Project Report No 343

July 2004

Price: £10.50

DEVELOPMENT OF A COST-EFFECTIVE METHOD FOR ASSESSING VARIETAL PURITY OF BARLEY AND MALT USING MOLECULAR GENOTYPING

By

J WHITE, N GLYNN and H JONES

NIAB, Huntingdon Road, Cambridge CB3 OLE

This is the final report of a 13-month project that started in September 2002. This work was funded by the HGCA with a grant of £31,474 (Project 2728).

TABLE OF CONTENTS

1. Abstract

This study demonstrates that it is possible to accurately estimate the level of varietal admixture in a sample of grain – specifically barley – using DNA microsatellite markers without having to analyse grains individually.

The analytical laboratory results obtained are analysed using a Bayesian statistical approach applied to attribute data from counted batches of grain. The attributes in question are 'contains no admixture' and 'contains at least one grain of admixture'.

The advantage of this approach is that it offers the potential to reduce the cost of molecular genotyping since expensive per grain extraction and amplification of DNA can be replaced by extraction from a composite bulk of grains. For example, 100 grains could be analysed as 10 batches, each of 10 grains – a reduction of 90% in some aspects of input cost.

The use of this approach is limited by two factors 1) that where polymorphism exists that off-type alleles can be detected above a background of alleles which are as expected. 2) That the number of grains examined does not exceed that number in which one off-type grain can be detected i.e. the limit of detection. Methodologies for the detection of off-type alleles and determination of limits of detection for a suite of microsatellite markers are reported.

The study develops this bulking theme and shows that by arranging samples into a two dimensional matrix an increase in precision is obtained if the attributes of both rows and columns are considered as a matrix.

The statistics are explained in detail and validated using computer modelling.

Composite data, such as these are difficult to interpret if the varietal identity of admixture is to be determined. Some rules for database construction and interrogation are presented to assist but it is found that unambiguous identification of admixture is unlikely to be possible where levels of admixture are above a few percent.

The method is applied to artificially created samples and 'blind samples'. The results of this testing phase are good with only one sample giving results which are not consistent with the known sample composition.

Some difficulty was encountered obtaining reliable amplification of some molecular markers and it is concluded that at least one alternative marker should be sought for routine use.

2. Summary

2.1. The Importance of Variety in the Marketing of Malt.

To maximise the efficiency with which industrial plant is used maltsters strive to obtain barley which will process in a predictable manner. One of the key variables which must be controlled, if predictability is to be achieved, is the variety of barley supplied. Once variety has been selected quick tests, like nitrogen content, can be used to fine tune malting conditions.

So, by receiving the variety he specifies, the maltster can increase the efficiency with which he works and simplify testing at intake. Further down the supply chain some brewers also value the predictability offered by particular varieties and may specify a sub-set of available varieties which they are prepared to buy. Varietal purity and identity is therefore critical to both grain purchase and malt sales.

2.2. Identifying Variety

Varietal identity and purity is currently managed by a process of quality assurance through farm, merchants, maltsters and exporters. Quality assurance must be supported by quality control. Currently quality control of grain and malt in trade is by visual examination of grain and by protein electrophoresis.

Neither visual examination nor protein electrophoresis give an unambiguous identification of barley variety. (For example only 18% of varieties on the UK recommended list can be unambiguously identified by protein electrophoresis.) The problem is made worse by the fact that variety identification from grain characters is not possible after barley has been malted and protein electrophoresis on malted grains is also problematic with about 20% of grains examined giving no readable information.

DNA genotyping offers the potential to discriminate between all varieties in trade. Unfortunately DNA techniques are currently expensive to deliver.

2.3. Overview of the Project

The key objective of this project was to demonstrate the feasibility of reducing the cost of estimating the level of varietal admixture in samples of barley by analysing counted batches of grain and determining whether a batch contained admixture or did not contain admixture.
To achieve this objective a series of tasks were undertaken. The tasks were as follows:

- Establish a suitable set of molecular markers (3.1).
- Establish a database of varieties genotyped using these molecular markers (3.2).
- Establish the reliable limit of detection of admixture (3.3) this value determined the largest counted batch which could be used.
- Determine the batch size which could most practically be handled.
- Fully document the statistical approach. (3.6).
- Consider the possibility of identifying admixture from composite results (3.8).
- Validate the statistical approach using computer simulation (3.11)
- Evaluate the detection of admixture in artificially created samples (3.9.2.2)
- Demonstrate the application of the counted batch approach using 'blind' samples (3.9.2.3).

This report documents three linked themes which are brought together to demonstrate the feasibility of this new approach to assessing varietal purity. The themes are statistical, biochemical and informational. Each is reported separately and then the findings from each are applied to samples in the testing phase.

In the report the maximum permitted admixture was assumed to be in the range 0-10%. Look-up tables for 5%, 7% and 10% admixture are supplied, look-up tables for any level of admixture and any batch size or matrix layout can be calculated using the mathematical models reported.

2.3.1. Statistical Approach

A method is described which allows an estimate to be made of the 'most probable admixture' in a population by reference to the attributes of counted batches of grain. The attributes in question are simply 'contains no admixture' or 'contains at least one grain of admixture'. Provided batches containing both attributes are obtained an estimate of admixture may be made together with a statement of the credible range within which that estimate lies.

The batches used can be independent of each other but there is some merit in obtaining batches by sampling grain arranged in a matrix. If both rows and columns are sampled it is possible to make a more precise estimate of admixture than from rows or columns alone.

2.3.2. Molecular Genetics

To determine the presence or absence of admixture in a batch SSR (sometimes called micro-satellite) markers were used. A database of genotypes of common varieties was established such that when a batch from a sample was tested its 'expected genotype' – as determined by a statement from the customer – could be compared with the observed genotype. Off-type (unexpected) alleles observed in any batch suggested the presence of at least on grain of admixture and so allowed the attribute of the batch to be determined.

2.3.3. Informational Analysis

The data produced from anything but the simplest mixture of varieties in a matrix becomes very complex. The problems and limitations of interpretation are described and the broad conclusion is that where levels of admixture are low some useful information about the contaminating variety may be gleaned. As the level of admixture rises it is possible to say what varieties are absent but the list

of varieties which might be present soon grows to a point where it is unlikely to give much guidance to customers. It should be remembered, however, that the primary decision relating to malting barley purchase is one of acceptance or rejection. Once a decision is made to reject the nature of the admixture becomes almost academic except in analytical examination of the supply chain to determine the source of error.

2.3.4. Summary of Findings

From this study it is clear that a counted batch sampling approach can be used to estimate, with useful accuracy, the level of admixture in a grain sample. The estimate is strictly speaking a 'most probable admixture' and the precision of the estimate should be viewed as a 'most credible range' however, for practical purposes these values can be used as 'mean' and 'confidence interval'.

To apply the counted batch approach it is only necessary to establish which batches contain at least one grain of admixture and which contain no grains of admixture. This makes an analytical method based on the detection of alleles very attractive. The presence of off-type alleles in any batch will classify that batch as containing at 'least one grain of admixture'. The use of SSRs as a tool for making this classification has been demonstrated with some success. It is noted, however, that finding molecular markers which amplify strongly and reliably is critical if the method is to be used routinely. The authors report that two markers, 'Bmac 209' and 'Bmag 211' are very suitable, others tried have proved less reliable.

A critical factor in determining the presence of the off-type allele is an objective assessment of the capillary electrophoresis trace. Considerable effort was directed at detecting off-types where stutter around particular allelic forms overlap on traces. An area normalisation approach has proved effective in detecting hidden admixture. Where alleles are well separated in terms of No. of base pairs the recognition of the off-type is much simpler although good amplification is essential if low levels admixture is to be reliably detected above baseline noise.

For the method to be fully deployed it is important that the approach is highly discriminating (between varieties). The discrimination between varieties by any combination of methods is in itself a probability function. Thus even highly discriminating methods will fail to achieve 100% discrimination within large populations of varieties. In this study the molecular marker combination

used was 100% discriminating within the variety set although failure to collect data for any one marker (such as might be the case if amplification failed) would reduce the discrimination to less than 100%.

Careful database interpretation can be used to identify admixture grains where the concentration of admixture is low. As level of admixture rises it rapidly becomes impossible to make any unambiguous statements about the genotype of the admixture, although some genotypes (varieties) can be excluded as possibilities. However, provided the varieties in the relevant population of varieties can be distinguished it should always be possible to detect and measure admixture.

It would be fair to say that this method is better at quantifying admixture than identifying varieties comprising that admixture. This may limit the technique to situations where a screen is required – such as at mill or malting intake – and where the contract specifies a variety and a maximum permitted admixture.

Using 96 grains in batches, quite fortuitously, gives a useful measurements in the range 0-10% admixture within which most grain contracts are set. The method becomes less accurate and the varietal identity of admixture much less certain as the level of admixture rises. However, this is not relevant to contracts which, once admixture is clearly 'too high' result in rejection: it is not really relevant whether the grain was rejected for 15% admixture or 25% admixture.

There is some risk to both buyer and seller from sampling errors. This is a familiar problem in grain trading which can be dealt with by setting production targets at a higher standard than the contract specifies. It is noted that buyers of large quantities of grain can afford to take an 'average view' and reduce the confidence with which any single purchase can be said to meet the contract specification. Evidence is presented that the statistical approach employed allows buyer and seller risk to be controlled in a predictable manner.

Finally, on the issue of cost. The majority of equipment available for work in molecular biology is built around the 8 x 12 micro-titre plate. This format allows a set of 96 grains to be considered as 12 batches of 8, 8 batches of 12 or indeed a matrix. In this study the matrix approach has been developed most fully although the results from matrix analysis can be re-interpreted as simple batch analysis if required. The matrix approach does allow a more accurate estimate of admixture than simple batches – using the same number of grains. Most cost effective solutions are likely to be

based around standardised equipment but batched grains reduce the variable costs of molecular biology reagents and clean-up columns which make up a considerable proportion of the total cost of molecular biology testing.

This approach is technically feasible, statistically robust and offers a reduced cost relative to a grain by grain approach. Accuracy is sufficient for decision making when purchasing grain against a maximum permitted admixture in the range 0-10%.

Some further effort is required to find two more molecular markers which offer discrimination and reliability. These are needed to replace BLYRCAB and Bmag 135 which initially appeared suitable but in practice proved problematic.

The authors also note that the approach can be applied to any batched analysis giving ‘attribute data’.

3. Technical Detail

3.1. Molecular Genetics

3.1.1. Microsatellites

Microsatellites or simple sequence repeats (SSRs) are DNA regions which are composed of single or relatively few short sequence motifs usually in tandem ('pure simple sequences'). They are thought to accumulate through DNA polymerase slippage and mispairing during replication and recombination or extension of single-strand ends. Microsatellite regions are present abundantly in genomes of most higher organisms (Tautz 1989) and have been utilised extensively in human genetics. A large numbers of microsatellites have been reported in a range of crop species and have become a popular marker system in linkage mapping projects, high-throughput genotyping as well as in studies on population genetics and gene flow. Analysis of SSR regions has a number of advantages over other marker systems such as high levels of polymorphism, locus specificity, co-dominance, reproducibility, ease of analysis through PCR and random distribution throughout the genome. Several hundred SSRs have been reported in barley (Becker and Heun, 1995; Liu et al., 1996; Struss and Plieske, 1998; Ramsay et al., 2000 and Macaulay et al., 2001) allowing the selection of loci from specific regions of the genome. The application of microsatellite based genotyping in barley has been focused on the production of genetic maps and identification of quantitative trait loci (QTLs) for use in marker-assisted selection breeding programmes. The fact that polymorphisms in microsatellite regions exist between crop varieties combined with the advantages generally associated with microsatellite genotyping also makes them useful for variety identification purposes.

3.1.2. Methods for Analysing Microsatellite Regions

Analysis of microsatellite regions is usually performed by amplification of the repeat units using the polymerase chain reaction (PCR) with primers designed for the sequences flanking the repeat area. Thus during the PCR the repeat area is amplified. The size of microsatellite regions can be small (<100 b.p.) and often the differences between alleles is also small (sometimes as low as 1 b.p.) therefore the resolving power of the detection system used must be high. Methods for the detection of SSR alleles have broadly speaking evolved around technologies used for DNA sequencing. Usually the primers used are produced with a dye which is then incorporated into the amplified PCR product as the reaction progresses. This allows the visualisation of amplified PCR products during electrophoresis usually in a capillary or a polyacrylamide gel.

3.2. Materials and Methods

3.2.1. Context

The aim of this project is to use the analysis of microsatellite loci for the detection of 'off-type' alleles in counted batches of grain. Previous work at NIAB has identified a suite of microsatellite loci and varieties of barley which show polymorphisms (Table 2). Combinations of these varieties which would show the greatest number off-types were chosen for experimentation.

3.2.2. Selection of Microsatellite Markers.

Four microsatellite primer pairs were chosen from an initial set of 16 (Table 1) which were examined against DNA from a small number of barley varieties. Primer pairs were chosen for further experimentation based on the quality of amplified products resulting.

Table 1. Microsatellite primers examined initially in this study.

Primer name
[1]Bmac0093
[1]Bmac0156
[1]Bmac0209*
[1]Bmac0399
[1]Bmag0225
[1]Bmag0013
[1]Bmag0218
[1]Bmag0211*
[1]Bmag0353
[1]Bmag0223
[1]Bmag0135*
[1]Bmag0120
[2]BMS02
[2]BMS40
[3]BLYRCAB*
[1]HvHvA1

* Used for further experimentation in this project.

[1] Ramsay L, Macaulay M, Ivanissevich S. degli, MacLean K., Cardle L., Fuller J., Edwards K. J., Tuvesson S., Morgante M., Massari A., Maestri E., Marmiroli N., Sjakste T., Ganal M., Powell W., and Waugh R. (2000).

[2] Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W. (1997)

[3]Rundle,S.J. and Zielinski,R.E. (1991)

Table 2. Alleles exhibited by barley varieties used in this study (alleles are expressed as b.p. (base pairs).

Variety	Bmac 209	Bmag 209	Bmag 135	BLYRCAB
Alexis	195	202/3	164	232
Angela	195	208/9	180	206/7
Cellar	195	202/3	166	208/9
Century	210	202/3	162	192
Chalice	195	202/3	166	203/4
Dandy	210	204/5	166	203/4
Heligan	195	208/9	166 & 162	232
Maris Otter	195	204/5	162	192
Muscat	212	204/5	180	210/1
Optic	210	202/3	180	232
Peridot	212	208/9 & 204/5	166	210/1
Riviera	212	204/5	166	210/1
Tyne	195	206/7	162	192

3.2.3. Methodology used to Extract, Amplify and Detect Microsatellites in this study.

In order to genotype multiple grains in a single analysis using SSR markers, grains or extracts must be bulked together. A matrix bulking approach was used (Figure 1) which produced 20 samples for analysis from 96 grains. Three methods were examined in order to identify the most efficient and accurate method of bulking.

Bulks were produced for analysis using DNA extracted from 96 individual grains.

Grains were bulked together prior to DNA extraction, milled to a flour and DNA extracted.

Bulks of grain lysates were produced from part-extracted samples of individual grains.

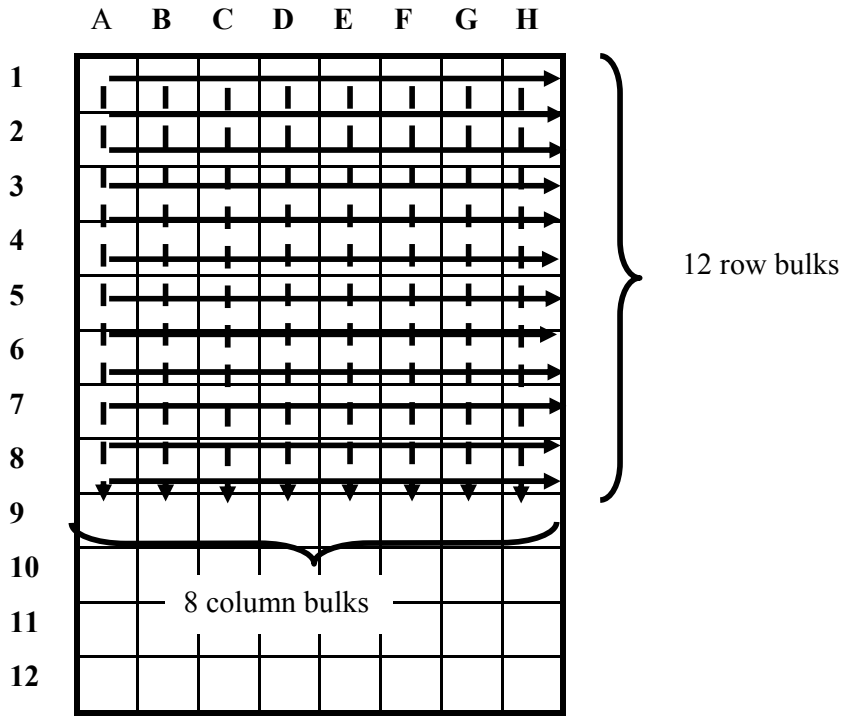


Figure 1. Matrix approach to bulking; seed, lysate or DNA bulks were made from the 12 rows and 8 columns.

3.2.3.1.Individual Grain Extraction

DNA was extracted from individual grains using DNeasy® 96 Plant extraction kit (Qiagen, UK). Individual grains (96) were part crushed using a pair of pliers. Approximately one quarter of each part crushed grain was transferred to a collection microtube in a 96 well plate. A ball bearing was added, the plate was sealed and attached to a Mixer Mill 300 at a setting of 130 oscillations per second for 30 seconds, the plate was rotated and the process repeated. The remainder of the DNA extraction procedure was carried out according to the manufacturers' recommendations, DNA was eluted in 200µl of elution buffer provided. Following DNA extraction, bulks were made from the

extracted DNA from each of the 12 rows and 8 columns by combining 5µl of eluted DNA from each well (Figure 1). Therefore for the ten micro-titre plates constructed, 200 bulk samples were generated for analysis using the four microsatellite markers.

3.2.3.2.Bulked Grain Extraction

An alternative approach to grain bulking was investigated whereby single grains were placed into each well of a 96 well micro-titre plate. Grains were removed, sliced in half and half of each grain placed in one of 12 row bulks, the remaining half was placed in one of the corresponding 8 column bulks. The 12 bulks of 8 half grains and 8 bulks of 12 half grains were ground to a flour using a mortar and pestle. A small amount was removed to an Eppendorf tube and the DNA extracted using a Plant Mini kit (Qiagen, UK) according to the manufacturers recommendations. DNA was eluted into 200µl of elution buffer provided.

3.2.3.3.Bulks of Grain Lysate.

Individual grains (96) were part crushed using a pair of pliers. Approximately one quarter of each part crushed grain was transferred to a collection microtube in a 96 well plate. A ball bearing was added, the plate was sealed and attached to a Mixer Mill 300 at a setting of 130 oscillations per second for 30 seconds, the plate was rotated and the process repeated. The resulting lysate was sub-sampled and bulked by row and by column to create the composite samples. Using these composites, the remainder of the DNA extraction procedure was carried out using Qiagen Dneasy 96 well extraction kits used according to the manufacturers' recommendations. DNA was eluted in 100µl of elution buffer provided. Following DNA extraction, bulks were made from the extracted DNA from each of the 12 rows and 8 columns by combining 5µl of eluted DNA from each well (Figure 1). Therefore for the ten micro-titre plates constructed, 200 bulk samples were generated for analysis using the four microsatellite markers.

3.3. Limit of Detection Studies

In order to utilise a bulked sample approach either from a matrix of rows and columns or from counted batches, the limit of off-type detection must be determined. The limit of detection determines the maximum number of grains that can be analysed in each bulk. In order to determine the limit of detection for the microsatellite markers used in this study, two approaches were employed. Firstly, two micro-titre plates were constructed by filling the wells with grains of variety Muscat as the main type and variety Tyne as the off-type. The level of admixture ranged from 1 grain in 12 to 4 grains in 12 in one plane and 1 grain 8 to 4 grains in 8 in the other plane. The same method was employed for a second plate though a second off-type variety (Alexis) was introduced. DNA was extracted from each individual seed and bulked along the twelve row bulks and eight column bulks for analysis (Figure 1).

Secondly, spiked DNA solutions were prepared from DNA extracted from the varieties Angora (A) and Angela (B) and Dandy (C) and Decanter (D). Ratios of extracted DNA of A:B and C:D were prepared at 1:5, 1:10, 1:20, 1:30 and 1:40. DNA was amplified using two sets of microsatellite primers (Bmac 209 and Bmag 209) as described previously and analysed by capillary electrophoresis traces were examined visually for the presence of off-type alleles. The dilutions were constructed again and the analysis repeated.

3.4. Methods used for Amplification and Marking Amplicons

The primers used were a sub-set of those described by Ramsay et al (2000). Oligonucleotide primers for the markers Bmac 209, Bmag 209, Bmag 135 and BLYCAB were labelled with one of four dyes (6-FAM, PET, TET or NED) for detection by Applied Biosystems genetic analyser. PCR reaction conditions were; 100 nM each primer, 2.5 mM $MgCl_2$, 100µM of dNTPs, 0.8 U of *Taq* Polymerase, reactions were performed in volumes of 10µl and 2µl of the eluted DNA extract was used per reaction. Thermocycling was performed in a Perkin Elmer 9600, cycling conditions were the same for all primers and consisted of an initial denaturation of 94°C for 3 min followed by 30 cycles of 94°C for 30 s, 58°C for 30 s and 72°C for 1 min followed by a final extension of 72°C for 5 min.

Following thermocycling, 2µl of each PCR reaction was removed to a fresh well of a 96 well micro-titre plate containing 10µl of Hi-Di® and 0.5µl Liz®. Plates were placed in a Applied Biosystems 3100 Genetic Analyser for analysis using Genscan® analysis software according to the manufacturers recommendations, results were visualised using Genotyper® software.

3.4.1. Detection Methods

3.4.1.1. Methods for Identifying Admixture in the Presence of Peak Stutter.

The results from Genotyper® for the main type and off-type alleles were recorded for each marker and plate combination. Where an allele for an off-type variety is close in size to that of the major type (e.g. 1 base difference) recognition of the off-type can be difficult. The situation is complicated as a characteristic of many microsatellite alleles is the production of 'stutter peaks' these peaks are smaller in height than the main peak though can be either larger or smaller in size than the main peak (Figure 2). The potential therefore exists for stutter peaks to be misidentified as off-type peaks and *vice versa*.

The ratio between peak area for the main-type peak and the stutter peaks was determined and compared between samples known to contain admixture and also for those containing no admixture. In this way, the subjectivity related to the identification of off-types over and above stutter peaks could be reduced. In Figure 3 the off-type peak is 204 bases, stutter gives a small peak at 204 bases in all samples. By using the ratio approach Columns G and C can be seen to have excess peak area relative to the 202 base peak indicating the presence of the off-type allele in the bulk.

