Check for updates

METHOD ARTICLE

REVISED Developing and deploying an efficient genotyping workflow for accelerating maize improvement in developing countries [version 3; peer review: 3 approved, 1 approved with reservations]

Queen Offornedo (iD), Abebe Menkir, Deborah Babalola, Melaku Gedil (iD)

Bioscience Center and Maize Improvement Program, International Institute of Tropical Agriculture (IITA) Headquarters, Ibadan, Oyo State, 200001, Nigeria

Abstract

Background: Molecular breeding is an essential tool for accelerating genetic gain in crop improvement towards meeting the need to feed an ever-growing world population. Establishing low-cost, flexible genotyping platforms in small, public and regional laboratories can stimulate the application of molecular breeding in developing countries. These laboratories can serve plant breeding projects requiring low- to medium-density markers for marker-assisted selection (MAS) and quality control (QC) activities.

Methods: We performed two QC and MAS experiments consisting of 637 maize lines, using an optimised genotyping workflow involving an in-house competitive allele-specific PCR (KASP) genotyping system with an optimised sample collection, preparation, and DNA extraction and quantitation process. A smaller volume of leaf-disc size plant samples was collected directly in 96-well plates for DNA extraction, using a slightly modified CTAB-based DArT DNA extraction protocol. DNA quality and quantity analyses were performed using a microplate reader, and the KASP genotyping and data analysis was performed in our laboratory.

Results: Applying the optimized genotyping workflow expedited the QC and MAS experiments from over five weeks (when outsourcing) to two weeks and eliminated the shipping cost. Using a set of 28 KASP single nucleotide polymorphisms (SNPs) validated for maize, the QC experiment revealed the genetic identity of four maize varieties taken from five seed sources. Another set of 10 KASP SNPs was sufficient in verifying the parentage of 390 F_1 lines. The KASP-based MAS was successfully applied to a maize pro-vitamin A (PVA) breeding program and for introgressing the aflatoxin resistance gene into elite tropical

Open Peer Review

Approval Status ? ✓ ✓ ✓

	1	2	3	4
version 3 (revision) 03 Aug 2022		✓ view	✓ view	✓ view
version 2 (revision) 10 Jun 2022	? view	? view		
version 1 17 Jan 2022	✗ view			

1. **John Damien Platten**, International Rice Research Institute, Los Baños, Philippines

2. **Bhoja R. Basnet** (iD), International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

3. **Karthikeyan Adhimoolam** (iD), Tamil Nadu Agricultural University, Madurai, India
   Jeju National University, Jeju-si, South Korea

4. **Ruairidh J H Sawers**, The Pennsylvania State University - University Park Campus, University Park, USA

maize lines.

Conclusion: This improved workflow has helped accelerate maize improvement activities of IITA's Maize Improvement Program and facilitated DNA fingerprinting for tracking improved crop varieties. National Agricultural Research Systems (NARS) in developing countries can adopt this workflow to fast-track molecular marker-based genotyping for crop improvement.

Keywords
Molecular breeding, KASP, Genotyping workflow, Marker-assisted selection, Quality Control, National Agricultural Research Systems (NARS), Developing countries

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Melaku Gedil (M.Gedil@cgiar.org)

How to cite this article: Offornedo Q, Menkir A, Babalola D and Gedil M. **Developing and deploying an efficient genotyping workflow for accelerating maize improvement in developing countries [version 3; peer review: 3 approved, 1 approved with reservations]** Gates Open Research 2022, **6**:3 https://doi.org/10.12688/gatesopenres.13338.3

First published: 17 Jan 2022, **6**:3 https://doi.org/10.12688/gatesopenres.13338.1

REVISED   **Amendments from Version 2**

Version 3 has been updated based on the reviewers' comments on the previous version.

We have included additional context to the Introduction section (P3) to articulate the use cases better and provided a genotyping cost comparison of the procedure as indicated by Reviewer 1.

As pointed out by Reviewer 2, we have rephrased the indicated sentences in the Introduction and Methods section. The misplaced reference has been replaced and listed in the reference table. We have included additional context to suitably articulate the study objective towards the end of the Introduction section, which is "This study aims to develop a genotyping workflow optimized for cost-effective and fast turn-around time that can be deployed by less sophisticated and reasonably equipped laboratories in developing countries, to accelerate maize improvement research." We have furnished Table 1 with details of the exact number of genotypes and samples used for the experiments. We have also provided a new figure (Figure 5) and table (Table 5) to aptly articulate the KASP genotyping analysis for the hybrid verification experiment.

Any further responses from the reviewers can be found at the end of the article

Introduction

Agriculture is the mainstay of millions of low-income households in Sub-Saharan Africa (SSA). However, productivity is way below the yield potential of significant crops due to several interacting factors contributing to the yield reduction. The paucity of nutritionally improved resilient crop varieties is a crucial constraint. This constraint can be mitigated by the rapid development of cultivars adapted to specific agroecology zones[1]. The current yield gain trend in major food crops has shown that relying on conventional breeding alone is insufficient to meet the food needs of an estimated nine billion people in 2025[2]. There is a need to accelerate genetic gain by deploying new breeding strategies[3,4]. This need has led to the scientific community's massive investment in developing genomic resources and support systems, to provide valuable tools to accelerate breeding processes[5].

Various bottlenecks have hindered the substantial impact of molecular breeding for crop improvement, particularly in developing countries[6,7]. The major limiting factors are a lack of infrastructure and capacity for genomics resources and poor information flow, resulting in reduced access to operational and decision support tools[8]. Private companies in developed countries usually own the proprietary rights to many emerging genomics resources and systems, making it difficult for public research sectors, non-profit research institutes, and small laboratories in developing countries to have direct access. These challenges are being curbed by various international initiatives such as the Excellence in Breeding (EiB) platform, which coordinates its activities with the Genomic and Open-source Breeding Informatics Initiative (GOBii), and High Through-Put Genotyping (HTPG). In addition, the Integrated Breeding Platform (IBP)-hosted Generation Challenge Program (GCP) and the Breeding Management System (BMS)[9] target the development and adoption of molecular breeding in developing countries. These and other consultative group-hosted initiatives and

platforms galvanise worldwide partners drawn from public, private, and governmental institutions towards the common goal of increasing agricultural productivity through efficient tools, technologies, and data management systems[6].

Despite the availability of many low-cost genotyping platforms and resources, it is not easy to meet the genotyping needs of many users who work on different crops, different locations, and often fewer samples due to cost implications[7,8]. The current available genotyping platforms have a minimum sample size requirement. For instance, the EiB facilitated genotyping at Intertek offers reduced cost if the user orders genotyping of 1536 samples; fewer samples are acceptable, but the price increases. Intertek's standard cost for routine KASP genotyping is $2.6 per sample per 10 SNPs, excluding shipping costs, compared to our in-house genotyping at $2.95. Even though large volume sizes can be consolidated and shipped for genotyping, there are times when breeders and partners may want to fingerprint a few dozen lines for identity or parentage analysis for quick decision making. In such cases, sending less than the minimum number of samples is not only more priced per datapoint but entails shipping cost and a turn-around time of 2–3 weeks. Using other markers, such as SSR, is more expensive and cumbersome. The use of genotyping systems such as KASP in-house alleviates all these issues. Also, the issue of inefficient courier services in this part of the world, which often results in reduced or damaged perishable specimens, can be circumvented if a reasonably affordable system is available locally. More so, we re-purposed standard laboratory instruments for the genotyping workflow. For instance, the qPCR machine, which is mostly used for expression analysis, was adapted to KASP genotyping with the installation of appropriate software for SNP calling. Likewise, the Fluostar plate reader was used for plate-level DNA quantification in lieu of single sample analysis by Spectrophotometer.

For these reasons it is imperative to devise a sustainable strategy for routine, cost-effective, and easily accessible genotyping services to complement these international outsourcing initiatives by providing in-house or local (regional) genotyping platforms, where possible, to accelerate the genotyping workflow. One such regional initiative in Africa is the Integrated Genotyping Support Services (IGSS) genotyping facility at Biosciences eastern and central Africa/International Livestock Research Institute (BeCA/ILRI), Kenya. This strategy will allow breeders to outsource to a regional genotyping service provider or set up a core facility in-house.

One factor that influences breeders' choice of genotyping platform is the level of throughput. Other factors considered are the data turn-around time, ease of data analysis (available informatics), reproducibility, flexibility, and cost per datapoint or cost per sample[10,11]. For high and ultra-high throughput markers, breeders outsource to array- and sequenced-based genotyping service providers. These platforms are suitable for discovery applications and approaches requiring hundreds to thousands of samples to be genotyped with tens to thousands of markers, such as genome-wide association studies (GWAS), gene mapping, and large-scale genomic selection[10,12] They are also suitable for genotyping a few samples with many markers

(multiplexing), such as genetic diversity analysis or background selection. While multiplex platforms provide higher throughput with lower reagent consumption, it limits scientists to using a multiplexed set of several thousand single nucleotide polymorphisms (SNPs) per assay[13]. They are also demanding in informatics resources and presently produce datasets with a significant percentage of missing data[13]. The high cost per sample and the initial assay development time of highly multiplexed platforms can be problematic for crop improvement applications, usually requiring low- to medium-density markers[11]. For these low- to mid-density genotyping approaches, a uniplex SNP genotyping platform is appropriate[14].

Uniplex genotyping assays are low-throughput genotyping systems that are ideally flexible regarding assay design, ease of running, and cost-effectiveness[15]. These systems provide plant breeders with the flexibility to mix and match different SNPs for a given sample set. They allow breeders to use a smaller subset of informative SNPs such as functional SNPs and trait-specific haplotypes, thereby eliminating the generation of unintended datapoints when using fixed-array SNPs. Even though a range of uniplex SNP genotyping assays exists, the most competitive uniplex systems that have been successfully applied in crop improvement research are TaqMan[16–19], competitive allele-specific PCR (KASP)[11,20], Amplifuor[21], and rhAmP[22] assays. These uniplex genotyping systems vary in reaction chemistry, detection method, and reaction format. Uniplex systems can either be outsourced or installed in-house.

In this study, we utilised the KASP assay, as it is one of the most used assays among plant breeders and biologists[15,19]. KASP is an endpoint PCR-based SNP genotyping method from KBiosciences, now LGC Biosearch Technologies, UK. KASP uses fluorescently-labelled allele-specific primers for the bi-allelic discrimination of SNPs and insertion-deletion mutations (INDELs)[23]. KASP was developed to reduce cost, mainly from probe design, and improve genotyping efficiency, becoming a preferred alternative to TaqMan[11,24]. The KASP genotyping system has been successfully applied in crops such as maize[11,15,25], wheat[10,20,26], rice[27], soybean[28], peanut[29], amongst others. KASP has developed into a global benchmark technology for genotyping crop plants[11,23,29–31] following the validation of KASP markers across crops of global importance (such as maize - 1250 markers, wheat - 1864 markers, and rice - 2015 markers) by the Generation Challenge Program of the Integrated Breeding Platform[9]. The International Maize and Wheat Improvement Center (CIMMYT) has successfully utilised the 1,250 maize KASP markers for various genetic applications, including quantitative trait loci (QTL) mapping, marker-assisted recurrent selection (MARS), allele mining, and QC analysis[11]. The Maize Improvement Program of the International Institute of Tropical Agriculture (IITA) has generated over 2,000 datapoints using KASP in-house for different genotype analyses, including QC and MAS.

However, some bottlenecks in the genotyping workflow slow down the genotyping process, delaying crop improvement: (1) method of sample collection and processing, (2) level of DNA extraction and quantitation, and (3) DNA-based genotyping[8]. Gedil and Menkir (2019) provided a thorough review of the Maize Improvement Program's (MIP) molecular marker-based crop improvement activities. However, reports of research accelerating the entire genotyping process by minimizing these bottlenecks and providing a cost-effective genotyping workflow suitable for small-scale breeders and laboratories in developing countries are lacking. This study aims to develop a genotyping workflow optimized for cost-effective and fast turn-around time which can be deployed by less sophisticated and reasonably equipped laboratories in developing countries, to accelerate maize improvement research.

Methods
Plant materials
The overall genotyping workflow was applied in some experiments representative of the genotyping activities common in small to medium breeding programs. Table 1 below describes the plant materials used in each experiment. The genetic identity experiment was performed using four well-adapted maize varieties originating from IITA but regenerated at four locations. For the hybrid verification experiment, 60 maize F_1 progenies originating from five bi-parental crosses were used. Lines KS23-3, KS23-5, and KS23-6 are resistant to maize lethal necrosis (MLN) disease, while IITATZI1653 and IITATZI1667 are IITA-adopted elite maize lines with high PVA content. Another 330 F_1 plants originating from four sets of bi-parental crosses involving *Striga*-susceptible (TZdEEI 102, TZdEEI 99, TZdEEI 4, and TZdEEI 13) and *Striga*-resistant (TZEEI 29, and TZEEI 79) parents were also screened to identify true hybrids. A total of 70 PVA-QPM enriched maize inbred lines were genotyped to select lines harbouring the favourable allele for the *crtRB1* gene associated with PVA content in maize. In the fourth breeding cycle of the maize enrichment project using marker-assisted backcrossing to introgress resistance to aflatoxin accumulation in elite tropical maize lines, we genotyped a total of 159 BC_1S_2 maize lines. We applied a 15% selection intensity to identify lines harbouring the favourable alleles of the QTLs associated with resistance to aflatoxin accumulation. These plants were grown in maize fields at IITA Ibadan, Nigeria.

Sample collection and preparation, and DNA extraction and quantitation
A total of 16 to 20 leaf discs were collected from young leaves of each tagged plant, directly into Corning 96-well Polypropylene 1.2 ml cluster tubes with strip caps (Merck, Germany) using Haris Uni-core 4.0 mm puncher and cutting mat (Merck, Germany). Two 4.0 mm stainless steel grinding balls (SPEX SamplePrep) were placed in each tube. Plant tissues were preserved on ice for transport from the field to the laboratory. They were stored in a -80°C freezer before lyophilising for 48 hours using FreeZone Freeze Dryer (Labconco) following the manufacturer's manual. Lyophilised leaf tissues were ground into powder by shaking at 1,500 strokes per minute for 1.5 min using an automated high-throughput tissue homogeniser, Geno/Grinder 2010 (SPEX SamplePrep).

