PLOS ONE

SNP-based assessment of genetic purity and diversity in maize hybrid breeding

Chimwemwe Josia[1], Kingstone Mashingaidze[2], Assefa B. Amelework[3], Aleck Kondwakwenda[1]*, Cousin Musvosvi[1], Julia Sibiya[1]

1 School of Agriculture, Earth and Environmental Sciences, University of KwaZulu-Natal, Scottsville, Pietermaritzburg, South Africa, 2 Agricultural Research Council- Grain Crops Institute (ARC-GCI), Potchefstroom, South Africa, 3 Agricultural Research Council -Vegetable and Ornamental Plants (ARC-VOP), Roodeplaat, Pretoria, South Africa

* alekondwa@gmail.com

Abstract

Assessment of genetic purity of parental inbred lines and their resultant F_1 hybrids is an essential quality control check in maize hybrid breeding, variety release and seed production. In this study, genetic purity, parent-offspring relationship and diversity among the inbred lines were assessed using 92 single-nucleotide polymorphism (SNP) markers. A total of 188 maize genotypes, comprising of 26 inbred lines, four doubled haploid (DH) lines and 158 single-cross maize hybrids were investigated in this study using Kompetitive Allele Specific Polymerase Chain Reaction (KASP) genotyping assays. The bi-allelic data was analyzed for genetic purity and diversity parameters using GenAlex software. The SNP markers were highly polymorphic and 90% had polymorphic information content (PIC) values of > 0.3. Pairwise genetic distances among the lines ranged from 0.05 to 0.56, indicating a high level of dissimilarity among the inbred lines. A maximum genetic distance of (0.56) was observed between inbred lines CKDHL0089 and CML443 while the lowest (0.05) was between I-42 and I-40. The majority (67%) of the inbred lines studied were genetically pure with residual heterozygosity of <5%, while only 33% had heterozygosity levels of >5%. Inbred lines, which were not pure, require purification through further inbreeding. Cluster analysis partitioned the lines into three distinct genetic clusters with the potential to contribute new beneficial alleles to the maize breeding program. Out of the 68 hybrids (43%) that passed the parent-offspring test, seven hybrids namely; SCHP29, SCHP95, SCHP94, SCHP134, SCHP44, SCHP114 and SCHP126, were selected as potential candidates for further evaluation and release due to their outstanding yield performance.

Introduction

Maize (*Zea mays* L.) is the principal crop for food security and nutrition for many people in sub-Saharan Africa (SSA) and Latin America [1]. However, adequate production of the crop is hampered by low grain yields due to biotic and abiotic stresses. Hence, there is a need to improve grain yield through hybrid breeding to exploit heterosis.

Assessment of genetic purity of parental inbred lines and parent-offspring test for the resultant F_1 hybrids is an essential quality control function in maize hybrid breeding. These

functions are now more critical due to the stringent intellectual property requirements governing plant breeding and variety registration in many countries [2]. Additionally, the maintenance of high levels of genetic purity is critical for the robust agronomic performance of the genotype. Parent-offspring test helps to prove parentage for a specific hybrid whether it is a true derivative of the original parental inbred lines without pollen contamination [3]. Inbred lines' genetic purity and parentage can be proved using three approaches namely; grow out test (GOT), use of biochemical markers and use of molecular markers.

Grow out test is a morphologically based approach using a set of descriptors, biochemical markers approach analyses the protein profiles of the genotype and the molecular marker approach detects variation of the genotypes directly at the DNA level [3]. Unlike GOT and biochemical marker methods, which have low polymorphism and high environmental influences, molecular markers are polymorphic, independent of the environment, reproducible, expressed at all developmental stages, known position in the genome, linked to traits of interest and automation is possible [3].

They are several types of molecular markers that are available for detection of polymorphism [4]. The main ones include restriction fragment length polymorphism (RFLPs), random amplified polymorphic DNA (RAPD), simple sequence repeats (SSRs), amplified fragment length polymorphism (AFLP) and single nucleotide polymorphism (SNP). In this study, SNP markers were used to determine the genetic purity of the maize parental inbred lines and to prove the parentage of the resultant single-cross hybrids. Recent advances in molecular technology have emphasized the use of SNP markers because they are cost-effective per data point, locus–specificity, codominant, have adequate genomic abundance, and potential for high throughput unlike the other markers [5].

The application of molecular markers is more efficient, saves time and resources [6] and they are free from environmental influences compared to morphological markers. It is often assumed that the use of a large number of markers results in higher accuracy. In most sequence-based marker systems, the levels of missing data can lead to wrong interpretation, hence, the selection of fewer markers with high and repeatable representation across samples is desired and is cost-effective. Fewer markers with high excepted heterozygosity, missing value of <20%, and observed heterozygosity of <6% are ideal markers for accurate quality control genotyping [7]. A set of 50–100 single plex assay SNPs are adequate for molecular-based quality control genotyping [2]. It is against this background that in the present study, a set of 92 SNPs were used to genotype 30 parental lines and 158 single-cross hybrids.

According to [3], parental inbred lines are expected to be pure with residual heterozygosity of less than 5%. Inbred lines having residual heterozygosity above 5% are either not pure due to genetic contamination or not fixed unless if they were deliberately maintained at early generation during development. Genetic contamination reduces the genetic and physiological quality of the seeds leading to decreased crop productivity [8]. Hence, inbred line genetic purity assessment and parent offspring test are important quality control procedures for a successful hybrid breeding program.

Materials and methods

Experimental material

A total of 188 maize genotypes, comprising 26 elite parental inbred lines, four doubled haploid lines and 158 experimental single-cross hybrids (S1 Table) were genotyped using 92 single nucleotide polymorphism (SNP) markers (S2 Table). These markers are a subset of the 100 SNP markers recommended by CIMMYT for routine quality control genotyping in maize and cover the 10 pairs of the maize chromosomes [3]. All the genotypes used in this study were

sourced from Agriculture Research Council-Grain Crops Institutes (ARC-GCI), Potchefstroom, South Africa.

