J O U R N A L O F
Biochemistry
—— and ——
Molecular Biology

Different Real Time PCR Approaches for the Fine Quantification of SNP's Alleles in DNA Pools: Assays Development, Characterization and Pre-validation

Elia Mattarucchi[†], Milena Marsoni[†], Giorgio Binelli[‡], Alberto Passi[†], Francesco Lo Curto[†], Francesco Pasquali[†] and Giovanni Porta[†,*]

[†]Department of Experimental and Clinical Biomedical Sciences, University of Insubria. Via Dunant 5, 21100 Varese Italy
[‡]Department of Biotechnology and Molecular Sciences, University of Insubria, Via Dunant 3, 21100 Varese Italy

Single nucleotide polymorphisms (SNPs) are becoming the most common type of markers used in genetic analysis. In the present report a SNP has been chosen to test the applicability of Real Time PCR to discriminate and quantify SNPs alleles on DNA pools. Amplification Refractory Mutation System (ARMS) and Mismatch Amplification Mutation Assay (MAMA) has been applied. Each assay has been pre-validated testing specificity and performances (linearity, PCR efficiency, interference limit, limit of detection, limit of quantification, precision and accuracy). Both the approaches achieve a precise and accurate estimation of the allele frequencies on pooled DNA samples in the range from 5% to 95% and don't require standard curves or calibrators. The lowest measurement that could be significantly distinguished from the background noise has been determined around the 1% for both the approaches, allowing to extend the range of quantifications from 1% to 99%. Furthermore applicability of Real Time PCR assays for general diagnostic purposes is discussed.

Keywords: DNA pools genotyping, Pre-validation, Real Time PCR, Single nucleotide polymorphism, SNP

Introduction

Single nucleotide polymorphisms (SNPs) provide a powerful tool for a number of genetic investigations: they are the preferred choice for association studies due to their abundance and stability (Risch, 2000), they may be used to monitor minimal residual disease and post-transplantation chimerism (Fredriksson *et al.*, 2004) and even to identify chromosome abnormalities (Pont-Kingdon and Lyon, 2003). Association studies in non-Mendelian complex disease have emphasized cost benefits of allele frequency estimation in DNA pools (Norton, 2004) and raised the interest in reliable, very sensitive and high throughput screening assays such as Real Time PCR (Germer *et al.*, 2000; Elmaagacli *et al.*, 2002; Shifman *et al.*, 2002). This technique enables allele discrimination by either, allele-specific probes (Breen *et al.*, 2000) or allele-specific forward primers (Newton *et al.*, 1989; Breen *et al.*, 2001). When fine quantifications are required, allele-specific forward primers are preferred because of their specificity could be directly regulated by the annealing temperature of the reaction. In Amplification Refractory Mutation Systems (ARMS) (Newton *et al.*, 1989) the specificity of forward primers is given by the terminal 3' nucleotide; if this nucleotide mismatches the template the primer will not be extended. Discrimination power can be improved using allele-specific forward primers with a deliberate added mismatched nucleotide located near the 3' terminal region (Newton *et al.*, 1989; Glaab and Skopek, 1999). This system is named Mismatch Amplification Mutation Assay (MAMA). Allele-specific primers should be expected to amplify only their own allele. Nevertheless also the mismatched allele is amplified (aspecific amplification), but less efficiently. For this reason an accurate estimation of the background noise is a relevant part in testing quantitative assays. Allele quantification may be achieved by comparing two reactions, each specific for one of the two alleles (Germer *et al.*, 2000). Alternatively, one of the two allele-specific reactions may be normalized by a reference sequence (Schwarz, 2004). In both cases, the Comparative Threshold-cycle (Ct) Method could be applied (Livak and Schmittgen, 2001). Standard curves are not required if the reactions have similar efficiencies (Rutledge and Cote, 2003). Two reactions

*To whom correspondence should be addressed.
Tel: 39 0332 217106; Fax: 39 0332 217119
E-mail: giovanni.porta@uninsubria.it

with the same efficiency, may give different Ct values when the same amount of template is amplified. This is due to probe, primers and target differences (chemistry and sequence). In this case a sample at a known concentration (calibrator) may be used to adjust measurements. E.g. heterozygous and homozygous DNA are, by definition, a 50% and 100% calibrators (Alizadeh *et al.*, 2002).

