

Competitive enzymatic reaction to control allele-specific extensions

Emilie Hultin, Max Källér, Afshin Ahmadian* and Joakim Lundeberg

Royal Institute of Technology (KTH), AlbaNova University Center, School of Biotechnology, SE-106 91 Stockholm, Sweden

Received January 20, 2005; Revised and Accepted February 18, 2005

ABSTRACT

Here, we present a novel method for SNP genotyping based on protease-mediated allele-specific primer extension (PrASE), where the two allele-specific extension primers only differ in their 3'-positions. As reported previously [Ahmadian, A., Gharizadeh, B., O'Meara, D., Odeberg, J. and Lundeberg, J. (2001), *Nucleic Acids Res.*, 29, e121], the kinetics of perfectly matched primer extension is faster than mismatched primer extension. In this study, we have utilized this difference in kinetics by adding protease, a protein-degrading enzyme, to discriminate between the extension reactions. The competition between the polymerase activity and the enzymatic degradation yields extension of the perfectly matched primer, while the slower extension of mismatched primer is eliminated. To allow multiplex and simultaneous detection of the investigated single nucleotide polymorphisms (SNPs), each extension primer was given a unique signature tag sequence on its 5' end, complementary to a tag on a generic array. A multiplex nested PCR with 13 SNPs was performed in a total of 36 individuals and their alleles were scored. To demonstrate the improvements in scoring SNPs by PrASE, we also genotyped the individuals without inclusion of protease in the extension. We conclude that the developed assay is highly allele-specific, with excellent multiplex SNP capabilities.

INTRODUCTION

The most common type of genetic diversity in the human genome is single nucleotide polymorphisms (SNPs). These bi-allelic alterations are distributed across the genome and have been implicated in genetic disorders, susceptibility to

different diseases, predisposition to adverse reactions to drugs and use in forensic investigations. The selection of suitable strategies for monitoring and scoring of these SNPs requires validated SNPs and accurate methodologies. A number of techniques are now available for rapid and automated genotyping (1,2). The use of miniaturized assays, such as microarrays with oligonucleotide reagents immobilized on small surfaces is a frequently proposed approach for large-scale mutation analysis and high-throughput genotyping of SNPs. Despite considerable efforts, some of the microarray techniques, however, suffer from relatively low accuracy compared with conventional DNA sequencing technologies. For example, microarray hybridization of PCR fragments to allele-specific oligonucleotide probes relies on the thermal stability of the PCR fragment and the short probe (3). The limitation lies in the fact that the minute differences in the duplex stability between a perfect match and a mismatch at one base may be difficult to distinguish. However, improvements in the technology have been made by monitoring, in real-time, the dynamics of hybridization (4,5). Another approach to enhance the discrimination power of microarray-based polymorphism analysis is to employ enzymatic events on the arrays, such as ligation assays (6,7), minisequencing (8–10) and allele-specific extension (ASE) (11,12).

ASE on microarrays takes advantage of discrimination properties of DNA polymerase in the extension of a 3'-terminus mismatch primer. Two allele-specific primers with alternating 3'-termini are designed to match one allele perfectly but mismatch the other allele at the 3'-terminus. In this way, allele-specific primer extension with fluorescent-labeled nucleotides provides information about the presence or absence of an allele that can be analyzed by a laser fluorescence scanner. In the case of homozygous template, one of the allele-specific primers results in detectable product, while poor extension of the other primer will be observed owing to 3'-terminus mismatched primer-template. A heterozygous sample results in equal extension of both primers. Thus, fluorescent signal ratio between allele-specific primers scores the genotype. This method has previously been employed to identify single base variations, but it has been shown that

*To whom correspondence should be addressed. Tel: +46 8 5537 8333; Fax: +46 8 5537 8481; Email: afshin@biotech.kth.se

some mismatches are poorly discriminated by the DNA polymerase (12–15). We have previously demonstrated that the extension of mismatched configurations occurs with slower reaction kinetics in comparison with the extension of the matched primer-template configurations (16). In addition, we demonstrated the use of a nucleotide-degrading enzyme (apyrase) to kinetically distinguish genotypes in array-based ASE assays (17–20). In the apyrase-mediated allele-specific extension (AMASE), mismatch extensions are minimized since apyrase degrades the nucleotides and prevents the extension of the slower mismatched primer-template. However, in this work, we have further developed and modified the AMASE approach by replacing apyrase with a thermostable protein-degrading (digesting) enzyme (Proteinase K). In this method, denoted as protease-mediated allele-specific extension (PrASE), we still take advantage of the fact that DNA polymerase extends 3'-termini mismatches with slower reaction kinetics. The use of Proteinase K allows genotyping at elevated reaction temperature as compared with thermolabile apyrase, a feature that may be necessary when a thermostable DNA or RNA polymerase is used for the extension of allele-specific primers. In addition, a well-known factor in molecular biotechnology is that low temperature is the major cause of non-specific hybridization (21,22). This factor can cause problems in the AMASE assay especially in multiplex genotyping assays, giving rise to signals from spurious hybridized primers. Consequently, the novel PrASE assay permits stringent temperature conditions that will lead to higher multiplex capability. Here, we demonstrate the use of Proteinase K in a 13-plex SNP analysis. Simultaneous detection is achieved on a tag microarray using individual tags of the extension primers.

MATERIALS AND METHODS

Microarray preparation and spatial separation of samples

Forty-eight oligonucleotide tags functioning as probe captures on the glass slide were taken from www.genome.wi.mit.edu (9) (Table 1). The tags containing a 5'-poly(T) spacer of 15 thymine residues were synthesized by MWG-Biotech AG (Ebersberg, Germany) with a 5'-terminus amino link with a C6 spacer (to facilitate covalent immobilization to the pre-activated slides). The oligonucleotides were suspended at a concentration of 20 μ M in 150 mM sodium phosphate, pH 8.5 and 0.06% sarkosyl and were spotted with a Q-array (Genetix, Hampshire, UK) on Code Link activated slides (Amersham Biosciences, Uppsala, Sweden). Sarkosyl was added to the spotting solution as it improved spot uniformity. After printing, the arrays were incubated overnight in a humid chamber followed by post coupling as outlined by the manufacturer. Briefly, the slides were incubated at 50°C for 15 min in a blocking solution (50 mM ethanol amine, 0.1 M Tris, pH 9 and 0.1% SDS), rinsed twice in dH₂O and washed with pre-warmed (50°C) 4 \times SSC/0.1% SDS for 15–60 min on a shaker (20 \times SSC contains 175.3 g/l NaCl, 88.2 g/l sodium citrate, pH 7; and 10% SDS contains 100 g/l SDS, pH 7.2). The arrays were then rinsed in dH₂O and dried by centrifugation for 3 min at 800 r.p.m.

