Theor Appl Genet (2011) 122:819–829
DOI 10.1007/s00122-010-1489-0

ORIGINAL PAPER

A flexible quantitative methodology for the analysis of gene-flow between conventionally bred maize populations using microsatellite markers

P. R. H. Robson · R. Kelly · E. F. Jensen · G. D. Giddings · M. Leitch · C. Davey · A. P. Gay · G. Jenkins · H. Thomas · I. S. Donnison

Received: 18 December 2009 / Accepted: 4 November 2010 / Published online: 26 November 2010

Abstract Previous studies of gene-flow in agriculture have used a range of physical and biochemical markers, including transgenes. However, physical and biochemical markers are not available for all commercial varieties, and transgenes are difficult to use when trying to estimate gene flow in the field where the use of transgenes is often restricted. Here, we demonstrate the use of simple sequence repeat microsatellite markers (SSRs) to study gene flow in maize. Developing the first quantitative analysis of pooled SSR samples resulted in a high sampling efficiency which minimised the use of resources and greatly enhanced the possibility of hybrid detection. We were able to quantitatively distinguish hybrids in pools of ten samples from non-hybrid parental lines in all of the 24 pair-wise combinations of commercial varieties tested. The technique was used to determine gene flow in field studies, from which a simple model describing gene flow in maize was developed.

Abbreviations

ANOVA	Analysis of variance
PAGE	Polyacrylamide gel electrophoresis
SSR	Simple sequence repeat

Communicated by J.-B. Veyrieras.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1489-0) contains supplementary material, which is available to authorized users.

P. R. H. Robson (✉) · R. Kelly · E. F. Jensen · G. D. Giddings · M. Leitch · C. Davey · A. P. Gay · G. Jenkins · H. Thomas · I. S. Donnison
IBERS, Aberystwyth University, Aberystwyth, UK
e-mail: ppr@aber.ac.uk

Introduction

Cultivation of transgenic crops has been approved worldwide. In some areas, regulatory requirements and adverse public perception have resulted in demand for segregation between transgenic and conventionally bred crops. Effective segregation requires prevention of gene flow between sexually compatible crops. Studies of the likelihood of gene flow at a particular geographical location require possible contaminating crops to contain a genetic or physical marker which allows for the identification of hybrids. A common marker used in gene flow studies has been a transgene conferring resistance to herbicide, which provides a readily assayable difference between often very genetically similar varieties (Weekes et al. 2008). The presence of transgenic markers can also be screened using event-specific real-time PCR (e.g. Hernández et al. 2004; Yang et al. 2006). However, use of a transgenic marker has limitations, since often the regulatory burden on experiments is greater and participation by farms may be reduced due to perceived adverse public reaction and/or direct action by protest groups.

Non-transgenic markers such as kernel colour have successfully been used to determine separation distances (Halsey et al. 2005), rates of gene flow within fields (Bannert et al. 2008) and in combination with transgenes to validate transgene detection methods (Pla et al. 2006). However, physical markers such as kernel colour or waxiness are rarely available in commercial varieties and studies using them do not therefore reflect the range of performance of current varieties under current agricultural practices. Transferring such traits into suitable varieties would involve costly and time-consuming introgression. It would clearly be beneficial if gene flow studies could reflect the potential impact of exchanging conventionally

bred for transgenic crops without impacting upon existing agriculture or having to employ transgenic crops as part of the methodology.

Microsatellites or simple sequence repeats (SSRs) are polymorphic markers present throughout eukaryote genomes (Tautz and Renz 1984). Repeat units may comprise single or multiple nucleotides, with the number of repeat units at a locus varying between genotypes. SSR loci display high mutation rates of 10^{-2}–10^{-3} per locus per gamete per generation (Weber and Wong 1993). Both genetic and environmental effects on SSR polymorphism rates have been reported in plants (e.g. Li et al. 2002, 2003), which may vary for different loci (Wierdl et al. 1997), repeat units (Chakraborty et al. 1997) and SSR lengths. The predominant mechanism by which variation in repeat unit numbers is generated is thought to be through DNA slippage (Levinson and Gutman 1987), although other mechanisms such as unequal crossover and sister chromatid exchange may also contribute (Wolff et al. 1991). DNA slippage generates a loop structure which is recognised and removed by the mismatch repair system. SSRs have been used successfully in evolutionary studies and also to generate genetic maps for a number of plants, including maize (Senior et al. 1996; Sharopova et al. 2002). There are publicly available databases of SSRs for a number of species including maize (Lawrence et al. 2005).

Despite extensive use of SSRs for the identification of hybrids, their use in gene-flow studies in agricultural populations has been limited. SSRs have been used in a study of geneflow to isolated male sterile oilseed rape plants (Devaux et al. 2007) and have been extensively applied to population studies of trees (e.g. Morand et al. 2002; Gérard et al. 2006). We describe the use of SSRs to determine gene-flow in agricultural populations of conventional maize crops. Maize is a suitable crop for this study because it is a major global crop (USDA 2010) and as with many intensively cultivated crops there are concerns regarding geneflow from transgenic maize to non-transgenic maize including to landraces (Pineyro-Nelson et al. 2009). Besides, there are extensive genetic resources available to studies in maize including a large number of SSR markers (Lawrence et al. 2005). The approach does not require transgenic crops to be cultivated prior to risk assessment, permits minimal disruption to existing cultivation practices and is applicable to a very wide range of cultivars. We have used a screen of publicly available microsatellite markers to distinguish between conventional varieties of maize. We used hybrids generated under glasshouse conditions to test the sensitivity of the analysis performed on field samples. A quantitative PCR-based assay of complex microsatellite amplifications was used to estimate the proportion of hybrids within pooled samples. Using pooled samples allowed us to increase sampling tenfold, while minimising costs and increasing the sensitivity of hybrid detection.

Materials and methods

Plant growth conditions

Plants (variety details listed in Suppl. 1) were either grown in a polytunnel, or, when temperatures and light levels were low, in a glasshouse at 25°C supplemented with a 16-h photoperiod (high pressure sodium son-T plus lights). There was no observable difference in fertility between plants grown under the two conditions. Seed was sown into John Innes number 2 compost. For seedling analysis, approximately 50–60 seeds were sown into trays (37 × 23.5 cm). To obtain fertile plants, seeds were sown individually into 10-l pots. Receptive silks were bagged to reduce the likelihood of polycross progeny and artificially cross-pollinated using manually harvested pollen.

