


Article

Utilization of Phytochemical and Molecular Diversity to Develop a Target-Oriented Core Collection in Tea Germplasm

Do Yoon Hyun ¹, Gwang-Yeon Gi ², Raveendar Sebastin ¹, Gyu-Taek Cho ¹, Seong-Hoon Kim ¹, Eunae Yoo ¹, Sookyeong Lee ¹, Dong-Mo Son ² and Kyung Jun Lee ^{1,*} 

¹ National Agrobiodiversity Center, National Institute of Agricultural Sciences (NAS), RDA, Jeonju, Jeollabuk-do 54874, Korea; dyhyun@korea.kr (D.Y.H.); raveendars@gmail.com (R.S.); gtcho@korea.kr (G.-T.C.); shkim0819@korea.kr (S.-H.K.); eung77@korea.kr (E.Y.); xsanta7@korea.kr (S.L.)

² Tea Industry Institute, Jeollanamdo Agricultural Research and Extension Service, Boseong-gun, Jeollanam-do 59455, Korea; gygi4975@korea.kr (G.-Y.G.); dmson0810@korea.kr (D.-M.S.)

* Correspondence: lkj5214@korea.kr; Tel.: +82-63-238-4861

Received: 10 October 2020; Accepted: 26 October 2020; Published: 29 October 2020



Abstract: Tea has received attention due to its phytochemicals. For the direct use of tea germplasm in breeding programs, a core collection that retains the genetic diversity and various phytochemicals in tea is needed. In this study, we evaluated the content of eight phytochemicals over two years and the genetic diversity through 33 SSR (simple sequence repeats) markers for 462 tea accessions (entire collection, ENC) and developed a target-oriented core collection (TOCC). Significant phytochemical variation was observed in the ENC between genotypes and years. The genetic diversity of ENC showed high levels of molecular variability. These results were incorporated into developing TOCCs. The TOCC showed a representation of the ENC, where the mean difference percentage, the variance difference percentage, the variable rate of coefficient of variance percentage, and the coincidence rate of range percentage were 7.88, 39.33, 120.79, and 97.43, respectively. The Shannon's diversity index (I) and Nei's gene diversity (H) of TOCC were higher than those of ENC. Furthermore, the accessions in TOCC were shown to be selected proportionally, thus accurately reflecting the distribution of the overall accessions for each phytochemical. This is the first report describing the development of a TOCC retaining the diversity of phytochemicals in tea germplasm. This TOCC will facilitate the identification of the genetic determinants of trait variability and the effective utilization of phytochemical diversity in crop improvement programs.

Keywords: catechin; genetic diversity; phytochemicals; SSR; targeted-oriented core collection; tea germplasm

1. Introduction

Since the International Board for Plant Genetic Resources (IBPGR) was established in 1974 to coordinate the global efforts to systematically collect and conserve the world's threatened genetic plant diversity, many countries and organizations have founded gene banks, and millions of crop resources have been preserved [1,2]. As a result of the global efforts to conserve plant genetic resources for food and agriculture, the number and scale of ex situ germplasm collection has increased tremendously in the last 40 years [3]. However, the large sizes of redundant collections, either individually or collectively, for particular species have become an obstacle to the characterization, evaluation, utilization, and maintenance of those species [2,3]. As part of this solution, the authors in [4] proposed that collections could be pruned to core collections, which could "represent with a minimum of repetitiveness, the genetic diversity of a crop species and its relatives." This core collection

serves as a working collection that can be extensively examined, while the accessions excluded from such collections can be preserved as preliminary collections [5]. Therefore, core collections can facilitate the use of crop germplasm and manage the entire collections [2].

Tea (*Camellia sinensis* (L.) Kuntze) is a woody evergreen plant in the family Theaceae and is native to the region covering the northern part of Myanmar, as well as the provinces of Yunnan and Sichuan in China. It is one of the most popular beverages and has become a daily drink for many people around the world [6]. To combat climate change, biological threats, and market fluctuations, the main tea-producing countries of China, Sri Lanka, and India have managed and preserved their tea and genetic resources both in situ and ex situ [7]. In addition, they have developed core collections or subsets of tea germplasm that maintain the original diversity of the collections but at a size that facilitates the evaluation, use, and conservation of the entire collections using geographical origin, phenotypic traits, and molecular markers [8–12]. Knowledge and understanding of the genetic background, genetic diversity, relationships, and identification are important for the collection, preservation, characterization, and utilization of tea resources [13]. The proper characterization and evaluation of genetic resources via systematic preservation and maintenance is the most important factor in utilizing such resources for improving crops [14]. The characterization of germplasm can be carried out using morphological, biochemical, and molecular descriptors according to the standard criteria contained in the tea descriptors [15]. Among the characteristics in tea descriptors, morphological traits and phytochemical content tend to be most affected by environmental factors. In addition, since phytochemical content can have very large variations depending on the environment, these characteristics need to be evaluated for multiple years, yielding more precise data. On the other hand, molecular markers are rarely influenced by the environment and thus directly offer an observation of genomic diversity.

The phytochemical characterization of plant germplasm is an acceptable method to define biochemical diversity [16]. The composition of phytochemicals in tea is important, as these chemicals contribute to tea's quality and pharmacological properties [15]. Tea consists of compounds rich in polyphenols, theanine, and caffeine, which not only determine the quality of tea but also provide tremendous health benefits [17]. Among tea polyphenols, catechins account for 8% to 26% of the tea leaves' dry weight [18]. Previous studies reported that because each catechin monomer has a different chemical structure they each have unique bioactivity, bioavailability, and physiological pharmacokinetic properties [19,20]. In addition, the origin and growing conditions of the tea plant affect the contents of the tea's phytochemicals, which changes bioactivity [21,22]. The leaves of the tea tree have been primarily cultivated as a source of tea beverages, in which phytochemicals such as catechin and caffeine are the main functional compounds. The development of a new variety that contains enhanced phytochemical contents (qualitatively or quantitatively) is the ultimate objective of tea breeding programs. Therefore, along with the evaluation of phytochemical diversity, the development of a core collection that can represent the diversity of the entire germplasm is very important not only for the conservation and management of germplasm but also for tea breeding programs.

To assess the genetic diversity and/or develop new cultivars in many countries, molecular markers such as restriction fragment of length polymorphism (RFLP) [23,24], random amplified polymorphic DNA (RAPD) [25–27], amplified fragment length polymorphism (AFLP) [24], and simple sequence repeats (SSR) [11,28–30] were used. [31] reported that morphological traits have drawbacks such as the influences of environment on trait expressions, epistatic interactions and pleiotrophic effects among others despite the value of their advantages. On the other hand, molecular markers are used because they are least affected by environmental factors and are almost unlimited in number. In addition, they offer a possibility to observe the genome directly, and thus eliminate the shortcomings inherent in a phenotype observation [32].

In our previous study, we analyzed the genetic diversity of tea accessions collected in Korea using 21 SSRs [28]. In this study, we evaluated the content of eight phytochemicals over two years (2018 and 2019) and analyzed the genetic diversity through 33 SSR markers for 462 tea accessions collected from

Korea, China, Japan, and Indonesia. In addition, a target-oriented core collection was developed using both the phytochemical content and genetic diversity. This core collection will be used to efficiently preserve, manage, and evaluate tea germplasm in the genebank of Korea and to be provided to the tea breeding program as breeding materials.

2. Materials and Methods

2.1. Plant Material

A total of 462 tea accessions were obtained from the National Agrobiodiversity Center (NAC) at the Rural Development Administration in South Korea (Table S1). These accessions are currently preserved as genetic resources in the Tea Industry Institute (34°46' N, 127°5' E) and are maintained through similar horticultural practices. Fresh tea buds and young leaves of the first flush were harvested between 09:00 a.m. and 12:00 a.m. on 24 May in 2018 and 2019. All samples were stored in a freezer at $-80\text{ }^{\circ}\text{C}$ until analysis.

2.2. Phytochemical Analysis

The powdered tea samples (0.1 g) were extracted by intermittent shaking with 1 mL of 70% (*v/v*) methanol at $70\text{ }^{\circ}\text{C}$, and the mixture was centrifuged at 12,000 rpm for 10 min. The supernatant was then diluted 1:10 (sample: 70% methanol) and filtered through a $0.45\text{-}\mu\text{m}$ syringe filter. The diluted samples were analyzed with an Agilent 1260 Infinity HPLC system (Agilent Technology, Santa Clara, CA, USA). The analysis was performed using a COSMOSIL 2.5 Cholesterol (2.5 μm , $2.0 \times 50\text{ mm}$, NACALAI TESQUE, INC., Kyoto, Japan). The HPLC conditions were as follows: solvent A, Acetonitrile/20 mmol/L Phosphate buffer (pH2.5) = 10/90; solvent B, Acetonitrile/ 20 mmol/L Phosphate buffer (pH2.5) = 30/70; B concentration, 0% to 100% 5 min linear gradient; column temperature, $40\text{ }^{\circ}\text{C}$; and flow rate, 0.6 mL/min. The filter detector was set to 280 nm.