The 48 oligonucleotides were printed in 48 identical arrays (an array of arrays) on the slide and each array contained

Table 1. Tag sequences (named T2–T49) on the microarray slide (T14–15, T28–35 and T38–49 were not used in this project)

Tag name	5'-tag sequence
T2	CGCAGGTATCGTATTAATTGATCTGC
T3	CCTCATGTCAACGAAGAACAAGAAC
T4	ATTGAAGCCTGCCGTGCGAGACTAA
T5	AGACTGCGTGTGGCTCTGTACACAA
T6	TTATGGTGTATCAGTCAACACCAGG
T7	GAGACACCTTATGTTCTATACATGC
T8	TCCATGCGCTTGCTTTCATCTAGC
T9	GCCTTACATACATCTGTCGGTTGTA
T10	CACAAGGAGGTCAGACCAGATTGAA
T11	GCCACAGATAATATTCACATCGTGT
T12	ACACATACGATTCTGGGAACCTCAA
T13	TTACAGGATGTGCTCAACAGACGTT
T14	GCTCACAAATTTGCAATGATGTGCC
T15	CTGCACTGCTCATTAAATATACTTCTGG
T16	TTCACGCACTGACTGACAGACTGCTT
T17	CAACATCATCACGCAGAGCATCATT
T18	GCATCAGCTAACTCCTTCGTGTATT
T19	GGCGTTATCACGGTAATGATTAACAGC
T20	ACATCAATCTCTGACCGTTCCGC
T21	GCCTTATGCTCGAAGTACGCAATAAC
T22	CGGATATCACACGATCAATCATAGGTAA
T23	CCTTAATCTGCTGCAATGCCACAGC
T24	TAGCTCTCCGCCTACAATGACGTCA
T25	AGGAACGCCTTACGTTGATTATGA
T26	GAGTCAGTACCGATGTAGCCGATAA
T27	ACTCGAATGAACCAGGCGATAATGG
T28	ATTATATCTGCCGGAAGGTACGCC
T29	GGACAGACAGTGGCTACGGCTCAGTT
T30	CGGTATTCGCTTAATTCAGCACAAAC
T31	GCTCTTACCTGTTGTGAGATATAA
T32	GAATCGGTAGTAAGCGCCGCTCTT
T33	CAATCATGCTGCTAACGTTGTGACCG
T34	ATGTTGTCTGCGATTGACTCTTCTT
T35	ATTAGCATCGTACAGCGCGATATG
T36	ATGTTGCGAGCACTTGCAGTACCTT
T37	AGTTCGTGCTTACCGCAGAATGTCAG
T38	CCTGTACCTGAATCAATGTTAGGTT
T39	CCAGATTACCTGCTGATGATCAACTG
T40	ATAATCGGCGTGGCAGATAACATAT
T41	TTGTCGAATCCAATCGTATCCAGTT
T42	AACCATGAGAACGTCGTTGACGATG
T43	GCTTCAACGAGCATGTCTGGAATGG
T44	CACTCTTCGATAGTCCAGCGATGGC
T45	AGTTCAGCGGATACGGAACGAGAT
T46	CGCATTGCTTGTGAATATTGCGTTC
T47	GATACATGGTAAGTCGCGTCTCTT
T48	TATTCGGATCGTCTATCAGTTCTGC
T49	CAGGTGAGCCGTGTAGTTGAAGGTT

triplicates of each oligonucleotide. The 48 sub-arrays were separated during hybridization by a reusable silicone mask (Elastosile[®] RT 625 A/B, Wacker-Chemie GmbH, Munich, Germany), molded in an inverted 384-well plate and excised to fit the slide (11). A Custom made rack was used to press the silicone firmly to the slide and keep it in place during the reactions.

SNP positions

Thirteen SNP positions were selected for this study. The reference sequence numbers, the allele alternatives and the chromosomal locations of the SNPs are rs17429 (C/T) (Xq28), rs1799841 (T/C) (20p11), rs203319 (A/G) (22q13), rs752118 (C/T) (20q13), rs1800598 (C/T) (11q23), rs752744 (A/G)

(11p13–11p12), rs752471 (A/G) (12q), rs714784 (C/T) (15q21.2), rs760589 (C/T) (1p35), rs737820 (A/G) (22q11.2), rs746713 (A/G) (22q12.3–22q13.2), rs758593 (A/G) (5q33) and rs740841 (A/G) (12p13.3), referred as SNP numbers 1–13 in the rest of the text.

Samples and multiplex PCR amplification

Genomic DNA extracted from blood from 36 healthy Swedish individuals (named 1–36) was used in the analysis. The 13 SNPs were amplified in a nested PCR by a multiplex outer PCR (30 cycles of 95°C for 20 s, 55°C for 40 s and 72°C for 30 s) followed by multiplex inner PCR (40 cycles of 95°C for 20 s, 55°C for 40 s and 72°C for 30 s). The rationale for employing a nested PCR approach was to perform multiplex amplification without putting efforts in PCR optimization. In addition, since a small fraction of the outer products is transferred to the inner reactions, the outer PCRs can be used for different analytical investigations without the use of original genomic material. Prior to PCR cycles, the DNA was denatured at 95°C for 5 min and after PCR cycling the reactions were incubated at 72°C for 10 min. The reaction volume in the outer PCR was 50 µl, containing ~2.5 ng genomic DNA as starting material. An aliquot of 0.5 µl of the outer PCR product was transferred to the inner PCR with a reaction volume of 50 µl. The outer and inner multiplex PCR contained 1 U AmpliTaq Platinum DNA polymerase (Invitrogen AB, Lidingö, Sweden), 1× PCR buffer [20 mM Tris–HCl, pH 8.4, 50 mM KCl], 2 mM MgCl₂, 0.2 mM dNTPs and 0.05 µM of each primer.

The outer forward and reverse PCR primers were located 16 nt upstream and downstream of the SNP positions, respectively, generating outer PCR products of 71 bp. The inner biotin-labeled primers were located only 2 nt from the SNP position, while the non-biotinylated primers had a distance of 6 nt to the SNPs. This approach generated inner PCR products between 46 and 52 bp. The 5' end biotinylated primers are used to allow immobilization of PCR products to super paramagnetic beads. The sequence of all PCR primers is outlined in Table 2.

Sample preparation, hybridization and specific extensions

The procedures of PCR product immobilization, washing, annealing of 3'-terminus allele-specific primers and the

multiplex ASEs were automated by the use of a Magnatrix 1200 pipetting robot system capable of handling magnetic beads (Magnetic Biosolutions, Stockholm, Sweden). An aliquot of 200 µg streptavidin-coated super paramagnetic beads (Dynabeads M280; Dynal, Oslo, Norway) were used to immobilize each multiplexed inner PCR product. After immobilization and wash with a binding/washing buffer (10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 2 M NaCl, 1 mM β-mercaptoethanol and 0.1% Tween-20), single-stranded DNA (ssDNA) was obtained by alkali elution (0.1 M NaOH for 5 min at room temperature) of the non-biotinylated strand. The supernatant was discarded and the beads were washed once with Tris-EDTA. A mixture (total volume of 60 µl) containing 0.08 µM of each extension primer (two 3'-terminus allele-specific primers per SNP) (Table 3), 1× annealing buffer (AB) (10 mM Tris-acetate pH 7.75, 2 mM Mg-acetate) and 0.5 µg single-stranded binding protein (SSB) [in-house produced (23) and also available from Amersham Biosciences] was added to the immobilized ssDNA (SSB was included to obtain specific annealing). Each ASE primer contained a specific tag at its 5' end. A comparison of Tables 1 and 3 shows that each tag on the glass slide is complementary to one of the tags on the ASE primers. The ASE primers were allowed to anneal to the captured strands at 72°C for 3 min, 50°C for 7 min and 40°C for 1 min. The excess of primers was discarded and the immobilized ssDNA was washed once with 1× AB and then resolved in 1× AB to a volume of 20 µl.

The PrASE reaction was performed at 45°C by adding first 20 µl of a solution containing 10 U exonuclease-deficient (exo⁻) Klenow DNA polymerase [in-house produced (24) and also available from New England Biolabs, Beverly, MA], 1× extension buffer (EB) (42.5 mM Tris–HCl, pH 8, 5 mM MgCl₂ and 1 mM DTT) and 0.25% BSA to the ssDNA. This mixture was set to incubate for 1 min before 20 µl of a second mixture containing 1.5 µM of each dNTP [50% Cy5-labeled dCTP and dTTP (Amersham Biosciences)], 2× EB, 0.5% BSA and 20 µg of Proteinase K was added to initiate the extension by exo⁻ Klenow polymerase and simultaneously terminate the extension by degradation of the Klenow polymerase by the protease. This competitive enzymatic reaction was incubated at 45°C for 1 min. The conventional allele-specific primer extension was carried out with the same reaction conditions but the enzyme Proteinase K was omitted.