A pre-validation process is needed to ensure that the method can be successfully transferred to other laboratories. With this aim the following performance parameters have been here evaluated: linearity (proportionality of the signal to the amount of DNA, tested by regression analysis), PCR efficiency (actual efficiency calculated from the slope of the regression line obtained to asses linearity), interference limit (lowest measurement that can be significantly distinguished from the background noise due to aspecific amplifications), limit of detection (lowest amount of target DNA that the tested assay can detect reliably, but not necessarily quantify), limit of quantification (lowest amount of target DNA that the tested assay can quantify with an acceptable level of accuracy and reliability), precision (degree of mutual agreement among quantifications of the same sample) and accuracy (closeness of agreement between the reported concentration and the correspondent nominal value of the reference sample measured) (Horwitz, 1995). The dynamic range is the interval of concentrations within which the analytical procedure has a suitable level of accuracy and precision. Nevertheless ARMS and MAMA real time PCR assays are becoming critical diagnostic tools (Punia *et al.*, 2004) rising the importance of pre-validation protocols. Performance parameters to ensure a solid pre-validation protocol have been here discussed.

Materials and Methods

Genetic marker selection and samples preparation A SNP at the *MetH locus* (GenBank access number M37520) on chromosome 7q31 showing a C/T transition in position 187, has been chosen as model in this study. The two alleles have been named here *MetHc* and *MetHt*. Human genomic DNA was isolated from peripheral blood cells of 10 healthy subjects using FlexiGene DNA kit (Qiagen S.p.a, Milano, Italy) and subsequently characterized by direct sequencing. An homozygote for each of the two alleles of the chosen polymorphism was selected. Sequencing was performed by the CRIBI Biotechnology Centre of the University of Padova; further technical details can be obtained at the CRIBI Website (http://bmr.cribi.unipd.it). Genomic DNA was accurately quantified using TaqMan® RNase P Control Reagents (Appliedbiosystems Foster City, USA) according to manufacturer's protocol and then diluted to 40 ng/µl with 10 mM Tris buffer pH 7.5; concentrations were then checked again. Equimolar DNA samples were used to establish the Real Time allele-specific assays (primers and probe set-up and optimisation). Subsequently, part of each homozygous DNA was used to obtain two serial dilutions curves; one (170, 17, 1.7 ng/µl and 0.17 ng/µl of DNA in 10 mM Tris buffer pH 7.5) to assess linearity and efficiency and a second curve (367, 183.5, 91.75, 45.87, 22.93, 11.47 pg/µl and 5.73 pg/µl of DNA in 10 mM Tris buffer pH 7.5) to evaluate limits of detection and quantification. A series of reference samples was produced diluting the remaining of the *MetHc* homozygous DNA in the *MetHt* according to the following volume/volume (v/v) percentages: 1, 5, 10, 20, 40% and 60%. A second series of reference samples has been produced diluting at the same concentrations the *MetHt* homozygous DNA in the *MetHc*. As the DNA solutions were equimolar, v/v percentages correspond to the percentage of one allele relatively to the other. Furthermore taking advantage from the biallelic system, the two series of reference samples cover a concentration range from 1% to 99% of both the alleles. Every dilution has been obtained mixing solutions at 4°C for at least 30 min; once equilibrated, all dilutions were subdivided in single-use aliquots and stored at –20°C to avoid repeated freezing and thawing cycles.

Primers and probes ARMS and MAMA allele-specific forward primers have been designed for each of the two *MetH* alleles. The specificity of the ARMS primers is given by the third last nucleotide at 3' region. An additional deliberate mismatch in fifth from the 3' end improve specificity of MAMA primers. A forward primer just outside the polymorphic region of the *MetH* (Tot primer) has been used to asses the total amount of the target sequence (Tot assay). All the established assays use the same probe and reverse primer and have been designed by Primer Express™ software version 2.0 (Applied Biosystems, Foster City, USA) on the complementary strand of the *MetH*. The Probe is a TaqMan® Minor Groove Binder (MGB) labelled at the 5' end with 6-carboxyfluorescein (FAM). Descriptions and sequences of primers and probe are listed in Table 1.

Table 1. Real Time PCR primers and probe. Polymorphic nucleotide in bold and introduced artificial mismatch base in MAMA primers are underlined.

Description	Name	Sequence (5'→3')
ARMS forward primer MetHc specific	ARMSc	CGT CAG AAA AAT CC**G** GG
ARMS forward primer MetHt specific	ARMSt	ATC GTC AGA AAA ATC **A**G G
MAMA forward primer MetHc specific	MAMAc	GAT CGT CAG AAA AAT <u>G</u>C**G** GG
MAMA forward primer MetHt specific	MAMAt	AAG ATC GTC AGA AAA AT<u>G</u> C**A**G G
Forward primer MetH total amount	Tot	CTA GCA GTT GAA AGA TCG TCA GAA AAA
Common reverse primer	REV	CCC GAA TCT CAG GAA GTC TCT GTC
Common probe	Allele3	6-FAM-AAT CAG CAA CTT AGA CCA AAC-MGB