Genomic DNA extraction

A bespoke leaf disc harvester was used to sample standardised areas of leaf tissue of 28 mm^2. When individual plants were sampled, genomic DNA was extracted from 20 leaf discs per plant. Pools consisted of 20 leaf discs, two from each of ten plants. Hybrid and non-hybrid leaf discs were mixed to simulate the occurrence of between one and nine hybrid events per pool. Genomic DNA was extracted using an Autogenprep 740 (Autogen, Holliston, Massachusetts, USA) and manufacturer-recommended conditions, with one exception that DNA was re-suspended in 200 µl. Following extraction, 2 µl of genomic DNA was analysed on an agarose gel to test for the presence of intact high Mw DNA.

Microsatellite (SSR) analysis

Previously identified, publicly available, primers were used to amplify SSR markers (Lawrence et al. 2005). To reduce potential bias, a panel of 70 SSR markers were chosen on the basis of a range of repeat sizes and base compositions. Taq polymerase (Roche diagnostics) and manufacturer-recommended PCR conditions were used to amplify both unlabelled and fluorescent-labelled products from 2 µl of genomic DNA. Unlabelled amplification products were analysed using polyacrylamide gel electrophoresis (Bert et al. 1999). SSR products were visualised using a silver nitrate staining method (Tixier et al. 1997). Fluorescent-labelled amplification products were analysed on an ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City,

USA). The resulting chromatograms were analysed and quantified using ABI Genemapper software.

Identification of suitable SSR markers for genomic fingerprinting of commercial varieties of maize

A total of 24 pair-wise combinations of commercial maize cultivars, listed in Tables 1 and 2, were analysed. The collection was split into two groups, of either pollen donors or recipients, according to how they were used in our gene flow studies (representing the transgenic and non-transgenic crops, respectively). Cultivar NK Bull was a donor to 11 recipient cultivars (Table 1); other donor-recipient combinations are shown in Table 2.

The suitability of individual microsatellite products to act as markers of gene flow was determined as follows: Varietal uniformity of microsatellite amplifications was tested using 16 independent genomic extractions from each variety. Donor varieties were selfed and 48 progeny screened for segregation of the SSR amplification product.

Markers were tested for biased amplification of genomic DNA in mixtures. Some SSR markers displayed preferential amplification bias in mixtures of different genotypes, which could make amplification of SSRs from rare hybrid events difficult to detect. Genomic DNA extractions from donor and recipient varieties were mixed in ratios of 1:9, 5:5, and 9:1. Any SSR markers displaying preferential amplification bias in these mixtures were excluded from further study.

Table 1 Parameter values, standard errors and goodness of fit (R^2) of the fitted polynomial calibration curves for NK-Bull as the pollen donor

Maternal parent	Parameters			Standard errors			R^2	SSR code
	a	b	c	a	b	c		
Agreement	0.59	33.90	55.28	2.51	13.11	13.11	0.9649	umc2068
Cameron	−17.19	170.30	−85.11	8.60	46.06	46.26	0.6328	umc2025
Justina	1.67	−12.48	100.40	3.76	19.67	19.44	0.9118	umc2068
Vogue	6.45	−30.86	108.30	4.01	21.67	21.96	0.8451	umc2025
Meribel	0.196	70.72		3.395	6.605		0.7351	umc2025
Nescio	2.86	9.08	103.90	2.89	15.20	15.27	0.9631	umc2068
Gazelle	−4.27	64.61		1.54	2.90		0.9100	umc2025
Fjord	−1.24	29.31	20.64	2.58	13.78	14.20	0.8316	umc2025
Axxor	4.19	−7.95	119.80	4.88	25.22	25.21	0.8857	umc2025
Baluga	3.28	13.76	87.97	3.03	16.30	17.03	0.9406	umc2068
Sparticus	9.60	−44.03	115.00	3.05	16.76	17.35	0.8874	umc2068
Candir	−8.17	84.65		2.40	4.33		0.9185	umc2068

Table 2 Parameter values, standard errors and goodness of fit (R^2) of the fitted polynomial calibration curves for pollen donors other than NK-Bull

Pollen donor	Maternal parent	Parameters			Standard errors			R^2	SSR
		a	b	c	a	b	c		Code
Fabius	Vernal	−1.157	97.51		2.639	4.613		0.8994	umc2068
	Crescendo	6.36	37.67	90.48	5.12	30.59	35.53	0.8991	umc2068
Crescendo	Vernal	−11.02	87.19		3.57	6.13		0.8283	umc2025
Diplomat	Caruso	10.28	7.53	73.38	3.02	15.40	15.48	0.9220	umc2068
Gazelle	Antares	−2.29	72.00		2.40	4.97		0.7809	umc2025
Justina	NK-Bull	1.96	28.18	69.19	4.31	22.09	21.69	0.9050	umc2025
	Gazelle	−0.03	79.59	−24.23	5.20	26.50	26.46	0.6549	umc2025
Caruso	Diplomat	0.18	55.69	43.48	4.11	21.18	21.16	0.9064	umc2047
	Meribel	2.14	107.70	−41.90	4.24	20.88	20.70	0.7299	umc1818
Meribel	Caruso	−20.92	270.70		6.754	11.89		0.9024	umc1818
Vernal	Fabius	3.27	8.68	79.08	2.95	15.90	15.81	0.9414	umc2068
Vogue	Caruso	2.95	26.28	99.19	3.12	16.25	16.28	0.9610	umc2068
Pendove	Vogue	−3.05	93.88	3.20	4.85	25.94	25.51	0.8800	ucm2047

Table 3 Details of the subset of SSRs used in calibration curves (data from Maize Genetics and Genomics Database (Lawrence et al. 2005))

SSR code	Number of alleles	GenBank code	Repeat unit	Length
umc2025	4	BE453974	AGCT	16
umc2047	2	BE575311	GACT	16
umc2068	3	BE511684	CCG	15
umc2067	4	AJ251018	CATG	20
umc1818	4	AF148498	CAG	18

Markers were tested to confirm that SSR amplifications were sufficiently sensitive to detect a single hybrid in a pool of ten non-hybrids (2 leaf discs in 20).

