Identification of true male parents in F1 populations of cacao using SSR markers

N K Izzah*, I Sulistiyorini and I N A Wicaksono

Indonesian Industrial and Beverage Crops Research Institute, IAARD, Jl. Raya Pakuwon Km. 2 Parungkuda, Sukabumi, 43357, Indonesia

Corresponding author email: lila_ref@yahoo.co.id

Abstract. Cacao is widely cultivated in Indonesia because of its economic benefits. The development of new varieties with desirable traits is required to meet industrial demands. In the present study, two hundred and four F1 cacao progenies have been generated from crossing between ten parental combinations to improve high-yielding varieties. However, progeny paternity of each clone remained undetermined. The study aimed to identify the male parents of F1 cacao populations using SSR markers. A total of 38 SSR markers were applied to screen polymorphism among ten parental combinations. Eleven polymorphic markers were used to amplify 204 cacao F1 hybrids. The genotype data were analyzed using the Cervus program to identify the true genotype of male parents. The result showed that 87 F1 progenies were identified as the true genotype of male parents at a 95% confidence level. Therefore, out of 204 F1 progenies, 87 F1 progenies have been identified their true male parents' identity. This study demonstrated the utility of SSR markers to detect the true identity of male parents, which helps breeders select F1 progenies known to their parents' identities.

1. Introduction

Cacao (*Theobroma cacao* L.) is one of the most popular estate crops among farmers in Indonesia due to its high economic value. The main part of cacao needed by the industrial sector is cocoa beans that can be processed for various products. Currently, high-yielding cacao varieties have become a major priority to increase cacao production as expected by farmers. The government has released many cacao varieties, including Sulawesi 1, Sulawesi 2, MCC 01, MCC 02, and BL 50. Of which many of them were obtained through participatory plant breeding (PPB) as a strategy to strengthen on-farm conservation by encouraging and consolidating the role of farmers in setting breeding goals and selecting diverse genetic materials [1]. On the other hand, assembling new high-yielding varieties can also be achieved by forming a set of the population generated from crossing between superior clones followed by selection. Selection can be carried out individually to get new superior clones and the population to get new superior hybrids [2].

Crossing between two superior clones is an important step in obtaining superior F1 progenies from both parents as well as increasing genetic diversity in nature or existing collections. A high level of genetic diversity obtained through crosses provides an opportunity to create new varieties with desirable traits. Therefore, selecting the ideal parental combination is crucial so that each parent's superior trait or heterosis effect can be inherited onto their progenies [3]. Breeders usually select parental clones from different heterotic groups and reduce blindness in the selection of crossing parents to increase breeding efficiency [4]. Theoretically, variation in the progeny would be high in crosses between genetically more...
distant parents due to the maximum number of segregating loci. The genotypic variation within a family indicates the informativeness of the family for genetic studies [5]. In addition, parental selection based on molecular fingerprinting helps breeders reduce crossing among related lines or reduce the chances of introgression of undesirable alleles [6].

Accurate parents’ information is very important in tree breeding programs, particularly plantation crops. It is pivotal because it can reduce genotyping costs through pedigree-based imputation, reduce genomic estimation bias of breeding values, and combine genotyped non-genotyped individuals into a joint analysis [7]. However, in the field, human errors often occur during plotting and planting in cacao plantations [8,9]. This condition led to several problems, such as the impurity of F1 hybrids resulting from artificial crosses and the diversity of cacao seeds distributed to the farmers. Hence, the solution is to use genetic data to reconstruct parent-progeny relationships in the parentage assignment [7].

Individuals’ parentage analysis can be identified at molecular levels through the application of DNA markers. Previous studies reported the success of applying genetic data generated from microsatellite markers or SNP arrays to analyze parentage assignment. In addition, many scientists have used microsatellite markers for distinguishing groups and quantifying the differences between individual species [10]. Here, microsatellite markers were used to obtain genetic data from 204 cacao F1 hybrids in the present study. This study aimed to identify the true male parents in F1 cacao populations using SSR markers. The information obtained from this study would be useful for breeders to assemble new high-yielding cacao varieties.

