2017 International Symposium on Cocoa Research (ISCR), Lima, Peru, 13-17 November 2017

Identification of a core SNP panel for cacao identity and population analyses.

A. Mahabir, L.A. Motilal, D. Gopaulchan, A. Sankar and P. Umaharan.

Cocoa Research Centre, Sir Frank Stockdale Bldg., The University of the West Indies, St. Augustine, 330912, Trinidad, Trinidad and Tobago.

Abstract
It is becoming increasingly easier to obtain genetic data from hundreds to thousands of single nucleotide polymorphisms (SNPs) in cacao (*Theobroma cacao* L.) plants. Yet, a consensus panel of SNPs for diversity, identity or population ancestry studies remains to be adopted by the cacao community. SNP panels were assembled based on major allele frequency (MjAF), polymorphism information content (PIC) and linkage group (LG) distribution. These panels were assessed on a test panel of 155 accessions to determine the minimum number and best combination of SNPs that could unambiguously separate reference cacao genetic profiles and simultaneously detect the correct population structure. Five designer panels, building on the results of the previous panels that achieved full resolution on the test case of 155 accessions were also assessed on a real world dataset of 1231 accessions. Increasing the number of SNPs generally resulted in improved resolution of genetic identities with concomitant reduction of synonymous groups. Retention of SNPs for panel inclusion relied on informativeness and PIC but did not need to be distributed equally among the ten chromosomes. A panel of 96 SNPs was suggested as a minimal core set of SNPs for adoption by the international cacao community.

Keywords: DNA fingerprinting; match analysis; genetic ancestry; cocoa genebanks; *Theobroma cacao*;

Introduction
Theobroma cacao L. (2n = 2x = 20) of the Malvacecae family (Alverson et al. 1999, Bayer et al. 1999) has its centre of origin and diversity in Amazonian South America (Cuatrecasas 1964; Motamayor et al. 2008). Cacao is a commercially important industrial tree crop that is among the top ten global agricultural commodities (Utro et al. 2012). The fermented and dried cotyledons of the seeds (beans) are the raw ingredients in the multibillion dollar confectionery industry. Cacao is an important cash crop in over 50 countries, mainly on small-holder farms, particularly in West Africa where over 70% of the world's cacao is produced (ICCO 2017). Genetic resources are present in over 60 germplasm collections held in various countries (Motilal In Press), in farmers' fields and endemically in Amazonian South America (Zhang and Motilal 2016).

Cacao has been traditionally classified into three agromorphological groups: Criollo, Forastero, and Trinitario (Cheesman 1944; Cuatrecasas 1964; Toxopeus 1985). Descriptions of genetic diversity in cacao classically and traditionally relied on morphological traits until the development of molecular markers. Molecular marker information was used to sort cacao into 10 ancestral groups (Motamayor et al. 2008). New collections from the wild in Bolivia enabled the identification of an additional population (Zhang et al. 2012) and allowed a reclassification into 13 genetic clusters (Motamayor et al. 2010). The ancestral groups are distributed across a variety of accession groups. For instance, the accession PA 120 [PER] belongs to the Parinari accession group which fits into the Marañon population group of Motamayor et al. (2008). Accession groups are named according to the collection expedition (Turnbull and Hadley 2017) with the result that some accession groups contain individuals belonging to more than one ancestral group. There are 29,666 accession names in the International Cocoa Germplasm Database (ICGD; Turnbull and Hadley 2017) with over 24,000 cacao accessions being distributed over 40 collections (CacaoNet 2012). While some redundancy is expected and there are DNA fingerprints for some accessions in the ICGD, the majority of the accessions remain to be fingerprinted or have multilocus SNP profiles deposited. In addition, germplasm from new collecting expeditions and progenies from breeding trials also need to be genetically identified. Having a common SNP panel would greatly facilitate comparative and transferable results among international collaborators.

A variety of molecular markers have been employed to genotype and assess genetic diversity in cacao with continual adoption of the latest technologies (Motilal et al. 2017 and references therein). The most abundant molecular marker type is the single nucleotide polymorphism (SNP). The use of DNA markers, such as SNPs, in plant breeding can increase the efficiency and precision of breeding (Collard and Mackill 2008). SNPs have been used to characterize crops such as maize (Van Inghelandt et al. 2010) and soybean (Liu et al. 2017). Cacao has also been subjected to SNP genotyping in several studies (Lukman et al. 2014; Livingstone et al. 2015; Padi et al. 2015; Motilal et al. 2017), with over 6000 SNPs being identified by Livingstone et al. (2015).

