



Polymorphism Ratio Sequencing: A New Approach for Single Nucleotide Polymorphism Discovery and Genotyping

Robert G. Blazej, Brian M. Paegel and Richard A. Mathies

Genome Res. 2003 13: 287-293

Access the most recent version at doi:[10.1101/gr.396203](https://doi.org/10.1101/gr.396203)

References This article cites 31 articles, 9 of which can be accessed free at:
<http://genome.cshlp.org/content/13/2/287.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, and the Cellecta logo, which is a green molecular structure with the word "CELLECTA" below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Methods

Polymorphism Ratio Sequencing: A New Approach for Single Nucleotide Polymorphism Discovery and Genotyping

Robert G. Blazej,¹ Brian M. Paegel,² and Richard A. Mathies^{1,2,3}

¹University of California, Berkeley/University of California, San Francisco Joint Bioengineering Graduate Group, Berkeley, California 94720, USA; ²Department of Chemistry, University of California, Berkeley, California 94720, USA

Polymorphism ratio sequencing (PRS) combines the advantages of high-throughput DNA sequencing with new labeling and pooling schemes to produce a powerful assay for sensitive single nucleotide polymorphism (SNP) discovery, rapid genotyping, and accurate, multiplexed allele frequency determination. In the PRS method, dideoxy-terminator extension ladders generated from a sample and reference template are labeled with different energy-transfer fluorescent dyes and coinjected into a separation capillary for comparison of relative signal intensities. We demonstrate the PRS method by screening two human mitochondrial genomes for sequence variations using a microfabricated capillary array electrophoresis device. A titration of multiplexed DNA samples places the limit of minor allele frequency detection at 5%. PRS is a sensitive and robust polymorphism detection method for the analysis of individual or multiplexed samples that is compatible with any four-color fluorescence DNA sequencer.

Genetic studies are expanding past the identification of rare Mendelian diseases to encompass common complex disorders such as cancer, neurological diseases, autoimmune diseases, and cardiovascular diseases. The abundance and stability of single nucleotide polymorphisms (SNPs) has made them the DNA variation of choice for such endeavors. Due to the low penetrance of individual alleles, association studies designed to uncover significant genetic risk factors require the scoring of hundreds of thousands of SNPs (Kruglyak 1999) in several thousand samples (Dahlman et al. 2002). Multiplexed high-throughput genotyping methods are essential to accommodate this challenging number of SNPs and samples.

Current methods of SNP detection include microarrays (Chee et al. 1996), allele-specific polymerase chain reaction (PCR; Liu et al. 1997), oligonucleotide ligation assays (Grossman et al. 1994; Samiotaki et al. 1994; Tobe et al. 1996), single-base extension (Syvanen et al. 1993; Pastinen et al. 1996; Shumaker et al. 1996), TaqMan (Livak et al. 1995), and molecular beacons (Tyagi et al. 1998). These methods are designed to interrogate SNPs in one sample at a time. Modifications such as kinetic PCR (Germer et al. 2000) and detection using denaturing high-performance liquid chromatography (HPLC; Giordano et al. 2001) can extend these procedures to multiplexed sample analysis. Each method offers a trade-off between sensitivity, throughput, parallel processing, user accessibility, and cost, and they all experience limitations due to local DNA hybridization characteristics. Selection of allele-specific probes is confined to the DNA sequence immediately adjacent to or surrounding the targeted polymorphism. This restriction can lead to nonuniform hybridization characteristics, ambiguous results, or the rejection of otherwise informa-

tive SNPs in a significant percentage of the markers. Additionally, novel polymorphisms are not detected by these approaches. These limitations are particularly important when studying functional or rare SNPs.

Quality-based resequencing approaches do not experience the limitations described above and are considered the gold standard for SNP scoring (Rieder et al. 1998). Despite this advantage and the availability of automated DNA sequencing instruments, resequencing has not proven to be efficient or cost-effective for large SNP studies. This failing arises because of the inability to multiplex conventional sequencing reactions and because the sequencing assay was not specifically designed for polymorphism detection.

In this report we describe polymorphism ratio sequencing (PRS), a method that combines the sequence independence and assay reliability of DNA sequencing with the sensitivity necessary to perform parallel, multiplexed genotyping. PRS is a dideoxy terminator-based method that employs standard sequencing chemistry in a novel labeling and pooling scheme to compare sample and reference DNA templates directly in a single lane. The result is an internally controlled method for polymorphism discovery, multiplexed allele frequency determination, and genotyping. The compact (16,569 bp) and well-characterized human mitochondrial genome was selected as a test system for PRS development. Two mitochondrial genomes from unrelated Centre d'Etude Polymorphisme Humaine (CEPH) individuals were simultaneously scanned for sequence variations using PRS and analyzed in <30 min using a microfabricated capillary array electrophoresis device (μ CAE; Paegel et al. 2002). Because standard sequencing chemistry is used, PRS is also compatible with conventional four-color fluorescence DNA sequencing instruments. Our analysis identified 30 previously characterized and six novel polymorphisms. One polymorphism was chosen to demonstrate the sensitivity of the method in determining the minor allele frequency in a multiplexed sample.

³Corresponding author.

E-MAIL rich@zinc.cchem.berkeley.edu; **FAX** (510) 642-3599.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.396203>.