Real time PCR assays Quantitative PCR reactions were performed on ABI Prism 7000 SDS and subsequently analysed using ABI Prism 7000 SDS software version 1.0 (Appliedbiosystems Foster City, USA). The reaction mixture contained: 12.5 µl of the TaqMan® Universal PCR MasterMix (Appliedbiosystems Foster City, USA); reverse primer 900 nM; probe 200 nM; DNA depending on the experiment and nuclease free water up to 25 µl. ARMS and MAMA forward primers and the Tot primer were used at 2500 nM, 3000 nM and 900 nM respectively. Reactions were prepared and run in triplicate on 96-well plates (Appliedbiosystems Foster City, USA). A threefold reaction mixture was thoroughly mixed by vortexing and then aliquoted in order to minimise the variability among wells. PCR thermal profile was: 2 min at 50ºC followed by 10 min at 95ºC and 40 amplification cycles (95ºC for 15 s and 61ºC for 60 s).

Linearity and efficiency Linearity of each assay was evaluated by the coefficient of determination (R^2) of the regression curve obtained by plotting the Ct values of 170, 17, 1.7 ng and 0.17 ng of homozygous DNA, versus the logarithm of the correspondent DNA amount. PCR efficiencies (E) were calculated by $E = 10^{(-1/b)} -1$ (Rutledge and Cote, 2003) where b is the slope of the regression line. R^2 and b parameters were estimated directly by the ABI Prism 7000 SDS software.

Interference limit (IL) We estimated the background noise due to aspecific amplifications. The two homozygous DNA (80 ng each) were amplified by the two allele-specific assays of both the ARMS and the MAMA systems. Every experiment has been repeated three times and average Ct values of specific ($\overline{Ct}_s$) and aspecific ($\overline{Ct}_a$) amplifications and the relative standard deviations ($\acute{o}_s$ e $\acute{o}_a$), were acquired. The 99% interval of confidence of the Ct_a values, was estimated as $\overline{Ct}_a \pm t_{0.01} \cdot \sigma_a / \sqrt{N}$, where N is the number of observations and $t_{0.01}$ is the tabulated value of Student's t for the 1% probability level and N-1 degree of freedom. The background threshold cycle for a given amount of DNA (Ct_{lim}) was therefore set at $Ct_{lim} = \overline{Ct}_a - t_{0.01} \cdot \sigma_a / \sqrt{N}$. The lowest allele concentration that could be significantly distinguish from the background (IL) has been calculated by $IL = 100/(2^{(\overline{Ct}_s - Ct_{lim})} + 1)$ (Germer et al., 2000).

Limit of detection (LOD) Different amounts (367, 183.5, 91.75, 45.87, 22.94, 11.47 pg and 5.73 pg) of the two homozygous DNA have been amplified with the correspondent allele-specific ARMS and MAMA assays. Samples were run in triplicate and every experiment has been repeated three times. For each assay the detection limit was fixed at the lowest DNA amount at which signals of all the replicates were still detectable.

Limit of quantification (LOQ) Average Ct values obtained from the amplification of each DNA sample used to evaluate LODs have been analysed by Randomised Complete Block Design Analysis of Variance. In particular each DNA amount was tested against the lower ones in a series of orthogonal comparisons. Limit of quantification was defined as the lowest DNA amount whose average specific Ct was significantly different from the average Ct values of the lower amounts.

Precision and accuracy Each reference sample was amplified by the ARMS and the MAMA allele-specific assays. The relative concentrations (%) of the *MetHc* or the *MetHt* allele were estimated by the following formula: $\% = 100/(2^{\Delta Ct} +1)$ where ΔCt is the Ct difference between the *MetHc* and the *MetHt* specific reactions or *vice versa* respectively (Germer et al., 2000). Triplicates of each allele-specific reaction were run on the same plate and each experiment was repeated three times. Concentrations were analysed by Completely Random Analysis of Variance. Precision was evaluated by the Relative Standard Deviation (RSD); $RSD = 100 \cdot (\sigma/MC)$, where MC is the average Measured Concentration and σ the correspondent standard deviation. Accuracy was assessed plotting Nominal Concentrations (NC) versus the correspondent MC and performing a regression analysis. Furthermore the closeness of agreement between nominal and measured concentrations was described by the Relative Difference (RD); $RD = 100 \cdot (MC - NC)/NC$. The accuracy and precision limits proposed by the Community Reference Laboratory for GM food and feed of the European Union, have been here acknowledge (Method acceptance criteria and method performance requirements available at http://gmo-crl.jrc.it/doc/Method%20requirements.pdf). Therefore the dynamic range of each assay has been defined as the interval of concentrations within which MC is ±25% of the NC and RSD <25%.

Conversion of DNA amounts in *MetH* copy number According to the Database of Genome Sizes of the Technical University of Denmark (available at http://www.cbs.dtu.dk/databases/DOGS/) 3.676 pg of human genomic DNA correspond to one *MetH* copy.