SSR markers meeting all the above criteria were tested again against 16 independent genomic extractions per variety using fluorescently labelled SSRs. This checked for the presence of very low amplification products, not discernable on PAGE gels, which co-migrate with SSR products. Such products may be confused with SSR amplifications from low product concentration generated from pools containing single hybrid events and thus were excluded from further study. SSR markers meeting all criteria were chosen for further study (Table 3).

Quantitative analysis of hybrid proportions for complex SSR marker amplifications

Hybrid and non-hybrid leaf discs were mixed to simulate ratios of hybrid and non-hybrid samples that might be expected in field samples (1:9, 2:8, 3:7, 4:6, 5:5, 7:3, and 9:1). DNA was extracted as above and analysed using amplification of fluorescent-labelled SSRs. SSR amplifications are complex and consequently not amenable to real-time quantitative PCR. Output from the ABI 3100 was analysed using ABI Genemapper software (Applied Biosystems, Foster City, USA) which generates peak heights of fluorescence for amplification products resolved by size. Peak heights from amplification products indicative of hybrids were expressed relative to those for amplification products common to both maternal and paternal varieties. The relationships between peak height ratios and the proportion of hybrid kernels per pool were determined using first- or second-order polynomial curve fits. At least eight replicate mixtures were analysed per ratio. These relationships were used to determine the number of hybrid kernels per pool in the field data.

Field sampling

The quantitative method was validated using gene flow data collected in 2003 from fields of the Jealott's Hill Research Station, Berkshire, UK, and Birchenfields Farm, Chester, Clwyd, UK. The field layouts are shown in Figs. 1 and 2. At Jealott's Hill a central 0.4 Ha plot of variety Gazelle was sown into a 7.1 Ha field of variety Antares. Samples were analysed from distances of 1, 10, 40 and 70 m along transects in seven directions roughly at 45° intervals from North (except to the South West). At each direction and distance 12 cobs were sampled with three replicate measures of 10 kernels sampled per cob (total of 360 kernels per sample distance; 10,080 kernels analysed overall). In Chester the source was a 0.4 Ha plot of cultivar NK-Bull, sown into a 6.4 Ha field of cultivar Vogue. Samples were analysed at 0, 2, 5, 10 and 15 m from the source, along transects at 45° angles from North. At each direction and distance 6 cobs were sampled, with three replicate measures of 10 kernels sampled per cob (total of 180 kernels per sample distance; 7,200 kernels analysed overall).

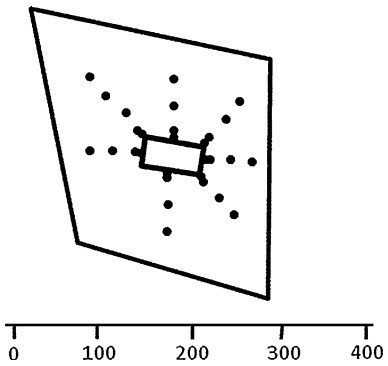


Fig. 1 Layout of the Jealott's Hill field in 2003. *Solid block* represents the pollen source and the *black circles* show the positions of sample collection (scale is in m)

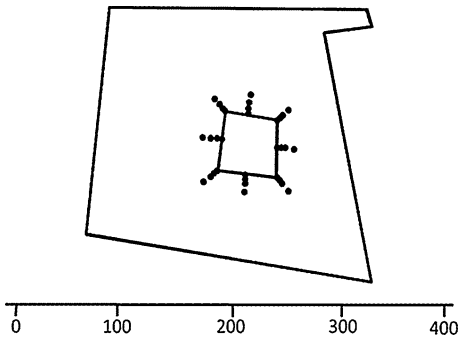


Fig. 2 Layout of the Chester field in 2005. *Block* represents the pollen source and the *black circles* show the positions of sample collection (scale is in m)

Statistical analysis and modelling of the gene flow data

The relationship between gene-flow and distance and direction was analysed. Analyses of variance were used to determine if there was a significant effect of distance and/

or direction. As the data are proportions based on counts, the error distribution was explored prior to curve fitting. It was not normal or the same for all Y (proportion positive pools) and Maximum Likelihood Program software (Ross 1987) was used to fit models with binomial errors.

The gene-flow model fitted was as follows: The data were in the form of proportions (of positive pools or cross pollinated kernels, C). If S was the amount of pollen from the source (donor) plots and R the amount of pollen from the fields containing the recipient (test) plants then $C = S/(R + S)$, assuming that both sources of pollen were equally capable of pollinating kernels. R was assumed to be constant and S a simple exponential decay function of distance D so that $S = S_0 e^{-rD}$ where S_0 was S at distance zero and r was the rate of decay of pollen deposition with distance. Then $C = S_0 e^{-rD}/(R + S_0 e^{-rD})$, and $C = S_0/(S_0 + R)$ when $D = 0$.

To deal with the problem of estimating the parameters of a proportion, with a very large number of possible solutions (i.e. of C) S_0 and R were replaced by a single parameter expressing their relative sizes, by multiplying both the top and bottom of the right side by $1/R$. Then, after some simplification,

$$C = \frac{S_{0R} \cdot e^{-r.D}}{1 + S_{0R}.e^{-r.D}}$$

where $S_{0R} = S_0/R$, with $C = S_{0R}/1 + S_{0R}$ when $D = 0$, and $S_{0R} = C/(1-C)$. The equation was uniquely defined for each and any value of C. S_0 and R were aliased (described together in the same variable) and could not be uniquely determined.

Results

SSR amplification to assay gene flow in commercial maize varieties

Genetic fingerprinting using SSR analysis was used to distinguish donor and recipient maize from commercially available varieties. In all cases tested it was possible to distinguish hybrid genomic DNA from non-hybrid maternal genomic DNA using SSR amplification and to detect a single hybrid event in a pool of 10. This result was achieved using paternal variety NK Bull and 12 other randomly chosen commercial varieties, and in a further 12 combinations of maternal and other paternal varieties.

For the majority of varieties, uniformity was observed in the amplification products. Some exceptions were noted, for example the population of variety Crescendo segregates for most microsatellites indicating the presence of three genomes in this variety. Several SSRs displayed biased amplification, and were rejected from further study.