After polymerization, the enzymes and dNTPs were discarded and immobilized DNA was washed with 1× AB.

Table 2. List of PCR primers (5'→3'); one of the inner primers is biotin labeled

SNP	Outer forward primer	Outer reverse primer	Inner forward primer	Inner reverse primer
1	ctcccagaagaggggtgg	cgtgtgtgcacggcctgt	gggtagcccttgagacc	cacggcctgtctctctctct
2	tccatgccacagacctcaat	cttgttactcgtctgatgg	cctcaatgacaagagtgtgca	tactcgtctatggcaagctccag
3	tgaatataacctcagtgagg	tgctccactactctcaga	agtgaggtgaggagacaatta	ccactactcagaacctcttgaa
4	acctgggggctaagaaga	aggacgcactctgctgtg	gaagaagagctgtctctggg	tcttctgtgtacggaggagg
5	gatgcctgggggtgtgaag	tagcagaggaaggagtggc	tgaagaaaagagccttgggtt	aaggagtggccccaggttccc
6	agaactactctgcaaaagct	gcaaggctggctcccaag	gctcccgtgaggggtgg	gggtcccaagaagaactctgac
7	cttggacttggtgagcgata	caaccattgccaggcaagg	gcgatagggaagtgttagc	gccaggcaaggagaccgggg
8	ctcagcagtgcatgatgaa	gataaggctactaccagtg	tgatgaaatgagtgtctgac	ctagctaccagtgccaccacac
9	ccctcctaagtggcagcca	cagtgctcactctctgtgg	gccatgtctccataacc	ctctgtggcgttctgcgga
10	agctctgcagcgtgtccc	gctgagcccgtgggtgg	gccctcaagctgtaggcc	ggtgggtgggtggagggactg
11	tgctccgctcttccctcc	actgcatgaccaccggag	cctctctgtaaaaatgggga	gaccaccggagacataagctact
12	gactgctgttaggagatttc	tacaggcaggttaagacca	gatttctccagataggctc	gcaggttaagaccaaaacttca
13	gggtcacattctgcttaag	aggagagacacatggggcc	cattctggcttaagcgaatttc	tggggcgagaggggacgggaaa

Table 3. Allele-specific extension primers (5'→3') with 5'-tags complementary to the tags on the chip, named after the complementary tag sequence and the SNP allele

Name	Primer sequence
T2_1G	GCAGATCAATTAATACGATACCTGCG_GTGTC-CTGTCTCTGAGACG
T3_1A	GGTTCTGTCTTCTGTTGACATGAGG_GTGTCC-TGTCCTCTGAGACA
T4_2A	TTAGTCTCCGACGGCAGGCTTCAAT_ATGGCA-AAGTCCAGGGCACA
T5_2G	CTGTGACAGAGCCAACACGCAGTCT_ATGGCA-AAGTCCAGGGCACC
T6_3G	CCTGGTGGTTGACTGATCACCATAA_CTCAGA-AACTTCTTGAACCCAG
T7_3A	GCATGTATAGAACATAAAGGTGTCT_CTCAGA-AACTTCTTGAACCCAA
T8_4G	GCTAGATGAAGAGCAAGCGCATGGA_TGTACG-GGAGGAGGCATGG
T9_4A	TACAACCGACAGATGTATGTAAGGC_TGTACG-GGAGGAGGCATGA
T10_5G	TTCAATCTGGTCTGACCTCCTTGTG_GCCCA-GGTTCCCTAGTCG
T11_5A	ACACGATGTGAATATTATCTGTGGC_GCCCA-GGTTCCCTAGTCA
T12_6T	TTGAAGTTCGAGAATCGTATGTGT_CAAGAA-AGCACTGACCTGGAT
T13_6C	AACGTCTGTTGAGCACATCCTGTAA_CAAGAA-AGCACTGACCTGGAC
T16_7T	AAGCAGTCTGTCAGTCAGTGCCTGAA_AAGGA-AGCACCGGGGCAAAT
T17_7C	AATGATGCTCTGCGTGATGTTG_AAGGAA-GCACCGGGGCAAAC
T18_8G	AATACACGAAGGAGTTAGCTGATGC_GTGTCC-ACCACACCCCTGG
T19_8A	GCTGTTAATCATTACCGTGATAACGCC_GTGT-CCACCACACCCCTGA
T20_9G	GCGGAACGGTCAGAGAGATTGATGT_TGGCG-GTTGTGCGGATTCAG
T21_9A	GTTATGGTCAGTTCGAGCATAAGGC_TGGCGG-TTGTGCGGATTCAA
T22_10T	TTACCTATGATTGATCGTGGTGATATCCG_GG-TGGAGGGACTGGAACAT
T23_10C	GCTGTGGCATTGACGAGATTAAGG_GGTGGA-GGGACTGGAACAC
T24_11T	TGACGTCATTGTAGGCGGAGAGCTA_GGAGAC-ATAAGTACTTCTCT
T25_11C	TCAATAATCAACGTAAGGCGTTCCT_GGAGAC-ATAAGTACTTCTCCC
T26_12T	TTATCGGCTACATCGGTAAGGACT_AAAGAC-CAAAACATTTCAAGACAT
T27_12C	CCATTTCGCTGGTTTCATTCGTGA_AAAGAC-CAAAACATTTCAAGACAC
T36_13C	AAGTACTGCAAGTGCTCGCAACAT_CTTAAG-CGCCAATTTCTTTTT
T37_13T	CTGCATTCTGCGGTAAGCACGAACT_CTTAAG-CGCCAATTTCTTTTT

To release the extended primers, the immobilized DNA was treated with 7 μ l of 0.1 M NaOH (5 min at room temperature) and the supernatants, i.e. the PrASE products were neutralized with 0.1 M HCl and 10 \times AB to a total volume of 16 μ l. An aliquot of 16 μ l of 2 \times hybridization buffer (HB) (10 \times SSC, 0.4% SDS) and 0.5 μ g SSB was finally added to a total volume of 32 μ l.

The fluorescently labeled PrASE products, each containing a specific signature tag at the 5' end, were then hybridized to the generic tag arrays on the glass slide for 60 min at 50°C.

After hybridization, the slide was washed with 50°C prewarmed 2 \times SSC/0.1% SDS for 6 min, then with 0.2 \times SSC at room temperature for 1 min and finally with 0.1 \times SSC at room temperature for 1 min. After the washing steps, the slide was dried by a brief centrifugation.

Data analysis

Data were obtained by scanning the slide with an Agilent scanner (Agilent Technologies, CA), generally optimal with the laser power at 40% (to avoid saturated signals at 100%). Data were analyzed using the GenePix Pro 5.0 software (Axon instruments). The median local background intensities were subtracted from the median intensities of the spots by GenePix Pro 5.0 and the data were analyzed in Microsoft Excel, where the mean values of fluorescence intensities of the triplicates for each signature tag were used to calculate the allelic fractions of the 13 SNPs for each individual.