Nevertheless, there is as yet no firm consensus on a minimal SNP panel for cacao accession identification although Motilal et al. (2017) recommended two panels of 96 SNPs for identity analysis of cacao. Saunders et al. (2004) recommended a set of 15 microsatellite (SSR) loci for identity analysis to the cacao community. Motilal et al. (2009) showed that the composition of the SSR primer panel (quantity and choice of marker) was critical to obtaining full resolution among unique accessions. Ji et al. (2013) found that 26 most informative SNPs could distinguish among 115 accessions with 99.999% certainty but did not indicate what measure was used for the information content or which of the 70 SNPs in their study were to be retained. Fang et al. (2014) used 48 SNPs to demonstrate the feasibility of SNPs in cacao authentication and traceability. Lukman et al. (2014) used 53 SNPs in a genetic diversity study of 136 accessions but did not provide details on the SNPs. Takrama et al. (2014) used 53 SNPs to reliably separate 39 accessions but like Lukman et al. (2014) the panel could not differentiate between the Ucayali and Morona clusters. Livingstone et al. (2015) found that 30 SNP loci were adequate to differentiate between three pairwise combinations of closely related individuals. However, these authors did not indicate whether the 30 loci could discriminate amongst all the 1,152 accessions that were screened with their 6k SNP chip nor did they identify the loci used. In contrast Padi et al. (2015) reported 64 SNPs that discriminated amongst 2424 individuals, although their panel could not resolve a set of three Amelonado accessions.

Thus to facilitate the adoption and use of a common set of SNP markers, the SNP panels suggested by Motilal et al. (2017) were reassessed in order to:

(1) ascertain whether the method used to compile a panel for identity resolution affected the effectiveness of the panel;
(2) identify a minimum panel of SNPs that would offer the same resolution as the maximal set;
(3) identify a minimum panel of SNPs that would generate the same ancestral profiles; and hence
(4) construct a panel which could be used for both identity and ancestry analysis.

Materials and Methods

Leaf tissue samples were collected from a training set of 155 clones from the International Cocoa Genebank Trinidad (ICGT), Jamaica and Haiti. The samples were submitted to LGC Genomics for SNP genotyping, using 192 SNPs developed by CIRAD (Motilal et al. 2017). This panel of 192 SNPs was reduced to a set of 182 SNPs containing less than 5.85% missing data. In-house records showed that the 155 reference samples could be grouped into nine genetic clusters Amelonado (and Amelonado hybrids; 15), Criollo (4), Contamana (12), Guiana (10), Iquitos (19), Marañon (13), Nanay (27), Nacional/Curaray (19) and Refractario (36) when the maximum set of 182 SNPs is used.

The genotype data generated was analysed using GenAlEx version 6.503 (Peakall and Smouse 2006, 2012) to obtain the major allele frequency (MjAF). The polymorphism information content (PIC) was generated using Cervus version 3.03 (Marshall et al. 1998). The reduced set of 182 SNPs were then ranked based on their MjAF and PIC values and SNP panels were constructed as follows:

1. Five panels were constructed based on the MjAF and PIC respectively starting with a set of 24 and in stepwise increments of twelve. The SNPs with highest values for each statistic were preferentially included in these panels. The MjAF values ranged from 0.503 to 1.0 and PIC values ranged from 0.006 to 0.375. Panels based on the MjAF value contained SNPs with a MjAF $\geq$ 0.69, while those based on PIC had a PIC value between 0.323 – 0.375. The loci selected for each panel were not restricted by linkage group.
2. Five panels were constructed based on equal distribution across the ten chromosomes (linkage groups; LGs) using 6, 7, 8, 9, and 10 random SNPs per chromosome.
3. Five upgraded designer panels were constructed based on the performances of the aforementioned panels to achieve full separation on the training set of 155 samples.

All panels constructed were subjected to identity analysis and ancestry analysis. Identity analysis was done using Cervus (Marshall et al. 1998) to determine the ability of each panel to unambiguously distinguish amongst the 155 reference samples. Fuzzy matching amongst the reference types was set at five loci. The resolution abilities of the five designer panels were tested on a real world set of 1231 accessions from the ICGT and for which data on 170 SNPs were available.

Ancestry analysis was determined in STRUCTURE v2.3.4 (Pritchard et al. 2000) to ascertain whether the designer panels were able to assign samples to the correct genetic cluster. Samples were considered allocated to an ancestry group if the results showed 85% or higher membership in a group. An admixed model under independent allele frequencies was fitted using a burn-in of 500,000 followed by a MCMC of 750,000 with 10 iterations for K = 9 groups.