All the data collected from three replicate experiments. The summarized phytochemicals in the tea accessions were calculated, and evaluation of the annual variation in the phytochemicals under consideration was conducted through a multivariate analysis of variance (MANOVA) using PAST 3 [33]. Hierarchical clustering was performed using the R statistical software (<http://www.r-project.org>).

2.3. DNA Extraction

Genomic DNA was extracted from the leaves of the tea accessions using a Qiagen DNA extraction kit (Qiagen, Hilden, Germany). The DNA quality and quantity were measured using 1% (*w/v*) agarose gel and spectrophotometry (Epoch, BioTek, Winooski, VT, USA). The extracted DNA was diluted to 30 ng/ μL and stored at $-20\text{ }^{\circ}\text{C}$ until further PCR amplification.

2.4. SSR Genotyping

For the SSR analysis, a total of 33 SSRs were selected from previous studies [11,34] based on linkage groups and PIC value (Supplementary Table S2). They were fluorescently labelled (6-FAM, HEX, and NED) and used to facilitate the detection of the amplification products. The PCR reactions were carried out in a 25 μL reaction mixture containing 30 ng template DNA, 1.5 mM MgCl_2 , 0.2 mM of each dNTP, 0.5 μM of each primer, and 1 U Taq polymerase (Inclone, Korea). Amplification was performed with the following cycling conditions: initial denaturation at $94\text{ }^{\circ}\text{C}$ for 5 min, followed by 35 cycles of denaturation at $95\text{ }^{\circ}\text{C}$ for 30 s, annealing at $55\text{--}62\text{ }^{\circ}\text{C}$ (depending on the primers, Table S2) for 30 s, extension at $72\text{ }^{\circ}\text{C}$ for 1 min, and a final extension step at $72\text{ }^{\circ}\text{C}$ for 10 min. Each amplicon was resolved on an ABI prism 3500 DNA sequencer (ABI3500, Thermo Fisher Scientific Inc., Wilmington, DE, USA) and scored using the Gene Mapper Software (Version 4.0, Thermo Fisher Scientific Inc.).

2.5. Genetic Diversity and Population Structure

The number of alleles (N_a), number of genotypes (N_g), Shannon–Wiener index (S), Expected heterozygosity (H_e), and Evenness were calculated using the poppr package for the R software [35]. An analysis of molecular variance (AMOVA) within and between the gene pools was performed using the GenAlEx software v. 6.5 [36].

The population structure was analyzed using STRUCTURE v.2.3.4 [37] and DAPC. In the STRUCTURE analysis, Bayesian-based clustering was performed, testing three independent runs, with K ranging from 1 to 10. Each run had a burn-in period of 50,000 iterations and 500,000 Monte Carlo Markov iterations, assuming an admixture model. The output was subsequently visualized with the STRUCTURE HARVESTER v.0.9.94 [38]. The most likely number of clusters was inferred according to Evanno [39]. The DAPC analysis was performed using the adegenet package for the R software [40,41] according to Lee et al. [28]. A Mantel test was performed using R software [40] in order to investigate the relationship between the genetic and phytochemical distances of tea accessions.

2.6. Development and Evaluation of the Core Collection

The POWERCORE program [42] was used to develop the independent core collection using the phytochemical data for two years (2018 and 2019) and the genotypic data of 33 SSR markers. The mean difference percentage (MD%), variance difference percentage (VD%), variable rate of coefficient of variance (VR%), and coincidence rate of range (CR%) were calculated to assess the level of diversity captured in the core collection compared to the entire collection [43]. In addition, the representation of the core collection was evaluated by estimating Shannon's diversity index (I) and Nei's diversity index (H). The distance matrix was used to construct a dendrogram via the neighbor-joining (NJ) method with 1000 bootstrap replicates. The principal coordinate analysis (PCoA) was performed using DARwin v. 6.0 [44].

3. Results

3.1. Phytochemical Diversity of 462 Tea Accessions

The variation and distribution of the catechin and caffeine (Caf) content in the 462 tea accessions are summarized in Table 1. Wide variation and diversity of the phytochemicals in the tea germplasms were observed. The average total catechin (TC) content in 2018 and 2019 was 70.04 ± 15.37 mg/g and 89.77 ± 19.04 mg/g, varying from 34.14 to 125.25 and 42.03 to 152.30 mg/g, respectively. Epigallocatechin 3-gallate (EGCG) was the most abundant catechin (63.9% in 2018 and 55.9% in 2019), averaging 44.74 ± 12.31 mg/g in 2018 and 50.20 ± 15.01 mg/g in 2019; epigallocatechin 3-gallate (ECG), epicatechin (EC), gallic catechin (GC), catechin (C), catechin 3-gallate (CG), and gallic catechin 3-gallate (GCG) were the next most common in abundance.

The levels of Caf, EC, ECG, EGCG, and TC in the 462 tea accessions between 2018 and 2019 demonstrated a normal distribution, and their H' levels were high (≥ 2.00). In contrast, minor components such as GC, C, CG, and GCG did not show a normal distribution, and their H' values were lower, although their coefficients of variation were high.

All nine phytochemicals showed highly significant differences between tea accessions ($p < 0.001$) and experimental years ($p < 0.001$) (Table S3). In the year \times accessions interactions, C and CG did not show significant differences, while the other phytochemicals showed highly significant differences ($p < 0.001$).

Table 1. Variation and distribution of phytochemical contents (mg/g) in the 462 tea accessions.

Phytochemical	Year	Min	Max	Mean	SD	Median	Skewness	Kurtosis	CV (%)	H' ¹
C	2018	0.32	42.53	3.14	4.38	2.03	5.12	31.94	139.37	1.14
	2019	1.05	19.13	4.76	2.87	4.04	1.75	4.34	60.26	1.86
Caf	2018	0.44	36.64	17.42	5.20	16.96	−0.02	1.41	29.83	2.00
	2019	0.39	28.79	15.95	5.24	16.29	−0.46	0.15	32.86	2.07
CG	2018	0.15	5.61	1.30	1.03	0.97	1.27	1.49	79.22	1.88
	2019	0.34	21.76	2.47	1.72	2.06	4.60	39.43	69.53	1.65
EC	2018	3.42	26.58	10.40	3.21	10.07	1.05	2.50	30.83	2.02
	2019	1.98	22.88	8.12	2.97	7.73	0.98	1.88	36.59	2.00
ECG	2018	1.96	17.98	7.88	2.84	7.51	0.71	0.53	36.01	2.03
	2019	3.15	34.42	14.68	4.89	14.64	0.38	0.51	33.30	2.07
EGCG	2018	13.15	95.90	44.74	12.31	43.74	0.39	0.40	27.51	2.06
	2019	11.49	91.34	50.20	15.01	50.38	0.02	−0.05	29.90	2.08
GC	2018	0.67	13.71	3.05	1.86	2.61	2.22	6.74	60.88	1.74
	2019	3.21	15.57	7.72	1.84	7.53	0.53	0.73	23.87	2.03
GCG	2018	0.01	3.49	0.50	0.43	0.38	2.84	12.26	85.69	1.67
	2019	0.33	9.75	1.89	1.07	1.62	1.81	6.92	56.70	1.87
TC	2018	34.14	125.25	70.04	15.37	68.50	0.53	0.23	21.94	2.05
	2019	42.03	152.30	89.77	19.04	89.00	0.08	0.07	21.21	2.06

¹ H' , Shannon-Weave index; C, (+)-Catechin; Caf, Caffeine; CG, (−)-Catechin 3-gallate; EC, (−)-Epicatechin; ECG, (−)-Epigallocatechin 3-gallate; EGCG, (−)-Epigallocatechin 3-gallate; GC, (−)-Gallocatechin; GCG, (−)-Gallocatechin 3-gallate; TC, total catechin.

3.2. Clustering Analysis

In total, 462 tea accessions were classified into four clusters according to their phytochemicals (Table 2 and Figure 1.). Cluster I contained 108 tea accessions and had higher content of Caf, ECG, EGCG, GCG, and TC between 2018 and 2019. Cluster II had 111 accessions and showed lower contents of C, EC, GC, and GCG in 2018 and higher contents of Caf, ECG, EGCG, GC, and TC in 2019. Cluster III consisted of 59 tea accessions and showed higher C and EC and lower Caf, ECG, and EGCG content between the two years. Cluster IV had 184 tea accessions with lower C and GCG contents between 2018 and 2019.

Table 2. Average cluster values of the phytochemicals in the 462 tea accessions for the two years (mg/g).

