Research Report

Genotyping of SNPs in a Polyploid Genome by Pyrosequencing™

Andreas M. Rickert[1,2]**,**
Andreas Premstaller[1]**,**
Christiane Gebhardt[2]**, and**
Peter J. Oefner[1]

[1]Stanford Genome Technology
Center, Palo Alto, CA, USA and
[2]Max-Planck-Institute for
Breeding Research, Cologne,
Germany

ABSTRACT

Single-nucleotide polymorphisms (SNPs) are the most frequent DNA sequence variations, and they have become increasingly popular markers for association studies. Allelic discrimination of the mostly binary SNPs has been reported for diploid species, mainly the human, but not for polyploid genomes such as the agriculturally important crops. In the present study, we analyzed the applicability of pyrosequencing to genotyping SNPs in tetraploid potatoes. Out of 94 polymorphic loci tested, 76 (81%) proved to be amenable to allelic discrimination by pyrosequencing. An additional locus could be genotyped by the addition of an ssDNA binding protein to the pyrosequencing reaction. Of the remaining 17 loci, two failed because of the presence of paralogs in the genome, while in the other cases, self-annealing of the primer or template at the low reaction temperature (28°C) employed in pyrosequencing rendered allelic discrimination impossible. The quantitative precision of pyrosequencing was found to be similar to that of conventional dideoxy sequencing and single-nucleotide primer extension. Except for some sequence-specific limitations, pyrosequencing appears to be an appropriate method for genotyping SNPs in polyploid species because it is possible to distinguish not only between homo- and heterozygosity but also between the different heterozygous states.

INTRODUCTION

Numerous methods have been evaluated on diploid species for the allelic discrimination of single-nucleotide polymorphisms (SNPs). Among others, they include allele-specific PCR (3,19), oligonucleotide ligation (16,20), nick translation PCR (18), allele-specific hybridization (4,28,30), serial invasive signal amplification reaction (11), T_m-shift genotyping (7), pyrosequencing (1,2, 21), and minisequencing (14). However, no single method for scoring SNPs in diploid genomes has been widely accepted (29), and documented studies of their applicability to polyploid organisms are lacking. In contrast with diploid genomes, the ratio of SNP alleles can vary in polyploid genomes. In tetraploids, for instance, five allelic ratios are distinguishable: 4:0, 3:1, 2:2, 1:3, and 0:4. The exact determination of allele frequency in individuals and, eventually, in populations is essential in association studies; therefore, the quantitative accuracy of genotyping is critical.

The present study demonstrates that pyrosequencing can reliably distinguish the five different allelic states of a binary SNP in tetraploid potato, provided that the ratio of incorporated nucleotides does not exceed a ratio of 1:10 as noted previously (25–27). Finally, the quantitative precision of pyrosequencing was compared with that of conventional dye-terminator cycle sequencing and single-nucleotide primer extension in combination with completely denaturing HPLC (DHPLC).

MATERIALS AND METHODS

Sequences and Plant Material

The SNPs 44/114, 44/174, and 66/212 had been identified by comparative Sanger DNA sequencing of the two potato cultivars, Leyla and Nikita (Rickert et al., unpublished data). Genotyping was carried out in the parents and 52 F1 plants of a cross between Leyla and Nikita that had been provided by SAKA-RAGIS Pflanzenzucht GbR, Windeby, Germany. Genomic DNA was extracted with the DNeasy® Plant Maxi Kit (Qiagen, Valencia, CA, USA). The DNA concentrations were measured with the PicoGreen® dsDNA Quantitation Kit (Molecular Probes, Eugene, OR, USA) and diluted to 50 ng/µL.

PCR and Primers

All primers were designed using the program Oligo version 4.0 (Molecular Biology Insights, Cascade, CO, USA) and were purchased from Invitrogen (Carlsbad, CA, USA). For pyrosequencing, fragment 44 (216 bp), which contains SNPs 44/114 and 44/174, was amplified with the unlabeled primer 44-F and the 5′-biotinylated primer 44-R-biotin (Table 1). Fragment 66 (270 bp), which contains SNP 66/212, was pyrosequenced on both the sense and the antisense strands, and, accordingly, either the reverse or the forward amplimer was labeled with biotin. Unlabeled primers were used to amplify the same fragments in preparation for conventional and single-nucleotide extension sequencing, respectively. The amplifications were performed in a GeneAmp® PCR System 9700 Thermal Cycler (Applied Biosystems, Foster City, CA, USA). Each 50-µL PCR contained 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 10 pmol each primer, 0.2 mM each dNTP, 1.5 U AmpliTaq Gold® DNA polymerase (Applied Biosystems), and 50 ng ge-

nomic DNA. The thermal cycling protocol comprised an initial denaturation at 95°C for 5 min to activate AmpliTaq Gold, 50 cycles of 30 s at 94°C, 45 s at 56°C, and 1 min at 72°C, followed by a final 8-min extension at 72°C. PCR products were analyzed by 2% agarose gel electrophoresis.

Pyrosequencing

We used the multi-magnet PSQ™ 96 Sample Prep Tool (Pyrosequencing AB, Uppsala, Sweden) for transferring the immobilized samples during post-PCR sample preparation. Biotin-labeled PCR products were immobilized on streptavidin-coated paramagnetic beads (Dynabeads® M-280; Dynal AS, Oslo, Norway) by mixing 25 µL of the PCR with 12 µL Dynabeads (10 µg/µL) and 37 µL 2× binding-washing buffer (0.05% Tween® 20, 0.5 mM EDTA, 1 M NaCl, 5 mM Tris-HCl, pH 7.6). The samples were incubated at 65°C for 20 min under constant agitation using a thermal mixer (Thermomixer®; Eppendorf AG, Hamburg, Germany). The samples were then washed for 1 min in 50 µL 0.5 M NaOH. They were released and recaptured once in 100 µL 1× annealing buffer (20 mM Tris-acetate, 5 mM magnesium acetate, pH 7.6). Subsequently, the immobilized templates were transferred to 40 µL 1× annealing buffer plus 5 µL sequencing primer (20 pmol in water). For primer annealing, the samples were heated at 80°C for 2 min using the GeneAmp 9700 thermal cycler. After equilibration to room temperature, the sequencing reaction was performed.