2. Materials and methods

2.1. Plant materials and DNA extraction

Two hundred and four individuals’ progenies generated from 10 cross combinations (Table 1) along with five parental clones (ICCRI 03, TSH 858, ICS 13, DR 1, and SCA 6) were used in this study. First, DNA was extracted from 3 g of fresh young leaves according to the modified cetyltrimethylammonium bromide (CTAB) method [11]. The extracted DNA was then measured its quality and quantity using NanoDrop ND-1000 (NanoDrop Technologies, Inc., Wilmington, DE, USA). Afterward, each DNA sample was diluted using 1x TE buffer solution until the concentration reached 10 ng/µl.

| No | Crossing combination | No. of progenies |
|----|----------------------|------------------|
| 1. | ICCRI 03 x TSH 858   | 16               |
| 2. | ICCRI 03 x DR 1      | 18               |
| 3. | ICCRI 03 x ICS 13    | 24               |
| 4. | ICCRI 03 x SCA 6     | 20               |
| 5. | TSH 858 x DR 1       | 23               |
| 6. | TSH 858 x ICS 13     | 25               |
| 7. | TSH 858 x SCA 6      | 24               |
| 8. | DR 1 x ICS 13        | 14               |
| 9. | DR 1 x SCA 6         | 25               |
| 10.| ICS 13 x SCA 6       | 15               |
|    | Total number         | 204              |

2.2. SSR markers analysis

A total of 38 SSR markers were used to amplify five parental clones of cacao (ICCRI 03, TSH 858, DR 1, ICS 13, and SCA 6). These 38 SSR markers consisted of 25 markers obtained from the Indonesian Center of Agricultural Biotechnology and Genetic Resources Research and Development (ICABIOGRAD). They can be accessed through the Indonesian Agricultural Genome Center (PGPI)
database (http://genom.litbang.pertanian.go.id), and 13 others were designed by [12] and [13]. PCR amplification was carried out with a total volume of 15 µl consisting of 10 ng DNA sample, 1x PCR mix, and SSR primers with a final concentration of 0.2 mM. The PCR amplification step was conducted as follows: pre denaturation at 94°C for 3 minutes, followed by 35 cycles consisting of denaturation at 94°C for 15 seconds, annealing at 55°C for 15 seconds, and extension at 72°C for 15 seconds, the next step is the final extension at 72°C for 10 minutes, and the process ends with cooling at 15°C for 5 minutes.

The amplification results were then checked on 1% agarose gel to determine the success of the amplification process. Finally, the PCR products were electrophoresed using 6% non-denatured polyacrylamide gel and 1x TBE buffer solution. The gels were stained with ethidium bromide solution for 20 minutes. Afterward, the DNA bands were visualized on a UV trans-illuminator lamp and documented with a digital camera.

2.3. **Genotyping of F1 cacao populations using polymorphic SSR markers**
The 204 F1 cacao populations were then targeted for genotyping using polymorphic SSR markers. The genotyping process of the 204 F1 cacao progenies was carried out using the standard protocols of PCR amplification and electrophoresis using 1% agarose, as well as electrophoresis using 6% non-denatured polyacrylamide gel. Genotyping analysis is the first step to determine the parental assignment of the F1 individual progenies. The true male parents were indicated by the type of allele in F1 progenies that corresponds to both parents.

2.4. **Data analysis**
The genotyping results of 204 cacao F1 populations were scored based on the allele type of both parents. The scoring results of all F1 progenies and all-male parents were analyzed to identify pollen donors. Next, molecular analysis was performed using CERVUS 2.0 program [14]. Paternity analysis was conducted in three steps, i.e., calculating allele frequency, undertaking simulation to determine the success rate of the candidate male parents, and choosing the selected male parents for each progeny. The molecular parameters analyzed were alleles frequency, Polymorphic Information Content (PIC), observed heterozygosity (Ho), and Expected heterozygosity (He) of the F1 cacao population. The next step was to conduct data simulation and parent analysis. The identity of male parents can be known by all parent with plus LOD method. The analysis results were then used to determine the identity of male parents based on the highest LOD value. The LOD value determines candidate male parents at 95% confidence level indicated by (*) symbol.