RESULTS

Coding and Pooling Scheme

Figure 1 presents the PRS base coding scheme for A/C and G/T companion traces. Unlike conventional DNA sequencing, PRS simultaneously compares *two* DNA templates, sample and reference, using different fluorescent dyes to label identical base termination ladders. Each trace is composed of four electropherograms that concurrently interrogate two bases: A and C in Figure 1A and G and T in Figure 1B. Because each trace compares only two sets of bases at a time, gaps between peaks occur in regions of nonrepresented bases. Two complementary PRS reactions yield complete sequence coverage. Sample and reference templates produce coincident peaks at each base position when the sequences are identical and noncoincident peaks when there are polymorphisms.

Polymorphism Identification

Processing of PRS data is considerably simplified compared to conventional sequence data analysis because base calling is not necessary to identify polymorphisms. After color matrixing and normalization, sample and reference data are aligned to each other such that all nonvariant peaks overlap. A variance plot is then generated from the aligned data by squaring the signal intensity difference between the sample and reference electropherograms. This Δ^2 plot produces excellent signal-to-noise ratios and is used for detection of polymorphic bases. Polymorphisms were identified at peak positions where the Δ^2 plot exceeded three standard deviations (SDs) from the mean. Figure 2 illustrates how polymorphisms are identified along with examples of the detected variations. At a polymorphic site, the sample and reference traces are not coincident. This deviation manifests itself in different ways depending on the type of polymorphism. A transition polymorphism (e.g.,

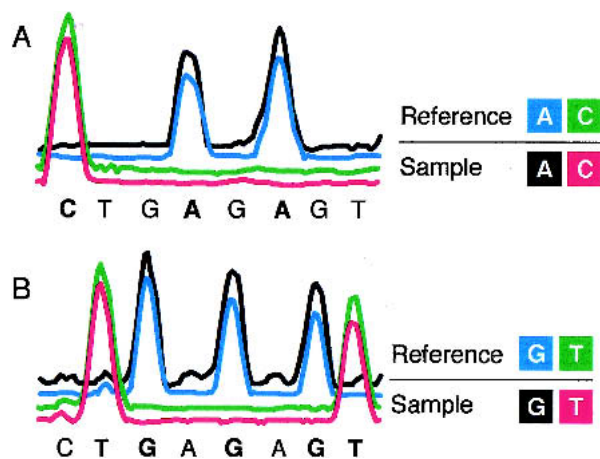


Figure 1 Polymorphism ratio sequencing labeling and reaction tube pooling scheme. *Insets* present the four-color labeling scheme for bases A/C and G/T in sample and reference sequences. (A) A and C extension products from the sample template are respectively labeled with R110 (λ_{\max} 525 nm) and ET-TAMRA (λ_{\max} 580 nm). Similarly, A and C extension products from the reference template are labeled with ET-R6G (λ_{\max} 555 nm) and ET-ROX (λ_{\max} 605 nm). Products from the four reactions are combined and analyzed in a single lane. (B) G and T extension products from sample template are labeled with R110 and ET-TAMRA, and reference template G and T extension products are labeled with ET-R6G and ET-ROX, respectively. The baseline signal intensity of the four channels has been offset to show coincident peaks more clearly.

G→A) is exemplified by the loss of the sample G peak in the G/T trace and the appearance of a new A peak in the A/C trace (Fig. 2A). Whereas a transition mutation is detected as a one-base change in two PRS traces, a transversion mutation (e.g., C→A) is detected as two base changes within one PRS trace (Fig. 2B). Although the C and A peaks overlap, they are mismatched relative to the defined coding scheme and therefore produce conspicuous squared signal differences. Due to the dependence of base incorporation frequency on the preceding base, a C→A base change creates a “footprint” on the following base, seen as different sample/reference peak heights for the following C base. Incorporation frequency parity is restored for subsequent bases. Insertion-deletion polymorphisms result in a sample/reference frame shift after the variant site (Fig. 2C). Realignment of the traces is achieved by gapping and displacing the sample or reference traces until the Δ^2 plot is minimized. The loss of single-base resolution for higher-molecular-weight fragments and for compressed regions has only a modest effect on the relative sample/reference signal. Figure 2D demonstrates that positive variant identification, which could not be accurately scored by conventional sequencing, was possible in what is typically considered very low-quality sequence.

Genomic PRS data from two individuals aligned to a map of the human mitochondrial genome are presented in Figure 3. The genome encodes 13 proteins, 22 tRNAs, the 12S and 16S rRNA, and contains 1025 characterized genetic variations, of which over 50 have known pathogenesis. One individual was arbitrarily chosen as the reference and the other as the sample. Sequence variations between the two individuals produce definitive signals in the squared difference plot. Variants were localized onto the Anderson human mitochondrial sequence (Anderson et al. 1981) and compared to known polymorphisms in the MITOMAP database (<http://www.mitomap.org>; Kogelnik et al. 1996). Polymorphic sites were reconfirmed using conventional fluorescence-based DNA sequencing methods (Ju et al. 1995a). The outermost black circle in Figure 3 presents a binary representation of the squared difference in signal for all four bases. Thirty known and six novel polymorphisms (*n3720*, *n3849*, *n4553*, *n4736*, *n5426*, and *n16291*) were identified. Table 1 lists the identified polymorphisms and corresponding amino acid changes. Mutation *n13708* is a secondary Leber’s hereditary optic neuropathy risk factor (Brown et al. 1992).