Effect of Proteinase K on extension length

To investigate the effect of Proteinase K on primer extension, we designed four synthetic oligonucleotide templates based on the sequence of SNP 13. All four templates were designed to match the extension primer for the C-allele of SNP 13. Two extra nucleotides were included at the 3' end of the synthesized templates to generate mismatches to the extension primer and avoid extension of the synthesized templates. Each of the four oligonucleotide templates was designed to contain only one G-nucleotide (downstream of the annealed primer). This G-nucleotide was situated 5, 10, 15 or 20 nt downstream of the annealed primer, respectively. To be able to investigate the effect of Proteinase K on the extension event, the reactions were carried out with only Cy5-labeled dCTPs together with native dTTP, dATP and dGTP. The extension length was analyzed by both ASE and PrASE with three different amounts (20, 40 and 80 μ g) of Proteinase K.

Removal of extension primer excess

To evaluate the importance of a liquid-phase PrASE, where it is possible to immobilize and wash biotinylated templates by robotic systems capable of handling magnetic beads, we performed the PrASE and hybridization to the tag microarray with and without wash after primer annealing. The wash step thus serves to remove the excess of extension primers. Six individuals (individuals 1, 7, 13, 19, 25 and 31) were investigated and PrASE was performed for all 13 SNPs with and without removal of the excess of primers. This comparison was practically possible by the addition of extension primers after the PrASE reaction. After the PrASE reaction, the mixture was divided into two halves of 15 μ l each. To one half of the PrASE products, we added an extension primer solution (0.08 μ M of each primer) and to the other half 1 \times AB. The primer solution in the first mixture (half) corresponds to the amounts that are washed away by the robot after primer annealing. The solutions were then hybridized to the generic tag arrays and the data were analyzed. The comparison data analysis (for each individual) was performed by calculating the signal intensities from all spots (26 allele-specific primers in triplicates) and then taking an average of total signal intensities.

Effect of multiplex PrASE genotyping

In order to evaluate the effect of multiplexing in terms of sensitivity of the PrASE, 12 different PrASE reactions were performed in a ladder from a simplex with only one pair of ASE primers to a multiplex of 12 primer pairs, i.e. a multiplex PrASE of 12 SNPs for sample (individual) number 27. One of the 13 SNPs, SNP 13 (rs740841), was left out owing to the design of the robot.

Investigation of SNP 1 (rs17429)

To control that the assay works for SNP 1 (rs17429), which only generated homozygous genotypes for the wild-type allele C in the investigated individuals, a synthetic oligonucleotide template for each allele (i.e. SNP 1 C-template and SNP 1 T-template) was ordered. The templates correspond to the inner PCR products of the SNP. The synthetic templates were investigated by PrASE in three different reactions. The reactions contained 2 pmol of SNP 1 C-template, 2 pmol of SNP 1 T-template and 1 pmol of each C- and T-templates, respectively.

Pyrosequencing

Pyrosequencing was performed according to manufacturer's instructions (Biotage, Uppsala, Sweden) on 8 individuals (18, 20, 21, 22, 23, 28, 33 and 34), where 12 of the 13 SNPs were genotyped (SNP 7 excluded), generating 96 genotypes, i.e. one 96-well plate. The individuals were first amplified with simplex inner PCR for each of the 12 SNPs. The pyrosequencing primers were located 2 nt from the SNP position.

RESULTS AND DISCUSSION

Different assays have been developed for high-throughput SNP genotyping, where microarray-based formats are preferable. Here, we have utilized the polymerase ability to discriminate between extensions of 3'-terminus matched and mismatched allele-specific primers in the presence of Proteinase K, a protein-degrading enzyme. This is facilitated by the differences in kinetics of matched and mismatched primer elongation. Only perfectly matched primers are thus extended, while in the case of slower reaction kinetics (mismatched primer extension) the protease degrades the polymerase before any nucleotides are incorporated. The method of PrASE was employed to analyze 13 SNPs in 36 individuals in a multiplex assay using generic tag arrays for the detection of the SNP alleles.

The principle of the novel PrASE method for multiplex SNP genotyping is shown in Figure 1. A nested PCR, in which both outer and inner reactions are multiplexed, is designed to yield 71 bp and 46–52 bp outer and inner PCR products, respectively. One of the inner PCR primers is biotinylated to facilitate immobilization onto streptavidin-coated magnetic beads and to render the target DNA single stranded. Immobilized ssDNA can be obtained by alkali elution of the non-biotinylated strand and, thus, ASE may be performed either on the eluted non-biotinylated strand or on the immobilized strand. To maximize the signal intensity as well as to develop an automated protocol, we chose to perform the extensions on the solid support (see below). Nevertheless, after hybridization of

3'-terminus allele-specific primers, the extension reaction with fluorescently labeled nucleotides accompanied with protease is performed on the beads. As shown in Figure 1, the allele-specific extended products are released by alkali elution, neutralized and hybridized to a generic tag array.

The efficiency of the protease was investigated by the analysis of the number of incorporated nucleotides in the PrASE reaction. Without the inclusion of protease in the reaction, the extension would continue to the end of the template. Instead, the addition of Proteinase K in the reaction is expected to terminate the extension after incorporation of a limited number of nucleotides (owing to the degradation of the polymerase). Thus, to perform this investigation, four synthetic oligonucleotide templates corresponding to the sequence of SNP 13 were designed. The templates were modified to only contain one G-nucleotide, positioned 5, 10, 15 or 20 nt downstream of the annealed primer, respectively. Extensions with different amounts of Proteinase K (0, 20, 40 and 80 μ g) were performed where only nucleotide dCTP was Cy5-labeled, while the other 3 nt were native. In this way, the number of nucleotides that are incorporated during the polymerase extension can be estimated by comparing the signals for the four different templates. If Proteinase K is omitted in the reaction, the extension would continue to the end of the template and, therefore, we would detect signals from all four templates. On the other hand, with inclusion of Proteinase K, the extension reactions will terminate after incorporation of some nucleotides and only generate signals from the templates with a G-nucleotide close to the extension primer. As shown in Figure 2, the signal intensities decline with higher amounts of Proteinase K. This decrease is more evident for the templates with an incorporated labeled nucleotide that is distanced 15 and 20 nt from the start of the polymerization. The normal amount of Proteinase K that we use in the PrASE reactions is 20 μ g and as it can be seen, this amount allows incorporation of at least 10 nt. However, in a normal PrASE reaction, 2 nt (C and T) are fluorescently labeled to ensure incorporation of at least one labeled nucleotide. With higher amounts of Proteinase K, such as 40 and 80 μ g, it is still possible to generate signals from the '5' and '10' nt templates, but too weak for the templates with G-nucleotide at longer distance.

A robotic procedure capable of performing different steps was developed. This automated system facilitates immobilization of the biotin-labeled fragments and consequently allows washing of the buffers and other solutions. One of these steps involves annealing of allele-specific primers and removal/wash of the excess of the primers (i.e. the primers that have not been annealed) and has proven to be important in improving the sensitivity. Since the excess and potentially non-extended primers also contain signature tags, these will hybridize to the tags on the microarray. Thus, by removing the excess of primers, a considerably higher signal is obtained. This is exemplified in Figure 3, in which six individuals (number 1, 7, 13, 19, 25 and 31) were compared with and without removal of extension primers. Figure 3 shows a direct comparison between average signal intensities for the 13 investigated SNPs (see below) using the PrASE protocol. At least a 15-fold difference in average signal intensity can be seen in all six individuals. This demonstrates the advantage to remove the excess of unannealed primers to achieve a system with higher sensitivity. However, it should be mentioned that