Year	Group	N ¹	C	Caf	CG	EC	ECG	EGCG	GC	GCG	TC
2018	I	108	3.66b ²	19.83a	1.95a	11.16ab	10.94a	55.78a	3.55a	7.99a	92.18a
	II	111	2.06c	17.84b	1.22b	8.75c	7.10b	45.62b	2.40c	4.63bc	65.56b
	III	59	10.71a	12.88c	1.50b	11.62a	5.70c	33.59d	3.21ab	5.66b	67.46b
	IV	184	2.63bc	17.21b	1.27b	10.55b	7.24b	41.25c	3.09b	3.95c	65.36b
2019	I	108	5.01b	18.66a	2.35b	7.67b	18.34a	59.04a	7.83b	25.25a	124.85a
	II	111	4.70b	19.80a	2.25b	7.43b	17.54a	61.10a	8.89a	20.74b	121.84a
	III	59	8.60a	9.37c	4.32a	11.11a	9.64c	32.21c	6.68c	19.52b	91.59b
	IV	184	3.50c	13.76b	2.21b	7.91b	12.15b	43.23b	7.24bc	13.89c	89.60b

¹ N, number of accessions; C, (+)-Catechin; Caf, Caffeine; CG, (−)-Catechin 3-gallate; EC, (−)-Epicatechin; ECG, (−)-Epigallocatechin 3-gallate; EGCG, (−)-Gallocatechin 3-gallate; GC, (−)-Gallocatechin; GCG, (−)-Gallocatechin 3-gallate; TC, total catechin. ² The same letter in each column indicates no significant difference according to a least significant difference test, $p < 0.05$.

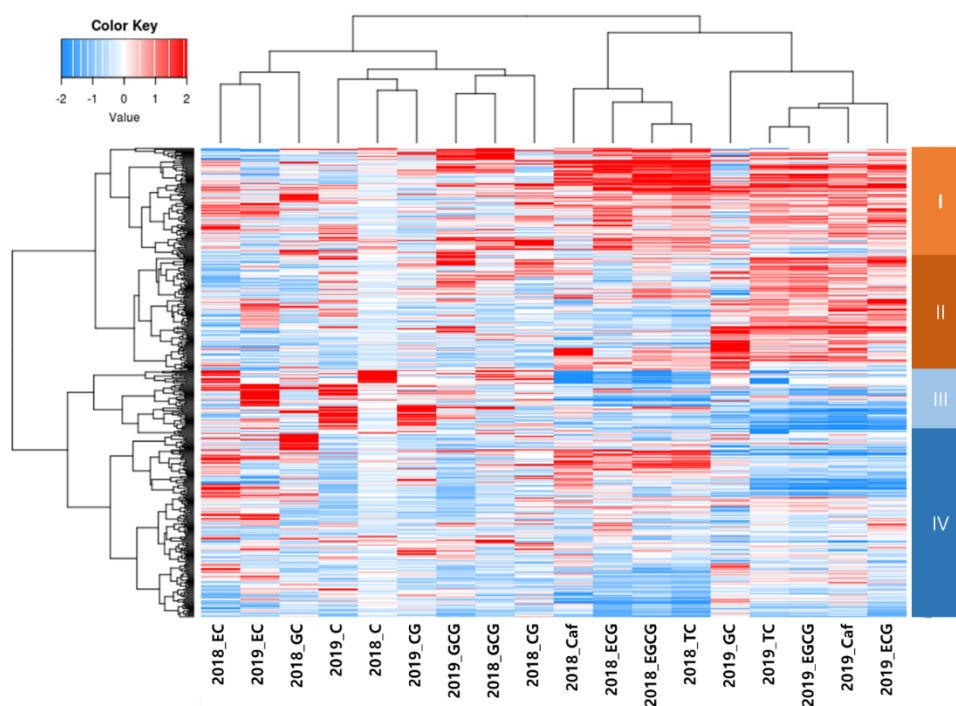


Figure 1. Hierarchical clustering analysis of the phytochemicals in the 462 tea accessions. The colours in the heatmap indicate the z-score which was calculated by subtracting the mean of phytochemicals across different samples and dividing it by the standard deviation of the phytochemicals across all the samples. The red color indicates positive z-score, the white color indicates zero z-score, whereas the blue colour indicates negative z-score. Higher intensity of the color in the scale indicates a higher magnitude of the z-score. The dendrogram on the x-axis indicates the degree of similarity between the phytochemicals, the closer the phytochemicals the higher the level of similarity in them and the phytochemicals have been clustered using hierarchical clustering. Similarly, the dendrogram on the y-axis indicates the degree of similarity between the different samples, the closer the samples the higher the level of similarity in them and they have been clustered using hierarchical clustering (Ward, Euclidean distance).

3.3. SSR Fingerprinting

A total of 428 alleles were detected in 33 SSR loci among the 462 tea accessions (Table 3). The number of observed alleles (N_a) and the number of genotypes (N_g) ranged from 5 (TM324 and TM480) to 23 (MSE0083), with an average of 13.0, and 10 (TM324 and TM480) to 103 (MSE0083), with an average of 50.2. The Shannon–Wiener index (S) and expected heterozygosity (H_e) ranged from 0.92 (TM461) to 2.42 (TM422), with an average of 1.78, and 0.54 (MSE0237) to 0.88 (MSE0083 and TM422), with an average of 0.77. The evenness was calculated from 0.57 (TM576) to 0.90 (TM351), with an average of 0.75.

The diversity indices among the four origins are calculated in Table 4. The N_a and N_g contents ranged from 3.2 (IDN) to 11.8 (KOR) and 2.4 (IDN) to 43.6 (KOR), respectively. The S and H_e contents were calculated to be 1.05 (IDN) to 1.91 (CHN) and 0.73 (JPN) to 0.81 (CHN), respectively. The Evenness ranged from 0.76 (KOR) to 0.90 (IDN), with an average of 0.79.

Table 3. Genetic diversity parameters of the 33 SSR markers in the 462 tea accessions.

Locus	Na ¹	Ng	S	He	Evenness	Locus	Na	Ng	S	He	Evenness
MSE0029	14	63	2.25	0.86	0.79	MSG0699	20	101	1.02	0.86	0.70
MSE0083	23	103	1.92	0.88	0.74	TM241	9	30	2.03	0.75	0.81
MSE0107	19	66	2.03	0.82	0.66	TM324	5	10	2.05	0.56	0.84
MSE0113	18	63	1.91	0.84	0.76	TM337	14	68	1.83	0.87	0.83
MSE0173	12	47	2.40	0.82	0.80	TM341	9	22	1.48	0.71	0.76
MSE0237	9	19	1.15	0.54	0.65	TM351	6	13	1.09	0.7	0.90
MSE0291	13	59	1.35	0.78	0.69	TM382	11	49	2.17	0.82	0.85
MSE0313	17	59	2.30	0.83	0.73	TM422	22	94	2.42	0.88	0.73
MSE0403	16	58	2.15	0.79	0.64	TM428	13	59	1.17	0.8	0.67
MSG0258	18	101	1.79	0.87	0.79	TM447	7	22	1.44	0.75	0.88
MSG0361	18	68	1.82	0.87	0.74	TM461	8	28	0.92	0.77	0.78
MSG0380	15	79	2.18	0.86	0.77	TM480	5	10	1.93	0.63	0.85
MSG0423	20	76	1.88	0.84	0.70	TM530	7	25	1.20	0.68	0.73
MSG0429	13	44	2.26	0.76	0.61	TM576	7	14	1.56	0.56	0.57
MSG0470	17	58	1.82	0.81	0.74	TM581	7	16	1.65	0.62	0.71
MSG0610	10	40	2.16	0.78	0.73	TM604	7	16	1.26	0.65	0.85
MSG0681	19	78	2.14	0.87	0.82	Mean	13.0	50.2	1.78	0.77	0.75

¹ Na, Number of observed alleles; Ng, Number of genotypes; S, Shannon–Wiener index; He, Expected heterozygosity.

Table 4. Genetic diversity parameters of the four origins in the 462 tea accessions using 33 SSRs.

Origin	N ¹	Na	Ng	S	He	Evenness
KOR	408	11.8	43.6	1.73	0.76	0.76
JPN	13	5.9	6.8	1.42	0.73	0.78
IDN	3	3.2	2.4	1.05	0.76	0.90
CHN	38	10.5	18.4	1.91	0.81	0.79

¹ N, Number of accessions; Na, Average number of alleles per accessions; Ng, Number of genotypes; S, Shannon2013Wiener index; He, Expected heterozygosity; KOR, Korea; JPN, Japan; IDN, Indonesia; CHN, China.

Genetic and phytochemical distance differences among tea accessions were concordant based on the Mantel test ($r = 0.0899$, $p = 0.017$) indicating that these two analyses (genetic and phytochemicals) grouped the genotypes in a different manner.

3.4. Population Structure

The relatedness among genotypes and their rooting with geographical designation were studied using a population structure analysis. Determination of the log mean probability and change in the log probability (ΔK) (following [29]) provided two subpopulations ($K = 2$) (Figure 2A). STR_C1 was dominated by genotypes belonging to 191 accessions from KOR, 12 accessions from JPN, 3 accessions from CHN, and 2 accessions from IDN (Figure 2B). STR_C2 contained 217 accessions from KOR, 35 accessions from CHN, and one accession from JPN and IDN. The mean alpha value (an estimate of the degree of admixture) for the analyzed samples was 0.2759.