Calibration

In all cases tested the fluorescence peak height ratios for the diagnostic SSR markers show quantitative relationships according to the relative amounts of hybrid events in the DNA samples (Fig. 3a pollen donor NK-Bull, Fig. 3b pollen donors other than NK Bull, with parameters in Tables 1 and 2, respectively). For all comparisons the data were adequately modelled by linear or quadratic polynomials and Run's tests showed that the residuals were well distributed and that there was no significant deviation from the model ($P = 0.5609$ and 0.0622, respectively) (Motulsky and Christopoulos 2003). Thus, the number of hybrids present in the pools of ten was quantitatively determined in all maternal varieties tested using the relationships given.

Jealott's Hill, 2003 field data

Most pools returned a zero gene flow score and the majority of positive pools occurred within 10 m (Fig. 4). Gene flow was detected at the farthest sample points in the north-easterly transect which correlates approximately with the largest integral of wind strength and direction during the pollination period (meteorological data not shown) (Fig. 5).

The proportion of positive pools was greater than the proportion of gene-flow, as might be expected (Fig. 4). Statistical analysis (ANOVA) indicated that there was a significant effect of both distance and direction and also a significant interaction between the two (Table 4). This was true for both the proportion of pools containing hybrid kernels and the proportion of hybrid kernels estimated from the calibration. Most of the variation was due to the differences between the proportions of positive pools or hybrid kernels with distance (84.68 and 67.97% of the total variance, respectively). The difference between data from different directions was bigger for the proportion of kernels than for the proportion of positive pools (5.97% compared with 2.92%). Since the direction effect was significant it was not appropriate to fit one model to the combined data. Fits to data from the north and south were not appropriate since hybridisation was only detected at one distance in each case. Some of the models for other data sets also have large errors relative to the parameter values because there are few data values much above zero. For the north-west there were two non-zero values, one close to zero (at 10 m the proportion of positive pools was only 0.028, $s = 0.048$).

Figures 6 and 7 show the variation with distance and direction in the proportions of positive pools and hybrid kernels, respectively. The overall relationship between the gene-flow, distance and direction is similar for both measures. It was not possible to fit either parameter

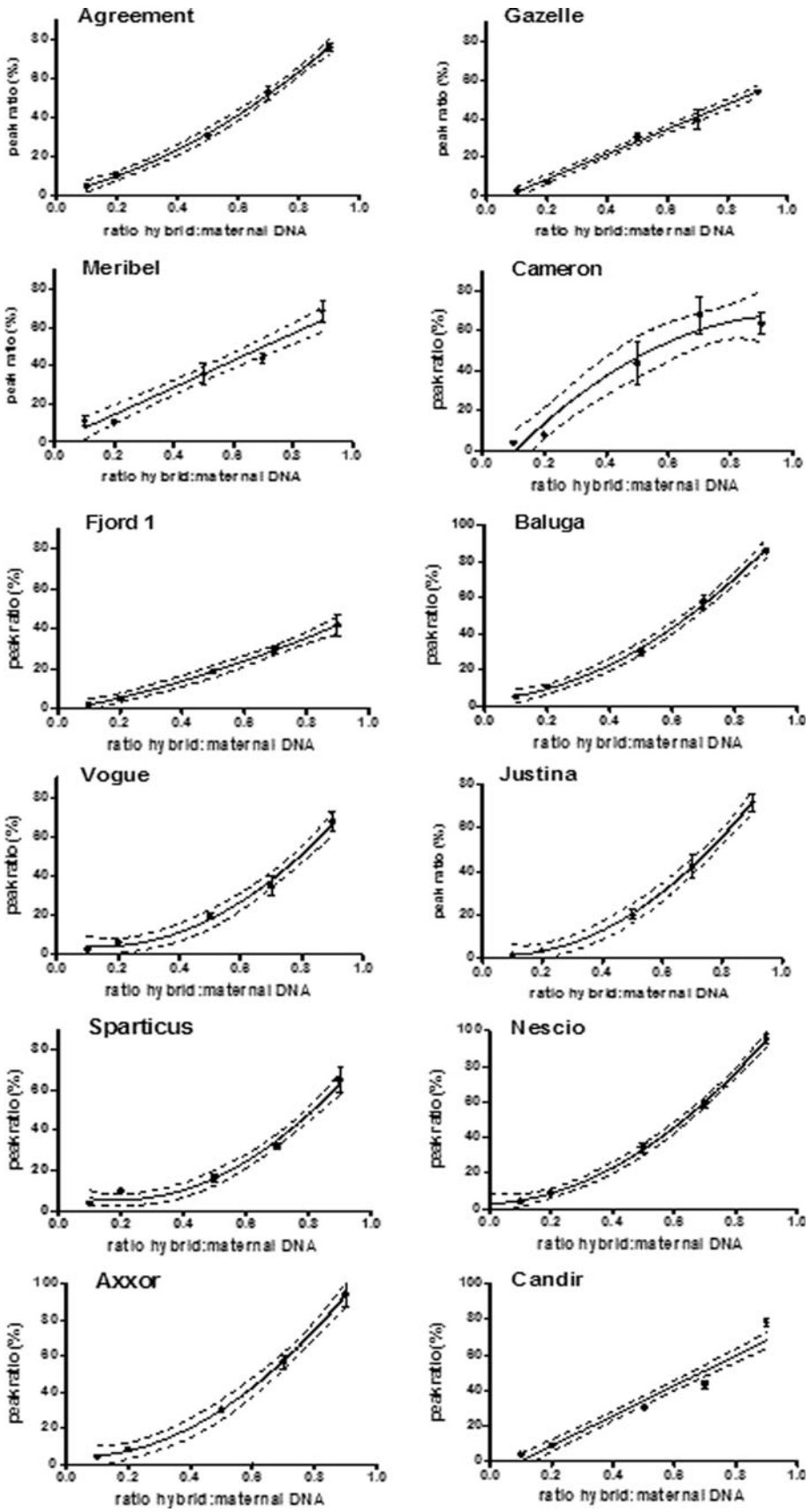
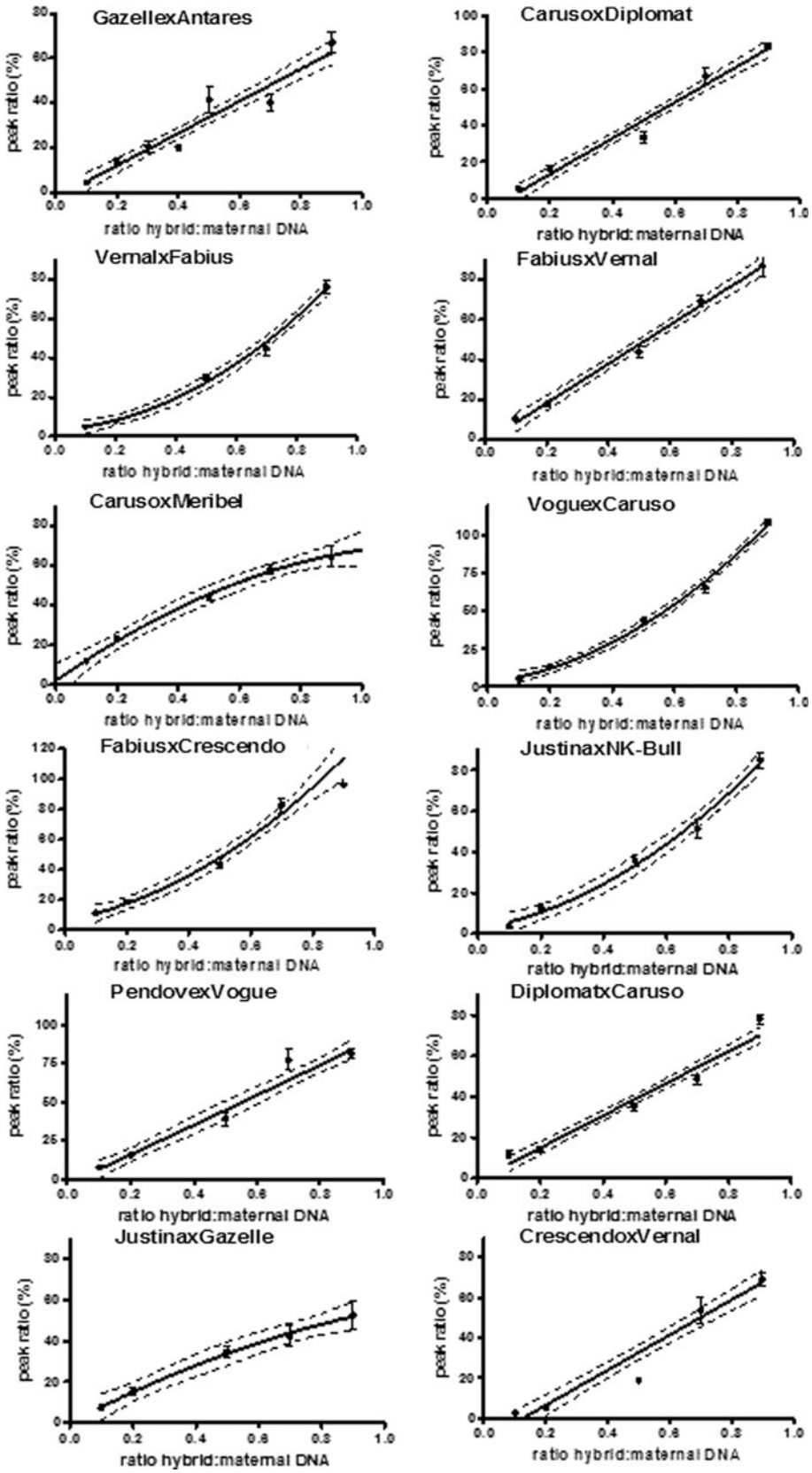