Results

Establishment of the real time PCR assay outside the polymorphic region of *MetH* A Real Time PCR assay to detect the *MetH* sequence (Tot assay) was designed using Primer Express™ default parameters. The forward primer of this assay (Tot primer) match the *MetH* just before the SNP at position 187. The final concentrations of the forward and the reverse primers were tested independently at 50, 300 and 900 nM. Steepest amplifications were achieved at 900 nM for both the primers. Specificity was confirmed by agarose gel electrophoresis and direct sequencing (data not shown).

Establishment of the real time PCR assays specific for the *MetHc* and the *MetHt* alleles An extensive number of different ARMS and MAMA allele-specific forward primers were combined with the same reverse primer and probe used in the Tot assay. ARMS and MAMA primers containing the allele-specific nucleotide in the last, penultimate and third last position from the 3' end have been considered. The MAMA deliberate mismatch have been positioned in the sixth, fifth and fourth position from the 3' end. Primers were tested amplifying samples of the two homozygous DNA; for both the ARMS and the MAMA systems the *MetHc* and the *MetHt*
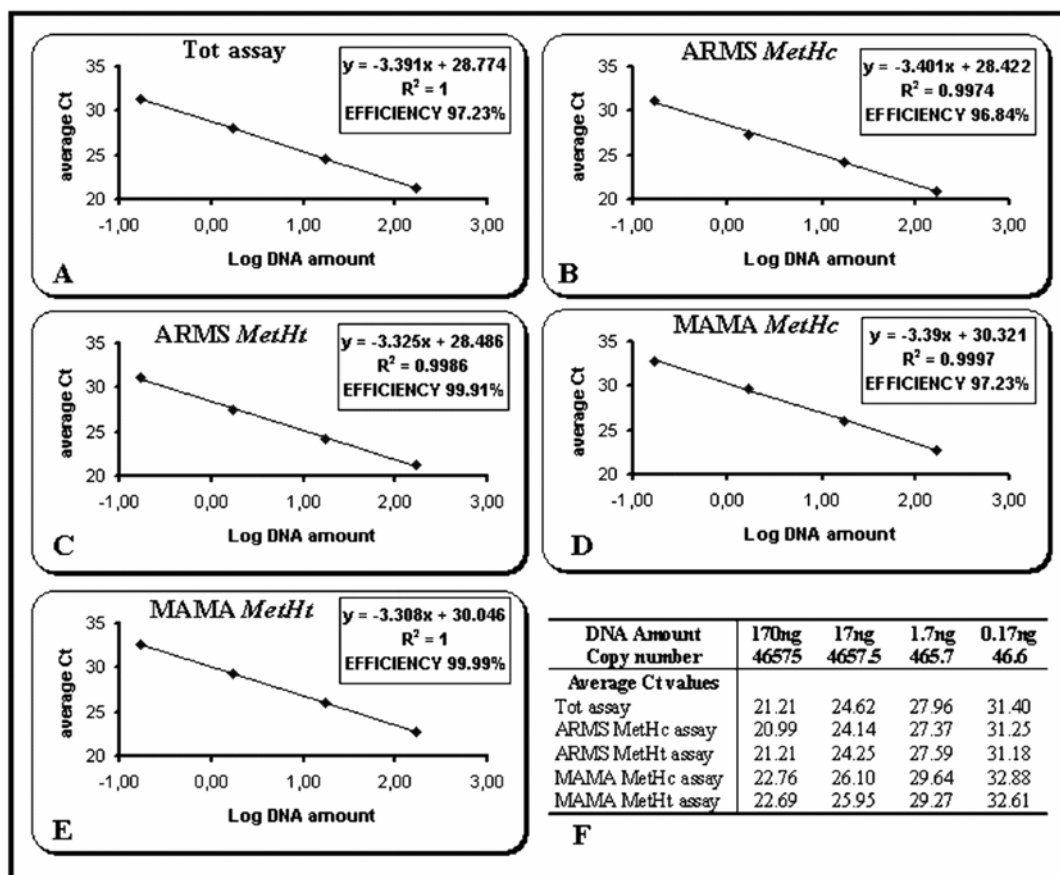
Fig. 1. Linearity plot of the Tot assay (A), the ARMS *MetHc* assay (B), the ARMS *MetHt* assay (C), the MAMA *MetHc* assay (D) and the MAMA *MetHt* assay (E). Equations of the regression curves, coefficient of determinations (R^2) and PCR efficiencies were calculated. For each DNA amount the correspondent copy number of target sequence and the average Ct values obtained by the tested assays have been reported (F). Threshold was set at 0.12