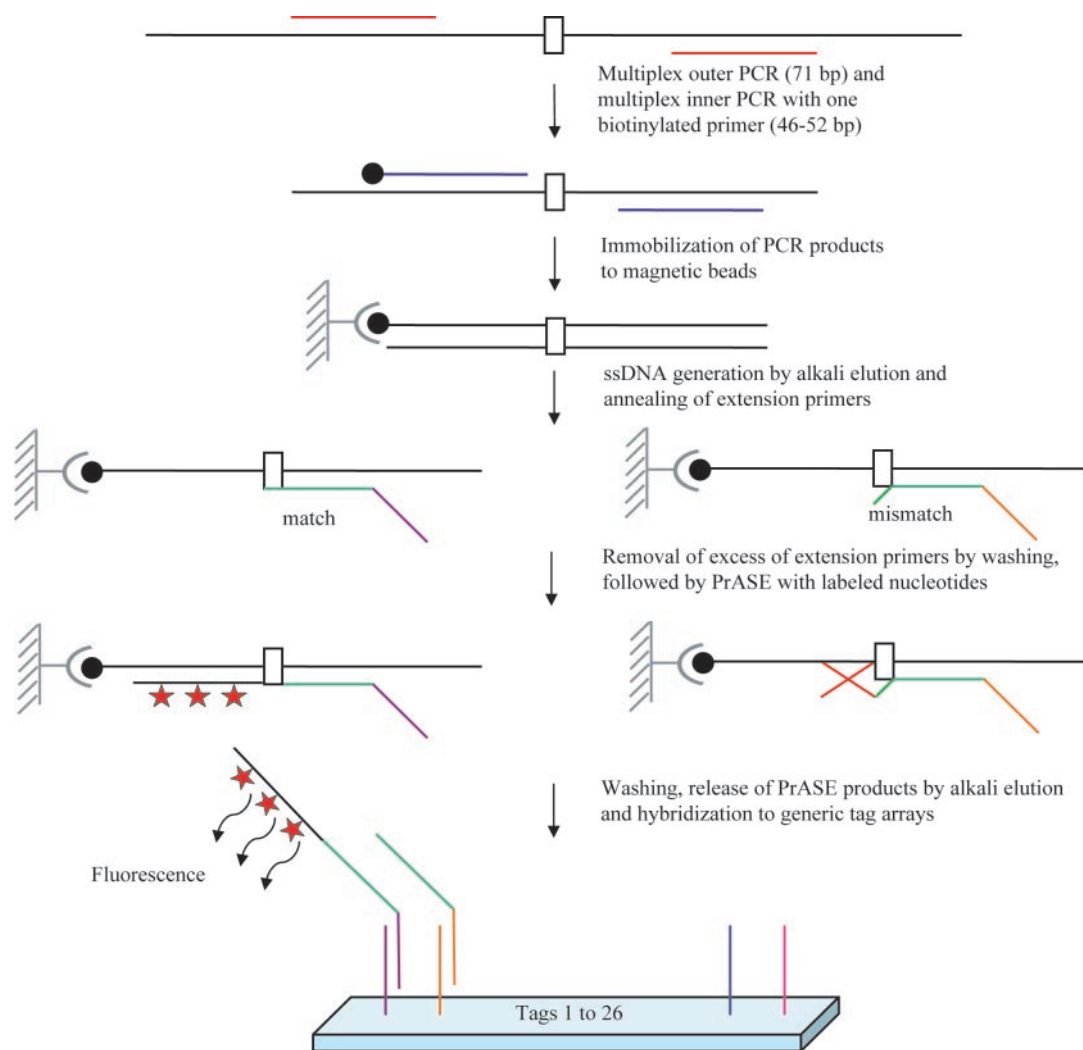


Figure 1. Schematic drawing of the PrASE procedures. A nested multiplex PCR amplification is performed with biotinylated inner PCR primers, generating short PCR products. The biotinylated PCR products are immobilized to streptavidin-coated magnetic beads, and ssDNA is generated by alkali elution of the non-biotinylated strand followed by annealing of ASE primers to the immobilized DNA strands. A washing step, where unannealed extension primers are removed, is then followed by PrASE with Cy5-labeled dCTPs and dUTPs. The last steps involve removal of unreacted nucleotides and PrASE enzymes, release of PrASE products by alkali elution and hybridization to tag microarrays. All procedures except the last step of hybridization to tag arrays are automated.

the increased sensitivity did not affect the specificity and in both cases the PrASE reactions gave similar genotyping results.

A high-throughput SNP genotyping assay requires high multiplexing capacity with high sensitivity. In order to evaluate the effect of multiplexing in the detection steps, 12 different PrASE reactions were performed in a ladder from a simplex with only one extension primer pair to a multiplex of 12 primer pairs (i.e. a multiplex PrASE of 12 SNPs). One of the SNPs, SNP 13, was not included owing to the design of the robot with a 12-pipette head. Figure 4 shows the logarithmic light signal intensities of all the 12 SNPs at different degrees of multiplexing. As shown in Figure 4, the signal intensities are linear and do not decrease by increased level of multiplexing. Consequently, as the multiplex PrASE assay shows no decrease in sensitivity in comparison with a simplex assay, the limitations of the degree of multiplexing rather lies in the multiplex PCR.

As mentioned above, a total of 36 individuals were analyzed by the ASE assay based on the extension kinetics of the

polymerase and the protease. In order to evaluate the accuracy of the technique, 12 of the 13 SNPs (SNP 7, rs752471 excluded) were analyzed with pyrosequencing (25,26) on 8 of the 36 individuals used in this study. The sequences obtained by pyrosequencing were all manually checked and edited. The PrASE results correlated 100% with the unambiguous pyrosequencing results (data not shown). However, one genotype obtained by pyrosequencing (SNP 13, rs740841, for individual 22) generated an ambiguous sequence (data not shown).

To evaluate the effect of a protease on genotyping and accuracy of ASE reactions, the genotype results of the 36 individuals were compared both with and without inclusion of protease. Figure 5 exemplifies raw data image of PrASE (left) and conventional ASE (right) for individual 13. As shown, each array block contains identical spots in triplicate. Here in Figure 5, the first replicate has been used to illustrate the position of pairs of signature tags representing allele-specific primers for each SNP. For example, the spots in

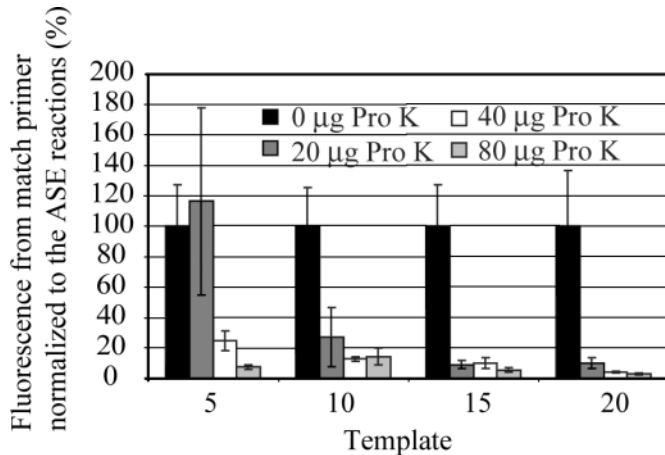


Figure 2. The effect of a protease on extension length. Four synthetic templates have been analyzed by conventional ASE and PrASE with three different amounts of Proteinase K (20, 40 and 80 µg). The templates only contain one G-nucleotide, at different distances downstream of the extension primer, and differ from each other only by the G position. The primers are then extended with Cy5 labeled dCTPs together with native dGTPs, dATPs and dUTPs, generating a signal only if the incorporated nucleotides cover the G-nucleotide position. The fluorescent signal obtained by ASE (0 µg Proteinase K), where the extension is not hindered, has been used to normalize the fluorescence signals acquired by the PrASE reactions. Note that the fluorescent signals from the extended primers are normalized to the signals from the ASE reactions (black bars). The standard deviations are based on analysis of nine data points.