To understand the genetic relationship among the 462 tea accessions, a DAPC analysis was performed (Figure 3). Four clusters were detected in coincidence with the lowest BIC values using the *find.clusters* function. The DAPC analysis was carried out using the detected number of clusters. Typically, the 50 first PCs (60.3% of variance conserved) of PCA and three discriminant eigenvalues were retained. These values were confirmed via a cross-validation analysis. The four clusters were titled D1–4. A major shift in accessions from STR_C1 to D1 and D3 was observed, and the main tea accessions of D2 and D4 were located in STR_C2. D1 contained 132 (126 accessions from STR_C1 and six accessions from STR_C2) accessions, with 123 (117 from STR_C1 and 6 from STR_C2) from KOR, 6 (STR_C1) from JPN, two (STR_C1) from IDN, and one (STR_C1) from CHN. The major accessions of D2 included 43 accessions (42 from STR_C2 and one from STR_C1) from KOR and 32 accessions (STR_C2) from CHN, with one accession from JPN. D3 consisted of 83 accessions, with 75 accessions

(73 in STR_C1 and two in STR_C2) from KOR, six (STR_C1) from JPN, and two (STR_C1) from CHN. D4 comprised 171 accessions, with 167 accessions (STR_C2) from KOR, three (STR_C2) from CHN, and one (STR_C2) from IDN.

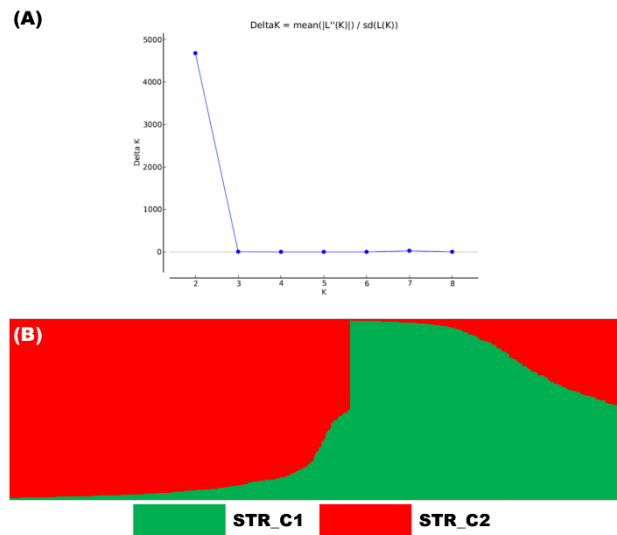


Figure 2. Model-based populations in association panels consisted of 462 tea accessions: (A) Delta K values for different numbers of populations (K) assumed in analysis completed with the STRUCTURE software (B) Classification of 462 tea accessions into two populations using STRUCTURE Version 2.3.4, where the numbers on the y-axis show the subgroup membership, and the x-axis shows the different accession. The distribution of accessions into different populations is indicated by the color coding (Cluster 1, STR_C1, is red round; and Cluster 2, STR_C2, is green).

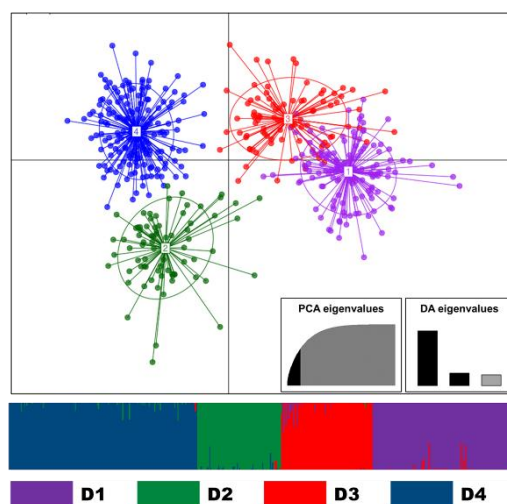


Figure 3. Discriminant analysis of the principal components (DAPC) for 462 tea accessions. The axes represent the first two Linear Discriminants (LD). Each circle represents a cluster, and each dot represents an individual. The numbers represent different subpopulations identified by the DAPC analysis.

The molecular variance within and between the regional pools, as well as the sub-populations derived from the clustering analysis, STRUCTURE, and DAPC analysis, was evaluated (Table 5). For the regional gene pools, percentage of variance within and among populations was found to be 94% and 6% of the total variation, respectively. The clustering analysis and DAPC provided a variance of 99% and 1% for within and among sub-populations, respectively, while the two sub-population derived from STRUCTURE showed only 100% of total variance within the groups. Among the four

AMOVA results, regional pools showed minimum within-population variance (94%) and maximum among-population variance (6%), indicating that the regional pools are fairly structured groups for the panel under consideration. The genetic differentiation (PhiPT) of the four subpopulations showed a range from 0.005 (STRUCTURE) to 0.056 (regional pools).

Table 5. Analysis of the molecular variance (AMOVA) among and within populations in the regional pools and sub-populations derived from the clustering analysis, STRUCTURE, and DAPC.

	df ¹	SS	MS	Est. Var.	%	PhiPT	p Value
Regional pools							
Among Populations	3	326.490	108.830	2.199	6%	0.056	0.001
Within Populations	458	16,882.545	36.861	36.861	94%	-	-
Total	461	17,209.035	-	39.061	100%	-	-
Sub-population derived from clustering analysis							
Among Populations	3	211.739	70.580	0.305	1%	0.008	0.001
Within Populations	458	16,997.296	37.112	37.112	99%	-	-
Total	461	17,209.035	-	37.417	100%	-	-
Sub-population derived from STRUCTURE							
Among Populations	1	77.344	77.344	0.175	0%	0.005	0.003
Within Populations	460	17,131.690	37.243	37.243	100%	-	-
Total	461	17,209.035	-	37.418	100%	-	-
Sub-population derived from DAPC							
Among Populations	3	263.207	87.736	0.456	1%	0.012	0.001
Within Populations	458	16,945.828	37.000	37.000	99%	-	-
Total	461	17,209.035	-	37.456	100%	-	-

¹ df, degree of freedom; SS, sum of squares; MS, mean squares; Est. Var., estimates of variance; %, percent of variance.

3.5. Development and Evaluation of a Core Collection

The MANOVA analysis indicated significant year effects, as well as significant interaction effects between the year and accession effects by considering all quantitative traits together (Table S4). Therefore, the phytochemical data for both years (2018 and 2019) and molecular marker data were treated independently for the development of the core collection. The target-oriented core collection (TOCC) was developed with phytochemicals and molecular data using POWERCORE. TOCC included 100 accessions (21.6% of the entire collection) belonging to four origins, with 73 accessions from KOR, 22 from CHN, 4 from JPN, and 1 from IDN.

Differences between the means of the entire collection (ENC) and TOCC were found to be not significant for all traits (Table 6). The mean difference percentage (MD%), coincidence rate of range (CR%), variance difference percentage (VD%), and variable rate of the coefficient of variance (VR%) were used to comparably evaluate the properties of TOCC with ENC (Table 7). Overall, the nine phytochemicals, MD%, VD%, VR%, and CR% were 7.88%, 39.33%, 120.79%, and 97.43%, respectively.

Table 6. Phytochemical diversity comparison between the entire collection and core collection of tea accessions (mg/g).

		C ¹	Caf	CG	EC	ECG	EGCG	GC	GCG	TC
2018	Entire Collection	3.14	17.42	1.30	10.40	7.88	44.74	3.05	0.50	70.04
	Target-oriented core collection	3.93	16.71	1.26	10.45	7.43	43.42	2.91	0.47	68.58
		ns	ns	ns	ns	ns	ns	ns	ns	ns
2019	Entire Collection	4.76	15.95	2.45	8.12	14.68	50.20	7.72	1.89	89.78
	Target-oriented core collection	4.61	15.91	2.45	8.04	14.53	50.73	7.56	1.91	89.85
		ns	ns	ns	ns	ns	ns	ns	ns	ns

¹ C, (+)-Catechin; Caf, Caffeine; CG, (-)-Catechin 3-gallate; EC, (-)-Epicatechin; ECG, (-)-Epigallocatechin 3-gallate; EGCG, (-)-Gallocatechin 3-gallate; GC, (-)-Gallocatechin; GCG, (-)-Gallocatechin 3-gallate; TC, total catechin; ns, not significant at $p = 0.05$.

Table 7. Evaluation indices for the developed core collection.

	N ¹	MD% ¹	VD%	VR%	CR%	Na	I	H
Entire Collection (ENC)	462	-	-	-	-	2.11	0.308	0.195
Target-oriented core collections (TOCC)	100	7.88	39.33	120.79	97.43	2.11	0.335	0.209

¹ N, number of accessions; MD%, mean difference percentage; VD%, variance difference percentage; VR, variable rate of coefficient of variance; CR%, coincidence rate of range; I, Shannon's diversity index; H, Nei's diversity index; Na, Average number of alleles per locus.

To evaluate the quality of TOCC, Shannon's diversity index (I), Nei's diversity index (H), and the number of alleles (Na) were calculated using the molecular data (Table 7). The number of alleles (Na) in TOCC was the same as that in ENC. The genetic diversity of TOCC revealed by these markers was compared with that of ENC. The I and H of TOCC were higher (0.335, 0.209) than those of ENC (0.308, 0.195).

