Screening of yield and growth-related SSR markers in Philippine coffee

Miguel V. Mondragon*[1,2], Dominique Anne B. Bongala[2], Ernelea P. Cao[2], and Daisy May C. Santos[1,2]

[1]Natural Sciences Research Institute, College of Science, University of the Philippines, Diliman, Quezon City
[2]Institute of Biology, College of Science, University of the Philippines, Diliman, Quezon City

ABSTRACT

Yield is a complex trait affected by numerous factors, such as disease resistance, bean size and shape. Molecular markers such as simple sequence repeats (SSRs) can identify quantitative trait loci (QTL) associated with yield. For this study, SSR markers related to yield and growth, such as genes coding for certain protein kinases and ligases, were mined from the available *Coffea canephora* reference genome. A total of 100 SSR markers were screened using *C. arabica* (commonly known as Arabica coffee), *C. canephora* (commonly called as Robusta coffee), *C. liberica* var. *liberica* (locally known as Liberica coffee or Kapeng Barako) and *C. liberica* var. *dewevrei* (locally called as Kapeng Excelsa) as samples. These types have their own unique growth and yield characteristics making them suitable for comparison. Of the 82 loci successfully amplified, 46 loci were polymorphic. A total of 66 bands from 37 loci were specific to certain types. Nineteen (19) bands were specific to Arabica samples, five for Robusta, three for Liberica, and another three for Excelsa. The other remaining bands were shared by a combination of two or three types. Forty-six bands were found near genes annotated as enzymes that promote plant growth and the ability to respond and adapt to environmental signals like light, hormones and stress factors. The bands specific to Robusta might explain its moderate resiliency to abiotic stress conditions and better yield compared to the other species. These bands are indicative of genetic differences in the yield-related target loci. As such, these markers have great potential in marker-assisted selection for yield and growth improvement of Philippine coffee. This study is the first to identify potential yield and growth-related SSR markers in Philippine coffee types.

INTRODUCTION

Coffee consumption in the Philippines has been steadily increasing since 1990. In contrast, coffee production has been widely inconsistent. During the crop year 2019 – 2020, coffee consumption was at 3.250 million 60-kg bags, while coffee production was only at 0.307 million 60-kg bags (Retrieved from https://www.ico.org/new_historical.asp on 15 March 2023). This wide gap between local supply and demand strains exportability, which in turn affects profitability of the Philippine coffee industry.

Maximizing productivity while minimizing loss leads to high yield outcomes. Aside from abiotic and biotic factors, yield is also influenced by numerous genetic factors which directly or indirectly affect it. In *Coffea arabica*, it has been shown that the average green bean yield is positively correlated with the

KEYWORDS

Coffea, Philippine coffee, simple sequence repeats (SSRs), yield, growth

---

*Corresponding author
Email Address: mvmondragon@up.edu.ph
Date received: March 15, 2023
Date revised: May 21, 2023
Date accepted: June 1, 2023

following: hundred green bean weight, plant height, stem diameter, length of leaves and primary branches, resistance to coffee berry disease and percentage of bearing primary branches (Atinafu and Mohammed 2017). Narrow sense heritability of bean weight has been found to be high in the case of *C. canephora*. This means that it is highly affected by additive genetic factors.

Numerous studies have been done trying to pinpoint regions responsible for yield. Simple sequence repeats (SSRs) are short, tandem repeats of DNA distributed throughout the genome (Wang et al. 2009). SSRs are a useful tool in identifying quantitative trait loci (QTL) responsible for a specific trait in question. It has been shown that SSR and EST-SSR markers from *C. canephora* are easily transferable across different species within *Coffea* (Poncet et al. 2007). This makes these markers a practical tool to use when dealing with *Coffea* species which have limited molecular information available, such as *C. liberica*. SSR markers are also relatively fast and inexpensive compared to newer methods, such as next-generation sequencing.

The whole genome sequences of *C. arabica* and *C. canephora* are both available today. The genetic closeness of *C. canephora* with *C. liberica* (Lashermes et al. 1997; Cros et al. 1998), as well as the origin of *C. arabica* from the hybridization of *C. canephora* and *C. eugenioides* (Scalabrin et al. 2020), make *C. canephora* a better template for a multi-species study. In 2014, the team of Denoeud was able to draft the whole genome of *C. canephora*, as well as annotate over 25,000 protein-coding genes (Denoeud et al. 2014). This provided a high-density genetic map for use in SSR data mining *in silico* and primer design as exhibited by the work of Ogutu et al. (2016).

In 2019, Santos et al. were able to mine and screen SSRs associated with aroma using the *C. canephora* genome. Using a similar method, it would also be possible to mine SSRs associated with yield and screen them using Philippine coffee types. Due to the complexity of yield as a trait, this study focused on genes which were involved with growth, such as plant protein kinases and ligases that eventually influence the amount of berries and therefore seeds (or the coffee beans) produced by the plant. Plant protein kinases are enzymes that facilitate the transfer of phosphate from ATP to the serine, threonine, or tyrosine residues of the protein substrate (Lehti-Shiu and Shiu 2012). These enzymes play a role in the ability of plants to respond and adapt to different signals such as light, hormones, and stress factors. (Stone and Walker 1995; Wang et al. 2020). On the other hand, protein ligases are involved in protein ubiquitination, degradation, and are associated with cellular responses to signals (Zheng and Shabek 2017). These enzymes are also involved in the signal transduction of plant hormones such as auxin (Robinson and Ardley 2004). Changes in growth, whether through overexpression of growth promoting hormones or an adaptive stress response, will in turn affect subsequent yield.

The Philippines commercially produces four types of coffee, namely, Arabica, Robusta, Liberica and Excelsa. Data on yield parameters of individual coffee trees are limited, therefore it is more practical to utilize the innate yield characteristics of the different *Coffea* species which make them distinguishable from one another. For example, Robusta exhibits high yield, in terms of grams of dried beans produced per tree, according to the National Seed Industry Council (NSIC) data (Retrieved from https://nsic.buplant.da.gov.ph/registry.php on 9 March 2023). On the other hand, Liberica has the largest bean and leaf size among the *Coffea* species (Gibson 2018).

Using the different Philippine coffee types, namely, Arabica, Robusta, Liberica, and Excelsa, this study aimed to screen and identify potential yield and growth-related SSR markers using the method of Santos et al. (2019). Differences in the band amplification sizes between the types would indicate genetic differences in the specific yield and growth-related locus. The scope of this study is limited to screening these yield and growth-related markers. Therefore, future studies would be needed to determine whether or not these genetic differences are responsible for the phenotypic differences in yield between these Philippine coffee types.

MATERIALS AND METHODS

Sample Collection and DNA Extraction

Young leaf samples from the Arabica type were collected from the Bureau of Plant Industry in Baguio City, Benguet. Three trees each from the following Arabica varieties, Yellow Bourbon, Red Bourbon, and Caturra, were used. Leaf samples from five trees each of the Robusta, Liberica, and Excelsa types were sourced from the National Coffee Research, Development and Extension Center of the Cavite State University (CvSU) in Indang, Cavite. The samples for each type represented replicates to verify the genetic homogeneity in each case. DNA from the young leaves of each type was extracted following the protocol from the DNeasy Plant Mini Kit (QIAGEN). Purity and concentration of the extracted DNA were assessed using the Epoch Microplate Spectrophotometer (BioTek).

SSR Loci, Primer Design, and Filtering of Yield and Growth-related Primer Pairs

Yield and growth-related SSR loci were determined using the same methods employed by Santos et al. (2019), with a few minor modifications stated below. In the filtration part, primer pairs which hit or have at least one base pair overlap with exons of genes involved with growth and yield based on the annotations by Denoeud et al. (2014) were collated. One hundred primer pairs were then randomly chosen and synthesized by a commercial company.