3. **Results and discussion**
3.1. **Screening of polymorphic markers**
Polymorphic DNA markers can be defined as variations of single base pair, multiple base pairs, and repeated sequences in a DNA sequence, contributing to different arrangements among individuals, groups, or populations [15]. Genetic studies usually require polymorphic DNA markers for their analysis. Hence, screening of polymorphic markers is an important step in genetic applications. In the present study, polymorphic SSR markers were screened using five cacao clones as parents, namely ICCRI 03, TSH 858, DR 1, ICS 13, and SCA 6 and 38 SSR markers. The five cacao clones were combined into ten parental combinations, as shown in Table 1. The result showed that 11 markers were polymorphic in the ten parental combinations (Table 2). These polymorphic markers were then applied to 204 F1 cacao progenies.
Table 2. Eleven SSR polymorphic markers.

| No | SSR loci | Forward sequences | Reverse sequences |
|----|----------|-------------------|-------------------|
| 1  | mTcCIR33 | TGGGTGGAAGATTTGGT | CAACAATGAAAATAGGCA |
| 2  | mTcCIR69 | TCGGTGTTCATCAGTA  | CATGCTATGAGATTTGAAG |
| 3  | mTcCIR1 | GCAGGCAGGCTCAGTGAAGCA | TGGCAACCCAGAAAACGAT |
| 4  | mTcCIR15 | CAGCGCCCTTTTGTAG | TATTTGGGATTCTTGATG |
| 5  | mTcCIR109 | GGAATGTAGGAGAAGATAG | GGACAAAAGAGACATA |
| 6  | mTcCIR24 | TTTGGGTGATTTCTTCTGA | TCTGTCTGCTTTTGTGA |
| 7  | mTcCIR167 | GTAGAACATTAACAAACTT | ACAATCTAATAAAATACGAG |
| 8  | mTcCIR184 | GGTGGGTTAGTCTCTCCA | TCTGTCTGCTTTTGTGA |
| 9  | SSRCc4.8 | GCAATAAATTTTCCAAGAGG | CTTGGTAGTCTCAGAGATG |
| 10 | SSRCc2.5 | GGGAACCCATAATGTAGAATC | CCTTTCCTCTCTTGATAG |
| 11 | SSRCc5.48 | TTCTAATGCAAGCTCAAAAG | GGCTTTTGGTTCCTTAGATAG |

3.2. Parentage assignment using molecular genetic data

Molecular genetic data of 204 cacao F1 hybrids obtained using SSR markers were subjected to parentage analysis to identify the correct male parents. Assignment of these 204 F1 genotypes was determined based on the highest positive LOD value. The Cervus program recognized 202 F1 hybrids and excluded two genotypes. Out of 202 identified F1 progenies, 87 progenies (43.07%) were observed to have true male parents at a 95% confidence level (Table 3). In detail, it was found that 68.75% of the total number of F1 progenies from the ICCRI 03 x TSH 858 crosses combination was really the offspring of this cross, 38.89% of the total number of F1 progenies from the parent combination of ICCR I 03 x DR 1 were true F1 progeny of this cross, 37.5% of the total number of ICCRI 03 x ICS 13 descendants were the original descendants of this cross, 15% of the total number of F1 hybrids produced from ICCRI 03 x SCA 6 cross were truly offspring of this cross, 30.43% of the total number of progeny from the combination of TSH 858 x DR 1 cross were the number of real progenies of this cross, 44% of the total number of F1 hybrids from TSH 858 x ICS 13 parental combination were direct descendant of this cross, 50% of the total number of F1 hybrids generated from the crosses between TSH 858 x SCA 6 were the original progenies of this cross, 35.71% of the total number of F1 hybrids produced from DR 1 x ICS 13 crosses were the descendants of this cross, 56% of the total number of F1 progenies resulted from a cross between DR 1 x SCA 6 were the real progenies of this cross, and 53.33% cacao progenies of the F1 hybrids were originate from the a cross between ICS 13 x SCA 6.