Multiplexed Sample Analysis

Figure 4A presents two-color PRS trace data from a separate experiment in which a reference template is compared to multiplexed sample templates containing varying ratios of a C→T transition. The reference template contains DNA from a single individual and therefore produces identical peaks in all traces. The polymorphic base in the multiplexed templates produces peak heights consistent with the template ratios. The peak height ratio at the polymorphic base is used to determine the allele frequency. Figure 4B presents the calibration curve generated from eight multiplexed samples ranging in concentration from 0%–100%. For each of the eight samples, four separate PCR and PRS reactions were performed and analyzed in independent lanes. The results are linear through all concentrations ($R^2 = 0.9991$). The minor allele frequency limit of detection achievable by PRS is set by the lowest concentration in this study (5%) in which the signal-to-noise ratio is 7.0.

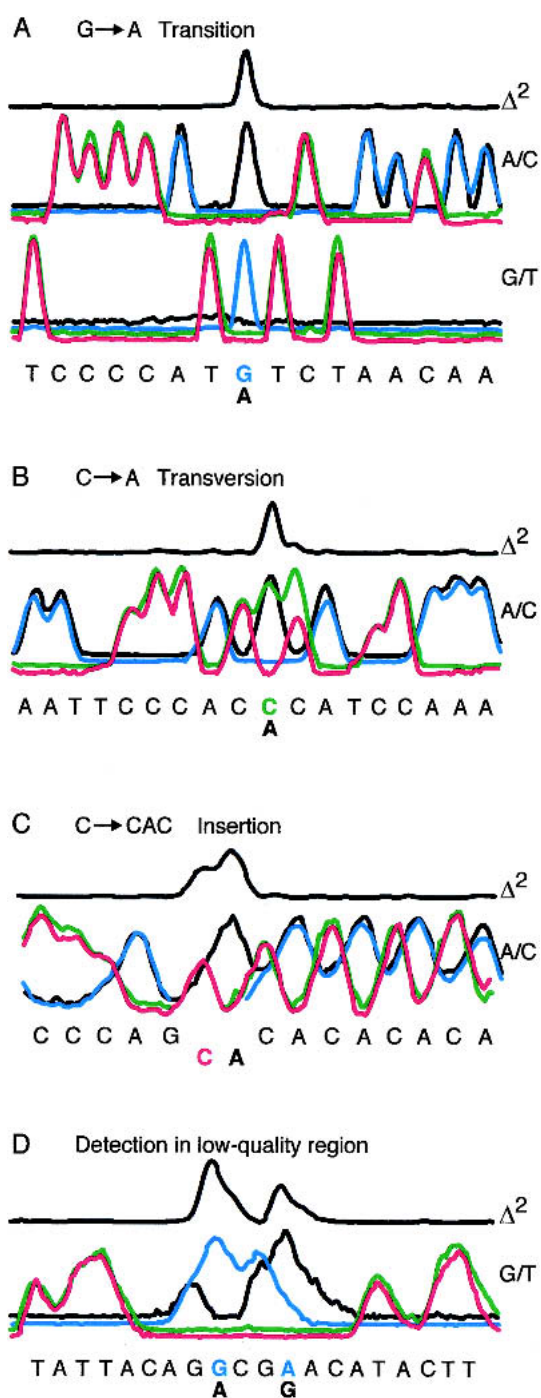


Figure 2 Representative traces of various polymorphisms. The Δ^2 plot is generated by squaring the sum of the differences between the sample and reference traces for all four comparisons. (A) Two companion traces for a G→A transition polymorphism. Transition events detected in one trace are confirmed in the companion trace. (B) A/C trace for an C→A transversion displays mismatched coincident peaks. (C) Example of a C→CAC insertion. Insertion results in a sample/reference frame shift, which is corrected by realigning the traces after the insertion. (D) Positive polymorphism identification in a “low-quality” sequence region. Single-base resolution is not required for accurate base variation detection.

DISCUSSION

The results presented here demonstrate that PRS is an accurate, high-throughput approach for SNP discovery and genotyping of individual and multiplexed samples that is compatible with all four-color fluorescence DNA sequencing instruments. This method derives its unique capabilities from the *simultaneous* comparison of sample and reference template in the *same* lane, producing trace data that are internally controlled at every base position and resistant to the lane-to-lane injection, mobility, and detection artifacts inherent to any DNA sequencing method. Although peak height and base spacing are not uniform throughout a sequencing run, the relative signal intensity between the sample and reference electropherograms is remarkably consistent, even in the low-quality region seen in Figure 2D. As a result, PRS data quality is not impaired by conventional DNA sequencing anomalies resulting from local sequence content (e.g., GC compressions). Because terminator incorporation rates are consistent for a given sequence (Kwok et al. 1994), PRS signal intensities from nonvariant sample and reference templates align nearly perfectly, permitting direct polymorphism detection and accurate quantitation of multiplexed DNA samples. Although sequence artifacts and alignment may also be corrected in conventional four-color sequencing through postrun software, it is our experience that performing such corrections on separate traces is problematic because the additional data manipulations reduce quantitation and accuracy.