allele-specific assays achieving the highest difference between specific and aspecific amplifications were selected. Increasing concentrations of forward primers were tested up to 4000 nM for the purpose of determining, for each assay, the minimum primer concentration that yield the minimum Ct value of the specific amplification. This correspond to the complete amplification (saturation) of the target sequences in the sample. Saturation has been reach at 2500 nM for the two ARMS assays and at 3000 nM for the two MAMA ones. Annealing temperature was raised up to 61°C, gaining an average Ct difference between specific and aspecific amplifications of 8.58 and 9.27 for the ARMS and MAMA systems respectively. Homozygous DNA samples have been amplified by the correspondent ARMS and MAMA allele-specific assays and the Tot assay. Previously optimized PCR conditions have been followed. The same copy number of targets (*MetHc*, *MetHt* and *MetH*) were contained in each amplified sample, as batch solutions were equimolar. Ct values of the *MetHc* and the *MetHt* specific amplifications were 21.85 and 21.90 for the ARMS assays and 23.13 and 23.31 for the MAMA assays, respectively. The Ct value of the Tot assay was 21.98. Therefore allelic quantification may be

achieved by direct comparison of the two ARMS or MAMA allele-specific assays. Alternatively each allele-specific assay may be normalized by the Tot assay (using *MetH* as reference sequence). In this case a calibrator must be use to take in account the average delay of 1.24 Ct between the MAMA assays and the Tot assay.

Linearity and efficiency   To evaluate the linearity and the efficiency of the assays under study, 170, 17, 1.7 ng and 0.17 ng of the two homozygous DNAs were amplified. Average Ct values of the specific amplifications were plotted versus the logarithm of the correspondent DNA amount. As shown in Fig. 1 each assay is linear over three magnitudes of DNA amount, ranging from 170 to 0.17 ng and efficiencies are closed to the average value of 98.24%.

Interference limit (IL)   Samples of the two homozygous DNA (80 ng each) were amplified by both the ARMS and MAMA *MetHc* assays and both the ARMS and MAMA *MetHt* assays. Samples have been analysed in triplicate and experiments repeated three times. Average Ct values of specific and aspecific amplifications are listed in Table 2 with

Table 2. Interference limits. Average specific and aspecific Ct values obtained amplifying samples of the two homozygous DNA (80 ng each) by the ARMS (grey columns) and the MAMA (white columns) allele-specific assays; σ, standard deviation; Ct_{lim}, background threshold cycle; IL, interference limit, * indicate aspecific amplifications. Threshold was set at 0.2

Sample	MetHc Homozygous DNA				MetHt Homozygous DNA			
System	ARMS		MAMA		ARMS		MAMA	
Assay	MetHc	MetHt	MetHc	MetHt	MetHc	MetHt	MetHc	MetHt
Average Ct	22.78	32.59*	24.38	31.59*	30.45*	23.10	35.69*	24.35
σ	0.147	0.621	0.070	0.397	0.267	0.066	0.430	0.058
Ct_{lim}	30.39		30.19		29.51		34.17	
IL	0.51%		1.75%		1.16		0.11%	

Table 3. Average Ct values obtained amplifying different amounts of MetHc (grey columns) or MetHt (white columns) homozygous DNA by the correspondent ARMS and MAMA allele-specific assays. In square brackets are values obtained when at least an undetectable amplification has been recorded; σ, standard deviation. Threshold was set at 0.15

DNA amount	367pg		183.5pg		91.75pg		45.87pg		22.94pg		11.47pg		5.73pg	
Target copy number	100		50		25		12.5		6.25		3.12		1.56	
Assay	MetHc	MetHt	MetHc	MetHt	MetHc	MetHt	MetHc	MetHt	MetHc	MetHt	MetHc	MetHt	MetHc	MetHt
System	ARMS		ARMS		ARMS		ARMS		ARMS		ARMS		ARMS	
Average Ct	30.75	30.87	31.72	31.67	35.89	32.99	34.95	33.92	[38.62]	35.29	[38.26]	35.89	---	[38.86]
σ	0.13	0.17	0.24	0.16	1.47	0.24	1.26	0.23	[1.13]	0.61	[1.9]	0.51	---	[0.68]
System	MAMA		MAMA		MAMA		MAMA		MAMA		MAMA		MAMA	
Average Ct	31.87	31.85	33.01	32.60	34.33	34.35	35.43	35.42	36.35	36.67	36.69	37.10	[38.59]	[39.35]
σ	0.24	0.12	0.20	0.13	0.17	0.31	0.44	0.56	0.78	0.66	1.56	0.46	[0.50]	[0.33]

the correspondent standard deviations. The IL of the ARMS MetHc and the MetHt allele-specific assays resulted of 0.51% and 1.16% respectively. The IL of the correspondent MAMA assays resulted of 1.75% and 0.11%. The average IL of both the ARMS and MAMA systems have been settled at the 1%.