The result demonstrated that parentage assignment using the Cervus program was very helpful in identifying F1 progenies with correct male parents. It is expected that the known identity of F1 hybrids is very important in the cacao breeding program, making it easier for breeders to make selections based on the traits inherited from both parents. Subsequently, the F1 individuals identified as progenitors of their true male parents can be grouped according to the combination of parents. The parentage analysis had been successfully identified 400 cacao accessions in Cameroon. Their correct parents, i.e., 25.5% of cacao accessions were found to be closely related to Amelonado variety, 46.3% of the farm accessions were direct descendants of 24 parental clones used in biclonal seed gardens (BSGs), and another 28.3% accessions were generated from uncontrolled pollination [16]. Thus, parentage analysis using molecular genetic data can be used to verify the genetic identity and pedigree information of cacao [8], to characterize pollination and gene flow patterns [17], as well as understand domestication and artificial selection processes [18].
Table 3. The result of male parents’ identification in F1 cacao populations based on the Cervus program.

| Crossing combination          | Total number of progenies | No. of identified progeny | No. of progeny with appropriate male parents |
|------------------------------|---------------------------|---------------------------|---------------------------------------------|
| 1. ICCRI 03 x TSH 858        | 16                        | 11                        | 11                                          |
| 2. ICCRI 03 x DR 1           | 18                        | 7                         | 7                                           |
| 3. ICCRI 03 x ICS 13         | 24                        | 9                         | 9                                           |
| 4. ICCRI 03 x SCA 6          | 20                        | 3                         | 3                                           |
| 5. TSH 858 x DR 1            | 23                        | 8                         | 7                                           |
| 6. TSH 858 x ICS 13          | 25                        | 11                        | 11                                          |
| 7. TSH 858 x SCA 6           | 24                        | 12                        | 12                                          |
| 8. DR 1 x ICS 13             | 14                        | 5                         | 5                                           |
| 9. DR 1 x SCA 6              | 25                        | 14                        | 14                                          |
| 10. ICS 13 x SCA 6           | 15                        | 8                         | 8                                           |
| Total number                 | 204                       | 88                        | 87                                          |

On the other hand, 56.93% of the identified F1 progenies did not correspond to the known male parents used in this study. This condition is possibly due to uncontrolled hybridizations caused by pollen contamination during hand pollination or self-compatibility [16]. Even though cacao is known to be a cross-pollinating species, some varieties showed high self-compatibility. Thus, we assumed that some of the F1 cacao progenies were generated from parental clones with high self-compatibility, so they were not progenies from any of the male parents used in crosses. In the case of pollen contamination, the problems generally occurred during the hand pollinations process, even between clearly labeled parents, which may be due to pollen mislabeling, leading to illegitimate progenies (with one or both parents wrongly used) [9].

Among 87 F1 progenies possessing the true male parent identity, 48 F1 progenies showed the same identity with the existing parental combination label (Table 4). Meanwhile, the remaining 39 other progenies differed from the existing label (Table 5). The results showed that erroneous labels were often found in cacao collections in the field. The study found that 44.83% of F1 hybrids with true male parents exhibited different labels from the existing ones. This phenomenon was also found in many cacao germplasm collections worldwide, estimated to be around 15-44% of mislabeled clones [19]. The mislabeling problem greatly affects the reliability, accuracy, and efficiency of the conservation of cacao germplasm collections [9]. In addition, mislabeling is also one of the main factors contributing to the high rate of unwanted progenies produced in a seed garden [20]. This condition contributed to the mixing of seeds that are distributed to farmers. Later, the impact of this problem will be reduced farmer adoption of the recommended varieties if it fails to produce the expected yield. Therefore, parentage analysis using molecular genetic data is strongly recommended in the F1 hybrids generated either from hand pollination or open pollination to ensure proper decision-making regarding their use in the breeding program [9].