Genomic DNA was extracted from ground leaf tissues using a cetyltrimethylammonium bromide (CTAB)-based DNA extraction method as described by Diversity Array Technology (DArT)[32] with minor modifications (Table 2). Dry leaf tissues

Table 1. Plant materials used for the experimentation of the optimized genotyping workflow.

S/N	Experiments	Genotypes (Parental maize lines: traits)	Population Development (crosses)	No of samples
1	**Genetic identity**	• **SAMMAZ 15 (1WDC2SynF2):** Medium maturing, good seed quality, high yield potential, tolerance to Striga hermonthica. (Y-6.9t/ha) • **SAMMAZ 16 (TZLComp1SynW-1):** Late maturing, good seed quality, high yield, resistance to Striga hermonthica. (6.4t/ha) • **SAMMAZ 27 (EV99DT-W-STR):** Drought tolerant and Striga resistant. (5.5t/ha) • **SAMMAZ 39 (PVA SYN8):** Intermediate-level pro-vitamin A content (6.4µg/g), high yield potential. (6.8t/ha)	Performed using four well-adapted maize varieties originating from IITA but regenerated at four locations. Maize seedlings were grown in pots for about two weeks until they reached the three-four-leaf stage in a screen house at the Bioscience Center of IITA Ibadan, Nigeria.	20 maize lines resulting from 4 genotypes by 5 locations.
2	**Hybrid verification:**	• **KS23-3, KS23-5, and KS23-6:** Maize lethal necrosis (MLN) resistant maize lines • **IITATZI1653 and IITATZI1667:** Maize lines with high PVA content • **TZdEEI 102, TZdEEI 99, TZdEEI 4, and TZdEEI 13:** Striga susceptible maize inbred lines • **TZEEI 29 and TZEEI 79:** Striga resistant maize inbred lines.	**Set 1a:** KS23-3 x IITATZI1653; **Set 2a:** KS23-5 x IITATZI1653; **Set 3a:** KS23-6 x IITATZI1653; **Set 4a:** KS23-3 x IITATZI1667; **Set 5:** KS23-5 x IITATZI1667; **Set 1b:** TZEEI 29 x TZdEEI 99; **Set 2b:** TZdEEI 4 x TZEEI 79; **Set 3b:** TZEEI 79 x TZdEEI 13; **Set 4b:** TZdEEI 102 x TZEEI 29 Seedlings for the F_1 plants were grown in a maize field at IITA Ibadan, Nigeria.	• **Set a:** 60 F_1 maize lines originating from five crosses involving three KS23 (MLN-resistant) lines and two PVA enriched maize lines. • **Set b:** 330 F_1 maize lines originating from four bi-parental crosses involving two Striga resistant maize lines and four Striga susceptible lines.
3	**Marker-assisted selection**	• **PVA-QPM** enriched maize inbred lines were genotyped to select lines harbouring the favourable allele for the crtRB1 gene associated with PVA content in maize. • **Backcross (BC_1S_2)** maize lines in the fourth breeding cycle of the maize enrichment project; using marker-assisted backcrossing to introgress resistance to aflatoxin accumulation in elite tropical maize lines	Ten plant stands per row were planted for each inbred, and leaf tissues were collected from each row for DNA extraction by bulking leaves from all ten plant stands per row. For the aflatoxin population, we applied a 15% selection intensity to identify lines harbouring the favourable alleles of the QTLs associated with resistance to aflatoxin accumulation. All maize lines were grown at IITAs maize field, Ibadan, Nigeria.	• 70 PVA-QPM maize lines • 159 BC_1S_2 maize lines

Legend: PVA = Pro-vitamin A; QTL = Quantitative trait loci. Source of plant materials: Maize Improvement Program, International Institute of Tropical Agriculture (IITA) Headquarters, Ibadan, Nigeria.

Table 2. DArT DNA extraction protocol with minor modification.

The chemicals and reagents used were as outlined in the Diversity Array Technology (DArT) Plant DNA extraction protocol (Accessed on June 2, 2020).

Extraction procedure:

1. Aliquot freshly prepared, well-mixed "fresh buffer solution" and preheat in a 65°C water bath.
2. Grind sample leaf discs in 1.2 ml cluster tubes using a Geno/Grinder 2010 (Spex Sample Prep) to a fine powder
3. Add 500 µl buffer solution to dissolve the powder completely
4. Incubate at 65°C for 1 hr, with gentle shaking
5. Cool down for 5 min and add 500 µl of chloroform: isoamyl alcohol (24:1) mixture
6. Mix well by gentle inversion for 30 min, and spin for 20 min, at 10,000 x g, at room temperature
7. Transfer about 400 µl of the water phase to a fresh 1.2 ml tube, add the same volume of ice-cold isopropanol and invert the tube approximately ten times, nucleic acids should become visible
8. Incubate for 30 min at -20 °C, and spin for 30 min, at 10,000 x g, at room temperature
9. Discard supernatant, and wash pellet with 400 µl 70 % EtOH
10. Discard EtOH, dry pellet and dissolve in 100 µl of nuclease-free water-RNAseA solution in a 90:10 ratio.

were used instead of fresh ones; we included a 30-minute incubation period during the alcohol precipitation step; the DNA pellet was resuspended in a nuclease-free water and RNaseA solution. The DNA quality and quantity were determined by spectrophotometry using the FLUOstar Omega Microplate Reader (BMG LABTECH) following the manufacturer's manual.

KASP genotyping and data analysis

The isolated genomic DNA was diluted to a working concentration of 30 ng/µl and used as template DNA for the KASP genotyping reaction. A total of 28 KASP SNPs were used to determine the selected maize varieties' genetic identity, while 10 KASP SNPs were used to verify true hybrids among the F_1 maize lines. The SNPs (Table 3) were taken from a maize QC SNP panel[9] recommended by CIMMYT[7,33] and chosen for their high polymorphic information content (PIC) and uniform maize genome coverage. Trait-specific KASP markers (Table 4) were used to screen BC_1S_2 lines carrying the favourable allele for resistance to aflatoxin accumulation and identify inbred lines with high PVA content. The KASP reaction was performed in 96- and 384-well plates. For the 96-well plate, a total reaction volume of 10 µl consisting of 5 µl template DNA and 5 µl of the prepared genotyping mix (2×KASP master mix and primer mix) was used. In contrast, for the 384-well plate, a total reaction volume of 5 µl consisting of 2.5 µl template DNA and 2.5 µl of the prepared genotyping mix was used. All reaction was performed following the KASP manual (accessed on June 24, 2020). The KASP assay and master mix were purchased from LGC Biosearch Technologies (LGC Group). The amplification reaction was run in-house (Bioscience Centre of IITA Ibadan, Nigeria) using the LightCycler 480 II PCR System (Roche Life Sciences, Germany) and GeneAmp PCR System 9700 (Applied Biosystems, USA). The description of the parameters for the LC480 II qPCR machine is outlined in the LC480 operator's manual. To perform the KASP genotyping experiment on the LC480 II machine, we used the Endpoint Genotyping Analysis module within the LightCycler software, adjusting the parameters as outlined in the KASP genotyping protocol provided by LGC Biosearch Technologies. The Endpoint genotyping analysis module is based on the use

of dual hydrolysis probes, which are designed for wild-type and mutant target DNA and are labelled with different dyes (FAM and HEX). However, when using a non-qPCR machine (such as the GeneAmp PCR System 9700) for amplification, a third colour probe (ROX) normalizes the fluorescence measurement. The LightCycler software within the LC480 II machine determines the sample genotypes automatically by measuring the intensity distribution of the two probes after a PCR amplification step. The relative dye intensities are then visualized in a scatter (cluster) plot that discriminates them as wild-type, heterozygous mutant, or homozygous mutant samples. The LightCycler software automatically groups similar samples and assigns genotypes based on the intensity distribution of the two dyes. The KASP amplification conditions included one cycle of KASP unique Taq activation at 94°C for 15 min, followed by 36 cycles of denaturation at 94°C for 20 s, and annealing and elongation at 60°C (dropping 0.6°C per cycle) for 1 min. Endpoint detection of the fluorescence signal was acquired for 1 min at 30°C when using the LightCycler 480 II real time-PCR System or read using the FLUOstar Omega Microplate reader (BMG Labtech, SA) when using the GeneAmp PCR System 9700. For fluorescence detection, the filter combination for the Excitation and Emission wavelength of both dyes was set at 465 – 533 (FAM) and 523 – 568 (HEX), respectively, when using LC480 II, and 485 - 520 (FAM), 544 - 590 (HEX) and 584 - 620 (ROX) when using FLUOstar Omega Microplate reader. The genotype calls were exported from the LightCycler software as fluorescent intensities of each sample in ".txt" file format and imported for analysis in the KlusterCaller analysis software (LGC Biosearch Technologies). The KlusterCaller software adjusted the cluster plot axes to enable the proper calling of genotypes. The genotype calls were grouped as homozygous for allele X (allele reported by FAM, X-axis), homozygous for allele Y (allele reported by HEX, Y-axis), heterozygous (alleles reported by FAM and HEX, between X- and Y-axis), or uncallable. The result from the KlusterCaller was exported in two file formats (".csv" and ".txt"). The ".csv" file was imported into the SNPviewer2 version 4.0.0 software (LGC Biosearch Technologies), where the cluster plot image was viewed and downloaded for publication. The genotype calls in the ".txt"

Table 3. List of KASP single nucleotide polymorphisms (SNPs) used in the QC experiments.

SNP ID	Linkage group	Position (cM)	Allele X	Allele Y	Trait category	Analysis	Dataset
ae1_7	5	79	A	G	QC	GID & HV	GCP/IBP-Maize
PHM15331_16	10	28	A	G	QC	GID	GCP/IBP-Maize
PHM2438_28	4	12	A	G	QC	GID	GCP/IBP-Maize
PHM2770_19	10	36	A	C	QC	GID	GCP/IBP-Maize
PHM3466_69	6	108	A	G	QC	GID	GCP/IBP-Maize
PHM5181_10	9	26	C	T	QC	GID & HV	GCP/IBP-Maize
PHM5502_31	3	58	A	G	QC	GID & HV	GCP/IBP-Maize
PZA00413_20	3	60	A	C	QC	GID & HV	GCP/IBP-Maize
PZA00726_10	4	55	A	C	QC	GID	GCP/IBP-Maize
PZA01216_1	1	116	A	G	QC	GID & HV	GCP/IBP-Maize
PZA01456_2	10	61	A	G	QC	GID	GCP/IBP-Maize
PZA01477_3	4	81	C	T	QC	GID	GCP/IBP-Maize
PZA01533_2	7	112	A	G	QC	GID	GCP/IBP-Maize
PZA01885_2	2	115	A	G	QC	GID & HV	GCP/IBP-Maize
PZA01919_2	10	44	C	G	QC	GID & HV	GCP/IBP-Maize
PZA02090_1	3	15	A	T	QC	GID & HV	GCP/IBP-Maize
PZA02164_16	5	70	A	G	QC	GID & HV	GCP/IBP-Maize
PZA02269_3	1	149	C	T	QC	GID & HV	GCP/IBP-Maize
PZA02358_1	4	31	A	G	QC	GID	GCP/IBP-Maize
PZA02378_7	2	64	A	G	QC	GID	GCP/IBP-Maize
PZA02741_1	1	91	C	T	QC	GID	GCP/IBP-Maize
PZA02746_2	8	94	G	T	QC	GID	GCP/IBP-Maize
PZA02779_1	4	108	A	G	QC	GID & HV	GCP/IBP-Maize
PZA03135_1	8	57	A	C	QC	GID & HV	GCP/IBP-Maize
PZA03363_1	7	49	A	G	QC	GID & HV	GCP/IBP-Maize
PZA03605_1	10	75	A	G	QC	GID	GCP/IBP-Maize
PZB01658_1	6	28	A	T	QC	GID & HV	GCP/IBP-Maize
sh1_12	9	18	A	G	QC	GID & HV	GCP/IBP-Maize

LEGEND: QC = Quality control; GID = Genetic Identity; HV = Hybrid verification; GCP/IBP = Generation Challenge Program/Integrated Breeding Platform. **Source**: Integrated Breeding Platform (Accessed June 26, 2020).

file were used to calculate the genetic distance using the PowerMaker 3.25 statistical software[34].

Source data
The list of KASP SNPs for genotyping maize was obtained freely from the Integrated Breeding Platform website.

The trait-specific KASP SNPs (Supplementary Table 1, *Underlying data*) and QC KASP SNPs (Supplementary Table 2, *Underlying data*) were purchased as KBDs (KASP-by-Design)

from LGC Biosearch Technologies, UK, for use in our laboratory.

Results
Optimising in-house genotyping workflow
Our laboratory's routine sampling procedure spans seven days, from plant sampling and preparation to DNA extraction and quantitation. We present an expedited workflow (Figure 1) that ensures a good sample tracking system. Firstly, barcoding software, barcode readers, barcode labels, and barcode

Table 4. List of trait-specific KASP single nucleotide polymorphisms SNPs used in the MAS experiment.