Fig. 3 **a** Calibration curves for NK-Bull as the pollen donor. The maternal parent is indicated for each curve. The *error bars* show the standard errors for each data point (between 8 and 13 replicates per point). The *dotted lines* show the 95% confidence intervals of the fitted curves, which are either first- or second-order polynomials (i.e. empirical). The percentage peak ratios (ratios of peak heights) are for the particular diagnostic SSR fragments. **b** Calibration curves for crosses involved in the gene-flow studies. The maternal parent is indicated first for each curve and the paternal parent second. The *error bars* show the standard errors for each data point (between 8 and 13 replicates per point). The *dotted lines* show the 95% confidence intervals of the fitted curves, which are either first or second order polynomials (i.e. empirical). The percentage peak ratios (ratios of peak heights) are for the particular diagnostic SSR fragments

Fig. 3 continued



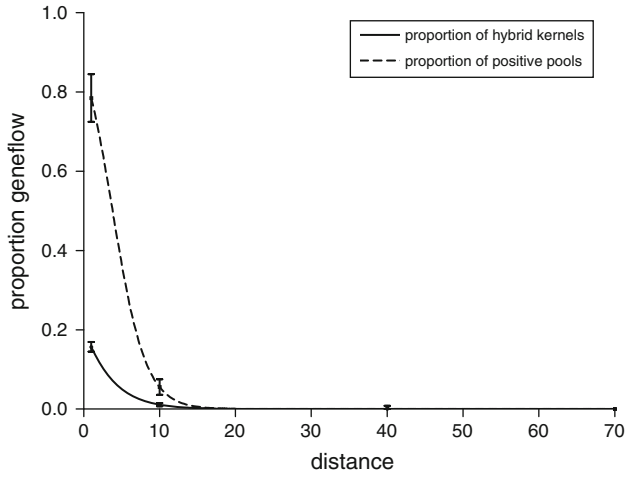


Fig. 4 The variation in the proportion of positive pools and hybrid kernels over distance (m), along with fitted curves, for the data from Jealott's Hill in 2003. *Error bars* show the standard errors of the observed means