Amplification of SSR Primers

One hundred SSR primers located in genes related to yield and growth were screened using PCR (Supplementary Table 1). The PCR mixture contained 10 µL of the following: 2.0 µL 5x MyTaq Reaction buffer, 0.5 µL 10 µM primer, 0.05 µL of 5 U/µL MyTaq Polymerase, and 0.3 µL 20ng/uL DNA. The following PCR conditions were used: initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 55 °C for 30 s, and extension at 72°C for 2 min and a final extension at 72°C for 7 min. Initial screening was done by running the resulting amplicons through 1% agarose gel. Amplified samples were then subjected through 10% polyacrylamide gel using a voltage of 85V for 85 min for better resolution.

Data Analysis

Based on the polyacrylamide gel photographs, amplicon size was measured using the software, GelAnalyzer 19.1 (Lazar and Lazar 2019). The SSR primers were evaluated based on amplification success, as well as the number and sizes of each amplicon.

RESULTS AND DISCUSSION

Mining of the yield and growth-related SSRs has identified a total of 1255 loci consisting of 66 dinucleotide repeats, 151 trinucleotide repeats, 681 tetranucleotide repeats, 253 pentanucleotide repeats, and 104 hexanucleotide repeats. A total of 1143 primer pairs were designed for these loci. From this list, 100 primer pairs were randomly chosen for screening. Only 82 primer pairs were successfully amplified. However, five of these pairs were not able to amplify all the Liberica and Excelsa samples, while one pair was not able to amplify Robusta samples. After PAGE examination, 46 of these primer pairs were found to be polymorphic, which were then used for further analysis. Table 1 summarizes the details of each of these loci. A range of two to nine alleles per loci were found in each of these.

Table 1: List of loci related to yield and growth found to be polymorphic in *Coffea arabica, C. canephora, C. liberica* var. *liberica and C. liberica* var. *dewevrei.*

SSR MARKER	MOTIF	Gene	Gene Description	Primer Sequence	Number of Alleles	Size Range
Y9	(CCAT)2	Cc00_g18440	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	F: aagacgggcttgaatcctgg R: agccgtagtgttgatcaggc	4	292-369
Y11	(CCAT)2	Cc00_g18440	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	F: aagacgggcttgaatcctgg R: tgcaactgggagcttcacat	9	428-934
Y12	(CTTTC)2	Cc00_g18440	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	F: acatggtgggacgatcactg R: gttccttgcgtttctcagcc	5	433-911
Y17	(GATT)2	Cc00_g18470	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	F: agggacgttcactagggtca R: aggcatggagattgtggcaa	4	141-294
Y19	(TAATT)2	Cc01_g00650	Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD3-1	F: aaggcacgctttcagttgac R: gcagtcatcgcgcatgataaa	4	384-569
Y22	(TTAC)2	Cc01_g00880	Hypothetical protein	F: agcttgggctggagatttgg R: tgcatcttcatcagtgcccat	5	527-638
Y24	(AAAAG)4	Cc01_g01850	E3 ubiquitin-protein ligase SINAT2	F: aaggccagatcaaccgaacc R: tgccttgactaagtcggtcg	4	282-412
Y25	AC(11.5)	Cc01_g01850	E3 ubiquitin-protein ligase SINAT2	F: actgacaactgcagccactt R: atggctgtagctgaggaagc	6	483-497
Y26	(AAAC)2	Cc01_g01850	E3 ubiquitin-protein ligase SINAT2	F: aggcgaaattggcaaagctg R: tgttgtagggtgcacaagca	3	392-433
Y27	(TCCA)2	Cc01_g02750	E3 ubiquitin-protein ligase SINAT2	F: aagagtgccaagtctcgacg R: atgcgagtacttgtgtggca	2	370-422
Y29	(TTTCT)2	Cc01_g02750	E3 ubiquitin-protein ligase SINAT2	F: tgtttcctagctcatcgcgg R: aggattgttgcctagcgtgt	6	414-1005
Y32	(GCGACC)2	Cc01_g02810	Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g11300	F: ccgagtcgtatctggaacgg R: tccgccagtaaggtctacca	2	472-477
Y33	(GGAT)2	Cc01_g02940	G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	F: aacggaaggaggtagaccga R: ccaacataccgaccactgct	3	176-369
Y34	(TTGGC)2	Cc01_g02940	G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	F: accgtaactgtggaccaagc R: ctttcaatgcagcagcccag	3	263-416
Y35	(GGAT)2	Cc01_g02940	G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	F: agtggcgccaagtttgtcta R: ccaacataccgaccactgct	3	393-447
Y36	AAAATAAAATAAAA; (GTA)4	Cc02_g01560	E3 ubiquitin-protein ligase SINAT3	F: aaacgagccctgcatagcaa R: gttaaggatcgccgttgagc	5	418-458
Y38	AT(10.5)	Cc02_g02140	G-type lectin S-receptor-like serine/threonine-protein kinase At5g24080	F: aagcagaagacttccggtgg R: cagattgacaccctgccagt	4	252-460
Y39	AT(10.5)	Cc02_g02140	G-type lectin S-receptor-like serine/threonine-protein kinase At5g24080	F: aagcagaagacttccggtgg R: ccagtgcttgagcgttcttg	4	178-426
Y41	(CAAG)2	Cc02_g11990	Putative G-type lectin S-receptor-like serine/threonine-protein kinase B120	F: agagggtggtcatttcgcag R: taaacggatgctgggagagc	5	365-742
Y42	(TCGAT)2	Cc02_g12000	G-type lectin S-receptor-like serine/threonine-protein kinase RKS1	F: aagttgggtcggaaattcctct R: tctgcaagcataggagctgt	3	444-523
Y43	(AGA)3	Cc02_g12000	G-type lectin S-receptor-like serine/threonine-protein kinase RKS1	F: aatcacccgtcccaggatca R: ggacagtcccgtaaatggca	2	325-331
Y44	TC(8.5)	Cc02_g12010	Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD1-1	F: actcagtcagcccttgcttg R: cgggagttcattgggctcat	6	283-488
Y45	(TGGT)2	Cc02_g12010	Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD1-1	F: actcagtcagcccttgcttg R: tgggagagctcaaatcaccg	5	458-614
Y50	(TGGA)2	Cc03_g12580	Putative receptor protein kinase ZmPK1	F: aagaagcggagctgaccaaa R: agtttgaaccaccgaagcca	2	200-246
Y52	(GTTCA)2	Cc03_g12580	Putative receptor protein kinase ZmPK1	F: acgggcctgagatatcgagt R: cctgaggattggaccgactg	2	378-450