Results

Generally, the resolution ability of the SNP panels increased with increasing number of SNPs leading to full resolution among the training set of 155 accessions and less closely matched fuzzy equivalents (Figs. 1 and 2). Of the original sets of SNP panels constructed (MjAF, PIC and LG), the panels based on the

PIC and LG provided better identity resolution (82.23 – 93.55% and 84.52 to 96.13 % respectively) overall as compared to the MjAF panels (13.55 to 85.16%) as seen in Figure 1. The method of choice affected the discriminant ability with panels based on PIC generally performing better than those based on MjAF for the same number of SNPs. As the number of allocated fuzzy loci increased there was a general increase in the amount of fuzzy matched samples (Fig. 2). The panels selected based on LG seemed to perform better based on overall fuzzy matches than those based on MjAF and PIC values especially when fuzzy matches up to three loci were considered. However, PIC panels with at least 60 SNPs had less fuzzy matches at four and five loci than did LG panels with similar numbers. The best performing of the non-designer panels were those from MjAF containing 96 SNPs; from PIC containing 60, 72 and 96 SNPs and LG panels containing 90 or 100 SNPs .

Five designer panels based on the separation ability of the MjAF, PIC and LG panels were constructed as a set of 48 (LAM48), a set of 60 that were generally able to separate closely related samples (LOW60), two sets of 96 (AM96 and LAM96) and a set of 106 (AM106). These panels did not contain any samples with fuzzy matching at one locus (Fig. 1) and generally outperformed the other panels in having less fuzzy matches (Fig. 2). The LAM48 panel had the highest incidence of fuzzy matching among the designer panels. The PID_{sib} values arising out of the AM96, LAM96 and AM106 panels on the 1231 accessions were at most 10^{-18} (Fig. 3A) which was twelve orders of magnitude higher than with the set of 170 loci. Assessing the designer panels on the larger dataset of 1231 accessions revealed that AM106 and LAM96 achieved the highest resolution (Fig. 3B) and least number of duplicate groups (Fig. 3C) that were closest to the maximal set of 170 SNP loci.

Ancestry for K = 9 were best in AM96 and LOW60 in having only one unresolved pair of ancestral groups in each iteration (Table 1). The panel LAM96 had the highest number of runs with unresolved ancestral groups and LAM48 partitioned the Nanay cluster into two groups.

Discussion

In this study, we examined a recommended panel of 182 SNPs (Motilal et al. 2017) for the minimum number of SNPs that could perform similar to the full panel for identity analysis and ancestry. Discriminant SNPs based on MjAF, PIC and LG were used to create 15 test panels of differing numbers of SNPs. The 15 test panels showed that those based on LG were good choices when fuzzy matching was not considered. Furthermore, for the same number of markers and considering fuzzy matches, as well as full resolution, panels based on PIC were better at resolving identities than those based on MjAF or LG distribution. Of these panels, the set of 96 based on PIC could be considered to be the best based on resolution ability and fuzzy matching. Our results corroborate that of Yoon et al. (2007) who reported that the efficacy of the SNPs, as well as, the size and diversity of the population being investigated would influence the composition of the SNP panel. These results indicate that there is an ascertainment bias in selecting SNP panels for identity resolution. Earlier studies with low numbers of SNPs and accessions (Ji et al. 2013; Fang et al. 2014; Lukman et al. 2014; Takrama et al. 2014) were therefore fortunate in achieving reliable separation. Ascertainment bias may occur when studies use widely divergent samples. In this case, it may be easy to find fewer SNPs that can discriminate amongst all accessions. Ascertainment bias can also occur when the focus is only on closely related individuals, as the selected SNPs may not be able to discriminate amongst a wider set of diverse accessions. Nevertheless, our results show that for identity analysis in cacao, the number of SNP markers must be complemented by choosing SNPs that can resolve closely related samples even at the expense of having low discriminatory power. A similar result for SSR markers in cacao was previously reported (Motilal et al. 2009).

Increasing the number of samples highlighted the importance of choosing both the correct number of SNPs and the choice of SNP in creating a SNP panel. Although a designer minimum set of 48 SNPs (LAM48) could completely resolve the identities of 155 accessions, this set performed poorly in a larger set of 1231 accessions. However, a deliberate designer panel (LAM96) afforded better resolution on the training set of 155 accessions and performed well on a real data set of 1231 accessions. Furthermore, even when the number of SNPs was nearly doubled, a set of 170 SNPs did not achieve full resolution among the 1231 accessions (Fig. 3B). This indicated that the matched accessions within the groups may be duplicates of each other or be very closely related and the specific markers needed to reveal the differences were not present. A PID_{sib} of approximately 10^{-18} was obtained for this panel which was only five orders of magnitude less than the maximal threshold for PID (Motilal et al. 2009). There is therefore scope for the inclusion of other SNPs to improve the LAM96 panel to obtain more stringent PIDsib values and 100% resolution of identities.