Limit of detection (LOD) and limit of quantification (LOQ)
Two series of samples containing 367, 183.5, 91.75, 45.87, 22.94, 11.47 pg and 5.73 pg of the MetHc or MetHt homozygous DNA were amplified with the correspondent ARMS and MAMA allele-specific assays (Table 3). Samples were run in triplicate and experiments repeated three times recording average Ct values and standard deviations. LODs were determined in 45.87 pg for the ARMS MetHc assay and in 11.47 pg for the ARMS MetHt and the MAMA MetHc and MetHt assays.

For each series, average Ct values were tested by Randomised Complete Block Design Analysis of Variance, where each experiment was considered as a block. Differences between experiments were never significant ($F_{[2,10]} = 1.532$, p = 0.263 and $F_{[2,12]} = 0.226$, p = 0.801 respectively for the MetHc and MetHt DNA analyzed by the correspondent ARMS allele-specific assays; $F_{[2,12]} = 0.455$, p = 0.645 and $F_{[2,12]} = 0.744$, p = 0.496 for the MetHc and MetHt DNA analyzed by the correspondent MAMA assays), vice versa they were always

significant between DNA concentrations ($F_{[5,10]} = 44.56$, p = $1.62 \cdot 10^{-6}$; and $F_{[6,12]} = 237.22$, p = $9.4 \cdot 10^{-12}$ respectively for the MetHc and MetHt DNA analyzed by the correspondent ARMS allele-specific assays; $F_{[6,12]} = 112.6$, p = $7.67 \cdot 10^{-10}$ and $F_{[6,12]} = 37.48$, p = $4.33 \cdot 10^{-7}$ for the MetHc and MetHt DNA analyzed by the correspondent MAMA assays). In particular each DNA amount was tested against the lower ones in a series of orthogonal comparisons. LOQs have been determined at 183.5 pg for the ARMS MetHc assay and at 45.87 pg for the ARMS MetHt and the MAMA MetHc and MetHt assays. These DNA amounts represent the 0.23% and the 0.057% of a sample containing 80 ng of DNA pools.
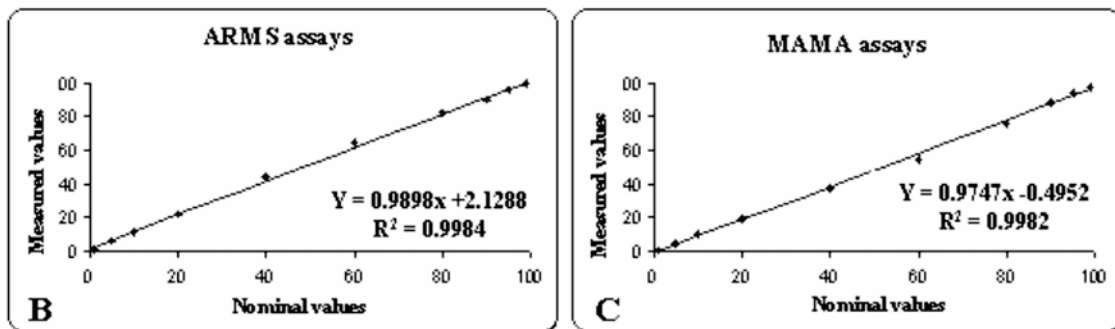
Precision and accuracy The two series of reference samples, covering a concentration interval ranging from 1% to 99% of both the alleles, were analysed by the ARMS and the MAMA systems. Each sample (80 ng) was amplified in triplicate and each quantification was repeated three times. Completely Random Analysis of Variance demonstrated that the difference between concentrations were always significant, accounting for more than the 99% of the total variability of the data (Table 4). Relative Standard Deviation (RSD) and Relative Difference (RD) values are reported for each reference sample (Fig. 2 (A)). Plots of the nominal concentrations versus the correspondent average concentrations measured by the

Table 4. Completely Random Analysis of Variance of the measured concentrations. Different concentrations have been considered as the source of variability. The percentage of the total variability explained by the different DNA concentrations is given by the R^2 values

Experiment		F value	Significance	R^2
Samples	System			
MetHc 1%-60%	ARMS	$F_{[5,45]} = 6069.49$	$p = 1.88 \cdot 10^{-62}$	99.85%
MetHt 1%-60%	ARMS	$F_{[5,41]} = 3673.68$	$P = 0$	99.28%
MetHc 1%-60%	MAMA	$F_{[5,48]} = 1052.15$	$p = 8.39 \cdot 10^{-48}$	99.10%
MetHt 1%-60%	MAMA	$F_{[5,48]} = 1372.60$	$p = 1.5 \cdot 10^{-50}$	99.31%