The distribution of tea accessions in TOCC was determined via a line graph obtained through the phytochemicals of ENC, along with a Principal coordinate analysis (PCoA) and Neighbor Joining (NJ) obtained through the genetic analysis of ENC (Figure 4 and Supplementary Figure S1). The accessions in TOCC were shown to be selected proportionally, accurately reflecting the distribution of the overall accessions for each phytochemical. For the distribution of molecular data, TOCC showed a balanced distribution in PCoA and NJ.

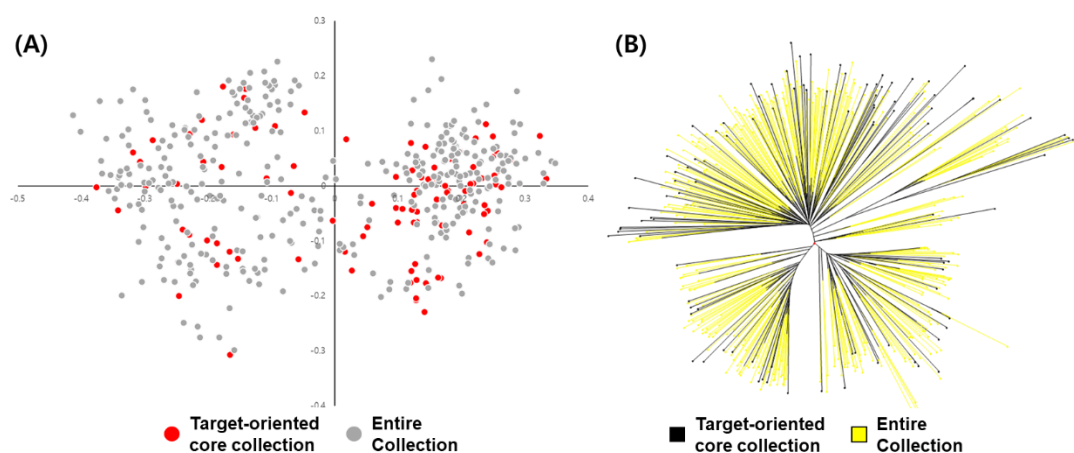


Figure 4. Distribution of tea accessions in the core collections using molecular data. (A) Principal coordinate analysis; (B) Neighbor-Joining Tree.

4. Discussion

A vast collection consisting of 15,234 accessions of tea is available in 23 gene banks around the world [7]. The biochemical characterization of tea germplasm in earlier studies demonstrated significant variability [18,45–48]. Despite the substantial diversity of compounds in tea germplasm, the development of tea cultivars was limited due to bottlenecks in tea breeding, such as long gestation periods, high inbreeding depression, and self-incompatibility [49]. In addition, the tea quality and yield in the main tea producing countries, such as China, India, Sri Lanka, Kenya, Japan, etc., were significantly improved with an increase in the ratio of clonal tea acreage [50]. Breeding strategies often focus on a limited set of target traits, resulting in cultivars with a narrow genetic base. Yao et al. [51] reported that the developed tea cultivars from China, Japan, and Kenya have a narrow genetic basis due to the popularity of only a few cultivars for breeding and planting. This has produced several problems, such as the spread of specific diseases and insects, the concentration of plucking time in the tea season, the non-uniformity of taste and flavor, and susceptibility to environmental changes [40,51]. Meegahakumbura et al. [29] noted that a molecular analysis that can discern not

only patterns of lineage, but the origin of tea germplasm is also required because the morphological characteristics that are traditionally used to define cultivars are highly plastic and easily influenced by environmental conditions. The present study attempted to address the above issue by generating a core collection of tea germplasm that includes data on the molecular variability of the crop, in addition to biochemical characterization.

4.1. Phytochemical Diversity of Tea Germplasm

Significant variation was observed among the 462 tea accessions for catechin and caffeine content in this study (Table 1). In addition, significant differences between the two years were observed (Table S3). Catechins and caffeine serve as secondary metabolite defense compounds in tea plants. They provide sessile plants with protection against pathogens and predators, oxidative stress, and other environmental variables. Thus, the content of catechins and caffeine varied in the tea samples based on environmental variability [45]. Many previous studies reported a large variation in catechin and caffeine contents in tea accessions [15,18,52,53]. The authors in [54] noted that a biochemical characterization with different proportions of total catechins and their components would be a useful tool for the development of quality-tea clones. The authors in [55] reported that differences between locations were far larger than the variations among cultivars, implying that environmental effects should be taken into consideration when total catechin and its component contents are utilized as biochemical markers in tea breeding programs.

There are six major catechins in tea leaf: (+)-catechin (C), (–)-epicatechin (EC), (–)-epicatechin gallate (ECG), (+)-gallocatechin (GC), (–)-epigallocatechin (EGC), and (–)-epigallocatechin gallate (EGCG) [56]. The concentration of catechins in tea was determined as follows: EGCG>ECG>EGC>EC>GC>C [52,53,57,58]. In addition, the authors determined antioxidant activity in the following order: ECG>EGCG>EC>EGC [59]. The variation of catechin contents in tea accessions depends on the condition of the tea germplasm, such as the number of samples and the origin of the tea accessions, in each study. The range of each catechin's content in the previous studies was as follows: EGCG, 13.0 to 139.0 mg/g; ECG, 3.2 to 89.1 mg/g; EGC, 2.1 to 249 mg/g; EC, 2.0 to 54.5 mg/g; GC, 1.4 to 22.7 mg/g; and C, 0.3 to 30.9 mg/g [52,54,57,59,60]. In this study, the 462 tea accessions also showed a similar level of catechin content to that in previous studies (Table 1). The concentration of catechins in tea germplasm is important for tea quality. For instance, the ratio of $(EGCG + ECG) \times 100/EGC$ has been suggested as a quality index for measuring the difference in the catechin levels of fresh tea shoots across growing seasons [60]. In addition, the catechin index (CI) $(EC + ECG)/(EGC + EGCG)$ has been used as a biochemical marker for studying the genetic diversity of tea germplasms [54]. The tea accessions with desirable compositions of catechins in this study could be incorporated into breeding programs for crop improvement.

Caffeine is the most abundant alkaloid in tea, with content usually between 15 and 50 mg/g [15]. In this study, the caffeine content of 462 tea accessions ranged from 0.4 to 36.6 mg/g (2018) and 0.4 to 28.8 mg/g (2019) (Table 1). Kottawa-Arachchi et al. [15] noted that various amounts of caffeine have been observed in different tea growing countries. Due to the pharmacological properties of caffeine on the central nervous system, the demand for low-caffeine tea is increasing greatly, from 2% of total tea consumption in 1980 to 15% in the early twenty-first century [61]. Although many countries have invested in methods and techniques to make decaffeinated tea, such techniques can remove the tea's unique aroma and taste, which will worsen the quality. It is thus important to develop low caffeine clones through breeding and selection, as such clones could be a solution to the problem of high caffeine levels and contribute tremendously to the provision of natural low-caffeine tea [18]. The tea accessions with a lower caffeine content in this study could be used as naturally low-caffeine genetic resources for crossbreeding parents.

4.2. Genetic Diversity of Tea Germplasm

In our previous study, we analyzed the genetic diversity and population structures of 410 tea accessions collected from South Korea using 21 SSR markers and revealed the narrow genetic base of South Korean tea accessions [28]. In the present study, the genetic diversity and population structure of 462 tea accessions from China, Japan, Indonesia, and Korea (conserved in NAC) were analyzed using 33 SSR markers. As shown in Table 4, higher diversity was detected among the tea accessions in China ($H = 1.91$, $I = 0.81$) than among those in Korea ($H = 1.73$, $I = 0.76$), Japan ($H = 1.42$, $I = 0.73$), and Indonesia ($H = 1.05$, $I = 0.76$). Other studies also similarly reported that the Chinese tea population exhibited a higher level of genetic diversity than that of other tea populations from other countries [24,51]. In general, China is thought to be the origin of tea, so Chinese tea populations are the most likely to account for the largest proportion of diversity [51]. Our previous study noted that Korean tea germplasm showed low genetic diversity because of limitations in the gene stock from China, political and religious reasons, and extreme environmental conditions [45]. Tanaka et al. [62] reported that the tea plant in Japan was first introduced from China about 1200 years ago and that the country's original tea populations were established based on only a few of seeds from a restricted source. In addition, the authors in [23,25] suggested that the low genetic diversity of tea accessions in Japan could be attributed to long and intensive selection and breeding from the genetically limited tea stock in Japan.