Leaf sampling, hybrids' field evaluation, DNA extraction and genotyping

Maize genotypes were planted at ARC-GCI research farm, Potchefstroom (26˚74"S; 27˚8'E) during the 2017/18 summer season. Leaf sampling was done using the supplied LGC sampling kit (LGC Genomics Laboratory, United Kingdom). Five to eight leaf discs were taken per entry five weeks after planting for DNA extraction. Leaf samples from the same entry were placed in a specific 2 x 96-well plate with each well representing an individual genotype. Each well was sealed using a perforated trip cap and the desiccant sachet was placed directly on top of the strip cap-sealed well and the plastic lid was replaced on top. The storage rack was secured using an elastic band and was placed inside a sealable plastic bag. The sealed bag was placed into the plant kit box and the samples were shipped to LGC Genomics Laboratory, in the United Kingdom for genotyping. DNA extraction, amplification and visualization were done according the LGC protocol (www.lgcgroup.com). Genomic DNA was extracted from the leaf disc samples and the quality and quantity of the extracted DNA was determined before genotyping. Genotyping was done using the 92 SNP markers, following the Kompetitive Allele Specific Polymerase Chain Reaction (KASP) protocol used by LGC Genomics (www.lgcgroup.com).

Field evaluation of single-cross hybrids was done at three locations namely; Potchefstroom (ARC-GCI) in North West province, Cedara in KwaZulu-Natal province and Vaalharts in the Northern Cape province, South Africa, during the 2017/18 summer season. The trial constituted of five production environment; Potchefstroom and Cedara representing two environments (low N and optimum), while Vaalharts had an optimum environment. One hundred and seventy single-cross maize hybrids including three checks were laid out in a 34 x 5 (0,1) alpha lattice experimental design replicated twice in each environment. However, for the purpose of this study only 158 set of single cross hybrids were used due to limited availability of seeds for the dropped hybrids.

Statistical analyses

Data filtering for monomorphic SNPs and/or SNPs with a minor allele frequency of less than 2% were performed and all the 92 SNPs were polymorphic and of high quality. Genetic purity of the parental inbred lines was calculated as percentage residual heterozygosity using the formula shown in Eq (1) [3]. Genetic diversity parameters such as observed heterozygosity (Ho), expected heterozygosity (He), and fixation index (F_{IS}) were determined using GenAlex version 6 [9]. The formula: PIC = 1 - ΣP_{ij}^2, where P_{ij} is the frequency of jth allele of the ith locus [10], were used to calculate Polymorphic information content (PIC).

$$\% \, Heterozygosity = \left(\frac{SNP \, with \, heterozygous \, Loci}{Total \, number \, of \, SNPs} \right) 100 \tag{1}$$

Genetic relationships within and among the inbred lines were assayed with a neighbor-joining algorithm, using the unweighted pair group method (UWPGM) in DARwin 6.0 software [11]. Pairwise dissimilarity matrices were obtained from the Jaccarrd's coefficient and a dendrogram was generated. For the node construction, a bootstrap analysis was performed, based on 10000 bootstrap values, using DARwin. The distinctiveness of the clusters was checked, using the cophenetic correlation coefficient (r). The parent-offspring relationship for each parent-hybrid pair was tested according to methods described by [3]. Parameters such as the

proportion of SNPs from parent A and parent B, SNPs shared by both parents and SNPs that do not belong to either of the parent were estimated.

Results

SNP characterization

The distributions of values for polymorphic information content, gene diversity, inbreeding coefficient and minor allele frequencies of the 92 SNPs estimated on the 188 maize genotypes are shown in Fig 1. Inbreeding coefficient displayed contrasting values ranging from -0.17 to 1.00, with a mean of 0.10. About 24% of the SNPs showed negative F_{IS} values. Nearly 39% of the SNPs had F_{IS} values between 0.10 and 0.40 (Fig 1A). The SNPs diversity ranged from 0.11 to 0.50; however, the vast majority (92%) fell between 0.30 and 0.50 and eight SNPs revealed moderate gene diversity (Fig 1B). Approximately 90% of the markers used in this study had PIC values exceeding 0.30. Most of the values (77%) were between 0.40 and 0.50 and only one marker (PZA03527_3) displayed a PIC value of less than 0.2 (Fig 1C). The minor allele frequency ranged from 0.06 for the marker PZA03527_3 to 0.50 for the marker sh1_12, with a mean of 0.35 (Fig 1D). More than 55% of the SNPs revealed a minor allele frequency exceeding the mean (0.35). Observed heterozygosity (H_o) values ranged from 0.0 to 0.56 with a mean of
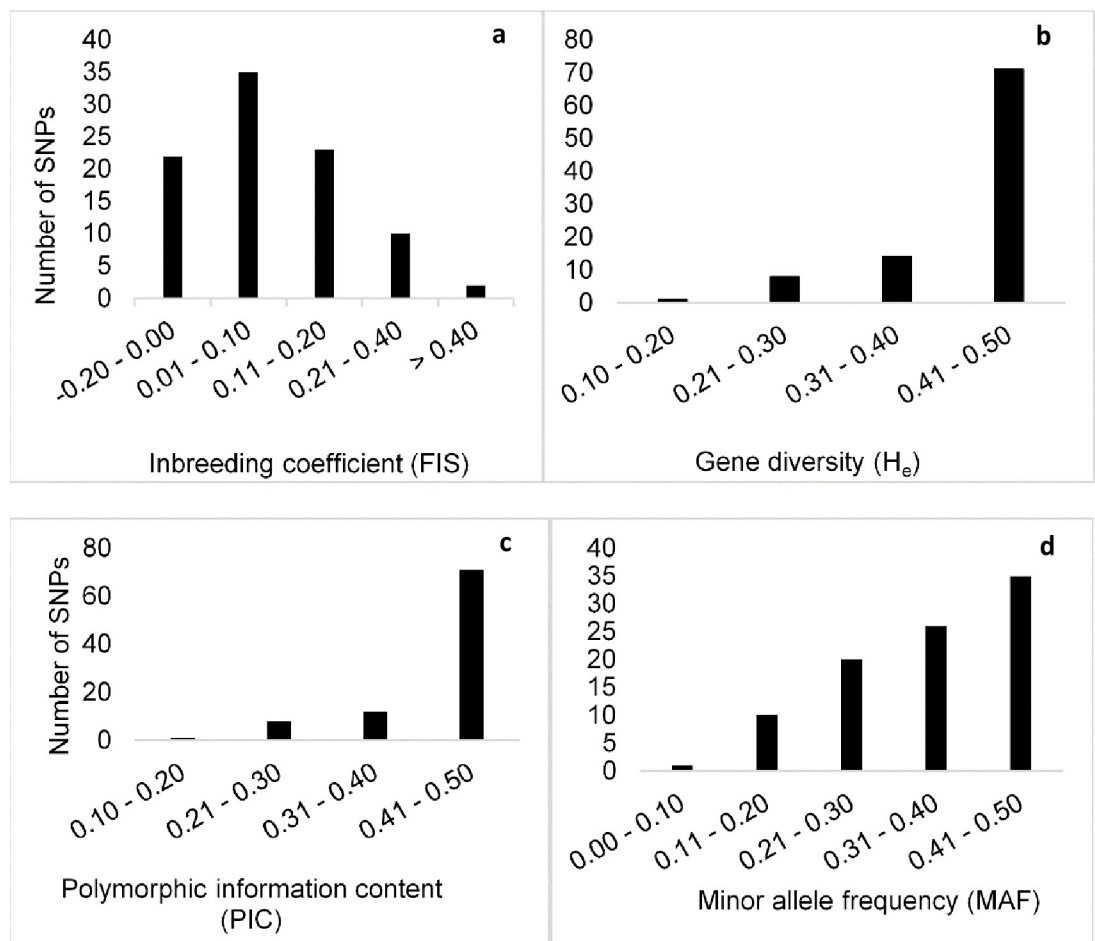