PRS provides facile and unambiguous polymorphism detection in low-quality as well as high-quality sequence regions. There is no need to align PRS data to a consensus sequence and compare the sometimes ambiguous base calls, as in conventional sequence-based analysis. Although probabilistic base call quality scores are useful in resolving such ambiguities (Rieder et al. 1998), polymorphisms in low-quality conventional sequence data are indistinguishable from sequencing errors (Marth et al. 1999). PRS relies on relative and not absolute sequence quality, thereby enabling positive polymorphism detection in regions well below single-base resolution. Additionally, relative sequence quality is enhanced in the PRS method because gaps produced by comparing only two bases in each trace result in fewer overlapping peaks. Every variant base also generates two distinct signals—the loss of one peak and the appearance of another, as seen in Figure 2A. This complementarity provides internal, robust error checking not possible with single-read approaches.

Our analysis of the human mitochondrial genome demonstrates the feasibility of using PRS for high-throughput genotyping, SNP discovery, and diagnostics. In this study, 30 previously characterized and six novel polymorphisms were identified, and all classes of simple genetic variation (i.e., transition, transversion, and insertion-deletion) were detected. The PRS method has important advantages over hybridization-based approaches that require the design of probes in the immediate vicinity of the variant base. This requirement precludes the use of probes in regions where the local sequence content does not permit efficient or consistent hybridization. The PRS method provides the flexibility to design optimized probes up to several hundred bases away from the target SNP. As a result, all SNPs including important rare or functional polymorphisms can be scored. Additionally, sequences surrounding the targeted SNP are also scanned, providing complete and unbiased detection of novel or unanticipated variant sites. Although each polymorphism was visually

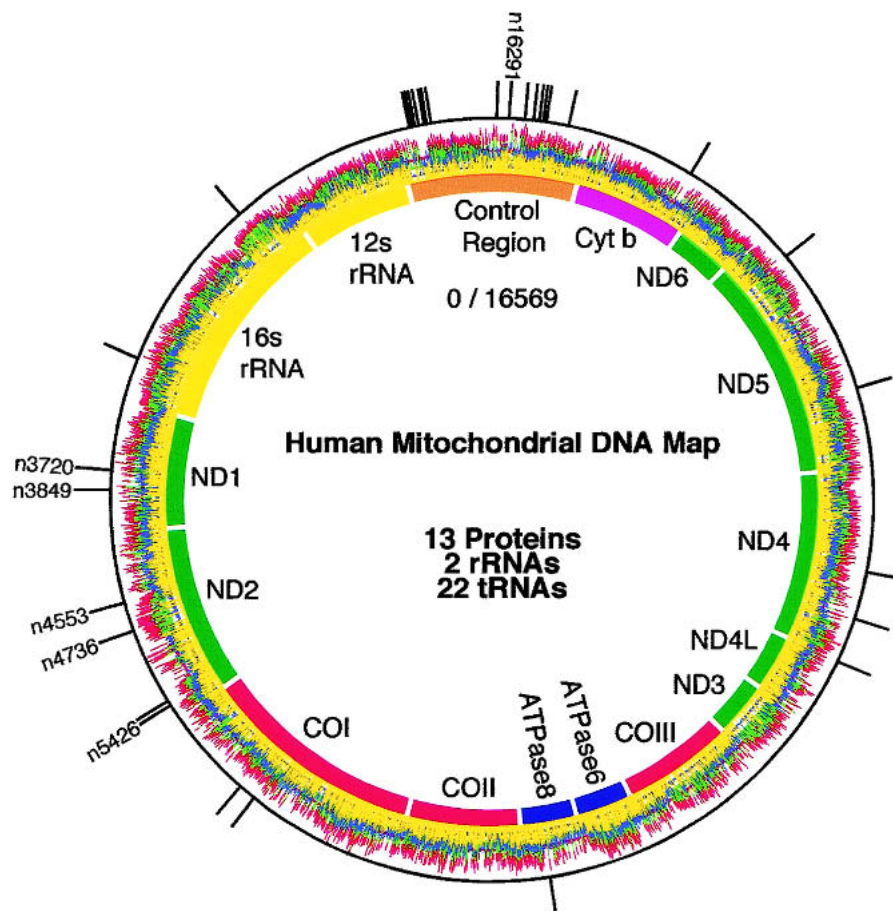


Figure 3 Whole mitochondrial genome analysis. Genomic PRS data is presented wrapped around a map of the human mitochondrial genome. For clarity, traces are offset and black has been replaced with yellow. The outer circle presents a binary representation of the squared cumulative difference (Δ^2) plot. Novel polymorphisms are labeled with the mtDNA nucleotide position. Data were collected in <30 min. using a microfabricated capillary array electrophoresis sequencing device.

inspected in this study, the high signal-to-noise ratio intrinsic to the squared difference (Δ^2) plot and two-signal error checking should facilitate the development of software for automated detection.

The PRS method is particularly well suited for clinical genetic diagnostics, in which sensitivity and a low error rate are critical. For example, unambiguous detection of mtDNA mutation *n13708*, a secondary Leber's hereditary optic neuropathy mutation, was achieved. Detection of genetic mutations in different tissue types within an individual is also possible. Indicative and possibly causative mitochondrial genetic variations are frequently observed between healthy and cancerous tissues (Polyak et al. 1998; Fliss et al. 2000). Application of the PRS method to detect such variations requires only the use of reference DNA from patient blood and sample DNA from tumor tissue. The sensitivity of the PRS method to as little as 5% genetic variations addresses important problems associated with tumor sample heterogeneity. Furthermore, the ability to scan for known and unknown variations is important, because many of the mutations seen in cancerous tissue are novel (Fliss et al. 2000).