It is important to identify the correspondence between the genetic diversity of tea accessions and their origins. In this study, the different approaches (STRUCTURE and DAPC) used to analyze the population structures of the 462 tea accessions were able to provide complementary information. However, the structuring of tea accessions at $K = 2$ (based on the estimated ΔK value in STRUCTURE) and $K = 4$ (based on the BIC and DAPC) clearly did not segregate the accessions based on geographical distinctions. The Evanno method is artificially maximal at $K = 2$, in some cases, because it finds the highest level of structure in the data by focusing only on the changes in slope [39,63]. Similar results were obtained in previous studies on tea germplasm structures based on SSR ($K = 2$) [11,30,64,65]. The DAPC method does not require that populations be in HW equilibrium and can handle large sets of data without using parallel processing software, so it provides an interesting alternative to the STRUCTURE software [66]. In addition, the DAPC analysis provided more detailed clusters compared to the STRUCTURE analysis in previous analyses using SSR [28,66,67]. Our results also agree with those of previous studies where the DAPC analysis ($K = 4$) provided more detail than STRUCUTRE ($K = 2$). However, these results indicated lower genetic differentiation (PhiPT, DAPC = 1.2%; Clustering analysis of phytochemicals = 0.8%; STRUCTURE = 0.5%) than that in the collection area (5.6%). This might be due to an imbalance in the distribution of tea accessions used in this study, as 88.3% of tea accessions in this study were collected from South Korea. In our previous study, the genetic differentiation in the DAPC analysis of Korean tea germplasm was 1.4% [23]. This affected the low genetic differentiation between groups resulting from an analysis of the population structure, although the genetic differentiation of tea origins was also shown to be low (5.6%).

4.3. Development of a Target-Oriented Core Collection

To develop core collections, various methods, such as phenotypes, proteins, and molecular markers, have been used. However, there is no universally accepted method to construct a core collection because every method has advantages and disadvantages [68]. Previous studies have proven that phenotypes are useful parameters for developing core collections [2,12,69]. Kumar et al. [70] reported that the use of molecular markers in the development of a core collection is more effective than the use of other data, such as morphological traits sensitive to environmental effects. In addition, molecular markers are more effective in identifying and minimizing redundancy. Le et al. [71] suggested that the use of both phenotypic and molecular data together is more effective than their use individually when constructing a core collection. In this study, molecular markers and biochemical contents were utilized for the construction of a core collection in tea germplasm using the POWERCORE program,

which was successfully used to build a core collection for various plant species, including olive [69], safflower [71], and tea [9].

In this study, seasonal data sets were handled independently to develop the core collections because the MANOVA analysis presented noticeable Genotype X environmental interactions. In addition, the evaluation indices (MD%, VD%, VR%, CR%) were comparable and reflected their effectiveness in capturing diversity to validate the core collection. MD%, VD%, and VR% were used to evaluate the statistical consistency between the core and entire collections [42], while MD% was used to represent the difference in the accession averages between the core and entire collections, which should be <20% for a representative core collection. VD% indicates the variance captured by the core collection, and VR% indicates a comparison between the coefficient of variation values present in the core and entire collections. CR% indicates whether the distribution ranges of each variable in the core set are well represented when compared to the entire collection, which should be greater than 80% [12,42,43,70]. In this study, the core collections yielded a CR% of more than 80% (97.43%) and an MD% of less than 20% (7.88%) (Table 7). Similar results for other species were reported in core collections developed with a lower MD% or higher CR%, which were more representative of the entire collections [72,73]. In addition, the distributions of each phytochemical in the tea accessions showed similarities to those of the entire collection (Figure S1). In general, the core collections can be classified into three types or categories: core collections representing (1) individual accessions, (2) extremes, and (3) the distribution of accessions in the entire collection [3]. Odong et al. [3] suggested that a core collection of type 3 (distribution of accessions) is only of interest if the aim is to provide an overview of the composition of the whole collection using only a part of the collection. The authors in [23,74] suggested that this type of core collection can be obtained by maximizing the representativeness of the pattern of trait variations in the whole collection. Considering these reports, the core collection developed in this study showed a similar pattern of type 3, which could represent the entire collection.

By integrating genetic diversity and phytochemical content, we developed a target-oriented core collection that we have not tried before in tea germplasm. The main targets for tea breeding and use are mostly related to catechin content; therefore, the phytochemical analysis and development of TOCC allow us to extend the use of tea germplasm broadly. Furthermore, the TOCC retained the phytochemical and genetic diversity of ENC, as we extracted the accessions after analyzing the variation of the content over two years using molecular marker data. The genetic diversity indices (I and H) and the distribution of accessions (NJ and PCoA) also indicate that the TOCC is well developed and reflects the whole diversity of ENC. Throughout this process, we developed a greater value-added core collection, which will not only provide useful materials to breeders but also aid in the efficient management of genebank. This target-oriented core collection is distinguished from the previous core collection in which accessions were selected based on their agronomic traits and molecular markers. Our upgraded core collection focused on the phytochemical content in tea germplasm suggests new directions for the use and conservation of tea germplasm.

5. Conclusions

Evaluating a plant germplasm and establishing a core collection will enhance the proper utilization of plant genetic resources [73]. Especially, core collections have been developed in various crop collections because they have a size that facilitates evaluation, use, and conservation while maintaining existing genetic diversity [7]. In this study, phytochemicals content and genetic diversity on 462 tea accessions were evaluated and the target-oriented core collection was constructed based on these results. The phytochemical contents of 462 tea accessions showed varying distributions, although the genetic diversity was low. In addition, this is the first attempt to combine molecular diversity data with phytochemicals to develop a core collection of tea germplasm conserved in NAC. This target-oriented core collection will provide access to genetic diversity and phytochemical traits, which will be useful for characterizing the genetic determinants of the traits of interest. Furthermore, it could be used to design more effective breeding programs to increase the global utility of tea as a functional crop.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/2073-4395/10/11/1667/s1>, Figure S1. Distribution of phytochemicals between the entire collection and the core collection in 2018 (A) and 2019 (B). Blue line, entire collection; Orange line, core collection. C, (+)-Catechin; Caf, Caffeine; CG, (−)-Catechin 3-gallate; EC, (−)-Epicatechin; ECG, (−)-Epigallocatechin 3-gallate; EGCG, (−)-Gallocatechin 3-gallate; GC, (−)-Gallocatechin; GCG, (−)-Gallocatechin 3-gallate; TC, total catechin. Table S1. List and the content of phytochemicals of 462 tea accessions in this study. Table S2. List of 33 SSR primers in this study. Table S3. Mean squares for phytochemicals according to year, accessions, and year x accession interactions. Table S4. Multivariate analysis of variance (MANOVA) to study yearly differences in all quantitative traits together.

Author Contributions: Conceptualization, K.J.L. and D.Y.H.; data curation, K.J.L. and D.Y.H.; formal analysis, K.J.L. and R.S.; resources, G.-Y.G. and D.-M.S.; investigation, S.-H.K., E.Y., and S.L.; writing—original draft, K.J.L.; writing—review and editing, D.Y.H.; funding acquisition, G.-T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research Program for Agricultural Science and Technology Development, National Institute of Agricultural Sciences, Rural Development Administration, Republic of Korea, grant number [PJ01355702].

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

TOCC	Target-oriented core collection
ENC	Entire collection
C	(+)-Catechin
CG	(−)-Catechin 3-gallate
EC	(−)-Epicatechin
ECG	(−)-Epigallocatechin 3-gallate
EGCG	(−)-Epigallocatechin 3-gallate
GC	(−)-Gallocatechin
GCG	(−)-Gallocatechin 3-gallate
TC	Total catechin

References

- Engels, J.; Fassil, H. Plant and Animal Genebanks. In *Encyclopedia of Life Support Systems*; UNESCO, Ed.; Eolss Publishers: Paris, France, 1999; p. 34.
- Qi-Lun, Y.; Fan, P.; Zou, S.-X. Constructing a Core Collection for Maize (*Zea mays* L.) Landrace from Wuling Mountain Region in China. *Agric. Sci. China* **2008**, *7*, 1423–1432. [[CrossRef](#)]
- Odong, T.L.; Jansen, J.; Van Eeuwijk, F.A.; Van Hintum, T.J.L. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theor. Appl. Genet.* **2012**, *126*, 289–305. [[CrossRef](#)]
- Frankel, O.H. Genetic perspectives of germplasm conservation. In *Genetic Manipulation: Impact on Man and Society*; Arber, W., Illemensee, K., Peacock, W.J., Starlinger, P., Eds.; Cambridge University Press: Cambridge, UK, 1984; pp. 161–170.
- Brown, A.H.D. Core collections: A practical approach to genetic resources management. *Genome* **1989**, *31*, 818–824. [[CrossRef](#)]
- Wambulwa, M.C.; Meegahakumbura, M.K.; Kamunya, S.; Muchugi, A.; Möller, M.; Liu, J.; Xu, J.-C.; Ranjitkar, S.; Olinirina, N.; Gao, L.-M. Insights into the Genetic Relationships and Breeding Patterns of the African Tea Germplasm Based on nSSR Markers and cpDNA Sequences. *Front. Plant Sci.* **2016**, *7*, 1244. [[CrossRef](#)]
- Bramel, P.; Chen, L. *A Global Strategy for the Conservation and Use of Tea Genetic Resources*; The Crop Trust: Bonn, Germany, 2019.
- Chen, L. Tea genetic resources in China. *Int. J. Tea Sci.* **2012**, *8*, 1–10.
- Raina, S.N.; Ahuja, P.S.; Sharma, R.K.; Das, S.C.; Bhardwaj, P.; Negi, R.; Sharma, V.; Singh, S.S.; Sud, R.K.; Kalia, R.K.; et al. Genetic structure and diversity of India hybrid tea. *Genet. Resour. Crop. Evol.* **2011**, *59*, 1527–1541. [[CrossRef](#)]
- Ranatunga, M.; Gunasekare, M. Stratification of Camellia Germplasm to Facilitate Construction of Core Collection; A Prerequisite for Tea Crop Improvement. *Sri Lanka J. Tea Sci.* **2009**, *74*, 62–73.