Fig 1. Distribution of the 92 SNPs estimated for all 188 maize genotypes for (a) Inbreeding coefficient, (b) Gene diversity, (c) Polymorphic information content and (d) Minor allele frequency.

Table 1. Summary of genetic diversity parameters of 92 SNPs per chromosome measured in a set of 188 maize genotypes.

Chromosome	No. of SNPs	Genetic parameters							
		Maize Inbred lines				Hybrids			
		Ho	He	F_{IS}	PIC	Ho	He	F_{IS}	PIC
1	9	0.07	0.42	0.82	0.42	0.41	0.40	-0.03	0.40
2	12	0.11	0.45	0.75	0.44	0.44	0.42	-0.05	0.42
3	11	0.08	0.45	0.80	0.44	0.47	0.44	-0.07	0.43
4	9	0.07	0.45	0.84	0.44	0.45	0.45	-0.02	0.45
5	12	0.11	0.46	0.76	0.45	0.51	0.45	-0.13	0.45
6	9	0.09	0.45	0.79	0.44	0.44	0.42	-0.05	0.42
7	7	0.10	0.42	0.75	0.30	0.49	0.44	-0.12	0.44
8	9	0.08	0.49	0.84	0.49	0.39	0.43	0.17	0.43
9	8	0.06	0.44	0.85	0.43	0.46	0.42	-0.08	0.42
10	6	0.09	0.46	0.81	0.45	0.47	0.44	-0.06	0.44
Overall mean		0.09	0.45	0.80	0.44	0.45	0.43	-0.05	0.43
SE		0.00	0.01	0.01	0.01	0.01	0.01	0.02	0.01

N = Number of individuals tested; H_o = observed heterozygosity; H_e = expected heterozygosity; F_{IS} = inbreeding coefficient; PIC = polymorphic information content; SE = Standard error.

https://doi.org/10.1371/journal.pone.0249505.t001

0.40 (data now shown). SNP markers PZA00793_2 and PHM2350_17 had H_o value of 0.0 indicating the alleles of these SNPs were 100% fixed among the maize genotypes, however, 97% of the SNPs had H_o values exceeding 15%. SNP markers PHM2350_17, and PZA00793_2 showed inbreeding coefficient value of 1.

The genetic diversity parameter estimates of the 92 SNPs used in this study summarized per chromosome are presented in Table 1. The number of SNPs on each chromosome ranged from six on chromosome 10 to 12 on chromosomes 2 and 5, with a mean of 9.2 SNPs per chromosomes. The observed heterozygosity of the SNP loci for the inbred lines ranged from 6% to 11%, while the hybrids revealed H_0 values ranging from 39% to 49%. The gene diversity values for the inbred lines ranged from 0.42 to 0.49, with a mean gene diversity of 0.45. However, no significant differences were observed in PIC and gene diversity values among the ten chromosomes. The mean inbreeding coefficient (F_{IS}) was significantly higher for inbred lines ranging from 0.75 to 0.85, with a mean value of 0.80. The hybrids on the other hand, revealed very low F_{IS} values ranging from -0.02 to 0.17, with a mean value of -0.05.

Genetic purity of parental maize inbred lines

The percentage of missing data per SNP in this study was below 3% and varied from 0 to 2.17%, with the overall mean of 1.16%. Based on the 92 SNPs, genetic purity among the 30 inbred lines varied from 0.0 to 57.6, with an overall mean of 10.43 (Table 2). All the 92 SNP loci tested in this study were fixed in 60%. Out of the 18 genotypes that showed 100% genetic purity, four lines (CKDHL0089, CKDHL0295, CKDHL0378 and CKDHL0470) were doubled haploids. Inbred lines CKL05022 and CB323 had a heterozygous percentage of less than 5% and these inbred lines were considered to be fixed. However, 33.3% of the inbred lines had residual heterozygosity ranging from 5.43 to 57.61%.