SNP ID	Chromosome No.	FAM allele	HEX allele	Trait category	analysis	Source
S1_85016181	1	C	G	Aflatoxin	MAS	CIMMYT/IITA
S3_14863214	3	G	A	Aflatoxin	MAS	CIMMYT/IITA
S3_90027035	3	A	G	Aflatoxin	MAS	CIMMYT/IITA
S3_90023939	3	T	A	Aflatoxin	MAS	CIMMYT/IITA
S3_179639685	3	C	G	Aflatoxin	MAS	CIMMYT/IITA
S3_14229695	3	T	C	Aflatoxin	MAS	CIMMYT/IITA
S5_182519023	5	A	G	Aflatoxin	MAS	CIMMYT/IITA
S5_63229636	5	C	A	Aflatoxin	MAS	CIMMYT/IITA
S5_198883041	5	T	A	Aflatoxin	MAS	CIMMYT/IITA
PHM12859_7	3	C	T	Aflatoxin	MAS	CIMMYT/IITA
PZA02792_16	5	T	C	Aflatoxin	MAS	CIMMYT/IITA
MZA4145_18	3	A	G	Aflatoxin	MAS	CIMMYT/IITA
snpZM0015	10	A	G	PVA	MAS	CIMMYT

LEGEND: MAS = Marker-assisted selection; PVA = Provitamin A; CIMMYT = International Maize and Wheat Improvement Center; IITA = International Institute of Tropical Agriculture.
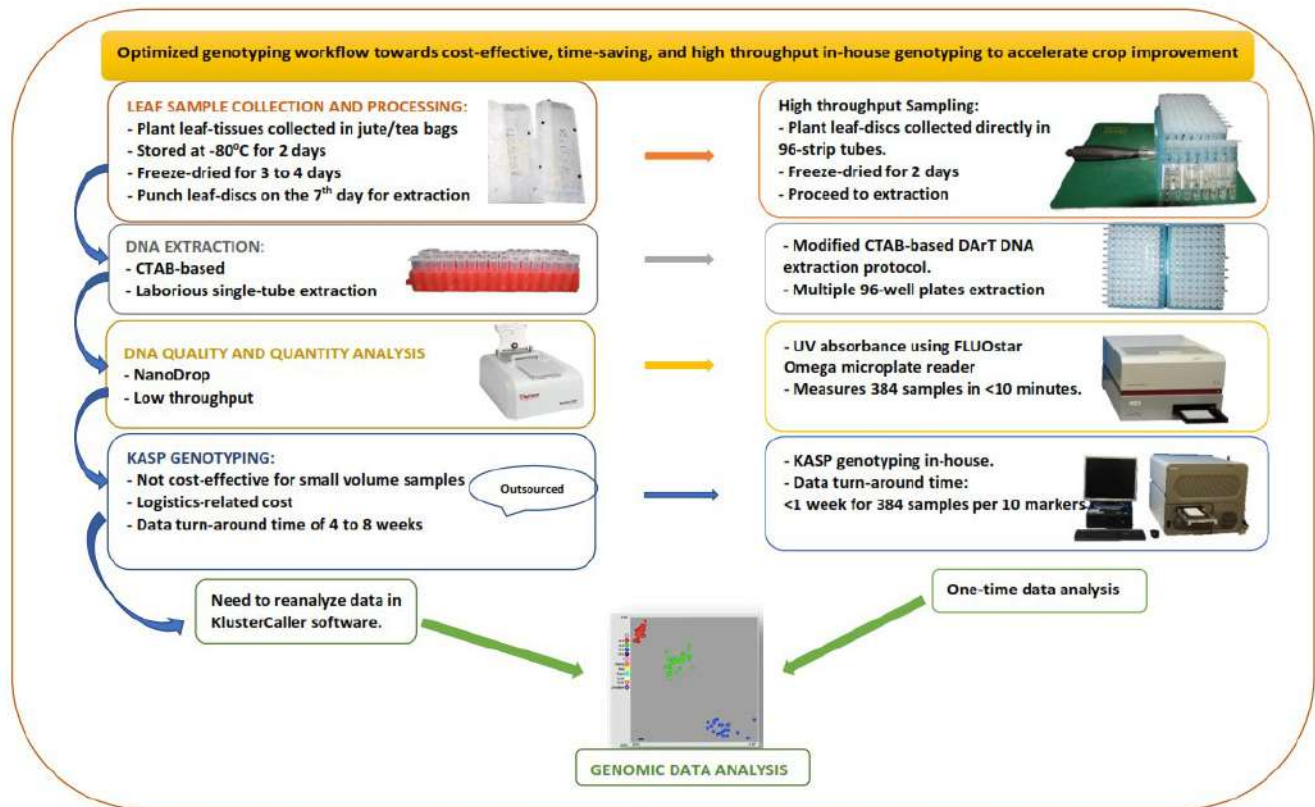

Figure 1. Diagram showing improvement to minimize bottlenecks in the genotyping workflow.

printers were introduced to facilitate sample tracking and data management. Waterproof/tear-proof tags and labels designed using BarTender barcoding software (Seagull Scientific) were printed using ZT230 Printer (Zebra, USA) and attached to plants before sample collection. Plate maps created in the BarTender software were linked to the sample location on the field and in the lab storage facility. Next, young plant leaf tissues were collected by punching leaf discs directly into the 96-well 1.2 mL polypropylene cluster tubes in wet-ice cooler bags, which reduced the sampling time and the time required for freeze-drying.

The sample DNA was extracted using the DArT DNA extraction protocol, slightly modified to maximise reagent and increase throughput, by using a reduced volume of reagents optimised to extract maize DNA from a smaller amount of leaf tissue (16–20 leaf discs, 4.0 mm). We also used freeze-dried leaf tissue, which allowed grinding using an automated high-throughput tissue homogeniser, Geno/Grinder 2010, with a 384-samples grinding capacity (4 × 96-sample plates) in two minutes.

The UV absorbance protocol for the FLUOstar Omega microplate reader (BMG LABTECH) was used to measure the concentration and purity of the DNA samples. By using this method, the 637 DNA samples were quantified in less than 10 minutes. The DNA purity (A260/A280 ratio) ranged from 1.7 to 2.0, with an average concentration of 985 ng/μl.

Following the optimized workflow, the total time from sampling and processing to DNA extraction and quantitation of the 637 leaf samples was reduced from seven to five days.

In order to optimise and use the KASP system in-house, KASP assays and allele-calling software (KlusterCaller) were purchased from LGC, UK. The amplification parameters on the compatible PCR (GeneAmp 9700) and real-time PCR machines (Roche LightCycler 480 II) were optimised. Microtiter 96- and 384-well plates compatible with the different machines were acquired from Roche, Germany. We also optimised the FLUOstar Omega microplate reader for fluorescence measurement of amplified products following the manufacturer's manual. Then, we ran a KASP trial kit provided freely by LGC Biosearch to test for functionality with the different amplification equipment.

Application of the optimised genotyping workflow

Following the KASP set-up, we genotyped plant samples for QC and MAS in-house, with low-density markers. The QC genotyping ensured on-time identification of errors and mislabeling in inbred lines and false hybrids in F_1 maize breeding populations. Using the in-house KASP genotyping platform significantly reduced genotyping cost and time compared to outsourcing.

Genetic identity. Using a subset of 28 maize QC KASP SNPs, we were able to identify the genetic origin of a set of twenty well-adapted maize varieties originating from IITA, which were regenerated at four other locations. Genetic identification was performed using the original maize varieties' molecular marker profile and the genetic distance approach. Seed sources having <5% genetic distance were considered the same. The genetic distance among the four original maize lines, and between lines from IITA and each of the four seed sources, was calculated using PowerMaker 3.25 statistical software. The genetic distance among the four designation lines from IITA ranged from 0.0563 to 0.1239, indicating that the lines were different. The genetic distance among the different seed sources of the same line designation was: 0.0105-0.0314 (SAMMAZ15), 0.0105–0.0418 (SAMMAZ16), 0.0105–0.0837 (SAMMAZ27), and 0.000–0.0563 (SAMMAZ39). The SNPviewer, a tool that enables viewing genotyping data as a cluster plot, was used to view and generate an image of the genotyping result. The SNPviewer image showed that designated lines from three out of the four seed sources grouped with lines from IITA (Figure 2). The dendrogram image (Figure 3) also showed a grouping of different seed sources of the same line designation except for SAMMAZ39-1, SAMMAZ16-3, and SAMMAZ27-4. This clustering pattern indicates that all seeds from the same line had a common origin. SAMMAZ27-4 appeared to be genetically distant from SAMMAZ27-IITA by 0.0837. However, it grouped with SAMMAZ15 (Figure 3: blue circle), suggesting a possible mislabeling or mix-up of seeds during harvesting and storage. SAMMAZ16-2 and SAMMAZ39-1 grouped on a different tree limb (Figure 3: red circle), indicating possible pollen contamination or seed mix-up during handling.

Hybrid verification. In another QC experiment using our workflow, we screened two groups of F_1 plants for hybrid verification, including their parental inbred lines, with 10 KASP SNP markers. The parental inbred lines were screened with an initial 50 KASP SNP taken from a defined panel of maize QC KASP markers to identify polymorphic markers. Only 10 KASP markers polymorphic between the parental lines were used to screen the F_1 plants to verify their parentage. The KASP genotyping assay was useful in distinguishing between the parental genotypes and identifying the true hybrid lines. Cluster analysis of Group1 F_1s (Figure 4) grouped the genotypes into three clusters. The heterozygous F_1 progenies were in the middle of the plot, and the homozygous parental inbred lines diverged from each other (along the X- and Y-axis of the plot) for all markers. The genotyping result (Table 5) and the clustering pattern indicate that the F_1 progenies were true hybrids. Similar clustering was observed among F_1s in Group 2 except in Set 3b, where 38 F_1s grouped with parental genotypes. The homozygous F_1s could be due to contamination from foreign pollens during the crossing in the field or seed mix-up during storage or planting.

Nonetheless, the KASP genotyping assay suffers some genotyping errors, especially during the automatic calling of genotypes. For instance, one F1 line (SCH-4) developed from the bi-parental cross, KS23-6 and IITATZI1653, appeared to cluster with the parent 2 (IITATZI1653) when genotyped with marker PZB01658_1 (Figure 5). The datapoint representing IITATZI1653
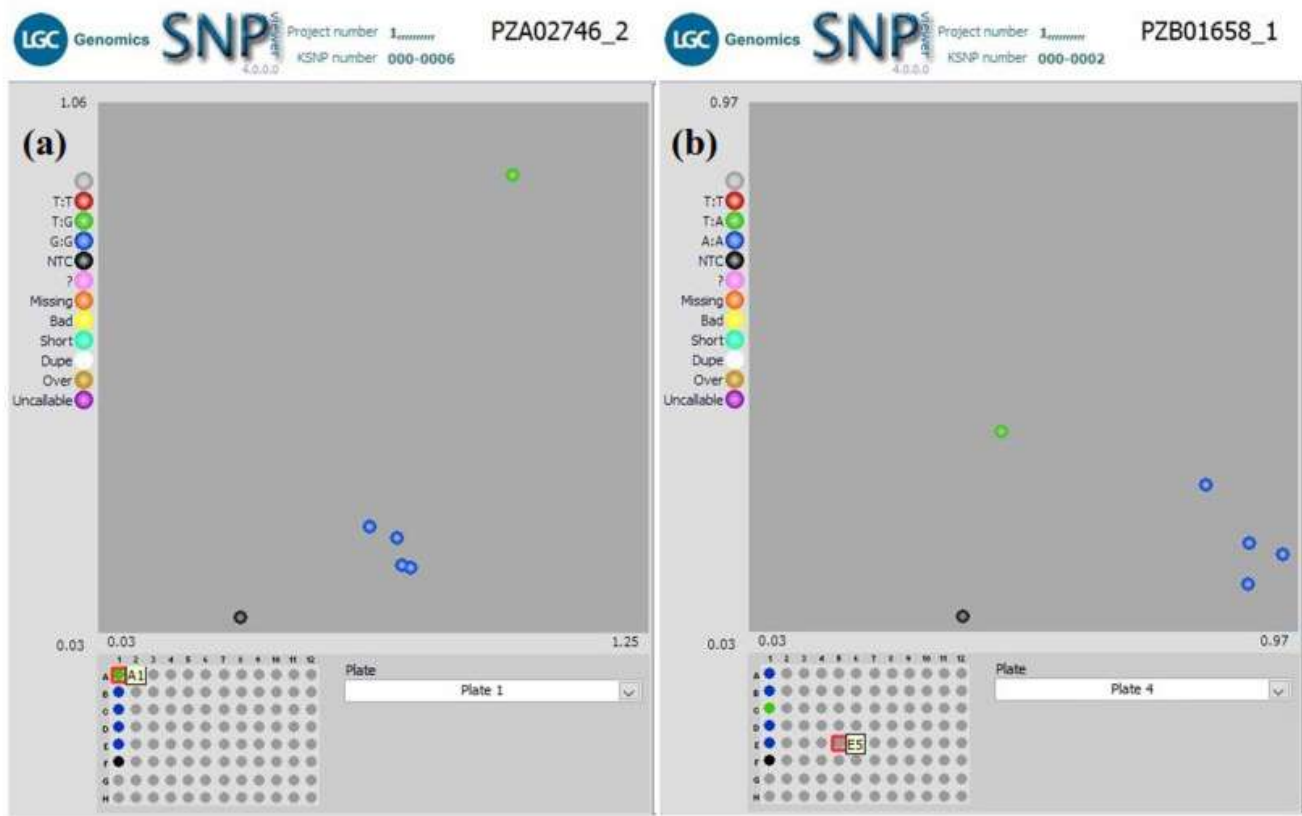
Figure 2. SNPviewer screenshot showing clustering of IITA's maize lines with same lines from three out of four seed sources.
(**a**) Sammaz15-2, -3, and -4 grouped with IITA's Sammaz15 (blue dots) using SNP PZA02746_2; (**b**) Sammaz39-2, -3, and -4 grouped with IITA's Sammaz39 (blue dots) using SNP PZB01658_1. For each SNP marker, blue dots represent homozygous genotypes, green dots represent heterozygote genotypes, and the black dots represent no-template controls (NTC) as indicated on the left side of each image.

(Figure 5, information in the yellow square) was plotted higher up, away from the X-axis, which brought it closer to the datapoint representing SCH-4 plotted slightly away from the other F1s in the middle. Because genotype calls are generated based on the relative position of datapoints on the plot, SCH-4 was automatically called as the nearby parental genotype, A:A, which was an error seeing that line SCH-4 was heterozygous (true hybrid) for the rest of the markers. The upward positioning of line IITATZI1653 away from the X-axis could be possibly due to trace contamination of line IITATZI1653 sample DNA with line KS23-6 sample DNA during sample preparation. A monomorphic marker is seen in the genotyping of F1 lines developed from the bi-parental crosses KS23-3 x IITATZI1653 using marker PHM5502_31.