11. Taniguchi, F.; Kimura, K.; Saba, T.; Ogino, A.; Yamaguchi, S.; Tanaka, J. Worldwide core collections of tea (*Camellia sinensis*) based on SSR markers. *Tree Genet. Genomes* **2014**, *10*, 1555–1565. [[CrossRef](#)]
12. Wang, X.; Chen, L.; Yang, Y. Establishment of core collection for Chinese tea germplasm based on cultivated region grouping and phenotypic data. *Front. Agric. China* **2011**, *5*, 344–350. [[CrossRef](#)]
13. Ni, S.; Yao, M.-Z.; Chen, L.; Zhao, L.-P.; Wang, X.-C.; Apostolides, Z.; Chen, Z.-M. Germplasm and Breeding Research of Tea Plant Based on DNA Marker Approaches. In *Uncertainty Modeling for Data Mining*; Springer Science and Business Media LLC: Berlin, Germany, 2012; pp. 361–376.
14. Gunasekare, M.; Ranatunga, M.; Piyasundara, J.H.N.; Kottawa-Arachchi, J. Tea genetic resources in Sri Lanka: Collection, Conservation and Appraisal. *Int. J. Tea Sci.* **2012**, *8*, 51–60.
15. Kottawa-Arachchi, J.D.; Gunasekare, M.T.K.; Ranatunga, M.A.B. Biochemical diversity of global tea [*Camellia sinensis* (L.) O. Kuntze] germplasm and its exploitation: A review. *Genet. Resour. Crop. Evol.* **2018**, *66*, 259–273. [[CrossRef](#)]
16. Ullah, J.; Shah, A.; Nisar, M.; Khan, U.; Khan, J.; Ahmad, H.; Rokhan, G.; Jabalok, B. Biochemical characterization of lentil germplasm for genetic diversity. *Plant Cell Biotech. Mol. Biol.* **2016**, *17*, 7–13.
17. Xia, E.-H.; Tong, W.; Wu, Q.; Wei, S.; Zhao, J.; Zhang, Z.-Z.; Wei, C.; Wan, X. Tea plant genomics: Achievements, challenges and perspectives. *Hortic. Res.* **2020**, *7*, 1–19. [[CrossRef](#)] [[PubMed](#)]
18. Chen, L.; Zhou, Z.X.; Chen, L.; Zhou, Z.X. Variations of main quality components of tea genetic resources [*Camellia sinensis* (L.) O. Kuntze] preserved in the China National Germplasm Tea Repository. *Plant Foods Hum. Nutr.* **2005**, *60*, 31–35. [[CrossRef](#)]
19. Yü, Y.; Deng, Y.; Lu, B.-M.; Liu, Y.-X.; Li, J.; Bao, J.-K. Green tea catechins: A fresh flavor to anticancer therapy. *Apoptosis* **2013**, *19*, 1–18. [[CrossRef](#)]
20. Higdon, J.V.; Frei, B. Tea Catechins and Polyphenols: Health Effects, Metabolism, and Antioxidant Functions. *Crit. Rev. Food Sci. Nutr.* **2003**, *43*, 89–143. [[CrossRef](#)] [[PubMed](#)]
21. Musial, C.; Kuban-Jankowska, A.; Gorska-Ponikowska, M. Beneficial Properties of Green Tea Catechins. *Int. J. Mol. Sci.* **2020**, *21*, 1744. [[CrossRef](#)] [[PubMed](#)]
22. Masek, A. Antioxidant and Antiradical Properties of Green Tea Extract Compounds. *Int. J. Electrochem. Sci.* **2017**, *12*, 6600–6610. [[CrossRef](#)]
23. Matsumoto, S.; Kiriwa, Y.; Yamaguchi, S. The Korean Tea Plant (*Camellia sinensis*): RFLP Analysis of Genetic Diversity and Relationship to Japanese Tea. *Breed. Sci.* **2004**, *54*, 231–237. [[CrossRef](#)]
24. Wachira, F.; Tanaka, J.; Takeda, Y. Genetic variation and differentiation in tea (*Camellia sinensis*) germplasm revealed by RAPD and AFLP variation. *J. Hortic. Sci. Biotechnol.* **2001**, *76*, 557–563. [[CrossRef](#)]
25. Kaundun, S.S.; Zhyvoloup, A.; Park, Y.-G. Evaluation of the genetic diversity among elite tea (*Camellia sinensis* var. *sinensis*) accessions using RAPD markers. *Euphytica* **2000**, *115*, 7–16. [[CrossRef](#)]
26. Chen, L.; Yamaguchi, S. RAPD markers for discriminating tea germplasms at the inter-specific level in China. *Plant Breed.* **2005**, *124*, 404–409. [[CrossRef](#)]
27. Young-Goo, P.; Kaundun, S.S.; Zhyvoloup, A. Use of the bulked genomic DNA-based RAPD methodology to assess the genetic diversity among abandoned Korean tea plantations. *Genet. Resour. Crop. Evol.* **2002**, *49*, 159–165. [[CrossRef](#)]
28. Lee, K.J.; Lee, J.-R.; Raveendar, S.; Shin, M.-J.; Kim, S.-H.; Cho, G.-T.; Hyun, D.Y. Assessment of Genetic Diversity of Tea Germplasm for Its Management and Sustainable Use in Korea Genebank. *Forest* **2019**, *10*, 780. [[CrossRef](#)]
29. Meegahakumbura, M.K.; Wambulwa, M.C.; Li, M.-M.; Thapa, K.K.; Sun, Y.; Moller, M.; Xu, J.-C.; Yang, J.-B.; Liu, J.; Liu, B.-Y.; et al. Domestication Origin and Breeding History of the Tea Plant (*Camellia sinensis*) in China and India Based on Nuclear Microsatellites and cpDNA Sequence Data. *Front. Plant Sci.* **2018**, *8*. [[CrossRef](#)]
30. Fang, W.; Cheng, H.; Duan, Y.; Jiang, X.; Li, X. Genetic diversity and relationship of clonal tea (*Camellia sinensis*) cultivars in China as revealed by SSR markers. *Plant Syst. Evol.* **2011**, *298*, 469–483. [[CrossRef](#)]
31. Korir, N.K.; Han, J.; Shangguan, L.; Wang, C.; Kayesh, E.; Zhang, Y.; Fang, J. Plant variety and cultivar identification: Advances and prospects. *Crit. Rev. Biotechnol.* **2012**, *33*, 111–125. [[CrossRef](#)] [[PubMed](#)]
32. Lai, J.A.; Yang, W.C.; Hsiao, J.Y. An assessment of genetic relationships in cultivated tea clones and native wild tea in Taiwan using RAPD and ISSR markers. *Bot. Bull. Acad. Sin.* **2001**, *42*, 93–100.
33. Hammer, O.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *4*, 1–9.