Y54	(CT)4	Cc03_g12580	Putative receptor protein kinase ZmPK1	F: cgctgaagtgaaccaccaga R: gttcgccatacaaccatgcc	3	256-507
Y59	(TATCTG)2	Cc03_g13490	G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	F: acaggaggactaggctcagg R: ggtttgcagttgtccgtcac	2	356-368
Y62	(TATCTG)2	Cc03_g13490	G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	F: caccaccaccacgtactacc R: ggtttgcagttgtccgtcac	2	473-485
Y63	(TCGG)2	Cc03_g13490	G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	F: cagttgaccgtagccgatga R: ggacttctccaagccagacc	2	130-582
Y66	(TCT)3	Cc04_g07480	Putative receptor protein kinase ZmPK1	F: aactgctcacgccctaaaca R: tgcacgcagctaatggaaga	3	366-500
Y68	(CTGAT)2	Cc04_g07480	Putative receptor protein kinase ZmPK1	F: agttggaaatctggctcgca R: gcctgacccatggaaacgta	3	358-486
Y73	(TTGTGG)2; (ATGC)2; (CGGT)2	Cc04_g07480	Putative receptor protein kinase ZmPK1	F: gagcagggaaagcctggaat R: tctgcatgactgcaacaggt	4	434-800
Y76	(TTGCAG)2; (TCCTT)2	Cc04_g13490	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	F: acaggttaaggtgccagcaa R: agccaatctgcagttcacca	2	388-407
Y79	(CTTTT)2; (TCCTT)2	Cc04_g13490	G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	F: agggtcgtttgccacattga R: ccagattcaagctgtgctgc	4	433-994
Y86	(CTTT)2; (CTT)3; (CTT)3	Cc05_g05720	G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130	F: ttgccacagtcaagagggaa R: tggagacgagggtctcatca	4	490-981
Y88	(TCC)3	Cc05_g08960	Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5	F: ttcagcagcagttgtcacca R: gctgcactctgcaggtatga	2	366-478
Y99	(CCACTT)2	Cc05_g12870	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	F: gtggtgtcctgggtgttcaa R: ctcgaggaatgcagaaccgt	4	463-637
Y105	GGA(5.0)	Cc06_g05290	Putative phytosulfokine 3 precursor	F: acgtggctcgaactcgtaat R: tgatcgctgaggctcacttg	4	450-468
Y107	AGC(6.0)	Cc06_g05290	Putative phytosulfokine 3 precursor	F: tggttggtagaagacgagcg R: ggacggagggcgtattactt	2	395-404
Y123	(ATA)3	Cc07_g08260	Putative E3 ubiquitin-protein ligase SINA-like 10	F: aaacagcctgcctaccactc R: aagtgggttgatagcgctcc	4	279-476
Y133	(TGTGAT)2; (GCT)3	Cc07_g17480	Putative G-type lectin S-receptor-like serine/threonine-protein kinase RLK1	F: acagcatgactcctgtgagc R: cctgaatccctttgtcgcct	2	207-478
Y135	(TGATTT)2	Cc07_g17500	Putative G-type lectin S-receptor-like serine/threonine-protein kinase RLK1	F: agcttgcagaaactacacgga R: tgcctgaaggtgaaagccat	8	184-558
Y139	(ATTTCT)2	Cc08_g02480	Putative Eyes absent homolog 4	F: ataccgccaccgggttatag R: acactggttgactgaaggtcat	3	374-468
Y163	(AGA)3	Cc09_g02700	G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	F: gctgcaatgcgacagaaaca R: acgtcagggagttgaagctg	5	78-532
Y180	CTTTCTTTCTTTC	Cc11_g05110	G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	F: catcttcggtgggctctctc R: gatacagaagtccgcgagca	4	167-740
Y199	(ACTCAA)2	Cc11_g06400	Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	F: cgtggagatcattgagccga R: acaaccaaggttggaatcacct	4	341-472

Monomorphic markers, or those which amplify bands of similar sizes, give little information in this type of study. On the other hand, polymorphic markers, or those which amplify bands of different sizes, and the presence or absence of amplification are more informational. Different sized bands or the absence of amplification would indicate that there are genetic differences between the samples. Small changes in size may indicate an increase or decrease in the repeating elements, while changes of a 100 bp or more might indicate insertions or deletions within the target region. As these SSR markers have been pre-screened to be located in or near yield-related genes, these genetic changes might have an effect on the phenotypic characteristics of the types involved. For example, a band specific to Liberica may or may not be located near the gene responsible for its large bean size.

A total of 66 bands from 37 loci were found to be specific to certain types (Table 2). Nineteen (19) bands were specific to Arabica samples, five bands for Robusta, three bands for Liberica, and another three bands for Excelsa (Table 2). Seventeen (17) bands were shared by both Arabica and Robusta, while 15 bands were shared by both Liberica and Excelsa. One band from SSR Y86 was shared by Arabica and Excelsa. There is also one band from SSR Y127 which was shared by all types except for Robusta, and another band from SSR Y105 which was shared by all except for Arabica. Other interesting discoveries include the two bands at SSR Y99 which were shared by Arabica samples (Figure S1), and a band at SSR Y27 specific to the Yellow Bourbon variety of Arabica (Figure S2). The high number of bands shared by Arabica and Robusta was expected because of the close relationship between the two. *C. arabica* originated from the hybridization of *C. canephora* and *C. eugenioides* (Scalabrin et al. 2020). Likewise, shared bands between Liberica and Excelsa were also expected as they fall under the same species, *C. liberica* (Davis et al. 2006). However, it was noteworthy to discover bands which were specific to either Liberica and Excelsa. Morphological differences in the leaf have been found between these two types (Cao et al. 2014; Baltazar and Buot 2019), however genetic differences have been limited except for markers such as AFLPs (N'diaye et al. 2005) and ISSRs (Panaligan et al. 2020).

Table 2: List of loci that distinguishes between the four coffee types (Arabica, Robusta, Liberica, and Excelsa).

SSR Marker	Type
Y9 302bp	Liberica/Excelsa
Y9 369bp	Robusta
Y11 551bp	Arabica
Y11 580bp	Arabica
Y11 655bp	Arabica
Y11 697bp	Arabica
Y11 805bp	Liberica/Excelsa
Y11 889bp	Liberica/Excelsa
Y12 439bp	Robusta
Y12 788bp	Arabica
Y12 899bp	Arabica
Y17 158bp	Arabica
Y17 258bp	Arabica
Y17 290bp	Arabica/Robusta
Y19 465bp	Robusta
Y19 511bp	Arabica
Y19 562bp	Arabica
Y25 486bp	Liberica
Y26 428bp	Liberica/Excelsa
Y27 422bp	Arabica(Yellow Bourbon)
Y28 527bp	Arabica/Robusta
Y29 418bp	Liberica/Excelsa
Y32 472bp	Arabica/Robusta
Y34 265bp	Arabica/Robusta
Y34 413bp	Arabica
Y36 418bp	Arabica/Robusta
Y36 433bp	Arabica/Robusta
Y36 441bp	Liberica/Excelsa
Y36 458bp	Excelsa
Y38 278bp	Liberica
Y39 366bp	Arabica/Robusta
Y39 411bp	Arabica/Robusta
Y42 479bp	Liberica/Excelsa
Y43 331bp	Robusta
Y44 476bp	Arabica/Robusta
Y52 400bp	Liberica/Excelsa
Y54 504bp	Robusta
Y59 357bp	Arabica/Robusta
Y62 473bp	Arabica/Robusta
Y63 133bp	Liberica/Excelsa
Y66 493bp	Liberica
Y68 475bp	Arabica
Y76 388bp	Liberica/Excelsa
Y79 483bp	Liberica/Excelsa
Y79 698bp	Liberica/Excelsa
Y86 618bp	Excelsa
Y86 874bp	Arabica/Excelsa
Y86 970bp	Excelsa
Y99 478bp	Arabica/Robusta
Y99 521bp	Liberica/Excelsa
Y99 585bp	Arabica
Y99 637bp	Arabica
Y105 451bp	Arabica/Robusta
Y105 459bp	Arabica
Y105 465bp	Robusta/Liberica/Excelsa
Y107 404bp	Arabica/Robusta
Y123 289bp	Arabica/Robusta
Y123 310bp	Liberica/Excelsa
Y123 365bp	Arabica
Y127 236bp	Arabica/Liberica/Excelsa
Y135 231bp	Liberica/Excelsa
Y163 519bp	Arabica/Robusta
Y180 170bp	Arabica
Y180 592bp	Arabica
Y180 718bp	Arabica
Y199 358bp	Arabica/Robusta

It was also interesting to find a band specific to the Yellow Bourbon variety of Arabica. Genetic differences within Arabica are rare because of autogamy and its nature as an allotetraploid (Anthony et al. 2002). Therefore, studying this band further might enable us to understand the genetic nuances between the different Arabica varieties and might also explain the distinct yellow cherry coloration of the Yellow Bourbon variety.