Marker-assisted backcrossing. We performed multiple field selections annually by applying our workflow in MAS projects, which accelerated the maize breeding process. For instance, in the MABC project, a set of trait-specific KASP SNPs was used to select 24 BC_1S_2 maize lines potentially introgressed with resistance to aflatoxin accumulation after four selection cycles in less than two years. Potentially introgressed lines

are undergoing field evaluation under artificial infestation for resistance to aflatoxin accumulation. The result of the MAS of high PVA lines, on the other hand, identified nine out of 70 inbred maize lines harbouring favourable alleles of the *crtRB1* gene, which is associated with high PVA content in maize.

Discussion

There are different methods of plant tissue sampling, including collecting samples in silica gel[35], NaCl/CTAB[36], alcohol[37], blotter paper, gel pack, dry ice, and liquid nitrogen[38]. These methods provide reasonably good quality and quantity of DNA for molecular marker genotyping. However, deciding which method to use is based on the number of samples and distance from the field to the laboratory[38]. We routinely use wet ice in Styrofoam boxes and cooler bags. It is cost-effective and suitable for close-proximity sample collection, and leaf samples are preserved by freeze-drying[39] before DNA extraction. We collected fresh leaf tissues directly into 96-well extraction tubes rather than the traditional jute or tea bags, which means our procedure provides high throughput sampling. This sampling process also ensured that sample DNA was not degraded by prolonged exposure of leaf tissues to moisture as it occurs
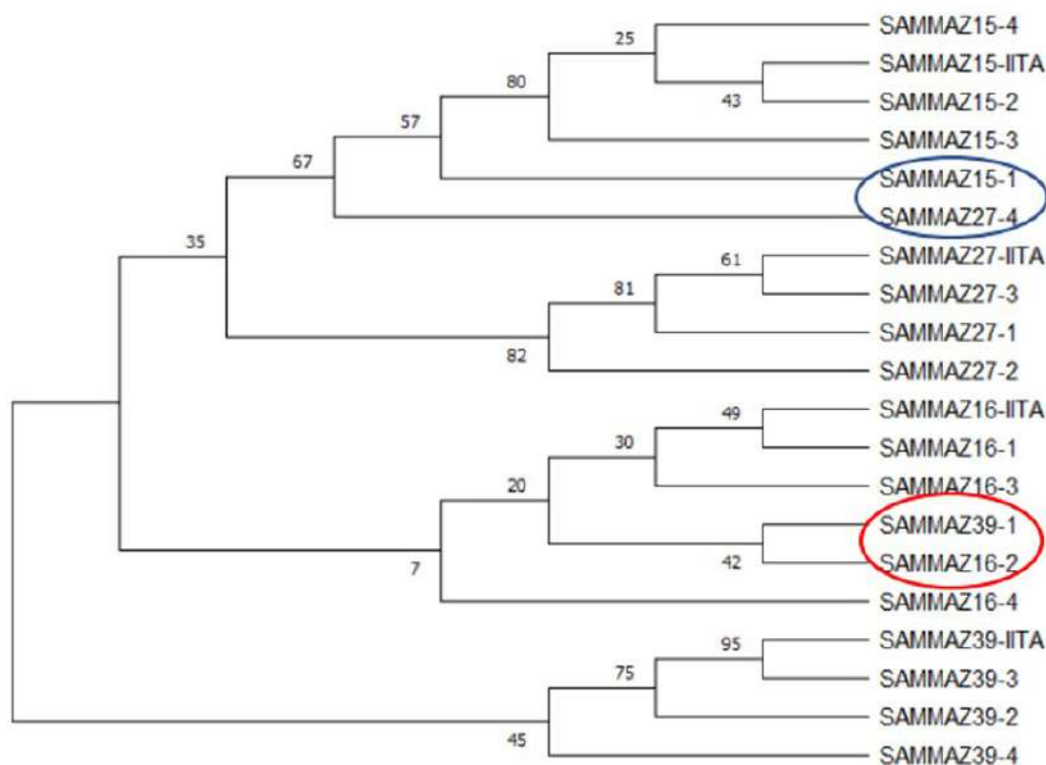
Figure 3. Neighbor-joining tree for four maize varieties taken from five seed sources based on genetic distance, performed with 1,000 bootstrap. Bootstrap values are indicated on the tree branches. The suffixes "-1", "-2", "-3", "-4", and "-IITA", after line name indicate seed source 1, 2, 3, 4, and IITA.

in post-freeze drying cutting of leaf tissues stored in jute and tea bags.

Our protocol aimed to extract high-quality DNA suitable for KASP genotyping from a smaller amount of leaf tissues. The reduced sample volume lowered the cost of reagents and the time for DNA extraction. The automated grinding in 96-well plates increased throughput and minimised the time required for manual grinding. Thus, this method would benefit MAS breeding programs that often screen thousands of plant samples each season[40]. A similar high-throughput result was achieved by Anderson *et al.* (2018)[41]. They optimised the DNA extraction method by Whitlock *et al.* (2008)[42], used a 96-well plate for extraction and achieved a consistent yield across the plate with a low failure rate.

Three steps of the original DArT DNA extraction method were slightly modified to achieve our aim. The first modification was made in the sample grinding step, where we used dried leaf tissues instead of fresh ones;—using dried samples enabled high-throughput grinding using a Geno/Grinder, reducing the time used in manual grinding with liquid nitrogen. The second modification was at the alcohol precipitation step: the sample tubes were incubated at -20°C for 30 minutes after adding the ice-cold isopropanol, instead of only mixing

by inversion. This incubation is necessary for slow and complete DNA precipitation. The third modification was reconstituting the DNA pellet: we dissolved the DNA in a solution of nuclease-free water and RNaseA instead of using a Tris-EDTA (TE) buffer to prevent the chelating effect of EDTA on Mg^{2+} during PCR[43,44]. The success of the KASP genotyping experiment is dependent on the quality and quantity of genomic DNA. Usually, a final minimum DNA concentration of 5 ng/µl is required for maize, to generate clear and consistent allele calls using the KASP assay[45]. Our slightly modified DNA extraction method provided good quality DNA, suitable for KASP genotyping. Jain *et al.* (2013) extracted suitable quality DNA from honey that was amplifiable by PCR, using an optimised DArT DNA extraction protocol.

Some commonly used DNA quality and quantity analysis methods include agarose-gel electrophoresis, fluorescence, and Ultraviolet (UV) absorbance-based measurement[38]. Fluorescence-based measurement using DNA-binding dyes such as PicoGreen is fast, sensitive, and dsDNA-specific; however, it comes with the DNA-binding reagent's added cost[46,47]. Agarose gel electrophoresis is laborious and carries the risk of exposure to hazardous chemicals like ethidium bromide[47]. The UV absorbance measurement is the most common DNA quantitation method. It is based on DNA absorbing UV light
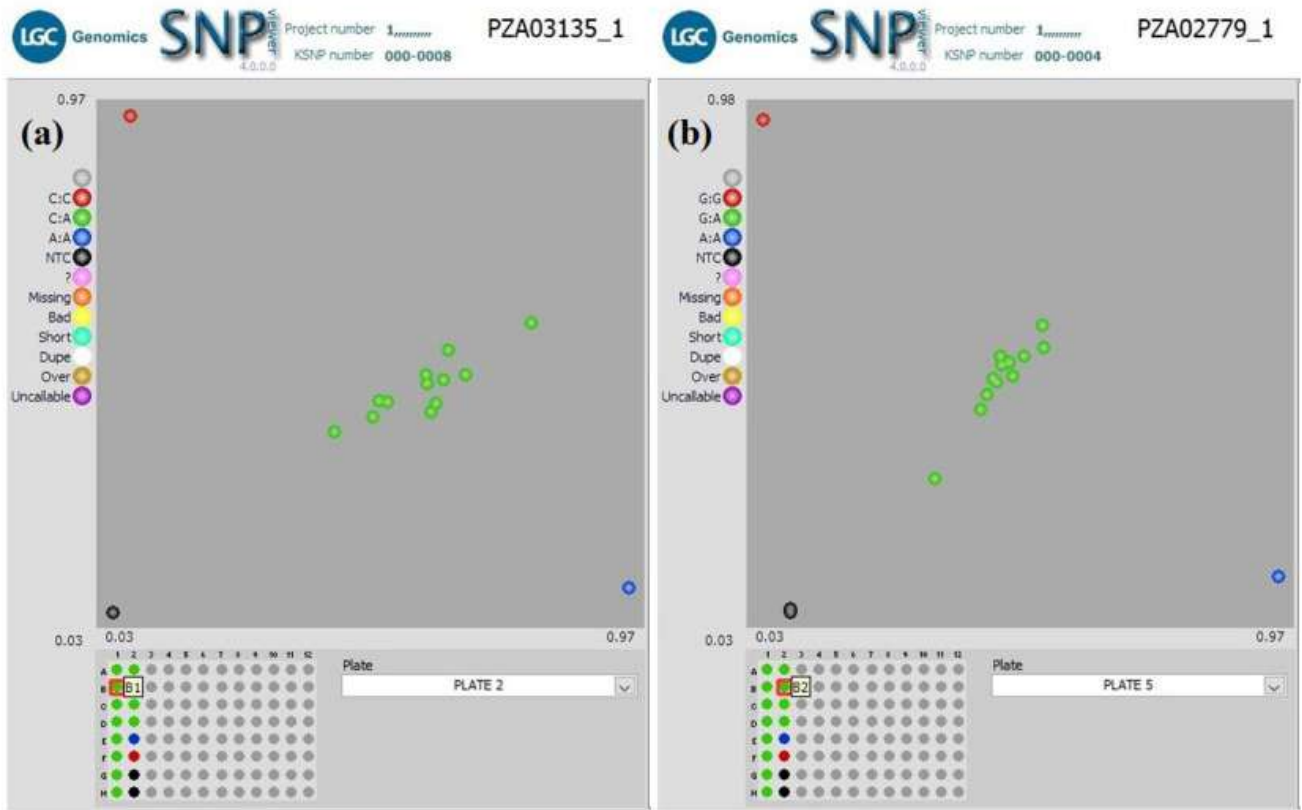
Figure 4. SNPviewer screenshot showing the result of hybrids verification in two sets of F$_1$ Plants. (**a**) Genotyping 12 F1 lines produced from a cross between KS23-5 and IITATZI1653, using SNP PZA03135_1. (**b**) Genotyping of 12 F1 lines produced from a cross between KS23-5 and IITATZI1667, using SNP PZA02779_1. For each SNP marker, blue dots represent homozygous parental genotype reported by FAM, red dots represent homozygous parental genotype reported by HEX, green dots represent heterozygous hybrid genotypes, and the black dots represent no-template controls (NTC). **Legend**: FAM = Carboxyfluorescein; HEX = Hexachloro-fluorescein.

Table 5. KASP genotyping result for the hybrid verification experiment.

Subject ID: SNP ID	KS23-6	IITATZI1653	SCH-1	SCH-2	SCH-3	SCH-4	SCH-5	SCH-6	SCH-7	SCH-8	SCH-9	SCH-10	SCH-11	SCH-12
ae1_7	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PHM5181_10	C:C	T:T	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C
PZA01216_1	G:G	A:A	G:A	G:A	G:A	A:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA01885_2	G:G	A:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA01919_2	C:C	G:G	G:C	G:C	G:C	G:G	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C
PZA02090_1	T:T	A:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A
PZA02779_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA03135_1	A:A	C:C	C:A	C:A	C:A	C:C	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A
PZA03363_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A

Subject ID:	KS23-6	IITATZI1653	SCH-1	SCH-2	SCH-3	SCH-4	SCH-5	SCH-6	SCH-7	SCH-8	SCH-9	SCH-10	SCH-11	SCH-12
SNP ID														
PZB01658_1	T:T	A:A	T:A	T:A	T:A	A:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A
True Hybrid?	Parent 1	Parent 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject ID:	KS23-3	IITATZI1667	SCH-1	SCH-2	SCH-3	SCH-4	SCH-5	SCH-6	SCH-7	SCH-8	SCH-9	SCH-10	SCH-11	SCH-12
SNP ID														
ae1_7	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA00413_20	C:C	A:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A
PZA01885_2	G:G	A:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA01919_2	C:C	G:G	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C
PZA02269_3	C:C	T:T	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C
PZA02779_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA03135_1	A:A	C:C	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A
PZA03363_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZB01658_1	T:T	A:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A
sh1_12	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
True Hybrid?	Parent 1	Parent 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject ID:	KS23-5	IITATZI1667	SCH-1	SCH-2	SCH-3	SCH-4	SCH-5	SCH-6	SCH-7	SCH-8	SCH-9	SCH-10	SCH-11	SCH-12
SNP ID														
PZA00413_20	C:A	A:A	A:A	A:A	C:A	C:A	C:A	A:A	C:A	C:A	A:A	C:A	C:A	C:A
PZA01885_2	G:G	A:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA01919_2	C:C	G:G	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C
PZA02164_16	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA02269_3	C:C	T:T	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C
PZA02779_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA03135_1	A:A	C:C	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A
PZA03363_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZB01658_1	T:T	A:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A
sh1_12	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
True Hybrid?	Parent 1	Parent 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject ID:	KS23-3	IITATZI1653	SCH-1	SCH-2	SCH-3	SCH-4	SCH-5	SCH-6	SCH-7	SCH-8	SCH-9	SCH-10	SCH-11	SCH-12
SNP ID														
ae1_7	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PHM5502_31	G:G	G:G	G:G	G:G	G:G	G:G	G:G	G:G	G:G	G:G	G:G	G:G	G:G	G:G
PZA01885_2	G:G	A:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA01919_2	C:C	G:G	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C
PZA02269_3	C:C	T:T	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C

Subject ID:	KS23-6	IITATZI1653	SCH-1	SCH-2	SCH-3	SCH-4	SCH-5	SCH-6	SCH-7	SCH-8	SCH-9	SCH-10	SCH-11	SCH-12
SNP ID														
PZA02779_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA03135_1	A:A	C:C	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A
PZA03363_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZB01658_1	T:T	A:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A
sh1_12	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
True Hybrid?	Parent 1	Parent 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Subject ID:	KS23-5	IITATZI1653	SCH-1	SCH-2	SCH-3	SCH-4	SCH-5	SCH-6	SCH-7	SCH-8	SCH-9	SCH-10	SCH-11	SCH-12
SNP ID														
PZA00413_20	C:A	C:C	C:A	C:C	C:A	C:C	C:A	C:A	C:C	C:C	C:C	C:C	C:C	C:A
PZA01885_2	G:G	A:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA01919_2	C:C	G:G	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C	G:C
PZA02164_16	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA02269_3	C:C	T:T	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C	T:C
PZA02779_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZA03135_1	A:A	C:C	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A	C:A
PZA03363_1	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
PZB01658_1	T:T	A:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A	T:A
sh1_12	A:A	G:G	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A	G:A
True Hybrid?	Parent 1	Parent 2	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Legend: SCH = Single cross hybrid, and the suffixes '-1 to -12' represent the number of F1s genotyped for each cross.

at a specific wavelength; DNA concentration is calculated by measuring the absorbance at 260nm and using the relationship A260 of 1.0 equals 50 µg/ml pure dsDNA[46]. DNA purity is estimated based on two UV absorbance ratios: A260/A280 ≥1.7 and A230/A260 ≥ 1.5 for pure DNA[46]. Our workflow optimized the nucleic acid quantitation method to a high throughput using a microplate reader and 96- and 384-well plates. The FLUOstar microplate reader uses ultrafast UV/Vis spectrometers for absorbance measurements, measuring 96 samples (96-well plate) to 384 samples (384-well plate) simultaneously within one second per well. It combines speed and the acquisition of complete absorbance spectra (220 to 1000 nm), making it ideal for nucleic acid quantification[48].

