

## Developing KASP Markers for Identification of Basmati Rice Varieties

Steele, Katherine; Quinton-Tulloch, Mark; Burns, Malcolm; Nader, Werner

### Food Analytical Methods

DOI:

[10.1007/s12161-020-01892-3](https://doi.org/10.1007/s12161-020-01892-3)

Published: 01/04/2021

Publisher's PDF, also known as Version of record

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*

Steele, K., Quinton-Tulloch, M., Burns, M., & Nader, W. (2021). Developing KASP Markers for Identification of Basmati Rice Varieties. *Food Analytical Methods*, 14(4), 663-673.  
<https://doi.org/10.1007/s12161-020-01892-3>

#### Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# Developing KASP Markers for Identification of Basmati Rice Varieties

Katherine Steele<sup>1</sup> · Mark Quinton Tulloch<sup>1</sup> · Malcolm Burns<sup>2</sup> · Werner Nader<sup>3</sup>

Received: 27 May 2020 / Accepted: 12 October 2020

© The Author(s) 2020

## Abstract

Authentication of Basmati rice has relied on microsatellite markers since 2004, but microsatellites cannot distinguish between all of the forty-one Basmati varieties approved in 2017. This study investigated whether single nucleotide polymorphisms (SNP) and insertion/deletion (InDel) variations developed into KASP™ (Kompetitive Allele Specific PCR; LGC Biosearch Technologies) could be used to distinguish between commercial Basmati varieties. Suitable loci were identified by comparing whole genome sequences of 120 diverse rice accessions. Sequences flanking these loci were standardized across a wide range of rice genomes to produce optimal KASP designs. We selected 364 KASP designs to use for genotyping; they were either near to informative microsatellite markers, within the *Badh2* and *Waxy* genes, or distributed throughout the rice genome. Genotypes for 327 KASP were obtained with 255 loci revealing polymorphism in up to 41 samples of approved Basmati varieties and 20 non-Basmati varieties. The varieties genotyped had not been used in the KASP design process. KASP were able to distinguish between commercial Basmati varieties that could not be distinguished with currently available microsatellites. Thirty-seven Basmati varieties could be distinguished from all others with between 3 and 8 KASP markers out of a pool of 98 informative markers. A reduced set of 24 KASP markers could determine whether a sample belongs to one of eight family groups. All of the KASP markers used in this study can be purchased from LGC Biosearch Technologies. These markers have potential to be used by industry for routine testing and regulation.

**Keywords** Authenticity · DNA fingerprinting · Fragrance · Genomics · KASP · *Oryza sativa*

## Introduction

Basmati is a type of long-grained fragrant rice, and only certain named varieties grown in specified regions of India and Pakistan are permitted to carry the name Basmati (Singh et al. 2000). Authorities in India and Pakistan have recently approved specific modern varieties derived from crosses with Basmati landraces to be sold labeled as Basmati if they conform to certain physical and aromatic characteristics (Mahajan et al. 2018; Roy et al. 2018). Basmati rice imported into the

EU and UK is subject to regulations (whereby nine approved Basmaties have tariff free status) and by the industry Code of Practice (CoP) to ensure that it is accurately described and presented for sale so as not to mislead the consumer. In 2017, the number of approved Basmati varieties was revised and increased from 15 to 41, and the CoP was updated to reflect this (Rice Association 2017). However, concerns were raised within the rice industry that the increased number of Basmati rice varieties coming onto the market may not all be identifiable with currently used DNA markers. There were also concerns that some of the new varieties could have a more distant relationship to landrace varieties than the previously approved Basmati varieties.

Two previous studies were funded by the Food Standards Agency (UK) to identify DNA markers for Basmati authentication. Initially, Bligh (2000) identified 12 microsatellite markers (also known as simple sequence repeats, SSR) that together could be used to distinguish Basmati varieties from non-Basmati varieties. Then, between 2003 and 2008, over 200 publically available rice microsatellite markers and 50 InDel markers were tested at Bangor University on the 15 Basmati varieties listed in the 2005 CoP, and informative

**Supplementary Information** The online version of this article (<https://doi.org/10.1007/s12161-020-01892-3>) contains supplementary material, which is available to authorized users.

✉ Katherine Steele  
k.a.steele@bangor.ac.uk

<sup>1</sup> School of Natural Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, UK

<sup>2</sup> LGC, Queens Road, Teddington, Middlesex TW11 0LY, UK

<sup>3</sup> Eurofins Global Control GmbH, Hamburg, Germany

markers were identified and used in surveillance in the UK (Steele et al. 2008; Woolfe and Steele 2019). Subsequently, Eurofins Genomics (Ebersberg, Germany) selected 13 SSR markers from these two previous studies and added two additional SSR markers to come up with a set of 15 microsatellites. These 15 SSRs have been widely used by industry regulators in the EU for over a decade to determine between the Basmati varieties and quantitatively estimate the ratios of Basmati to non-Basmati grains present in mixtures.

An additional PCR marker known as *fgr* (Bradbury et al. 2005) that is based on InDel and SNP variation is commonly used to detect non-fragrant rice. Together, these 16 markers were used to test the recently approved Basmati varieties that were added to the CoP in 2017 (Nader et al. 2019). The results revealed that five of the newly approved Basmati varieties (Yamini CSR-30, Pusa Basmati 3, Vallabh Basmati 22, Vallabh Basmati 23, and Pusa 1637) had the same allele profiles as other permitted varieties in the 2017 CoP (Rice Association 2017).