Obtaining reliable ancestry information was easily achieved with a lower number of SNPs but at the expense of unresolved individuals. A designer set of 60 SNPs (LOW60) or 96 SNPs (AM96) could obtain the same ancestral allocation as that of 182 SNPs on the panel of 155 accessions. Similar to Takrama et al. (2014) and Lukman et al. (2014) two genetic groups were unresolved. In the former two studies, the clusters were the Contamana and the Nacional whereas in the present study, Nacional and Curaray were unresolved. According to Pritchard et al. (2000), sample size, number of molecular markers, as well as,

admixture affects the output produced by STRUCTURE and the program works well with a small number of markers. Our results indicate that the composition of the SNP panel is also another factor. We have observed that increasing the K value; can eventually allocate the samples into its respective groups at the expense of having sub-clusters, unknown clusters with few samples or unknown clusters with minimal ancestral contributions (data not shown). The designer panels of LAM96 or AM106 could therefore at higher K values perform as well as the full complement of 182 SNPs on the 155 accession set. Furthermore, all three panels may therefore fully assign all individuals to their correct respective clusters but at some as yet undetermined higher K value. The influence of SNP panel composition on ancestry allocation in cacao has not been reported as yet to the best of our knowledge.

A common panel of SNPs that can reliably discriminate amongst accessions and allocate ancestry would be valuable to the cacao community in comparing and sharing diversity data. The designer panel LAM96 (Table 2) is recommended as the base panel to which additional SNPs can be added as needed.

References

Alverson WS, Whitlock BA, Nyffler R, Bayer C, and Baum DA (1999) Phylogeny of the core Malvales: evidence from ndhF sequence data. Am J Bot 86:1474–1486

Bayer C, Fay MF, De Bruijn PY, Savolainen V, Morton CM, Kubitzki K, Alverson WS, and Chase MW (1999) Support for an expanded family concept of Malvaceae within a recircumscribed order Malvales: a combined analysis of plastid atpB and rbcL DNA sequences. Botanical Journal of the Linnaean Society 129:267-303.

CacaoNet (2012) A global strategy for the conservation and use of cacao genetic resources, as the foundation for a sustainable cocoa economy (B. Laliberté, compiler), Bioversity International, Montpellier, France, http://cacaonet.org/.

Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cacao populations. Tropical Agriculture (Trinidad) 21:144-159.

Cuatrecasas J (1964) Cacao and its allies. A taxonomic revision of the genus Theobroma. In Contributions to the U.S. National Herbarium 35(6):375-614. Washington, DC: Smithsonian Institution.

Fang W, Meinhardt LW, Mischke S, Bellato CM, Motilal L, and Zhang D (2014) Accurate determination of genetic identity for a single cacao bean, using molecular markers with a nanofluidic system, ensures cocoa authentication. Journal of Agricultural and Food Chemistry 62(2): 481-487 doi: 10.1021/jf404402v.

ICCO (2017) Quarterly Bulletin of Cocoa Statistics, Vol. XLIII, No. 3, Cocoa year 2016/17. https://www.icco.org/about-us/international-cocoa-agreements/cat_view/30-related-documents/46-statistics-production.html?limit=35&limitstart=0&order=date&dir=ASC

Ji K, Zhang D, Motilal LA, Boccara M, Lachenaud P, and Meinhardt LW (2013) Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. Genetic Resources and Crop Evolution 60: 441-453. doi 10.1007/s10722-012-9847-1.

Liu Z, Li J, Fan X, Htwe NMPS, Wang S, Huang W, Yang J, Xing L, Chen L, Li Y, Guan R, Chang R, Wang D, and Qiu L (2017) Assessing the numbers of SNPs needed to establish molecular IDs and characterize the genetic diversity of soybean cultivars derived from *Tokachi nagaha*.The Crop Journal 326-336. http://dx.doi.org/10.1016/j.cj.2016.11.001

Livingstone D, Royaert S, Stack C, Mockaitis K, May G, Farmer A, Saski C, Schnell R, Kuhn D, and Motamayor JC (2015) Making a chocolate chip: development and evaluation of a 6K SNP array for *Theobroma caca*. DNA Research, 2015, 1–13 doi: 10.1093/dnares/dsv009

Lukman, Zhang D, Susilo AW, Dinarti D, Bailey B, Mischke S, and Meinhardt LW (2014) Genetic identity, ancestry and parentage in farmer selections of cacao from Aceh, Indonesia revealed by single nucleotide polymorphism (SNP) markers. Tropical Plant Biology 7(3-4): 133-143.

Marshall TC, Slate J, Kruuk LEB, and Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Molecular Ecology 7:639-655.

Motamayor, J. C., P. Lachneaud, J. W. da Silva e Mota, R. Loor, D. N. Kuhn, J. S. Brown, and R. J. Schnell. 2008. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). PLoS ONE 3 (10): e3311. doi:10.1371/journal.pone.0003311.

Motamayor JC, Lachenaud P, da Silva e Mota JW, Loor R, Martinez WJ, Graham J, Kuhn DN, Brown S, and Schnell RJ (2010) 'No mas 'Forastero': a new protocol for meaningful cacao germplasm classification, in Proceedings of the 16th International Cacao Research Conference, COPAL, Nigeria, pp. 179-185.