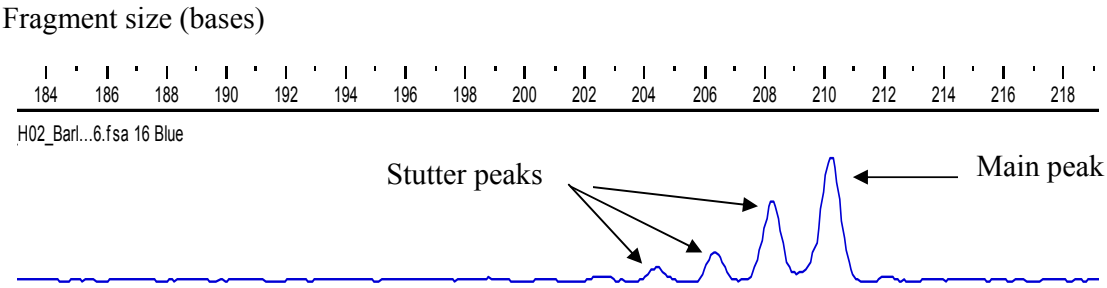


Figure 2. Example of barley microsatellite profile from ABI 3100 showing main peak and characteristic stutter peaks.

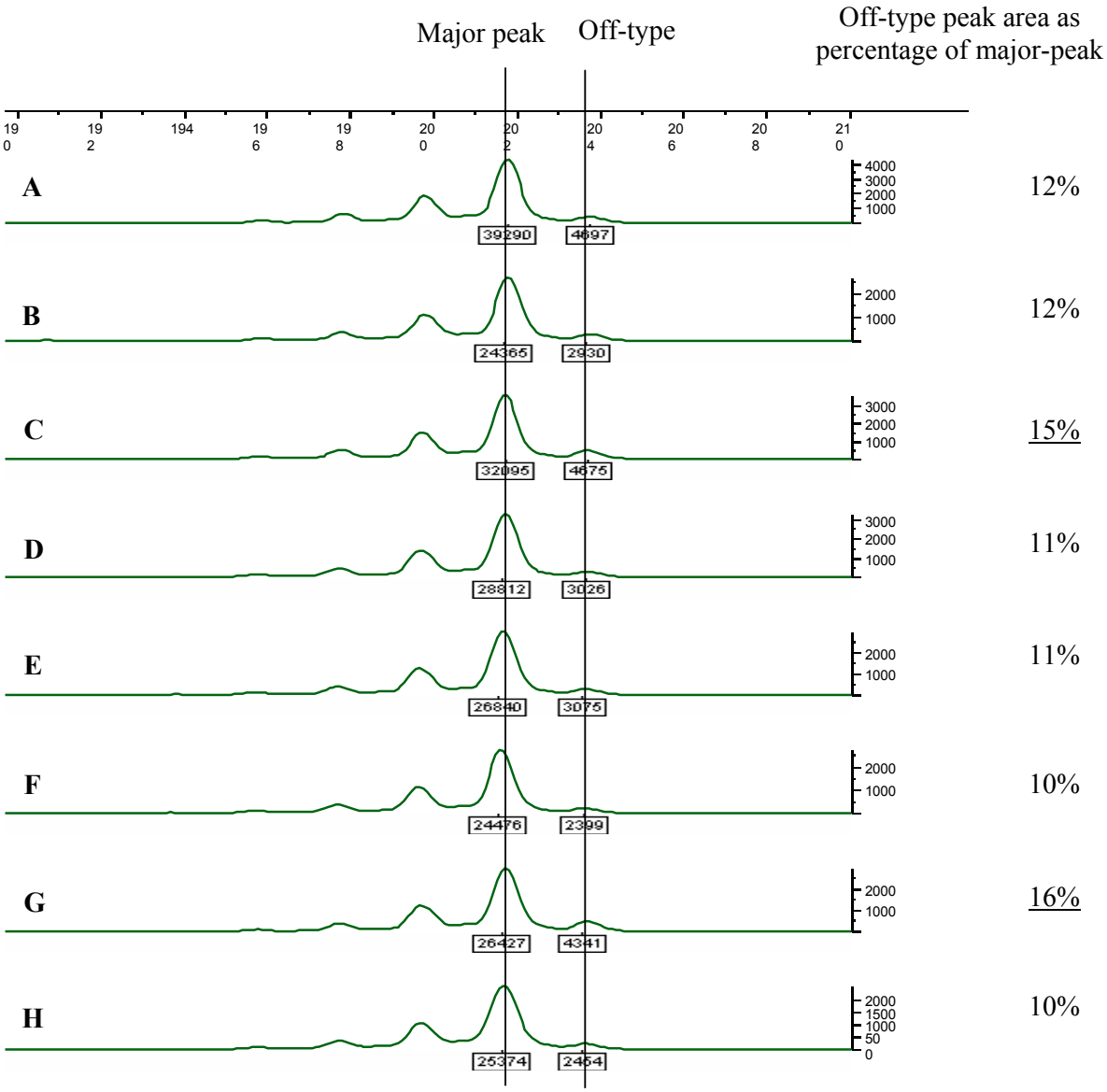


Figure 3. ABI Traces for marker Bmag 209 showing major (202 base) and off-type (204 base) peaks for column bulks A-H from artificial mixture plate AM2. The proportion of the off-type peak area as a percentage of the major peak for each bulk is given on the right. Off-type alleles are present in bulk C and G.

3.5. Construction of Test Samples

3.5.1. Construction of Known Grain Mixtures

In order to test whether off-type alleles could be accurately detected in bulk extracts ten micro-titre plates containing grains made from combinations of the varieties outlined in Table 2 were constructed (Table 3). Grains of off-type varieties were placed in known positions across the plate. In this way, when grains or grain extracts are bulked together across columns or rows, the number of off-type grains varied from 0 off-type grains in 8 or 12 to a maximum of 5 off-type grains in 8 or 4 off-type grains in 12. Some plates also contained a second off-type variety (Table 3).

Table 3. Combinations of barley varieties used for construction of ten micro-titre plates.

Plate name	**Main type variety**	**Off-type(s) variety**
CP1	Muscat	Tyne
AM2	Optic	Riviera
AM3	M. Otter	Heligan
AM4	Century	Angela
AM5	Chalice	Alexis
CP6	Muscat	Alexis/ Tyne
AM7	Optic	Alexis/ Riviera
AM8	M. Otter	Heligan/ Dandy
AM9	Century	Angela/ Peridot
AM10	Chalice	Muscat/ Alexis

The exact layout of these artificial mixtures is given in section 3.9.2.

3.5.2. Construction of Grain Mixtures for Blind Testing

Analysis of plates where the identity and position of grains is known was a useful training and method validation exercise. Unfortunately such an experiment is always subject to unintentional bias by the operator who may reinterpret results until the observed and expected outcomes coincide.

The real test, therefore, is the analysis of unknown samples. The limitation of truly unknown samples would, of course, be that the 'right answer' would also be unknown.

To achieve a meaningful test artificial mixtures 'blind samples' were created each comprising 96 grains. The varietal identity of each of the grains was known so, for example, a sample might have contained 90 grains of Halcyon and six grains of Optic. The samples were prepared in such a way that the known composition of each sample was unknown until after the experiment was complete and results had been written down.

To set up the blind test eight samples were created and labelled with a statement of the majority variety or, in one case, the statement 'malting barley' – as if the samples had been presented for testing by a third party. Each sample was labelled with a post-it numbering it 1-8. The record of the exact composition of each numbered sample was then sealed in an envelope.

The samples were given to a second person who removed the post-it notes and randomly re-labelled the samples A-H. A record was kept of the relationship between the 1-8 and the A-H codes. Thus no single person could know the true composition of any sample without opening the sealed envelope. Table 4 summarises the process.

Table 4 Identity and other data for samples used in blind testing.

Post-it No.	A-H	Statement of Contents	Composition of sample.			
			No. Grains	No. Grains	No. Grains	No. Grains
1	C	Cellar	94 Cellar	2 Optic		
2	D	Optic	91 Optic	2 Halcyon	1 Chalice	2 Cellar
3	B	Chalice	89 Chalice	4 Muscat	3 Cellar	
4	H	Alexis	76 Alexis	20 Muscat		
5	E	Alexis	86 Alexis	10 Riviera		
6	F	Malting Barley	81 Chariot	10 Optic	5 Cellar	
7	A	Muscat	92 Muscat	4 Optic		
8	G	Alexis	91 Muscat	5 Halcyon		

3.6. Statistical Considerations

3.6.1. Buyer and Seller Risk

In all trade it is imperative that the product offered for sale meets the buyers expectations. In some contexts the definition of a buyer's expectation is highly subjective. In the trade of grain and seed, however, it is often possible to minimise the subjectivity by the use of contracts which specify measurable quality indexes and the acceptable range within which the values for these indexes may lie.

In the case of malting barley two important indexes of quality are varietal identity and purity. The identity is usually stated as 'Optic', 'Cellar', etc. The purity is usually expressed as a maximum permitted level of admixture – for example - 5%.

Clearly there is error associated with any laboratory measurement, the error comes from several sources but the consequence is that a measured value of 5% is really an estimate of the true value and the estimate has a standard error associated with it.

Suppose the standard error was such that a measured value of 5% might represent a true value in the range 2.5% to 7.5%. A buyer requiring grain with not more than 5% admixture and who bought grain with a measured content of 5% admixture would run a risk that he would receive a product which was actually below the required standard. By the same token, the seller, presenting a sample of grain for analysis which had a true level of admixture of 4.99% might find that the laboratory reported a level of admixture in excess of 5%.

The preceding paragraph outlines the buyers risk – the risk that an analytical result will overstate the quality resulting in false acceptance of a product - and the sellers risk, that an analytical result will understate the quality of a product resulting in false rejection.

There are a number of things which can be done to minimise either buyer or seller risk but only an improvement to the accuracy of the analysis benefits both parties equally and even then the cost of such additional analyses may fall wholly on one party or the other.

It is not the purpose of this study to resolve the dilemma but it is important that the study does offer evidence that both buyer and seller risk can be quantified and managed. The following statistical approaches will show how admixture can be measured and analytical errors estimated. In the latter part of the study the methods are validated using computer simulation and the buyer and seller risks are highlighted.

3.6.2. Measuring Admixture

This study is aimed at finding a cost effective way to estimate the level of varietal admixture in a sample of barley or malt offered for sale using DNA genotyping. This estimate is to be used in the quality assurance of barley trading.

The general question of estimating admixture divides into two independent technical problems; statistical method and discrimination method. This division is possible because all discrimination methods will give a result which states that the sample examined is either consistent with the stated variety or inconsistent with the stated variety. The discrimination power of the method may allow further information to be gleaned about the identity of any admixture. Indeed, some techniques allow laboratories to identify the variety and admixture in a sample without the need for a statement of expected variety to be given. However, in all cases the result given is in the form of an attribute – the sample examined is either one thing or another – ‘variety A’ or ‘not variety A’.

The universality of this attribute feature means that the statistical approach described below is of much wider application than measuring varietal admixture and the techniques described are certainly not specific to DNA based data.

Whatever the nature of admixture in question there are always four critical values which must be known before any decision on acceptance or rejection can be made. These are:

a) The maximum level of admixture acceptable to the customer.
b) The statistical confidence required that the acceptable maximum level of admixture has not be exceeded.
c) The estimated level of admixture in the bulk in question.
d) The statistical confidence limits of the estimate.

Using these four values any sample can be accepted or rejected.

There is a direct relationship between the testing cost and the statistical confidence limit of the estimate. In practical terms this means that as the true level of admixture approaches the maximum acceptable it becomes increasingly expensive to demonstrate that a bulk is acceptable to the desired level of confidence. The result is that purity standards are set at a high level to minimise testing cost without incurring unacceptable risk of purchasing unusable grain. However, an inevitable consequence is that some bulks are rejected unnecessarily.

Hitherto purity of a seed bulk has been determined by individual examination of a counted number of grains (see; ISTA, 1999). The examination may be visual, chemical or biochemical. Recently molecular biology techniques have been developed can equally be applied to bulks of seed milled into a flour (see; Farid 2002). In its simplest form such analysis of the flour will give crude attribute data – either positive or negative for admixture. This section describes a method for using the attribute data from a series of such bulks to obtain estimates of (c) and (d) above. The estimates may be used both to demonstrate that a bulk meets the criteria set in (a) and (b) and to estimate the likelihood that a sample which has initially failed to satisfy (b) might prove to be acceptable if further testing was conducted.

3.6.3. Batched Seed Methods

It should be noted that the technique outlined is a Bayesian method which deals with estimates of likelihood. Thus the best estimate of the mean is the 'most probable admixture' and the statistical error of the estimate is described as a 'credible range'. In practice these differences in terminology do not effect the way the estimates can be used but they do serve to reflect the different assumptions used to derive them. Some effort has been made to validate the approach empirically by computer simulation and the results of this successful validation are reported in 3.11.

Using a laboratory test which gives either a positive or negative result (e.g. ‘wheat’ or ‘not wheat’, ‘variety A’ or ‘not variety A’, etc) creates attribute data. If this is applied to a counted batch of seed which is smaller than the analytical limit of detection (and the batch is milled to form a homogenous sample) then attribute data can be obtained for such counted batches in a similar way. In this case the result must be interpreted as, for example, no admixture in the counted batch or at least one grain of admixture in the counted batch. The data are qualitative, not quantitative.

The results from such a test may readily be used to estimate the confidence level that the maximum level of admixture has not be exceeded.

Equation 1

$$P = 1 - \left(\frac{(100 - x)}{100} \right)^{n}$$

Equation 1 derives from the binomial distribution (Wetherill and Brown, 1991) and gives the probability (P) of detecting admixture at least once in a counted batch of seeds, where the batch contains n seeds, selected at random from a sample with admixture at x%.

By substituting the maximum acceptable level of admixture for x% and the number of seeds in the batch for n the probability that the acceptable maximum has not been exceeded when the batch analysed contains no admixture can be calculated. Table 5 gives some example outcomes.

Table 5. Probability (P) of detecting at least one grain of admixture in counted batches of various sizes

	Batch size (n)							
Admixture (x%)	50	56	60	70	80	90	96	100
0.5%	0.22	0.24	0.26	0.30	0.33	0.36	0.38	0.39
1.0%	0.39	0.43	0.45	0.51	0.55	0.60	0.62	0.63
1.5%	0.53	0.57	0.60	0.65	0.70	0.74	0.77	0.78
2.0%	0.64	0.68	0.70	0.76	0.80	0.84	0.86	0.87
2.5%	0.72	0.76	0.78	0.83	0.87	0.90	0.91	0.92
3.0%	0.78	0.82	0.84	0.88	0.91	0.94	0.95	0.95
3.5%	0.83	0.86	0.88	0.92	0.94	0.96	0.97	0.97
4.0%	0.87	0.90	0.91	0.94	0.96	0.97	0.98	0.98
4.5%	0.90	0.92	0.94	0.96	0.97	0.98	0.99	0.99
5.0%	0.92	0.94	0.95	0.97	0.98	0.99	0.99	0.99
5.5%	0.94	0.96	0.97	0.98	0.99	0.99	1.00	1.00
6.0%	0.95	0.97	0.98	0.99	0.99	1.00	1.00	1.00
6.5%	0.97	0.98	0.98	0.99	1.00	1.00	1.00	1.00
7.0%	0.97	0.98	0.99	0.99	1.00	1.00	1.00	1.00
7.5%	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
8.0%	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
8.5%	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
9.0%	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00
9.5%	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10.0%	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00

The size of the batch (n) is constrained by the limit of detection. This constraint may be circumvented by using multiple batches each of a size lower than the limit of detection. Equation (1) can be adapted to take account of multiple batches.

Equation 2

$$P = 1 - \left(\left(\frac{(100 - x)}{100} \right)^{n} \right)^{b}$$

Equation (2) gives the probability (P) of detecting admixture at least once in b batches of seeds, where each batch contains n seeds selected at random from a sample with admixture at x%

Clearly the number of batches may theoretically be increased indefinitely. The consequence would be that if admixture were present some batches would be negative (i.e. no admixture detected) and some batches would be positive (i.e. at least one grain of admixture detected). The number of positive and negative batches and the size of those batches can be used to make an estimate of the level of admixture in the bulk in question and credible limits placed on the estimate.

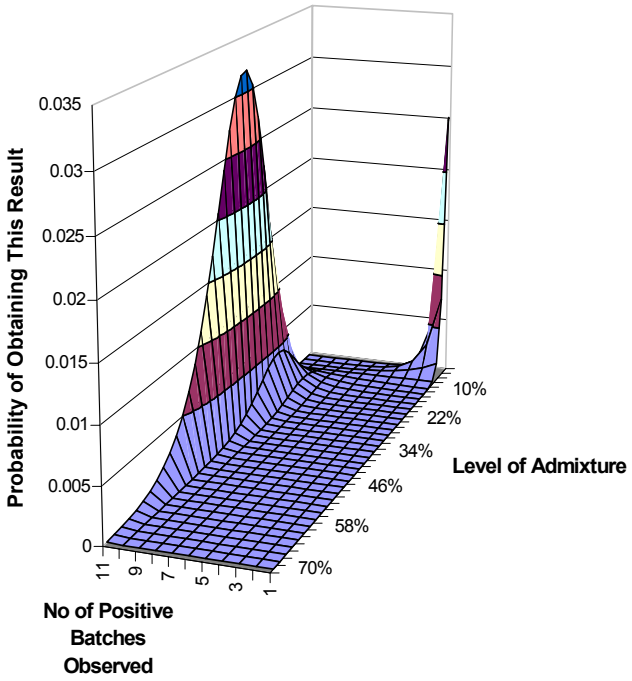
To make this estimate it is necessary to examine the likelihood of observing any specific outcome at a range of possible levels of admixture. Equation 3 gives the probability ($P(x,n,b^{+ive},b^{-ive})$) of observing a particular number of positive (b^{+ve}) and negative (b^{-ve}) batches of size (n) at a given level of admixture (x%).

Equation 3

$$P\left(x, n, b^{+ive}, b^{-ive}\right) = \left(\left(\frac{(100 - x)}{100} \right)^{n} \right)^{b^{-ive}} \bullet \left(1 - \left(\left(\frac{(100 - x)}{100} \right)^{n} \right)^{b^{+ive}} \right)$$

Figure 4 shows the probability distribution which this equation gives in an example where 12 batches each of 8 grains have been examined. To simplify the illustration the extreme cases (zero positive batches and 12 positive batches have been omitted).

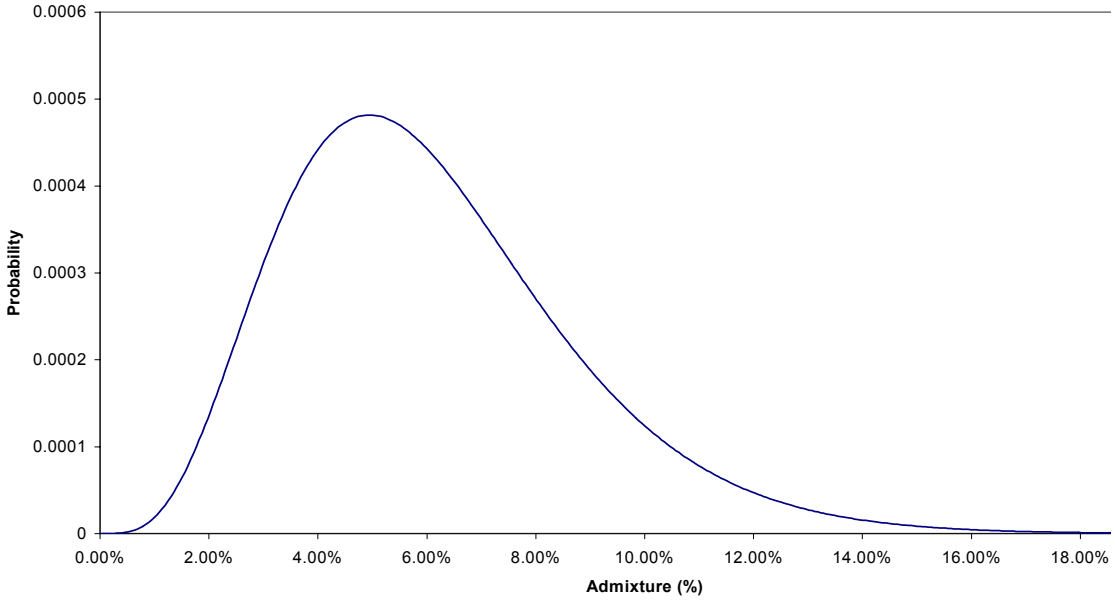
Figure 4 Probability Distribution for 12 Batches of 8 Seeds



The probability distribution in Figure 4 can be used to examine any observed outcome (in terms of positive and negative batches observed) and obtain a probability distribution for that outcome.

Figure 5 gives such a curve and is a two dimensional slice through the three-dimensional plane shown in Figure 4.

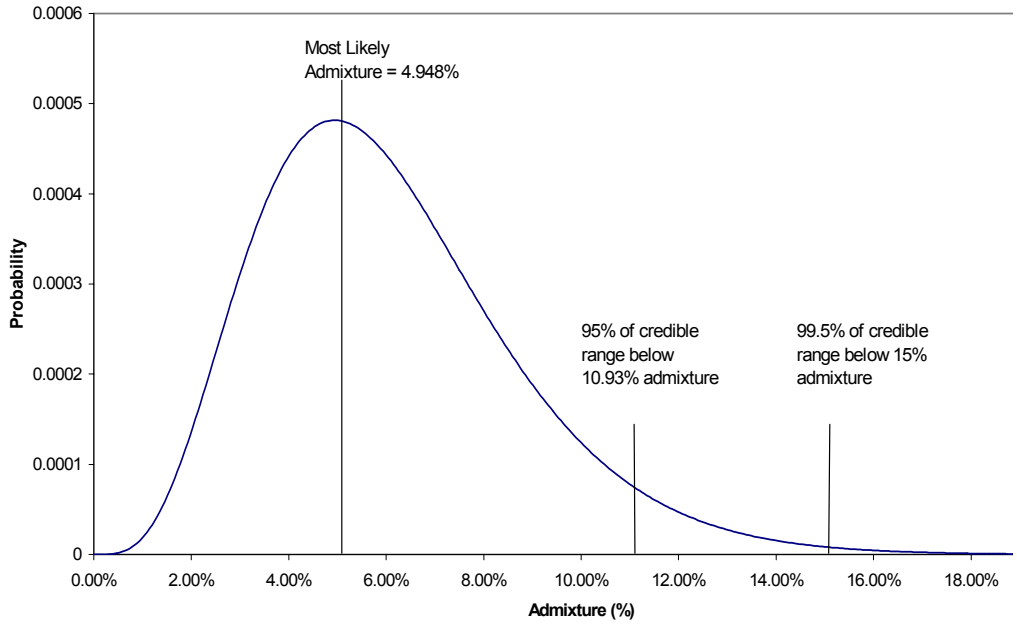
Figure 5 Probability distribution where 4 batches, each of 8 grains were found to contain admixture and a further 8 batches did not.



By integrating the area under the curve in Figure 5 it is possible to estimate the most probable admixture and the credible limits of the estimate. The most probable admixture is the value of x corresponding to the maximum probability.