Although outsourcing KASP offers a lower cost per data point, this lower genotyping cost is usually driven by a high volume of samples, impracticable for most MAS projects genotyping smaller sample volumes with select markers[49]. Our in-house genotyping system provides reduced cost, mainly from logistics, and faster data turn-around times, ultimately accelerating the genotyping workflow.

A few studies serve as the benchmark for QC analysis in maize using the KASP genotyping system. Semagn *et al.* (2012) suggested using a subset of 50 to 100 KASP markers for routine QC; Chen *et al.* (2016) used a smaller subset of markers (10 markers) to assess mislabeling of entries across a panel of CIMMYT Maize Lines (CMLs) achieving up to 99% detection probability. The latter also proposed using a rapid QC approach, with a smaller subset of markers, to ensure effective QC, lower genotyping costs, and shorten data turn-around time during seed production. Using a subset of markers, we were able to identify seed mix-up and labelling errors. For instance, the grouping of SAMMAZ27-4 with SAMMAZ15 (Figure 3: blue circle) suggests a possible mislabeling or mix-up of seeds during harvesting and storage. Also, the grouping of SAMMAZ16-2 and SAMMAZ39-1 (Figure 3: red circle) indicates possible pollen contamination or seed mix-up during handling. Similar errors due to seed mix-up and contamination were reported in Semagn *et al.* (2012), where 50 KASP SNPs were used to determine genetic identity among two to four seed sources of the same inbred line. Ertiro *et al.* (2015) also reported a high discrepancy in genetic purity and identity by the
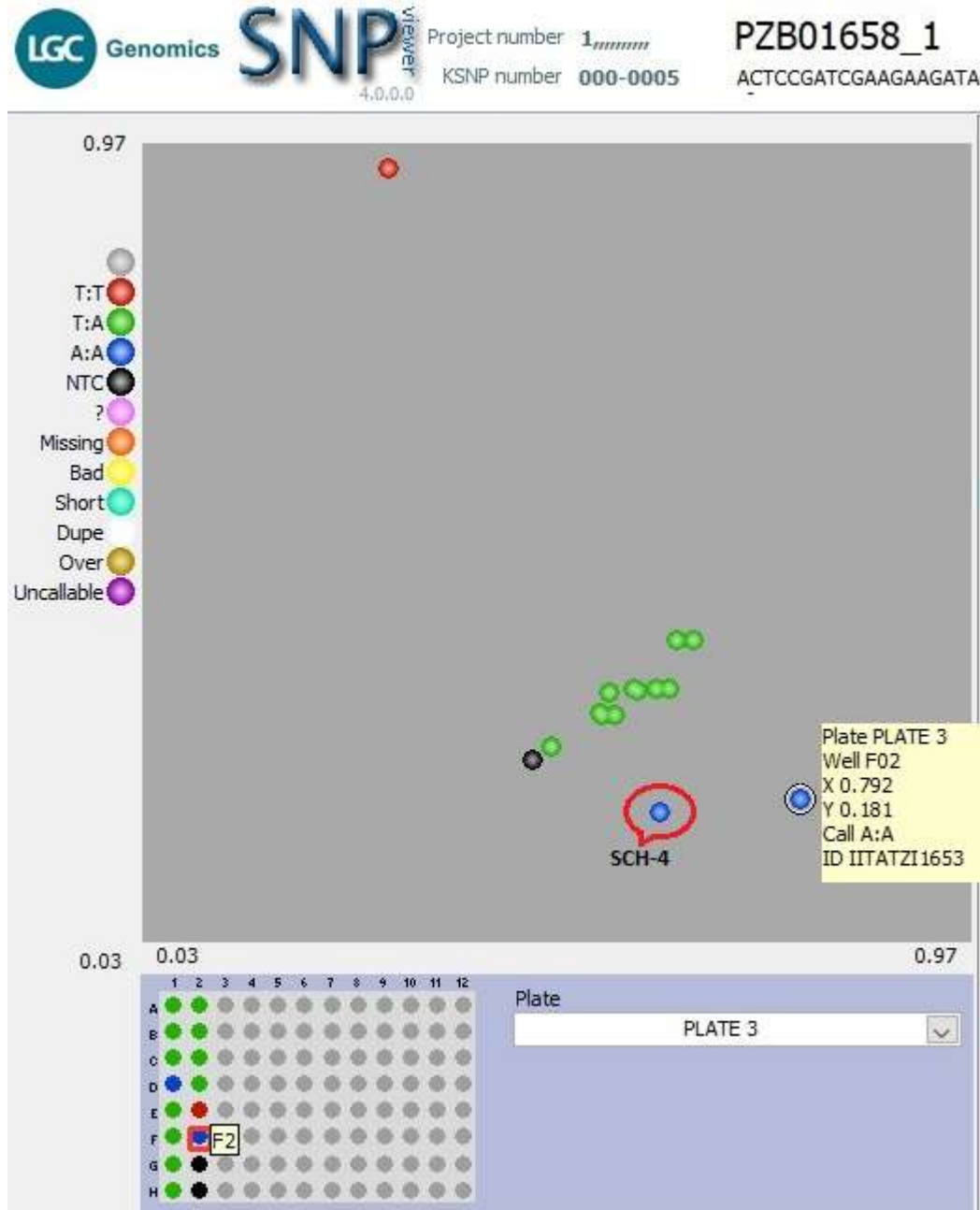
Figure 5. SNPviewer screenshot showing the result of hybrid verification of F$_1$ Plants. Genotyping 12 F$_1$ lines produced from a cross between KS23-6 and IITATZI1653, using SNP PZB01658_1.The blue dots represent homozygous parent 2 (IITATZI1653) genotype reported by FAM, red dots represent homozygous parent 1 (KS23-6) genotype reported by HEX, green dots represent heterozygous (F$_1$s) genotypes, and the black dots represent no-template controls (NTC). **Legend:** SCH-4 = Single cross hybrid (F$_1$) sample 4; FAM = Carboxyfluorescein; HEX = Hexachloro-fluorescein.

origin of seed sources irrespective of the genotyping platform used. They concluded that using a small subset of pre-selected high-quality markers was sufficient for performing QC analysis using low-marker density genotyping platforms like KASP. This study showed that the rapid QC method using 28 KASP SNPs efficiently distinguished the four maize varieties taken from five seed sources.

Hybrid verification is often performed during seed production or population breeding to confirm that a particular hybrid is derived from the intended parental lines (free from contamination by foreign pollens). Reducing the data turn-around time is essential to ensure that an accurate hybrid is selected to be carried forward in breeding programs or dissemination to farmers in seed production[33]. A reduced

turn-around time also saves the cost of inputs applied to undesired genotypes since they can be discarded as soon as they have been identified upon genotyping. Our expedited workflow was able to achieve this. The possibility of contamination by self-pollination or foreign pollen exists; as such, hybrid verification is necessary to enable a seed producer to check whether accurate crosses are made for the production of the hybrid; this increases the confidence of the end-users on the quality and integrity of seeds produced[33]. Our results showed that 10 KASP markers were sufficient in distinguishing between maize parental inbred lines and identified true hybrid lines, residual contaminations, and possible sampling errors. A small subset of KASP markers has also been used to verify hybrids in other plant species. Patterson *et al.* (2017)[50] achieved a highly accurate picture of *Myriophyllum* species distribution dynamics in North American lakes by genotyping 39 individuals from both parental watermilfoil and their hybrids, using a subset of three KASP markers. Osei *et al.* (2020)[51] used 38 KASP markers to screen tomato genotypes to identify true F_1 hybrids for the possible development of inbreds with long shelf life through marker-assisted backcrossing (MABC).

Following our optimised workflow, we were able to identify high-PVA maize lines harbouring the favourable allele of the *crtRB1* gene, which could serve as donor lines for the maize PVA breeding program. The KASP-based selection of aflatoxin-resistant maize lines promises to fast-track the development of tropical lines resistant to aflatoxin, which will contribute to genetic gain in maize production. Similar success was achieved by the Biotechnology Center of the University of California, Davis, USA, where KASP SNPs associated with *Phytophthoria capsici* resistance were used to identify and selectively breed pepper strains[52]. So far, we have generated over 2,000 data points using our in-house genotyping workflow. Applying our optimised workflow to the QC and MAS experiments outlined above reduced the volume of reagents and consumables used, shortened the data turn-around, and ultimately accelerated the crop improvement process.

Conclusions

This study describes for the first time an improvement of an entire conventional DNA-based genotyping workflow, including the benchmark KASP genotyping platform in-house in our facility to fast-track molecular marker-based selection for crop improvement. We acknowledge the initial capital investment to procure some of these instruments. However, it is not always necessary to equip each lab or breeding program. The use of shared facilities locally and regionally, and the re-purposing of existing equipment such as the PCR machine and the spectrophotometer, help overcome the high cost of essential instruments. The improved genotyping workflow promises to accelerate the marker-assisted selection process and push crop improvement activities to attain the yield potential over a shorter time period. The result of this work can be readily adopted by national institutions, public and small plant breeding laboratories in developing countries to accelerate molecular marker-based genotyping for crop improvement activities, including QC and MAS. The results will also be helpful to accelerate the QC activities of seed producers and facilitate cultivar identification and adoption-tracking studies.

Data availability

Underlying data

Figshare: SNP data for "Developing and deploying an efficient genotyping workflow for accelerating maize improvement in developing countries.", https://doi.org/10.6084/m9.figshare.17157914[53].

This project contains the following underlying data:

- Supplementary Table 1. List of trait-specific KASP SNPs used in the MAS experiment with sequence information

- Supplementary Table 2. List of KASP SNPs used in the QC experiments with sequence information

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

References

1. Delannay X, McLaren G, Ribaut JM: **Fostering molecular breeding in developing countries.** *Mol Breeding.* 2012; **29**(4): 857–873.
   Publisher Full Text

2. Ray DK, Mueller ND, West PC, *et al.*: **Yield Trends Are Insufficient to Double Global Crop Production by 2050.** *PLoS One.* 2013; **8**(6): e66428.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Xu Y, Li P, Zou C, *et al.*: **Enhancing genetic gain in the era of molecular breeding.** *J Exp Bot.* 2017; **68**(11): 2641–2666.
   PubMed Abstract | **Publisher Full Text**

4. Barabaschi D, Tondelli A, Desiderio F, *et al.*: **Next generation breeding.** *Plant Sci.* 2016; **242**: 3–13.
   PubMed Abstract | **Publisher Full Text**

5. Varshney RK, Singh VK, Hickey JM, *et al.*: **Analytical and decision support tools for genomics-assisted breeding.** *Trends Plant Sci.* 2016; **21**(4): 354–363.
   PubMed Abstract | **Publisher Full Text**

6. Gedil M, Tripathi L, Ghislain M, *et al.*: **Biotechnology Success Stories by the Consultative Group on International Agriculture Research (CGIAR) System.** *Science Policy Reports.* 2014; 95–114.
   Publisher Full Text

7. Ribaut JM, de Vicente MC, Delannay X: **Molecular breeding in developing countries: challenges and perspectives.** *Curr Opin Plant Biol.* 2010; **13**(2): 213–218.
   PubMed Abstract | **Publisher Full Text**

8. Gedil M, Menkir A: **An Integrated Molecular and Conventional Breeding**

Scheme for Enhancing Genetic Gain in Maize in Africa. *Front Plant Sci.* 2019; **10**: 1430.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. **Integrated Breeding Platform.** Accessed on June 30 2020.
Reference Source

10. Rasheed A, Hao Y, Xia X, *et al.*: **Crop breeding chips and genotyping platforms: progress, challenges, and perspectives.** *Mol Plant.* 2017; **10**(8): 1047–1064.
PubMed Abstract | **Publisher Full Text**

11. Semagn K, Olsen M, Babu R, *et al.*: **Single nucleotide polymorphism genotyping using Kompetitive Allele SpeciFic PCR (KASP): Overview of the Technology and its Application in Crop Improvement.** *Mol Breed.* 2014; **33**: 1–14.
Publisher Full Text

12. Semagn K, Beyene Y, Makumbi D, *et al.*: **Quality control genotyping for assessment of genetic identity and purity in diverse tropical maize inbred lines.** *Theor Appl Genet.* 2012; **125**(7): 1487–1501.
PubMed Abstract | **Publisher Full Text**

13. Beissinger TM, Hirsch CN, Sekhon RS, *et al.*: **Marker Density and Read Depth for Genotyping Populations Using Genotyping-by-Sequencing.** *Genetics.* 2013; **193**(4): 1073–1081.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Jain SA, de Jesus FT, de Marchioro GM, *et al.*: **Extraction of DNA from honey and its amplification by PCR for botanical identification.** *Food Sci Tech (Campinas).* 2013; **33**(4): 753–756.
Publisher Full Text