KASP is a different type of PCR-based marker system that identifies DNA polymorphism at specific genomic loci; these can be either an InDel (insertion/deletion of one or more nucleotides) or a single nucleotide polymorphism (SNP). It is not possible to develop KASP markers that directly target SSR sequences. KASP requires endpoint PCR with detection of two fluorescent dyes indicating the genotype depending on the amount of each of two alternative primers that become incorporated into amplicons. The reaction results in co-dominant allele calls for the target variant SNP or InDel. Successful application of KASP technology relies on primer designs that account for variations between non-target sequences of different rice varieties to ensure detection of the target polymorphic alleles (Steele et al. 2018). Suitable variations occur throughout the genome, so for this study, it was considered strategic to use target SNP or InDel variants located within 10 kb of the microsatellites currently used for Basmati testing and rice authentication or that have previously shown high levels of polymorphism in rice.

In non-aromatic rice a protein product, BADH2 (also known as BAD2) is produced as the gene product of the *BADH2* gene. But in fragrant rice, BADH2 is not produced, and its absence enables the biosynthesis of 2-acetyl-1-pyrroline which accounts for the characteristic aroma of cooked Basmati rice (Chen et al. 2008). Basmati varieties, along with some other fragrant rice varieties, contain an 8 bp deletion and two SNPs (*badh2.1*) in exon 7 of the *BADH2* gene, and this is the most common allele of *BADH2* responsible for aroma in rice (Kovach et al. 2009). The PCR marker *fgr* (Bradbury et al. 2005) targets the *badh2.1* variant, and it is widely used in Basmati authentication tests, so SNP and InDel variants in the *BADH2* gene were predicted to be potential informative KASP markers.

Glutinous rice does not contain granule-bound starch synthase (GBSS), the protein product encoded by the *Waxy* gene, because only immature 3.3 kb *Wx* pre-mRNA containing intron I is produced. Basmati is an intermediate between glutinous and non-glutinous rice having both splice variants of the *Waxy* gene (mature 2.3 kb *Wx* mRNA and immature 3.3 kb *Wx* pre-mRNA) resulting in a phenotype with intermediate gelatinization temperature (70–74 °C) and medium gel strength, two important traits that characterize Basmati rice. Variants in the *Waxy* gene were, therefore, predicted to be potential informative KASP markers.

The objective of this study was to test 364 selected KASP marker designs for detection of polymorphism and utility in discrimination among the varieties listed as Basmati in the 2017 CoP and between approved Basmati and non-approved rice varieties.

## Methods

### Sourcing of Reference Materials

The UK Rice Association requested samples of all approved Basmati varieties from the All India Rice Exporters' Association (AIREA) and Rice Exporters Association of Pakistan (REAP). Thirty-three samples representing different varieties of polished grains were received from them in 2017. India was the original source of 25 samples, and Pakistan was the original source of 8 samples (Table 1). The 2017 CoP lists Punjab Basmati 2 twice, but only one sample with that name was received (from India). The Indian Punjab Basmati is also known as Bauni Basmati (Roy et al. 2018), while unrelated the Pakistani Punjab Basmati is genetically close to Chenab Basmati. Only one approved sample labeled as Punjab Basmati was received, and it was from Pakistan; therefore, Punjab Basmati (Bauni Basmati) was missing from the analysis. Other approved varieties (Table 1) were from Bangor University's collection of rice reference materials that had been obtained previously by the Food Standards Agency between 2003 and 2011 from verified sources in India and Pakistan, IRRI, and from a collection at Nottingham University collected before 2000 (Bligh et al. 1999). The final set of 41 approved Basmati varieties available for testing included two approved samples labeled Shaheen and Shaheen Basmati.

Reference materials of 19 non-approved varieties were selected from Bangor's collection (Table 1), either because they have previously been detected as adulterants to commercial Basmati or because they could potentially be used as adulterants, or alternatively, in certain cases and subject to appropriate agreements, they could