The first samples of each type, except for Excelsa, are NSIC-registered individuals. Succeeding samples were from trees derived from either tissue culture or cuttings of these registered individuals. Yield data, represented as dried bean grams per tree per year, was 3070 and 2700 for CvSU *Kapeng Manipis* (Robusta) and CvSU Barako (Liberica), respectively. Meanwhile, yield data from the BPI Baguio Arabica individuals was 714.10, 614, and 293.30 for Red Bourbon, Caturra, and Yellow Bourbon, respectively (Retrieved from https://nsic.buplant.da.gov.ph/registry.php on 9 March 2023). Average yield of the 18 registered Robusta trees nationwide is 1854 g/tree, ranging from 990 g/tree up to 3462 g/tree. On the other hand, average yield for the seven Arabica individuals is 687.08 g/tree, ranging from 293.30 g/tree up to 987.21 g/tree. Unfortunately, only one Liberica tree is registered in the NSIC. From this data, we can observe the apparent difference in yield between the different *Coffea* species.

The main markers found in this study were associated with genes described as G-type lectin S-receptor-like serine/threonine-protein kinases, putative receptor protein kinase ZmPK1, E3 ubiquitin-protein ligases, and putative phytosulfokine 3 precursor.

In a study by Sun et al. (2013), the overexpression of a G-type lectin S-receptor-like serine/threonine-protein kinase from *Glycine soja* showed growth benefits to the plant as it promoted higher yield, higher chlorophyll content and higher plant height under salt stress in *Arabidopsis*. A total of 46 specific bands were found near genes annotated as G-type lectin S-receptor-like serine/threonine-protein kinases, while another four bands were near genes described as putative receptor protein kinase ZmPK1. The bands specific to Robusta might explain its moderate resiliency to abiotic stress conditions compared to the other species. Robusta also has a wider tolerable altitude range and greater distribution (Davis et al. 2006).

E3 ubiquitin ligases have been found to be involved in a variety of plant biological processes such as floral development, pollen tube extension, ethylene biosynthesis, stress response, and hormone signaling among others (Mazzucotelli et al., 2006). Twelve specific bands were observed near genes designated as E3 ubiquitin-protein ligases. These might explain the slight differences in the reproductive development between the species. One study recorded the development from inflorescence up to fruit maturation of these four coffee types (Salazar et al. 2019). Duration of the major stages, such as inflorescence emergence, flowering, development of fruit, and ripening of fruit and seed, slightly vary between the species. A longer maturation period was observed for Liberica and Excelsa, possibly due to their relatively larger fruits compared to Arabica and Robusta. Duration from start of inflorescence up to fruit maturation was 364 d, 416 d, 448 d, and 455 d, for Arabica, Robusta, Excelsa, and Liberica, respectively (Salazar et al. 2019).

Phytosulfokine (PSK) is a plant peptide which promotes cell proliferation, growth and cellular longevity *in vitro* and *in vivo* (Yang et al. 2001; Matsubayashi et al. 2006). Aside from these growth-promoting functions, PSK also regulates plant defense against bacterial and fungal infections (Mosher et al. 2013). PSK mediates between the switch to either plant growth or defense in response to environmental changes (Igarashi et al. 2012; Ding et al., 2022). One study has found that overexpression of a certain PSK precursor gene increases root growth and grain yield in rice (Geng et al. 2020). In this study, four bands were found near a putative phytosulfokine 3 precursor gene. A band specific for Robusta, Liberica and Excelsa might be responsible for the increased disease resistance of these species compared to Arabica. On the other hand, the absence of two bands in Liberica and Excelsa indicates a distinguishable molecular characteristic compared to other coffee species. Such molecular feature might explain their characteristic larger leaf and fruit phenotype when compared with other species.

Further studies into these genes might give us a better understanding on how these specific genes are involved and if these genetic differences are translated into downstream processes.

CONCLUSION

This study is the first to identify potential yield and growth-related SSR markers in the Philippine coffee types. Although not exhaustive, the study has identified a few genes of interest, such as G-type lectin S-receptor-like serine/threonine-protein kinases, putative receptor protein kinase ZmPK1, E3 ubiquitin-protein ligases, and a putative phytosulfokine 3 precursor. Within or near these genes, the study has found bands specific for each of the different types, as well as some bands which are shared by two or three of these types. It is still unclear if the genetic differences in these genes are directly involved in the complex phenotypic expressions of yield in these coffee types. Further investigation of these potential markers might be beneficial for use in marker-assisted selection. This would greatly benefit the Philippine coffee industry in improving our coffee stocks as well as promoting our local coffee, “Kapeng Barako”. The current trend in genetic studies is to utilize whole genome sequencing in characterizing species and understanding traits. This study has shown that SSR markers are still an effective tool for these kinds of studies with the added benefit of being fast and inexpensive.

ACKNOWLEDGMENT

The authors would like to thank the following: Natural Sciences and Research Institute, College of Science, UP Diliman (Project code: BIO-21-1-03) and the Philippine Council for Agriculture, Aquatic, and Natural Resources Research and Development of the Department of Science and Technology (PCAARRD DOST) for the research funding; Institute of Biology, College of Science, UP Diliman; Dr. Ruel M. Mojica of the National Coffee Research Development and Extension Center of the Cavite State University in Indang, Cavite, and the Bureau of Plant Industry in Baguio City, Benguet for allowing us to collect coffee leaf samples.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

Anthony F, Combes M, Astorga C, Bertrand B, Graziosi G, Lashermes P. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theoretical and Applied Genetics 2002; 104(5):894-900.

Atinafu G, Mohammed H. Association and path coefficient analysis of yield and yield attributes of coffee (*Coffea arabica* L.) under Sidama Specialty Coffee Growing Area, Awada, Southern Ethiopia. Advances in Crop Science and Technology 2017; 5: 307.

Baltazar A, Buot, Jr I. Leaf architectural analysis of taxonomically confusing coffee species: *Coffea liberica* and *Coffea liberica* var. *dewevrei*. Biodiversitas 2019; 20(6):1560-1567.

Cao E, Constantino-Santos D, Ramos L, Santos B, Quilang J, Mojica R. Molecular and morphological differentiation among *Coffea* (Rubiaceae) varieties grown in the farms of Cavite Province, Philippines. Philippine Science Letters 2014; 7(2):387- 397.

Cros J, Combes M, Trouslot P, Anthony F, Hamon S, Charrier A, Lashermes P. Phylogenetic analysis of chloroplast DNA variation in *Coffea* L. Molecular Phylogenetics and Evolution 1998; 9:109-117.

Davis A, Govaerts R, Bridson D, Stoffelen P. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). Botanical Journal of the Linnean Society 2006; 152:465-512.

Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G, Aury J, Bento P, Bernard M, Bocs S, Campa C, Cenci A, Combes M, Crouzillat D, Da Silva C, Daddiego L, De Bellis F, Dussert S, Garsmeur O, Gayraud T, Guignon V, Jahn K, Jamilloux V, Joët T, Labadie K, Lan T, Leclercq J, Lepelley M, Leroy T, Li L, Librado P, Lopez L, Muñoz A, Noel B, Pallavicini A, Perrotta G, Poncet V, Pot D, Priyono, Rigoreau M, Rouard M, Rozas J, Tranchant-Dubreuil C, Vanburen R, Zhang Q, Andrade A, Argout X, Bertrand B, De Kochko A, Graziosi G, Henry R, Jayarama, Ming R, Nagai C, Rounsley S, Sankoff D, Giuliano G, Albert V, Wincker P, Lashermes P. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 2014; 345: 1181–1184.