Case-control association studies employing genome-wide screening require the interrogation of many thousands

of SNPs (Risch and Merikangas 1996; Merikangas 1999). The low phenotypic effect of any single gene in complex disease traits necessitates the use of large numbers of samples to obtain significant association. The examination of multiplexed samples using PRS will significantly reduce sample preparation and analysis to a tractable level in such studies (Barcellos et al. 1997). The PRS method provides sufficient quantitative sensitivity to multiplex 20 haploid (e.g., mitochondrial) or 10 diploid samples and to detect as little as 5% allele frequency in population pools. The standardized sequencing chemistries and electrophoretic separation technologies used in PRS will also facilitate parallel polymorphism screening and allele frequency determination. Such analysis should prove useful for SNP validation, rare allele identification, and genetic association studies involving affected and unaffected sample pools.

The PRS coding scheme presented here requires more tube manipulations than conventional dye-terminator sequencing and is not beneficial for all sequence comparisons. Bulk preparation of reference reactions, sample multiplexing, and alterations to the coding/pooling scheme⁴ all provide avenues to simplify the assay. Nonetheless, there are a number of situations where the benefits of PRS analysis justify the additional overhead. In critical sequence analysis

such as diagnostic sequencing and tumor/normal comparisons, error rates must be low. To achieve high-quality sequence information over large regions using conventional sequencing methods, multiple, overlapping sense and antisense traces must be obtained and aligned. By comparison, the PRS method has the potential to provide, in a single run, high-quality sequence comparison over an extended region of the trace. More importantly, in situations such as association studies where allele frequency is the desired information, one can take advantage of the remarkable sensitivity of the PRS method to multiplex many samples. This quantitative, multiplexed capability is not provided by any existing resequencing method. The present study highlights the potential of alternative base labeling and pooling schemes to enhance

⁴Straightforward modifications to the PRS coding/pooling scheme such that the forward and reverse extension reactions are performed simultaneously would reduce the number of reaction tubes by a factor of two. Another factor of two reduction is provided by the use of uniquely-labeled energy transfer (ET)-cassette primers (Berti et al. 2001). For example, by pooling sample and reference templates amplified with uniquely-tailed PCR primers and performing PRS analysis using corresponding cassette-labeled primers, it is feasible to obtain forward and reverse PRS data on both templates with only four reaction tubes—the same number of tubes as dye-terminator sequencing.

Table 1. Mitochondrial DNA Variation Detection

Position ^a	Base change ^b	Gene ^c	Amino acid change ^d	Position ^a	Base change ^b	Gene ^c	Amino acid change ^d
185	G → A	CTRL	—	8473	T → C	ATP8	Syn
188	A → G	CTRL	—	10876	A → G	ND4	Syn
217	T → C	CTRL	—	11251	A → G	ND4	Syn
228	G → A	CTRL	—	11467	A → G	ND4	Syn
462	C → T	CTRL	—	12612	A → G	ND5	Syn
489	T → C	CTRL	—	13708	G → A	ND5	Ala → Thr
508	A → G	CTRL	—	14798	T → C	CTYB	Phe → Leu
514	CAC → C	CTRL	—	15907	A → G	CTRL	—
1811	A → G	16S	—	16051	A → G	CTRL	—
3010	G → A	16S	—	16069	C → T	CTRL	—
3720	A → G*	ND1	Syn	16092	T → C	CTRL	—
3849	G → A*	ND1	Syn	16126	T → C	CTRL	—
4553	T → C*	ND2	Syn	16129	G → C	CTRL	—
4736	T → C*	ND2	Syn	16183	A → C	CTRL	—
5390	A → G	ND2	Syn	16189	T → C	CTRL	—
5426	T → C*	ND2	Syn	16291	C → A*	CTRL	—
6045	C → T	COI	Syn	16362	T → C	CTRL	—
6152	T → C	COI	Syn	16366	C → T	CTRL	—

^aPosition in the Anderson reference sequence (Anderson et al. 1981).

^bArrow direction indicates base change from the Anderson reference sequence; asterisk indicates novel polymorphism.

^cGene name abbreviations: CTRL, control region; 16S, 16S rRNA; ND, NADH dehydrogenase; ATP, ATP synthase; CO, cytochrome oxidase; Cytb, cytochrome b.

^dAbbreviations: —, non-coding; Syn, synonymous mutation.

conventional sequencing. The ultimate utility of PRS will depend on a balance between the benefits afforded by the method and any additional reaction setup complexity.