**Table 1** Names and origins of rice samples used in this study

Variety name	Bangor accession ID	Classification	Sample obtained from
Basmati 217		Approved Basmati—ZID	IRRI seed bank
Basmati 370		Approved Basmati—ZID	India (Rice Association)
Basmati 386		Approved Basmati—ZID	India (Rice Association)
Kernal 574***		Approved Basmati—ZID	Food Standards Agency
Pusa Basmati 1		Approved Basmati—ZID	India (Rice Association)
Ranbir		Approved Basmati—ZID	India (Rice Association)
Super Basmati		Approved Basmati—ZID	Food Standards Agency
Taraori		Approved Basmati—ZID	India (Rice Association)
Type 3		Approved Basmati—ZID	Food Standards Agency
Basmati 198	S51	Approved Basmati	Food Standards Agency
Basmati 2000		Approved Basmati	Pakistan (Rice Association)
Basmati 385	S29	Approved Basmati	Food Standards Agency
Basmati 515		Approved Basmati	Pakistan (Rice Association)
Basmati 564		Approved Basmati	India (Rice Association)
Haryana Basmati (HKR 228/IET 10367)		Approved Basmati	India (Rice Association)
Improved Pusa Basmati 1 (Pusa 1460)		Approved Basmati	India (Rice Association)
Kasturi (IET 8580)		Approved Basmati	India (Rice Association)
Kissan Basmati		Approved Basmati	Pakistan (Rice Association)
Mahi Sugandha	S94	Approved Basmati	Food Standards Agency
NIAB Basmati 2016		Approved Basmati	Pakistan (Rice Association)
Noor Basmati		Approved Basmati	Pakistan (Rice Association)
Pant Basmati 2		Approved Basmati	India (Rice Association)
Punjab Basmati (Pakistan)*		Approved Basmati	Pakistan (Rice Association)
Punjab Basmati 2**		Approved Basmati	India (Rice Association)
Punjab Basmati 3		Approved Basmati	India (Rice Association)
Pusa Basmati 1121		Approved Basmati	India (Rice Association)
Pusa Basmati 1509		Approved Basmati	India (Rice Association)
Pusa Basmati 1609		Approved Basmati	India (Rice Association)
Pusa Basmati 1637		Approved Basmati	India (Rice Association)
Pusa Basmati 1728		Approved Basmati	India (Rice Association)
Pusa Basmati 6 (Pusa 1401)		Approved Basmati	India (Rice Association)
Shaheen Basmati		Approved Basmati	Pakistan (Rice Association)
Shaheen	S73	Approved Basmati	Food Standards Agency
Vallabh Basmati 22		Approved Basmati	India (Rice Association)
Vallabh Basmati 23		Approved Basmati	India (Rice Association)
Yamini (Basmati CSR 30)		Approved Basmati	India (Rice Association)
Chenab Basmati		Approved Basmati	Pakistan (Rice Association)
Malviya Basmati Dhan 10-9		Approved Basmati	India (Rice Association)
Pant Basmati 1		Approved Basmati	India (Rice Association)
Vallabh Basmati 21		Approved Basmati	India (Rice Association)
Vallabh Basmati 24		Approved Basmati	India (Rice Association)
American long-grain aromatic	S69	Non-Basmati	Nottingham University
Basmati 105 (BLB 105)		Non-Basmati	NIBGE, Pakistan
Basmati C22 Peshwari	S-99.3	Non-Basmati	Nottingham University
Mugad Sugandha	S24	Non-Basmati	BAU, India
Niab/221-9	S36	Non-Basmati	University of Faisalabad, Pakistan
Pak386	S75	Non-Basmati	Grain Tech, Pakistan
Pusa Suganda 2	S87	Non-Basmati	United Riceland, India
Pusa Suganda		Non-Basmati	United Riceland, India

**Table 1** (continued)

Variety name	Bangor accession ID	Classification	Sample obtained from
Pusa Sugandha	2009	Non-Basmati	United Riceland, India
Sherbati awns	S86	Non-Basmati	United Riceland, India
Sherbati	S49	Non-Basmati	Grain Tech, Pakistan
Sherbati short	S85	Non-Basmati	United Riceland, India
Sugandha	S97	Non-Basmati	Food Standards Agency
S. Shabnam	S95	Non-Basmati	Food Standards Agency
Superfine	S77	Non-Basmati	Grain Tech, Pakistan
IR6	S36	Non-fragrant	University of Faisalabad, Pakistan
KS282	S37	Non-fragrant	University of Faisalabad, Pakistan
Basmati_1	S-105.1 (147)	Non-verified Basmati	Nottingham University
Commercial Sample		Non-verified Basmati	Commercial source

\*Punjab Basmati is listed in the 2017 CoP as “Punjab Basmati (Bauni Basmati)”; however, in this study, no sample with either of these names was obtained from India but a sample labeled Punjab Basmati was obtained from Pakistan. Punjab Basmati from India is also known as Bauni Basmati (Roy et al. 2018), and a sample of it obtained from India by the Food Standards Agency in 2004 was tested with microsatellites by Nader et al. (2019), revealing that Punjab Basmati and Bauni Basmati were distinct genotypes. Only Punjab Basmati (Pakistan) was tested in the current study

ZID zero import duty required for import into the European Union, *IRRI* International Rice Research Institute, *BAU* Birsa Agricultural University, *NIBGE* National Institute for Biotechnology and Genetic Engineering

\*\*Punjab Basmati 2 appears twice in the CoP 2017, but only one sample with this name was received (from India). It is likely that the editors of the 2017 CoP intended one of them to be Punjab Basmati (Pakistan)

\*\*\*Kernal is also known as Kernel Basmati or Pakistani Basmati

potentially be approved as Basmati in the future. All reference materials were stored at 4°C.

## DNA Extraction

The study was designed to maximize the number of KASP genotyping assay designs tested and not the number of replicate samples. Nineteen non-replicated samples were extracted using the CTAB method described by Saghai-Marooof et al. (1984) from 20 g of multiple grains of a single sample ground in a coffee grinder (cleaned and decontaminated between every sample, with the first ground material from each separate sample discarded). DNA was also extracted from 60 rice samples from single grains (up to five replicates per sample) following the sampling and extraction protocol described in the standard operating procedure for the identification of selected Basmati rice varieties (FSA 2005) using Phytopure reagents (GE Healthcare) supplied by Sigma-Aldrich.