Figure 6 illustrates these points. Figure 6 shows that the observed outcome (4 positive and 8 negative batches each of 8 grains) gives an estimated admixture of 4.948%. The credible range of this estimate, however, is quite wide and one could only be 95% confident that the true level of admixture was below 10.93%.

Figure 6 Probability distribution as Figure 5 with some critical statistics illustrated.



The technique illustrated here has already given both (c) and (d) above.

Before examining how the approach might be applied in the context of this study there is one further mathematical formula which must be given. There are circumstances in which the counted batch sizes chosen may be uneven, for example five batches each of eight grains and ten batches each of five grains might have been examined. Information from such experiments can be used and the equation of the probability distribution is given in Equation 4.

Equation 4

$$P = P\left(x, n, b^{+ive}, b^{-ive}\right) \bullet P\left(x, n_1, b_1^{+ive}, b_1^{-ive}\right) \bullet P\left(x, n_2, b_2^{+ive}, b_2^{-ive}\right)$$

$P(x,n,b^{+ive},b^{-ive})$ as defined in Equation 3, b_1^{+ive}, b_1^{-ive}, etc refer to the outcomes for a batch of size n_1.

3.7. Application to DNA Genotyping of Barley

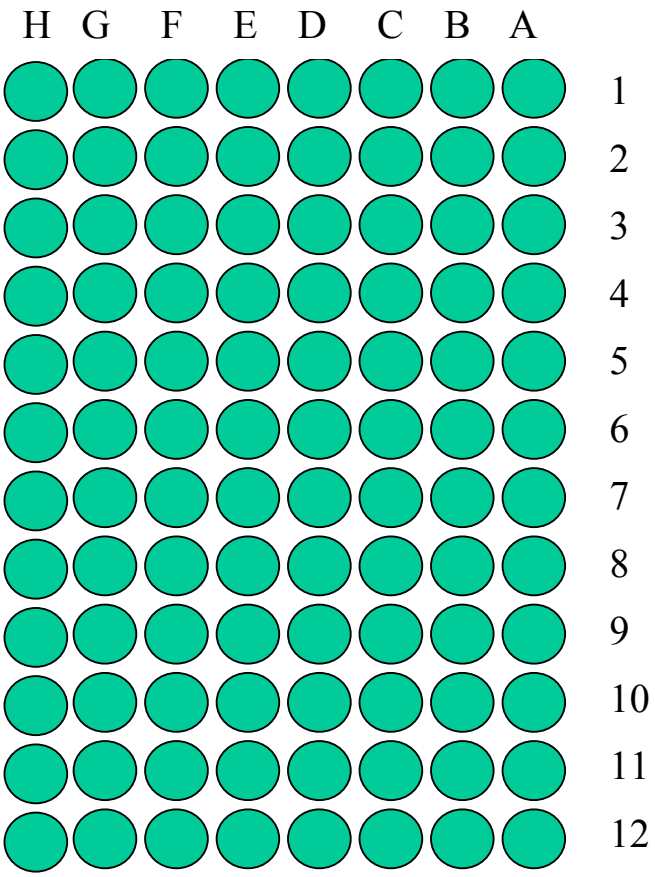
3.7.1. Equipment Considerations

Since the goal of this study is a reduction in cost of delivery it is sensible, in the first instance, to attempt to adapt apparatus which is already available rather than to design apparatus *de novo*.

In this context the 96 well micro-titre plate is an almost inevitable starting point. The reason being that all commercially available systems for processing DNA assume a micro-titre plate format using either 96 or 384 wells. Extraction, purification, amplification and detection systems all assume samples will be presented in this way. Indeed, key automation steps (like robotic pipetting and automated sample loading into electrophoresis equipment) is all formatted to this ubiquitous standard.

Figure 7 illustrates the micro-titre plate format which contains 96 wells in an 8 x 12 grid. Each position can be described by its (Column, Row) co-ordinates. Many current biochemical and molecular biological methods for varietal purity assessment use this format and, when positions occupied for control grains are taken into account, these methods are frequently implemented using sample sizes which are multiples of 8 or 16.

Figure 7 The ubiquitous micro-titre plate.



3.7.2. Using Batches

The micro-titre plate format suggests the use of batches of either eight or 12 grains. Reference to Table 5 shows that 96 grains examined without the detection of admixture would suggest a maximum level of admixture below 3% (P= 0.95).

In practical terms; 8 batches of 12 grains all found to be free of admixture would be good evidence that the sample did not contain admixture above 3%. Thus using a batching approach has reduced the number of potential analyses from 96 to 8.

If, however, some batches were found to be positive for admixture then the equations above could be applied to give an estimate of the level of admixture and the credible range of the estimate. Table 6 gives the estimates of admixture and the proportion of the credible range of estimates below 5, 7 and 10% admixture these have been chosen as representing levels of admixture which would be of some concern to a maltster. Table 6 may readily be re-calculated for any level of admixture of interest.

Table 6. The estimate of most likely admixture (%) and the proportion of the credible range below 5% admixture using 8 columns of 12 grains or 12 rows of 8 grains.

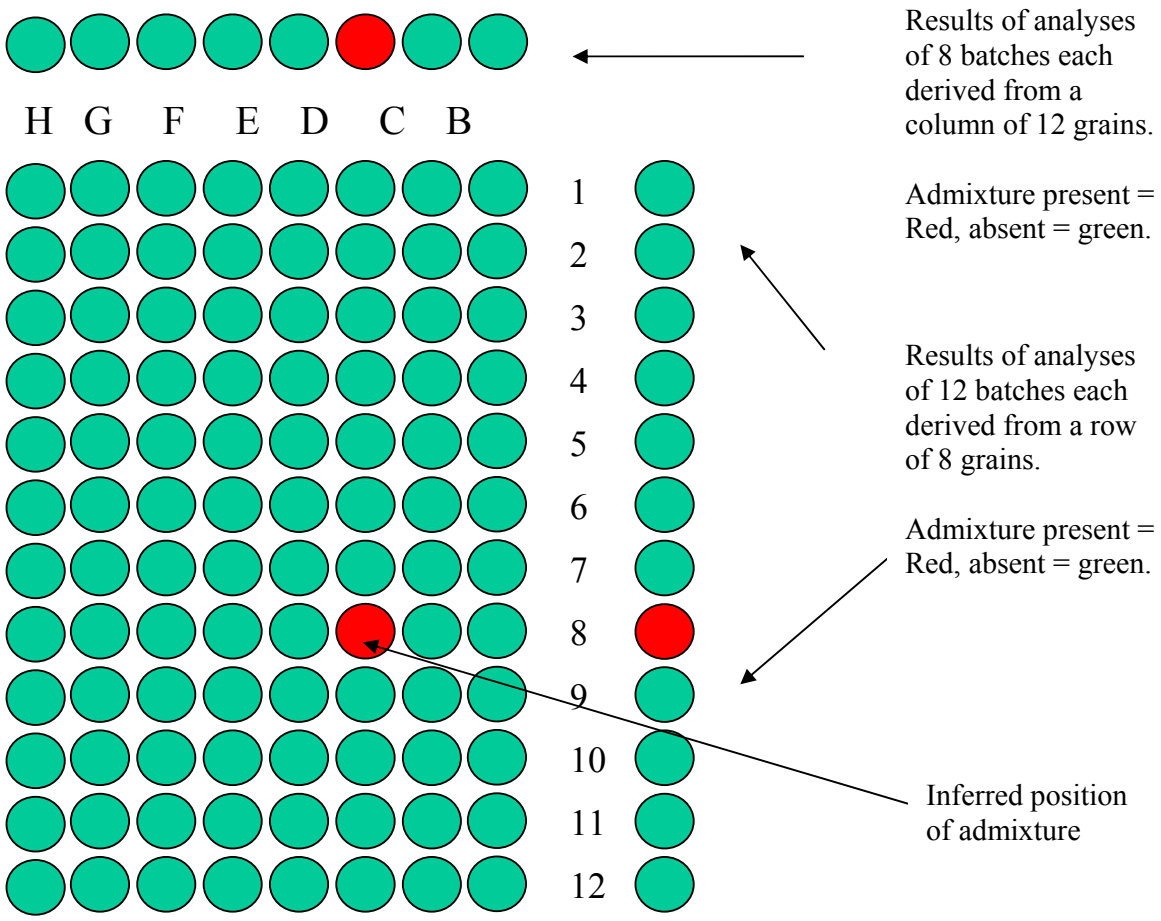
No of Positive Columns	Most Likely Level of Admixture (%)	Proportion of the Credible Range below:		
		5% Admixture	7% Admixture	10% Admixture
0	na	0.99	0.99	0.99
1	1.09	0.94	0.99	0.99
2	2.34	0.80	0.94	0.99
3	3.85	0.55	0.81	0.96
4	5.63	0.29	0.56	0.84
5	7.87	0.10	0.29	0.61
6	10.89	0.02	0.10	0.31
7	15.89	0.00	0.02	0.08

No of Positive Rows	Most Likely Level of Admixture (%)	Proportion of the Credible Range below:		
		5% Admixture	7% Admixture	10% Admixture
0	na	0.99	0.99	0.99
1	1.1	0.95	0.99	0.99
2	2.24	0.83	0.95	0.99
3	3.54	0.62	0.86	0.98
4	4.95	0.38	0.68	0.92
5	6.51	0.20	0.46	0.79
6	8.28	0.07	0.25	0.60
7	10.37	0.02	0.11	0.37
8	12.81	0.01	0.03	0.18
9	15.89	0.00	0.01	0.06
10	20.05	0.00	0.00	0.01
11	26.72	0.00	0.00	0.00

3.7.3. Exploiting the Matrix Approach

Hitherto it has been assumed that grains and batches are all independent of each other. In the micro-titre plate example above rows or columns were tested and all that could be said of a positive row or column was that it contained at least one grain of admixture. If , however, both the rows and the columns are tested then it becomes possible to infer more information about the grains tested. Figure 8 illustrates a simple case where one row and one column contain at least one positive grain. In this example it is evident that only one grain of admixture is present in the matrix at position (C,8). This is a significant finding because, although only 20 analyses have been conducted complete information on 96 grains has been obtained. This allows the re-estimation of admixture and credible interval on the basis of a batch size of 1 grain. This revised observation reduces the estimate of most likely admixture to 1.042 and the proportion of the credible range below 5% admixture increases to 0.957.

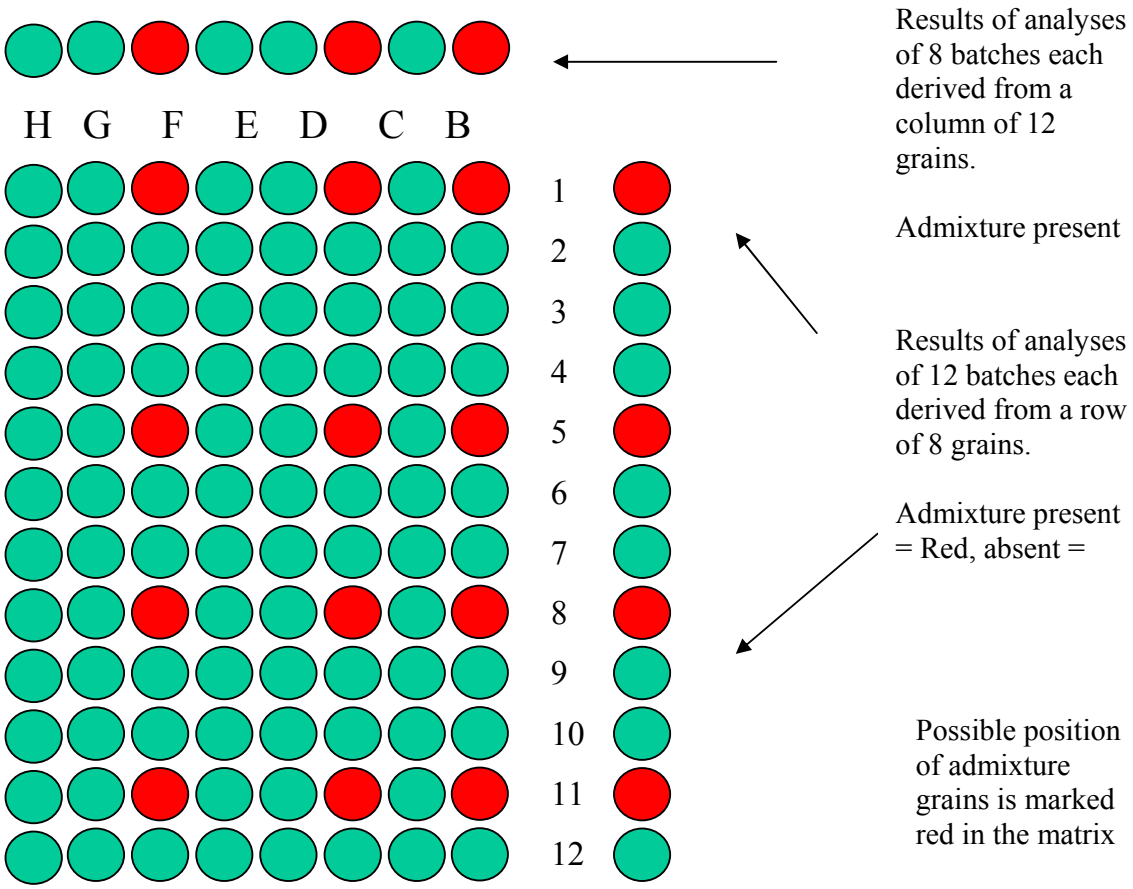
Figure 8 A simple matrix solution



Complete information may be obtained in a similar way wherever there is only one row or one column exhibiting evidence of admixture.

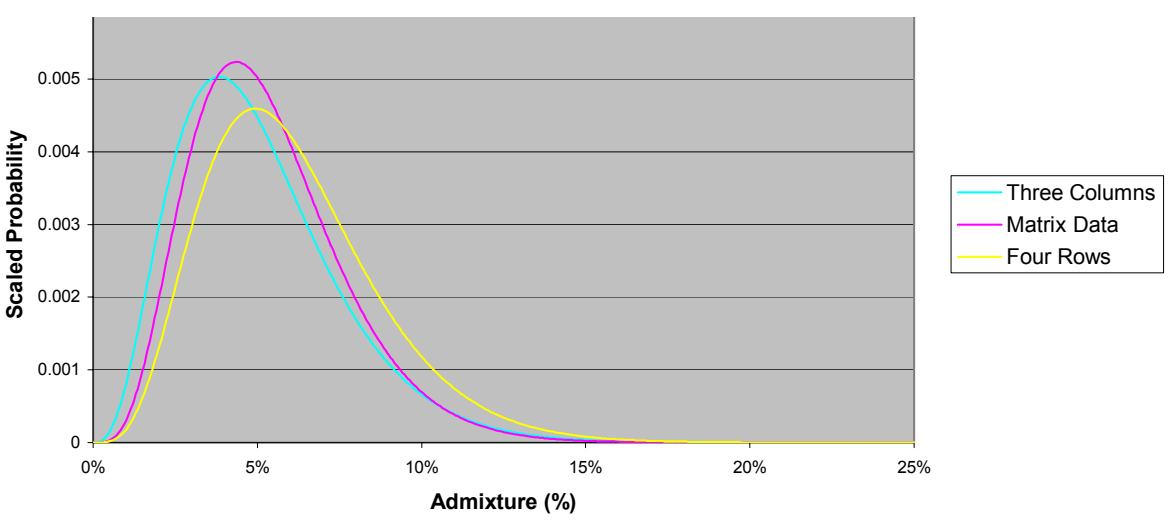
In cases where both the number of rows and columns exhibiting evidence of admixture is greater than one then there is a limited degree of ambiguity in the interpretation of the observed result. Figure 9 illustrates an ambiguous outcome. There are four rows and three columns found to contain admixture, at the intersection of each is the position of a possible grain of admixture. This apparently intractable outcome may be simplified if all the positions where admixture might be present are imagined to be grouped together into a four by three grid. Clearly of the 96 positions on the micro-titre plate 12 may contain admixture but 84 positions do not.
Looking at the ambiguous outcomes the minimum number of grains of admixture necessary to create this outcome would be four - each row must contain at least one grain of admixture. The position regarding the columns is ambiguous – all must contain at least one grain of admixture but one must contain at least two.

Figure 9 A Complex Matrix Solution



In order to apply the probability model to this outcome it is essential to interpret the data in a strictly logical way. The result we can infer from the outcome in Figure 9 is that 84 batches, each of one grain, all contain no admixture, additionally four batches each of three grains were found to contain at least one grain of admixture. This outcome may be modelled using Equation 4 and will be found to yield a most likely level of admixture of 4.375% and the proportion of the credible range below 5% admixture is 0.493. Looking at Table 6 it can be seen that without the matrix approach the estimate of admixture would have been either been 3.854 (three columns) or 4.948 (four rows). It is perhaps unsurprising that this revised estimate lies between the two previous estimates. The result obtained is also more precise. Figure 10 shows the three relevant probability distributions overlaid on one graph and it can be seen that the credible range of the matrix analysis corresponds to the higher estimate of the lower limit (the four row result) and the lower estimate of the higher limit (the three column result). In fact the estimate from the matrix analysis is also more accurate since it includes an allowance for the fact that there must be at least four grains of admixture on the matrix. The column only view of the matrix is blind to this fact.

Figure 10 Probability distributions based on a rows, columns and matrix approach.



Tables 3 and 4 give estimates of most likely admixture and proportion of the credible range below 5% admixture when the matrix approach is fully applied (see Appendix 1. for equivalent tables for 7% and 10% admixture).

Table 7. Estimates of the most likely level of admixture (%) obtained using the matrix analysis method on an 8x12 micro-titre plate.

		No. of columns observed to contain admixture							
		1	2	3	4	5	6	7	8
No. of rows observed to contain admixture	1	1.04	2.08	3.13	4.17	5.20	6.30	7.30	8.30
	2	2.08	2.08	3.18	4.30	5.40	6.50	7.60	8.70
	3	3.13	3.18	3.23	4.40	5.50	6.70	7.90	9.20
	4	4.17	4.30	4.40	4.48	5.68	6.93	8.28	9.64
	5	5.20	5.40	5.50	5.68	5.83	7.24	8.65	10.20
	6	6.30	6.50	6.70	6.93	7.24	7.65	9.17	10.89
	7	7.30	7.60	7.90	8.28	6.65	9.17	9.69	11.77
	8	8.30	8.70	9.20	9.64	10.20	10.89	11.77	12.81
	9	9.38	9.84	10.42	11.09	11.88	12.87	14.17	15.89
	10	10.42	11.02	11.72	12.60	13.70	15.10	17.03	20.05
	11	11.46	12.19	13.13	14.22	15.62	17.66	20.68	26.71
	12	12.50	13.40	14.48	15.89	17.81	20.63	25.68	-

Table 8. Estimates of the proportion of the credible range below 5% admixture obtained using the matrix analysis method on an 8x12 micro-titre plate. Shaded area would be the zone of acceptance for P=0.05.

			No. of columns observed to contain admixture							
		0	1	2	3	4	5	6	7	8
No. of rows observed to contain admixture	0	0.993	-	-	-	-	-	-	-	-
	1	-	0.96	0.87	0.72	0.53	0.35	0.21	0.11	0.05
	2	-	0.87	0.86	0.71	0.51	0.33	0.19	0.94	0.42
	3	-	0.72	0.71	0.69	0.49	0.31	0.17	0.08	0.03
	4	-	0.53	0.51	0.49	0.47	0.28	0.15	0.06	0.02
	5	-	0.35	0.33	0.31	0.28	0.26	0.13	0.05	0.02
	6	-	0.21	0.19	0.17	0.15	0.13	0.11	0.04	0.01
	7	-	0.11	0.09	0.08	0.06	0.05	0.04	0.03	0.01
	8	-	0.05	0.04	0.03	0.02	0.02	0.01	0.01	0.01
	9	-	0.02	0.02	0.01	0.01	0.01	0.00	0.00	0.00
	10	-	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	11	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	12	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

3.7.4. Example of General Rules for Applying Batch and Matrix Analysis.

Applying the above technique requires a structured approach which must start with a definition of the critical value for admixture (C_x) and the proportion of samples with admixture greater than this which one is prepared to accept (R).

Separately it is essential to know the maximum batch size within which one grain of admixture can be reliably detected. The batch size used should not exceed this value.

Proceed as follows:

1. Using Equation 1 determine the minimum number of grains which must be tested such that if no admixture is found then any admixture present is below C_x with P>=R.

2. If the number of grains in (1) is below a logical unit – say 96 – then consider analysing more grains to give greater flexibility in the event that admixture is detected.

3. Divide the chosen number of grains into counted batches of size not greater than that determined by the limit of detection. If the matrix analysis method is to be used structure the test to create rows and columns.

4. Analyse each batch (or row and column) and record whether admixture is detected or not.

5. If no admixture is detected then **ACCEPT THE GRAIN.**

6. If admixture is detected in all batches use equation 3 to estimate the effect of analysing one more batch (matrix analysis is not meaningful) and finding it to be negative. If the outcome shows the grain would be unacceptable then **REJECT THE GRAIN** otherwise go to 8.

7. If the cost of one additional analysis is acceptable analyse one more batch (matrix analysis is not meaningful). If the result is positive for admixture go to 6. If the result is negative for admixture **ACCEPT THE GRAIN.**

8. If some batches are positive and some are negative use Equation 3 to estimate the proportion of the credible range which would contain admixture greater than Cx. Refine this estimate using matrix analysis if possible. If the proportion is less than R then **ACCEPT THE GRAIN.**

9. If the result does not show the sample to be acceptable use Equation 3 to estimate the effect of analysing one additional batch and finding it to be free of admixture. If the sample would still be unacceptable then **REJECT THE GRAIN.**

10. If the cost of one additional analysis is acceptable analyse one more batch (it will not be practical to add this result to a matrix so use Equation 4 to incorporate the additional data). If the result is positive for admixture go to 11. If the result is negative for admixture **ACCEPT THE GRAIN.**

11. **REJECT THE GRAIN.**

3.8. Use of Variety Database in Analysis of Results

3.8.1. Varietal Identification

NIAB maintains genotyping databases for a wide range of crop species. The information held covers a full suite of techniques including phenotypic markers, biochemical markers and a number of DNA assays including RAPD, SSR, SNP and AFLP data. An SSR database is under development, which may, in the future, underpin variety, testing services. Varieties are selected for inclusion in this database from all available varieties within the NIAB reference collection on the basis of their recent status as commercially significant by virtue of acreage sown, importance to premium markets or significance within a niche market. Varieties included in the database are tabulated below. The subset of varieties examined in the course of this study is indicated in bold typeface. The varieties given in parentheses do not have complete data for the markers used in this study and are not included in the statistical treatment.