Samples		ARMS assays				MAMA assays			
NC	NCN	MC	σ	RSD	RD	MC	σ	RSD	RD
MetHc									
1%	218	1.44	0.12	8.44	44.00	1.01	0.21	20.43	1.31
5%	1090	6.15	0.36	5.83	22.95	5.10	0.42	8.25	2.06
10%	2180	11.77	0.75	6.34	17.69	9.91	1.36	13.78	-0.86
20%	4360	22.09	0.52	2.38	10.44	19.61	2.38	12.16	-1.94
40%	8720	44.54	1.66	3.73	11.35	37.21	2.99	8.04	-6.98
60%	13080	64.10	0.77	1.20	6.84	54.63	2.35	4.31	-8.94
MetHt									
1%	218	1.01	0.12	12.40	0.47	2.82	0.61	21.60	182.50
5%	1090	4.57	0.34	7.62	-8.59	6.26	0.43	6.87	25.00
10%	2180	10.13	0.75	7.36	1.31	11.44	2.17	19.00	14.46
20%	4360	18.20	1.11	6.09	-8.97	24.55	1.64	6.69	22.75
40%	8720	35.90	0.77	2.14	-10.25	45.36	2.35	5.19	13.41
60%	13080	55.46	1.66	3.00	-7.56	62.79	2.99	4.76	4.65

Fig. 2. precision and accuracy results. The two series of reference samples have been analyzed by the ARMS and the MAMA assays. For each sample (80 ng) Nominal *MetHc* or *MetHt* Concentration (NC), Nominal Copy Number of target sequence (NCN), Average Measured Concentration (MC), standard deviation (ó), Relative Standard Deviation (RSD) and Relative Difference (RD) are reported (A). Plots of MC values measured with the ARMS and MAMA assays versus the correspondent NC values have been reported in (B) and (C) respectively, where the whole range from 1% to 99% have been considered. For each plot the function of the regression curve and the coefficient of determination (R^2) have been reported. Threshold was set at 0.2.

ARMS and the MAMA assays, are reported in Fig. 2 (B) and (C) respectively.

Thresholds for Relative Difference and Relative Standard Deviation were fixed at ±25% and <25% respectively, resulting in a dynamic range from 5% to 95% for both the ARMS and MAMA assays.

Discussion

Newton *et al.* (1989) first used oligoncleotide primers to discriminate SNPs by PCR, visualizing results on agarose gel. The technique was improved by introducing 5' fluorogenic probes (Glaab and Skopek, 1999) to suit rapid SNPs genotyping and assessing hotspot mutations. A number of

applications of great interest, as the monitoring of minimal residual disease in the course of malignant haemopathies, the quantitative assessment of post-transplant chimerism and studies on complex disease, require allele frequency estimations on pooled DNA samples. Real Time PCR may be applied to this kind of investigations, but should undergo a rigorous performance evaluation (pre-validation). In this study ARMS and MAMA allele-specific assays were used to explore the parameters needed to establish and test quantitative methods. Forward primers of the allele-specific assays have been selected following two criteria: specificity and saturation of all the target sequences in the sample under analysis. Specificity has been evaluated by the difference between the specific and aspecific amplifications (ΔCt) of an homozygous DNA sample. Average ΔCt value was 8.58 and 9.27 for the

ARMS and MAMA assays respectively. As reported by Germer *et al.* (Germer *et al.*, 2000), those values are fully adequate for allele frequency estimations on DNA pools. High forward primers concentrations (2500 nM for the ARMS and 3000 nM for the MAMA ones) were used to saturate the reactions, and balance the low melting temperature used to improve primers specificity. MAMA primers are more specific than ARMS ones, however we found that their annealing strength is strongly affected by the position of the added mismatch (data not shown). MAMA primers holding the mismatch in the fifth nucleotide from the 3' end have been selected because this configuration gains a satisfactory saturation and specificity. Tested assays have provided to be linear over tree $\log_{10}$ of magnitude and have efficiencies close to the average of 98.24%. Each single allele-specific assay is characterized by its own limit of detection (LOD) and quantification (LOQ). Allele frequencies on pooled DNA samples are measured by using both the allele-specific assays. Interference limit (IL) has been here introduced to describe how much the aspecific amplifications of one single allele-specific assay affect the specific amplifications of the other allele-specific assay, resulting in false positive results or quantification bias. Average IL values of both the ARMS and MAMA systems have been considered around the 1% of a sample containing 80 ng of pooled DNA. The average IL value significantly exceed the LOQs for all the tested assays. For that reason the IL represents the most relevant parameter to fix the lower limit of allele frequency estimations. Another key factor to consider in performing a real time PCR assay is how representative the DNA template added to the reaction (test sample) is compared to the bulk sample. If a very small quantity of DNA is added to the reaction the sampling error will be larger. According to Kay and Van den Eede (Kay and Van den Eede, 2001) at least 20 copies of target sequence should be present in the reaction to guaranty that it contains relevant DNA. Following their suggestion we added 80 ng of DNA to each reaction to test accuracy and precision. In the weakest dilution (1%) this represents 218 copies of the less abundant allele. As illustrated in Fig. 2, both the systems show an acceptable level of accuracy and precision within the whole range from 1% to 99%. However we acknowledge the strict accuracy and precision limits proposed by the Community Reference Laboratory for GM Food and Feed of the European Union, reducing the dynamic range of each system within a range from 5% to 95%. Nevertheless accuracy and precision limits can be adapted case-by-case to the goal the assay has been planed for, extending or reducing the dynamic range. In conclusion the allele-specific assays here introduced have been demonstrated to be effective in the allele quantification within a range at least from 5% to 95%. This study is relevant, not only for basic research studies, but also for clinical investigations. All the critical issues to approach allele frequency estimation on pooled DNA samples by Real Time PCR (assay set up and pre-validation) have been discussed. Furthermore the pre-validation procedure described in this paper can be applied, not only to the allele quantification, but also to many other Real Time PCR applications.