DNA samples were tested for concentration and quality (A260:A280 and A260:A230) with a NanoDrop™ 1000 spectrophotometer (Thermo Fisher). In addition, the size and integrity of the DNA was checked by running 5 µl of each extracted sample (not normalized) mixed with 2-µl loading buffer, on a 0.8% agarose gel stained with Safeview (Bioline). A control of IR64 rice leaf DNA extracted with a Qiagen DNeasy kit was run alongside the samples on the gel. The extract from one replicate of each sample, extracted with either method, with the highest DNA concentration measured

with the NanoDrop™ 1000, was selected for use in the first round of KASP screening. Only samples extracted with the Phytopure methods were selected for the second screen.

## Design and Selection of KASP Markers

To design KASP, sequence data from 120 publically available re-sequenced rice genomes were mined to identify flanking sequences of variants due to SNPs (single nucleotide polymorphisms) or InDels (insertion/deletions). The bioinformatics approaches used were alignment of each genome against the Shuhui498 indica reference genome (Du et al. 2017), Bowtie2 for sequence read alignment (Langmead and Salzberg 2012), and SAMtools for genotype likelihood calculation and variant calling (Li et al. 2009). The appropriate International Union of Pure and Applied Chemistry (IUPAC) nucleotide ambiguity codes were used to incorporate all detected variation into KASP primer design to improve versatility of the markers compared to the earlier designs that were based on only nine re-sequenced genomes (Steele et al. 2018). For each KASP design a .bed file containing the sequences spanning 60 bp on each side of the target variant was used to generate primers with the Kraken™ software by LGC, Biosearch Technologies (Hoddesden, UK).

From over 1 million new designs, 364 KASP markers were chosen for this study of which 139 were validated in this study for the first time. They were selected to target variations

considered to be relevant to Basmati varieties using the following three criteria:

- (i) Within 10 kb of the genome alignment positions of primers for the microsatellite markers widely used in Basmati authentication testing, including RM1, RM223, RM202, RM44, RM201, RM229, RM241, RM171, RM55, RM263, RM212, RM252, RM282, RM339, RM55, RM5432, and RM72. Up to 5 KASP were selected that were within 5 kb for each microsatellite.
- (ii) In proximity of the fragrance-determining gene *BADH2* on chromosome 8 (all designs targeting variations within 5 kb of *BADH2* were selected). One KASP targeted the 8 bp InDel within *Badh2.1*. Twelve KASP targeted other functional polymorphisms associated with other *Badh2* variants: a 7 bp deletion in exon 2 (Shi et al. 2008); a 7 bp insertion in exon 8 (Amarawathi et al. 2008); a 3 bp insertion in exon 13 (Myint et al. 2012); and a 3 bp deletion in 5' UTR (Shi et al. 2014). Due to insufficient sequence information in the original publication, it was not possible to design KASP in the positions relating to the specific *badh2* sequence variations described by Shao et al. (2013). A further six KASP designs were located in close proximity to microsatellites that show co-segregation with the major aroma locus.
- (iii) Located within the amylose gene (*Waxy*) locus on chromosome 6 (at 1.64 Mbp) that controls biosynthesis of GBBS which produces amylose, a major protein that binds to starch granules. Five KASP were designed for sequences within *Waxy* including two for target functional variants.
- (iv) Located in or near to genes considered of interest by breeders in India and Pakistan. A total of 225 KASP markers were selected from successfully validated new designs used for rice genotyping in Innovate UK project 103711.
- (v) Selected to give an even distribution of background markers across all 12 chromosomes.

## Genotyping

KASP genotyping was carried out by LGC Biosearch Technologies, Hoddesdon, UK, via their all-inclusive service for plant samples (<https://www.biosearchtech.com/services/genotyping-services/all-inclusive-services>) using their high-throughput workflow and genotyping instrumentation.

An initial main screen used 364 KASP markers to test 60 DNA samples (41 extracted with Phytopure from a single grain and 19 extracted using the CTAB method from multiple grain samples). Some CTAB-extracted samples failed to amplify with any markers or were only successful with a limited number of KASP markers. For this reason, a second screen with only 23

KASP was run on 21 varieties using Phytopure-extracted single grain samples, most in replicate (two to five). A maximum of five replicates were used for all varieties that had shown within variety microsatellite variation (Nader et al. 2019).

A data file containing called genotypes was provided for both screens by LGC Biosearch Technologies, and these could be analyzed using their SNPViewer™ software (<https://www.biosearchtech.com/support/tools/genotyping-software/snpviewer>). Where no allele could be called, a “?” was given in the spreadsheet; this could either be due to a null allele (due to a variant that is incompatible with the primer binding) or due to a failed PCR reaction which was most likely to be caused by poor quality DNA. Genotyping results from both rounds were combined into a single dataset in Microsoft Excel containing a maximum of 327 marker genotypes for each variety. In the few instances where different replicates of the same sample differed, the heterozygous genotype was called and used for subsequent analysis.

## Data Analysis

From 327 successful KASP markers, all monomorphic loci and loci with only one homozygous genotype present among the entire dataset were removed, leaving 255 polymorphic KASP to be analyzed with the aim of identifying the smallest and most informative subsets of KASP markers.

The markers were sorted according to the frequency of the rarest (minor) alleles. Those with the lowest minor allele frequencies (MAF) were then considered first, in a test procedure that added each marker genotype in turn, stepwise for each variety, until additional markers did not further distinguish the variety from other varieties in the test set. Any markers that were uninformative for all varieties were discarded. This procedure identified the greatest number of unique varieties that could be separated with as few markers as possible.