Alexis, Angela, Angora, (Antonia), Avenue, Baton, (Brise), **Cellar**, **Century, Chalice, Chariot**, Chime, Cooper, County, **Dandy**, (Decanter), (Delibes), Derkado, (Epic), Extract, (Fanfare), Felice, Ferment, Fighter, (Flute), (Gaelic), (Gleam), Haka, **Halcyon**, (Hanna), (Hart), **Heligan**, Intro, Jewel, Landlord, (Leonie), (Linnet), Livet, **Maris Otter**, (Manitou), Maresi, (Melanie), **Muscat**, Opal, **Optic**, (Pacific), (Pastoral), (Pearl), **Peridot**, Pipkin, Prisma, (Puffin), (Regina), Rifle, **Riviera**, Saloon, (Sevilla), Siberia, Spice, (Spirit), (Sprite), Static, Sumo, (Sunrise), Tankard, Tavern, **Tyne**, (Vanessa), (Vertige)

The ability to tell varieties, one from another, can be measured for any system of genotyping markers. In a simple system of two varieties, where the varieties can be told apart by presence or absence of a characteristic, the discrimination is absolute. If the number of varieties increases to three, one variety will differ from the other two. The varieties can be compared, one to another, in three combinations where two comparisons will discriminate and one will not; in this instance the discrimination could be expressed as 67%. In the general case discrimination (%D) may be calculated using Equation 5

Equation 5

$$\%D = \left(\frac{Number\_of\_distinct\_pairs}{Number\_of\_pairs} \right) \times 100$$

This approach can be adopted to calculate the discrimination for a single marker when applied to a population and to markers used in combination. When a number of markers are used in combination a prediction of the combined discrimination power may be calculated using Equation 6.

Equation 6

$$ {}^{n}_{1}D\% = 100 - \left(\frac{(100 - \%D_1) \times (100 - \%D_2) \times .... \times (100 - \%D_n)}{100^{n-1}} \right) $$

Should the predicted discrimination rate be significantly greater than that calculated from observed data the conclusion that one or more of the markers used are subject to underlying linkage must be considered.

In addition to the discrimination rate, the polymorphism information content (PIC) for each marker may be calculated using Equation 7.

Equation 7

$$ PIC = 1 - \sum_i p_i^2 - \sum_i \sum_{j=i+1} p_i^2 p_j^2 $$

where p is the frequency of each observed state of a marker.

For the forty-three varieties examined in the study with four markers the discrimination and PIC data are reported in Table 9.

Marker	**Bmag 0135**	**Bmac 209**	**Mmag 0211**	**BLYR CAB**	**Combined (D%)**	**Predicted (D%)**
Discrimination (D%)	59%	48%	65%	86%	99%	99%
PIC	0.536	0.447	0.592	0.825		

Table 9 Discrimination power of markers used in this study.

3.8.2. Identifying an Unknown Grain

In any genotyping system the ultimate ambition is to collect sufficient data for a grain or collection of grains and assign them each to a unique variety from all varieties in common knowledge. In some circumstances the aim of genotyping grains might be to exclude the possibility of it belonging to a variety or group of varieties. The probability of being able to assign a grain to a unique variety will be proportional to the number of independent genotypic data points acquired for the grain and inversely proportional to the number of varieties in common knowledge. Similarly, the average number of varieties in a group (Groupsize) can be calculated using Equation 8.

Equation 8

$$Groupsize = PopSize\left(\frac{(100 - \%D)}{100}\right)$$

Where Popsize is the number of varieties examined and %D is the discrimination power of the genotyping data

In practice the cost of acquiring sufficient data for each grain to provide an unique designation and maintaining a reference collection of all known varieties is prohibitive. In reality it is sufficient to maintain a reference collection of those varieties that are actively traded in any market. This list of actively traded varieties must be subject to periodic reviews to ensure it reflects additions to or deletions from the portfolio of varieties in trade.

Where genotypic data are sufficient to assign a grain to a unique variety from among the 'market list' an unequivocal identification is made. Where genotypic data are only sufficient to assign a grain to a group of varieties an equivocal identification is made.

Table 10. Variety groups for the combination of varieties and makers used in this study

Number of varieties in group	Number of groups	Varieties
1	30	Alexis, Angela, Angora, Avenue, Baton, Cellar, Century, Chariot, Dandy, Derkado, Felice, Ferment, Fighter, Haka, Halcyon, Jewel, Livet, Maresi, Muscat, Optic, Peridot, Pipkin, Prisma, Riviera, Siberia, Spice, Sumo, Tankard, Tavern, Tyne
2	3	Maris Otter, Opal, Heligan & Intro, Extract & Saloon
3	1	Landlord, Chime & Chalice
4	1	Cooper, County, Rifle & Static

As Table 10 shows, using the four markers described thirty out of forty-three varieties can be identified unequivocally. The remaining thirteen varieties are assigned to groups of varieties giving an equivocal identification. To arrive at an unequivocal identification for these three varieties, additional markers, selected intelligently on a case by case basis, would be needed.

For any collection of grains there may be two outcomes from collecting genotypic character data. If the grains in the collection are similar to each other then the collection will exhibit only one genotype. Alternatively, if the grains in the collection are not similar to each other the collection will exhibit more than one genotype.

When ninety-six grains taken from a bulk for analysis and distributed randomly into the wells of a microtitre plate and the extracted DNA pooled systematically as described in section 3.6 above,

variety identification proceeds by examining collections of grains by rows and by columns. The problem of identifying varietal admixture can be separated into two distinct processes.

Firstly, the major varietal component must be identified. Identification of the major component requires that some of the pooled extracts exhibit only one genotype. As can be seen from Table 7, where the major component dominates the sample there is a high probability of finding pooled extracts from either rows, columns or both that will exhibit only one genotype. Where no variety predominates, at a level of 80% or greater, the power of counted batches begins to break down because of the frequent occurrence of outcomes where all batches contain at least one grain of admixture. This effect is clearly seen in Table 25 (below) where computer modelling has been used to validate the statistical method.

Secondly, the identity of any admixture should, where possible, be determined.

3.8.3. Fine-tuning the Laboratory Process.

Where the laboratory receives a sample with a variety description the process is simplified to that of confirming the major variety. In this instance the pooled DNA would be examined using a marker set selected to uniquely identify the variety from all others in active trade in the local market. It is conceivable that this could be achieved with a single marker though it is more likely to require an overlaying of data from several markers. For example, Fighter and Alexis can be uniquely identified from the set of forty-three varieties given above with the BlyCab marker alone. Conversely a unique identification of Chalice would require additional markers to those described in this study.

Where the laboratory receives a sample with no variety description the process requires an initial screen with a number of markers with the intention of assigning the main component to a unique variety or assignment to a small group of varieties. Where the initial screen assigns the sample to a group of varieties further markers can be selected intelligently to provide the required discrimination.

Where the laboratory receives a sample with a generic description such as ‘spring malting barley’ where a mix of varieties may be allowed as the ‘main component’ the number of counted batches examined might need to be increased (or the batch size reduced) until sufficient collections of grains

exhibiting only one genotype were observed. These single genotype collections would normally be required to give an unequivocal identification of each variety allowed by the generic description. The alternative approach of increasing the number of makers used would require iterative interpretations of the data for each well in the matrix until a 'most likely collection of varieties in the generic mix' could be arrived at. In either instance it may be more cost effective to analyse individual grains rather than treating pooled row and column extracts as counted batches.

Once the major component has been identified it is relatively straight-forward to record the number pooled extracts for rows and columns as exhibiting one genotype or exhibiting more than one genotype. Using Table 7 as a look-up table the 'most likely level of admixture (%)' is given and this can be reported alongside the variety of the major component as identified or confirmed.

3.8.4. Level of Discrimination Required.

As previously stated, the problem of quantifying varietal admixture can be separated into two distinct processes. Where admixture is detected above maximum level of admixture acceptable to the customer the decision for the buyer may be quite simple. The seller, however, may see value in identifying the 'off-type' varieties in an attempt to investigate a failure to maintain the value of the lot. Identification of the aberrant grains may be considered as the second process within quantifying varietal admixture.

The issues surrounding detection of 'off-type' alleles is discussed extensively above and illustrated in Figure 3, Figure 14, Figure 15, Figure 16 and Figure 17

Figure 17. Detection of off-type microsatellite allele using capillary electrophoresis system. (product size in bases given across the top). Admixture is between Muscat (majority) and Tyne.

. For any microtitre plate, the extracts, pooled by row and column will, when investigated have one genotype, two genotypes or several genotypes. Once the major variety genotype has been established by reference to the single genotype pooled extracts the remaining data can be considered. Where an extract appears to have a genotype in addition to the main variety, the second variety will be defined by an additional allele at least one locus. There is no requirement for the second genotype to differ from the main variety at all loci. For example Peridot and Riviera's genotypes differ only in marker 211, their character states for 135, 209 and BlyCab being identical; they are, however, differing genotypes. Where it is apparent that only one 'off-type' variety is present on a plate their likely number and position on the microtitre plate can be readily interpreted as illustrated in Figure 8 and Figure 9.

However, as the number of 'off-type' varieties within the collection of aberrant grains increases the greater the likelihood of pooled extracts exhibiting several genotypes. In this circumstance the possibility that complementary patterns among the genotypes will obscure varietal identities cannot be excluded.

3.8.5. Procedure for Identification of Admixture.

Record the observed alleles for all pooled extracts. The alleles should be recorded as present or absent irrespective of the size or intensity of the signal for the amplified fragment.

Identify all pooled extracts that exhibit only one genotype, that is, one allele only at each locus. The single genotype pooled extracts will normally share the same pattern. Should more than one pattern be observed this would be an indication of admixture at a significant level. The majority single genotype would then be considered as representing the major varietal component in the sample.

Estimate the most likely level of admixture by reference to Table 6.

Identify all pooled extracts that exhibit two genotypes. The second genotype in the extracts will be that which has alleles additional to those of the majority genotype at one or more locus. The data from these pooled extracts will be used to identify the minor components in the mixture. There may be more than one minor component in the sample. For the markers where no additional alleles are

observed for these pooled extracts the minor components will be assumed to have the same alleles as the major component.

Identify all pooled extracts that exhibit more than two genotypes. Where all observed alleles can be accounted for by the majority component and the additional bands contributed by the minority components previously identified there will be a tendency to assume that no further minority components have been identified; however, this assumption may not be justified.

3.8.6. Worked Example of Varietal Identification from Batched Samples.

Suppose a sample contained four varieties; A, B, C and D and four markers were used for discrimination, each having three alleles (Green Blue and Red).

Variety	Marker 1	Marker 2	Marker 3	Marker 4
A	Green	Green	Green	Green
B	Blue	Green	Blue	Blue
C	Red	Blue	Green	Red
D	Green	Blue	Green	Red

Figure 11. Illustrative variety set and colour coded genotypes.

In the above, each variety is uniquely identified. Consider the situation where these varieties are in a mixture and are examined in a 5 x 5 matrix. If variety A is present as the major component then an outcome might be as described in Figure 12.

Figure 12 shows a possible outcome from this scenario. The true genotype for each position on the matrix and for each marker is given. If one then considers how the data might be interpreted if the genotype of each row and column is all that is known then ambiguities begin to appear (Figure 13). There is only one grain of off-type which can be unambiguously genotyped (position B,2; see annotation), it is clearly variety B. All other off-types have some ambiguity in their genotype at one or more markers. If one assumes all possible permutations of alleles are possible then the ambiguity can suggest known or unknown varieties (see **Table 11**). The confidence with which any varietal identification could be made would be a matter of judgement based on the varieties in the database, the varieties likely to be present and the provenance of the sample. What may useful is the fact that some varieties can be excluded on the basis of the presence of unambiguous off-type alleles for some markers.

Position	Marker 1	Marker 2	Marker 3	Marker 4	Possible Variety	Impossible Variety
B, 2	Blue	Green	Blue	Blue	B	A,C,D
D, 1	Red	Blue	Green	Red	C	A,B,D
D, 1	Red	Green	Green	Red	unknown	A,B,D
D, 1	Red	Blue	Green	Green	unknown	A,B,D
D, 1	Red	Green	Green	Green	unknown	A,B,D

Table 11. Possible genotypes for some positions in Figure 13

It is clear from this example that detection of admixture is reliable, its identification, except at low concentrations, is not.

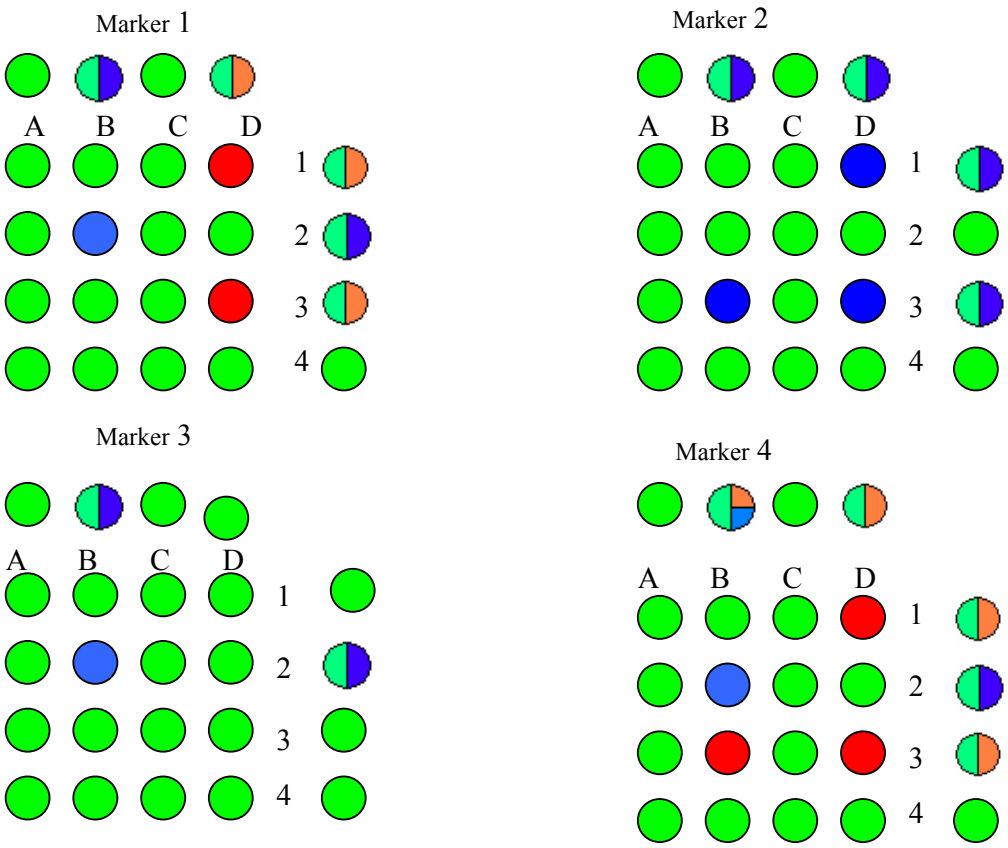


Figure 12. Actual position of off-type alleles and the result of analysis of rows and columns.

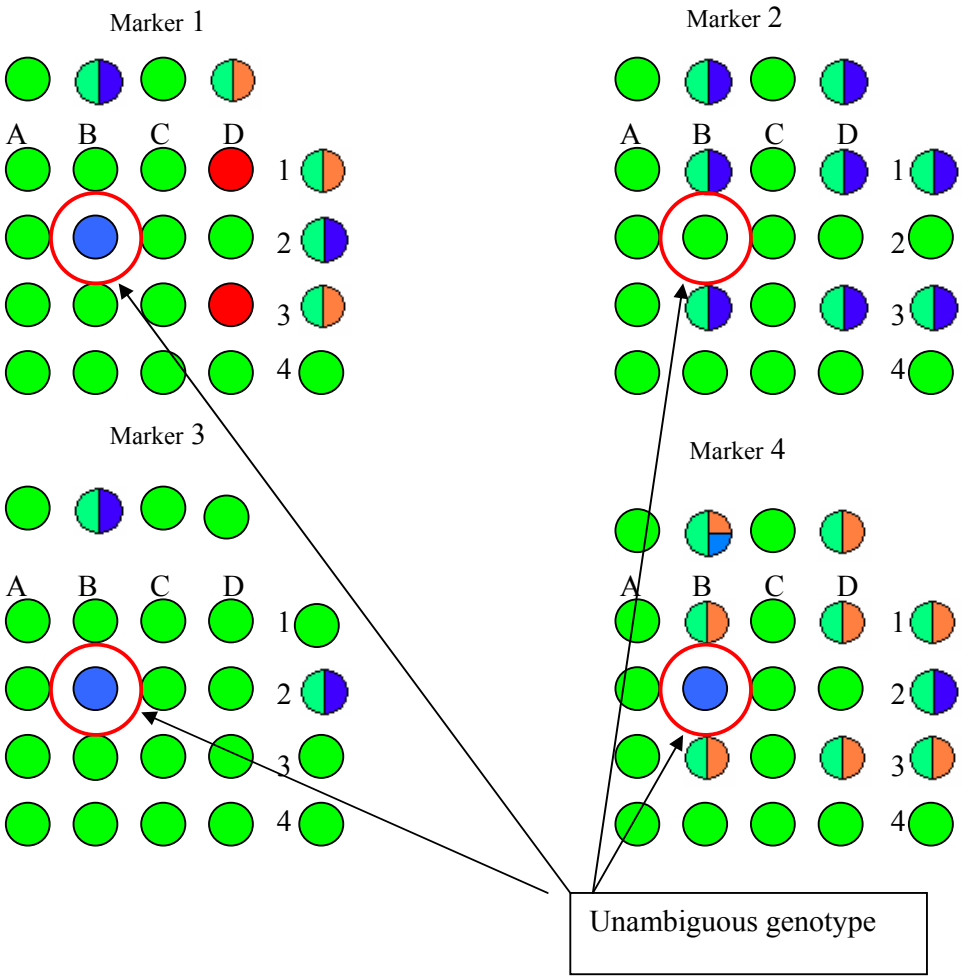


Figure 13. Ambiguous interpretation of genotypes at some positions in the matrix.

3.9. Results

3.9.1. Limit of Detection Studies

Examples of the ABI 3100 traces obtained for the limit of detection studies using markers Bmac 209 and Bmag 209 are given in Figure 14 and Figure 15 an off-type allele is visible at dilutions of 1:5, 1:10, 1:20 and 1:30 and is 210 bases in size for marker Bmac 209 and 208 bases in size for marker Bmag 209. Upon closer inspection, an off-type allele is also visible for the 1:40 dilution for both markers. In order for a matrix bulking approach to be employed in a micro-titre plate format the limit if detection must be at least 1:12 from the results obtained, this would appear to be within the detectable limit for markers Bmac 209 and Bmag 209.

Further dilution studies for admixture at 1part in 12 and greater, using grains rather than DNA, were conducted to check that the off-type grain would be detectable in routine use. Further dilution series are reported in Figure 16 and Figure 17.

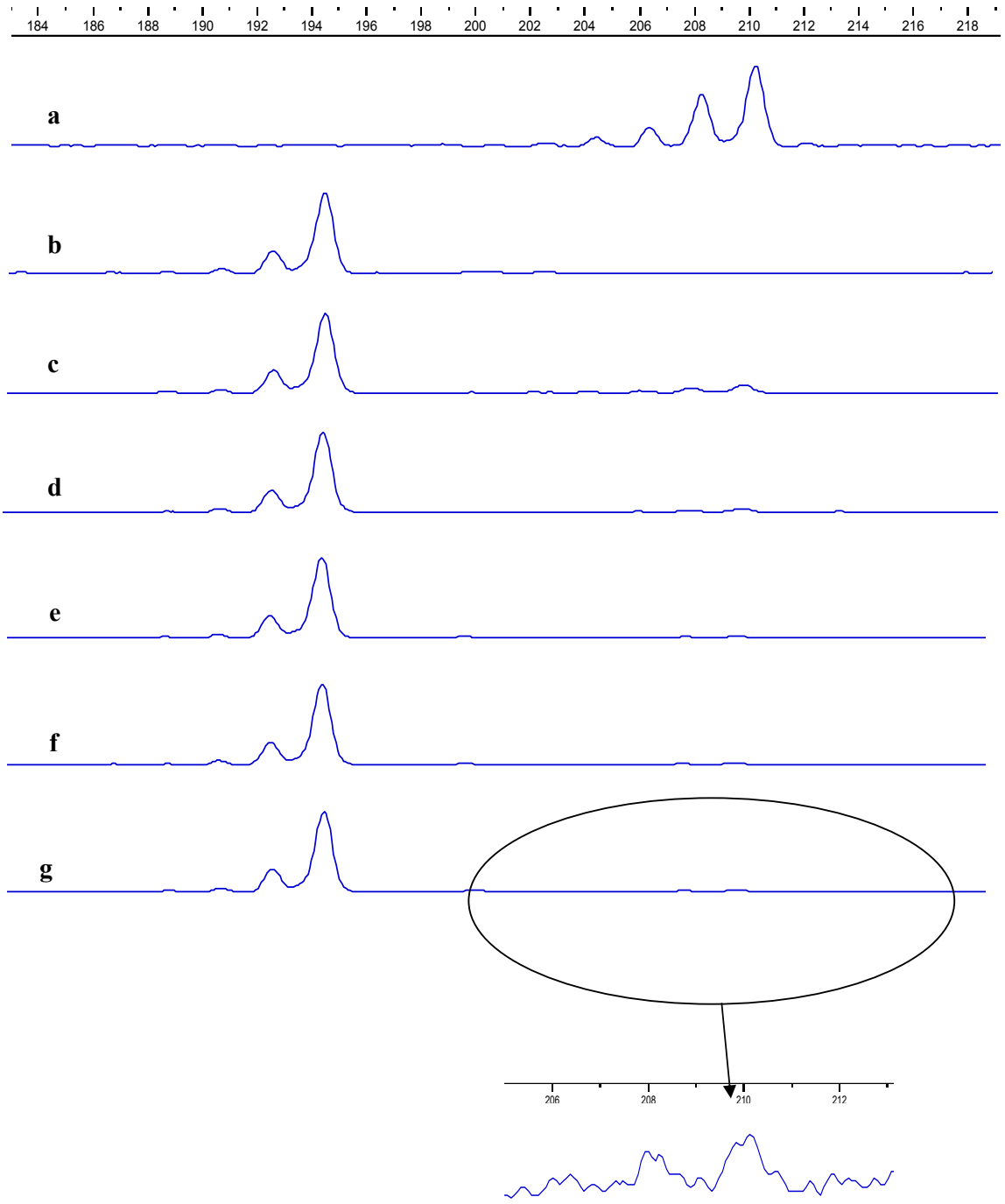


Figure 14. Detection of off-type microsatellite allele using capillary electrophoresis system. (product size in bases given across the top). a = Decanter, b = Dandy, c = 1:5 Dandy : Decanter DNA, d = 1:10, e = 1:20, f = 1:30, g = 1:40.