Solid-phase pyrosequencing was carried out in the automated 96-well PSQ 96 system. Its major components are a microplate-rotating mechanism, a disposable inkjet cartridge for the precise delivery of small volumes (200 nL) of the reagents, a cooled charged-coupled device (CCD) camera, data acquisition modules, and a PC interface. Before sequencing, we loaded the cartridge with the reagents and each of the four nucleotides (SNP Reagent Kit; Pyrosequencing AB). Finally, the PSQ 96 Plate that contained the primed DNA templates was placed on the plate-rotating socket. During the run, nucleotides were added, following a se-

quence-specific dispension protocol suggested by the manufacturer and given in the figure legends. For some SNPs, we evaluated the influence of single-stranded DNA binding (SSB) protein on pyrosequencing by adding 0.5, 1.0, and 1.5 µg SSB (Amersham Biosciences, Piscataway, NJ, USA) to the primed target immediately before the pyrosequencing reaction.

In pyrosequencing, the intensity of a light signal is directly proportional to the number of nucleotides incorporated. This information is presented in pyrograms. We used a new program (Peak Height Determination Prototype SW; Pyrosequencing AB) to convert the peak heights of the two polymorphic bases into percent.

Single-Nucleotide Primer Extension and DHPLC

To remove unincorporated primers and dNTPs, 10 µL PCR products were incubated as described previously (13) with 1 U each shrimp alkaline phosphatase (Amersham Biosciences) and exonuclease I (Invitrogen) at 37°C for 30 min, after which the enzymes were deactivated at 80°C for 15 min. The extension reactions were carried out in a 20-µL volume containing 50 ng input PCR product, 50 µM of the appropriate ddNTPs, 20 pmol FAM-labeled extension primer, and 0.5 U Thermo Sequenase™ DNA polymerase (Amersham Biosciences) in the buffer provided by the manufacturer. The reaction was performed in a thermal cycler with an initial denaturation step of 2 min at 95°C, followed by 50 cycles at 95°C, 43°C, and 60°C for 8 s each. At the end of thermal cycling, the reaction was heated to 96°C for 30 s and placed immediately on ice. Extension products and unextended primer were resolved on a monolithic poly(styrene-divinylben-zene) capillary column (i.d., 50 × 0.2 mm) and detected by laser-induced fluorescence on a recently described HPLC instrument (32). It has been shown that it is necessary to calibrate each extension reaction using a known heterozygote with equal allele ratio to correct for the typical non-equal representation of alleles in single-nucleotide extension reactions. This representation is caused by the differential efficiency

with which polymerase incorporates different nucleotides (12). In the absence of a potato DNA sample known from sequencing to contain alleles at equal ratio, two single-stranded oligo-nucleotide templates that represent the two alleles of interest were synthesized and mixed at an equimolar ratio to determine the correction factor.

Dye-Terminator Cycle Sequencing

Unlabeled products were purified enzymatically as described earlier. Using the amplimers as sequencing primers, we sequenced both strands according to the BigDye® chemistry reaction protocol (Applied Biosystems). Each cycle-sequencing reaction contained 3 µL purified PCR product, 4 µL dye-terminator reaction mixture, 0.9 µL primer (10 µM), and 5.6 µL water. Cycle sequencing was started at 96°C for 10 min, followed by 30 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. We purified the cycle-sequencing reactions using Centrifex™ gel filtration cartridges (Edge BioSystems, Gaithersburg, MD, USA) and analyzed them on an ABI PRISM® 3700 sequencer (Applied Biosystems).

RESULTS AND DISCUSSION

The principle of pyrosequencing is based on the detection of *de novo* incorporation of nucleotides (22,25–27). The incorporation process releases pyrophosphate (PP_i), which starts a cascade of enzymatic reactions that leads to the generation of visible light proportionally to the number of incorporated nucleotides (Figure 1). In contrast with single-nucleotide extension sequencing, it is unnecessary for the primer to anneal immediately adjacent to the polymorphic nucleotide site. In the present study, the 3′-terminal base of the sequencing primers annealed next to (SNP-44/114 and SNP-66/212as) or three (SNP-66/212s) or four bases (SNP-44/174) away from the polymorphic site. However, this flexibility in primer design is offset partly by the requirement to avoid any hairpin or dimer formation of the primer because the pyrosequencing reaction is performed at 28°C due to the thermal

Research Report

instability of luciferase. The same consideration also applies to the template and may affect the ability to genotype a given polymorphic site. In our hands, pyrosequencing failed to generate interpretable results in 15 of 94 (16%) unselected polymorphic sites because of the aforementioned reasons, while the presence of paralogs caused two additional failures. Overall, 77 of 94 (82%) polymorphic sites could be genotyped successfully. This compares favorably to other large-scale genotyping studies of SNPs in human (6,31) and *Arabidopsis thaliana* (5) using either hybridization to a high-density oligonucleotide tiling array (5,31) or single-nucleotide extension in combination with hybridization to a generic tag array (6). In those studies, robust scoring was obtained for 60%–80% of the SNPs tested.