Eight varieties (Basmati 217, Ranbir, Taraori, Chenab Basmati, Pusa Suganda, Sugandha, S. Shabnam, Niab/221-9) that could not be uniquely identified using the remaining markers were removed. Then, for each uniquely identifiable variety, the above analysis was run with only the markers that were informative for that variety. This produced a small (< 9) subset of markers for each target variety that could be used to identify that a single specific variety from all the others.

Nine families were assigned as follows: (A) Basmati traditional types; (B) Super Basmati and Basmati 2000 types; (C) Pusa Basmati types; (D) other Basmati types; (V) likely to be Basmati but not approved references; (W) non-Basmati (fragrant) closely related to approved varieties; (X) other non-Basmati (fragrant); (Y) American long-grain fragrant rice, non-approved; and (Z) non-fragrant potential adulterants. To identify a set of markers that could distinguish between these families, a similar stepwise procedure to that described above was used. However, the order in which markers were considered

was determined by the number of families containing a variety with the minor allele genotype for each marker, with those markers with a minor allele in the most families being tested first. Secondary sorting of marker order was based on descending MAF. The same stepwise procedure was carried out for each variety, but testing for each variety stopped when all possible varieties, based on the genotypes of the markers included in the test set, belonged to the same family. A dendrogram based on the smallest subset of markers that could distinguish between families was produced using a neighbor joining clustering method in Tassel (Bradbury et al. 2007).

## Results

### DNA Concentration and Quality

CTAB- and Phytopure-extracted DNA samples measured by the NanoDrop™ 1000 both showed concentrations ranging from 4 to 400 ng/ul. The agarose gel analysis showed that samples with measured concentrations of < 20 ng were either faint or not visible and all samples had lower molecular weight than the Qiagen DNeasy-extracted control sample used on the gel. For this reason, only samples that were both clearly visible on the gel and had measured concentrations of > 50 ng/ul were considered suitable for genotyping with KASP technology, and these were selected for use in the first KASP screen. However, a large proportion of the CTAB-extracted samples failed with all or some of the KASP used in the first screen; hence, a second screen was carried out with Phytopure extracts of the same samples, but with a reduced number of KASP.

### KASP Genotyping

From 364 KASP designs submitted, 327 (90%) were converted successfully to KASP markers and used for genotyping the rice samples (Supplementary Data Sheet 1). In total, 255 (78%) of the successful KASP showed polymorphism in the set of varieties tested. Thirteen KASP were monomorphic and 59 revealed only one homozygous allele and were therefore not considered informative in this population. There was full agreement between the first and second round of genotyping for the varieties that were genotyped in both rounds.

### Within Variety Variation

Replicate single grain samples for three of the varieties did not show consistent alleles, whereas the other 18 varieties that were tested in replicate were consistent. The three variable samples were as follows:

- Basmati 564—one out of four replicates had the allele corresponding to the 8 bp insertion at BADH2\_PM while a different replicate was homozygous for C (not T) at RM166\_SNP\_7.
- Chenab Basmati—heterozygous at RM209\_SNP\_nn\_1 in two of the five replicates.
- Basmati 198—the two replicate samples carried different alleles for c9bg\_ff\_SNP\_4.

The two different samples Shaheen and Shaheen Basmati differed at five markers.

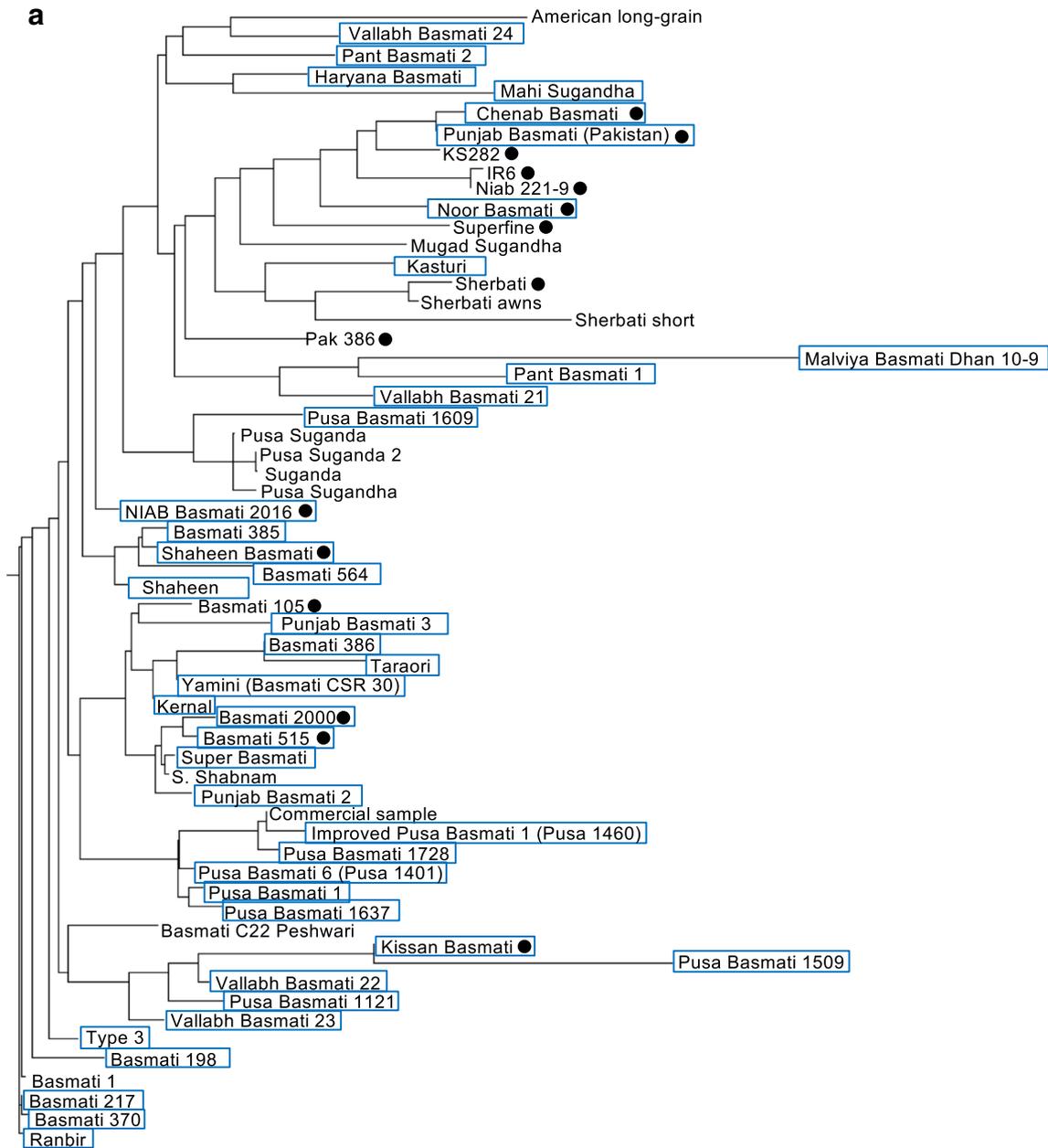
**Table 2** Varieties showing the same KASP genotypes as other varieties in this study, and therefore not included in the analysis of 52 for variety confirmation