Genetic relationship among 30 parental lines

The population structure for the parental lines was assessed using distance-based cluster analyses. Cluster analysis based on Jaccard's genetic distance values classified the 30 parental lines

Table 2. Genetic purity of 26 maize inbred lines and four doubled haploid lines based on 92 SNPs.

Name	% of missing alleles	% of heterozygote alleles
CKDHL0089	2.17	0.00
CKDHL0295	0.00	0.00
CKDHL0378	1.09	0.00
CKDHL0470	0.00	0.00
CML202	1.09	0.00
CML216	1.09	0.00
CML442	1.09	0.00
CML443	1.09	0.00
CB322	2.17	0.00
CML544	0.00	0.00
CML511	0.00	0.00
CZL068	0.00	0.00
I-38	2.17	0.00
I-42	2.17	0.00
CML540	2.17	0.00
CZL99017	1.09	0.00
CML312	2.17	0.00
I-40	1.09	0.00
CKL05022	1.09	2.00
CB323	2.17	4.35
CK21	1.09	5.43
RO549W	2.17	15.22
U2540W	1.09	16.30
CML444	2.17	18.48
CB339	0.00	19.57
CML543	2.17	39.13
CZL0919	1.09	42.39
CML488	0.00	45.65
CML547	0.00	48.91
CZL0718	1.09	57.61
Mean	1.16	10.43
SE	0.16	3.23

into three distinct clusters (Fig 2). The distinctiveness of the clusters was confirmed by the high cophenetic correlation coefficient for SNPs (r = 0.93). The highest genetic distance between the parental lines was 0.56 and the lowest was 0.05, while the mean was 0.47. The highest genetic distance (0.56) was found between parental lines CKDHL0089, CML443 and CB323. The lowest genetic distance (0.05) was found between inbred lines I-42, and I-40. The majority (92%) of the genetic distance values fell between 0.40 and 0.60, suggesting the genotypes were moderately and distantly related (Fig 3). Cluster I consisted of 10 parental lines and further sub-divided into two sub-clusters. Cluster II also had two sub-clusters comprising of 15 parental lines, while cluster III consisted of five parental lines. Overall, the cluster analysis was effective in discriminating the parental lines into groups and in providing genetic information for breeding and conservation. In this analysis, three sets of parental lines with different genetic backgrounds were included. The clustering patterns of the parental lines were according to their genetic background. All the CB inbred lines (CB323, CB322 and CB 339) were grouped in Cluster II sub-cluster 1, while all the I inbred lines (I-40, I-42, I-38) were
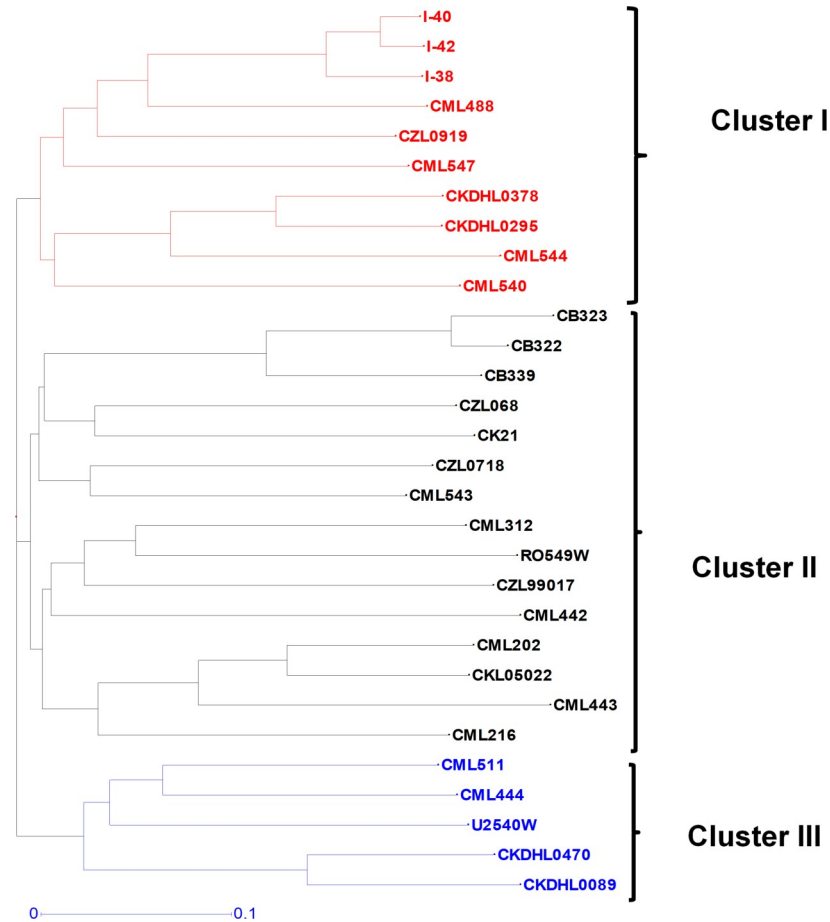
Fig 2. Neighbor-joining dendrograms based on UPGMA genetic dissimilarity depicting the genetic relationship between 30 parental maize lines based on 92 SNPs.

clustered in Cluster I sub-cluster 1. Similarly, all the CZL lines (CZL068, CZL0718 and CZL99017) except CZL0919 were clustered in Cluster I. The two CML heterotic tester lines, CML444 and CML202, which belong to heterotic groups A and B, respectively, were assigned in different clusters.