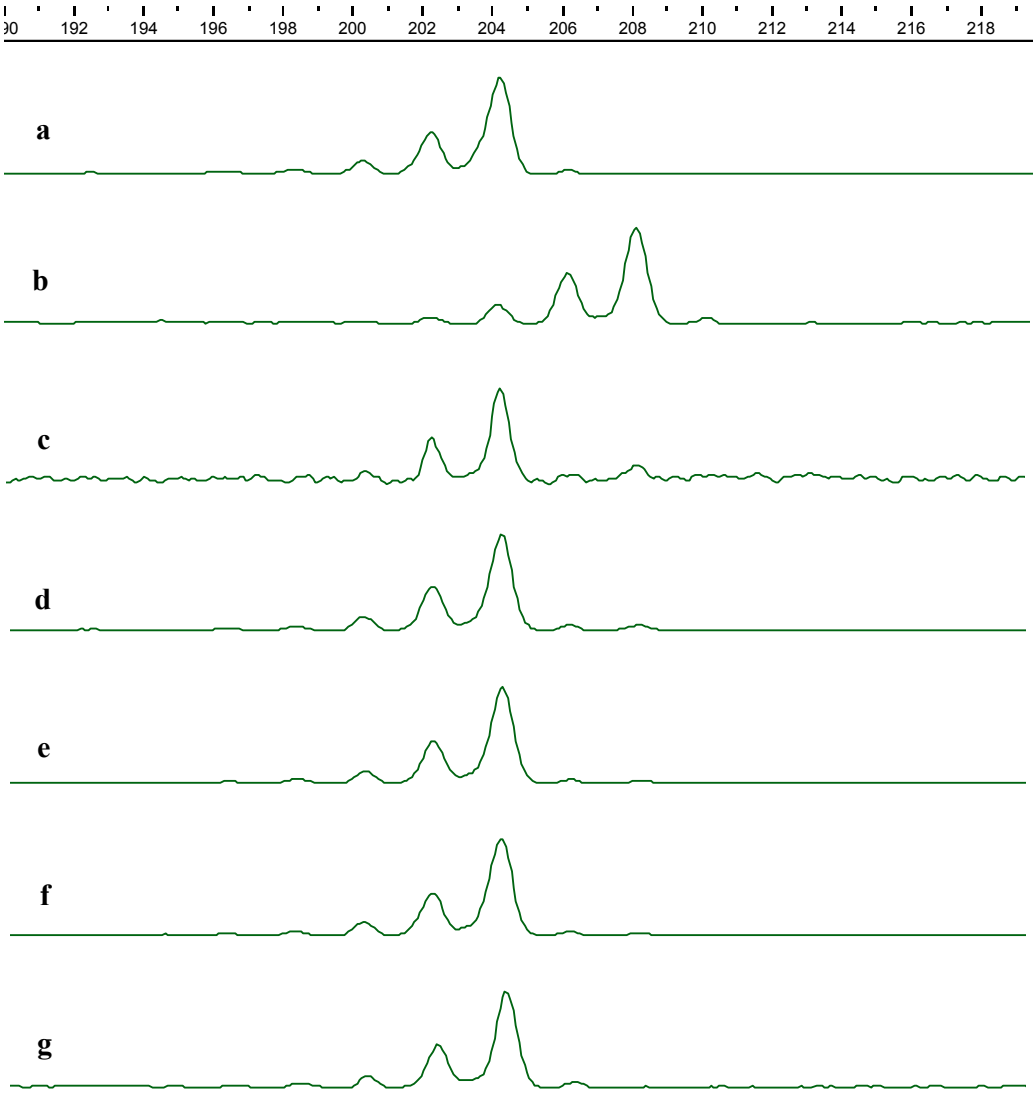


Figure 15. Detection of off-type microsatellite allele using capillary electrophoresis system. (product size in bases given across the top). a = Angora, b = Angela, c = 1:5 Angela : Angora DNA, d = 1:10, e = 1:20, f = 1:30, g = 1:40.

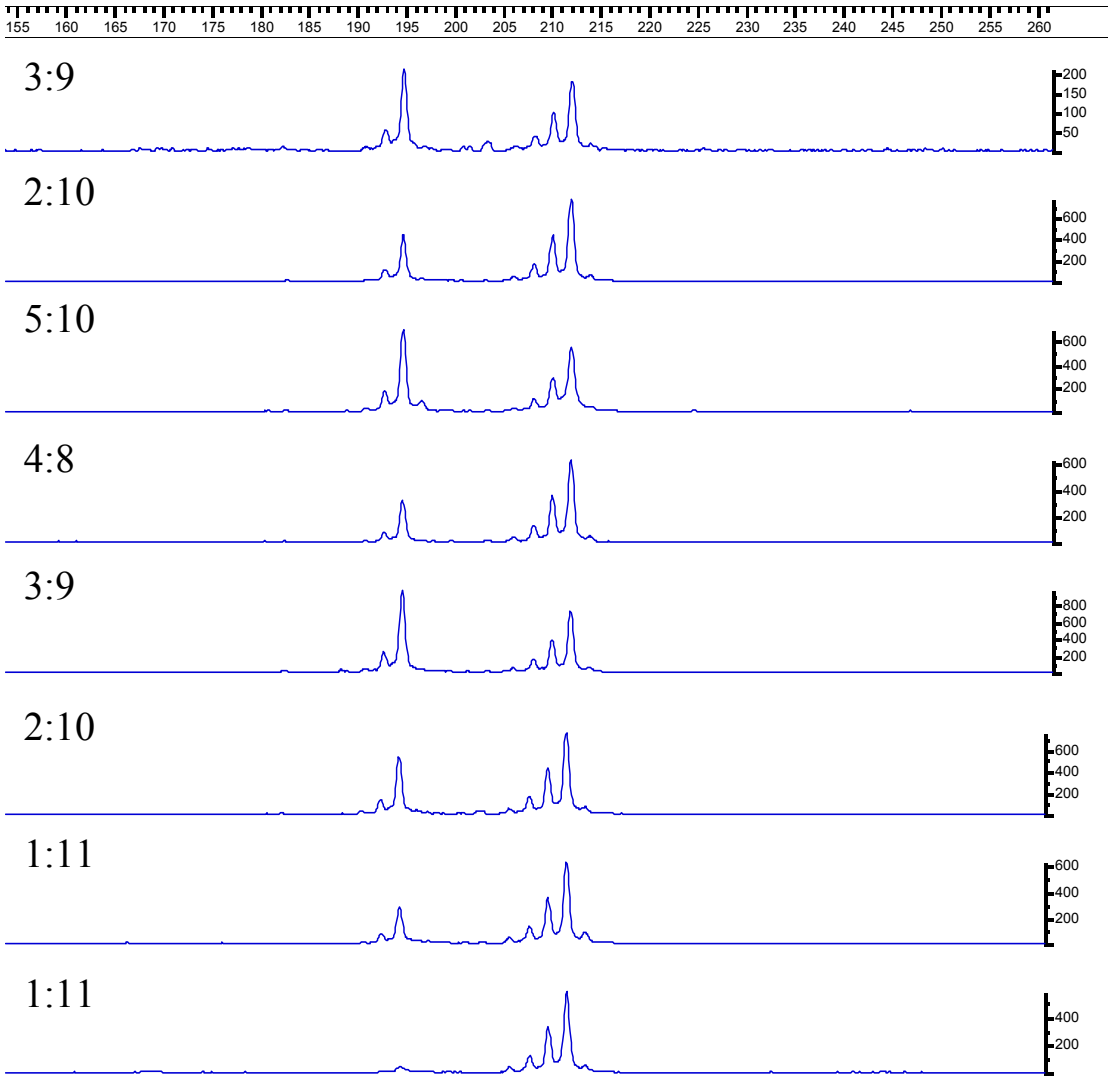


Figure 16. Detection of off-type microsatellite allele using capillary electrophoresis system. (product size in bases given across the top). Admixture is between Muscat (majority) and Alexis and/or Tyne.

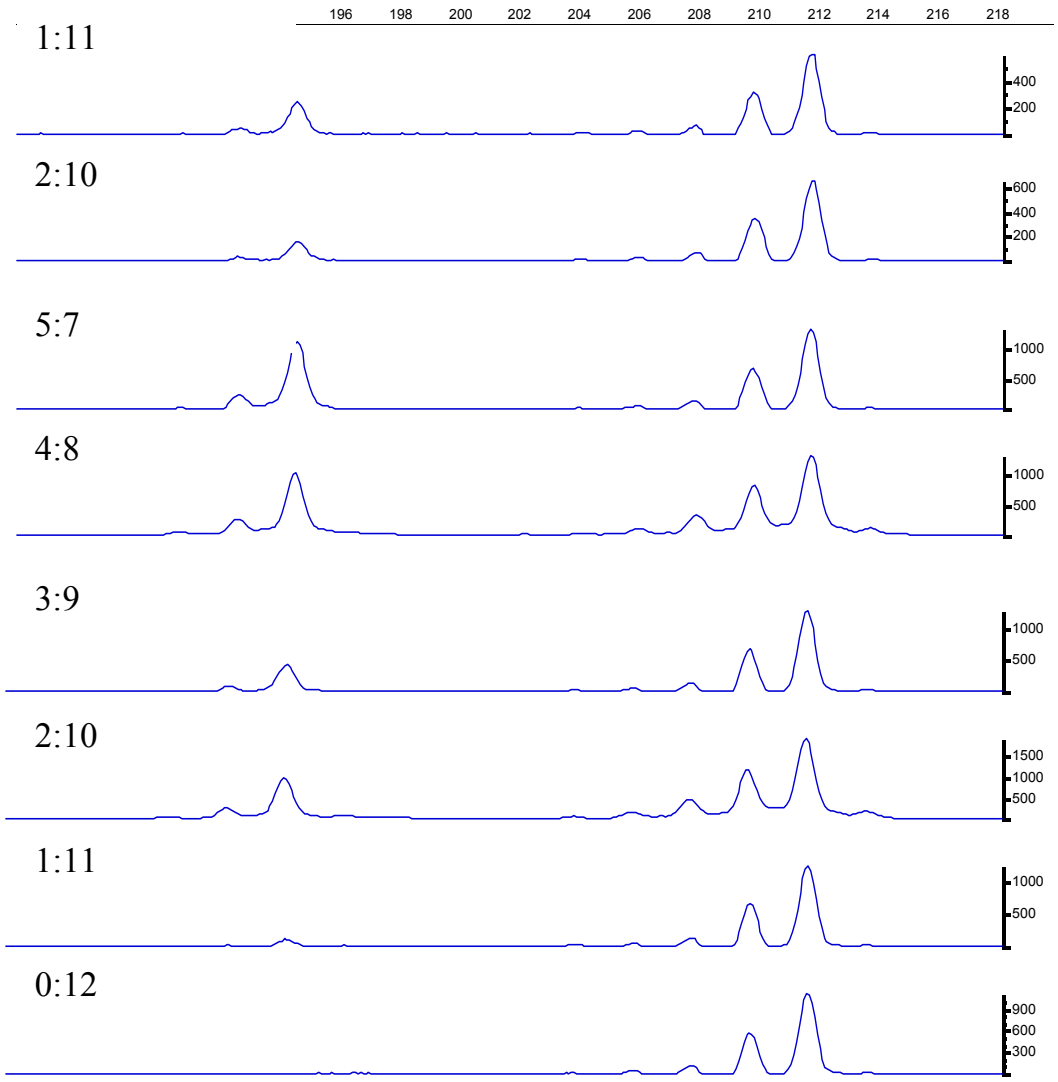


Figure 17. Detection of off-type microsatellite allele using capillary electrophoresis system. (product size in bases given across the top). Admixture is between Muscat (majority) and Tyne.

3.9.2. Artificial Mixture Studies

3.9.2.1.Experimental Design

As has been described in Section 3.5.1 artificial admixtures were constructed on micro-titre plates. Two were used in the limit of detection study (3.9.1) the remaining eight plates were used to check that known arrangements of admixture could be accurately detected using the matrix approach. (See 3.9.2.2)

3.9.2.2.Results from Artificial Mixture Plates

Results are summarised in a tabular format (see below). Major alleles are noted – these predominated in every row and every column. Minor alleles were seen in some rows and some columns and are recorded beside or below the main table. The known positions of off-type grains are given. In each table an interpretation of the result has been made using the statistical approach above.

Table 12. Positions of off-type grains for Artificial Mixture plate 2 (AM2) and result obtained for off-type detection.

	H	G	F	E	D	C	B	A		BLY[4]	Bmag 135[3]	Bmag 209[2]	Bmac 209[1]	Con.
									1		-	-	-	-
		R							2		166	204	212	+
									3		-	-	-	-
							R		4		166	204	212	+
									5		-	-	-	-
				R					6		-	204	212	+
									7		-	-	-	-
									8		-	-	-	-
						R			9		166	204	212	+
									10		-	-	-	-
									11		-	-	-	-
								R	12		166	204	212	+
Con.	-	+	-	-	-	+	-	-						
Bmac 209[1]	-	212	-	-	-	212	-	-						
Bmag 209[2]	-	204	-	-	-	204	-	-						
Bmag 135[3]	-	166	-	-	-	166	-	-						
BLY[4]														

Main variety type = Optic

Off-type = Riviera

Major peak for markers used

Con. = Consensus result

[1] Major peak 210

[2] Major peak 202

[3] Major peak 180

[4] Major peak 229

R = Known position of off-type (Riviera) grain

= No off-type detected

Conclusion: Observed off-type alleles consistent with Riviera. Majority Alleles consistent with Optic

Admixture present at:	**5.2%.**
Matrix estimate of admixture:	**5.37%**
% of such samples containing less than 5% admixture:	**33%**

Table 13. Positions of off-type grains for Artificial Mixture plate 3 (AM3) and result obtained for off-type detection.

	H	G	F	E	D	C	B	A
				H				
		H						
		H						
							H	
				H		H		
	H							
						H		
				H				
								H
Con.						+	+	+
Bmac 209[1]								
Bmag 209[2]			206			206	206	206
Bmag 135[3]								
BLY[4]	217	217		217		217	217	217

	BLY[4]	Bmag 135[3]	Bmag 209[2]	Bmac 209[1]	Con.
1	217		206		+
2	217		206		+
3	217		206		+
4	217		206		+
5					
6	217		206		+
7					
8	217		206		+
9	217				
10					
11	217		206		+
12	217		206		+

Main variety type = Maris Otter

Off-type = Heligan

Major peak for markers used

Con. = Consensus result

[1] Major peak 195

[2] Major peak 204

[3] Major peak 162

[4] Major peak 192

H = Position of off-type (Heligan) grain

= No off-type detected

Conclusion: Observed off-type alleles consistent with Heligan. Majority Alleles consistent with M.Otter.

Admixture present at:	**10.4%.**
Matrix estimate of admixture:	**9.2%**
% of such samples containing less than 5% admixture:	**3%**

Figure 18 to Figure 28 give automated genotyper traces for the analysis of sample AM4. In each figure the off-type alleles are indicated to show how these were detected. Similar approaches were used in all the other mixtures tested. Figure 3 shows the area normalisation approach used to find off-type peaks obscured by stutter.

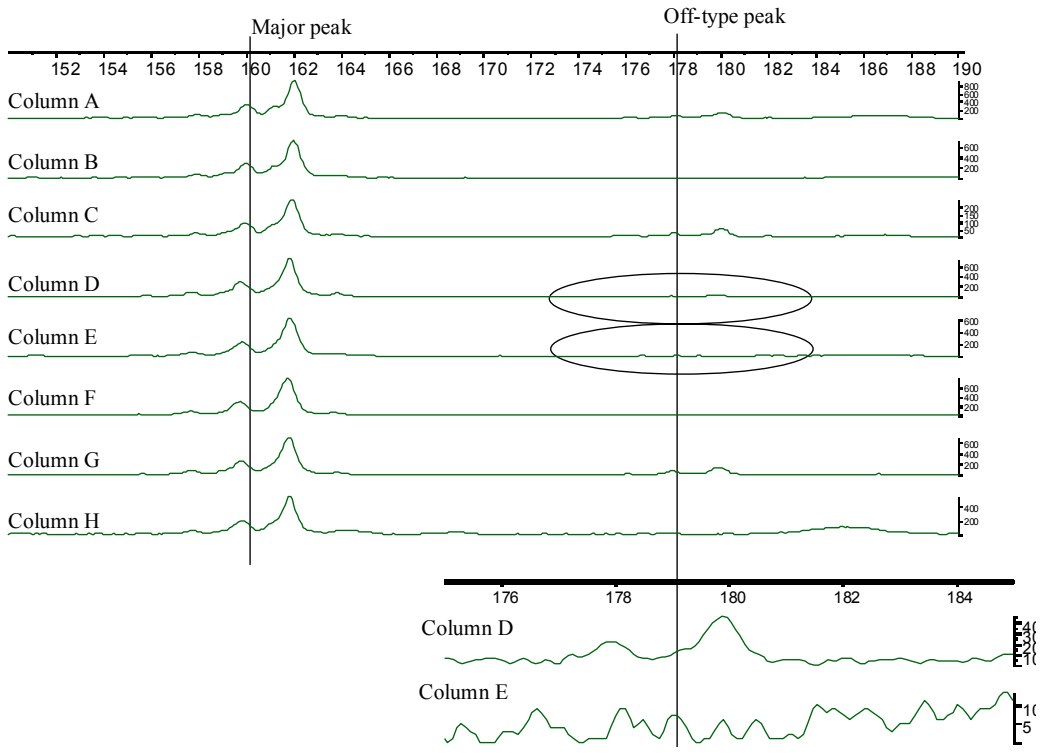


Figure 18. Output trace from ABI 3100 for column bulks A-H obtained from plate AM4 using microsatellite marker Bmag 135 showing major and off-type peaks. Zooming of column D reveals an off-type peak 180 in size, zooming of the same region of column E reveals no such peak. Therefore off-type peak 180 bases in size present in columns A, C, D and G.

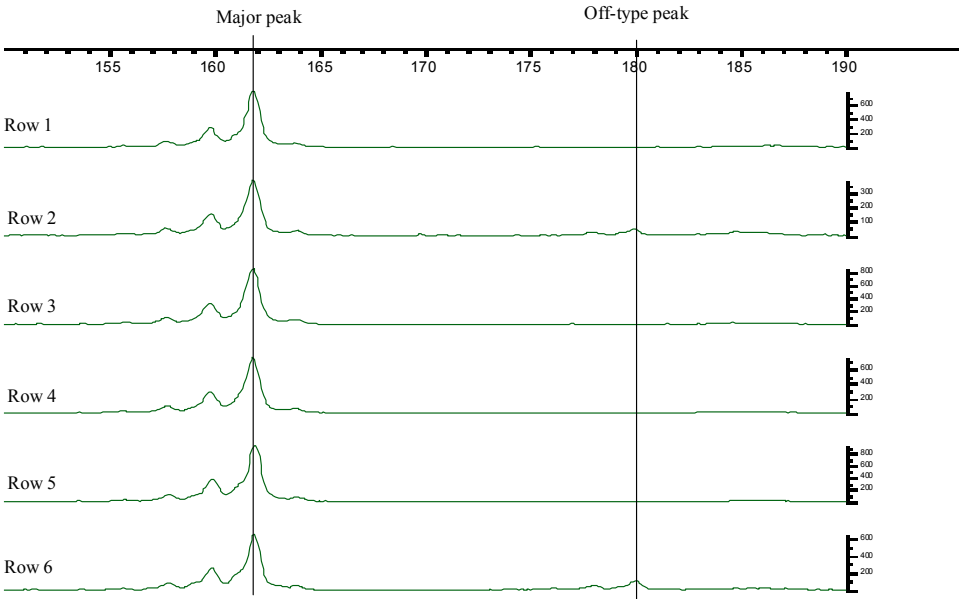


Figure 19. Output trace from ABI 3100 for row bulks 1-6 obtained from plate AM4 using microsatellite marker Bmag 135. Therefore off-type peak 180 bases in size present in rows 2 and 6.

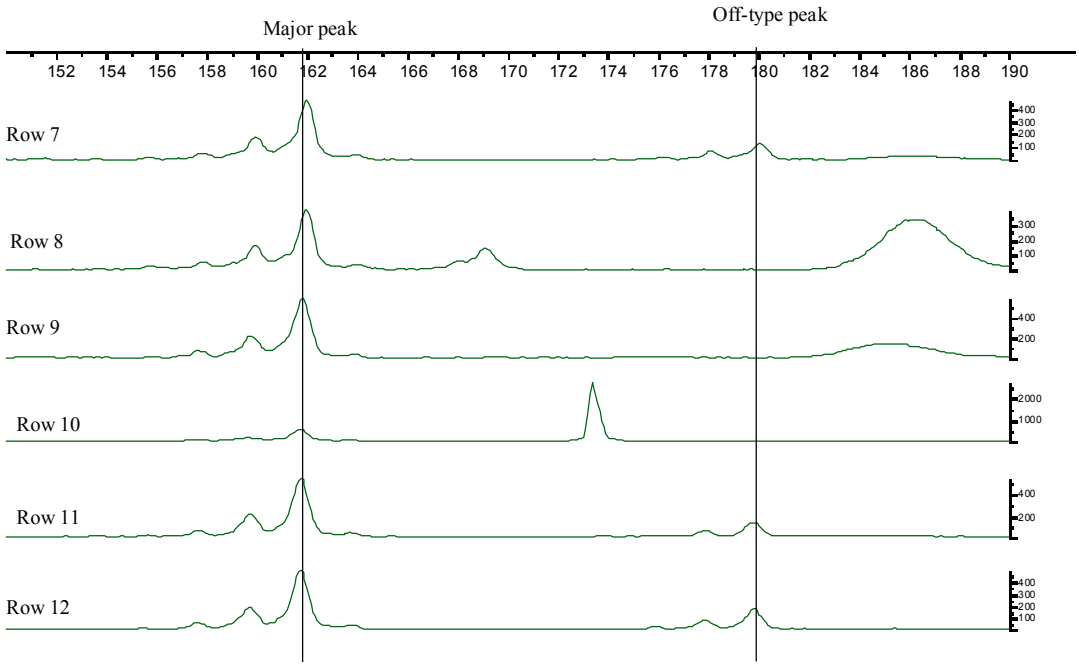


Figure 20. Output trace from ABI 3100 for row bulks 7-12 obtained from plate AM4 using microsatellite marker Bmag 135. Therefore off-type peak 180 bases in size present in rows 7, 11 and 12.

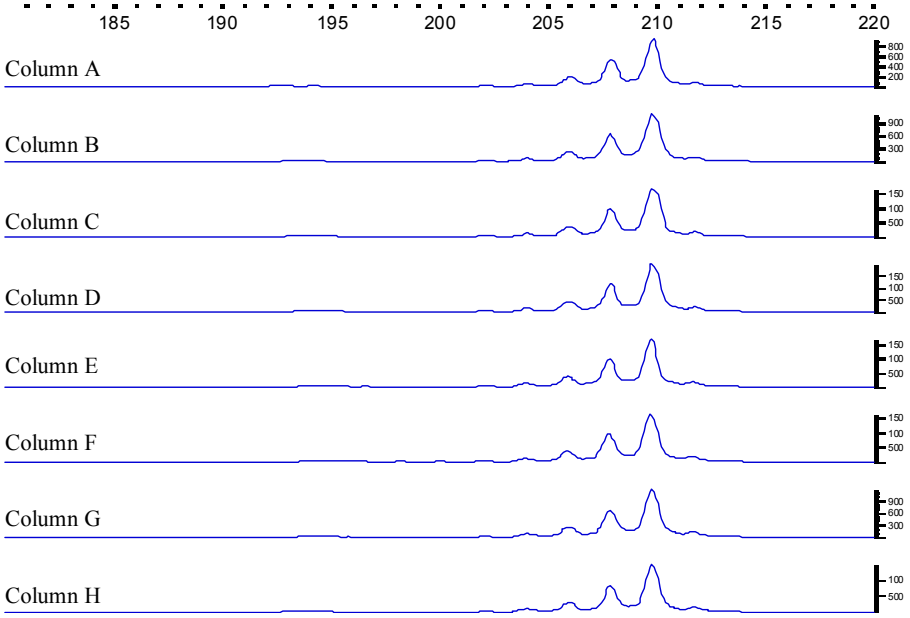


Figure 21. Output trace from ABI 3100 for row bulks A-H obtained from plate AM4 using microsatellite marker Bmac 209.