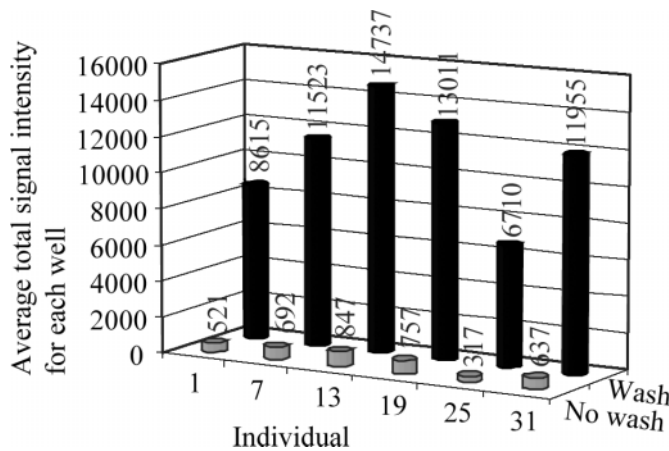


Figure 3. The effect of removal of unannealed extension primers on signal intensities on the arrays. Six individuals have been analyzed by PrASE on tag arrays both with and without the washing away of excess primers. The different individuals in this analysis are indicated on the x-axis. The y-axis shows an average of total signal intensities (for 13 SNPs) obtained for each individual.

the first square are representing allele-specific primers for SNP 1. The ovals in the second replicate indicate the SNPs with similar genotyping results in ASE and PrASE. The ovals in the third replicate, however, show where PrASE and ASE give conflicting genotypes. As compared with the cluster diagrams (see Figure 6), the PrASE results are correct when disagreement(s) with ASE occur.

To analyze the genotypes of the SNPs, the extension signals from the allele-specific primer pairs were used. The relative allelic fractions were calculated by taking the fluorescent

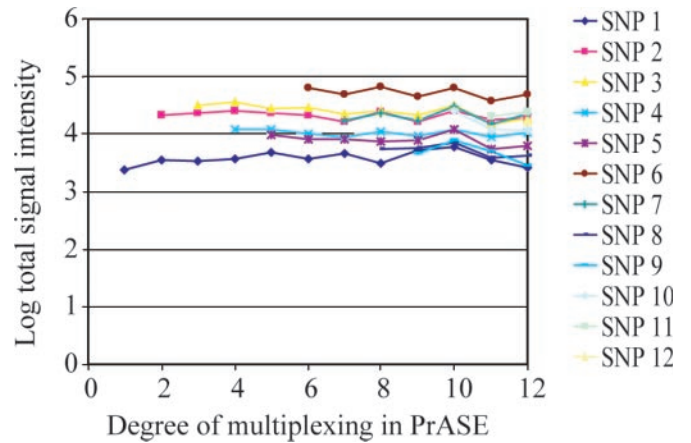


Figure 4. The effect of multiplexing on the sensitivity of PrASE. The logarithmic value of the total signal intensity for each SNP (y-axis) is plotted toward the number of SNPs, i.e. the degree of multiplexing in the PrASE reaction. A total of 12 different PrASE reactions were performed, from a simplex to a 12-plex, where SNP 1 has been analyzed in all 12 reactions down to SNP 12, which only has been analyzed in the 12-plex. To achieve the different degrees of multiplexing, simplex PCR products for each of the 12 SNPs were pooled.

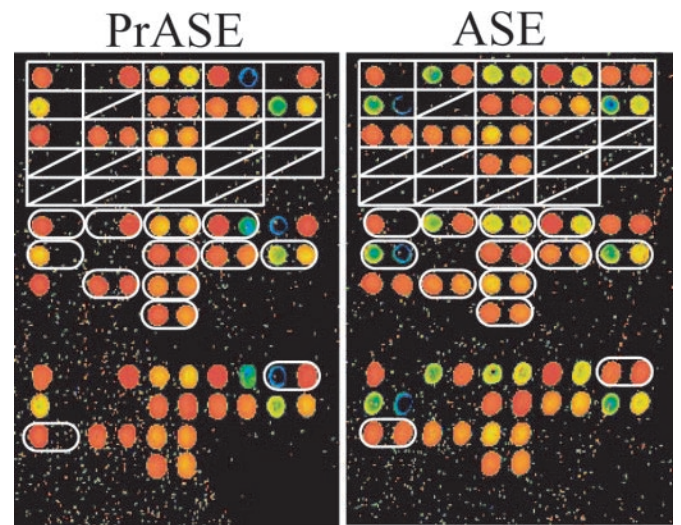


Figure 5. Array image for one individual analyzed by both PrASE (left) and ASE (right). The 13 SNPs are in triplicates, where in the first replicate the SNP positions are marked with white squares in the order of SNP 1 to 13 from upper left to lower right corner. In the second replicate, the SNPs with similar results in both PrASE and ASE are marked (white ovals), while the third replicate indicates conflicting results between PrASE and ASE (marked with white ovals). Note that all spotted tags (48) were not used in this assay (indicated with white slashed boxes).

signal intensity from spot1/(spot1 + spot2), where spot1 and spot2 correspond to primer extension of the first and the second allele, respectively (16), and are the mean value from the triplicates. This calculation gives allelic fractions of ~0.5 for heterozygous and close to 0 and 1, respectively, for the homozygous samples.

To illustrate the results, the 13 SNPs are plotted in separate diagrams in Figure 6. The cluster diagrams are shown both for PrASE (blue upper panel) and ASE (red lower panel), for the individual SNPs. As can be seen, when Proteinase K is omitted