Parent-offspring test and grain yield performance for the selected maize hybrids

Parent offspring verification test revealed that out of 158 single-cross hybrids tested,16 hybrids had 0% contamination, 96 hybrids registered contamination level within the range of 0.54% to 4.89% ($< 5\%$) and 46 hybrids had contamination greater than 5% (S3 Table). Further quality analysis revealed that some of the ten contaminated lines were used for 90 single-cross hybrids. The test also confirmed that the remaining 68 single-cross hybrids were generated using pure parental inbred lines with acceptable genetic contamination ($<5\%$).

The hybrids were evaluated for grain yield performance and the top and bottom 10% performing hybrids are presented in Table 3. Grain yield observed from 158 single-cross hybrids ranged from 2.83 t ha^{-1} for SCHP115 to 7.33 t ha^{-1} for SCHP17 and a mean yield of 5.55 t ha^{-1} was observed. The parent-offspring test was performed based on the criteria of $< 5\%$ genetic contamination on at least one of the parents and their hybrids. Based on the above criteria, of
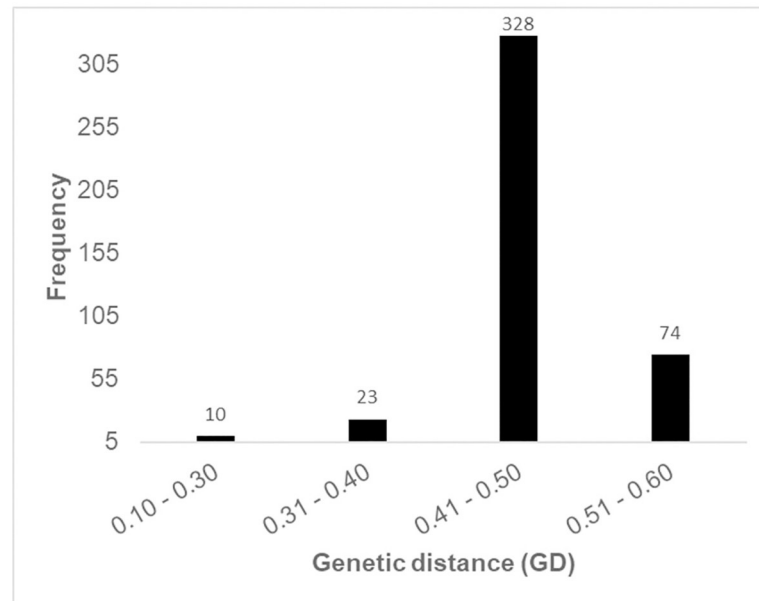
Fig 3. Genetic distance data summary of 30 parental maize lines using 92 SNPs.

the 15 top performing hybrids, seven hybrids (SCHP29, SCHP95, SCHP94, SCHP134, SCHP44, SCHP114 and SCHP126) passed the test and represented 47% of the hybrids. Similarly, among the bottom 15 performing hybrids, five hybrids (SCHP113, SCHP102, SCHP32, SCHP71, and SCHP15) fulfilled the requirement. Notably, hybrids SCHP29 and SCHP115 among the top 10% and bottom 10% performing hybrids, respectively, exhibited genetic contamination of 0%.

Discussion

In cross-pollinated crops such as maize, maintaining genetic purity for the inbred lines and the hybrids is vital for successful hybrid breeding and seed production. Significant changes in the genetic constitution of the inbred lines affect both the quality of newly developed hybrids and hybrid seed. The routine quality control genotyping in this study was done to detect any contamination which could have happened during hybrid development and inbred lines maintenance. Inbred lines are regarded pure or fixed if the proportion of heterozygous SNP loci does not go beyond 5% [2]. Inbred lines are also expected to maintain all the genetic characters that the breeder selected them for. Majority (67%) of the tested inbred lines were pure and fixed and 18 of them attained 100% genetic purity including four DH lines. The fact that all the DH lines (CKDHL0089, CKDHL0295, CKDHL0378 and CKDHL0470) used in this study exhibited 100% homozygosity indicates the advantages of using the DH approach in inbred line development. The DH approach enhances breeding efficiency through rapid generation of homozygous lines which are more fixed and predictive than those developed using conventional methods [12, 13]. This approach significantly shortens the breeding cycle through rapid development of fixed lines in two to three generations unlike the conventional approach that requires six to eight generations of inbreeding with approximately 99% homozygosity.

The remaining 14 inbred lines (CB323, CML202, CML216, CML442, CML443, CML511, CML544, CZL068, I-38, 1–42, CML 540, CZL99017, CML312 and I-40) with 100% genetic

Table 3. Parent-offspring test for top and bottom 10% performing hybrids based on grain yield.