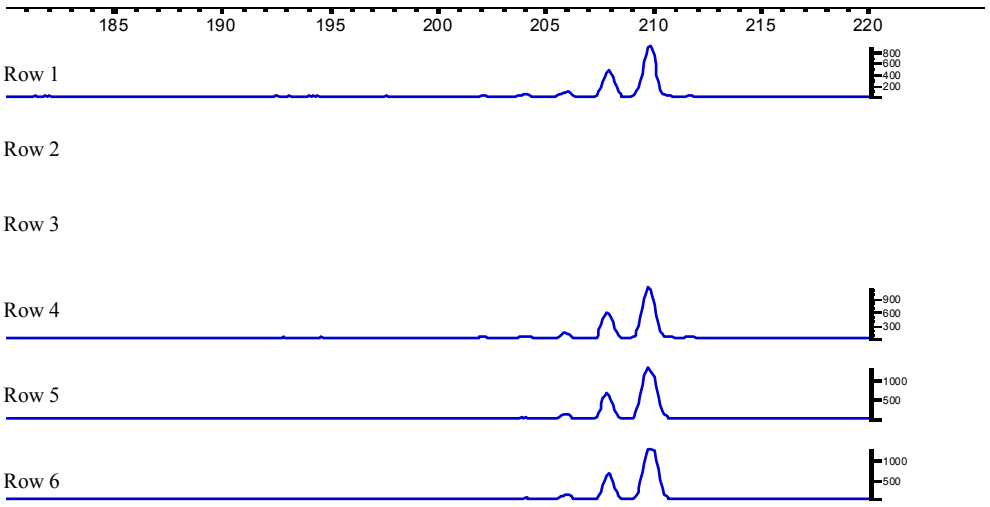


Figure 22. Output trace from ABI 3100 for row bulks 1-6 obtained from plate AM4 using microsatellite marker Bmac 209.

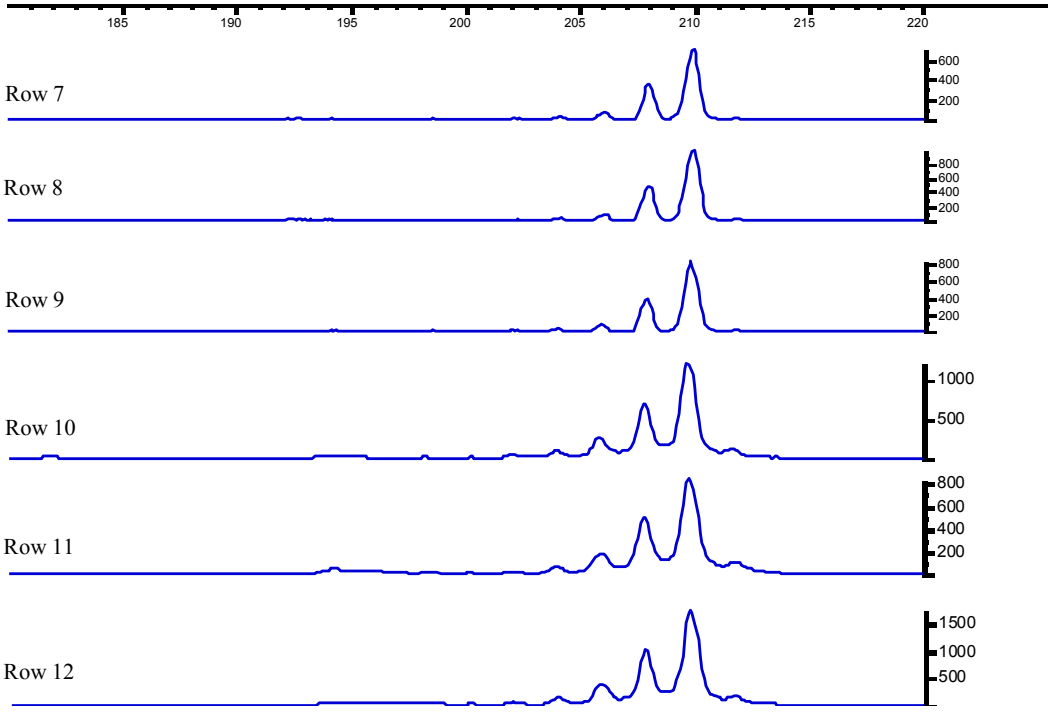


Figure 23. Output trace from ABI 3100 for row bulks 7-12 obtained from plate AM4 using microsatellite marker Bmac 209.

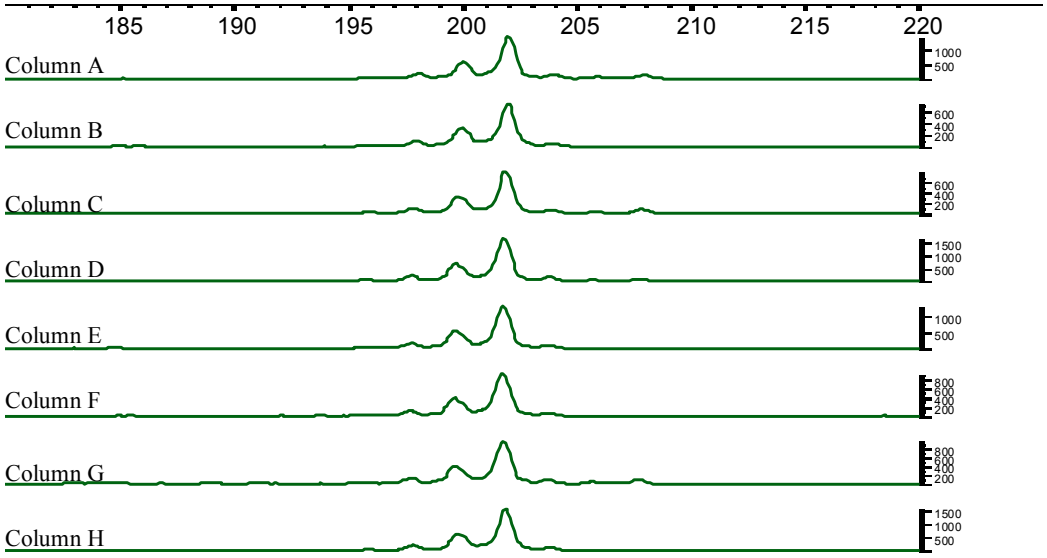


Figure 24. Output trace from ABI 3100 for column bulks A-H obtained from plate AM4 using microsatellite marker Bmag 209. Major peak size is 202 bases and off-type peaks 208 bases in size present in columns A, C, D and G.

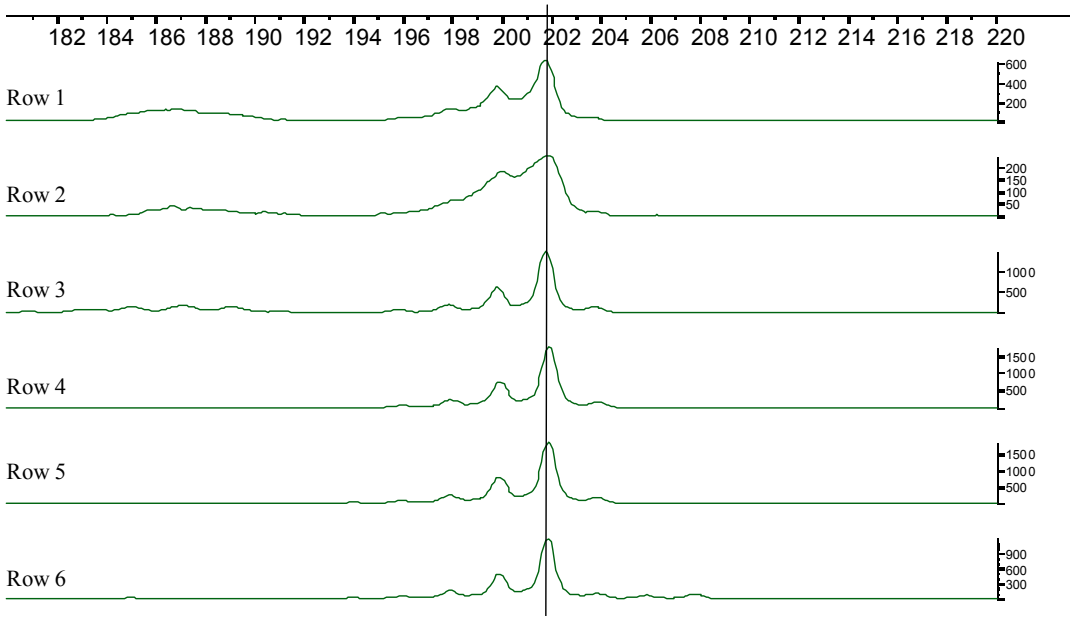


Figure 25. Output trace from ABI 3100 for row bulks 1-6 obtained from plate AM4 using microsatellite marker Bmag 209. Major peak size is 202 bases and off-type peaks 208 bases in size present in rows 2 and 6.

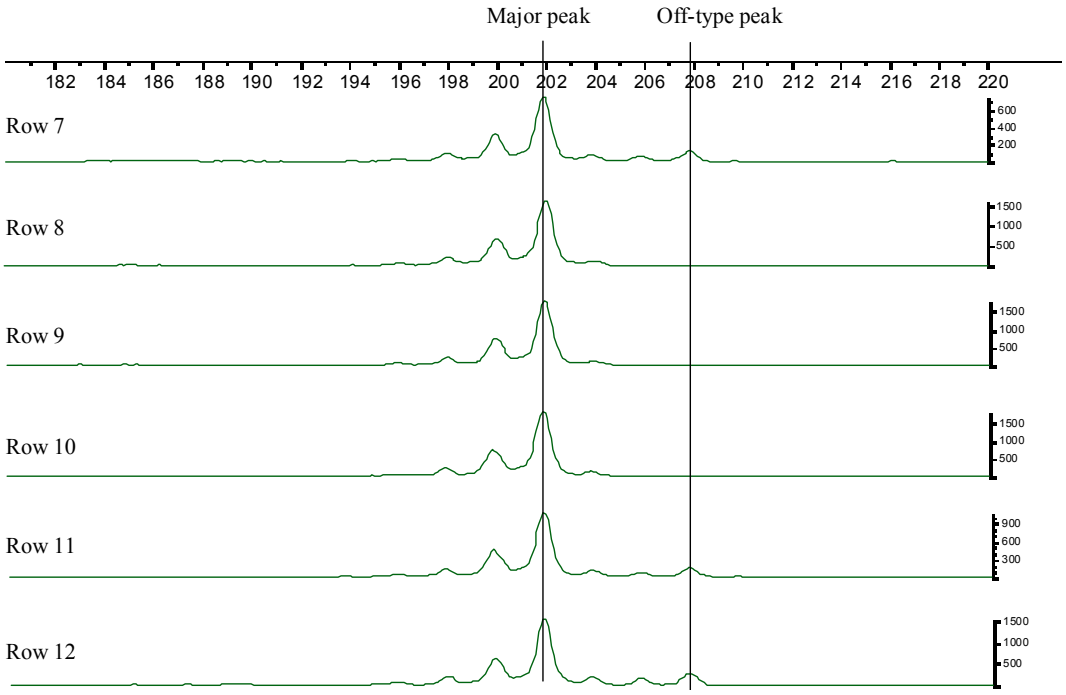


Figure 26. Output trace from ABI 3100 for row bulks 7-12 obtained from plate AM4 using microsatellite marker Bmag 209. Major peak at 202 bases, off-type peak at 208 bases in rows 7, 11 and 12.

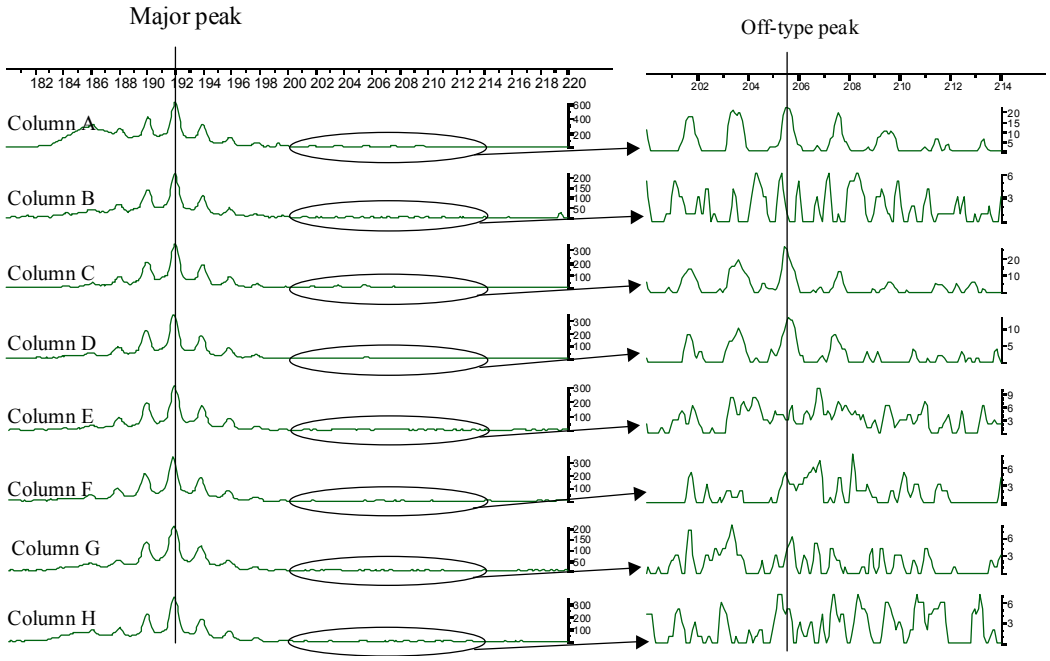


Figure 27. Output trace from ABI 3100 for column bulks A-H obtained from plate AM4 using microsatellite marker BLYCAB. Major peak at 192 bases, off-type peak 206 detected in columns A, C and D.

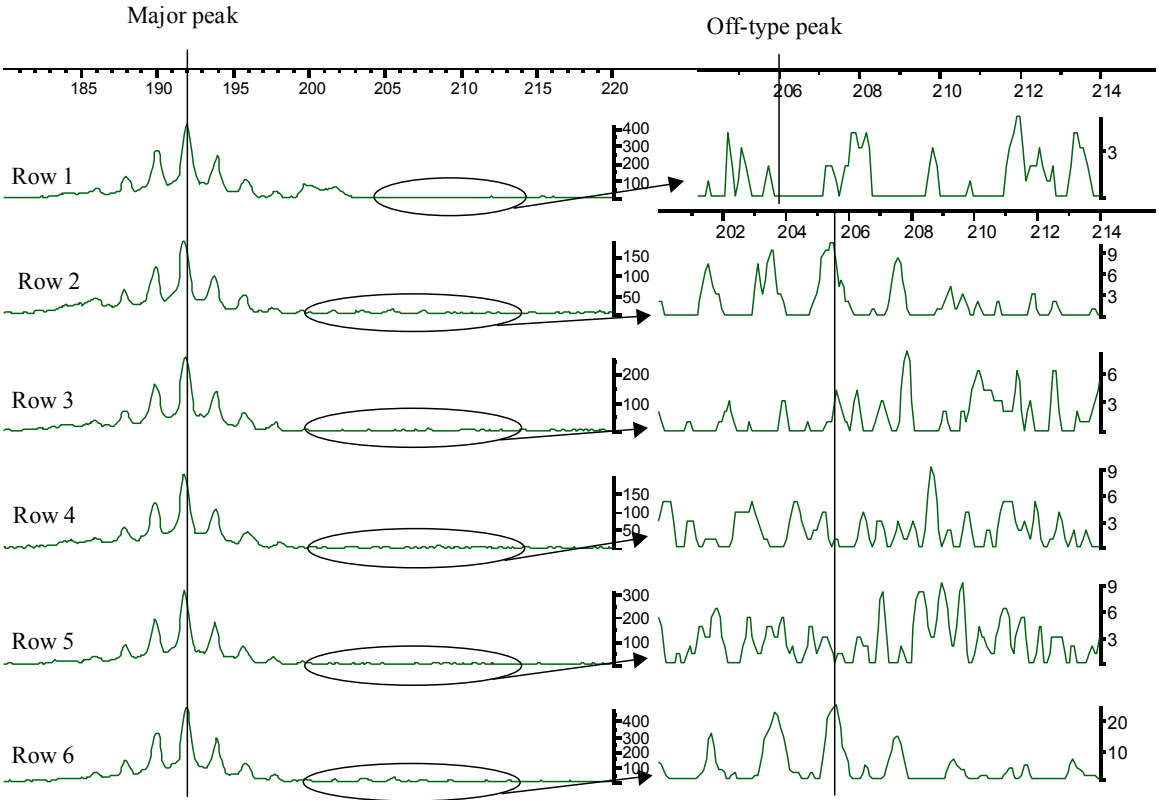


Figure 28. Output trace from ABI 3100 for row bulks 1-6 obtained from plate AM4 using microsatellite marker Bmag 209.

Table 14. Positions of off-type grains for Artificial Mixture plate 4 (AM4) and result obtained for off-type detection.

	H	G	F	E	D	C	B	A
					A			
		A						
						A		
								A
Con.		+			+	+		+
Bmac 209[1]								
Bmag 209[2]		208			208	208		208
Bmag 135[3]		180			180	180		180
BLY[4]					206	206		206

	BLY[4]	Bmag 135[3]	Bmag 209[2]	Bmac 209[1]	Con.
1					
2	206	180	208		+
3					
4					
5					
6	206	180	208		+
7	206	180	208		+
8					
9					
10					
11	206	180	208		+
12	206	180	208		+

Main variety type = Century

Off-type = Angela

Major peak for markers used

Con. = Consensus result

[1] Major peak Bmac 209

[2] Major peak 201

[3] Major peak 162

[4] Major peak 192

A = Position of off-type (Angela) grain

= No off-type detected

Conclusion: Observed off-type alleles consistent with Angela. Majority Alleles consistent with Century.

Admixture present at:	**4.2%.**
Matrix estimate of admixture:	**4.48%**
% of such samples containing less than 5% admixture:	**47%**

Table 15. Positions of off-type grains for Artificial Mixture plate 5 (AM5) and result obtained for off-type detection.

	H	G	F	E	D	C	B	A
					A			
		A					A	
						A		
				A				
	A							
		A						A
								A
Con.	+		+		+	+	+	+
Bmac 209[1]								210
Bmag 209[2]	204		204	204	204	204	204	204
Bmag 135[3]		182						162
BLY[4]		192	192		192	192	192	192

	BLY[4]	Bmag 135[3]	Bmag 209[2]	Bmac 209[1]	Con.
1			204		
2		162	204		+
3			204		
4			204	210	
5			204		
6	192		204		+
7	192	162	204		+
8	192	162	204		+
9		182	204		+
10		162	204		
11			204		
12			204		
Con.					
Bmac 209[1]					
Bmag 209[2]					
Bmag 135[3]					
BLY[4]					

Main variety type = Chalice

Off-type = Alexis

Major peak for markers used

Con. = Consensus result

[1] Major peak 195

[2] Major peak 202/3

[3] Major peak 166

[4] Major peak 203

A = Position of off-type (Alexis) grain

= No off-type detected

Conclusion: Observed off-type alleles consistent with Alexis. Majority Alleles consistent with Chalice

Admixture present at:	**9.4%.**
Matrix estimate of admixture:	**7.24%**
% of such samples containing less than 5% admixture:	**13%**

Table 16. Positions of off-type grains for Artificial Mixture plate 7 (AM7) and result obtained for off-type detection.

	H	G	F	E	D	C	B	A
		R						
							A	
				A				
						R		
								A
Con.		+		++		+	++	++
Bmac 209[1]				195			195	195
Bmag 209[2]		204				204		
Bmag 135[3]		166		164		166	164	164
BLY[4]								

	BLY[4]	Bmag 135[3]	Bmag 209[2]	Bmac 209[1]	Con.
1					
2	+	166	204		
3					
4	++	164		195	
5					
6	++	164		195	
7					
8					
9	+	166	204		
10					
11					
12	+++	164	204	195	

Main variety type = Optic

Off-type 1 = Alexis

Off-type 2 = Riviera

Major peak for markers used

Con. = Consensus result

[1] Major peak 210

[2] Major peak 202

[3] Major peak 180

[4] Major peak 232

A/ R = Position of off-type (Alexis/ Riviera) grain

= No off-type detected

+ = Riviera type off-type; ++ = Alexis type off-type; +++ Alexis and Riviera type off-type.

Conclusion: Observed off-type alleles consistent with Alexis and Riviera. Majority Alleles consistent with Optic.

Total admixture present at:	**5.2%.**
Matrix estimate of total admixture:	**5.83%**
% of such samples containing less than 5% admixture:	**26%**

Estimate and (actual) admixture with Alexis:	**3.23% (3.1%)**
Estimate and (actual) admixture with Riviera:	**3.18% (2.1%)**

Table 17. Positions of off-type grains for Artificial Mixture plate 8 (AM8) and result obtained for off-type detection.

	H	G	F	E	D	C	B	A
				H				
		H						
		D						
							D	
				D		H		
	H							
						D		
				H				
								D
Con.	+	+++		+++		+++	++	++
Bmac 209[1]		210		210			210	210
Bmag 209[2]	208	208		208		208		
Bmag 135[3]		166		166		166	166	166
BLY[4]		204		204		200		

	BLY[4]	Bmag 135[3]	Bmag 209[2]	Bmac 209[1]	Con.
1			208		+
2			208		+
3	204	166		210	++
4	204	166		210	++
5					
6	204	166	208	210	+++
7					
8			208		+
9					
10					
11					
12	204	166		210	++

Main variety type = Maris Otter

Off-type 1 = Heligan

Off-type 2 = Dandy

Major peak for markers used

Con. = Consensus result

[1] Major peak 195

[2] Major peak 204

[3] Major peak 162

[4] Major peak 192

H/ D = Position of off-type (Heligan/ Dandy) grain

= No off-type detected

+ = Heligan type off-type; ++ = Dandy type off-type; +++ Heligan and Dandy type off-type.

Conclusion: Observed off-type alleles consistent with Heligan and Dandy Majority Alleles consistent with M. Otter.