Ding S, Lv J, Hu Z, Wang J, Wang P, Yu J, Foyer C, Shi K. Phytosulfokine peptide optimizes plant growth and defense via glutamine synthetase GS2 phosphorylation in tomato. The EMBO Journal 2022; e111858: 1-17.

Geng Y, Jian C, Xu W, Liu H, Hao C, Hou J, Liu H, Zhang X, Li T. miR164-targeted TaPSK5 encodes a phytosulfokine precursor that regulates root growth and yield traits in common wheat (*Triticum aestivum* L.). Plant Molecular Biology 2020; 104(6):615-628.

Gibson M. Tea and Coffee. In: Food Science and Culinary Arts. Elsevier Inc., 2018: 353-372.

Igarashi D, Tsuda K, Katagiri F. The peptide growth factor, phytosulfokine, attenuates pattern-triggered immunity. The Plant Journal 2012; 71: 194–204.

Lashermes P, Combes M, Trouslot P, Charrier A. Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA. Theoretical and Applied Genetics 1997; 94:947-955.

Lazar Jr I, Lazar Sr I. 2019. GelAnalyzer 19.1. Accessed September 22, 2022, from http://www.gelanalyzer.com

Lehti-Shiu MD, Shiu SH. Diversity, classification and function of the plant protein kinase superfamily. Philosophical Transactions of the Royal Society B: Biological Sciences 2012; 367(1602): 2619-2639.

Matsubayashi Y, Ogawa M, Kihara H, Niwa M, Sakagami Y. Disruption and overexpression of *Arabidopsis* phytosulfokine receptor gene affects cellular longevity and potential for growth. Plant Physiology 2006; 142: 45–53

Mazzucotelli E, Belloni S, Marone D, De Leonardis AM, Guerra D, Di Fonzo N, Cattivelli L, Mastrangelo AM. The E3 ubiquitin ligase gene family in plants: regulation by degradation. Current Genomics 2006; 7(8): 509-522.

Mosher S, Seybold H, Rodriguez P, Stahl M, Davies K, Dayaratne S, Morillo S, Wierzba M, Favery B, Keller H, Tax F, Kemmerling B. The tyrosine-sulfated peptide receptors PSKR1 and PSY1R modify the immunity of *Arabidopsis* to biotrophic and necrotrophic pathogens in an antagonistic manner. The Plant Journal 2013; 73: 469–482.

N'diaye A, Poncet V, Louarn J, Hamon S, Noirot M. Genetic differentiation between *Coffea liberica* var. *liberica and C. liberica* var. *dewevrei* and comparison with *C. canephora*. Plant Systematics and Evolution 2005; 253:95-104.

Ogutu C, Fang T, Yan L, Wang L, Huang L, Wang X, Ma B, Deng X, Owiti A, Nyende A, Han Y. Characterization and utilization of microsatellites in the *Coffea canephora* genome to assess genetic association between wild species in Kenya and cultivated coffee. Tree Genetics & Genomes 2016; 12(3): 54.

Panaligan A, Baltazar M, Alejandro G. Genetic polymorphism of registered and popularly cultivated coffee (*Coffea* spp.) in the Philippines using inter-simple sequence repeats markers. Biodiversitas 2020; 21:4228–4233.

Poncet V, Dufour M, Hamon P, Hamon S, De Kochko A, Leroy T. Development of genomic microsatellite markers in *Coffea canephora* and their transferability to other coffee species. Genome 2007; 50:1156–1161.

Robinson PA, Ardley HC. Ubiquitin-protein ligases. Journal of Cell Science 2004; 117(22): 5191-5194.

Salazar B, Gunda D, Lagrimas AJ, Santos PJ, Del Rosario E. Profiling and analysis of reproductive phenology of four coffee (*Coffea* spp.) species in the Philippines using the BBCH Scale. Philippine Journal of Crop Science 2019; 44(3):10-19.

Santos D, Victoria A, Besa C, Cao E. Genome-wide profiling and characterization of terpene synthase-linked simple sequence repeats in *Coffea canephora* Pierre ex A. Froehner for identification of potential markers for aroma in Philippine-grown coffee varieties. Philippine Journal of Science: Special Issue on Genomics 2019; 148(S1):267-273.

Scalabrin S, Toniutti L, Di Gaspero G, Scaglione D, Magris G, Vidotto M, Pinosio S, Cattonaro F, Magni F, Jurman I, Cerutti M, Liverani F, Navarini L, Del Terra L, Pellegrino G, Ruosi M, Vitulo N, Valle G, Pallavicini A, Graziosi G, Klein P, Bentley N, Murray S, Solano W, Al Hakimi A, Schilling T, Montagnon C, Morgante M, Bertrand B. A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. Scientific Reports 2020; 10(4642):1-13.

Stone JM, Walker JC. Plant protein kinase families and signal transduction. Plant Physiology 1995; 108(2): 451-457.

Sun XL, Yu QY, Tang LL, Ji W, Bai X, Cai H, Liu XF, Ding XD, Zhu YM. GsSRK, a G-type lectin S-receptor-like serine/threonine protein kinase, is a positive regulator of plant tolerance to salt stress. Journal of Plant Physiology 2013; 170(5): 505-515.

Wang M, Barkley N, Jenkins T. Microsatellite markers in plants and insects, Part I: Applications of biotechnology. Genes, Genomes and Genomics 2009; 3(1): 1.

Wang P, Hsu CC, Du Y, Zhu JK. Mapping proteome-wide targets of protein kinases in plant stress responses. Proceedings of the National Academy of Sciences 2020; 117(6): 3270-3280.

Yang H, Matsubayashi Y, Nakamura K, Sakagami Y. Diversity of *Arabidopsis* genes encoding precursors for phytosulfokine, a peptide growth factor. Plant Physiology 2001; 127: 842–851.

Zheng N, Shabek N. Ubiquitin ligases: structure, function, and regulation. Annual Review of Biochemistry 2017; 86: 129-157.

SUPPLEMENTARY DATA

Table S1: List of yield-linked SSR markers screened for this study.

SSR CODE	Loc.	Motif	Gene	Primer sequences	Est. Product Size	Amplification rate	AGE Profile	PAGE profile	No. of alleles	Remarks
Y1	chr0	ACT(5.0)	Cc00_g15910 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD3-1	aaccaagagggacggaaacc tccttcactttctcggcagt	317	0%				
Y2	chr0	(AAGTA)2; (CCAT)2	Cc00_g15910 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD3-1	aactggctcctccagtaggt gcctgtttggactgttggtg	381	100%	Monomorphic	Monomorphic	1	
Y3	chr0	(AAGTA)2; (CCAT)2	Cc00_g15910 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD3-1	aactggctcctccagtaggt tgcctgtttggactgttggt	382	100%	Monomorphic	Monomorphic	1	
Y4	chr0	(ATC)3	Cc00_g15910 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD3-1	accaacagtccaaacaggca ggtttccgtccctcttggtt	187	100%	Polymorphic	Monomorphic	1	
Y5	chr0	(ATTGT)2	Cc00_g12500 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	acaataagcagcacccacca ttagcaagccactcaggagc	398	46%				
Y6	chr0	(CACC)2	Cc00_g12500 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	acagctggagggaattgagc cgatacctgtggtccttggg	393	100%	Monomorphic	Monomorphic	1	
Y7	chr0	(ATTGT)2; (CACC)2	Cc00_g12500 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	acagctggagggaattgagc ttagcaagccactcaggagc	450	50%				
Y8	chr0	(TGG)3	Cc00_g12500 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	agcgtggatagttgtggtcg ccctgctcaattccctccag	297	100%	Monomorphic	Monomorphic	1	
Y9	chr0	(CCAT)2	Cc00_g18440 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	aagacgggcttgaatcctgg agccgtagtgttgatcaggc	302	100%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y10	chr0	(CCAT)2	Cc00_g18440 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	aagacgggcttgaatcctgg gccgtagtgttgatcaggct	301	100%	Monomorphic	Polymorphic	4	Sequence very similar to Y9