METHODS

Template Amplification

Twenty-four pairs of overlapping PCR primers (Operon Technologies and BioSource International) were used to amplify mitochondrial DNA (mtDNA; Rieder et al. 1998). Universal dye-primer sequences, –40 M13 forward (GTT TTC CCA GTC ACG ACG) and –28 M13 reverse (AGG AAA CAG CTA TGA CCA TG), were added to the 5'-end of the forward and reverse primers, respectively. Mitochondrial DNA from two unrelated CEPH individuals (reference NA 13116, sample NA 10859) was amplified in 48 25- μ L PCR reactions [10 ng genomic DNA, 0.2 mM dNTPs, 1.5 mM MgCl₂, 0.3 μ M primers, 20 mM Tris-HCl pH 8.4, 50 mM KCl, and 1.25 U Platinum Taq DNA polymerase (Invitrogen)]. PCR reactions were incubated at 94° C for 1 min to denature the template and activate the polymerase, followed by 30 cycles at 94° C for 30 sec, annealing at 61° C for 45 sec, and extension at 72° C for 1.5 min with a final incubation at 72° C for 3 min using an Eppendorf Mastercycler (Brinkmann Instruments).

Polymorphism Ratio Sequencing

After thermal cycling, unincorporated primers and dNTPs were degraded by the addition of 30 U of exonuclease I (10 U/ μ L) and 2 U of shrimp alkaline phosphatase (1 U/ μ L, from USB). Reactions were incubated at 37° C for 30 min followed by a 15-min inactivation at 80° C. Treated PCR reactions were then diluted with water to a working volume of 90 μ L (~12 ng/ μ L). Dye-primer PRS was performed in separate tubes for each PCR amplicon as follows: 2 μ L C, A, G, or T Thermo Sequenase nucleotide mix (US 79525, Amersham Biosciences), 2 μ L 0.4 mM energy transfer (ET) primer (fluorescence emission maxima of 525, 555, 580, or 605 nm; Ju et al. 1995b), 4 μ L diluted PCR reaction, 2 μ L water. Reactions were cycled 25 times at 95° C for 30 sec, 55° C for 15 sec, and 72°

C for 1 min followed by pooling and precipitation (3 μ L 7.5 M ammonium acetate, 110 μ L 100% ethanol). Precipitated fragments were collected after 15 min by centrifugation at 16,000 \times g for 20 min followed by a 400 μ L wash with 70% ethanol. Samples were air-dried for 15 min and resuspended in 15 μ L 50% formamide. A 1 μ L aliquot from each of the 96 samples was robotically loaded (Microlab 4000) onto a μ CAE device for sequencing.

High-speed DNA sequencing on a μ CAE device is described in detail elsewhere (Paegel et al. 2002). Briefly, the radial μ CAE device contains 96 17-cm long capillaries sharing a common anode at the center of the device. Separate sample wells, which encircle the outer perimeter of the microplate, are connected to the 96 capillaries via a cross-channel injector. Prior to operation, all capillaries are simultaneously filled with linear polyacrylamide (CEQ Separation Gel—LPA I, Beckman Coulter) through the common anode port with a high-pressure gel loader (Scherer et al. 2001). Ninety-six unique PRS samples are then robotically loaded onto the μ CAE device. The loaded device is placed on a four-color rotary confocal fluorescence scanner (Shi et al. 1999) and fitted with an electrode array ring. Samples are then electrophoretically injected into the capillaries (500 V/cm for 30 sec) followed by fragment separation to the common anode at 150 V/cm for 29 min.

Data Analysis and Polymorphism Identification

Raw data preprocessing was performed with the program BaseFinder (Giddings et al. 1998). Leading low-quality data (primer peak and first 10 bases) and the trailing coelution peak were deleted. After baseline subtraction, color matrix parameters determined from a standard four-color M13 sequencing reaction were applied uniformly to the PRS data. A local intensity scaling factor was applied to normalize variations in template DNA concentration, dye-primer quantum yield, and photomultiplier tube gain. Next, traces were aligned to each other such that the squared difference between the sample and reference signal was minimized. A single Δ^2 plot was generated by squaring the sum of the ab-