Table 4. Cacao F1 Progenies identity with male parents matched to the initial label based on Cervus analysis at a 95% confidence level.

| No | Progeny code | LOD value | No | Progeny code | LOD value |
|----|--------------|-----------|----|--------------|-----------|
| 1  | 1-1 (3)      | 6.38E-01  | 25 | 8-11 (3)     | 3.62E+00  |
| 2  | 1-9 (3)      | 3.12E+00  | 26 | 8-1 (1)      | 4.43E+00  |
| 3  | 1-12 (3)     | 5.06E+00  | 27 | 8-5 (2)      | 6.36E+00  |
| 4  | 1-6 (2)      | 5.07E+00  | 28 | 8-3 (2)      | 9.66E-01  |
| 5  | 1-3 (2)      | 3.04E+00  | 29 | 9-7 (2)      | 5.53E-01  |
| 6  | 1-8 (1)      | 2.43E+00  | 30 | 9-6 (2)      | 8.96E+00  |
The 2nd International Conference on Sustainable Plantation  
IOP Conf. Series: Earth and Environmental Science 974 (2022) 012051  
doi:10.1088/1755-1315/974/1/012051

Table 5. Cacao F1 Progenies identity with male parents different from the initial label based on Cervus analysis at a 95% confidence level.

| No | Initial progeny code | Initial male parents | Male parents ‘identity based on Cervus analysis | Correct progeny code |
|----|----------------------|-----------------------|-----------------------------------------------|---------------------|
|  1 | 1-7 (1)              | TSH 858               | DR 1                                          | 2-7 (3)             |
|  2 | 1-1 (1)              | TSH 858               | ICS 13                                        | 3-1 (1)             |
|  3 | 2-4 (3)              | DR 1                  | TSH 858                                        | 1-4 (3)             |
|  4 | 2-1 (3)              | DR 1                  | ICS 13                                        | 3-4 (3)             |
|  5 | 3-7 (3)              | ICS 13                | DR 1                                          | 2-9 (3)             |
|  6 | 3-6 (3)              | ICS 13                | DR 1                                          | 2-6 (3)             |
|  7 | 4-4 (3)              | SCA 6                 | TSH 858                                        | 1-3 (3)             |
|  8 | 4-7 (2)              | SCA 6                 | TSH 858                                        | 1-7 (2)             |
|  9 | 4-7 (1)              | SCA 6                 | TSH 858                                        | 1-3 (1)             |
| 10 | 5-3 (3)              | DR 1                  | SCA 6                                         | 7-2 (3)             |
| 11 | 5-5 (3)              | DR 1                  | SCA 6                                         | 7-5 (3)             |
| 12 | 5-8 (3)              | DR 1                  | SCA 6                                         | 7-8 (3)             |
| 13 | 5-10 (2)             | DR 1                  | SCA 6                                         | 7-9 (2)             |
| 14 | 5-12 (2)             | DR 1                  | SCA 6                                         | 7-12 (2)            |
| 15 | 5-8 (2)              | DR 1                  | ICS 13                                        | 6-11 (2)            |
| 16 | 5-6 (1)              | DR 1                  | SCA 6                                         | 7-7 (1)             |
| 17 | 5-10 (1)             | DR 1                  | SCA 6                                         | 7-9 (1)             |
| 18 | 6-3 (3)              | ICS 13                | DR 1                                          | 5-7 (3)             |
| 19 | 6-5 (3)              | ICS 13                | SCA 6                                         | 7-10 (3)            |
| 20 | 6-6 (3)              | ICS 13                | SCA 6                                         | 7-13 (3)            |
| 21 | 6-8 (3)              | ICS 13                | SCA 6                                         | 7-14 (3)            |
| 22 | 6-10 (2)             | ICS 13                | DR 1                                          | 5-1 (2)             |
| 23 | 6-8 (2)              | ICS 13                | DR 1                                          | 5-4 (2)             |
| 24 | 6-6 (2)              | ICS 13                | DR 1                                          | 5-5 (2)             |
| 25 | 6-3 (2)              | ICS 13                | SCA 6                                         | 7-4 (2)             |
| 26 | 6-11 (1)             | ICS 13                | SCA 6                                         | 7-12 (1)            |
| 27 | 7-3 (3)              | SCA 6                 | ICS 13                                         | 6-1 (3)             |
The present study highlighted the importance of correct identification of cacao male parents for maintaining cacao germplasm collections and help breeders in generating new high-yielding cacao varieties. Overall, the results in this study demonstrated the effects of mislabeled clones on genetic purity of F1 progenies and seeds uniformity of cacao and strengthened the need for DNA fingerprinting to guide the selection of potential parents to minimize inbreeding in cacao breeding programs.