For all three SNPs shown in the present study, genotypes could be determined unequivocally in 99% of all reactions in the two parents and in the 52 individuals of the F1 generation. The reproducibility of pyrosequencing was evaluated by seven repeat measurements of all five possible genotypes of SNP-44/114 (Table 2). The percent distribution of the two alleles in case of the three heterozygous states could be reproduced with a mean relative standard deviation of 3.6%, with a range of 1.5%–9.1%. These values are similar to the mean standard deviation of ± 1.4 (90% confidence interval, 1.2–1.8) reported recently for 10 repeated measurements of five reference samples with allele ratios ranging from 0.18 to 9.0 by means of primer extension and DHPLC (9). The percent distribution of alleles, when determined in pools of 32, 64, 84, 96, and 112 chromosomes, respectively, differed from the distribution calculated on the basis of individual analyses by 0.1%–8.1% (data not shown). The discrepancy in allele frequency estimation between individual and pooled genotyping was, when compared with SNP-44/174, an average 2.7-fold larger for SNP-44/114. The most likely explanation is that the polymorphic site of SNP-44/114 (a C→T transition) is followed by a second C (Figure 2). Consequently, in addition to any C of the polymorphic site, four additional Cs will be incorporated, bring-

ing the totals to 4, 5, 6, 7, and 8 Cs, respectively, for the genotypes TTTT, TTTC, TTCC, TCCC, and CCCC. Table 2 shows that the standard deviations tend to increase the more the allele ratio deviates from 1. However, the accuracy is still sufficient to call all five possible allelic states.

Pyrosequencing was subsequently compared with single-nucleotide extension sequencing combined with DHPLC to separate the two fluorescent dye-labeled extension products from each other and from the unextended primer. DHPLC, typically performed at temperatures greater than or equal to 70°C, allows the resolution of isomeric extension products even when they differ only in a single base (12,23). Figure 2 shows that the incorporation of ddTTP causes a significantly greater increase in retention than that of ddCTP. In the case of a heterozygote that carries both alleles of a SNP with equal frequency, one would expect that the chromatographic peak heights or areas

of the two extension products would be equal. However, as can be seen in Figure 2, the height of the extension product of the T allele is (by a factor of 1.37) greater than that of the C allele. This is in accordance with a previous report on nine human SNPs (12). On no occasion was the expected allele ratio of 1 observed. Instead, allele ratios varied from 0.87 to 2.05; even identical SNP alleles embedded in different sequences yielded significantly different ratios. Fortunately, such differences are reproducible for a given SNP and, therefore, can be corrected. They can be caused by differences in efficiency with which the DNA polymerase incorporates ddNTPs (10) and/or differential PCR amplification of alleles (17). The latter could be excluded as a cause in the present study. For several SNPs, the deviation in allele ratio was identical whether a heterozygous PCR product or two synthetic oligonucleotides that represented the two alleles and mixed at an equimolar ratio were used as tem-
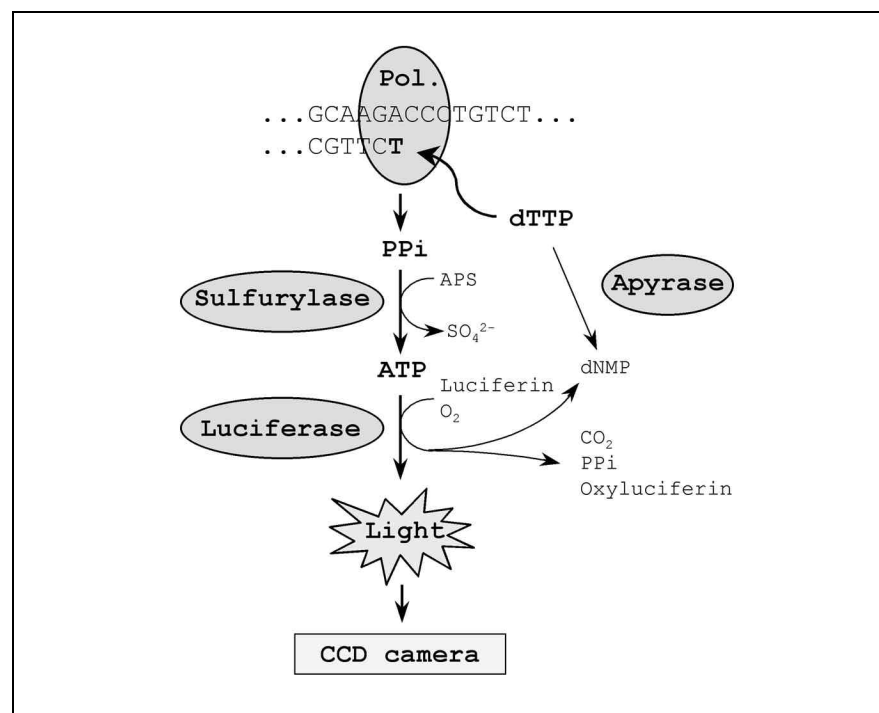

Figure 1. Scheme of the enzymatic reactions in liquid-phase pyrosequencing. The four different nucleotides are added stepwise to the primed DNA template. Following polymerase-mediated base incorporation, a proportional amount of PP$_i$ is released and converted to ATP by ATP-sulfurylase in the presence of adenosine 5′-phosphosulfate (APS). In turn, ATP is used in the luciferase reaction, during which a luciferin molecule is oxidized. The ensuing light, which is proportional to the number of nucleotides incorporated, is detected by a CCD camera. The iterative addition of nucleotides is possible as the excess of NTPs added to the reaction is continuously degraded between each cycle by apyrase into dNTPs and, subsequently, dNMPs.