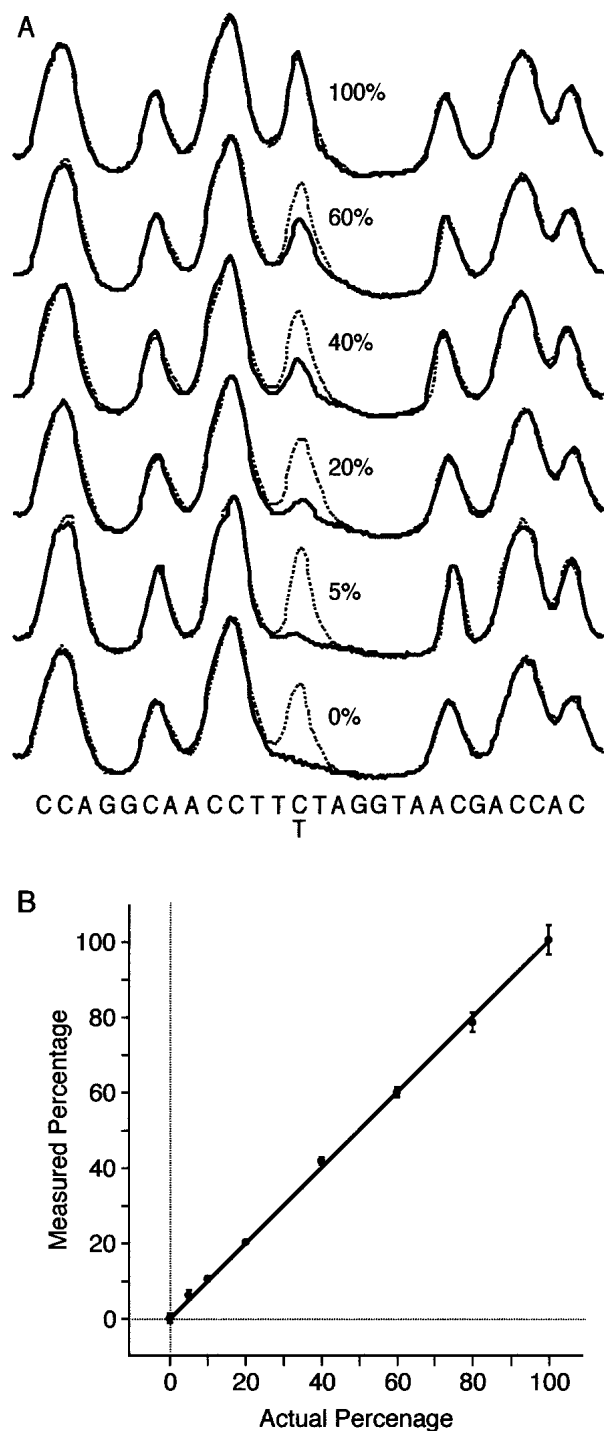


Figure 4 Minor allele frequency limit of detection analysis in multiplexed samples. (A) Two-color PRS traces comparing a single individual reference (dotted) and a multiplexed sample (solid). Multiplexed samples contain varying ratios of C→T polymorphic templates. (B) Allele frequency calibration curve generated from 32 independent traces. Error bars indicate 99% confidence interval ($R^2 = 0.9991$).

solute sample versus reference signal intensity differences from all eight traces. To identify polymorphisms, the standard

deviation of the Δ^2 plot was calculated from the observed values discretized at each peak position. Identified polymorphisms (≥ 3 SDs) were localized onto the Anderson human mitochondrial sequence (GenBank accession no. J01415; Anderson et al. 1981).

Multiplexed Sample Analysis

A SNP (*n6045* C→T) detected in the whole mitochondrial genome scan was selected to conduct a sensitivity study. For this analysis, genomic DNA from a reference sample (individual 1) and eight multiplexed samples (individuals 1 and 2) was prepared. Multiplexed samples consisted of 100%, 80%, 60%, 40%, 20%, 10%, 5%, and 0% DNA from individual 1 and a complementary amount of DNA from individual 2. Pools were prepared such that the total DNA concentration (10 ng/ μ L) remained constant. For each multiplexed and reference sample, four replicate PCR and two-color PRS reactions were performed as described above. For each trace, Gaussian curves were fit to the polymorphic and reference peak, and the ratio of peak heights was calculated. The signal-to-noise ratio was determined by dividing the average of each of the four ratios by the standard deviation.

ACKNOWLEDGMENTS

Microfabrication was carried out at the University of California Berkeley Microfabrication Laboratory. This research was supported by grants from the NIH (HG01399, P01 CA77664) and from the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under contract DEFG91ER61125. B.M.P. was supported by a National Human Genome Research Institute fellowship from the Berkeley Program in Genomics (T32 HG00047).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Anderson, S., Bankier, A.T., Barrell, B.G., Debruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* **290**: 457–465.
- Barcellos, L.F., Klitz, W., Field, L.L., Tobias, R., Bowcock, A.M., Wilson, R., Nelson, M.P., Nagatomi, J., and Thomson, G. 1997. Association mapping of disease loci, by use of a pooled DNA genomic screen. *Am. J. Hum. Genet.* **61**: 734–747.
- Berti, L., Xie, J., Medintz, I.L., Glazer, A.N., and Mathies, R.A. 2001. Energy transfer cassettes for facile labeling of sequencing and PCR primers. *Anal. Biochem.* **292**: 188–197.
- Brown, M.D., Voljavec, A.S., Lott, M.T., Torroni, A., Yang, C.C., and Wallace, D.C. 1992. Mitochondrial-DNA complex-I and complex-III mutations associated with Leber hereditary optic neuropathy. *Genetics* **130**: 163–173.
- Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S., and Fodor, S.P.A. 1996. Accessing genetic information with high-density DNA arrays. *Science* **274**: 610–614.
- Dahlman, I., Eaves, I.A., Kosoy, R., Morrison, V.A., Heward, J., Gough, S.C.L., Allahabadia, A., Franklyn, J.A., Tuomilehto, J., Tuomilehto-Wolf, E., et al. 2002. Parameters for reliable results in genetic association studies in common disease. *Nat. Genet.* **30**: 149–150.
- Fliss, M.S., Usadel, H., Cabellero, O.L., Wu, L., Buta, M.R., Eleff, S.M., Jen, J., and Sidransky, D. 2000. Facile detection of mitochondrial DNA mutations in tumors and bodily fluids. *Science* **287**: 2017–2019.
- 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.
- Giddings, M.C., Severin, J., Westphall, M., Wu, J.Z., and Smith, L.M. 1998. A software system for data analysis in automated DNA sequencing. *Genome Res.* **8**: 644–665.
- Giordano, M., Mellai, M., Hoogendoorn, B., and Momigliano-Richiardi,