15. Ertiro BT, Ogugo V, Worku 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**(1): 908.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Giancola S, McKhann HI, Berard A, *et al.*: **Utilization of the three high-throughput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants.** *Theor Appl Genet.* 2006; **112**(6): 1115–24.
PubMed Abstract | **Publisher Full Text**

17. Helguera M, Khan IA, Kolmer J, *et al.*: **PCR Assays for the Cluster of Rust Resistance Genes and Their Use to Develop Isogenic Hard Red Spring Wheat Lines.** *Crop Science.* 2003; **43**(5): 1839.
Publisher Full Text

18. Stevanato P, Broccanello C, Biscarin F, *et al.*: **High-throughput RAD-SNP genotyping for characterisation of sugar beet genotypes.** *Plant Mol Biol Rep.* 2014; **32**: 691 –6.
Publisher Full Text

19. Thomson MJ: **High-Throughput SNP Genotyping to Accelerate Crop Improvement.** *Plant Breeding and Biotechnology.* 2014; **2**(3): 195–212.
Publisher Full Text

20. Wu J, Liu S, Wang Q, *et al.*: **Rapid identification of an adult plant stripe rust resistance gene in hexaploid wheat by high-throughput SNP array genotyping of pooled extremes.** *Theor Appl Genet.* 2018; **131**(1): 43–58.
PubMed Abstract | **Publisher Full Text**

21. Myakishev MV, Khripin Y, Hu S, *et al.*: **High-Throughput SNP Genotyping by Allele-Specific PCR with Universal Energy-Transfer-Labeled Primers.** *Genome Res.* 2001; **11**(1): 163–169.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Broccanello C, Chiodi C, Funk A, *et al.*: **Comparison of three PCR-based assays for SNP genotyping in plants.** *Plant Methods.* 2018; **14**: 28.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. He C, Holme J, Anthony J: **SNP Genotyping: The KASP Assay.** *Crop Breeding.* 2014; **114**: 75–86.
Publisher Full Text

24. Ayalew H, Tsang PW, Chu C, *et al.*: **Comparison of TaqMan, KASP and rhAmp SNP genotyping platforms in hexaploid wheat.** *PLoS One.* 2019; **14**(5): e0217222.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

25. Chen J, Zavala C, Ortega N, *et al.*: **The Development of Quality Control Genotyping Approaches: A Case Study Using Elite Maize Lines.** *PLoS One.* 2016; **11**(6): e0157236.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26. Kaur B, Mavi GS, Gill MS, *et al.*: **Utilisation of KASP technology for wheat improvement.** *Cereal Res Commun.* 2020; **48**: 409–421.
Publisher Full Text

27. Steele KA, Quinton-Tulloch MJ, Amgai RB, *et al.*: **Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of *indica* rice.** *Mol Breed.* 2018; **38**(4): 38.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28. Shi Z, Liu S, Noe J, *et al.*: **SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance.** *BMC Genomics.* 2015; **16**(1): 314.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29. Neelam K, Brown-Guedira G, Huang L: **Development and validation of a breeder friendly KASPar marker for wheat leaf rust resistance locus *Lr21*.** *Mol Breeding.* 2013; **31**(1): 233–237.
Publisher Full Text

30. Khera P, Upadhyaya HD, Pandey MK, *et al.*: **Single Nucleotide Polymorphism–based Genetic Diversity in the Reference Set of Peanut (*Arachis* spp.) by Developing and Applying Cost-Effective Kompetitive Allele Specific Polymerase Chain Reaction Genotyping Assays.** *The Plant Genome.* 2013; **6**(3).
Publisher Full Text

31. Pariasca-Tanaka J, Lorieux M, He C, *et al.*: **Development of an SNP genotyping panel for detecting polymorphisms in *Oryza glaberrima/O. sativa* interspecific crosses.** *Euphytica.* 2014; **201**(1): 67–78.
Publisher Full Text

32. Diversity Array Technology Pty Ltd (DArT): **Plant DNA extraction protocol for DArT.** Accessed on June 24 2020.
Reference Source

33. Gowda M, Worku M, Nair SK, *et al.*: **Quality Assurance/ Quality Control (QA/ QC) in Maize Breeding and Seed Production: Theory and Practice.** CIMMYT: Nairobi. 2017.
Reference Source

34. Liu K, Muse SV: **PowerMarker: an integrated analysis environment for genetic marker analysis.** *Bioinformatics.* 2005; **21**(9): 2128–2129.
PubMed Abstract | **Publisher Full Text**

35. Chase MW, Hills HH: **Silica gel: an ideal material for Field preservation of leaf samples for DNA studies.** *Taxon.* 1991; **40**: 215–220.
Reference Source

36. Storchova H, Hrdlickova R, Chrtek J Jr, *et al.*: **An improved method of DNA isolation from plants collected in the Field and conserved in saturated NaCl/CTAB solution.** *Taxon.* 2000; **49**(1): 79–84.
Publisher Full Text

37. Flournoy LE, Adams RP, Pandy RN: **Interim and archival preservation of plant specimens in alcohols for DNA studies.** *Biotechniques.* 1996; **20**(4): 657–660.
PubMed Abstract

38. Semagn K: **Leaf Tissue Sampling and DNA Extraction Protocols.** In: Besse P, Editor. *Molecular Plant Taxonomy: Methods and Protocols.* New York: Human Press; 2014; **1115**: 53–67.
Publisher Full Text

39. Till BJ, Jankowicz-Cieslak J, Huynh OA, *et al.*: **Sample Collection and Storage.** *Low-Cost Methods for Molecular Characterization of Mutant Plants.* 2015; 9–11.
Publisher Full Text

40. Gu WK, Weeden NF, Yu J, *et al.*: **Large-scale, cost-effective screening of PCR products in marker-assisted selection applications.** *Theor Appl Genet.* 1995; **91**(3): 465–470.
PubMed Abstract | **Publisher Full Text**

41. Anderson CB, Franzmayr BK, Hong SW, *et al.*: **Protocol: a versatile, inexpensive, high-throughput plant genomic DNA extraction method suitable for genotyping-by-sequencing.** *Plant Methods.* 2018; **14**(1): 75.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

42. Whitlock R, Hipperson H, Mannarelli M, *et al.*: **A high-throughput protocol for extracting high-purity genomic DNA from plants and animals.** *Mol Ecol Resour.* 2008; **8**(4): 736–41.
PubMed Abstract | **Publisher Full Text**

43. Jim FH, Novok T, Garson JA, *et al.*: **Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon.** *BMC Res Notes.* 2008; **1**(1): 70.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

44. Lahiri DK, Schnabel B: **DNA isolation by a rapid method from human blood samples: Effects of $MgCl_2$, EDTA, storage time, and temperature on DNA yield and quality.** *Biochem Genet.* 1993; **31**(7–8): 321–328.
PubMed Abstract | **Publisher Full Text**

45. **LGC Biosearch Technologies.** Accessed on June 24, 2020.
Reference Source

46. **BMG LABTECH Floustar Omega Application Notes.** Accessed on June 2 2020.

47. Kalpana R, Stephen MC, Milan M, *et al.*: **Quantifying DNA concentrations using fluorometry: A comparison of fluorophores.** *Mol Vis.* 2002; **8**: 416–21.
PubMed Abstract

48. **BMG LABTECH Floustar Omega Operating Manual.** Accessed on June 2 2020.
Reference Source

49. Kumpatla SP, Buyyarapu R, Abdurakhmonov IY, *et al.*: **Genomics-assisted plant breeding in the 21st Century: Technological Advances and Progress.** In I. Y. Abdurakhmonov (Ed.), *Plant Breeding.* IntechOpen. 2012.
Publisher Full Text

50. Patterson EL, Fleming MB, Kessler KC, *et al.*: **A KASP Genotyping Method to Identify Northern Watermilfoil, Eurasian Watermilfoil, and Their Interspecific Hybrids.** *Front Plant Sci.* 2017; **8**: 752.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

51. Osei MK, Danquah E, Danquah A, *et al.*: **Hybridity testing of tomato F_1 progenies derived from parents with varying fruit quality and shelf life using single nucleotide polymorphism (SNPs).** *Scientific African.* 2020; **8**: e00267.
Publisher Full Text

52. Majeed U, Darwish E, Rehman SU, *et al.*: **Kompetitive Allele-Specific PCR (KASP): A Singleplex Genotyping Platform and Its Application.** *Journal of Agricultural Science.* 2018; **11**(1): 11.
Publisher Full Text

53. Offornedo Q, Menkir A, Babalola D, *et al.*: **SNP data for "Developing and deploying an efficient genotyping workflow for accelerating maize improvement in developing countries".** *figshare.* Dataset. 2021.
http://www.doi.org/10.6084/m9.figshare.17157914.v2

Open Peer Review

Current Peer Review Status: ❓ ✅ ✅ ✅

- -

Version 3

Reviewer Report 13 July 2023

https://doi.org/10.21956/gatesopenres.15059.r33596

✅
Ruairidh J H Sawers

Plant Science, The Pennsylvania State University - University Park Campus, University Park, PA, USA

A nice survey of the options and applications for genotyping in breeding programs with useful details and experience of establishing an in-house KASP genotyping platform. The manuscript is strong on specific details that will be helpful to other researchers attempting similar work.

I've included a few general comments/thoughts that might be useful to consider.

Other reviewers have commented on the usefulness of a more complete economic breakdown. I'm not sure specific numbers (such as pricing given in the introduction) are necessarily that useful as they will no-doubt change. However, I would agree that more discussion of the relative costs of given approaches and different scales would be helpful. More generally, although services for the same technology are compared, less attention is given to different approaches. Indeed, given the context of extending access to molecular platforms, if would be informative to say more about the costs of using these technologies at all in comparison to conventional methods. For the specific KASP application, more could be said about the costs of primer design – especially if not using a crop well served with existing sequences – and synthesis.

The case studies are informative but would benefit from providing more information on the markers (for example, map position) and the sample genotypes. More could be said about the selection of markers, and specifically more discussion of how many markers are actually needed for a given application (based on these empirical examples, as much as prior literature). Results for selected markers are presented visually. Are these "typical" examples? Can more of a summary be given as to how many markers "worked", and how reproducible and reliable the results were? The results presented clearly separate genotypic classes (except for one highlighted individual). Was this always the case? Was calling of heterozygotes always robust? Was any additional confirmation performed? Would it be possible to estimate the rate of miscalling, either per marker or generally for the platform?

The cluster/calling of the KASP signal is most robust when each genotypic class is represented by

multiple individuals. In Fig.2 the sample size is small, and the homozygous T/T class is not represented. In isolation such sampling may complicate "calibration" of the heterozygous calls. Similarly, in Fig. 4 only a single sample is typed for each of the two parental homozygotes. The examples presented look nice and clear, but was this always the case for all markers typed on these samples?

It wasn't entirely clear what was done in the dendrogram in Fig. 3. How many markers were used? How was this number/set determined? Were they spread throughout the genome? Do these lines show a level of heterozygosity? As above, was there any ambiguity/error in calling?

It's a small detail, but at times the use of the term "line" was a little confusing. It can help to keep "line" to refer to inbred (highly homozygous) stocks. While an F1 is typically a cross between two lines, don't refer the F1 as an "F1 line" etc. The expectation with regard to heterozygosity – and the requirement to accurately make het calls – is directly relevant to selection and use of a genotyping platform.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Maize diversity, genetics and genomics; crop nutrition; mycorrhizal symbiosis

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 30 June 2023

✔ **Karthikeyan Adhimoolam** (iD)
   [1] Tamil Nadu Agricultural University, Madurai, Tamil Nadu, India
   [2] Horticulture, Jeju National University, Jeju-si, Jeju-do, South Korea

This study demonstrates the improved genotyping workflow for maize improvement in developing countries. I recommend the manuscript for indexing.

However, I recommend drawing a better workflow figure instead of Figure 1. And present some data (i.e., Table) to support the cost-effectiveness. Also, I suggest the authors improve the language.

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Plant molecular breeding and plant genomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 16 September 2022

https://doi.org/10.21956/gatesopenres.15059.r32358

✔ **Bhoja R. Basnet** [iD]

International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Dear authors, thank you for sincerely handling the reviews and putting your best effort into addressing those comments and concerns. I do not have further queries. I believe we let readers, the fellow researchers, and the broader scientific community judge the merit of this research.

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Plant breeding and genetics, genomics, quantitative genetics, and breeding program optimization.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

- -

Version 2

Reviewer Report 08 July 2022

https://doi.org/10.21956/gatesopenres.14950.r32187

? **Bhoja R. Basnet** iD

International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

It should go through one more round of revision. Please address the following concerns:

1. Introduction P3L3: "However, breeders often want to fingerprint a few dozen lines urgently for identity or parentage analysis for example". Please rephrase this sentence.

2. Please provide well-articulated one or two objectives of this research (towards the end of the Introduction section).

3. Introduction section P6L1: "Here we utilized the KASP assay" - revise as "In this study, we utilized...".
   Introduction P6L6: reference 27 does not provide any account of soybean - the paper is about maize. Please verify this information and correct it as needed.

4. Methods P1L3: "Well-adopted" should be changed to "well-adapted".

5. Method P1 Last sentence: The plants were grown, not raised.

6. Table 1: This is a piece of good information. However, I ask you to provide the exact number of genotypes and the samples within each genotype for all the groups (please add additional columns as needed).

7. Use of BC1S2 does not seem to be reliable in this study unless you verify the selection with phenotypic data to estimate the sensitivity and specificity of the marker assessment. However, it doesn't seem to harm the manuscript either.

8. One important analysis I would like to suggest to add to this study is HYBRID VERIFICATION. Please prepare a data table for each sample - identified within each genotype (F1), such as order the column as F1 cross name / no, sample #, Marker gen _P1, Marker gen_P2, Observed F1 gen, True Hyb (Yes or no), if not if the F1 gen is observed as maternal or paternal type, etc. Then please assess the true to hybrid types or % hybridity within each genotype (using samples within cross) and across all samples. Then also revise your results section with a detailed discussion on how this assay is helpful to discriminate true-to-type hybrids and also describe potential bias caused by the assay itself - genotyping error or so using data on samples within each genotype.