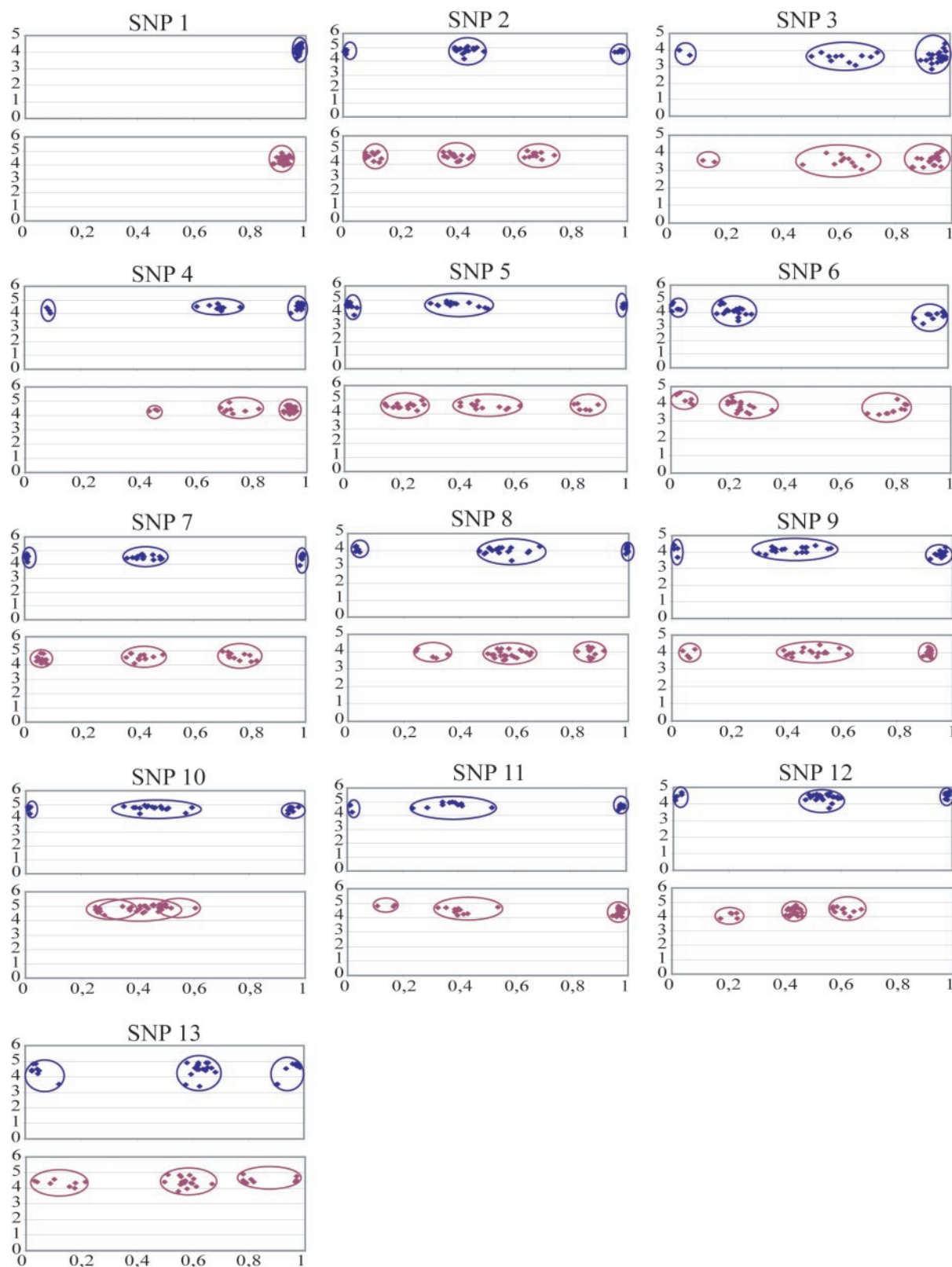


Figure 6. Genotyping results for PrASE (blue upper panels) and ASE (red lower panels). The 36 analyzed samples are visualized in 13 cluster diagrams for each SNP. The x-axes represent allelic fractions that are calculated by the equation $\text{spot1}/(\text{spot1} + \text{spot2})$, where spot1 and spot2 correspond to fluorescent signal intensity from primer extension of the first and the second allele, respectively. The y-axes represent logarithmic value of the total fluorescent signal intensity. Circles indicate the different genotypes, where samples scored as heterozygous are situated in the middle circle with an optimal allelic fraction close to 0.5. Homozygous samples for the first allele and the second allele are located in the circles with allelic fractions close to 1 and 0, respectively. Note that, for SNP 10, the scoring with ASE is impossible, while PrASE generates distinct clusters.

in the allele-specific reactions, the clusters are not as well distinguished as in the case of PrASE. In fact, for SNP 10, the genotyping by ASE is completely incorrect and no separation of the clusters can be observed. Furthermore, for SNP 13, the genotypes generated by ASE with allelic fractions around 0.8 could belong to either the heterozygous C/T or the homozygous C cluster, but are well separated with PrASE. For SNP 4, the clusters for homozygous C and heterozygous C/T are very close to each other. In addition, the cluster for homozygous T is situated approximately at allelic fraction 0.5 (an indication of heterozygous samples). The sum of these observations for SNP 4 makes it difficult to accurately and robustly genotype this SNP with ASE. The same problem can be seen for SNP 12, where the clusters are close to each other. However, the case of SNP 12 is not as difficult as SNP 4. Nevertheless, while many of the SNPs, genotyped with ASE, show difficulties in discrimination of matched and mismatched primers, addition of Proteinase K in the PrASE renders distinct and partitioned clusters.

Table 4 shows in detail the allele scores for each SNP in the 36 samples. These results show that all possible alleles of homozygous and heterozygous could be detected in 12 of

the 13 investigated SNPs. However, owing to low frequency of minor allele for SNP 1, it was not possible to score the T-allele in any of the analyzed 36 individuals. To ensure that the oligonucleotides for this SNP (tags on the microarrays and the ASE primer for the T-allele) were correctly synthesized and worked properly, a synthetic oligonucleotide template for each allele (i.e. SNP 1C template and SNP 1T template) were designed. The templates correspond to the inner PCR products of the SNP. The synthetic templates were investigated by PrASE in three different reactions, as homozygous for wild-type (CC), as homozygous for mutant (TT) and as heterozygous (CT). The PrASE reactions on C and T templates generated homozygous genotypes and the mixture of the templates was scored as heterozygous (data not shown).

In conclusion, this study demonstrates a novel approach for multiplex SNP genotyping based on 3'-terminus ASE. To enhance the accuracy of ASE reactions and prevent non-specific extension of 3'-termini primers, a protein-degrading enzyme was included in the reactions. Thirteen SNPs were accurately genotyped in 36 individuals by obtaining complete partitioned clusters for each SNP. The multiplex genotyping was facilitated by the use of generic tag arrays, which is a

Table 4. Genotyping results of the 36 individuals for the 13 SNPs

Individual	SNP genotype		3 C/T	4 C/T	5 C/T	6 A/G	7 A/G	8 C/T	9 C/T	10 A/G	11 A/G	12 A/G	13 C/T
	1 C/T	2 T/C											
1	CC	TT	TT	CC	CT	AA	AG	CT	CT	GG	AA	AG	CT
2	CC	TT	CC	CC	TT	GG	GG	CC	CC	AA	AG	AA	CT
3	CC	TC	CT	CC	TT	AA	GG	CT	CC	AA	AA	AG	CC
4	CC	CC	CC	CC	CT	AG	GG	CC	TT	GG	AA	AA	CT
5	CC	TC	CC	CT	TT	AG	GG	CT	CT	GG	AA	AG	TT
6	CC	TC	CC	CC	TT	AG	AA	CT	CC	AG	GG	AG	CC
7	CC	CC	CC	CC	TT	AA	AG	TT	CC	AA	AA	AG	CC
8	CC	TT	CC	CC	TT	GG	AG	CC	CC	AG	AA	AA	CC
9	CC	TC	CC	CC	CC	AG	AA	CC	CT	GG	AA	AA	CT
10	CC	CC	CC	CC	CT	AG	GG	CT	CT	AG	AA	AG	TT
11	CC	TC	CC	CC	CT	AG	AA	CT	CC	GG	AA	AA	CT
12	CC	CC	CT	CC	TT	AA	AG	CT	CC	GG	AA	AG	CT
13	CC	CC	CT	CC	TT	AA	AG	CT	CT	AA	AG	AG	CT
14	CC	CC	CC	CT	TT	GG	GG	CC	CT	AG	AA	AG	CT
15	CC	TC	CC	CC	CT	AG	GG	CT	CT	AA	AG	AA	CT
16	CC	TT	CT	CC	CC	AG	AA	TT	CC	AG	AA	AG	TT
17	CC	TT	CT	CC	CT	AG	AG	CT	CT	AG	AA	AG	CC
18	CC	CC	CC	TT	CT	AA	AG	TT	CT	AG	AA	AA	CC
19	CC	TT	CC	TT	TT	GG	GG	CT	TT	AG	AG	GG	CT
20	CC	TT	CC	CT	TT	AG	AA	TT	CT	AA	AA	AG	CT
21	CC	TC	CC	CC	CC	AG	AA	CC	CT	AG	AG	AG	CC
22	CC	TC	CC	CC	TT	AG	AA	CC	CC	AG	AG	AA	CC
23	CC	TC	CC	CT	CT	AA	AA	CT	CC	GG	GG	GG	TT
24	CC	TC	CC	CC	TT	AG	AG	CT	CC	AG	AA	AG	CT
25	CC	TC	TT	CC	CC	AG	AA	CT	CT	AG	AG	GG	CT
26	CC	CC	CC	CC	CT	GG	AG	CC	CC	AG	GG	AG	TT
27	CC	TT	CT	CT	TT	GG	AA	TT	TT	AG	AA	AA	CT
28	CC	TC	CC	CC	CT	AG	AA	CC	CC	GG	AA	AG	CT
29	CC	TT	CT	CT	CT	AA	AG	CC	CT	AG	AA	AG	TT
30	CC	TC	CT	CC	CT	AA	AG	CT	CT	AA	AG	AG	CT
31	CC	TC	CT	CC	TT	AA	GG	CT	CT	AG	AG	GG	TT
32	CC	CC	CC	TT	CC	AG	AG	CT	CC	AG	AG	AG	CT
33	CC	CC	CC	CC	TT	AG	AG	CT	TT	GG	AG	GG	CC
34	CC	TC	CT	CC	TT	AG	GG	CT	CC	AA	AG	AA	CC
35	CC	TT	CT	CT	CC	AA	GG	CT	CT	AG	AA	AG	TT
36	CC	TC	CC	CT	CT	AG	GG	CT	CT	AG	AA	AG	CT
Summary													
Homo 1	36	10	23	25	6	11	11	10	15	8	21	10	10
Hetero	0	16	11	8	13	19	13	21	17	19	12	21	18
Homo 2	0	10	2	3	17	6	12	5	4	9	3	5	8