Motilal, L.A. 2017 In Press. The role of gene banks in preserving the genetic diversity of cacao. In Achieving sustainable cultivation of cocoa, Umaharan, P. (ed.), Chapter 3, http://dx.doi.org/10.19103/AS.2018.0021.03

Motilal LA, Sankar A, Gopaulchan D, and Umaharan P (2017) 'Cocoa' in P. Chowdappa, A. Karun, M.K. Rajesh and S.V. Ramesh (eds.), Biotechnology of Plantation Crops, Daya Publishing House, New Delhi, India, pp. 313-354.

Motilal LA, Zhang D, Umaharan P, Mischke S, Boccara M, and Pinney S (2009) Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. Tropical Plant Biology 2:23-37. doi:10.1007/s12042-008-9016-z.

Padi FK, Ofori A, Takrama J, Djan E, OPoku SY, Dadzie AM, Bhattacharjee R, Motamayor JC, and Zhang D (2015) The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. Tree Genetics & Genomes 11: 44. https://doi.org/10.1007/s11295-015-0875-9

Peakall R, and Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Peakall R, and Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28:2537-2539.

Pritchard JK, Stephens M, and Donnelly P (2000) Inference of population structure from multilocus genotype data. Genetics 155:945-959.

Saunders JA, Mischke S, Leamy EA, and Hemeida AA (2004) Selection of international molecular standard for DNA fingerprinting of *Theobroma cacao*. Theoretical and Applied Genetics 110:41-47. doi:10.1007/s00122-004-1762-1.

Takrama J, Kun J, Meinhardt L, Mischke S, Opoku SY, Padi SK, and Zhang D (2014) Verification of genetic identity of introduced cacao germplasm in Ghana using single nucleotide polymorphism (SNP) markers. African Journal of Biotechnology 13(21): 2127-2136. doi: 10.5897/AJB2013.13331.

Toxopeus H (1985) Botany, types and populations. In Cocoa. 4th ed., edited by G. A. R. Wood and R. A. Lass, 11-37. London: Longman.

Turnbull CJ and Hadley P (2017), International Cocoa Germplasm Database (ICGD), [Online Database], CRA Ltd./ICE Futures Europe/University of Reading, UK, http://www.icgd.reading.ac.uk (17th July, 2017).

Utro F, Cornejo OE, Livingstone D, Motamayor JC, and Parida L (2012) ARG-based genome-wide analysis of cacao cultivars. BMC Bioinformatics 13(Suppl 19):S17.

Van Inghelandt D, Melchinger AE, Lebreton C, and Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor. Appl. Genet. 120:1289–1299, doi: 10.1007/s00122-009-1256-2.

Zhang D, Martínez WJ, Johnson ES, Somarriba E, Phillips-Mora W, Astorga C, Mischke S, and Meinhardt LW (2012) Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. Genet. Resour. Crop Evol., 59, 239-252, doi: 10.1007/s10722-011-9680-y.

Zhang D and Motilal L (2016) Origin, dispersal and current global distribution of cacao genetic diversity. In Cacao diseases: A history of old enemies and new encounters, eds. B Bailey and L Meinhardt, Chapter 1, pp. 3-32. New York, USA: Springer.