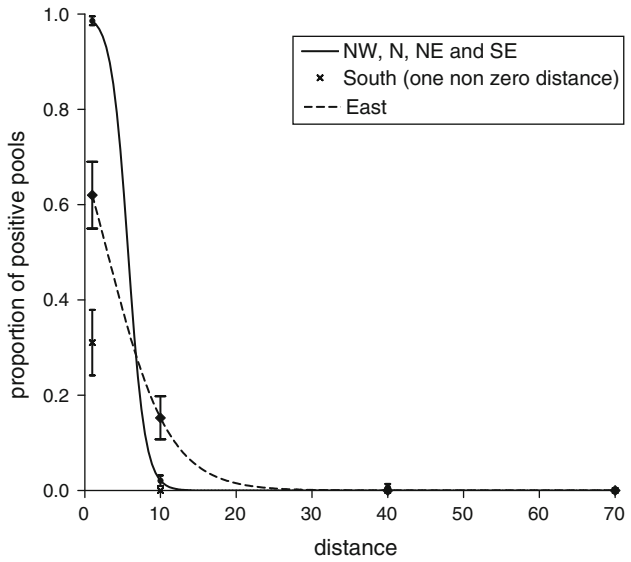


Fig. 5 The variation in the proportion of positive pools over distance (m) for the data from Jealott's Hill in 2003, along with fitted curves. *Error bars* show the standard errors of the observed means

simultaneously to the data from all directions (for S_{0R} $F = 8.612$, $df = 4{,}50$, $P < 0.0001$, for r $F = 5.559$, $df = 4{,}50$, $P = 0.0009$). The same curve fitted both the north-east and south-east data even when both curves were fitted with no constraints on the parameters (for positive pools $F = 6.8 \times 10^{-15}$, $df = 2.20$, $P = 1$). For data from the north-west, north, north-east and south-east one curve could be fitted to all without a loss of fit ($F = 0.8606$, $df = 6.40$ and $P = 0.5319$ for positive pools; $F = 0.8773$, $df = 6.40$ and $P = 0.5202$ for hybrid kernels). This was true also for the data from the east and west ($F = 1.975$, $df = 2.20$, $P = 0.1649$ in both cases; model parameters in Table 5).

Chester 2005 field data

Most gene-flow occurred within 5 m of the source (Figs. 7, 8). To the South there were quite a high proportion of pools containing hybrid kernels right up to 15 m, but these translated into low gene-flow levels when calibrated into proportions of hybrid kernels. Similarly a relatively high proportion of positive pools at 5 m to the East turned out to represent low levels of gene-flow after calibration. Statistical analysis (ANOVA) indicated that, for both the proportion of pools containing hybrid kernels and the proportion of hybrid kernels, there was a significant effect of both distance and direction, and also a significant interaction between the two (Table 6).

It was necessary to fit curves individually to data from the east and south-west (Figs. 7, 8; Table 7). For data from the north-east, south-east, west and north-west, one curve could be fitted to all without a loss of fit ($F = 0.1451$, $df = 3{,}100$ and $P = 0.9326$ for positive pools, $F = 0.1640$, $df = 3{,}100$, $P = 0.9203$ for hybrid kernels). This was true also for the data from the north and south ($F = 1.53$, $df = 1{,}50$ and $P = 0.2223$), but only after calibration (for positive pools $F = 4.74$, $df = 1{,}50$, and $P = 0.0343$). Fitting data to the northerly positive pools data alone was inappropriate due to there being only two non-zero values, with one of those close to zero. For the southerly positive pools data the large amount of variation in the curve tail means that the fit to the model also is not good, as shown by the relatively large standard errors (Table 7). After calibration, models can be fitted with the same rate of decline of gene-flow (r) in all directions ($F = 0.06929$, $df = 3.112$, P = 0.9762). By comparing the parameter value standard errors (Table 7), it can be seen that calibration has much improved the ability to fit gene-flow models to the data from all directions.

Table 4 Analysis of variance of gene-flow over distance and direction for the Jealott's Hill 2003 data

Source of variation	*df*	Proportions of positive pools			Proportions of hybrid kernels		
		F	*P* value	% of total variation	*F*	*P* value	Percentage of total variation
Interaction	18	15.18	<0.0001	10.29	13.44	<0.0001	19.48
Direction	6	12.95	<0.0001	2.92	12.35	<0.0001	5.97
Distance	3	749.8	<0.0001	84.68	280.6	<0.0001	67.79

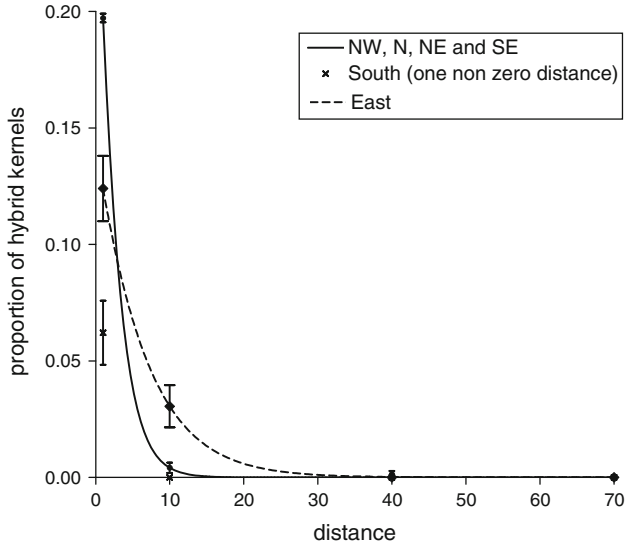


Fig. 6 The variation in the proportion of hybrid kernels over distance (m) for the data from Jealott's Hill in 2003, along with fitted curves. *Error bars* show the standard errors of the observed means

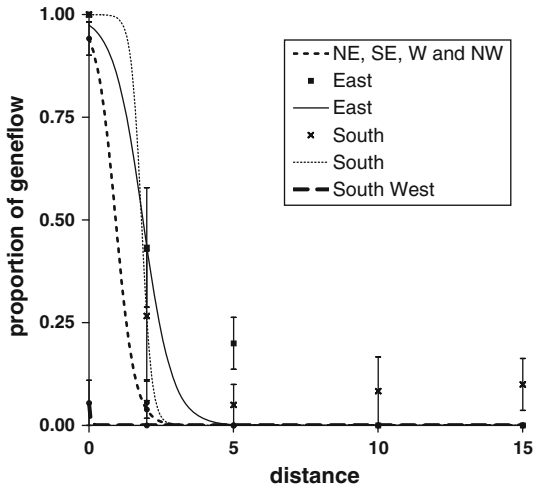


Fig. 7 The variation in the proportion of positive pools and hybrid kernels over distance (m), along with fitted curves, for the data from Chester in 2005. Error bars show the standard errors of the observed means

Discussion

We have demonstrated the use of SSR markers for detecting and measuring gene-flow between numerous pairs of commercial cultivars of maize. The method involves the bulking of samples into pools and the subsequent quantification of the proportion of hybrids in each pool. This allows greater detection sensitivity since larger numbers of samples can be checked faster and more economically than if each sample were typed individually. The use of commercially available varieties allows for a more realistic test of gene-flow in agriculture than would otherwise be available from the use of phenotypic markers, which tend to occur in novelty and other non-commercial varieties. It also allows for the testing of likely gene-flow between transgenic and conventionally bred cultivars without actually having to use transgenics in risk assessment studies, thus avoiding the inevitable constraints and problems this can incur. Our approach has been used in studies of gene-flow between agricultural maize at several different sites, two of which are presented here. In landscape studies it can be used to track gene-flow between a patchwork of fields of different varieties. It may also be used to provide quality assurance in other gene-flow studies, by confirming the genetic provenance of hybrids identified by other means.