Total admixture present at:	**10.4%.**
Matrix estimate of total admixture:	**9.17%**
% of such samples containing less than 5% admixture:	**4%**
Estimate and (actual) admixture with Heligan:	**4.48% (5.2%)**
Estimate and (actual) admixture with Dandy:	**4.48% (5.2%)**

Table 18. Positions of off-type grains for Artificial Mixture plate 9 (AM9) and result obtained for off-type detection.

H	G	F	E	D	C	B	A		BLY[4]	Bmag 135[3]	Bmag 209[2]	Bmac 209[1]	Con.
					A			1		180	208		+
								2					
			P					3		166	208	Bmag 209	++
								4					
								5					
	A							6		180	208		+
								7					
								8					
								9					
								10					
						P		11		166	208	Bmag 209	++
							P	12		166	208	Bmag 209	++

Con.								
Bmac 209[1]				Bmag 209			Bmag 209	Bmag 209
Bmag 211[2]		208		208		208	208	
Bmag 135[3]		180		166		180	166	166
BLY RCA B[4]		+		++		+	++	++

Main variety type = Century

Off-type 1 = Angela

Off-type 2 = Peridot

Major peak for markers used

Con. = Consensus result

[1] Major peak 210

[2] Major peak 202

[3] Major peak 162

[4] Major peak 192

A/ P = Position of off-type (Angela/ Peridot) grain

= No off-type detected

+ = Angela type off-type; ++ = Peridot type off-type;

Conclusion: Observed off-type alleles consistent with Angela and Peridot, Majority Alleles consistent with Century.

Total admixture present at:	**5.2%.**
Matrix estimate of total admixture:	**5.83%**
% of such samples containing less than 5% admixture:	**26%**

Estimate and (actual) admixture with Angela:	**2.04% (2.1%)**
Estimate and (actual) admixture with Peridot:	**3.23% (3.1%)**

Table 19. Positions of off-type grains for Artificial Mixture plate 10 (AM10) and result obtained for off-type detection.

	H	G	F	E	D	C	B	A		BLY RCA B[4]	Bmag 135[3]	Bmag 211[2]	Bmac 209[1]	Con.
	A							M	1	208	164	204	212	+++
									2	208				
						A			3					
									4					
				A					5					
									6			204	212	++
							M		7	208			212	++
									8	208				
		A							9	208	164			+
		M							10	208		204	212	++
				A		M			11	208	164	204	212	+++
									12	208				
Con.		+++	++	+		++	+++	++						
Bmac 209[1]		212	212			212	212	212						
Bmag 211[2]		204	204			204	204	204						
Bmag 135[3]		164		164			164							
BLY RCA B[4]	208	208			208			208						

Main variety type = Chalice

Off-type 1 = Alexis

Off-type 2 = Muscat

Major peak for markers used

Con. = Consensus result

[1] Major peak 195

[2] Major peak 202

[3] Major peak 166

[4] Major peak 203

A/ M = Position of off-type (Alexis/ Muscat) grain

- = No off-type detected

+ = Alexis type off-type; ++ = Muscat type off-type; +++ = Alexis and Muscat off-type.

Conclusion: Observed off-type alleles consistent with Alexis and Muscat, Majority Alleles consistent with Chalice.

Total admixture present at:	**9.4%.**
Matrix estimate of total admixture:	**7.65%**
% of such samples containing less than 5% admixture:	**11%**

Estimate and (actual) admixture with Alexis:	**3.18% (5.2%)**
Estimate and (actual) admixture with Muscat:	**5.83% (4.1%)**

3.9.2.3. Double Blind Mixture Studies

As has been described in Section 3.5.2 artificial admixtures were constructed in such a way as to allow the method to be tested ‘blind’.

3.9.2.4.Results of Analysis of 'Blind' Samples

Table 20. Results obtained for blind test plate E

	H	G	F	E	D	C	B	A
	X		X	X			X	
	X		X	X			X	
	X		X	X			X	
	X		X	X			X	
	X		X	X			X	
	X		X	X			X	
	X		X	X			X	
								Y
Con.								
Bmac 209[1]	212		212	212			212	
Bmag 211[2]	204		204	204			204	
Bmag 135[3]	166		166	166			166	182
BLY RCA B[4]	nd	nd	nd	nd	nd	nd	nd	nd

	BLY RCA B[4]	Bmag 135[3]	Bmag 211[2]	Bmac 209[1]	Con.
1	nd				
2	nd	(166)	204	212	
3	nd	166	204	212	
4	nd	166	204	212	
5	nd	166	204	212	
6	nd				
7	nd	166	204	212	
8	nd	166	204	212	
9	nd	166	204	212	
10	nd				
11	nd	182			
12	nd				

Major peak for markers used

[1] Major peak - 195

[2] Major peak - 202

[3] Major peak - 164

[4] Major peak – failed to amplify adequately to detect off-types.

Conclusions From Analysis

Sample supplied to lab as:	Alexis
Genotype of majority of grains:	**Consistent with Alexis**
Off types observed:	Appear to be two additional genotypes (X and Y)
Possible positions of off-types:	Indicated above by X and Y.
No Columns containing off-type:	4 columns contain X
	1 column contains Y
No Rows containing off-type:	7 rows contain X
	1 row contains Y
Most probable level of admixture:	X = **8.28%**
	Y = **1.04%**
Most probable level of total admixture:	**9.42%**
95% limits of credible range:	**4.9% - 17.7%**
Proportion of such results from bulks containing <=5% admixture:	**2.7%**

True Level of Admixture in the Sample

Sample contained 86 grains of Alexis and 10 grains of Riviera. 10.4% Admixture overall.

Comments

X is consistent with Riviera. Y is completely unexpected. Overall estimate of admixture is correct.

Table 21. Results obtained for blind test plate D

	H	G	F	E	D	C	B	A
		D	D	D	F			F
		D	D	A	B			B
		F	F	B	B			B
		F	F	B	B			B
		D	D	D	F			F
Con.								
Bmac 209[1]		195	(195)	195	195			195
Bmag 211[2]		204	204	204				
Bmag 135[3]				166	166			166
BLYRCAB[4]	nd	nd	nd	nd	nd	nd	nd	nd

	BLYRCAB[4]	Bmag 135[3]	Bmag 211[2]	Bmac 209[1]	Con.
1	nd				
2	nd		204	195	
3	nd				
4	nd	166	204	195	
5	nd				
6	nd	166		195	
7	nd				
8	nd				
9	nd	166		195	
10	nd				
11	nd		204	(195)	
12	nd				

Major peak for markers used

[1] Major peak - 210

[2] Major peak - 202

[3] Major peak - 180

[4] Major peak – failed to amplify adequately to detect off-types.

Conclusions From Analysis

Sample supplied to lab as:	Optic
Genotype of majority of grains:	**Consistent with Optic**
Off types observed:	Appear to be four possible genotypes (A,B,D and F)
Possible positions of off-types:	Indicated above by A,B,D and F.
No Columns containing off-type:	5 (impossible to assign columns to genotype)
No Rows containing off-type:	5 (impossible to assign rows to genotype)
Most probable level of total admixture:	**5.83%**
95% limits of credible range:	**2.6% - 13.1%**
Proportion of such results from bulks containing <=5% admixture:	**26%**

True Level of Admixture in the Sample

Sample contained 91 grains of Optic, 2 grains Halcyon, 1 grain Chalice, 2 grains Cellar. 5.2% Admixture overall.

Comments

Presence of multiple contaminants was noted. Estimate of level of admixture correct.

Table 22. Results obtained for blind test plate F

	H	G	F	E	D	C	B	A
Con.								
Bmac 209[1]	194	194	194					
Bmag 211[2]								
Bmag 135[3]								
BLYRCAB[4]	nd	nd	nd	nd	nd	nd	nd	nd

	BLYRCAB[4]	Bmag 135[3]	Bmag 211[2]	Bmac 209[1]	Con.
1	nd				
2	nd				
3	nd			194	
4	nd				
5	nd				
6	nd				
7	nd			194	
8	nd			194	
9	nd				
10	nd				
11	nd				
12	nd				

Major peak for markers used

[1] Major peak - 210

[2] Major peak - 202

[3] Major peak - 180

[4] Major peak – failed to amplify adequately to detect off-types.

Conclusions From Analysis

Sample supplied to lab as:	Chariot
Genotype of majority of grains:	**Consistent with Chariot**
Off types observed:	Appear to be one additional genotypes (X)
Possible positions of off-types:	Indicated above by X.
No Columns containing off-type:	3 columns contain X.
No Rows containing off-type:	3 rows contain X.
Most probable level of total admixture:	**3.23%**
95% limits of credible range:	**1.2% - 9.1%**
Proportion of such results from bulks containing <=5% admixture:	**69%**

True Level of Admixture in the Sample

Sample contained 81 grains of Chariot, 10 grains Optic and 5 grains Cellar. 15.6% Admixture overall. Cellar admixture present at 5.2%.

Comments

Optic was not detected because the BLYRCAB marker failed to amplify. Cellar grains were detected and the estimate of the level of Cellar admixture was correct.

Table 23. Results obtained for blind test plate A

	H	G	F	E	D	C	B	A
	X			X				
	X			X				
	X			X				
Con.								
Bmac 209[1]	210			210				
Bmag 211[2]	202			202				
Bmag 135[3]								
BL YR CA B[4]	nd	nd	nd	nd	nd	nd	nd	nd

	BL YR CA B[4]	Bmag 135[3]	Bmag 211[2]	Bmac 209[1]	Con.
1	nd				
2	nd				
3	nd				
4	nd				
5	nd		202	210	
6	nd				
7	nd		202	210	
8	nd				
9	nd				
10	nd				
11	nd		202	210	
12	nd				

Major peak for markers used

[1] Major peak - 212

[2] Major peak - 204

[3] Major peak - 180

[4] Major peak – failed to amplify adequately to detect off-types.

Conclusions From Analysis

Sample supplied to lab as:	Muscat
Genotype of majority of grains:	**Consistent with Muscat**
Off types observed:	Appear to be one possible genotype (X)
Possible positions of off-types:	Indicated above by (X)
No Columns containing off-type:	2
No Rows containing off-type:	3
Most probable level of total admixture:	**3.18%**
95% limits of credible range:	**1.2% - 9.0%**
Proportion of such results from bulks containing <=5% admixture:	**71%**

True Level of Admixture in the Sample

Sample contained 92 grains of Muscat and 4 grains Optic. 4.2% Admixture overall.

Comments

Admixture observed is consistent with Optic. Estimate of level of admixture is correct.

Table 24. Results obtained for blind test plate C

	H	G	F	E	D	C	B	A
			X	X	X	X		
			X	X	X	X		
			X	X	X	X		
			X	X	X	X		
			X	X	X	X		
Con.								
Bmac 209[1]			(197)	(197)	(197)	197		
Bmag 211[2]								
Bmag 135[3]								
BLY RCA B[4]	nd	nd	nd	nd	nd	nd	nd	nd

	BLY RCA B[4]	Bmag 135[3]	Bmag 211[2]	Bmac 209[1]	Con.
1	nd				
2	nd				
3	nd			197	
4	nd			197	
5	nd				
6	nd				
7	nd			(197)	
8	nd			(197)	
9	nd				
10	nd				
11	nd			197	
12	nd				

Major peak for markers used

[1] Major peak - 195

[2] Major peak - 202

[3] Major peak - 166

[4] Major peak – failed to amplify adequately to detect off-types.

Conclusions From Analysis

Sample supplied to lab as:	Cellar
Genotype of majority of grains:	**Consistent with Cellar**
Off types observed:	Appear to be one possible genotype (X)
Possible positions of off-types:	Indicated above by (X)
No Columns containing off-type:	4
No Rows containing off-type:	5
Most probable level of total admixture:	**5.68%**
95% limits of credible range:	**2.5% - 12.7%**
Proportion of such results from bulks containing <=5% admixture:	**28.3%**

True Level of Admixture in the Sample

Sample contained 94 grains of Cellar and 2 grains Optic. 2.1% Admixture overall.

Comments

Admixture observed is not consistent with Optic. Estimate of level of admixture is incorrect (P>0.95). Impossible to explain even when the data are examined knowing the admixture present. In effect the admixture detected was artifactual and not from the Optic grains added. Failure of BLYRCAB to amplify is probably a factor.

3.10. Criteria for Choosing Microsatellite Markers for this Type of Application.

Experience with the marker set chosen and its application to test samples reveals that BLYRCAB and to some extent Bmag 135 have been unreliable in routine use. The problems have been mainly due to poor amplification such that the major allele is visible but off-type alleles are insufficiently amplified to be visible above base-line noise. This appears to be because these markers do not amplify as readily as Bmag 211 and Bmac 209. The reason for this is obscure.

Initial concerns about stutter peaks masking off-type alleles proved only partly founded. In Bmac 209 and Bmag 211 it was relatively simple to detect hidden off-types by taking ratios of peak areas or peak heights. BLYRCAB alleles tended not to overlap significantly so stutter was not an issue, however poor amplification made the marker unreliable.

Thus the ideal marker for this application would ***amplify strongly and reliably***. As a rule of thumb, a peak height of about 100 times the baseline noise allows off-type alleles to be seen at 1:11 dilution as clear peaks above the baseline.

In terms of separation of alleles by size, it is certainly convenient ***if alleles differ by several base pairs in length*** and so separate completely. In this situation stutter is not a problem. However, overlapping alleles can be distinguished by careful examination of peak areas so this is not essential.

Stutter seems to be a feature of all barley microsatellites tested. It is not clear why this should be because, for example, microsatellite markers used potatoes do not exhibit this feature. Despite the relative ease with which the stutter problem can be overcome the **ideal molecular marker for use in this context would not exhibit stutter peaks.**

3.11. Computer Simulations of Replicate Sampling of Bulks

The mathematical approach outlined in Section 6 appears to be a valid way to estimate most likely level of admixture and the credible range around that estimate from counted batches of grain. In

order to validate the proposed statistical method computer simulation was used to generate virtual grain samples using a random number generator. Virtual grain bulks of infinite size were established with levels of admixture from 1% to 17%.

By repeatedly re-sampling the virtual grain bulks it is possible to generate large volumes of prediction data which can be used to test whether the method correctly estimates the most likely level of admixture and the extent to which the credible range includes the true level of admixture for any sample.

3.11.1. Experimental Design

A series of numbers from the random number generator was used to assign attributes to virtual grains. The attribute assigned as either 'true to type' or 'off-type'. The proportion of virtual grains in the latter category was adjusted by altering the value of the random number below which a grain would be assigned to the 'off-type' category.

In the experiment a population of 5000 virtual samples was created at each of 17 levels of admixture (1%, 2%, 3%…..17%). Each virtual sample consisted of 96 virtual grains, the virtual grains were held in an 8 x 12 array simulating the micro-titre plate format.

Software was written which examined each virtual row and virtual column in each virtual sample and identified those containing no off-types and those which contained at least one off-type.

From these data it was possible to make three estimates of the most likely admixture and for each virtual sample. The three estimates came from 'rows only', 'columns only' and the matrix approach. The means of these estimates for all samples are recorded in Table 25.

In addition to a sample by sample estimate of the means it was possible to make an overall estimate based on a sample of 100 rows or columns. This composite estimate avoids problems encountered when the data were examined as if from micro-titre plates.

3.11.2. Results of Computer Simulation.

3.11.2.1. Rows only Method

There are 12 rows on a micro-titre plate each containing 8 grains, thus there are 13 possible outcomes; 0,1,2,3…12 rows exhibiting off-types. Where 0 rows exhibit off-types or 12 rows all exhibit off-type it is impossible to estimate the level of admixture present except to say that admixture, if present, is unlikely to be greater (or less) than a particular value (see Equation 1 and Table 5). The existence of these two outcomes has a detrimental effect on the prediction of most likely admixture since all micro-titre plates containing no grains of off-type contain an estimated 0% admixture and all plates containing 12 rows each containing at least one grain of admixture may contain as much as 100% admixture. The consequence of assigning 100% admixture to all samples exhibiting 12 rows containing admixture is to substantially overestimate the most probable level of admixture for samples containing more than 10% admixture (see: Table 25).

3.11.2.2. Columns only Method

There are 8 columns on a micro-titre plate each containing 12 grains, thus there are 9 possible outcomes; 0,1,2,3…9 columns rows exhibiting off-types. Where 0 columns exhibit off-types or 9 columns all exhibit off-type it is impossible to estimate the level of admixture present except to say that admixture, if present, is unlikely to be greater (or less) than a particular value (see Equation 1 and Table 5). The existence of these two outcomes has a detrimental effect on the prediction of most likely admixture since all micro-titre plates containing no grains of off-type contain an estimated 0% admixture and all plates containing 8 columns each containing at least one grain of admixture may contain as much as 100% admixture. The consequence of assigning 100% admixture to all samples exhibiting 8 columns containing admixture is to substantially overestimate the most probably level of admixture for samples containing more than 5% admixture (see: Table 25).

3.11.2.3. Matrix Method

There are 8 columns and 12 rows on a micro-titre plate giving 69 possible permutations of rows and columns exhibiting off-types and no off-type. Where 0 rows and 0 columns exhibit off-type or where 8 columns and 12 rows exhibit off-type it is impossible to estimate the level of admixture present except to say that admixture, if present, is unlikely to be greater (or less) than a particular value (see Equation 1 and Table 5). The existence of these two outcomes has a detrimental effect on the prediction of most likely admixture since all micro-titre plates containing no grains of off-type contain an estimated 0% admixture and all plates containing 8 columns and 12 rows each containing at least one grain of admixture may contain as much as 100% admixture. The consequence of assigning 100% admixture to all such samples is to substantially overestimate the most probably level of admixture for samples containing more than 14% admixture (see: Table 25).

3.11.2.4. Larger Sample Size

The results reported in Table 25 for row, column and matrix methods are disappointing in that they show an overall bias as admixture concentration rises. To validate the statistics, and to demonstrate that it is the absence of a mixture of both positive and negative batches which has led to the bias the computer generated sample set was re-interrogated on a rows only or columns only basis such that 100 rows or columns were used to estimate the most probable level of admixture. The results of this analysis, given in Table 25, confirm that in situations where both positive and negative batches are present the estimate of most probable admixture is extremely accurate.

Table 25 Means Estimated by Various Methods

	Mean of 10000 simulated micro-titre plates			Prediction using 100 columns (12 grains per col) estimated admixture (%)	Prediction using 100 rows (8 grains per row) estimated admixture (%)
True Admixture (%)	Estimated admixture (%) from columns only	Estimated admixture (%) from rows only	Estimated admixture (%) using Matrix Method		
1.00	1.06	1.05	0.96	1.03	0.98
2.01	2.15	2.11	1.94	1.94	2.12
3.03	3.27	3.16	2.90	3.01	3.07
4.02	4.44	4.21	3.82	4.08	4.00
4.98	5.56	5.22	4.71	4.95	5.05
6.00	6.96	6.28	5.63	6.02	5.99
7.01	8.83	7.40	6.56	7.03	7.14
7.98	10.86	8.44	7.52	7.93	8.10
8.98	13.87	9.54	8.45	9.09	8.94
10.00	16.96	10.75	9.53	9.94	9.97
11.02	20.37	11.92	10.60	11.06	10.91
11.94	24.41	13.07	11.69	12.03	12.10
13.03	29.08	14.48	13.03	13.09	13.03
14.00	33.43	16.00	14.66	13.97	14.06
14.99	38.90	17.77	16.48	14.94	14.92
16.04	43.19	19.44	18.24	15.91	15.91
17.05	48.97	21.71	20.72	17.03	17.08

3.11.2.5. Confidence Limits and Credible Ranges

For each possible outcome at each level of admixture the number of samples exhibiting the particular outcome was noted and the result recorded as a frequency table (Table 26 to Table 30). This layout can be used to give confirmation that the credible ranges estimated for any sample reflect the outcome observed when many samples are tested. In addition, the buyer and seller risks may be interpolated from the frequency tables. from the Column Only Method.

Table 26 Proportion of samples exhibiting each estimated level of admixture which contained less than or equal to the actual level of admixture in the virtual bulk.

	Number of columns exhibiting the presence of at least one off-type grain.								
	0	1	2	3	4	5	6	7	8
	Estimate of admixture corresponding to the above number of positive columns (%)								
Admixture (%)	0	1.094	2.344	3.854	5.625	7.865	10.885	15.885	100
1%	62.85%	36.25%	12.96%	2.82%	0.38%	0.06%	0.00%	0.00%	0.00%
2%	85.40%	65.44%	36.19%	13.29%	3.21%	0.64%	0.07%	0.01%	0.00%
3%	94.39%	82.39%	58.19%	29.72%	10.72%	2.65%	0.49%	0.06%	0.01%
4%	98.00%	91.69%	73.94%	47.23%	22.54%	7.44%	1.77%	0.29%	0.03%
5%	99.32%	96.21%	85.19%	63.10%	36.23%	14.95%	4.30%	0.88%	0.11%
6%	99.74%	98.29%	91.96%	75.54%	50.41%	24.73%	8.61%	2.27%	0.35%
7%	99.93%	99.23%	95.84%	84.69%	62.71%	35.96%	14.96%	4.79%	1.06%
8%	99.97%	99.57%	97.96%	90.59%	73.42%	47.71%	22.85%	8.65%	2.39%
9%	99.97%	99.84%	98.96%	94.36%	81.65%	58.66%	32.24%	14.06%	4.96%
10%	99.98%	99.96%	99.50%	96.83%	87.60%	68.61%	42.28%	21.08%	8.90%
11%	100%	99.98%	99.75%	98.30%	91.89%	76.89%	52.72%	29.70%	14.45%
12%	100%	100%	99.89%	99.13%	94.95%	83.63%	62.85%	39.69%	22.07%
13%	100%	100%	99.94%	99.52%	96.94%	89.23%	72.07%	50.82%	32.21%
14%	100%	100%	99.98%	99.71%	98.35%	93.32%	80.70%	62.69%	44.74%
15%	100%	100%	99.99%	99.92%	99.17%	96.42%	88.21%	74.87%	60.41%
16%	100%	100%	99.99%	99.97%	99.68%	98.53%	94.65%	87.64%	78.49%
17%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 26 is essentially an experimental representation of buyers risk. Reading down a column the buyer can see what proportion of samples, exhibiting a particular number of positive columns in a set of 8 columns, would have a true level of admixture equal to or less than the value in the leftmost column. Thus if an analysis resulted in 6 negative batches and two batches containing at least one grain of admixture reading down the column gives a value of 85.2% of such samples would contain admixture at 5% or less.

Using this table a maltster who wanted 95% of his purchases to contain no more than 5% admixture would accept samples for which the analysis result showed either no positive columns or one positive column.

Table 27 Proportion of samples from virtual bulks containing each level of admixture which exhibited an observed level of admixture less than or equal to the mean.