References

Alizadeh, M., Bernard, M., Danic, B., Dauriac, C., Birebent, B., Lapart, C., Lamy, T., Le Prise, P. Y., Beauplet, A., Bories, D., Semana, G. and Quelvennec, E. (2002) Quantitative assessment of ematopoietic chimerism after bone marrow transplantation by Real-Time quantitative polymerase chain reaction. *Blood* **99**, 4618-4625.

Breen, G., Harold, D., Ralston, S., Shaw, D. and St Clair, D. (2000) Determining SNP allele frequencies in DNA pools. *Biotechniques* **28**, 464-470.

Elmaagacli, A. H. (2002) Real-Time PCR for monitoring minimal residual disease and chimerism in patients after allogeneic transplantation. *Int. J. Hematol.* **76**, 204-205.

Fredriksson, M., Barbany, G., Liljedahl, U., Hermanson, M., Kataja, M. and Syvanen A. C. (2004) Assessing hematopoietic chimerism after allogeneic stem cell transplantation by multiplexed SNP genotypi using microarrays and quantitative analysis of SNP alleles. *Leukemia* **18**, 255-266.

Germer, S., Holland, M. J. and Higuchi R. (2000) High-Throughput SNP Allele-Frequency determination in Pooled DNA Samples by Kinetic PCR. *Genome Res.* **10**, 258-266.

Glaab, W. E. and Skopek, T. R. (1999) A novel assay for allelic discrimination that combines the fluorogenic 5' nuclease polymerase chain reaction (TaqMan) and mismatch amplification mutation assay. *Mutat. Res.* **430**, 1-12.

Horwitz, W. (1995) Protocol for the design, conduct and interpretation of method performance studies. *Pure Apl. Chem.* **67**, 331-343.

Kay, S. and Van den Eede, G. (2001) The limit of GMO detection. *Nat. Biotechnol.* **19**, 405-406.

Livak, K. J. and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408.

Newton, C. R., Graham. A., Heptinstall, L. E., Powell, S. J., Summers, C., Kalsheker, N., Smith, J. C. and Markham, A. F. (1989) Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503-2516.

Norton, N., Williams, N. M., O'Donovan, M. C. and Owen M. J. (2004) DNA pooling as a tool for large-scale association studies in complex traits. *Ann. Med.* **36**, 146-152.

Pont-Kingdon, G. and Lyon, E. (2003) Rapid detection of aneuploidy (trisomy 21) by allele quantification combined with melting curves analysis of single-nucleotide polymorphism loci. *Clin. Chem.* **49**, 1087-1094.

Punia, P., Cane, P., Teo, C. and Saunders, N. (2004) Quantitation of hepatitis B lamivudine resistant mutants by real-time amplification refractory mutation system PCR. *J. Hepatol.* **40**, 986-992.

Ranade, K., Chang, M. S., Ting, C. T., Pei, D., Hsiao, C. F.,

Olivier, M., Pesich, R., Hebert, J., Chen, Y. D., Dzau, V. J., Curb, D., Olshen, R., Risch, N., Cox, D. R. and Botstein, D. (2001) High-throughput genotyping with single nucleotide polymorphisms. *Genome Res*. **11**, 1262-1268.

Risch, N. J. (2000) Searching for genetic determinants in the new millennium. *Nature* **405**, 847-856.

Rutledge, R. G. and Cote, C. (2003) Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Res*. **31**, 93.

Schwarz, G., Baumler, S., Block, A., Felsenstein, F. G. and Wenzel, G. (2004) Determination of detection and quantification limits for SNP allele frequency estimation in DNA pools using Real Time PCR. *Nucleic Acids Res*. **32**, 24.

Schwarz, G., Baumler, S., Block, A., Felsenstein, F. G. and Wenzel, G. (2004) Quantitative technologies for allele frequency estimation of SNPs in DNA pools. *Mol. Cell. Probes*. **16**, 429-434.