The commercial varieties used in this study were all F1 hybrids; however, commercially sensitive parental lines were unavailable and it was necessary to self the donors to test for segregation of the SSR products. This ensured that screening for one or a combination of markers would capture all of the possible gene-flow events derived from a donor variety. In addition, the SSR product was screened for preferential amplification from the genomic DNA of one variety. Amplifications were performed with mixtures of genomic DNA from each variety to test for amplification bias; <1% displayed bias and were rejected from further study. Hybrids between donors and recipients were generated. Such hybrids served two purposes. First, they enabled tests of the sensitivity of the assay, and proved that the analysis was sufficiently robust to detect single hybrid events in pooled samples. Second, the population of hybrid mixes representing different proportions of gene-flow events enabled quantification of gene-flow. In all variety combinations tested SSRs were identified which met the quality assurance thresholds described. None of the SSR combinations chosen were unable to detect a single hybrid in a pool of 10, suggesting that the methodology applied was within acceptable sensitivity limits. It has been

Table 5 The parameter values and their standard errors (SE) for models fitted to the different directions of the Jealott's Hill 2003 data

Direction	Proportions of positive pools				Proportions of hybrid kernels			
	S_{OR}	SE	r	SE	S_{OR}	SE	r	SE
NW, N, NE, and SE	174.9	111.8	0.9014	0.07685	0.3863	0.01682	0.4525	0.04214
E and W	2.083	0.3973	0.2447	0.03896	0.1673	0.0145	0.1669	0.03092

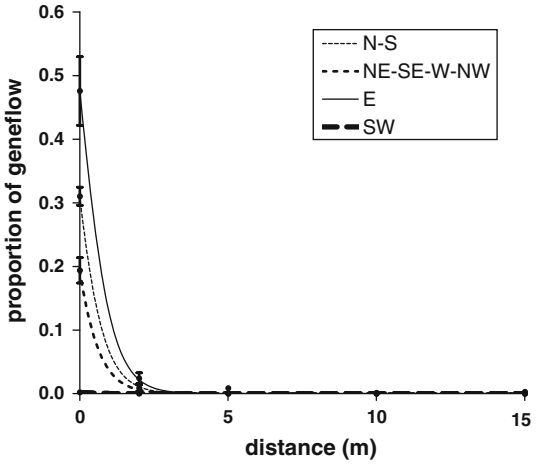


Fig. 8 The variation in the proportion of hybrid kernels over distance up to 15 m for the data from Chester in 2005, along with fitted curves. *Error bars* show the standard errors of the observed means for sampling transect directions: N–S, NE-SE-W-NW, E, and SW

previously reported that bulked samples sometimes result in "stuttered" SSR amplification products (Dubreuil et al. 2006); this was not seen in our experiments which may be due to the smaller pool size.

SSR comparisons between a single pollen donor (variety NK Bull) and 12 potential pollen recipients demonstrated that a single variety may be tracked using SSRs in a field-scale environment potentially containing many different varieties. Comparisons between multiple varieties and multiple donors demonstrated the general applicability to field-scale studies between many different potential donors and recipients, including some fields containing more than one recipient. In all cases, it was possible to find SSR amplification products that met the quality criteria and were able to distinguish potential pollen donors and recipients.

The peak height ratios for the SSR markers were indicative of the proportion of hybrid events in the DNA samples. The variation in peak ratios (between different ratios of hybrid to maternal DNA) was without plateaux at either end of the spectrum. This means that there was good resolution for transforming the proportions of positive pools into the proportions of hybrid kernels. This study thus demonstrates the usefulness of pooling samples for gene-flow determination, allowing increased sampling without increasing the cost of the laboratory analysis.

A simple model was developed to describe the proportion of gene-flow detected. Fitting models to data was not always appropriate if there were only one or two non-zero distances. In all cases there were few non-zero values beyond 5 m, indicating the low levels of short-distance gene-flow observed. As a result, some of the model parameters have large errors relative to their means.

The field data, especially that from Chester, showed the benefit of calibrating pooled data. Pooling data to increase sampling is a good idea for helping to detect rare gene-flow events. However, calibration is necessary to determine actual gene-flow levels, and to allow the modelling of gene-flow over distance and direction. Although there was a significant effect of both distance and direction for both the calibrated and un-calibrated data, the details varied between the two. In particular, the difference between directions was bigger for the calibrated data than for the proportion of positive pools. Whereas the same curve fitted both the north-east and south-east data before calibration, afterwards only the rates of decline of gene flow with distance were not significantly different. A common *r* value for the calibrated north-east, south-east and north-west data

Table 6 Analysis of variance of gene-flow over distance and direction for the Chester 2005 data

Source of variation	*df*	Proportions of positive pools			Proportions of hybrid kernels		
		F	*P* value	Percentage of total variation	*F*	*P* value	Percentage of total variation
Interaction	28	6.030	<0.0001	11.48	14.65	<0.0001	19.27
Direction	7	11.38	<0.0001	5.42	17.58	<0.0001	5.78
Distance	4	255.5	<0.0001	69.50	348.8	<0.0001	65.55

Table 7 The parameter values and their standard errors (SE) for models fitted to the different directions of the Chester 2005 data

Direction	Proportions of positive pools				Proportions of hybrid kernels			
	S_{OR}	SE	*r*	SE	S_{OR}	SE	*r*	SE
N and S	28292	7.3895e7	5.631	1268	0.4496	0.00669	1.852	0.2556
NE, SE, W, and NW	16.14	5.928	2.994	0.3265	0.2403	0.00644	1.852	0.2556
E	38.32	135.2	1.950	1.651	0.9079	0.04153	1.852	0.2556
SW	0.05820	0.02603	19.66	8.684e+15	0.0020	0.00034	1.852	0.2556

indicated the same approximate rates of decline of gene flow. Similarly a rate of decline of gene flow was appropriate for the models of the westerly and easterly data, but only after calibration.