- P. 2001. Determination of SNP allele frequencies in pooled DNAs by primer extension genotyping and denaturing high-performance liquid chromatography. *J. Biochem. Biophys. Methods* **47**: 101–110.
- Grossman, P.D., Bloch, W., Brinson, E., Chang, C.C., Eggerding, F.A., Fung, S., Iovannisci, D.A., Woo, S., and Winndeen, E.S. 1994. High-density multiplex detection of nucleic-acid sequences—Oligonucleotide ligation assay and sequence-coded separation. *Nucleic Acids Res.* **22**: 4527–4534.
- Ju, J.Y., Kheterpal, I., Scherer, J.R., Ruan, C.C., Fuller, C.W., Glazer, A.N., and Mathies, R.A. 1995a. Design and synthesis of fluorescence energy-transfer dye-labeled primers and their application for DNA-sequencing and analysis. *Anal. Biochem.* **231**: 131–140.
- Ju, J.Y., Ruan, C.C., Fuller, C.W., Glazer, A.N., and Mathies, R.A. 1995b. Fluorescence energy-transfer dye-labeled primers for DNA-sequencing and analysis. *Proc. Natl. Acad. Sci.* **92**: 4347–4351.
- Kogelnik, A.M., Lott, M.T., Brown, M.D., Navathe, S.B., and Wallace, D.C. 1996. MITOMAP: A human mitochondrial genome database. *Nucleic Acids Res.* **24**: 177–179.
- Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22**: 139–144.
- Kwok, P.Y., Carlson, C., Yager, T.D., Ankener, W., and Nickerson, D.A. 1994. Comparative-analysis of human DNA variations by fluorescence-based sequencing of PCR products. *Genomics* **23**: 138–144.
- Liu, V.W. S., Zhang, C.F., Linnane, A.W., and Nagley, P. 1997. Quantitative allele-specific PCR: Demonstration of age-associated accumulation in human tissues of the A→G mutation at nucleotide 3243 in mitochondrial DNA. *Hum. Mutat.* **9**: 265–271.
- Livak, K.J., Flood, S.J.A., Marmaro, J., Giusti, W., and Deetz, K. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic-acid hybridization. *PCR Methods Appl.* **4**: 357–362.
- Marth, G.T., Korf, I., Yandell, M.D., Yeh, R.T., Gu, Z.J., Zakeri, H., Stitzel, N.O., Hillier, L., Kwok, P.Y., and Gish, W.R. 1999. A general approach to single-nucleotide polymorphism discovery. *Nat. Genet.* **23**: 452–456.
- Paegel, B.M., Emrich, C.A., Weyemayer, G.J., Scherer, J.R., and Mathies, R.A. 2002. High throughput DNA sequencing with a microfabricated 96-lane capillary array electrophoresis bioprocessor. *Proc. Natl. Acad. Sci.* **99**: 574–579.
- Pastinen, T., Partanen, J., and Syvanen, A.C. 1996. Multiplex, fluorescent, solid phase minisequencing for efficient screening of DNA sequence variation. *Clin. Chem.* **42**: 1391–1397.
- Polyak, K., Li, Y.B., Zhu, H., Lengauer, C., Willson, J.K.V., Markowitz, S.D., Trush, M.A., Kinzler, K.W., and Vogelstein, B. 1998. Somatic mutations of the mitochondrial genome in human colorectal tumours. *Nat. Genet.* **20**: 291–293.
- Rieder, M.J., Taylor, S.L., Tobe, V.O., and Nickerson, D.A. 1998. Automating the identification of DNA variations using quality-based fluorescence resequencing: Analysis of the human mitochondrial genome. *Nucleic Acids Res.* **26**: 967–973.
- Risch, N. and Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science* **273**: 1516–1517.
- Samiotaki, M., Kwiatkowski, M., Parik, J., and Landegren, U. 1994. Dual-color detection of DNA-sequence variants by ligase-mediated analysis. *Genomics* **20**: 238–242.
- Scherer, J.R., Paegel, B.M., Wedemayer, G.J., Emrich, C.A., Lo, J., Medintz, I.L., and Mathies, R.A. 2001. High-pressure matrix loading device for electrophoresis microchips. *Biotechniques* **31**: 1150–1156.
- Shi, Y.N., Simpson, P.C., Scherer, J.R., Wexler, D., Skibola, C., Smith, M.T., and Mathies, R.A. 1999. Radial capillary array electrophoresis microplate and scanner for high-performance nucleic acid analysis. *Anal. Chem.* **71**: 5354–5361.
- Shumaker, J.M., Metspalu, A., and Caskey, C.T. 1996. Mutation detection by solid phase primer extension. *Hum. Mutat.* **7**: 346–354.
- Syvanen, A.C., Sajantila, A., and Lukka, M. 1993. Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing. *Am. J. Hum. Genet.* **52**: 46–59.
- Tobe, V.O., Taylor, S.L., and Nickerson, D.A. 1996. Single-well genotyping of diallelic sequence variations by a two-color ELISA-based oligonucleotide ligation assay. *Nucleic Acids Res.* **24**: 3728–3732.
- Tyagi, S., Bratu, D.P., and Kramer, F.R. 1998. Multicolor molecular beacons for allele discrimination. *Nat. Biotechnol.* **16**: 49–53.

WEB SITE REFERENCES

<http://www.mitomap.org>; MITOMAP database.

Received May 6, 2002; accepted in revised form December 4, 2002.