Variety removed from variety confirmation set	Number of markers with missing data	Varieties with similar homozygous genotypes	Total number of markers genotyped in all similar varieties	Number of Heterozygous variants
Basmati 217	22	Basmati 370 Basmati_1 S-105.1 Ranbir	304	5
Ranbir	13	Basmati 370 Basmati_1 S-105.1 Basmati 217	304	1
Chenab Basmati	153	Punjab Basmati (Pakistan)	162	9
Taraori	135	Basmati 386 Kernal 574 Yamini (Basmati CSR 30)	189	0
S. Shabnam	11	Super Basmati	318	0
Sugandha	8	Pusa Suganda 2	320	0
Niab/221-9	10	IR6	320	1

### Identification of Individual Varieties and Families

Forty-four of the 60 varieties could be distinguished with 69 KASP markers (Supplementary Data Sheet 1, Variety ID column). Eight varieties were removed from the dataset for subsequent analysis, including Taraori and Chenab Basmati, because this study could not distinguish between them exclusively with homozygous alleles (Table 2). The remaining set

of 52 varieties, containing 37 approved Basmatris, could all be distinguished in a pairwise comparison of variety confirmation, with between three to eight separate KASP markers from a pool of 98 KASP (Supplementary Data Sheet 2). Discrimination at the level of nine families was possible with only 24 markers, but a small proportion were based on differences between heterozygous and homozygous calls (Supplementary Data Sheet 3 and Fig. 1).



**Fig. 1 a** UPGMA clustering based on 24 markers that discriminate between families. Boxes denote 41 approved basmati varieties, and circles indicate samples with verified origin in Pakistan. **b** Simple matching coefficient values 0 = identical (dark blue) 1 = entirely

different (dark red) for all pairwise comparisons between rice varieties from 24 selected KASP that distinguish between eight families (A–D approved Basmati accessions, V–Z non-approved accessions)



**Table 3** KASP markers with potential to quantify components in specific mixtures of varieties

Varieties	Informative KASP markers	Genotype at informative marker
Basmati 370	RM263_SNP_nn_2	Basmati 370 GG; Type 3 CC
Type 3	Pibdom_NSb_func_5	Basmati 370 TC; Type 3 CC
	RM263_SNP_nn_1	Basmati 370 GG; Type 3 AA
	RM263_SNP_nn_defra_3	Basmati 370 AA; Type 3 GG
	Pib_SNP_nn_3	Basmati 370 TT; Type 3 CC
Ranbir	RM201_SNP_nn_defra_1	Basmati 217 AA; other two GA
Basmati 217	c4bg_fn_SNP_4	Ranbir GG; other two GA
Basmati 370	RM166_SNP_6	Basmati 217 TT; other two TG
	Pibdom_NSb_func_5	Basmati 217 CC; other two TC
	Pibdom_NSb_func_6	Basmati 217 CC; other two TC
	c12bg_ff_SNP_2	Basmati 217 CC; other two TC
Basmati 386	Amy_W2_R_1	Basmati 386 GG; Kernal AA
Kernal	ISA_SNP_nn_1	Basmati 386 CA; Kernal CC
Super Basmati	RM248_SNP_ff_2	Super Basmati GG; other two AA
Basmati 2000	RV223_SNP_ff_3	Super Basmati TT; other two AA
Basmati 515	c9bg_ff_SNP_4	Basmati 2000 TT; other two CC
Pusa Basmati 1121	RM166_SNP_7	Bas 1509 CC; other two TT
Pusa Basmati 1059	RM42B_SNP_nn_1	Sugandha GG; other two AA
Sugandha*		
Pusa Basmati 1728	Xa21_SNP_ff_2	Pusa Basmati 1728 TT; Pusa Basmati 1441 CC
Pusa Basmati 6 (Pusa 1404)		
Pusa Basmati 1	AP5659_5_SNP_nn_4	Pusa Basmati 1 GG; Pusa Basmati 1637 TT
Pusa Basmati 1637		