9. Did you sample multiple samples within each plant? If so, please revise the results section accordingly.

Is the rationale for developing the new method (or application) clearly explained?

Partly

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Plant breeding and genetics, genomics, quantitative genetics, and breeding program optimization.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 28 Jul 2022

Queen Offornedo

The authors thank the reviewers for taking the time to review the manuscript and raise critical issues. We trust that addressing these issues will immensely improve the paper. Below are the responses to the concerns raised.

Reviewer 2 comment: Introduction P3L3: "However, breeders often want to fingerprint a few dozen lines urgently for identity or parentage analysis, for example". Please rephrase this sentence.

Author's response: The sentence has been rephrased: "Even though large volume sizes can be consolidated and shipped for genotyping, there are times when breeders and partners may want to fingerprint a few dozen lines for identity or parentage analysis for quick decision making."

Reviewer 2 comment: Please provide well-articulated one or two objectives of this research (towards the end of the Introduction section).

Author's response: The objective of the research has been rephrased "This study aims to develop a genotyping workflow optimized for cost-effective and fast turn-around time that

---

can be deployed by less sophisticated and reasonably equipped laboratories in developing countries, to accelerate maize improvement research."

Reviewer 2 comment: Introduction section P6L1: "Here we utilized the KASP assay" - revise as "In this study, we utilized...".

Introduction P6L6: reference 27 does not provide any account of soybean - the paper is about maize. Please verify this information and correct it as needed.

Author's response: The phrase has been corrected.
The misplaced reference has been replaced. A more suitable reference has been attached to the statement. " Shi, Z., Liu, S., Noe, J. *et al.* SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance. *BMC Genomics* 16, 314 (2015). https://doi.org/10.1186/s12864-015-1531-3**"**

Reviewer 2 comment: Methods P1L3: "Well-adopted" should be changed to "well-adapted".

Authors' response: The phrase has been modified accordingly.

Reviewer 2 comment: Method P1 Last sentence: The plants were grown, not raised.

Author's response: The sentence has been modified as requested.

Reviewer 2 comment: Table 1: This is a piece of good information. However, I ask you to provide the exact number of genotypes and the samples within each genotype for all the groups (please add additional columns as needed).

Author's response: Table 1 has been modified to accommodate the required information.

Reviewer 2 comment: Use of BC1S2 does not seem to be reliable in this study unless you verify the selection with phenotypic data to estimate the sensitivity and specificity of the marker assessment. However, it doesn't seem to harm the manuscript either.

Author's response: The project is still ongoing. Phenotyping at different locations is currently underway. Definitely, we will utilize the phenotype data to verify the markers accuracy.

Reviewer 2 comment: One important analysis I would like to suggest to add to this study is HYBRID VERIFICATION. Please prepare a data table for each sample - identified within each genotype (F1), such as order the column as F1 cross name / no, sample #, Marker gen _P1, Marker gen_P2, Observed F1 gen, True Hyb (Yes or no), if not if the F1 gen is observed as

maternal or paternal type, etc. Then please assess the true to hybrid types or % hybridity within each genotype (using samples within cross) and across all samples. Then also revise your results section with a detailed discussion on how this assay is helpful to discriminate true-to-type hybrids and also describe potential bias caused by the assay itself - genotyping error or so using data on samples within each genotype.

Author's response: The genotyping analysis for the hybrid verification experiment is presented in Figure 5 and Table 5, under the Result section. The result section has also been furnished with a detailed discussion on using the KASP assay for hybrid verification and the potential drawback of the technology, as shown below:

Hybrid verification. In another QC experiment using our workflow, we screened two groups of F_1 plants for hybrid verification, including their parental inbred lines, with 10 KASP SNP markers. The parental inbred lines were screened with an initial 50 KASP SNP taken from a defined panel of maize QC KASP markers to identify polymorphic markers. Only 10 KASP markers, polymorphic between the parental lines, were used to screen the F_1 plants to verify their parentage. The KASP genotyping assay was useful in distinguishing between the parental genotypes and identifying the true hybrid lines. Cluster analysis of Group1 F_1s (Figure 4) grouped the genotypes into three clusters. The heterozygous F_1 progenies were in the middle of the plot, and the homozygous parental inbred lines diverged from each other (along the X- and Y-axis of the plot) for all markers. The genotyping result (Table 5) and the clustering pattern indicate that the F_1 progenies were true hybrids. Similar clustering was observed among F_1s in Group 2 except in Set 3b, where 38 F_1s were grouped with parental genotypes. The homozygous F_1s could be due to contamination from foreign pollens during the crossing in the field or seed mix-up during storage or planting.
Nonetheless, the KASP genotyping assay suffers some genotyping errors, especially during the automatic calling of genotypes. For instance, one F1 line (SCH-4) developed from the bi-parental cross, KS23-6 and IITATZI1653, appeared to cluster with the parent 2 (IITATZI1653) when genotyped with marker PZB01658_1 (Figure 5). The datapoint representing IITATZI1653 (Figure 5, information in the yellow square) was plotted higher up, away from the X-axis, which brought it closer to the datapoint representing SCH-4 plotted slightly away from the other F1s in the middle. Because genotype calls are generated based on the relative position of datapoints on the plot, SCH-4 was automatically called as the nearby parental genotype, A:A, which was an error seeing that line SCH-4 was heterozygous (true hybrid) for the rest of the markers. The upward positioning of line IITATZI1653 away from the X-axis could be possibly due to trace contamination of line IITATZI1653 sample DNA with line KS23-6 sample DNA during sample preparation. A monomorphic marker is seen in the genotyping of F1 lines developed from the bi-parental crosses KS23-3 x IITATZI1653 using marker PHM5502_31.

Reviewer 2 comment: Did you sample multiple samples within each plant? If so, please revise the results section accordingly.

Author's response: I am hoping that I got your question correct here. If you are referring

to whether or not we sampled by bulking, the answer is no, except for the MAS experiment for selecting PVA enriched lines, where we bulked ten leaf tissues from 10 plant stands per row.

Competing Interests: No competing interests.

Reviewer Response 25 Aug 2022

Bhoja R. Basnet

1. Before making the final decision, I am still unsure how you controlled or separated genetic purity and genotyping error in the assay. The markers seem to predict the hybrids almost perfectly, with few exceptions. How was that possible? I am not trying to deny the fact, but being curious as it was not the case in wheat.

2. My last question was about 'analyzing multiple samples from the same plant - without bulking.' Normal practice in QC for genetic purity and true-to-type hybrid verification is that multiple F1s samples are used (you have done it), and multiple samples within each plant are also used to control the genotyping or other handling errors that may arise during the genotyping workflow. It also gives confidence about the reproducibility of the same results for the same genetic materials.

Competing Interests: No competing interests were disclosed.

Author Response 06 Sep 2022

Queen Offornedo

Reviewer 2 comment: Before making the final decision, I am still unsure how you controlled or separated genetic purity and genotyping error in the assay. The markers seem to predict the hybrids almost perfectly, with few exceptions. How was that possible? I am not trying to deny the fact, but being curious as it was not the case in wheat.

Author's response: We are unsure what the case is when genotyping wheat using the KASP assay. However, the prediction of the maize hybrids could have been aided by the fact that maize is diploid, unlike wheat which is polyploid. The polyploid nature of wheat could make it challenging to distinguish between heterozygous and homozygous hybrid lines using KASP assays. Also, our genotyping is complemented by careful crossing and sample collection in the field and meticulous handling of the samples. The KASP assay genotyping prep is handled by select individuals and carried out in a dimmed light PCR workstation to minimize contamination and avoid activating the light-sensitive fluorophores in the KASP Mastermix. Furthermore, we first selected polymorphic markers that distinguished the parents, and only those were used to identify the hybrid lines. As such, any given F1 line is either a hybrid or not. Given the robustness and accuracy of the KASP assay, the use of multiple markers, and careful handling of the sample analysis, the chance of error is minimal.

Reviewer 2 comment: My last question was about 'analyzing multiple samples from the same plant - without bulking. 'Normal practice in QC for genetic purity and true-to-type hybrid verification is that multiple F1s samples are used (you have done it), and multiple samples within each plant are also used to control the genotyping or other handling errors that may arise during the genotyping workflow. It also gives confidence about the reproducibility of the same results for the same genetic materials."

Author's Response: Thank you for clarifying your question.

For this experiment, 12 different F1 plants per cross were sampled, although the number of samples per plant can vary depending on the breeder's request based on the number of seeds required for the subsequent experiment. We, however, did not do a duplicate analysis of the F1 plants. We acknowledge the importance of having technical replicates in an experiment; however, the accuracy of the KASP assay is well established (as cited in our manuscript). The specificity of the KASP assay means that even one validated marker can accurately distinguish between parents and offspring, and using up to 10 markers reduces the chance of genotyping error significantly. Therefore, using technical replicates may not be worth the cost; instead, we could increase the number of markers in case of any doubt or possible errors.

Competing Interests: No competing interest

Reviewer Report 20 June 2022

https://doi.org/10.21956/gatesopenres.14950.r32149

? **John Damien Platten**

International Rice Research Institute, Los Baños, Philippines

I thank the authors for taking the time to address the issues raised in the previous review. I do think there are still some outstanding issues that would substantially improve the case for the work presented:

- Articulation of what use cases are in mind needs to be better. Currently, the articulation is basically "low-throughput applications that may require fast turnaround time". This is true to a certain extent, but not the strongest case; it could easily be argued that any "low-throughput" request could either pay the extra up-front fees, as this will be cheaper than maintaining a lab just for this purpose, or samples aggregated with other larger jobs. In the end, this is essentially saying that because certain breeders were not organised in their workflow, we need to maintain an entire lab for them.

- I still don't see any costing of the procedures. This is sorely needed, even if this costing only includes consumables and not salaries of dedicated staff. It should be compared with

service providers for jobs of the same # samples and # SNPs.

Is the rationale for developing the new method (or application) clearly explained?
Yes

Is the description of the method technically sound?
Yes

Are sufficient details provided to allow replication of the method development and its use by others?
Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Marker design, marker evaluation, marker QC metrics, marker-assisted introgression, molecular breeding

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

> Author Response 28 Jul 2022
> **Queen Offornedo**
>
> The authors thank the reviewers for taking the time to review the manuscript and raise critical issues. We trust that addressing these issues will immensely improve the paper. Below are the responses to the concerns raised.
>
> **Reviewer 1 comment**: I thank the authors for taking the time to address the issues raised in the previous review. I do think there are still some outstanding issues that would substantially improve the case for the work presented:
> Articulation of what use cases are in mind needs to be better. Currently, the articulation is basically "low-throughput applications that may require fast turn-around time". This is true to a certain extent, but not the strongest case; it could easily be argued that any "low-throughput" request could either pay the extra up-front fees, as this will be cheaper than maintaining a lab just for this purpose, or samples aggregated with other larger jobs. In the end, this is essentially saying that because certain breeders were not organized in their workflow, we need to maintain an entire lab for them.

Authors' response: Besides the fact that the logistics for sending a low sample volume is not cost-effective, we also indicated in our previous response that there is a benefit of turn-around time when we genotype in-house. We had also indicated that neither the lab nor the equipment is procured solely for this genotyping purpose. We reiterate that the lab is a shared facility serving multiple activities, as are the equipment; We have re-purposed the real-time PCR for KASP assay. There is no maintenance cost for this workflow. The only thing dedicated to the workflow is the Klustercaller software.

We still outsource samples for the routine forward breeding application. However, there are applications where a quick genotyping of a small number of samples has to be done, for QC purposes, for breeders and partner seed companies. In summation, this study is focused on the scientific rigour rather than the business proposition of the technological workflow.

Reviewer 1 comment: I still don't see any costing of the procedures. This is sorely needed, even if this costing only includes consumables and not salaries of dedicated staff. It should be compared with service providers for jobs of the same # samples and # SNPs.

Authors' response: A statement detailing the cost comparison of the procedure has been included in the Introduction section (P3L6): "The standard cost for routine KASP genotyping by Intertek is $2.6 per sample per 10 SNPs, excluding shipping costs, compared to our in-house genotyping at $2.95."

Competing Interests: No competing interests.

Version 1

Reviewer Report 18 March 2022

https://doi.org/10.21956/gatesopenres.14581.r31823

✗ **John Damien Platten**
International Rice Research Institute, Los Baños, Philippines

Is the rationale for developing the new method (or application) clearly explained?
The value proposition for developing the 'method' is not well articulated. The authors mention several times the advantages of an in-house genotyping platform, and the general thrust of the paper is describing successful proof-of-concept application of some standard components of a SNP genotyping protocol. However it is not especially clear if the authors are aiming to establish

this service as a cost-effective alternative to outsourcing, to meet a specific need that outsourcing does not meet, or something else.

Alongside this, it is not clear what the novelty of the new method is. The entire workflow represents an implementation of standard technologies (CTAB extraction, DNA quantification, KASP genotyping). None of these are new techniques, nor is their combination into a genotyping workflow.

Is the description of the method technically sound?
As with point number 1, the overall description is technically sound, but several key details are overlooked. The machinery used in the critical step of plate scanning (actual data acquisition) is described, but key parameters are missing (please substitute equivalent parameters depending on the model of machine):

- What settings are used for lamp energy?
- What filters are used for excitation and emission spectra? This should include part numbers, and technical details of their performance.
- How is the analysis (clustering) done?
- How are results aggregated and conclusions drawn?

Are sufficient details provided to allow replication of the method development and its use by others?
See the previous comments. Some aspects are adequately described, but some others are sparse on critical technical details.

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Largely not applicable.

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
This does not seem to be the case. In particular, benchmarking data on the capacity, technical performance, cost etc. are lacking. This makes it impossible to judge the merits of this in-house system compared to outsourcing options.