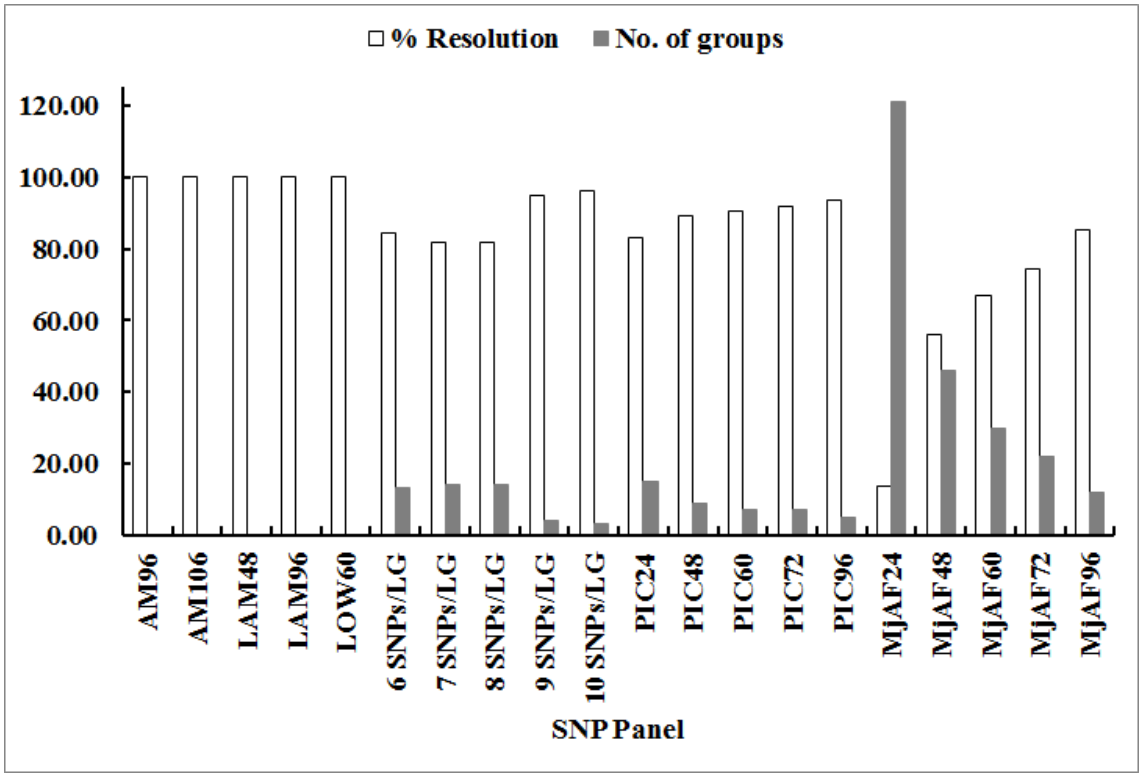

Figure 1. Twenty SNP panels on resolution ability relative to 182 SNPs on 155 cacao accessions Panels were based on distribution on linkage group (LG), major allele frequency (MAF), polymorphism information content (PIC) and designer panels (AM, LAM, LOW) based on separation ability. The number of SNPs are indicated in the suffix of the alphanumeric forms. In the LG panels, the total number of SNPs in each panel is by a factor of 10, since cacao has 10 chromosomes.

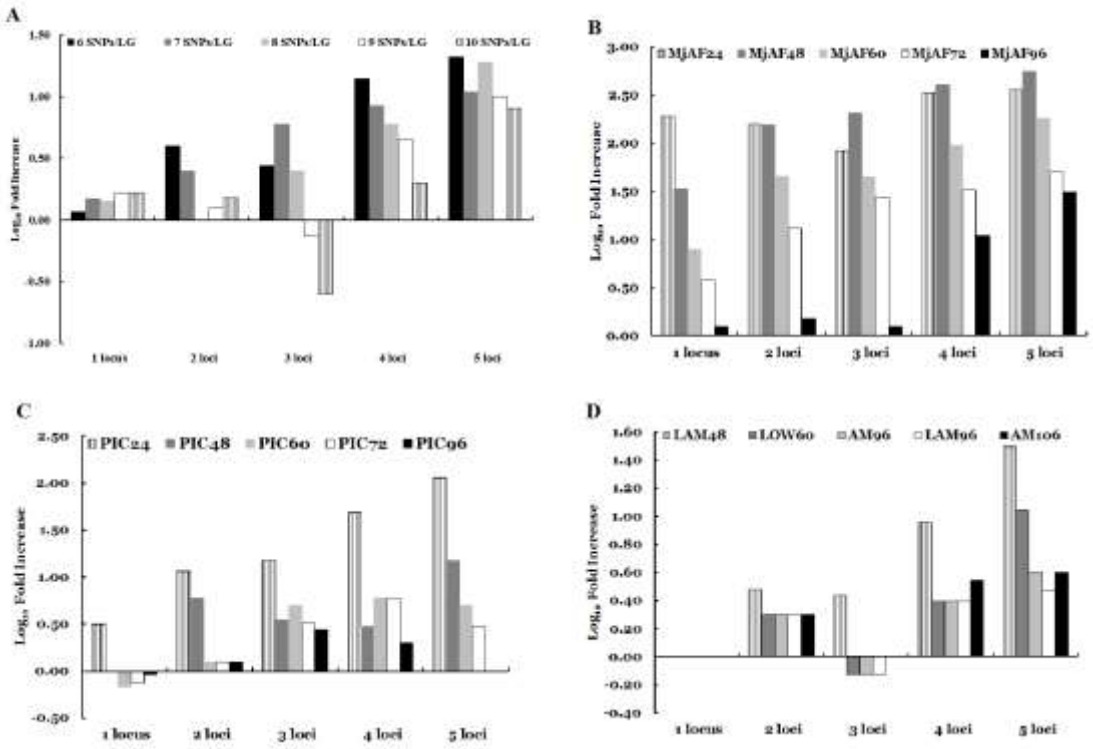

Figure 2 Increase in mismatches at 1-5 loci relative to that obtained with 182 SNPs on 155 cacao accessions.
Panels based on (A) distribution on linkage group (LG), (B) major allele frequency (MAF), (C) polymorphism information content (PIC) and (D) designer panels based on separation ability. The number of SNPs are indicated in the suffix of the alphanumeric forms. In the LG panels, the total number of SNPs in each panel is by a factor of 10, since cacao has 10 chromosomes.