Y11	chr0	(CCAT)2	Cc00_g18440 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	aagacgggcttgaatcctggtgcaactgggagcttcacat	447	100%	Monomorphic	Polymorphic	9	Scored for subsequent analysis
Y12	chr0	(CTTTC)2	Cc00_g18440 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	acatggtgggacgatcactggttccttgcgtttctcagcc	467	100%	Monomorphic	Polymorphic	5	Scored for subsequent analysis
Y13	chr0	(ACGG)2	Cc00_g18470 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	aaggatcgccacagttaggcctggctgcctgatttggact	431	42%				
Y14	chr0	(ACGG)2	Cc00_g18470 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	aaggatcgccacagttaggctggctgcctgatttggactt	430	96%	Monomorphic	Monomorphic	1	
Y15	chr0	(CCAT)2	Cc00_g18470 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	agacaggatgtccggtttccagtcaggttccgttgagctg	497	100%	Monomorphic	Monomorphic	1	
Y16	chr0	(CCAT)2	Cc00_g18470 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	agacaggatgtccggtttcctcagtcaggttccgttgagc	499	96%	Monomorphic	Monomorphic	1	
Y17	chr0	(GATT)2	Cc00_g18470 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	agggacgttcactagggtcaaggcatggagattgtggcaa	141	100%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y18	chr1	(ACTG)2; (TTG)3	Cc01_g00650 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD3-1	aagcagttgcagctcaaggtcccaaggcagaaacagcaac	430	71%				
Y19	chr1	(TAATT)2	Cc01_g00650 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD3-1	aaggcacgctttcagttgacgcagtcatcgcgcatgataaa	361	100%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y20	chr1	(AAAC)2	Cc01_g00650 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD3-1	aagtggttcagggtgccaaaccttgagctgcaactgcttg	312	100%	Monomorphic	Monomorphic	1	
Y21	chr1	(TTAC)2	Cc01_g00880 Hypothetical protein	agcttgggctggagatttggctgcatcttcatcagtgccc	500	0%				
Y22	chr1	(TTAC)2	Cc01_g00880 Hypothetical protein	agcttgggctggagatttggtgcatcttcatcagtgcccat	499	100%	Polymorphic	Polymorphic	5	Scored for subsequent analysis

Y23	chr1	(TTAC)2	Cc01_g00880 Hypothetical protein	agcttgggctggagatttgg tgcatcttcatcagtgccca	499	0%				
Y24	chr1	AAAAG(4.0)	Cc01_g01850 E3 ubiquitin-protein ligase SINAT2	aaggccagatcaaccgaacc tgccttgactaagtcggtcg	269	92%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y25	chr1	AC(11.5)	Cc01_g01850 E3 ubiquitin-protein ligase SINAT2	actgacaactgcagccactt atggctgtagctgaggaagc	492	96%	Monomorphic	Polymorphic	6	Scored for subsequent analysis
Y26	chr1	(AAAC)2	Cc01_g01850 E3 ubiquitin-protein ligase SINAT2	aggcgaaattggcaaagctg tgttgtagggtgcacaagca	344	96%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y27	chr1	(TCCA)2	Cc01_g02750 E3 ubiquitin-protein ligase SINAT2	aagagtgccaagtctcgacg atgcgagtacttgtgtggca	349	96%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y28	chr1	(AT)4	Cc01_g02750 E3 ubiquitin-protein ligase SINAT2	gcggattggagggaacattg gccagcctaacacctctctg	494	58%	Monomorphic	Monomorphic	1	Liberica/Excelsa not amplified
Y29	chr1	(TTTCT)2	Cc01_g02750 E3 ubiquitin-protein ligase SINAT2	tgtttcctagctcatcgcgg aggattgttgcctagcgtgt	450	100%	Polymorphic	Polymorphic	6	Scored for subsequent analysis
Y30	chr1	(CCGATA)2	Cc01_g02810 Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g11300	aatctacccggaacagcagc ccgttccagatacgactcgg	338	0%				
Y31	chr1	(CTTG)2	Cc01_g02810 Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g11300	acatgcctccaccctcaatg acagttcccgtcgcatcatc	271	79%				
Y32	chr1	(GCGACC)2	Cc01_g02810 Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g11300	ccgagtcgtatctggaacgg tccgccagtaaggtctacca	477	58%	Monomorphic	Polymorphic	2	Liberica/Excelsa not amplified Scored for subsequent analysis
Y33	chr1	(GGAT)2	Cc01_g02940 G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	aacggaaggaggtagaccga ccaacataccgaccactgct	348	100%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y34	chr1	(TTGGC)2	Cc01_g02940 G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	accgtaactgtggaccaagc ctttcaatgcagcagcccag	360	100%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y35	chr1	(GGAT)2	Cc01_g02940 G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	agtggcgccaagtttgtcta ccaacataccgaccactgct	408	100%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y36	chr2	AAAATAAAATAAAA; GTAGTAGTAGTA	Cc02_g01560 E3 ubiquitin-protein ligase SINAT3	aaacgagccctgcatagcaa gttaaggatcgccgttgagc	427	92%	Polymorphic	Polymorphic	5	Scored for subsequent analysis

Y37	chr2	(TA)4; (GATA)3	Cc02_g01560 E3 ubiquitin-protein ligase SINAT3	aagaagcactccctgcatgt gtgcgatttgaagcagcgaa	482	92%	Monomorphic	Monomorphic	1	
Y38	chr2	AT(10.5)	Cc02_g02140 G-type lectin S-receptor-like serine/threonine-protein kinase At5g24080	aagcagaagacttccggtgg cagattgacaccctgccagt	353	96%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y39	chr2	AT(10.5)	Cc02_g02140 G-type lectin S-receptor-like serine/threonine-protein kinase At5g24080	aagcagaagacttccggtgg ccagtgcttgagcgttcttg	338	58%	Monomorphic	Polymorphic	4	Liberica/Excelsa not amplified Scored for subsequent analysis
Y40	chr2	(ATTG)2	Cc02_g11990 Putative G-type lectin S-receptor-like serine/threonine-protein kinase B120	aaagtcgggacaaggcttcg aggccttgagcaataccgtc	327	100%	Monomorphic	Monomorphic	1	
Y41	chr2	(CAAG)2	Cc02_g11990 Putative G-type lectin S-receptor-like serine/threonine-protein kinase B120	agagggtggtcatttcgcag taaacggatgctgggagagc	363	92%	Monomorphic	Polymorphic	5	Scored for subsequent analysis
Y42	chr2	(TCGAT)2	Cc02_g12000 G-type lectin S-receptor-like serine/threonine-protein kinase RKS1	aagttgggtcggaaattcctct tctgcaagcataggagctgt	443	100%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y43	chr2	(AGA)3	Cc02_g12000 G-type lectin S-receptor-like serine/threonine-protein kinase RKS1	aatcacccgtcccaggatca ggacagtcccgtaaatggca	326	96%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y44	chr2	TC(8.5)	Cc02_g12010 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD1-1	actcagtcagcccttgcttg cgggagttcattgggctcat	303	96%	Monomorphic	Polymorphic	6	Scored for subsequent analysis
Y45	chr2	(TGGT)2	Cc02_g12010 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD1-1	actcagtcagcccttgcttg tgggagagctcaaatcaccg	463	96%	Monomorphic	Polymorphic	5	Scored for subsequent analysis
Y46	chr2	(ATTT)2	Cc02_g12020 Putative Receptor-like serine/threonine-protein kinase SD1-8	aaactggtccgaaaccacct aagaccttcaagcgaccagg	246	46%				
Y47	chr2	(TTCT)2	Cc02_g12020 Putative Receptor-like serine/threonine-protein kinase SD1-8	aaggttgtggtctggatggc aacaggaagggttgttggca	253	100%	Monomorphic	Monomorphic	1	
Y48	chr2	(CCTC)2	Cc02_g39500 Exportin-2	acaaacccaactacggcctc taacggccagagcttcactg	261	100%	Monomorphic	Monomorphic	1	