\*Non-approved

approved varieties (Fig. 1) distinguishing between nine pre-defined families, of which four families contain the approved Basmati varieties. Nader et al. (2019) showed that 15 microsatellites and *fgr* could distinguish between a different four families of approved Basmati varieties, including two families of traditional and two of evolved varieties. Alternatively, pairwise comparisons might be more cost-efficient, and for this markers that distinguish specific pairs can be selected from the set of 98 markers identified for variety confirmation (Supplementary File, Sheet 2). The industry needs to consider which non-approved varieties (including ones not available for this study) might be used as adulterants for specific approved varieties, and then the appropriate KASP markers can be selected for development into quantitative tests.

### Suitability of KASP Designs

These KASP markers, designed from comparison of 120 rice genomes, gave a 90% success rate for PCR amplification in 60 test samples which had not been used for primer design.

KASP markers selected within both the *fgr* and *Waxy* (amylose) genes were found to be useful in variety discrimination. The KASP marker targeting the 8 bp deletion in allele *badh2.1* (BADH2\_PM) gave the same pattern of variation as was detected in the same samples using the

*fgr* PCR marker (Nader et al. 2019), so it can be considered an equivalent KASP marker. The KASP marker *Frg8\_Ins\_Exon13* was designed to amplify a 3 bp insertion that is known to occur in Paw San, a fragrant Pearl rice variety from Myanmar (Myint et al. 2012), but only genotypes homozygous for the corresponding deletion (-,-) were detected. The absence of the alternative fragrance allele (3 bp insertion) in any of the CoP samples indicates that this alternative allele does not account for fragrance in any of them. There was no attempt made in this study to design a KASP assay for the second most common fragrance allele (*badh2.7*) which also has not been observed in Basmati varieties (Withana et al. 2020).

Deeper investigation into the scope and utility of KASP is needed. We recommend that large-scale studies should be designed to test a larger relevant set of non-approved varieties as well as to compare more replicates of each approved sample (and more representatives of each variety from different sources). This would give a better estimate of within variety variation and could fill the gaps in the current dataset.

Tens of thousands more marker designs could be screened using higher throughput sequencing-based genotyping technology for the same price as ~300 KASP, for example, by SeqSNP developed by LGC Biosearch Technologies ([biosearchassets.blob.core.windows.net/assetsv6/seqsnp-](https://biosearchassets.blob.core.windows.net/assetsv6/seqsnp-)

targeted-genotyping-by-sequencing-alternative-routine-breeding-programs.pdf). However, high-throughput sequencing requires leaf samples for extraction of higher molecular weight DNA than can currently be obtained from polished rice. Sequencing-based genotyping with much larger numbers of markers should reveal further discrimination between varieties, but KASP technology will be more appropriate for routine testing where fewer than 100 markers would be the ideal number utilized by official control labs or industry.

The most relevant set of KASP markers must ultimately be agreed by the industry, and then blind testing of samples across multiple labs should be instigated in order to further affirm the robustness and fitness for purpose of the KASP markers. All future work should avoid CTAB extractions for DNA preparation from polished rice grains, and we recommend the use of the Phytopure DNA extraction method.

This study paves the way for the development of procedures for quantitative testing of mixtures of specific sets of varieties that are of particular interest to the industry. The variety confirmation assays could be developed into a confirmation assay to distinguish among varieties, where smaller subsets of assays could be tested in a stepwise approach to confirm variety identification of a sample.

Primers for any of the KASP markers that were successfully amplified in this study (Supplementary Data Sheet 1) can be ordered from LGC Biosearch Technologies. The target loci used for KASP designs could also be adapted for amplification with technologies such as TaqMan or quantitative PCR.

## Conclusion

This work has established a framework which uses the emerging molecular biological system (KASP markers) to add further discrimination for Basmati rice above and beyond the traditional microsatellite approach. KASP markers are shown here to be flexible and effective analytical tools with demonstrable value for food authenticity testing by industry and regulators.

**Acknowledgments** We acknowledge the valuable advice from Victor Aguilera (Defra), Hadyn Thomas (Defra), and Alex Waugh (The Rice Association). We thank Dominique Fauvin and Rhian Gwilliam of LGC Biosearch Technologies and their teams for the rice KASP marker development funded by Innovate UK (Grant numbers 131781 and 103711). We acknowledge Dr. Mark Woolfe (former Head of Authenticity Branch, Food Standards Agency) and Dr. Mohammad Arif (NIBGE, Pakistan) for the additional reference material.

**Authors' Contributions** All authors (1) made substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data; (2) drafted the work or revised it critically for important intellectual content; (3) approved the version to be published; and (4) agreed to be accountable for all aspects of the work in ensuring

that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Funding** This project was funded by Defra (Defra Project code FA0166) and Innovate UK (Grant Numbers 131781 and 103711).