34. Ori, F.; Ma, J.-Q.; Gori, M.; Lenzi, A.; Chen, L.; Giordani, E. DNA-based diversity of tea plants grown in Italy. *Genet. Resour. Crop. Evol.* **2017**, *64*, 1905–1915. [CrossRef]
35. Kamvar, Z.N.; Tabima, J.F.; Grünwald, J. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2014**, *2*, e281. [CrossRef]
36. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics* **2012**, *28*, 2537–2539. [CrossRef]
37. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [PubMed]
38. Earl, D.A.; Vonholdt, B.M. Structure Harvester: A website and program for visualizing structure output and implementing the Evanno method. *Conserv. Genet. Resour.* **2011**, *4*, 359–361. [CrossRef]
39. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef] [PubMed]
40. Ivandic, V.; Hackett, C.A.; Nevo, E.; Keith, R.; Thomas, W.T.; Forster, B.P. Analysis of simple sequence repeats (SSRs) in wild barley from the Fertile Crescent: Associations with ecology, geography and flowering time. *Plant Mol. Biol.* **2002**, *48*, 511–527. [CrossRef]
41. Jombart, T. ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [CrossRef] [PubMed]
42. Kim, K.-W.; Chung, H.-K.; Cho, G.-T.; Ma, K.-H.; Chandrabalan, D.; Gwag, J.-G.; Kim, T.-S.; Cho, E.-G.; Park, Y.-J. PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* **2007**, *23*, 2155–2162. [CrossRef] [PubMed]
43. Hu, J.; Xu, H.M.; Zhu, J. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* **2000**, *101*, 264–268. [CrossRef]
44. Perrier, X.; Jacquemoud-Collet, J.P. DARwin Software. 2006. Available online: <http://darwin.cirad.fr> (accessed on 9 September 2020).
45. Ahmed, S.; Stepp, J.R. Chapter 5—Pu-erh Tea: Botany, Production, and Chemistry. In *Tea in Health and Disease Prevention*; Preedy, V.R., Ed.; Academic Press: Cambridge, MA, USA, 2013; pp. 59–71.
46. Anand, J.; Upadhyaya, B.; Rawat, P.; Rai, N. Biochemical characterization and pharmacognostic evaluation of purified catechins in green tea (*Camellia sinensis*) cultivars of India. *3 Biotech* **2014**, *5*, 285–294. [CrossRef]
47. Feng, L.; Gao, M.-J.; Hou, R.-Y.; Hu, X.-Y.; Zhang, L.; Wan, X.-C.; Wei, S. Determination of quality constituents in the young leaves of albino tea cultivars. *Food Chem.* **2014**, *155*, 98–104. [CrossRef] [PubMed]
48. Gai, Z.; Wang, Y.; Jiang, J.; Xie, H.; Ding, Z.; Ding, S.; Wang, H. The Quality Evaluation of Tea (*Camellia sinensis*) Varieties Based on the Metabolomics. *HortScience* **2019**, *54*, 409–415. [CrossRef]
49. Mondal, T.K. *Breeding and Biotechnology of Tea and Its Wild Species*; Springer Science and Business Media LLC: Berlin, Germany, 2014; pp. 1–167.
50. Chen, Z.-M.; Chen, L.; Apostolides, Z. Delicious and Healthy Tea: An Overview. In *Uncertainty Modeling for Data Mining*; Springer Science and Business Media LLC: Berlin, Germany, 2012; pp. 1–11.
51. Yao, M.Z.; Chen, L.; Liang, Y.R. 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 Breed.* **2008**, *127*, 166–172. [CrossRef]
52. Jin, J.-Q.; Ma, J.-Q.; Ma, C.-L.; Yao, M.-Z.; Chen, L. Determination of Catechin Content in Representative Chinese Tea Germplasm. *J. Agric. Food Chem.* **2014**, *62*, 9436–9441. [CrossRef] [PubMed]
53. Punyasiri, P.N.; Jeganathan, B.; Kottawa-Arachchi, J.D.; Ranatunga, M.A.B.; Abeysinghe, I.S.B.; Gunasekare, M.T.K.; Bandara, B.M.R. New Sample Preparation Method for Quantification of Phenolic Compounds of Tea (*Camellia sinensis* L. Kuntze): A Polyphenol Rich Plant. *J. Anal. Methods Chem.* **2015**, *2015*, 964341. [CrossRef] [PubMed]
54. Gulati, A.; Rajkumar, S.; Karthigeyan, S.; Sud, R.K.; Vijayan, D.; Thomas, J.; Rajkumar, R.; Das, S.C.; Tamuly, P.; Hazarika, M.; et al. Catechin and Catechin Fractions as Biochemical Markers to Study the Diversity of Indian Tea (*Camellia sinensis* (L.) O. Kuntze) Germplasm. *Chem. Biodivers.* **2009**, *6*, 1042–1052. [CrossRef] [PubMed]
55. Wei, K.; Wang, L.; Zhou, J.; He, W.; Zeng, J.; Jiang, Y.; Cheng, H. Catechin contents in tea (*Camellia sinensis*) as affected by cultivar and environment and their relation to chlorophyll contents. *Food Chem.* **2011**, *125*, 44–48. [CrossRef]
56. Robertson, A. The chemistry and biochemistry of black tea production—The non-volatiles. In *Tea: Cultivation to Consumption*; Willson, K.C., Clifford, M.N., Eds.; Springer: Dordrecht, The Netherlands, 1992; pp. 555–601.

57. Koch, W.; Kukula-Koch, W.; Komsta, Ł.; Marzec, Z.; Szwer, W.; Głowniak, K. Green Tea Quality Evaluation Based on Its Catechins and Metals Composition in Combination with Chemometric Analysis. *Molecules* **2018**, *23*, 1689. [[CrossRef](#)]
58. Leung, L.K.; Su, Y.; Chen, R.; Zhang, Z.; Huang, Y.; Chen, Z.-Y. Theaflavins in Black Tea and Catechins in Green Tea Are Equally Effective Antioxidants. *J. Nutr.* **2001**, *131*, 2248–2251. [[CrossRef](#)]
59. Chen, Z.-Y.; Zhu, Q.Y.; Tsang, D.; Huang, Y. Degradation of Green Tea Catechins in Tea Drinks. *J. Agric. Food Chem.* **2001**, *49*, 477–482. [[CrossRef](#)]
60. Yao, L.; Caffin, N.; D'Arcy, B.; Jiang, Y.; Shi, J.; Singanusong, R.; Liu, X.; Datta, N.; Kakuda, Y.; Xu, Y. Seasonal Variations of Phenolic Compounds in Australia-Grown Tea (*Camellia sinensis*). *J. Agric. Food Chem.* **2005**, *53*, 6477–6483. [[CrossRef](#)] [[PubMed](#)]
61. Gill, M. Speciality and herbal teas. In *Tea*; Springer Science and Business Media LLC: Berlin, Germany, 1992; pp. 513–534.
62. Tanaka, J.; Chen, L.; Apostolides, Z.; Chen, Z.-M. *Japanese Tea Breeding History and the Future Perspective*; Springer Science and Business Media LLC: Berlin, Germany, 2012; pp. 227–239.
63. Vigouroux, Y.; Glaubitz, J.C.; Matsuoka, Y.; Goodman, M.M.; Sánchez, G.J.; Doebley, J. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am. J. Bot.* **2008**, *95*, 1240–1253. [[CrossRef](#)]
64. Yao, M.-Z.; Ma, C.-L.; Qiao, T.-T.; Jin, J.-Q.; Chen, L. Diversity distribution and population structure of tea germplasms in China revealed by EST-SSR markers. *Tree Genet. Genomes* **2011**, *8*, 205–220. [[CrossRef](#)]
65. Zhang, Y.; Zhang, X.; Chen, X.; Sun, W.; Li, J. Genetic diversity and structure of tea plant in Qinba area in China by three types of molecular markers. *Hereditas* **2018**, *155*, 22. [[CrossRef](#)]
66. Campoy, J.A.; Lerigoleur, E.; Christmann, H.; Beauvieux, R.; Girollet, N.; Quero-García, J.; Dirlwanger, E.; Barreneche, T. Genetic diversity, linkage disequilibrium, population structure and construction of a core collection of *Prunus avium* L. landraces and bred cultivars. *BMC Plant Biol.* **2016**, *16*, 49. [[CrossRef](#)] [[PubMed](#)]
67. Lee, K.J.; Lee, J.-R.; Sebastin, R.; Cho, G.-T.; Hyun, D.Y. Molecular Genetic Diversity and Population Structure of Ginseng Germplasm in RDA-Genebank: Implications for Breeding and Conservation. *Agronomy* **2020**, *10*, 68. [[CrossRef](#)]
68. Liu, W.; Shahid, M.Q.; Bai, L.; Lu, Z.; Chen, Y.; Jiang, L.; Diao, M.; Liu, X.; Lu, Y. Evaluation of Genetic Diversity and Development of a Core Collection of Wild Rice (*Oryza rufipogon* Griff.) Populations in China. *PLoS ONE* **2015**, *10*, e0145990. [[CrossRef](#)]
69. Belaj, A.; Dominguez-García, M.D.C.; Atienza, S.G.; Urdíroz, N.M.; De La Rosa, R.; Satovic, Z.; Martín, A.; Kilian, A.; Trujillo, I.; Valpuesta, V.; et al. Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DARs, SSRs, SNPs) and agronomic traits. *Tree Genet. Genomes* **2011**, *8*, 365–378. [[CrossRef](#)]
70. Kumar, S.; Ambreen, H.; Variath, M.T.; Rao, A.R.; Agarwal, M.; Kumar, A.; Goel, S.; Jagannath, A. Utilization of Molecular, Phenotypic, and Geographical Diversity to Develop Compact Composite Core Collection in the Oilseed Crop, Safflower (*Carthamus tinctorius* L.) through Maximization Strategy. *Front. Plant Sci.* **2016**, *7*. [[CrossRef](#)]
71. Le, D.S.; Pagès, J. Analyse factorielle multiple hiérarchique. *Rev. Stat. Appl.* **2003**, *51*, 47–73.
72. Yanfang, Z.; DeChang, H.; Jincheng, Z.; Ping, Z.; Zhaohong, W.; Chuanjie, C. Development of a mulberry core collection originated in China to enhance germplasm conservation. *Crop. Breed. Appl. Biotechnol.* **2019**, *19*, 55–61. [[CrossRef](#)]
73. Lee, H.-Y.; Ro, N.-Y.; Jeong, H.-J.; Kwon, J.-K.; Jo, J.; Ha, Y.; Jung, A.; Han, J.-W.; Venkatesh, J.; Kang, B.-C. Genetic diversity and population structure analysis to construct a core collection from a large Capsicum germplasm. *BMC Genet.* **2016**, *17*, 1–13. [[CrossRef](#)]

74. Galwey, N.W. Verifying and validating the representativeness of a core collection. In *Core Collections of Plant Genetic Resources*; Hodgkin, T.B.A., van Hintum, T.J.L., Morales, E.A.V., Eds.; John Wiley and Sons: Chichester, UK, 2005; pp. 187–198.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).