Y49	chr2	GGATGAGGATGAGGA	Cc02_g39500 Exportin-2	acaactgtcagcacgagtgt aaatgaagtcccacccgctt	375	100%	Monomorphic	Monomorphic	1	
Y50	chr3	(TGGA)2	Cc03_g12580 Putative receptor protein kinase ZmPK1	aagaagcggagctgaccaaa agtttgaaccaccgaagcca	203	100%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y51	chr3	(TACC)2	Cc03_g12580 Putative receptor protein kinase ZmPK1	acaatggcttcggtggttca gcccttgacctttatggcct	483	54%				
Y52	chr3	(GTTCA)2	Cc03_g12580 Putative receptor protein kinase ZmPK1	acgggcctgagatatcgagt cctgaggattggaccgactg	429	100%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y53	chr3	(TGGA)2	Cc03_g12580 Putative receptor protein kinase ZmPK1	caggctcggagttctcaagg acgttgccacttgcttccta	300	100%	Monomorphic	Monomorphic	1	
Y54	chr3	(CT)4	Cc03_g12580 Putative receptor protein kinase ZmPK1	cgctgaagtgaaccaccaga gttcgccatacaaccatgcc	447	100%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y55	chr3	(GATT)2	Cc03_g12580 Putative receptor protein kinase ZmPK1	gacagcgacaacgctttgaG gttgaatctggccttgcagc	438	100%	Monomorphic	Monomorphic	1	
Y56	chr3	(GTTCA)2	Cc03_g12580 Putative receptor protein kinase ZmPK1	gacagcgacaacgctttgag cctgaggattggaccgactg	460	100%	Monomorphic	Monomorphic	1	
Y57	chr3	(GATT)2	Cc03_g12580 Putative receptor protein kinase ZmPK1	gcatggttgtatggcgaacc actcgatatctcaggcccgt	278	100%	Monomorphic	Monomorphic	1	
Y58	chr3	(TATC)2	Cc03_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	aactgacgaaggtgacgagg gctcctgagtggcttgctaa	456	96%	Monomorphic	Monomorphic	1	
Y59	chr3	(TATCTG)2	Cc03_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	acaggaggactaggctcagg ggtttgcagttgtccgtcac	357	100%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y60	chr3	(AGG)4	Cc03_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	actgccagacaagggagttg ggacttctccaagccagacc	225	100%	Monomorphic	Monomorphic	1	
Y61	chr3	(TATC)2	Cc03_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	agatccctctgtgacaccct gctcctgagtggcttgctaa	366	96%	Monomorphic	Monomorphic	1	
Y62	chr3	(TATCTG)2	Cc03_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	caccaccaccacgtactacc ggtttgcagttgtccgtcac	477	100%	Monomorphic	Polymorphic	2	Scored for subsequent analysis

Y63	chr3	(TCGG)2	Cc03_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	cagttgaccgtagccgatga ggacttctccaagccagacc	496	92%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y64	chr3	CACCACCACCAC	Cc03_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	ccctgctcaattccctccag acctgagcctagtcctcctg	360	0%				
Y65	chr3	(GGTG)2	Cc03_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	cgatacctgtggtccttggg acagctggagggaattgagc	393	96%	Monomorphic	Monomorphic	1	
Y66	chr4	(TCT)3	Cc04_g07480 Putative receptor protein kinase ZmPK1	aactgctcacgccctaaaca tgcacgcagctaatggaaga	463	96%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y67	chr4	(ATTT)2	Cc04_g07480 Putative receptor protein kinase ZmPK1	agtctccttggctccattgc gcactgcgatgtaaagcctc	294	100%	Monomorphic	Monomorphic	1	
Y68	chr4	(CTGAT)2	Cc04_g07480 Putative receptor protein kinase ZmPK1	agttggaaatctggctcgca gcctgacccatggaaacgta	381	100%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y70	chr4	(AGA)3	Cc04_g07480 Putative receptor protein kinase ZmPK1	caacatcggctgtgcacttt gcccatgagaccggaagaaa	418	96%	Monomorphic	Monomorphic	1	
Y73	chr4	(TTGTGG)2; (ATGC)2; (CGGT)2	Cc04_g07480 Putative receptor protein kinase ZmPK1	gagcagggaaagcctggaat tctgcatgactgcaacaggt	485	100%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y76	chr4	(TTGCAG)2; (TCCTT)2	Cc04_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	acaggttaaggtgccagcaa agccaatctgcagttcacca	408	92%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y79	chr4	(CTTTT)2; (TCCTT)2	Cc04_g13490 G-type lectin S-receptor-like serine/threonine-protein kinase At4g03230	agggtcgtttgccacattga ccagattcaagctgtgctgc	427	92%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y82	chr5	(ATC)3	Cc05_g05720 G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130	aggccatctatgggtcaggt gccaactgcatcattcgtgt	365	92%	Monomorphic	Monomorphic	1	
Y86	chr5	(CTTT)2; (CTT)3; (CTT)3	Cc05_g05720 G-type lectin S-receptor-like serine/threonine-protein kinase At2g19130	ttgccacagtcaagagggaa tggagacgagggtctcatca	471	100%	Monomorphic	Polymorphic	4	Scored for subsequent analysis