plates in the single-nucleotide exten-
sion reaction. After correction for un-
equal incorporation of nucleotides, one
would expect that extension reactions

performed on 1:3 and 3:1 heterozy-
gotes would yield the expected allele
frequencies of 0.25 and 0.75. However,
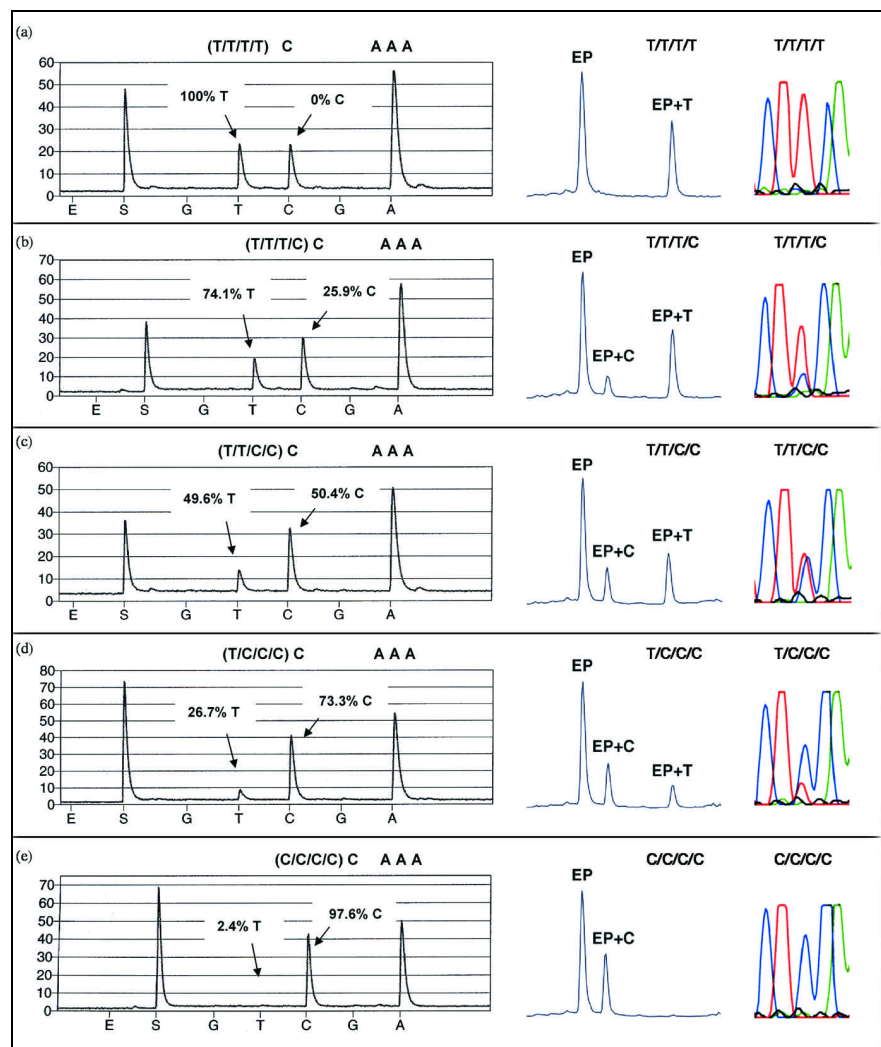the corrected allele frequencies for the

Figure 2. Comparison of the genotyping results obtained for the five possible different allele combinations of SNP-44/114 by pyrosequencing (left), single-nucleotide extension and DHPLC (center), and BigDye Terminator sequencing (right). Rows a and e show the two homozygotes, whereas rows b–d present the three different heterozygous allelic states. Pyrosequencing: the sequence to be analyzed was [TC]CAAA. Consequently, in a tetraploid genome, a total of 4–8 Cs will be incorporated, 0–4 Cs at the polymorphic site, depending on the number of C alleles present, and 4 Cs at the adjacent nucleotide site. The sequenced bases are shown above each pyrogram, and the allelic composition of the SNP is shown in brackets. The quantitative evaluation program calculated the percent distribution of the allelic bases T and C after the four Cs incorporated at the adjacent nucleotide site had been subtracted automatically. Arbitrary luminescence units are shown on the ordinate axis. According to the pre-specified dispensation scheme, nucleotides were added as indicated on the x-axis (GTCGA). The two Gs were included automatically in the dispensation to confirm the specificity of the sequencing reaction, with the first G corresponding to the 3′-terminal base of the sequencing primer. The background of the enzymatic reaction is assessed by the mere addition of enzyme cocktail (E) to the sample. The substrate (S) mixture containing APS and luciferin is added before the first addition of nucleotide to remove any contaminating PP$_i$ present in the sample or enzyme mixture. DHPLC of single-nucleotide extension reactions: column, monolithic poly(styrene-divinylbenzene) (i.d., 60×0.2 mm); mobile phase, 100 mM triethylammonium acetate, pH 7.0; gradient, 6.75%–8.00% acetonitrile in 5 min; flow rate, 2.4 µL/min; temperature, 72°C; detection of the FAM-labeled extension primer (EP) and its single-nucleotide extension products (EP+C and EP+T) was accomplished by argon ion laser-induced fluorescence detection, with emission monitored at 525 nm; injection volume, 1 µL.

Research Report

TTTC heterozygote were 68% T and 32% C, while they were 28% T and 72% C in the case of the TCCC heterozygote. This discrepancy was also observed in a larger analysis of 21 TTTC heterozygotes that yielded corrected frequencies of 70.3 ± 3.8% (mean ± SD) and 29.7 ± 3.8% for alleles T and C, respectively (Table 3). The corresponding values for six CCCT heterozygotes were 67.7 ± 1.1% and 32.3 ± 1.1% for alleles C and T, respectively. Interestingly, very similar estimates were obtained for the same samples using pyrosequencing (Table 3). Only dye-terminator sequencing yielded allele frequency estimates that were closer to the ones expected (Table 3). This was particularly true when the estimates were obtained from combined peak height measurements performed on both the forward and reverse strand sequencing trace. Accuracy decreased markedly when we used only the sequencing trace of the forward or reverse strand. Similar to single-nucleotide extension sequencing, it proved beneficial in conventional sequencing to correct the measured peak heights in the sequencing traces for the unequal incorporation of nucleotides.