Hybrid	Pedigree	% SNPs parent A	% SNPs parent B	% alleles shared by both	% contamination	GY (t ha^{-1})
Top 10%						
SCHP17	CK21/CKDHL0089	25.54	28.26	43.48	1.63	7.33
SCHP11	CB339/CML442	22.83	23.91	51.63	0.54	6.89
SCHP125	I-42/CB339	26.09	26.63	46.20	1.09	6.85
SCHP29	CKDHL0089/CML442	23.91	25.54	49.46	0.00	6.77
SCHP42	CKDHL0470/CK21	28.80	27.17	39.67	2.17	6.76
SCHP95	CML544/I-38	21.74	18.48	57.07	1.63	6.63
SCHP94	CML544 CML442	26.09	24.46	46.20	1.09	6.62
SCHP61	CML442/CB339	25.00	23.37	51.63	0.00	6.56
SCHP22	CK21/CML444	30.98	13.59	41.85	3.59	6.49
SCHP134	I-42/CML544	17.39	22.83	54.89	3.80	6.47
SCHP44	CKDHL0470/I-38	26.09	22.28	50.00	0.54	6.37
SCHP24	CK21/CML511	21.74	21.20	55.43	1.63	6.35
SCHP1	CB322/CK21	25.00	22.28	47.83	4.89	6.33
SCHP114	I-38/CML202	22.28	24.46	53.26	0.54	6.33
SCHP126	I-42/CKDHL0295	20.11	28.80	49.46	1.63	6.32
Bottom 10%						
SCHP26	CK21 / CML547	22.28	22.28	54.35	1.09	4.75
SCHP155	U2540W / CML511	11.96	23.91	50.00	10.87	4.68
SCHP113	I-38 / CKL05022	20.11	21.20	57.07	0.54	4.68
SCHP102	CZL068 / I-38	23.91	22.28	52.72	1.09	4.67
SCHP84	CML488 / U2540W	30.43	11.96	45.11	12.50	4.65
SCHP135	I-42 / CML547	14.67	21.20	60.87	2.17	4.53
SCHP32	CKDHL0089/ I-38	26.09	26.09	45.65	1.09	4.47
SCHP27	CK21 / CZL068	20.65	19.02	58.70	0.54	4.42
SCHP100	CML547 / U2540W	23.37	22.83	38.04	10.33	4.40
SCHP142	RO549W / CML511	20.11	29.35	44.02	5.43	4.39
SCHP71	CML442 / CZL068	23.37	21.74	51.09	0.54	4.38
SCHP139	RO549W / CML216	17.39	31.52	42.39	8.70	3.93
SCHP101	CZL068 / CK21	16.85	17.39	54.35	0.54	3.70
SCHP83	CML488 / RO549W	29.89	23.37	42.39	4.35	3.28
SCHP115	I-38 / CML216	19.57	21.74	57.61	0.00	2.83
Mean		23.73	23.48	48.05	3.6	5.55
SE		0.46	0.45	0.55	0.31	0.67

purity were generated through conventional system. This affirms that the maintenance of these inbred lines was carefully done for several generations of selfing for seed increase and they are suitable for high quality hybrid production. Similar findings were reported by [14] in his study using 1237 SNPs, where the majority of the inbred lines tested exhibited 100% homozygosity. However [15], reported that of the 265 inbred lines tested using 22,787 SNPs, only 22% of the inbred lines had 99.9% genetic purity. The higher levels of genetic purity observed in majority of the inbred lines used in this study indicates that the ARC-GCI's maize breeding program is efficient and quality oriented in terms of inbred line development and maintenance. However, inbred line CK21, which is currently used as a parent in some of the experimental hybrids had residual heterozygosity of 5.43%, which is slightly higher than the threshold of 5%. It should, therefore, be purified using ear-to-row selection methods. Lines with more than 15% residual heterozygosity are likely

to have been contaminated with pollen from unrelated genetic materials and should be discarded [3]. The higher level of heterogeneity observed in some of the inbred lines may be attributed to either that the inbred line is in the early generation of inbreeding or there was pollen contamination and/or seed admixture during maintenance breeding. An additional generation of inbreeding and extensive selection is highly recommended in order to fix these inbred lines. Some reports suggest that due to the strong inbreeding depression, higher levels of residual heterozygosity may have been deliberately maintained at early breeding level. Higher levels of residual heterozygosity were also reported among the inbred lines tested from Ethiopian Institute of Agricultural Research (EIAR) due to use of early generation inbred lines [15]. High level of residual heterozygosity may occur due to pollen contamination and/or seed mixture during seed regeneration, maintenance and bulking [16].

The rationale for doing parent-offspring test is to confirm if the particular hybrid is the true resultant F1 hybrid derived from the original inbred lines with no pollen contamination or within acceptable contamination levels [3]. Thus, the test provides a means to check if the pollination was done correctly during hybrid development. During hybridisation, there is a possibility of pollen contamination arising from self-pollination or cross-pollination from undesired neighbouring crops due to inadequate isolation distance. Therefore, parent-offspring test is important to ensure production of genuine quality hybrid seed.

Out of 158 experimental hybrids tested, 90 failed the test due to higher percentage of contamination (greater than 5%). This could be attributed partly to the use of genetically impure parental inbred lines or partly due to pollen contamination during hybrid production. The results revealed that the resultant hybrids derived from segregating inbred lines exhibited higher levels of genetic contamination of greater than 5%. Daniel et al. [17] reported a similar trend, where inbred lines with higher percentage of residual heterozygosity resulted in hybrids with higher contamination percentage. The results clearly confirm that the purity level of the parental inbred lines determines the purity of the resultant hybrids. In addition, it was evident that there was genetic contamination due to lack of pollen control between the crossing block and the neighbouring field as hybrids developed from pure inbred lines also revealed high levels of contamination. This could be due to failure in observing adequate isolation distance of not less than 360 m recommended in hybrid seed production. Genetic contamination of the experimental hybrids could also be as a result of incomplete detasselling [18]. However, this kind of impurity is usually tolerated at a level of 3–5%, without any effect on the yield performance. In this study, all of the top 10% high yielding hybrids had less than 5% contamination while some of the 10% low yielding hybrids had over 5% contamination. Use of genetically impure parental lines could also be another explanation for the observed generally low average hybrid grain yield (5.55 t ha^{-1}) in this study. This is because parental lines for the majority of top yielding hybrids passed purity test (Table 1) and the opposite was true for the bottom 10% hybrid. The finding concurs with Ipsilandis et al. [19], who reported a significant influence of contamination and pure pedigree to the ultimate yield performance of hybrids. This further confirms that pure hybrid seed has better competitive ability and yields better than low purity seed.