4. Conclusion
The result of parentage analysis demonstrated that 87 cacao F1 progenies (43.07%) had been identified their true male parents. In contrast, the other 115 progenies (56.93%) did not correspond to the known male parents used in the present study. Of the 87 identified F1 hybrids, 48 F1 progenies have the same identity as the existing label, while 39 F1 progenies have a different label. Thus, the present study results have proven the usefulness of molecular genetic data generated using SSR markers for parentage analysis in the cacao F1 population.

Acknowledgments
The authors would like to thank Januar Firmansyah, S.P., Tri Buana Dewi, and M. Dery Kurniawan who have assisted in conducting research in the laboratory. This research was supported by the State Government Budget of 2018 Fiscal Year.

References
[1] Sthapit B R, Joshi K D, Gyawali S, Subedi A, Shrestha K, Chaudhary P, Rana R, Rijal D, Upadhyaya M and Jarvis D 2002 Participatory Plant Breeding: Setting Breeding Goals and Choosing Parents for On-Farm Conservation 104–12
[2] Susilo A W 2007 War. Pus. Penelit. Kopi dan Kakao Indones. 23 11–24
[3] Izzah N K and Reflinur 2018 J. Hortic. 27 33–40
[4] Li X, Yu H, Li Z, Liu X, Fang Z, Liu Y, Yang L, Zhuang M, Lv H and Zhang Y 2018 Hortic. Plant J. 4 158–64
[5] Hung H Y, Browne C, Guill K, Coles N, Eller M, Garcia A, Lepak N, Melia-Hancock S, Orpeza-Rosas M, Salvo S, Upadaya Jula N, Buckler E S, Flint-Garcia S, McMullen M D, Rocheford T R and Holland J B 2012 Heredity (Edinb). 108 490–9
[6] Collard B C Y and Mackill D J 2008 Philos. Trans. R. Soc. B Biol. Sci. 363 557–72
[7] Whalen A, Gorjanc G and Hickey J M 2019 J. Anim. Breed. Genet. 136 102–12
[8] Takrama J, Kun J, Meinhardt L, Mischke S, Opoku S Y, Padi F K and Zhang D 2014 African J. Biotechnol. 13 2127–36
[9] Padi F K, Oforl A, Takrama J, Djan E, Opoku S Y, Dadzie A M, Bhattacharjee R, Motamayor J C and Zhang D 2015 Tree Genet. Genomes 11
[10] Tokarska M, Marshall T, Kowalczyk R, Wójcik J M, Pertoldi C, Kristensen T N, Loeschcke V, Gregersen V R and Bendixen C 2009 Heredity (Edinb). 103 326–32
[11] Allen G C, Flores-Vergara M A, Krasynanski S, Kumar S and Thompson W F 2006 Nat. Protoc. 1 2320–5
[12] Lanaud C, Risterucci A, Pieretti I, Falque M, Bouet A and Lagoda P 1999 Mol. Ecol. 8 2147–9
[13] Pugh T, Fouet O, Risterucci A M, Brottier P, Abouladze M, Deletrez C, Courtois B, Clement D, Larmande P, N’Goran J A K and Lanaud C 2004 Theor. Appl. Genet. 108 1151–61
[14] Marshall T C, Slate J, Kruuk L E B and Pemberton J M 1998 Mol. Ecol. 7 639–55
[15] Teama S 2016 IntechOpen i 13
[16] Efombagn M I B, Sounigo O, Eskes A B, Motamayor J C, Manzanares-Dauleux M J, Schnell R and Nyassé S 2009 Heredity (Edinb). 103 46–53
[17] Ashley M V 2010 CRC. Crit. Rev. Plant Sci. 29 148–61
[18] Motamayor J C, Risterucci A M, Lopez P A, Ortiz C F, Moreno A and Lanaud C 2002 Heredity (Edinb). 89 380–6
[19] Motilal L and Butler D 2003 Genet. Resour. Crop Evol. 50 799–807
[20] Lacombe T, Boursiquot J M, Lauco V, Di Vecchi-Staraz M, Péros J P and This P 2013 Theor. Appl. Genet. 126 401–14