The three methods were also employed for the genotyping of two pools of F1 plants that differed in phenotype and contained different ratios of the SNP-44/114 alleles. Based on individual genotyping results obtained by pyrosequencing and subsequently confirmed by single-nucleotide extension, the two pools of F1 plants were to contain the alleles C and T at ratios of 0.41:0.59 and 0.37:0.63, respectively. The three methods yielded very similar results that deviated in absolute values between 3.4% and 6.0% from the expected allele frequencies. The discrepancy between individual and pooled genotyping data was somewhat greater than the 0%–3.6% reported recently for human SNPs in pools of varying size (8,11). For the first pool (41.1% C and 58.9% T), the frequencies of alleles C and T, as determined by pyrosequencing, single-nucleotide extension, and conventional sequencing were 46.3%/53.7%, 44.5%/55.5%, and 45.3%/54.7%, respectively. This means that single-nucleotide extension came closest to the expected ratio, while pyrose-

quencing was off the most. In the case of the second pool (37.0% C and 63% T), pyrosequencing performed the best, while conventional sequencing performed the worst. The frequencies of alleles C and T, as determined by the three methods (same order as above), were 41.1%/58.9%, 42.3%/57.7%, and 43.0%/57.0%, respectively. Great care was taken in preparing the pools. Nevertheless, it cannot be excluded that at least part of the discrepancy between the expected and real allele frequencies were due to errors during the quantitation and mixing of the DNA samples. However, the data also seem to indicate that the different DNA polymerases used in the three assays (Klenow in pyrosequencing, Thermo Sequenase in single-nucleotide extension, and Ampli-Taq® Polymerase FS in conventional sequencing) exhibit very similar differential efficiencies of incorporation of the two alleles. The differential PCR amplification of the alleles as a source

of error is unlikely because we have observed similar deviations from the expected allele ratio using synthetic template for pyrosequencing and single-nucleotide extension (data not shown). Regarding the cost of genotyping 96 samples, single-nucleotide extension (approximately $180, using unlabeled extension primer) is approximately half as expensive as pyrosequencing (approximately $360), while conventional sequencing costs $800–$1500, depending on whether only one or both strands are being sequenced. In pyrosequencing, the biotinylated HPLC-purified primer that is used for immobilizing the PCR product to prepare a single-stranded template contributes about 60% to the total cost of the assay. Hence, cost per genotype drops with increasing the number of samples analyzed. However, the same argument also applies to single-nucleotide extension.

SNP-66/212 provided an interesting test case for evaluating the accuracy of

Table 1. Primers Used for Amplification of Target Sequence and for Pyrosequencing

Primer Name	Primer Sequence (5′→3′)
44-F	CTACATCACCTGCAGCACTCC
44-R(-biotin)	TTTCATGGAAGCTGGACGGCTAG
SNP-44/114	CCTAAACCCCTGAAATTCT
SNP-44/174	GAAAGAGCACTTGGATGC
66-F(-biotin)	CTTCATTCATACTATATGTTCAATGG
66-R(-biotin)	TTCAATCCGAGCTCTTGGAATTG
SNP-66/212-s	CCT TTG TGC TTC TAG
SNP-66/212-as	TACACCCACCCCCTTGTGAA
121-F	AATTGTACTAGGTAGTAATAGCGTG
121-R-biotin	ATCCTGACTCATACTCTCCTGC
SNP-121/289	CTG GTC TAC G[GC]T TTA ATA TAA

Table 2. Reproducibility of Pyrosequencing Was Determined by Seven Repeated Analyses of All Five Possible Genotypes of SNP 44-114

Genotype	T (%)			C (%)		
	Expected	Observed[a]	SD	Expected	Observed[a]	SD
TTTT	100	99.3	0.9	0	0.7	0.9
TTTC	75	72.6	2.5	25	27.4	2.5
TTCC	50	48.8	0.8	50	51.2	0.8
TCCC	25	25.2	1.1	75	74.8	1.1
CCCC	0	3.4	0.8	100	96.6	0.8

[a]Arithmetic mean of seven repeated pyrosequencing reactions

Research Report

Table 3. Allele Frequencies of the Three Possible Heterozygous States of SNP-44/114 as Determined by Pyrosequencing, Single-Nucleotide Extension, and Conventional Dideoxy Sequencing

Genotype (N)	TTTC (22)[a]		TTCC (24)[a]		TCCC (6)[a]	
Allele Frequency (%)	**T** Mean ± SD Range	**C** Mean ± SD Range	**T** Mean ± SD Range	**C** Mean ± SD Range	**T** Mean ± SD Range	**C** Mean ± SD Range
Pyrosequencing	71.1 ± 3.8 65.0–80.0	28.9 ± 3.7 20.0–35.0	48.4 ± 3.3 42.0–57.0	51.6 ± 3.3 43.0–58.0	26.8 ± 1.8 24.0–30.0	73.2 ± 1.8 70.0–76.0
Single-Nucleotide Extension ($k = 1.37$)	70.3 ± 3.8 63.2–76.1	29.7 ± 3.8 23.9–36.8	49.8 ± 3.9 41.0–57.0	50.2 ± 3.9 43.0–59.0	32.3 ± 1.1 31.1–34.1	67.7 ± 1.1 65.9–68.9
Sequencing ($k = 1.2$) Forward and Reverse	75.5 ± 2.9 71.6–78.5	24.5 ± 2.9 21.5–28.4	50.1 ± 1.8 48.0–52.1	49.9 ± 1.8 47.9–52.0	26.3 ± 4.0 22.4–30.8	73.7 ± 4.0 69.2–77.6
Forward Only	74.2 ± 2.5	25.8 ± 2.5	50.0 ± 2.0	50.0 ± 2.0	22.8 ± 0.5	77.2 ± 0.5
Reverse Only	76.8 ± 1.6	23.2 ± 1.6	50.0 ± 0.8	50.0 ± 0.8	29.7 ± 1.3	70.3 ± 1.3

For the latter two methods, a factor k was calculated on the basis of the ratio of peak heights obtained for a 2:2 heterozygous individual to correct for the unequal incorporation of alleles.