Overall conclusion:
The current manuscript shows ability to technically execute on a relatively small number of samples in a modest timeframe. However to be a substantial contribution in this space, more thought needs to be given to better articulate both the value proposition of the work, and provide some benchmarking data to back this up. For example if the overall purpose is to show the benefit of having an in-house genotyping platform as opposed to (or in addition to) outsourcing options, the following factors and results might be considered:

- What is the value of an in-house system? Turnaround time and flexibility are mentioned, which I agree with. However why is this particularly important, to justify the expense of setting up, maintaining and operating an in-house system? Are there logistical considerations that prevent the use of outsourced options? Is the in-house system functionally *superior* to outsourced options? Is there a particular part of the breeding process that does not lend itself to standard outsourced options – and if so, under what circumstances would it be advisable to use the in-house or outsourced options? See below

comments on benchmarking.

○ **Full** cost assessment of the in-house system, including salaries of technical staff, machine maintenance and depreciation. Some description of the staff involved (number of positions executing on various duties) would also be helpful.

○ Also an assessment of technology life-cycles; genotyping platforms are evolving rapidly. I have seen many cases of expensive machines being purchased, only to sit idle as the technology has moved on even before they are delivered. KASP is likely to be replaced in the next 5 years. How would the cost of staying up to date and current be factored in?

○ Exploration of capacity. The authors mention completing 3 jobs (637 samples) in two weeks. This is plausible based on personal experience, though I have seen in-house systems with far higher throughput. However this is a far cry from handling 20,000 samples at peak operating times. This relates back to the first point.

○ Also related to capacity, an exploration of current/anticipated peak demand for the system.

○ Technical performance metrics:
    ○ Average DNA quality.
    ○ Call rate: what percentage of samples×markers (datapoints) do not amplify?
    ○ Clarity: What proportion of datapoints are unscorable?
    ○ Reproducibility: Amongst technical and biological replicates, what proportion of datapoints are scored incorrectly/inconsistently?

○ Benchmarking against available outsourcing options. A comparison of parameters such as these:
    ○ Turnaround time.
    ○ **Full** cost per sample and per datapoint.
    ○ Monthly capacity (samples and datapoints).

In-house genotyping platforms can and do have merit and justification. However until these issues can be addressed, the manuscript in its current form offers no fundamental insights into how such a platform could add value to breeding over outsourcing options.

If the authors can better explain why their hub is superior over other options, backed up with benchmarking data such as specified, this would greatly enhance its value.

Is the rationale for developing the new method (or application) clearly explained?
Partly

Is the description of the method technically sound?
Partly

Are sufficient details provided to allow replication of the method development and its use by others?

Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Marker design, marker evaluation, marker QC metrics, marker-assisted introgression, molecular breeding

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 23 Mar 2022

Queen Offornedo

The authors are very grateful to the reviewer who thoroughly read the manuscript and raised critical issues. We believe that addressing these issues will immensely improve the paper. Below are the responses to the issues raised. We noticed that our core message, which is complementing outsourcing in some situations, was not clearly articulated. We await additional reviewers' comments. We plan to revise the manuscript based on our response below and additional reviewers' comments.

Is the rationale for developing the new method (or application) clearly explained? Reviewer's comment:
The value proposition for developing the 'method' is not well articulated. The authors mention several times the advantages of an in-house genotyping platform, and the general thrust of the paper is describing successful proof-of-concept application of some standard components of a SNP genotyping protocol. However it is not especially clear if the authors are aiming to establish this service as a cost-effective alternative to outsourcing, to meet a specific need that outsourcing does not meet, or something else.
Author's response:
The core message of our paper is to complement, not establish an alternative, to outsourcing (please see the third paragraph on page 3). We duly recognize the cost-effectiveness of the genotyping platform facilitated by CGIAR platforms such as HTPG/EiB. The need for developing such in-house workflow had been prompted by the following factors:

1. The current available genotyping platforms have a minimum sample size requirement. For instance, the EiB facilitated genotyping at Intertek costs $2 / sample if the user orders genotyping of 1536 samples (a set of 16 plates; four plates are

acceptable but price increases). Breeders often want to fingerprint a few dozen lines urgently for identity or parentage analysis. In such cases, sending less than the minimum number of samples is not only more priced per datapoint but entails shipping cost and a turn-around time of 2-3 weeks. Using other markers, such as SSR, is more expensive and cumbersome. The use of genotyping systems such as KASP alleviates all these issues.

2. Logistical issues related to shipping by courier: In this part of the world, courier services are not very satisfactory and reliable, often resulting in damage to samples in transit or longer than normal delays, which may reduce the quality of perishable specimens. If a reasonably affordable system is available locally, it can circumvent such problems.

3. The instruments used for this work are all standard instruments available in most molecular biology labs. Our workflow shows the re-purposing of these instruments for the genotyping workflow. For instance, the qPCR machine, which is mostly used for expression analysis, was adapted to KASP genotyping with the installation of appropriate software for SNP calling. Likewise, the Fluostar plate reader was used for plate-level DNA quantification in lieu of single sample analysis by Spectrophotometer.

Reviewer's comment:

Alongside this, it is not clear what the novelty of the new method is. The entire workflow represents an implementation of standard technologies (CTAB extraction, DNA quantification, KASP genotyping). None of these are new techniques, nor is their combination into a genotyping workflow.

Author's response:

This manuscript is about a workflow that combines carefully chosen and optimized best practices in lab techniques at different stages of genotyping to address pertinent problems faced by researchers in Sub-Saharan Africa (SSA). For users who want to genotype few samples quickly, some bottlenecks in the workflow has to be removed. Currently, the DNA extraction throughput has improved by isolating and quantifying DNA at a plate level (i.e., processing 96 samples simultaneously). Secondly, genotyping by other systems such as SSR is not cost-effective. Therefore, by implementing such a workflow, we could generate quality data quickly for application in the breeding pipeline. It should be noted that not many labs in developing countries are capable of using the KASP system in-house.

Is the description of the method technically sound?
Reviewer's comment:

As with point number 1, the overall description is technically sound, but several key details are overlooked. The machinery used in the critical step of plate scanning (actual data acquisition) is described, but key parameters are missing (please substitute equivalent parameters depending on the model of machine):

○ What settings are used for lamp energy?

○ What filters are used for excitation and emission spectra? This should include part numbers, and technical details of their performance.

○ How is the analysis (clustering) done?

How are results aggregated and conclusions drawn?

Author's response:

The required information will be incorporated in the revised manuscript under the

subsection "KASP genotyping and data analysis", as explained below:

The description of the parameters for the LC480 II qPCR machine is outlined in the LC480 manual. To perform the KASP genotyping experiment on the LC480 II machine, we used the Endpoint Genotyping Analysis module within the LightCycler software, adjusting the parameters as outlined in the KASP genotyping protocol provided by LGC Biosearch Technologies. The Endpoint genotyping analysis module is based on the use of dual hydrolysis probes, which are designed for wild-type and mutant target DNA and are labelled with different dyes (FAM and HEX). However, when using a non-qPCR machine (such as the GeneAmp PCR System 9700) for amplification, a third colour probe (ROX) normalizes the fluorescence measurement. The LightCycler software within the LC480 II machine determines the sample genotypes automatically by measuring the intensity distribution of the two probes after a PCR amplification step. The relative dye intensities are then visualized in a scatter (cluster) plot that discriminates them as wild-type, heterozygous mutant, or homozygous mutant samples. The LightCycler software automatically groups similar samples and assigns genotypes based on the intensity distribution of the two dyes. The KASP amplification conditions included one cycle of KASP unique Taq activation at 94°C for 15 min, followed by 36 cycles of denaturation at 94°C for 20 s, and annealing and elongation at 60°C (dropping 0.6°C per cycle) for 1 min. Endpoint detection of the fluorescence signal was acquired for 1 min at 30°C when using the LightCycler 480 II real time-PCR System or read using the FLUOstar Omega Microplate reader (BMG Labtech, SA) when using the GeneAmp PCR System 9700. For fluorescence detection, the filter combination for the Excitation and Emission wavelength of both dyes was set at 465 – 533 (FAM) and 523 – 568 (HEX), respectively, when using LC480 II, and 485 - 520 (FAM), 544 - 590 (HEX) and 584 - 620 (ROX) when using FLUOstar Omega Microplate reader. The genotype calls were exported from the LightCycler software as fluorescent intensities of each sample in ".txt" file format and imported for analysis in the KlusterCaller analysis software (LGC Biosearch Technologies). The KlusterCaller software adjusted the cluster plot axes to enable the proper calling of genotypes. The genotype calls were grouped as homozygous for allele X (allele reported by FAM, X-axis), homozygous for allele Y (allele reported by HEX, Y-axis), heterozygous (alleles reported by FAM and HEX, between X- and Y-axis), or uncallable. The result from the KlusterCaller was exported in two file formats (".csv" and ".txt"). The ".csv" file was imported into the SNPviewer2 version 4.0.0 software (LGC Biosearch Technologies), where the cluster plot image was viewed and downloaded for publication. The genotype calls in the ".txt" file were used to calculate the genetic distance using the PowerMaker 3.25 statistical software.

Are sufficient details provided to allow replication of the method development and its use by others?
Reviewer's comment:
See the previous comments. Some aspects are adequately described, but some others are sparse on critical technical details.

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Reviewer's comment:
Largely not applicable.

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Reviewer's comment:
This does not seem to be the case. In particular, benchmarking data on the capacity, technical performance, cost etc. are lacking. This makes it impossible to judge the merits of this in-house system compared to outsourcing options.
Author's response:
As mentioned above, the aim of the publication is not to replace outsourcing and we do not envisage competition with outsourcing. We are fully aware of the cost-effectiveness of the highly automated/robotic genotyping services accessible to us. We frequently use these services. The rationale for the workflow is to process fewer samples quickly. The time saved is invaluable. Loss of samples during shipments is a setback, which cannot be monetized easily.

Overall conclusion:
Reviewer's comment:
The current manuscript shows ability to technically execute on a relatively small number of samples in a modest timeframe. However to be a substantial contribution in this space, more thought needs to be given to better articulate both the value proposition of the work, and provide some benchmarking data to back this up. For example if the overall purpose is to show the benefit of having an in-house genotyping platform as opposed to (or in addition to) outsourcing options, the following factors and results might be considered:
Author's response:
We do not intend to establish a rival genotyping service to replace outsourcing. As explained above, this workflow is what we are using for a while now for processing a smaller number of samples (smaller than the minimum sample required by service vendors). This gives the flexibility to assay multiple crops with multiple markers in a single plate in a matter of hours. We are a small group dedicated to maize genomics. All we want is to share our methods with partners in the same situation. As can be seen from the stats of the preprint, our manuscript is already making an impact with 20 downloads and 150 views.

Reviewer's comment:
What is the value of an in-house system? Turn-around time and flexibility are mentioned, which I agree with. However why is this particularly important, to justify the expense of setting up, maintaining and operating an in-house system? Are there logistical considerations that prevent the use of outsourced options? Is the in-house system functionally superior to outsourced options? Is there a particular part of the breeding process that does not lend itself to standard outsourced options – and if so, under what circumstances would it be advisable to use the in-house or outsourced options? See below comments on benchmarking.
Author's response:
Addressed above and below.

Reviewer's comment:
Full cost assessment of the in-house system, including salaries of technical staff, machine maintenance and depreciation. Some description of the staff involved (number of positions

executing on various duties) would also be helpful.
Author's response:
Not applicable.

Reviewer's comment:
Also an assessment of technology life-cycles; genotyping platforms are evolving rapidly. I have seen many cases of expensive machines being purchased, only to sit idle as the technology has moved on even before they are delivered. KASP is likely to be replaced in the next 5 years. How would the cost of staying up to date and current be factored in?
Author's response:
We agree that the genotyping platforms are evolving rapidly. To make it clear, we have not purchased instruments solely for this technique. We have only re-purposed the existing machines. Both the qPCR machine and the plate reader have high demands for other uses. If we cease to use these machines for the KASP system, the normal utilization of the machines will continue.

Reviewer's comment:
Exploration of capacity. The authors mention completing 3 jobs (637 samples) in two weeks. This is plausible based on personal experience, though I have seen in-house systems with far higher throughput. However this is a far cry from handling 20,000 samples at peak operating times. This relates back to the first point.
Author's response:
Our response here is related to the above explanation. When we have a large volume of samples, we use low-density and mid-density genotyping service providers.

Reviewer's comment:
Also related to capacity, an exploration of current/anticipated peak demand for the system.
Technical performance metrics:
- Average DNA quality.
- Call rate: what percentage of samples × markers (datapoints) do not amplify?
- Clarity: What proportion of datapoints are unscorable?
- Reproducibility: Amongst technical and biological replicates, what proportion of datapoints are scored incorrectly/inconsistently?
- Benchmarking against available outsourcing options. A comparison of parameters such as
- these:
- Turn-around time.
- Full cost per sample and per datapoint.
- Monthly capacity (samples and datapoints).
- In-house genotyping platforms can and do have merit and justification. However until these issues can be addressed, the manuscript in its current form offers no fundamental insights into how such a platform could add value to breeding over outsourcing options.

If the authors can better explain why their hub is superior over other options, backed up with benchmarking data such as specified, this would greatly enhance its value.

Author's response:
As mentioned above, the aim of the publication is not to replace or compete with outsourcing, rather complement it, particularly in cases of a small volume of samples that are not cost-effective for outsourcing.

Is the rationale for developing the new method (or application) clearly explained?
Reviewer's comment: Partly

Is the description of the method technically sound?
Reviewer's comment: Partly

**Are sufficient details provided to allow replication of the method development and its use
by others?**
Reviewer's comment: Partly

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?
Reviewer's comment: No source data required

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?
Reviewer's comment: No

Competing Interests:
No competing interests were disclosed.

Competing Interests: No competing interests.

- -

Comments on this article

Version 1

Reader Comment 04 Apr 2022
Godfree g.chigeza@cgiar.org, IITA, Lusaka, Zambia

Great information laid down in a simple form. Congratulations to the authors for such a great job.

Competing Interests: Work at the same institution but this has not influenced my comment.

Author Response 14 Mar 2022

Queen Offornedo

The authors thank all who have viewed and downloaded our article.
We would appreciate it if you could take a minute to tell us what you think of the article in the "Add a comment" section.
Thank you.

Competing Interests: No competing interests were disclosed.