	Number of columns exhibiting the presence of at least one off-type grain.								
	0	1	2	3	4	5	6	7	8
	Estimate of admixture corresponding to the above number of positive columns (%)								
Admixture (%)	0	1.094	2.344	3.854	5.625	7.865	10.885	15.885	100
1%	37.97%	77.30%	94.66%	99.10%	99.84%	99.99%	99.99%	100%	100%
2%	13.62%	45.29%	76.42%	92.91%	98.37%	99.77%	99.99%	100%	100%
3%	5.43%	23.82%	53.29%	79.17%	93.67%	98.56%	99.81%	99.98%	100%
4%	2.18%	12.26%	33.37%	60.95%	83.80%	95.42%	99.22%	99.96%	100%
5%	0.80%	5.71%	20.78%	45.77%	72.21%	90.42%	97.99%	99.85%	100%
6%	0.25%	2.50%	11.57%	31.16%	58.56%	82.28%	95.15%	99.54%	100%
7%	0.12%	1.14%	6.33%	20.74%	44.49%	71.75%	90.70%	98.67%	100%
8%	0.02%	0.39%	3.24%	12.53%	33.23%	61.74%	85.29%	97.47%	100%
9%	0.00%	0.30%	1.63%	7.56%	23.46%	50.01%	78.01%	95.14%	100%
10%	0.01%	0.14%	0.87%	4.76%	16.25%	40.40%	70.37%	92.54%	100%
11%	0.01%	0.03%	0.37%	2.69%	10.98%	31.06%	62.24%	89.50%	100%
12%	0.00%	0.02%	0.20%	1.51%	7.42%	23.78%	54.01%	85.57%	100%
13%	0.00%	0.00%	0.07%	0.69%	4.52%	18.11%	45.61%	80.82%	100%
14%	0.00%	0.00%	0.05%	0.34%	3.07%	12.99%	38.76%	76.28%	100%
15%	0.00%	0.00%	0.02%	0.35%	1.93%	9.45%	31.87%	70.35%	100%
16%	0.00%	0.00%	0.00%	0.09%	1.09%	6.20%	25.40%	65.77%	100%
17%	0.00%	0.00%	0.01%	0.05%	0.66%	4.23%	20.21%	59.29%	100%

Table 29 is essentially an experimental representation of seller's risk. Reading across a row the seller can see what proportion of samples, containing a true level of admixture given in the left-most column would give an analytical result equal to or less than the estimate given in the fourth row. Thus if a sample containing a true level of admixture of 3% were tested using the 'columns only' approach 98.4% of such samples would be estimated to contain 5.625% admixture or less.

Using this table a seller who wanted 95% of his production to be accepted without problems against a maximum permitted level of admixture of 5.625% would need to aim to produce grain with a true level of admixture no greater than 2%.

3.11.2.6. Frequency Tables from the Rows Only Method.

Table 28 Proportion of samples exhibiting each estimated level of admixture which contained less than or equal to the actual level of admixture in the virtual bulk.

	Number of rows exhibiting the presence of at least one off-type grain.												
	0	1	2	3	4	5	6	7	8	9	10	11	12
	Estimate of admixture corresponding to the above number of positive rows (%)												
Admixture (%)	0	1.095	2.24	3.54	4.948	6.51	8.281	10.365	12.813	15.885	20.052	26.719	100
1%	62.85	37.29	14.47	3.72	0.64	0.10	0.01	0.00	0.00	0.00	0.00	0.00	0.00
2%	85.40	66.87	38.83	16.78	5.17	1.31	0.22	0.05	0.00	0.00	0.00	0.00	0.00
3%	94.39	83.54	61.00	35.95	15.49	5.59	1.49	0.33	0.07	0.01	0.00	0.00	0.00
4%	98.00	92.38	76.99	54.13	30.88	14.09	4.85	1.34	0.31	0.07	0.00	0.00	0.00
5%	99.32	96.54	87.50	69.80	47.07	26.35	10.70	3.74	1.10	0.25	0.06	0.02	0.00
6%	99.74	98.40	93.59	81.27	62.16	40.06	19.71	8.39	2.86	0.76	0.22	0.07	0.00
7%	99.93	99.29	96.80	89.25	73.85	53.89	30.91	15.27	6.33	2.09	0.68	0.17	0.22
8%	99.97	99.60	98.55	93.89	82.99	66.04	43.60	24.15	11.62	4.55	1.68	0.47	0.29
9%	99.97	99.87	99.23	96.64	89.52	76.11	55.62	34.82	18.85	8.67	3.69	1.42	0.37
10%	99.98	99.97	99.68	98.27	93.65	83.82	66.70	46.04	27.88	14.83	7.13	3.31	1.61
11%	100	99.99	99.83	99.07	96.25	89.66	75.94	57.30	38.54	22.69	12.78	6.51	3.65
12%	100	100	99.95	99.53	98.06	93.35	83.59	67.73	49.91	32.92	20.40	12.07	6.65
13%	100	100	99.98	99.78	98.89	95.94	89.48	77.07	61.31	44.76	30.99	21.07	12.20
14%	100	100	100	99.87	99.38	97.69	93.70	85.07	72.47	58.13	44.15	32.70	23.01
15%	100	100	100	99.97	99.75	98.96	96.61	91.55	82.48	72.18	60.14	49.83	40.39
16%	100	100	100	99.99	99.92	99.61	98.67	96.40	91.80	86.18	79.41	71.68	63.99
17%	100	100	100	100	100	100	100	100	100	100	100	100	100

Table 28 is essentially an experimental representation of buyers risk. Reading down a column the buyer can see what proportion of samples, exhibiting a particular number of positive rows in a set of 12 rows, would have a true level of admixture equal to or less than the value in the leftmost column. Thus if an analysis resulted in 10 negative batches and two batches containing at least one grain of admixture reading down the fourth column gives a value of 87.5% of such samples would contain admixture at 5% or less.

Using this table a maltster who wanted 95% of his purchases to contain no more than 5% admixture would accept samples for which the analysis result showed either no positive rows or one positive row.

Table 29 Proportion of samples from virtual bulks containing each level of admixture which exhibited an observed level of admixture less than or equal to the mean.

	Number of rows exhibiting the presence of at least one off-type grain.												
	0	1	2	3	4	5	6	7	8	9	10	11	12
	Estimate of admixture corresponding to the above number of positive rows (%)												
	0	1.095	2.24	3.54	4.948	6.51	8.281	10.365	12.8	15.89	20.05	26.72	100
Admixture (%)													
1%	37.97	76.49	93.94	98.88	99.81	99.97	99.99	100	100	100	100	100	100
2%	13.62	44.17	73.56	90.92	97.53	99.52	99.91	100	100	100	100	100	100
3%	5.43	22.65	49.39	74.86	89.91	96.96	99.27	99.83	99.98	100	100	100	100
4%	2.18	11.31	30.60	54.77	77.21	91.21	97.35	99.40	99.90	100	100	100	100
5%	0.80	5.10	17.78	38.60	62.21	82.38	93.08	97.95	99.58	99.92	99.99	100	100
6%	0.25	2.17	9.52	24.76	46.77	69.34	85.79	95.19	98.84	99.76	99.97	100	100
7%	0.12	1.04	4.91	15.51	32.57	55.34	75.80	89.74	96.93	99.33	99.91	99.97	100
8%	0.02	0.34	2.45	8.62	21.94	41.94	65.13	83.11	94.08	98.54	99.82	99.99	100
9%	0.00	0.28	1.10	4.76	14.28	30.86	52.83	74.43	89.41	96.88	99.44	99.99	100
10%	0.01	0.11	0.65	2.81	8.84	21.53	41.78	64.49	83.21	94.37	98.74	99.83	100
11%	0.01	0.03	0.22	1.29	5.08	14.69	31.56	54.35	76.44	90.67	97.87	99.72	100
12%	0.00	0.01	0.15	0.75	3.39	9.46	23.45	44.57	68.15	86.69	96.38	99.59	100
13%	0.00	0.00	0.03	0.37	1.58	5.84	16.59	35.50	59.12	80.56	94.04	99.24	100
14%	0.00	0.00	0.03	0.15	0.87	3.75	11.47	27.68	50.82	75.05	91.80	98.52	100
15%	0.00	0.00	0.00	0.13	0.67	2.76	8.08	21.19	41.94	67.38	87.73	97.62	100
16%	0.00	0.00	0.00	0.03	0.27	1.35	5.11	14.94	34.26	59.62	84.15	96.77	100
17%	0.00	0.00	0.00	0.01	0.13	0.77	3.20	10.48	27.47	52.51	78.71	95.07	100

Table 29 is essentially an experimental representation of seller's risk. Reading across a row the seller can see what proportion of samples, containing a true level of admixture given in the left-most column would give an analytical result equal to or less than the estimate given in the fourth row. Thus if a sample containing a true level of admixture of 3% were tested using the 'rows only' approach 89.91% of such samples would be estimated to contain 4.948% admixture or less. is essentially an experimental representation of seller's risk. Reading across a row the seller can see what

Using this table a seller who wanted 95% of his production to be accepted without problems against a maximum permitted level of admixture of 4.948% would need to aim to produce grain with a true level of admixture no greater than 2%.

3.11.2.7. Frequency Tables from the Matrix Method.

Table 30 Proportion of samples exhibiting each estimated level of admixture which contained less than or equal to the actual level of admixture in the virtual bulk.

	Positive Columns:	0	1	1	2	2	3	2	3	4
	Positive Rows:	0	1	2	2	3	3	4	4	4
	Estimated Mean (%):	0	1.04	2.08	2.08	3.18	3.23	4.3	4.4	4.48
Mean:										
1%		62%	38%	16%	15%	4%	4%	0%	0%	1%
2%		85%	68%	37%	39%	18%	17%	9%	6%	5%
3%		94%	84%	62%	61%	38%	36%	26%	17%	16%
4%		98%	93%	79%	77%	56%	54%	38%	33%	32%
5%		99%	97%	90%	88%	72%	70%	49%	49%	48%
6%		100%	98%	95%	94%	83%	81%	60%	64%	63%
7%		100%	99%	98%	97%	90%	89%	77%	76%	74%
8%		100%	100%	99%	99%	95%	94%	86%	84%	83%
9%		100%	100%	99%	99%	97%	97%	95%	90%	90%
10%		100%	100%	100%	100%	99%	98%	96%	95%	94%
11%		100%	100%	100%	100%	99%	99%	99%	97%	96%
12%		100%	100%	100%	100%	100%	100%	99%	99%	98%
13%		100%	100%	100%	100%	100%	100%	99%	99%	99%
14%		100%	100%	100%	100%	100%	100%	99%	100%	99%
15%		100%	100%	100%	100%	100%	100%	100%	100%	100%
16%		100%	100%	100%	100%	100%	100%	100%	100%	100%
17%		100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 30 is essentially an experimental representation of buyers risk. Reading down a column the buyer can see what proportion of samples, exhibiting a particular number of positive rows and columns, would have a true level of admixture equal to or less than the value in the leftmost column. Thus if an analysis showed 2 rows and 2 columns containing admixture reading down the sixth column gives a value of 88% of such samples would contain admixture at 5% or less.

Using this table a maltster who wanted 95% of his purchases to contain no more than 5% admixture would accept samples for which the analysis result showed either no positive rows or columns or one positive row and one positive column.

Table 31 Proportion of samples from virtual bulks containing each level of admixture which exhibited an observed level of admixture less than or equal to the mean.

	Positive Columns:	0	1	1	2	2	3	2	3	4	3	4	5	5	3	4	5	4	6	6
	Positive Rows:	0	1	2	2	3	3	4	4	4	5	5	5	7	6	6	6	7	6	7
	Estimated Mean(%):	0	1.04	2.08	2.08	3.18	3.23	4.3	4.4	4.48	5.5	5.68	5.83	6.65	6.7	6.93	7.24	7.28	7.65	9.17
Mean:																				
1%		38%	77%	78%	94%	95%	99%	99%	99%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
2%		14%	45%	46%	73%	77%	90%	91%	93%	97%	97%	99%	99%	100%	100%	100%	100%	100%	100%	100%
3%		5%	22%	24%	48%	54%	74%	74%	80%	90%	90%	94%	97%	97%	97%	97%	99%	99%	99%	100%
4%		2%	11%	13%	30%	35%	54%	54%	62%	76%	77%	84%	90%	91%	91%	92%	95%	95%	98%	99%
5%		1%	5%	6%	17%	21%	37%	37%	45%	60%	61%	71%	81%	83%	83%	85%	91%	91%	94%	97%
6%		0%	2%	3%	9%	12%	24%	24%	32%	45%	46%	58%	68%	71%	71%	73%	82%	82%	88%	93%
7%		0%	1%	1%	5%	7%	16%	16%	21%	32%	33%	43%	54%	58%	58%	60%	72%	73%	79%	87%
8%		0%	0%	0%	2%	3%	8%	8%	12%	21%	22%	30%	40%	45%	45%	48%	61%	61%	69%	78%
9%		0%	0%	0%	1%	2%	5%	5%	8%	14%	14%	22%	30%	35%	35%	38%	49%	50%	57%	69%
10%		0%	0%	0%	1%	1%	3%	3%	5%	8%	9%	14%	20%	26%	26%	29%	39%	40%	47%	60%
11%		0%	0%	0%	0%	0%	1%	1%	2%	5%	5%	9%	14%	19%	19%	21%	31%	31%	37%	50%
12%		0%	0%	0%	0%	0%	1%	1%	1%	3%	3%	6%	9%	13%	13%	15%	23%	23%	28%	40%
13%		0%	0%	0%	0%	0%	0%	0%	1%	1%	1%	3%	5%	10%	10%	11%	16%	17%	20%	30%
14%		0%	0%	0%	0%	0%	0%	0%	0%	1%	1%	2%	4%	7%	7%	8%	12%	12%	15%	23%
15%		0%	0%	0%	0%	0%	0%	0%	0%	1%	1%	1%	2%	5%	5%	5%	8%	8%	10%	17%
16%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	1%	3%	3%	3%	5%	5%	7%	12%
17%		0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	2%	2%	2%	4%	4%	4%	8%

Table 31 is essentially an experimental representation of seller's risk. Reading across a row the seller can see what proportion of samples, containing a true level of admixture given in the left-most column would give an analytical result equal to or less than the estimate given in the third row. Thus if a sample containing a true level of admixture of 3% were tested using the 'matrix' approach 90% of such samples would be estimated to contain 5.5% admixture or less.

Using this table a seller who wanted 95% of his production to be accepted without problems against a maximum permitted level of admixture of 5.5% would need to aim to produce grain with a true level of admixture no greater than 2%.

3.12. Conclusion

From this study it is clear that a counted batch sampling approach can be used to estimate, with useful accuracy, the level of admixture in a grain sample. The estimate is strictly speaking a 'most probable admixture' and the precision of the estimate should be viewed as a 'most credible range' however, for practical purposes these values can be used as 'mean' and 'confidence interval'.

To apply the counted batch approach it is only necessary to establish which batches contain at least one grain of admixture and which contain no grains of admixture. This makes an analytical method based on the detection of alleles very attractive. The presence of off-type alleles in any batch will classify that batch as containing at 'least one grain of admixture'. The use of SSRs as a tool for making this classification has been demonstrated with some success. It is noted, however, that finding molecular markers which amplify strongly and reliably is critical if the method is to be used routinely. The authors report that two markers, 'Bmac 209' and 'Bmag 211' are very suitable, others tried have proved less reliable.

A critical factor in determining the presence of the off-type allele is an objective assessment of the capillary electrophoresis trace. Considerable effort was directed at detecting off-types where stutter around particular allelic forms overlap on traces. An area normalisation approach has proved effective in detecting hidden admixture. Where alleles are well separated in terms of No. of base pairs the recognition of the off-type is much simpler although good amplification is essential if low levels admixture is to be reliably detected above baseline noise.

For the method to be fully deployed it is important that the approach is highly discriminating (between varieties). The discrimination between varieties by any combination of methods is in itself a probability function. Thus even highly discriminating methods will fail to achieve 100% discrimination within large populations of varieties. In this study the molecular marker combination used was 100% discriminating within the variety set although failure to collect data for any one marker (such as might be the case if amplification failed) would reduce the discrimination to less than 100%.

Careful database interpretation can be used to identify admixture grains where the concentration of admixture is low. As level of admixture rises it rapidly becomes impossible to make any unambiguous statements about the genotype of the admixture, although some genotypes (varieties) can be excluded as possibilities). However, provided the varieties in the relevant population of varieties can be distinguished it should always be possible to detect and measure admixture.

It would be fair to say that this method is better at quantifying admixture than identifying varieties comprising that admixture. This may limit the technique to situations where a screen is required – such as at mill or malting intake – and where the contract specifies a variety and a maximum permitted admixture.

Using 96 grains in batches, quite fortuitously, gives a useful measurements in the range 0-10% admixture within which most grain contracts are set. The method becomes less accurate and the varietal identity of admixture much less certain as the level of admixture rises. However, this is not relevant to contracts which, once admixture is clearly 'too high' result in rejection: it is not really relevant whether the grain was rejected for 15% admixture or 25% admixture.

There is some risk to both buyer and seller from sampling errors. This is a familiar problem in grain trading which can be dealt with by setting production targets at a higher standard than the contract specifies. It is noted that buyers of large quantities of grain can afford to take an 'average view' and reduce the confidence with which any single purchase can be said to meet the contract specification. Evidence is presented that the statistical approach employed allows buyer and seller risk to be controlled in a predictable manner.

Finally, on the issue of cost. The majority of equipment available work in molecular biology is built around the 8 x 12 micro-titre plate. This format allows a set of 94 grains to be considered as 12 batches of 8, 8 batches of 12 or indeed a matrix. In this study the matrix approach has been developed most fully although the results from matrix analysis can be re-interpreted as simple batch analysis if required. The matrix approach does allow a more accurate estimate of admixture than simple batches – using the same number of grains. Most cost effective solutions are likely to be based around standardised equipment but batched grains reduce the variable costs of molecular biology reagents and clean-up columns which make up a considerable proportion of the total cost of molecular biology testing.

This approach is technically feasible, statistically robust and offers a reduced cost relative to a grain by grain approach. Accuracy is sufficient for decision making when purchasing grain against a maximum permitted admixture in the range 0-10%.

Some further effort is required to find two more molecular markers which offer discrimination and reliability. These are needed to replace BLYRCAB and Bmag 135 which initially appeared suitable but in practice proved problematic.

The authors also note that the approach can be applied to any batched analysis giving ‘attribute data’.

3.13. Acknowledgements

The NIAB gratefully acknowledges the financial support of HGCA in the conduct of this study.

The Authors wish to acknowledge the skilled assistance given by Mr Andrew Hazelden.

3.14. Bibliography

Becker J and Heun M. (1995) Barley microsatellites: allele variation and mapping. *Plant Molecular Biology* 27: 835-845.

Diethard Tautz (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463-6471.

Farid E. Ahmed (2002). Detection of genetically modified organisms in foods. *Trends in Biotechnology*, **20:***5*:215-223

International Seed Testing Association (1999). International Rules for Seed Testing. *Seed Science and Technology**, 27, Supplement, 41 pp.***

International Seed Testing Association (1999a). International Rules for Seed Testing. Rule 2.7.2. *Seed Science and Technology**, 27, Supplement, 18 pp.***

Liu, Z. W., Biyashev, R. M. and Maroof, M. A. (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. *Theoretical and Applied Genetics* 93: 869-876.

Macaulay M, Ramasy L, Powell W. and Waugh R. (2001). A representative, high informative 'genotyping set' of barley SSR'sSSRs. *Theoretical and Applied Genetics* 102: 801-809.

Ramsay L, Macaulay M, Ivanissevich S. degli, MacLean K., Cardle L., Fuller J., Edwards K. J., Tuvesson S., Morgante M., Massari A., Maestri E., Marmiroli N., Sjakste T., Ganal M., Powell W., and Waugh R. 2000. A simple sequence repeat-based linkage map of barley. *Genetics* 156: 1997-2005.

Rundle,S.J. and Zielinski,R.E. 1991 Organization and expression of two tandemly oriented genes encoding ribulosebisphosphate carboxylase/oxygenase activase in barley. Journal of Biological Chemistry 266: 4677-4685.

Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W. 1997 Discriminating between barley genotypes using microsatellite markers. Genome 40:442-450

Struss, D; Plieske, J. (1998). The use of microsatellite markers for detection of genetic diversity in barley populations. *Theoretical and Applied Genetics*. 97: 308-315.

Wetherill G. B. and Brown, D.W. (1991). Statistical Process Control. Theory and Practice. pp. 14. Chapman and Hall, London.

3.15. Appendixes

Appendix 1. Additional Statistical Tables

Table 32. Estimates of the proportion of the credible range below 7% admixture obtained using the matrix analysis method on an 8x12 micro-titre plate. Shaded area would be the zone of acceptance for P=0.05.

			No. of columns observed to contain admixture							
		0	1	2	3	4	5	6	7	8
No. of rows observed to contain admixture	0	0.999	-	-	-	-	-	-	-	-
	1	-	0.99	0.97	0.92	0.82	0.68	0.52	0.37	0.24
	2	-	0.97	0.97	0.91	0.80	0.66	0.49	0.33	0.20
	3	-	0.92	0.91	0.90	0.79	0.63	0.45	0.29	0.16
	4	-	0.82	0.80	0.79	0.77	0.60	0.41	0.25	0.13
	5	-	0.68	0.66	0.63	0.60	0.57	0.37	0.21	0.10
	6	-	0.52	0.49	0.45	0.41	0.37	0.33	0.18	0.08
	7	-	0.37	0.33	0.29	0.25	0.21	0.18	0.14	0.05
	8	-	0.24	0.20	0.16	0.13	0.10	0.08	0.05	0.03
	9	-	0.14	0.11	0.09	0.06	0.04	0.03	0.02	0.01
	10	-	0.08	0.06	0.04	0.03	0.02	0.01	0.00	0.00
	11	-	0.04	0.03	0.02	0.01	0.00	0.00	0.00	0.00
	12	-	0.02	0.01	0.01	0.00	0.00	0.00	0.00	-

Table 33. Estimates of the proportion of the credible range below 10% admixture obtained using the matrix analysis method on an 8x12 micro-titre plate. Shaded area would be the zone of acceptance for P=0.05.

			No. of columns observed to contain admixture							
		0	1	2	3	4	5	6	7	8
No. of rows observed to contain admixture	0	0.999	-	-	-	-	-	-	-	-
	1	-	0.99	0.99	0.99	0.97	0.93	0.86	0.76	0.64
	2	-	0.99	0.99	0.99	0.97	0.92	0.84	0.72	0.59
	3	-	0.99	0.99	0.99	0.96	0.91	0.81	0.68	0.53
	4	-	0.97	0.97	0.96	0.95	0.89	0.78	0.63	0.46
	5	-	0.93	0.92	0.91	0.89	0.87	0.74	0.57	0.39
	6	-	0.86	0.84	0.81	0.78	0.74	0.70	0.51	0.32
	7	-	0.76	0.72	0.68	0.63	0.57	0.51	0.44	0.25
	8	-	0.64	0.59	0.53	0.46	0.39	0.32	0.25	0.18
	9	-	0.50	0.44	0.37	0.30	0.23	0.17	0.11	0.06
	10	-	0.37	0.31	0.24	0.18	0.12	0.07	0.04	0.01
	11	-	0.26	0.20	0.14	0.09	0.05	0.02	0.01	0.00
	12	-	0.17	0.12	0.07	0.00	0.00	0.00	0.00	-