flexible alternative to extension on the array surface (17). At least 15-fold increased sensitivity and consequently increased signal-to-noise was achieved by removal of the excess and unannealed signature tag primers. This was accomplished by an automated solid-phase magnetic bead system, which also enabled automation of all steps after PCR amplification until hybridization to the tag arrays. The effect of protease on the extension length was also investigated and we conclude that in the cases where a controlled extension length is desired the PrASE system can be employed. In addition, increased level of multiplexing did not affect the signal intensity, which is promising for future application of the PrASE assay with higher level of multiplex genotyping.

ACKNOWLEDGEMENTS

This work was supported by grants from the Swedish Research Council, the Knut and Alice Wallenberg Foundation (Wallenberg Consortium North) and Swedish agency for innovation systems (Vinova). Funding to pay the Open Access publication charges for this article was provided by the Swedish Research Council.

Conflict of interest statement. None declared.

REFERENCES

- Ahmadian, A. and Lundeberg, J. (2002) A brief history of genetic variation analysis. *BioTechniques*, **32**, 1122–1124, 1126, 1128 passim.
- Syvänen, A.C. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Rev. Genet.*, **2**, 930–942.
- Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J. *et al.* (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280**, 1077–1082.
- Howell, W.M., Jobs, M., Gyllensten, U. and Brookes, A.J. (1999) Dynamic allele-specific hybridization. A new method for scoring single nucleotide polymorphisms. *Nat. Biotechnol.*, **17**, 87–88.
- Prince, J.A., Feuk, L., Howell, W.M., Jobs, M., Emahazion, T., Blennow, K. and Brookes, A.J. (2001) Robust and accurate single nucleotide polymorphism genotyping by dynamic allele-specific hybridization (DASH): design criteria and assay validation. *Genome Res.*, **11**, 152–162.
- Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) A ligase-mediated gene detection technique. *Science*, **241**, 1077–1080.
- Nilsson, M., Malmgren, H., Samiotaki, M., Kwiatkowski, M., Chowdhary, B.P. and Landegren, U. (1994) Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science*, **265**, 2085–2088.
- Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. and Syvänen, A.C. (1997) Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.*, **7**, 606–614.
- Hirschhorn, J.N., Sklar, P., Lindblad-Toh, K., Lim, Y.M., Ruiz-Gutierrez, M., Bolk, S., Langhorst, B., Schaffner, S., Winchester, E. and Lander, E.S. (2000) SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. *Proc. Natl Acad. Sci. USA*, **97**, 12164–12169.
- Dubiley, S., Kirillov, E. and Mirzabekov, A. (1999) Polymorphism analysis and gene detection by minisequencing on an array of gel-immobilized primers. *Nucleic Acids Res.*, **27**, e19.
- Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L. and Syvänen, A.C. (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.*, **10**, 1031–1042.
- 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.
- Kwok, S., Kellogg, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C. and Sninsky, J.J. (1990) Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.*, **18**, 999–1005.
- Day, J.P., Bergström, D., Hammer, R.P. and Barany, F. (1999) Nucleotide analogs facilitate base conversion with 3' mismatch primers. *Nucleic Acids Res.*, **27**, 1810–1818.
- Ayyadevara, S., Thaden, J.J. and Shmookler Reis, R.J. (2000) Discrimination of primer 3'-nucleotide mismatch by *Taq* DNA polymerase during polymerase chain reaction. *Anal. Biochem.*, **284**, 11–18.
- Ahmadian, A., Gharizadeh, B., O'Meara, D., Odeberg, J. and Lundeberg, J. (2001) Genotyping by apyrase-mediated allele-specific extension. *Nucleic Acids Res.*, **29**, e121.
- O'Meara, D., Ahmadian, A., Odeberg, J. and Lundeberg, J. (2002) SNP typing by apyrase-mediated allele-specific primer extension on DNA microarrays. *Nucleic Acids Res.*, **30**, e75.
- Ericsson, O., Sivertsson, A., Lundeberg, J. and Ahmadian, A. (2003) Microarray-based resequencing by apyrase-mediated allele-specific extension. *Electrophoresis*, **24**, 3330–3338.
- Käller, M., Ahmadian, A. and Lundeberg, J. (2004) Microarray-based AMASE as a novel approach for mutation detection. *Mutat. Res.*, **554**, 77–88.
- Gharizadeh, B., Käller, M., Nyren, P., Andersson, A., Uhlen, M., Lundeberg, J. and Ahmadian, A. (2003) Viral and microbial genotyping by a combination of multiplex competitive hybridization and specific extension followed by hybridization to generic tag arrays. *Nucleic Acids Res.*, **31**, e146.
- Hall, B.L. and Finn, O.J. (1992) PCR-based analysis of the T-cell receptor V beta multigene family: experimental parameters affecting its validity. *BioTechniques*, **13**, 248–257.
- Tchernitchko, D., Legendre, M., Delahaye, A., Cazeneuve, C., Niel, F., Goossens, M., Amselem, S. and Girodon, E. (2003) Clinical evaluation of a reverse hybridization assay for the molecular detection of twelve MEFV gene mutations. *Clin. Chem.*, **49**, 1942–1945.
- Ehn, M., Nilsson, P., Uhlen, M. and Hober, S. (2001) Overexpression, rapid isolation, and biochemical characterization of *Escherichia coli* single-stranded DNA-binding protein. *Protein Expr. Purif.*, **22**, 120–127.
- Gräslund, T., Hedhammar, M., Uhlen, M., Nygren, P.Å. and Hober, S. (2002) Integrated strategy for selective expanded bed ion-exchange adsorption and site-specific protein processing using gene fusion technology. *J. Biotechnol.*, **96**, 93–102.
- Ahmadian, A., Gharizadeh, B., Gustafsson, A.C., Sterky, F., Nyren, P., Uhlen, M. and Lundeberg, J. (2000) Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.*, **280**, 103–110.
- Ronaghi, M., Uhlen, M. and Nyren, P. (1998) A sequencing method based on real-time pyrophosphate. *Science*, **281**, 363, 365.