**Data Availability** The full dataset of alleles is available from Defra for this publically funded project. The KASP™ primers are available from LGC Biosearch Technologies.

**Code Availability** Not Applicable.

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Ethics Approval** Not applicable.

**Consent to Participate** Not applicable.

**Consent for Publication** All authors have read and approved the final manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Amarawathi Y, Singh R, Singh AK, Singh VP, Mohapatra T, Sharma TR, Singh NK (2008) Mapping of quantitative trait loci for basmati quality traits in rice (*Oryza sativa* L.). *Mol Breed* 21:49–65
- Bligh HF (2000) Detection of adulteration of basmati rice with non-premium long-grain rice. *Int J Food Sci Technol* 35:257–265
- Bligh HFJ, Blackhall NW, Edwards KJ, McClung AM (1999) Using amplified fragment length polymorphisms and simple sequence length polymorphisms to identify cultivars of brown and white milled rice. *Crop Sci* 39:1715–1721
- Bradbury LM, Henry RJ, Jin Q, Reinke RF, Waters DL (2005) A perfect marker for fragrance genotyping in rice. *Mol Breed* 16:279–283
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
- Chen S, Yang Y, Shi W, Ji Q, He F, Zhang Z, Cheng Z, Liu X, Xu M (2008) Badh2, encoding betaine aldehyde dehydrogenase, inhibits the biosynthesis of 2-acetyl-1-pyrroline, a major component in rice fragrance. *Plant Cell* 20:1850–1861
- Du H, Yu Y, Ma Y, Gao Q, Cao Y, Chen Z, Ma B, Qi M, Li Y, Zhao X, Wang J (2017) Sequencing and de novo assembly of a near complete *indica* rice genome. *Nat Commun* 4:1–2

- FSA (2005) Protocol for the detection of certain basmati varieties using Beckman sequencer <https://documents.foodauthenticity.uk/index.php/sops?task=document.viewdoc&id=19>. Accessed 31 Aug 2020
- Kovach MJ, Calingacion MN, Fitzgerald MA, McCouch SR (2009) The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proc Natl Acad Sci* 106:14444–14449
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. *Nat Methods* 9:357–359
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079
- Mahajan G, Matloob A, Singh R, Singh VP and Chauhan BS, 2018. Basmati rice in the Indian subcontinent: strategies to boost production and quality traits. In *Advances in Agronomy* 151, 159–213. Academic Press
- Myint KM, Arikrit S, Wanchana S, Yoshihashi T, Choowongkamon K, Vanavichit A (2012) A PCR-based marker for a locus conferring the aroma in Myanmar rice (*Oryza sativa* L.). *Theor Appl Genet* 125: 887–896
- Nader W, Elsner J, Brendel T, Schubbert R (2019) The DNA fingerprint in food forensics: the basmati rice case. *Agro Food Ind Hi Tech* 30:6
- Rice Association (2017) Basmati code of practice, 2017. <http://www.riceassociation.org.uk/download/38/BasmatiCodeofPractice2017.pdf>. Accessed 31 Aug 2020
- Roy K, Mukherjee A, Maity A, Shubha K and Nag A, 2018. Protecting non-basmati indigenous aromatic rice varieties of West Bengal, India under geographical indication: a critical consideration. In *The Role of Intellectual Property Rights in Agriculture and Allied Sciences*, 295–318. Apple Academic Press
- Saghai-Marouf MA, Soliman KM, Jorgensen RA, Allard RWL (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci* 81:8014–8018
- Shao G, Tang S, Chen M, Wei X, He J, Luo J, Jiao G, Hu Y, Xie L, Hu P (2013) Haplotype variation at *Badh2*, the gene determining fragrance in rice. *Genomics* 101:157–162
- Shi W, Yang Y, Chen S, Xu M (2008) Discovery of a new fragrance allele and the development of functional markers for the breeding of fragrant rice varieties. *Mol Breed* 22:185–192
- Shi Y, Zhao G, Xu X, Li J (2014) Discovery of a new fragrance allele and development of functional markers for identifying diverse fragrant genotypes in rice. *Mol Breed* 33:701–708
- Singh VP, Singh U, Khush G (2000) The basmati rice of India. In: Singh INRK, Singh US, Khush GS (eds) *Aromatic rices*, IRRI, Manila and Oxford and IBH publishing, New Delhi
- Steele KA, Ogden R, McEwing R, Briggs H, Gorham J (2008) InDel markers distinguish Basmatris from other fragrant rice varieties. *Field Crops Res* 105:81–87
- Steele KA, Quinton-Tulloch MJ, Amgai RB, Dhakal R, Khatiwada SP, Vyas D, Heine M, Witcombe JR (2018) Accelerating public sector rice breeding with high-density KASP markers derived from whole genome sequencing of indica rice. *Mol Breed* 2018:38
- Withana WV, Kularathna R, Eshani MR, Kottearachchi NS, Kekulandra DS, Weerasena J, Steele K (2020) In silico analysis of the fragrance gene (*badh2*) in Asian rice (*Oryza sativa* L.) germplasm and validation of allele specific markers. *Plant Genet Resour.* 18:71–80
- Woolfe M, Steele K (2019) The authenticity of basmati rice—a case study DNA techniques to verify food authenticity: applications in food fraud (food chemistry, function and analysis). In: Burns M, Foster L, Walker M (eds) *Royal Society of Chemistry*

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.