Polymorphic information content (PIC) gives an estimate of the discriminatory ability and effectiveness of molecular markers with respect to the number of alleles that are expressed and their relative frequencies [20]. Lander and Botstein [21] described PIC mean value of >0.50 as highly informative, 0.25–0.50 moderately informative and <0.25 is slightly informative. Hence, the mean PIC value of 0.44 observed in this study for the inbred lines confirm that the markers used were effective in discriminating the genotypes,

reasonably informative and of good quality. This value is on the higher side compared to PIC values reported in some of the past related studies [22, 23]. Hao et al. [24] reported PIC values within the range of 0.01 to 0.38 using 1536 SNP markers on 95 parental inbred lines in maize. In another study [25], reported PIC values ranging from 0.27 to 0.38 with a mean of 0.34 while using 884 SNP markers. The mean PIC value observed in this study is comparable to the one reported by [26] of 0.43 using 66 SSR markers. The disparity between the mean PIC observed in this study and the findings of earlier scientist could be attributed to differences in the composition of the genetic materials, sample size as well as the choice of number of markers used in the studies. The good quality of the 92 SNPs used in this study could be attributed to the fact that they were a subset of the 100 SNPs distributed across the 10 chromosomes in the maize genome, which were carefully selected and recommended by CIMMYT for quality control genotyping.

The mean observed heterozygosity (H_0) of 0.09 and mean inbreeding coefficient of 0.80 reported in this study among the inbred lines indicate that the majority of the lines attained an appreciable level of homozygosity and were fixed. The low level of residual heterozygosity and high inbreeding coefficient in the present study further confirmed the existence of high level of inbreeding [27], which was expected among the parental inbred lines. The hybrids on the other hand, revealed very low F_{IS} values ranging from -0.02 to 0.17, with a mean value of -0.05 validating an excess of heterozygotes, which was expected for hybrids.

It is critical that genetic relatedness of parental inbred lines be established to facilitate efficient exploitation of heterosis in hybrid breeding. The use of distanced based cluster approach to partition parental lines into distinct clusters is a reliable and appropriate approach to measure relatedness among individuals within a population [28]. Higher genetic distance of 0.56 observed between any combination of the following three parental inbred lines namely, CKDHL0089, CML443 and CB323 indicate that these lines originate from different ancestors while lowest genetic distance of 0.05 between lines I-42 and I-40 suggest that these are sister lines with shared pedigree across most generations. On average, majority of the inbred lines studied (92%) had their genetic distance values falling within the range of 0.4 to 0.6. In a related study [29], reported that on average, 67.7% of 94 maize inbred lines had genetic distances within the range of 0.32 to 0.42. CIMMYT conducted a similar study on 450 inbred lines and 95% of the pair of inbred lines showed genetic distances ranging between 0.3 and 0.5 [2]. Genetic distances among the studied inbred lines revealed the uniqueness of the studied lines and presence of substantial genetic variability that could be exploited in the breeding program.

Conclusion

This study showed that the set of SNP markers recommended for quality control test were effective and reliable in assessing genetic purity. The results of study will be helpful in the verification of genetic purity of maize hybrid seed developed by ARC-GCI. The inbred lines used in the present study were expected to be genetically pure with not more than 5% residual heterozygosity, but 33% of inbred lines showed residual heterozygosity of greater than 5% which requires additional generations of purification. Parent-offspring test conducted on 158 experimental hybrids led to the elimination of 60% hybrids since at least one of their parental inbred lines failed the genetic purity test. Of the 30% of the hybrids that passed the quality control test, seven high yield potential hybrids were recommended for further evaluation and release. Failure of some genotypes to pass inbred line genetic purity test and parent-offspring test suggests the need for further quality improvement by ARC-GCI maize-breeding program in the breeding nurseries and during pollinations for hybrid production.

Supporting information

S1 Table. List of 30 parental maize lines and 158 experimental single-cross hybrids sourced from ARC-GCI.
(DOCX)

S2 Table. List of 92 SNPs markers used in this study and their chromosomal position.
(DOCX)

S3 Table. Parent-offspring test for the evaluated 158 maize hybrids.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Chimwemwe Josia, Kingstone Mashingaidze, Julia Sibiya.

Data curation: Chimwemwe Josia, Assefa B. Amelework, Aleck Kondwakwenda, Cousin Musvosvi.

Formal analysis: Chimwemwe Josia, Assefa B. Amelework, Aleck Kondwakwenda, Cousin Musvosvi.

Funding acquisition: Kingstone Mashingaidze, Julia Sibiya.

Investigation: Chimwemwe Josia, Kingstone Mashingaidze, Aleck Kondwakwenda, Julia Sibiya.

Methodology: Chimwemwe Josia, Kingstone Mashingaidze, Aleck Kondwakwenda, Julia Sibiya.

Project administration: Kingstone Mashingaidze, Julia Sibiya.

Resources: Kingstone Mashingaidze, Julia Sibiya.

Software: Chimwemwe Josia, Assefa B. Amelework, Aleck Kondwakwenda, Cousin Musvosvi.

Supervision: Kingstone Mashingaidze, Julia Sibiya.

Validation: Kingstone Mashingaidze, Assefa B. Amelework, Aleck Kondwakwenda, Cousin Musvosvi, Julia Sibiya.

Visualization: Kingstone Mashingaidze, Assefa B. Amelework, Aleck Kondwakwenda.

Writing – original draft: Chimwemwe Josia, Kingstone Mashingaidze, Aleck Kondwakwenda, Cousin Musvosvi.

Writing – review & editing: Chimwemwe Josia, Kingstone Mashingaidze, Assefa B. Amelework, Aleck Kondwakwenda, Cousin Musvosvi, Julia Sibiya.

References

1. Shiferaw B, Prasanna BM, Hellin J, Bänziger M. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. Food Security. 2011; 3:307–327.

2. Semagn K, Beyene Y, Makumbi D, Mugo S, Prasanna BM, Magorokosho C, et al. Quality control genotyping for assessment of genetic identity and purity in diverse tropical maize inbred lines. Theoretical

and Applied Genetics. 2012; 125:1487–1501. https://doi.org/10.1007/s00122-012-1928-1 PMID: 22801872

3. Gowda M, Worku M, Nair SK, Palacios-Rojas N, Huestis G, Prasanna B. Quality Assurance/Quality Control (QA/QC) in Maize Breeding and Seed Production: Theory and Practice. CIMMYT. 2017.