[a]For conventional dye-terminator sequencing, we performed only four sequencing reactions, two each on the forward and reverse strands of the same DNA sample. For pyrosequencing and single-nucleotide extension, reactions were carried out on the reverse strand only for the number of DNA samples given in parentheses after the genotype tested.

pyrosequening an SNP that contains an allele that is succeeded by two or more nucleotides of the same kind. Accurate genotyping of SNP-66/212 failed repeatedly when performed on the sense strand. Because the polymorphic T is part of a homooligomer of three Ts [GC(AT)TT], the peak height of T is the sum of the eight Ts at the two nonpolymorphic sites and the number of Ts in the SNP. Pyrosequencing only succeeded to determine the correct allele ratio as determined previously by conventional dideoxy sequencing in the cases of the AAAA and AAAT genotypes. In the case of the AATT genotype (Figure 3a), after the subtraction of the 8 Ts from the two nonpolymorphic sites, the software calculated a percent distribution of 87.7% A and 12.3% T, instead of the expected 50% A and 50% T. Since α-thio-ATP used in pyrosequencing yields generally a signal 20% higher than those of the other three nucleotides (8), the peak height for A should be reduced accordingly. However, even the corrected values (70.2% A and 29.8% T) suggest an AAAT rather than the actual AATT genotype. Further, it was impossible to distinguish the genotypes AAAT and AAAA (Figure 3, b and c). In both instances, the instrument's software cal-

culated allele frequencies from 100% A to 0% T. It is important to note that the failure of determining the correct allele frequencies is not caused by an inadequacy of the software supplied with the instrument. An analysis of the pyrograms by eye yielded the same results. It is rather a consequence of the limited dynamic range of pyrosequencing that does not extend over one order of magnitude for reasons that are not understood. Therefore, accurate genotyping by pyrosequencing is unfeasible when the ratio of two alleles exceeds 1:8. However, the five possible genotypes of SNP-66/212 could be genotyped unambiguously on the antisense strand. In this case, the sequencing primer could be designed so that the terminal base at the 3′ end annealed directly adjacent to the polymorphic site [(TA)GC].

Of 94 SNP loci tested, 18 yielded pyrograms of insufficient quality to assess the allele ratio with the standard protocol. It has been reported that the addition of 0.5 µg SSB to the primed template improves the read length in pyrosequencing (24). Therefore, we added increasing amounts of SSB (0.5, 1.0, and 1.5 µg, respectively) to all pyrosequencing reactions that had failed previously. Only in one case, SNP-121/289, did we observe a clear im-

provement in the quality of the readouts, with the degree of improvement depending on the amount of SSB added (Figure 4). It has been hypothesized (24) that SSB displaces apyrase from DNA, and, consequently, nucleotide degradation becomes more efficient (Figures 1 and 4). This is corroborated by the significant reduction in tailing of the peaks in the pyrograms.

In conclusion, all three methods evaluated in the present study have proven sufficiently quantitative to allow discrimination between the five possible allelic states of an SNP in a tetraploid genome. Although the unequal incorporation of nucleotides has been observed in pyrosequencing, its degree is far less than that of single-nucleotide extension, where incorporation efficiencies may differ by as much as a factor of two (12). Therefore, pyrosequencing allows the determination of the different allelic states with great confidence, even when a heterozygous control with an allele ratio of 1 is unavailable. Pyrosequencing is significantly more expensive than single-nucleotide extension sequencing in combination with DHPLC. This is mainly because of the need for a biotinylated PCR primer for subsequent preparation of single-stranded template for the pyrosequencing reaction. Its cost

Research Report

is comparable to that of nick translation PCR (18) with regard to both hardware and reagents, but significantly more sample manipulation is required. The low reaction temperature and the fact that allelic quantitation becomes inaccurate once the allele ratio exceeds 1:8 exert some constraints on the nature of sequences amenable to genotyping by pyrosequencing. However, overall, the rate of successful SNP scoring appears to be similar to those of other methods. It has also been found that cycle sequencing with dye terminators, particularly when it is performed on both the forward and reverse strands, is the most accurate method. Although its high cost excludes its use in individual genotyping of single SNPs, it may be the most appropriate method when pooled samples of both affected and unaffected individuals are used not only to discover

but also to estimate simultaneously the allele frequencies of binary SNPs (15). Finally, to date, single-nucleotide extension has been applied successfully to the individual genotyping of 7 of the 15

SNPs that had not been amenable to allelic discrimination by pyrosequencing. As mentioned earlier, the success or failure rate of pyrosequencing is similar to that of other widely used SNP genotyp-
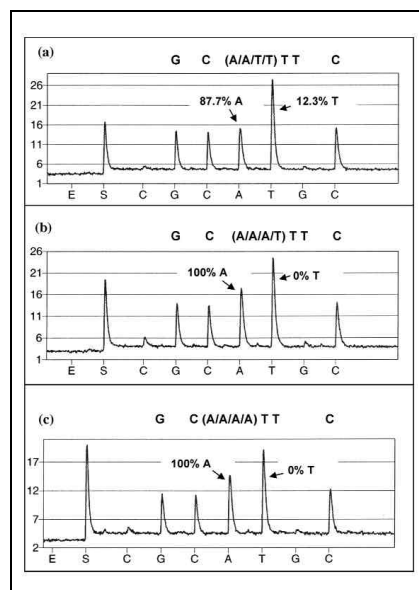

