet Genomes* 12:1–10. <https://doi.org/10.1007/S11295-015-0963-X/METRICS>
- Wang C, Han J, Pu Y, Wang X (2022) Tea (*Camellia sinensis*): a review of nutritional composition, potential applications, and Omics Research. *Appl Sci* 2022 12(12):5874. <https://doi.org/10.3390/AP12125874>
- Xia EH, Li FD, Tong W, Li PH, Wu Q, Zhao HJ et al (2019) Tea Plant Information Archive: a comprehensive genomics and bioinformatics platform for tea plant. *Plant Biotechnol J* 17:1938–1953. <https://doi.org/10.1111/PBI.13111>
- Xia E, Tong W, Hou Y, An Y, Chen L, Wu Q et al (2020) The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol Plant* 13:1013–1026. <https://doi.org/10.1016/J.MOLP.2020.04.010>
- Yagi K, Goto K, Nanjo F (2009) Identification of a major polyphenol and polyphenolic composition in leaves of *Camellia irrawadiensis*. *Chem Pharm Bull (Tokyo)* 57:1284–1288. <https://doi.org/10.1248/CPB.57.1284>
- Yang H, Wei CL, Liu HW, Wu JL, Li ZG, Zhang L et al (2016) Genetic divergence between *Camellia sinensis* and its wild relatives revealed via genome-wide SNPs from RAD sequencing. *PLoS ONE* 11:e0151424. <https://doi.org/10.1371/JOURNAL.PONE.0151424>
- Yao MZ, Chen L, Liang YR (2008) Genetic diversity among tea cultivars from China, Japan and Kenya revealed by ISSR markers and its implication for parental selection in tea breeding programmes. *Plant Breeding* 127:166–172. <https://doi.org/10.1111/J.1439-0523.2007.01448.X>
- MAAIF (2017) Ministry of Agriculture, Animal Industry and Fisheries: *Performance Review Report Financial year 2016/2017*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.