4. Semagn K, Bjørnstad A, Ndjiondjop M. An overview of molecular marker methods for plants. African Journal of Biotechnology. 2006; 5:2540–2568.

5. Hamblin MT, Warburton ML, Buckler ES. Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PloS one. 2007; 2:e1367. https://doi.org/10.1371/journal.pone.0001367 PMID: 18159250

6. Ertiro BT, Ogugo V, Worku M, Das B, Olsen M. Labuschagne M.et al. Comparison of Kompetitive Allele Specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize. BMC genomics. 2015; 16:908–910 https://doi.org/10.1186/s12864-015-2180-2 PMID: 26545737

7. Chen J, Zavala C, Ortega N, Petroli C, Franco J, Burgueño J, et al. The development of quality control genotyping approaches: A case study using elite maize lines. PLoS ONE. 2016; 11. https://doi.org/10.1371/journal.pone.0157236 PMID: 27280295

8. Salgado KC, Vieira DC, Von Pinho ÉR, Guimarães CT, Von Pinho RG, Souza LVD. Genetic purity certificate in seeds of hybrid maize using molecular markers. Revista Brasileira de Sementes. 2006; 28:169–175.

9. Peakall R, Smouse P. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes. 2006; 6:288–295.

10. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Nat. Acad. Sci. U. S. A. 1979; 76: 5269–5273. https://doi.org/10.1073/pnas.76.10.5269 PMID: 291943

11. Perrier X, Jacquemoud-Collet J. DARwin software: Dissimilarity analysis and representation for windows. Website http://darwin.cirad.fr/darwin. 2006. [accessed 1 September 2018].

12. Geiger H, Gordillo G. Doubled haploids in hybrid maize breeding. Maydica. 2009; 54: 485–499.

13. Chang MT, Coe EH. Doubled haploids. Molecular genetic approaches to maize improvement. Springer. 2009; 127–142.

14. Dao A. Genetic characterization, association mapping and evaluation of feterotic patterns of tropical maize (*Zea mays* L.) inbred lines under drought and non-drought conditions. PhD thesis,University of Ghana. 2013.

15. Ertiro BT, Semagn K, Das B, Olsen M, Labuschagne M, Worku M. Genetic variation and population structure of maize inbred lines adapted to the mid-altitude sub-humid maize agro-ecology of Ethiopia using single nucleotide polymorphic (SNP) markers. BMC genomics. 2017; 18:777–779. https://doi.org/10.1186/s12864-017-4173-9 PMID: 29025420

16. Warburton ML, Setimela P, Franco J, Cordova H, Pixley K, Bänziger M et al. Toward a cost-effective fingerprinting methodology to distinguish maize open-pollinated varieties. Crop Science. 2010; 50:467–477.

17. Daniel I, Adetumbi J, Oyelakin O, Olakojo S, Ajala M, Onagbesan S. Application of SSR markers for genetic purity analysis of parental inbred lines and some commercial hybrid maize *(Zea mays L.)*. American Journal of Experimental Agriculture. 2012; 2:597–606.

18. Moreira JN, Silva PSL, Silva K, Dombroski JL, Castro RS. Effect of detasseling on baby corn, green ear and grain yield of two maize hybrids. Horticultura Brasileira. 2010; 28:406–411.

19. Ipsilandis CG, Vafias BN, Karagiozopoulou A, Goulas CK. F1 Single-cross maize hybrid performance under low purity conditions. Asian Journal of Plant Sciences. 2005; 4: 75–82.

20. Abakemal D, Hussein S, Derera J, Semagn K. Genetic purity and patterns of relationships among tropical highland adapted quality protein and normal maize inbred lines using microsatellite markers. Euphytica. 2015; 204: 49–61.

21. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 1989; 121:185–199. PMID: 2563713

22. Dao A, Sanou J, Mitchell SE, Gracen V, Danquah EY. Genetic diversity among INERA maize inbred lines with single nucleotide polymorphism (SNP) markers and their relationship with CIMMYT, IITA, and temperate lines. BMC Genetics. 2014; 15:127. https://doi.org/10.1186/s12863-014-0127-2 PMID: 25421948

23. Nyombayire A, Derera J, Sibiya J, Gasura E, Aboyisonga CN. Genetic diversity among maize inbred lines selected for the mid- altitudes and highlands of Rwanda. Maydica. 2016; 61: 1–7.

24. Hao Z, Li X, Xie C, Weng J, Li M, Zhang D et al. Identification of functional genetic variations underlying drought tolerance in maize using SNP markers. Journal of integrative plant biology. 2011; 53:641–652. https://doi.org/10.1111/j.1744-7909.2011.01051.x PMID: 21564545

25. Yang X, Xu Y, Shah T, Li H, Han Z, Li J et al. Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica. 2011; 139:1045 https://doi.org/10.1007/s10709-011-9606-9 PMID: 21904888

26. Adeyemo O, Omidiji O. SSR-based and carotenoid diversity assessment of tropical yellow endosperm maize inbred lines. Plant Genetic Resources. 2013; 12: 67–73.

27. Labate J, Lamkey K, Sharon EM, Kresovich S, Sullivan H, Smith CK. Molecular and historical aspects of corn-belt dent diversity. Crop Science. 2003; 43:80–91.

28. Garcia AA, Benchimol LL, Barbosa AM, Geraldi IO, Souza CL Jr, Souza APD. Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. Genetics and Molecular Biology. 2004; 27:579–588.

29. Boakyewaa AG, Badu-Apraku B, Akromah R, Garcia-Oliveira AL, Awuku FJ, Gedil M. Genetic diversity and population structure of early-maturing tropical maize inbred lines using SNP markers. PloS one. 2019; 14:e0214810 https://doi.org/10.1371/journal.pone.0214810 PMID: 30964890