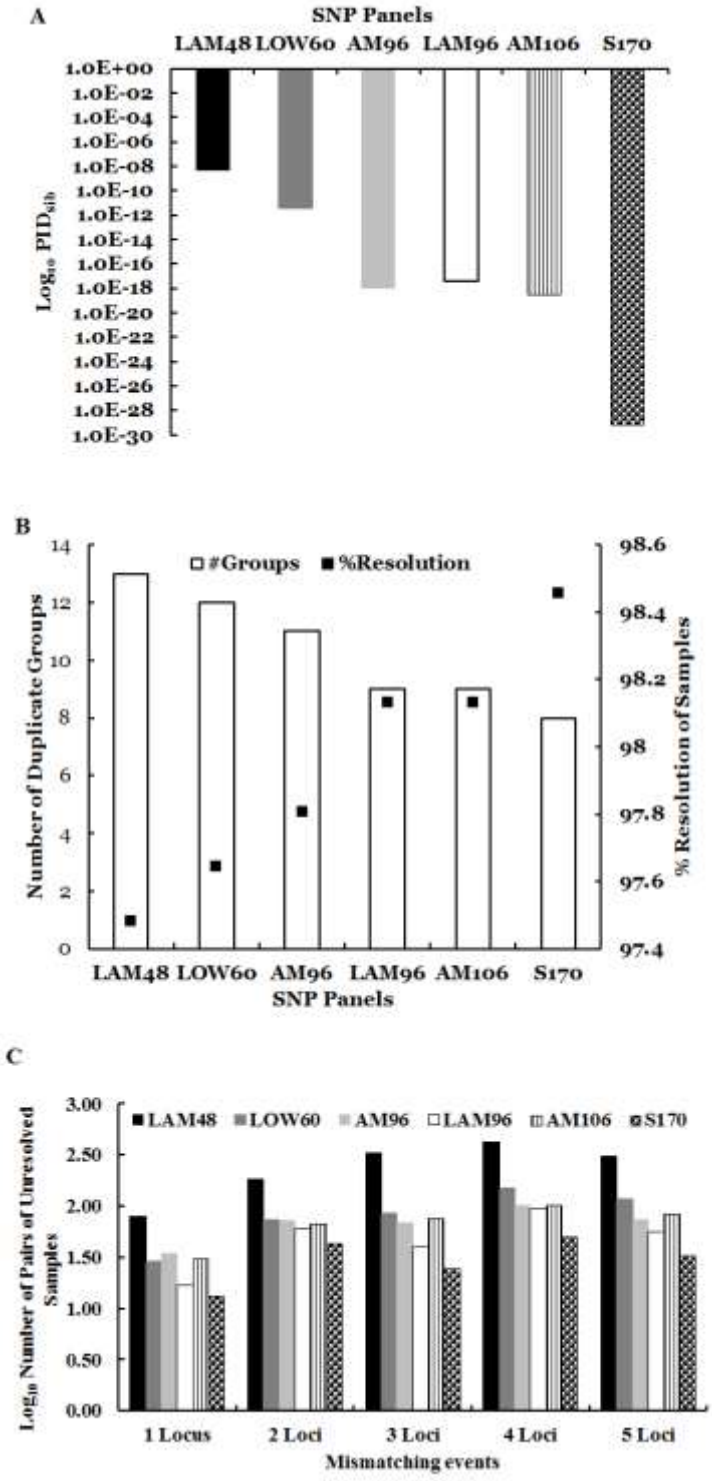

Figure 3 Effectiveness of five SNP designer panels and a maximal set of 170 SNPs on identity analysis of 1231 cacao accessions. Designer panels contain 48 (LAM48), 60 (LOW60), 96 (AM96, LAM96) or 106 (AM106) SNPs.