The flexibility of using SSRs allowed a wide range of maize varieties to be genetically distinguished in studies of gene flow in existing agricultural systems. We have demonstrated improvements in this technique by the quantitative association between SSR peak height and hybrid events in pooled samples. SSRs were successfully used as the primary screen for gene flow in this study, and this approach has been used in studies of gene flow between agricultural maize at several different sites of which two sites are presented here. However, in addition to using this approach as a primary screen it may be used also to provide quality assurance in other gene flow studies by confirming the genetic provenance of hybrids identified by some other means. This may also be used to distinguish the contributions of individual sources in landscape models where multiple sources are used with the same physical or chemical marker, but in backgrounds that can be genetically distinguished.

Acknowledgments This work was supported by UK Department of Environment Food and Rural Affairs grant (CTC0104) and by EU Framework VI programme grant Sustainable Introduction of GMOs into European Agriculture (SIGMEA). HT was supported by an emeritus fellowship from the Leverhulme Trust.

References

Bannert M, Vogler A, Stamp P (2008) Short-distance cross-pollination of maize in a small-field landscape as monitored by grain color markers. European J Agron 29:29–32

Bert PF, Charmet G, Sourdille P, Hayward MD, Balfourier F (1999) A high-density molecular map for ryegrass (*Lolium perenne*) using AFLP markers. Theor Appl Genet 99:445–452

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di- tri- and tetranucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041–1046

Devaux C, Lavigne C, Austerlitz F, Klein EK (2007) Modelling and estimating pollen movement in oilseed rape (*Brassica napus*) at the landscape scale using genetic markers. Mol Ecol 16:487–499

Dubreuil P, Warburton M, Chastanet M, Hoisington D, Charcosset A (2006) More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. Maydica 51:281–291

Gérard PR, Klein EK, Austerlitz F, Fernández-Majarrés JF, Frascaria-Lacoste N (2006) Assortive mating and differential male mating success in an ash hybrid zone population. BMC Evol Biol 6:96

Halsey ME, Redmund KM, Davis CA, Qualls M, Eppard PJ, Berberich SA (2005) Isolation of maize from pollen mediated gene flow by time and distance. Crop Sci 45:2172–2185

Hernández M, Duplan MN, Berthier G, Vaitilingom M, Hauser W (2004) Development and comparison of four real-time polymerase chain reaction systems for specific detection and quantification of *Zea mays* L. J Agric Food Chem 52:4632–4637

Lawrence CJ, Seigfried TE, Brendel V (2005) The Maize Genetics and Genomics Database. The community resource for access to diverse maize data. Plant Physiol 138:55–58

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Li YC, Röder MS, Fahima T, Kirzhner VM, Beiles A, Korol AB (2002) Climatic effects on microsatellite diversity in wild emmer wheat, *Triticum dicoccoides*, at Yehudiyya microsite. Heredity 89:127–132

Li YC, Fahima T, Röder MS, Kirzhner VM, Beiles A, Korol AB, Nevo E (2003) Genetic effects on microsatellite diversity in wild emmer wheat (*Triticum dicoccoides*) at the Yehudiyya microsite, Israel. Heredity 90:150–156

Morand M-E, Brachet S, Rossignol P, Dufour J, Frascaria-Lacoste N (2002) A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. Mol Ecol 11:377–385

Motulsky H, Christopoulos A (2003) Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting. GraphPad Software, Inc, San Diego

Pineyro-Nelson A, vanHeerwaarden J, Perales HR, Serratos-Hernández JA, Rangel A, Hufford MB, Gepts P, Garay-Arroyo A, Rivera-Bustamante R, Alvarez-Buylla ER (2009) Transgenes in mexican maize: molecular evidence and methodological considerations for GMO detection in landrace populations. Mol Ecol 18:750–761

Pla M, La Paz J-L, Penãs G, Garciá N, Palaudelmás M, Esteve T, Messeguer J, Melé E (2006) Assessment of real-time PCR based methods for quantification of pollen-mediated gene-flow from GM to conventional maize in a field study. Trans Res 15:219–228

Ross GJS (1987) Maximum likelihood program. The Numerical Algorithms Group Ltd. Rothampsted Experimental Station. Lawes Agricultural Trust, England

Senior ML, Chin ECL, Lee M, Smith JSC, Stuber CW (1996) Simple sequence repeat markers developed from maize sequences found in the GENBANK database: map construction. Crop Sci 36:1676–1683

Sharopova N, McMullen MD, Schultz L, Schroeder S, Sanchez-Villeda H, Gardiner J, Bergstrom D, Houchins K, Melia-Hancock S, Musket T, Duru N, Polacco M, Edwards K, Ruff T, Register JC, Brouwer C, Thompson R, Velasco R, Chin E, Lee M, Woodman-Clikeman W, Long MJ, Liscum E, Cone K, Davis G, Coe EH (2002) Development and mapping of SSR markers for maize. Plant Mol Biol 48:463–481

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl Acid Res 12:4127–4138

Tixier MH, Sourdille RM, Leroy P, Bernard M (1997) Detection of wheat microsatellites using a non radioactive silver-nitrate staining method. J Genet Breed 51:175–177

USDA (2010) World agricultural production. Circular series WAP 10–05. USDA. Foreign agricultural service, USA

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Weekes R, Allnutt T, Boffey C, Morgan S, Bilton M, Daniels R, Henry C (2008) A study of crop-to-crop gene flow using farm scale sites of fodder maize (*Zea mays* L.) in the UK. Transgenic Res 16:203–211

Wierdl M, Dominska M, Petes TD (1997) Microsatellite instabililty in yeast: dependence on the length of the microsatellite. Genetics 146:769–779

Wolff RK, Plaeke R, Jeffreys AJ, White R (1991) Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. Genomics 5:382–384

Yang L, Guo J, Pan A, Zhang H, Zhang K, Wang Z, Zhang D (2006) Event-specific quantitative detection of nine genetically modified maize using one novel standard reference molecule. J Agric Food Chem 55:15–24