Figure 3. Pyrograms of three different allele combinations of SNP-66/212 genotyped with the primer SNP-66/212s. The sequence to be analyzed was GC[AT]TTC, with the polymorphic T allele being part of a homooligomer run of three Ts. The expected sequences are shown above each pyrogram, and the allelic composition of the SNP is shown in brackets. The percent distribution of the A and T alleles was calculated by the software supplied with the instrument. Arbitrary luminescence units are shown on the ordinate axis. According to the pre-specified dispensation scheme, nucleotides were added as indicated on the x-axis (CGCATGC). In a pyrogram (a), an allele distribution of 50% A and 50% T would have been expected. Instead, the observed ratio was from 87.7% A to 12.3% T. Data analysis failed to distinguish AAAT (b) from AAAA (c).
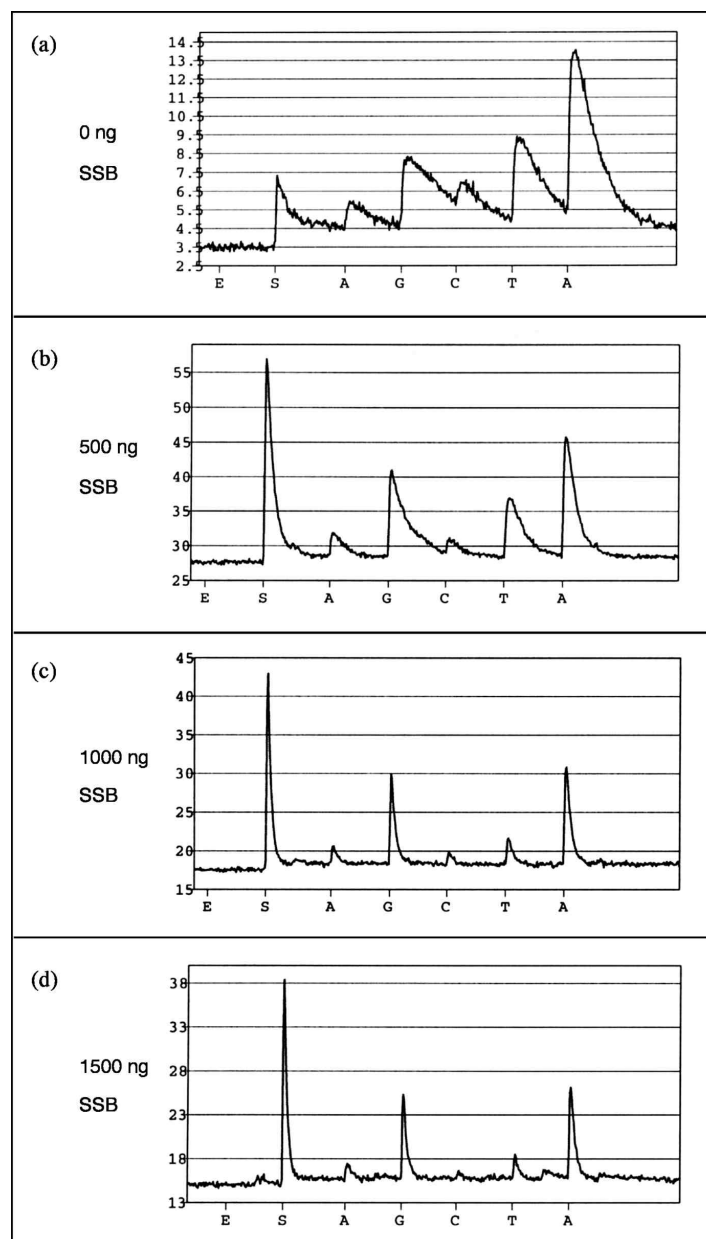

Figure 4. Effect of different amounts of SSB protein on pyrosequencing of SNP-121/289. The sequence to be analyzed was [GC]A. According to the pre-specified dispension scheme, nucleotides were added as indicated on the x-axis (AGCTA). The first A and T, the fourth nucleotide added, were included in the dispension to confirm the specificity of the pyrosequencing reaction. The pyrograms were obtained from four pyrosequencing reactions performed on the same homozyous GGGG potato with increasing amounts of SSB (0–1500 ng). The improvement in quality can be clearly seen. The pyrograms obtained in the absence of SSB (a) and in the presence of only 500 ng SSB (b) seem to indicate erroneously a heterozygous genotype of GGGC. Also notice the vanishing signals of the negative control nucleotides with the addition of increasing amounts of SSB.

ing techniques for which published data on large sets of SNPs are available. In practice, when an SNP is not amenable to genotyping with a particular method and cannot be replaced with another equally informative SNP, only a combination of technologies will ever succeed in scoring reliably all SNPs of interest. The choice of the primary method will vary from study to study and depends, among other things, on the number of SNPs and individuals to be genotyped, the ploidy of the genome, the availability of instrumentation, expertise, and appropriate controls, and the demand for individual versus pooled genotyping.