Y88	chr5	(TCC)3	Cc05_g08960 Putative G-type lectin S-receptor-like serine/threonine-protein kinase SD2-5	ttcagcagcagttgtcacca gctgcactctgcaggtatga	382	100%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y99	chr5	(CCACTT)2	Cc05_g12870 G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	gtggtgtcctgggtgttcaa ctcgaggaatgcagaaccgt	442	100%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y104	chr6	(TCCTAG)2	Cc06_g02170 Putative receptor protein kinase ZmPK1	atctccatccttgcacggtc tttcctgcgtggaggaagac	187	0%				
Y105	chr6	GGA(5.0)	Cc06_g05290 Putative PHYTOSULFOKINE 3 PRECURSOR	acgtggctcgaactcgtaat tgatcgctgaggctcacttg	467	92%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y107	chr6	AGC(6.0)	Cc06_g05290 Putative PHYTOSULFOKINE 3 PRECURSOR	tggttggtagaagacgagcg ggacggagggcgtattactt	404	58%	Monomorphic	Polymorphic	2	Liberica/Excelsa not amplified Scored for subsequent analysis
Y112	chr6	(AGA)3	Cc06_g15660 E3 ubiquitin-protein ligase SINAT2	ctttccaaatgcgccctgtc tctacatggccttcctccga	246	63%				
Y123	chr7	(ATA)3	Cc07_g08260 Putative E3 ubiquitin-protein ligase SINA-like 10	aaacagcctgcctaccactc aagtgggttgatagcgctcc	283	92%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y124	chr7	GCTGCTGCTGCTG	Cc07_g08260 Putative E3 ubiquitin-protein ligase SINA-like 10	cccatctctattgccgtccc tctctgggagatgaagcgga	348	100%	Monomorphic	Monomorphic	1	
Y127	chr7	(CTC)3	Cc07_g15480 Putative G-type lectin S-receptor-like serine/threonine-protein kinase RLK1	aatgcaagcagaccaaggga ggttgattgcaactgctggg	237	75%	Monomorphic	Monomorphic	1	Robusta not amplified
Y129	chr7	(GAT)3	Cc07_g15490 Putative G-type lectin S-receptor-like serine/threonine-protein kinase RLK1	cagtgacttcggcttggcta ttctgaccgggatgttgtcg	477	100%	Monomorphic	Monomorphic	1	
Y133	chr7	(TGTGAT)2; (GCT)3	Cc07_g17480 Putative G-type lectin S-receptor-like serine/threonine-protein kinase RLK1	acagcatgactcctgtgagc cctgaatccctttgtcgcct	452	100%	Monomorphic	Polymorphic	2	Scored for subsequent analysis
Y134	chr7	(TGCT)3; TGCTTGCTTGCTTG	Cc07_g17480 Putative G-type lectin S-receptor-like serine/threonine-protein kinase RLK1	cccacagcttggtactctcg aggccagacgattggtcatc	407	100%	Monomorphic	Monomorphic	1	
Y135	chr7	(TGATTT)2	Cc07_g17500 Putative G-type lectin S-receptor-like serine/threonin	agcttgcagaaactacacgga tgcctgaaggtgaaagccat	312	100%	Polymorphic	Polymorphic	8	Scored for subsequent analysis

			e-protein kinase RLK1							
Y139	chr8	(ATTTCT)2	Cc08_g02480 Putative Eyes absent homolog 4	ataccgccaccgggttatag acactggttgactgaaggtcat	386	100%	Monomorphic	Polymorphic	3	Scored for subsequent analysis
Y141	chr8	(TTTTAT)2; TTTTATTTTTATTTT	Cc08_g02480 Putative Eyes absent homolog 4	tgcatatccaccaaggtgct tttcagcacttcaccacact	479	0%				
Y145	chr8	(ACAGCA)2	Cc08_g04420 Putative Receptor-like serine/threonine-protein kinase SD1-8	acattggacatggtaggccg caggcttgggacttgtggat	152	100%	Monomorphic	Monomorphic	1	
Y150	chr8	(ACT)3	Cc08_g14820 Putative UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase	attgaggagacgttgggctg cacacctccctaacgccaat	364	100%	Monomorphic	Monomorphic	1	
Y161	chr9	(TAC)3	Cc09_g02680 Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	accacagttcgaaccagcaa ttagacgaccaaccggcaaa	497	71%	Polymorphic		Too many to count	
Y163	chr9	(AGA)3	Cc09_g02700 G-type lectin S-receptor-like serine/threonine-protein kinase At1g11330	gctgcaatgcgacagaaaca acgtcagggagttgaagctg	479	75%	Monomorphic	Polymorphic	5	Liberica/Excelsa not amplified Scored for subsequent analysis
Y170	chr10	(GAAGCT)2; (ATC)3	Cc10_g05160 Putative receptor protein kinase ZmPK1	cccatgaatcgagtaccgca gatggctaaccgagaccacc	452	100%	Monomorphic	Monomorphic	1	
Y171	chr10	(ATC)3	Cc10_g05160 Putative receptor protein kinase ZmPK1	cccatgaatcgagtaccgca ttgggcgactaacacagctt	353	79%	Monomorphic	Monomorphic	1	
Y172	chr10	ATATATATATA	Cc10_g07550 Putative Phytosulfokine	tcagcgttggatgaacgaca ccaaggagcggaggatgaag	484	63%	Polymorphic	Polymorphic	Too many to count	
Y180	chr11	CTTTCTTTCTTTC	Cc11_g05110 G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	catcttcggtgggctctctc gatacagaagtccgcgagca	423	100%	Monomorphic	Polymorphic	4	Scored for subsequent analysis
Y182	chr11	GA(14.0)	Cc11_g05120 G-type lectin S-receptor-like serine/threonine-protein kinase At4g27290	accagatgagcaacttacaaca cttgtgaccgatttgggttcc	500	75%	Polymorphic	Polymorphic	Too many to count	
Y199	chr11	(ACTCAA)2	Cc11_g06400 Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	cgtggagatcattgagccga acaaccaaggttggaatcacct	336	96%	Monomorphic	Polymorphic	4	Scored for subsequent analysis

Y200	chr11	(TATG)3	Cc11_g06400 Putative G-type lectin S-receptor-like serine/threonine-protein kinase At1g34300	gcaagccgtgttcaacttgtagtacctggagaccagagcc	367	100%	Monomorphic	Monomorphic	1	

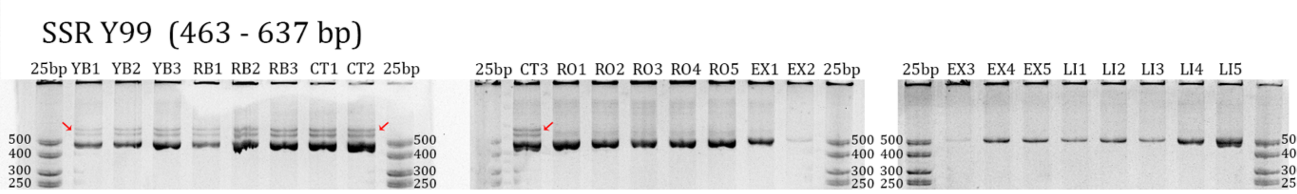

Figure S1: Polyacrylamide gel electrophoresis result showing the PCR products from the amplification of SSR Y99 primers using samples from Arabica varieties, namely Yellow Bourbon (YB), Red Bourbon (RB), and Caturra (Ct), as well as samples from Robusta (RO), Excelsa (EX), and Liberica (LI) varieties. The red arrows point to amplified bands (585 and 637 bp) specific to Arabica samples. Excelsa and Liberica samples also have shared specific bands (521 bp).

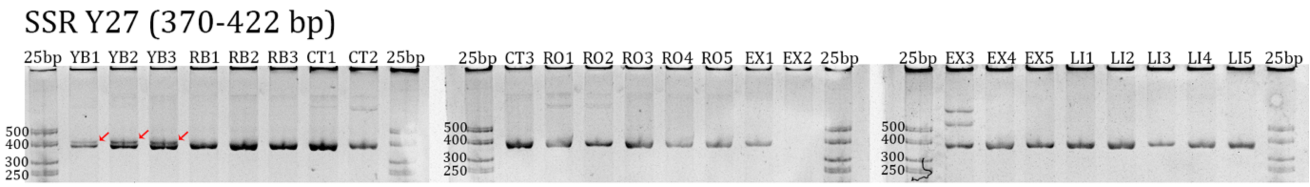

Figure S2: Polyacrylamide gel electrophoresis result showing the PCR products from the amplification of SSR Y27 primers using samples from Arabica varieties, namely Yellow Bourbon (YB), Red Bourbon (RB), and Caturra (Ct), as well as samples from Robusta (RO), Excelsa (EX), and Liberica (LI) varieties. The red arrows point to amplified bands (422 bp) specific to Yellow Bourbon samples.