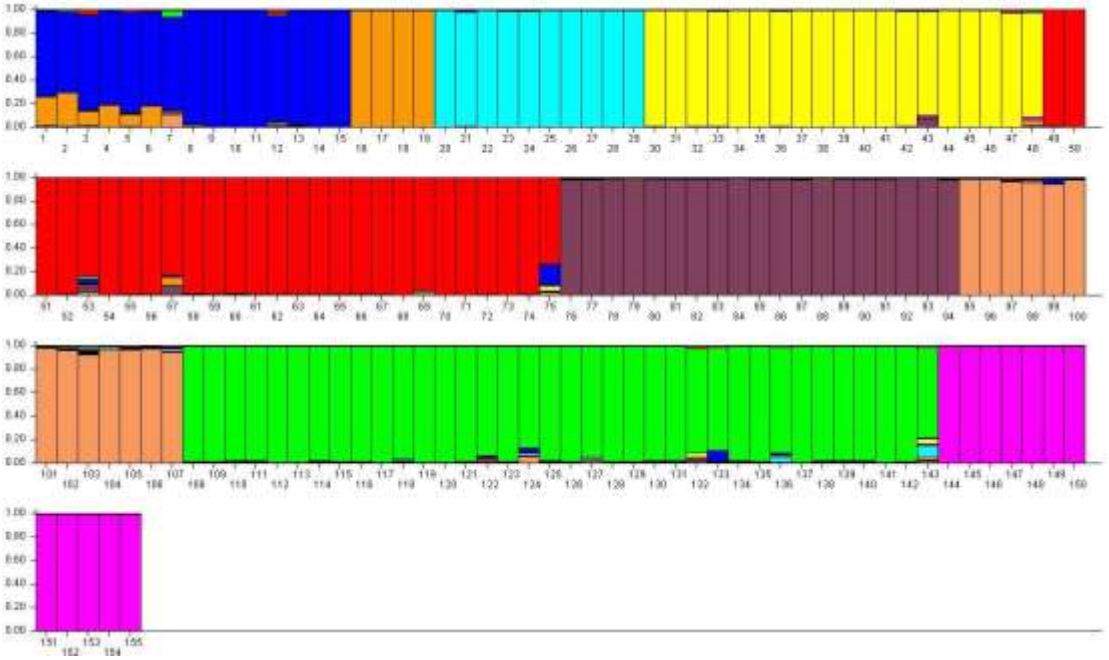

Figure 4. Example of ancestry output at K = 9 from designer panel LOW60 on 155 cacao accessions. Each individual bar represents an individual and each solid colour is a different genetic cluster. Genetic clusters from left to right are Amelonado, Criollo, Guiana, Iquitos, Nanay, Nacional/Curaray, Marañon, Refractario and Contamana.

Table 1 Ancestry allocation in 155 cacao accessions using selected SNP panels.

SNP Panel	# of 10 iterations with only Nacional and Curaray unresolved	Number of mixed genetic clusters observed		
		Minimum	Maximum	Mode
LAM48	7	1	2	1
LOW60	10	1	1	1
AM96	10	1	1	1
LAM96	0	1	4	2
AM106	4	1	3	3
S182	7	1	3	1

Table 2 Composition of the recommended LAM96 panel.

TcSNP0013	TcSNP0154	TcSNP0329[acd]	TcSNP0642	TcSNP0964	TcSNP1229
TcSNP0019	TcSNP0164	TcSNP0364[acd]	TcSNP0704	TcSNP0998[abcd]	TcSNP1237
TcSNP0032[bd]	TcSNP0176	TcSNP0372[abcd]	TcSNP0723[acd]	TcSNP1010	TcSNP1270[ab]
TcSNP0033	TcSNP0189[ac]	TcSNP0397	TcSNP0749	TcSNP1019	TcSNP1275
TcSNP0049	TcSNP0192	TcSNP0429[abd]	TcSNP0751[d]	TcSNP1028	TcSNP1293
TcSNP0064	TcSNP0193[abc]	TcSNP0456	TcSNP0791	TcSNP1038[abcd]	TcSNP1309[ac]
TcSNP0075[acd]	TcSNP0194	TcSNP0469[abcd]	TcSNP0814	TcSNP1053	TcSNP1331[ac]
TcSNP0097	TcSNP0214	TcSNP0519	TcSNP0823[acd]	TcSNP1058	TcSNP1362
TcSNP0105	TcSNP0226[abc]	TcSNP0534[abc]	TcSNP0835	TcSNP1074	TcSNP1401
TcSNP0131	TcSNP0230[abcd]	TcSNP0546	TcSNP0836[ab]	TcSNP1075[abd]	TcSNP1404
TcSNP0135	TcSNP0242[abcd]	TcSNP0577[abc]	TcSNP0841	TcSNP1112	TcSNP1414[abd]
TcSNP0139[abc]	TcSNP0256	TcSNP0591[abcd]	TcSNP0857	TcSNP1136	TcSNP1457
TcSNP0141	TcSNP0258	TcSNP0602[ac]	TcSNP0917[abd]	TcSNP1144[bd]	TcSNP1458[abcd]
TcSNP0143	TcSNP0259	TcSNP0606[d]	TcSNP0933	TcSNP1156[d]	TcSNP1484[abcd]
TcSNP0144[abcd]	TcSNP0280	TcSNP0607	TcSNP0953[ab]	TcSNP1195	TcSNP1524
TcSNP0150[abc]	TcSNP0313	TcSNP0640	TcSNP0954	TcSNP1205	TcSNP1527

Details of these SNPs can be found in Motilal et al. (2017).

SNPs common to other studies are as found in [a]Ji et al. (2013), [b]Fang et al. (2014), [c]Takrama et al. (2014) and [d]Padi et al. (2015).