REFERENCES

1. **Ahmadian, A., B. Gharizadeh, A.C. Gustafsson, F. Sterky, P. Nyren, M. Uhlen, and J. Lundeberg.** 2000. Single-nucleotide polymorphism analysis by pyrosequencing. Anal. Biochem. *280*:103-110.
2. **Alderborn, A., A. Kristofferson, and U. Hammerling.** 2000. Determination of single nucleotide polymorphisms by real-time pyrophosphate DNA sequencing. Genome Res. *10*:1249-1258.
3. **Bottema, C.D.K., G. Sarkar, J.D. Cassay, S. Ii, C.M. Dutton, and S.S. Sommer.** 1993. PCR-amplification of specific alleles: a general method of rapidly detecting mutations, polymorphisms, and haplotypes. Methods Enzymol. *218*:388-402.
4. **Chee, M., R. Yang, E. Hubbell, A. Berno, X.C. Huang, D. Stern, J. Winkler, D.J. Lockhart et al.** 1996. Accessing genetic information with high-density DNA arrays. Science *274*:610-614.
5. **Cho, R.J., M. Mindrinos, D.R. Richards, R.J. Sapolsky, M. Anderson, E. Drenkard, J. Dewdney, T.L. Reuber et al.** 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. Nat. Genet. *23*:203-207.
6. **Fan, J.B., X. Chen, M.K. Halushka, A. Berno, X. Huang, T. Ryder, R.J. Lipshutz, D.J. Lockhart et al.** 2000. Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. Genome Res. *10*:853-860.
7. **Germer, S. and R. Higuchi.** 1999. Single-tube genotyping without oligonucleotide probes. Genome Res. *9*:72-78.
8. **Gharizadeh, B., T. Nordström, A. Ahmadian, M. Ronaghi, and P. Nyrén.** 2002. Long-read pyrosequencing using pure 2′-deoxyadenosine-5′-O′-(1-thiotriphosphate)Sp-isomer. Anal. Biochem. *301*:82-90.
9. **Giordano, M., M. Mellai, B. Hoogendoorn, and P. Momigliano-Richiardi.** 2001. Determination of SNP allele frequencies in pooled DNAs by primer extension genotyping and denaturing high-performance liquid chromatography. J. Biochem. Biophys. Methods *11*:101-110.
10. **Haff, L.A. and I.P. Smirnov.** 1997. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Res. *7*:378-388.
11. **Hall, J.G., P.S. Eis, S.M. Law, L.P. Reynaldo, J.R. Prudent, D.J. Marshall, H.T. Allawi, A.L. Mast et al.** 2000. Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. Proc. Natl. Acad. Sci. USA *97*:8272-8277.
12. **Hoogendoorn, B., N. Norton, G. Kirov, N. Williams, M.L. Hamshere, G. Spurlock, J. Austin, M.K. Stephens et al.** 2000. Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools. Hum. Genet. *107*:488-493.
13. **Hoogendoorn, B., M.J. Owen, P.J. Oefner, N. Williams, J. Austin, and M.C. O'Donovan.** 1999. Genotyping single nucleotide polymorphisms by primer extension and high performance liquid chromatography. Hum. Genet. *104*:89-93.
14. **Jalanko, A., J. Kere, E. Savilahti, M. Schwartz, A. Syvanen, M. Ranki, and H. Soderlund.** 1992. Screening for defined cystic fibrosis mutations by solid-phase minisequencing. Clin. Chem. *38*:39-43.
15. **Kwok, P.Y., C. Carlson, T.D. Yager, W. Ankener, and D.A. Nickerson.** 1994. Comparative analysis of human DNA variations by fluorescence-based sequencing of PCR products. Genomics *23*:138-144.
16. **Landegren, U., R. Kaiser, J. Sanders, and L. Hood.** 1988. A ligase-mediated gene detection technique. Science *241*:1077-1080.
17. **Liu, Q., E.C. Thorland, and S.S. Sommer.** 1997. Inhibition of PCR amplification by a point mutation downstream of a primer. BioTechniques *22*:292-296.
18. **Livak, K.J., S.J. Flood, J. Marmaro, W. Giusti, and K. Deetz.** 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridisation. PCR Methods Appl. *4*:357-362.
19. **Newton, C.R., A. Graham, L.E. Heptinstall, S.J. Powell, C. Summers, N. Kalsheker, J.C. Smith, and A.F. Markham.** 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. *17*:2503-2516.
20. **Nickerson, D.A., R. Kaiser, S. Lappin, J. Stewart, L. Hood, and U. Landegren.** 1990. Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. Proc. Natl. Acad. Sci. USA *87*:8923-8927.
21. **Nordstrom, T., M. Ronaghi, L. Forsberg, U. de Faire, R. Morgenstern, and P. Nyren.** 2000. Direct analysis of single-nucleotide polymorphism on double-stranded DNA by pyrosequencing. Biotechnol. Appl. Biochem. *31*:107-112.
22. **Nyrén, P. and A. Lundin.** 1985. Enzymatic method for continuous monitoring of inorganic pyrophosphatase synthesis. Anal. Biochem. *151*:504-509.
23. **Oefner, P.J.** 2000. Allelic discrimination by denaturing high-performance liquid chromatography. J. Chromatogr. B *739*:345-355.
24. **Ronaghi, M.** 2000. Improved performance of pyrosequencing using single-stranded DNA-binding protein. Anal. Biochem. *286*:282-288.
25. **Ronaghi, M.** 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res. *11*:3-11.
26. **Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlén, and P. Nyrén.** 1996. Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. *242*:84-89.
27. **Ronaghi, M., M. Uhlén, and P. Nyrén.** 1998. Real-time pyrophosphate detection for DNA sequencing. Science *281*:363-365.
28. **Saiki, R.K., C.A. Chang, C.H. Levenson, T.C. Warren, C.D. Boehm, H.H. Kazazian, and H.A. Erlich.** 1988. Diagnosis of sickle cell anemia and β-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. N. Engl. J. Med. *319*:537-541.
29. **Shi, M.M.** 2001. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. Clin. Chem. *47*:164-172.
30. **Tyagi, S., D.P. Bratu, and F.R. Kramer.** 1998. Multicolor molecular beacons for allele discrimination. Nat. Biotechnol. *16*:49-53.
31. **Wang, D.G., J.B. Fan, C.J. Siao, A. Berno, P. Young, R. Sapolsky, G. Ghandour, N. Perkins et al.** 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science *280*:1077-1082.
32. **Xiao, W., D. Stern, M. Jain, C.G. Huber, and P.J. Oefner.** 2001. Multiplex capillary denaturing high-performance liquid chromatography with laser-induced fluorescence detection. BioTechniques *30*:1332-1338.

Address correspondence to

Dr. Peter J. Oefner
Stanford Genome Technology Center
855 California Avenue
Palo Alto, CA 94304, USA
e-mail: oefner@genome.stanford.edu