

Digital genotyping and haplotyping with polymerase colonies

Robi D. Mitra^{*†}, Vincent L. Butty[‡], Jay Shendure^{*}, Benjamin R. Williams^{*}, David E. Housman[§], and George M. Church^{*¶}

^{*}Lipper Center for Computational Genetics and Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115;

[‡]Graduate Program in Immunology, Harvard Medical School, Joslin Diabetes Center, 1 Joslin Center, Boston, MA 02215; and [§]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Melvin I. Simon, California Institute of Technology, Pasadena, CA, and approved March 6, 2003 (received for review October 21, 2002)

Polymerase colony (polony) technology amplifies multiple individual DNA molecules within a thin acrylamide gel attached to a microscope slide. Each DNA molecule included in the reaction produces an immobilized colony of double-stranded DNA. We genotype these colonies by performing single base extensions with dye-labeled nucleotides, and we demonstrate the accurate quantitation of two allelic variants. We also show that polony technology can determine the phase, or haplotype, of two single-nucleotide polymorphisms (SNPs) by coamplifying distally located targets on a single chromosomal fragment. We correctly determine the genotype and phase of three different pairs of SNPs. In one case, the distance between the two SNPs is 45 kb, the largest distance achieved to date without separating the chromosomes by cloning or somatic cell fusion. The results indicate that polony genotyping and haplotyping may play an important role in understanding the structure of genetic variation.

The study of genetic variation in the human population has the potential to greatly improve human health, both by predicting susceptibility to disease and guiding choice of therapy. The most common genetic variations in the human population are single-nucleotide polymorphisms (SNPs). By studying candidate genes and performing genome-wide linkage disequilibrium studies, scientists are trying to uncover the “causative SNP,” the SNP that alters gene function and thereby increases the risk of disease. It has been shown that deriving haplotypes increases the efficiency of linkage disequilibrium studies (1). Surprisingly, recent studies (2, 3) suggest that haplotypes will also be critical for candidate gene studies. These studies found that for some diseases, there is not one single SNP that is responsible for altering gene function, but instead, multiple SNPs interact to alter the function or expression of a protein (4). These alterations occur only when specific combinations of SNPs are present on the same chromosome, so one must determine the haplotype to find a correlation to the observed phenotype. In these cases, the focus has shifted from a causative SNP to a causative haplotype.

The most common approach for determining the haplotype, or phase, of a set of SNPs is computational inference from unphased data (3, 5–8). Although this methodology has greatly increased the power of both linkage and candidate gene studies, a recent study estimates the error rate to be between 19 and 48%, depending on the algorithm used (6). This error rate presents challenges for the use of this method as a research tool and makes it an unlikely candidate for use in the clinical setting. Recent findings suggest a way to improve the accuracy of haplotype inference. Daly *et al.* (9) and others (10, 11) have shown that SNPs tend to be inherited in larger haplotype blocks than previously thought, and that there are relatively few variants of each block. This observation has sparked a public effort to characterize all common haplotypes in the human population (12). Prior knowledge of common haplotype blocks may make it easier to infer the phase of SNPs that lie within the same haplotype block (13). However, even with this knowledge, it will be difficult to accurately predict the haplotype of two SNPs that

are in different haplotype blocks because these two SNPs will not typically be in linkage disequilibrium. These cases may not be rare because the genome contains ≈ 100 kb per gene and the average haplotype block is only 22 kb in European and Asian populations and 11 kb in Yoruban and African-American populations (11). This point is well illustrated by two known mutations, R347→H and A970→D, in the *CFTR* gene (14). These mutations are separated by 65 kb of genomic sequence and are not likely to be in linkage disequilibrium (http://pga.mbt.washington.edu/summary_data.html). Yet, when present in cis, they interact to produce more severe symptoms of cystic fibrosis than when present in trans, or when only one mutation is present. If haplotypes are to be used in the clinic as a prognostic marker, a direct molecular haplotyping technology that can phase SNPs separated by 50–100 kb is necessary.

Current methods for the experimental determination of haplotypes have limitations. Allele-specific PCR (15) and single-molecule PCR (16) require optimization and cannot routinely determine the phase of SNPs separated by >10 –15 kb. Atomic force microscopy (17) is a promising alternative, but it requires expensive equipment not commonly found in a molecular biology laboratory, and it is unclear how easily this technology can be scaled. Methods in which chromosomal fragments are cloned into bacterial artificial chromosomes, or in which somatic cell hybrids (1, 10) are made, are useful for studies in which a large number of SNPs must be phased. However, they are not cost-effective for the phasing of a small number of SNPs on a large number of samples (1), as would be necessary for a clinical diagnostic.

Here, we present a method to determine digital genotypes and haplotypes using polymerase colony (polony) technology (18). Using polonies, we show that the ratio of two allelic variants can be precisely quantified by counting molecules of each allele. Next, we determine the genotype and phase of three different pairs of SNPs up to 45 kb apart. We show this technology requires very little DNA as input; a buccal swab provides enough DNA to perform hundreds of reactions. We also provide evidence that a large number of polony assays can be performed on a single microscope slide, reducing the cost per assay.

Materials and Methods

Polony Amplification. Polony amplification was performed as described (18), except the acrylamide gels were 40 μ m thick and BSA was added to the gel mix at a final concentration of 0.2%. To pour these gels, we used slides partially covered with a Teflon coating (Erie Scientific, Erie, PA), which served as a spacer between the glass surface of the slide and a coverslip (Fisher Scientific). These slides were covered with mineral oil to prevent

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: polony, polymerase colony; SBE, single base extension; SNP, single-nucleotide polymorphism; LOD, logarithm of odds.

[†]Present address: Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110.

[¶]To whom correspondence should be addressed. E-mail: church@arep.med.harvard.edu.

evaporation and thermally cycled. For haplotyping reactions, four primers were used (two forward and two reverse primers) at a concentration of 0.25 μM each primer.

Polony Denaturation. After polony amplification, we removed the unattached DNA strand by incubating in 70°C denaturing buffer (70% formamide, 1× SSC) for 15 min and performing electrophoreses in 0.5× TBE (90 mM Tris/64.6 mM boric acid/2.5 mM EDTA, pH 8) with 42% urea for 1 h at 5–10 v/cm. We washed the slides 2 × 4 min in wash buffer 1 (10 mM Tris-HCl, pH 7.5/50 mM KCl/2 mM EDTA/0.01% Triton X-100).

For polony haplotyping reactions, we removed extended primers after the first single base extension (SBE) reaction by washing denaturing buffer at 70° and then washing 2 × 5' in dH₂O.

SBE Reactions. We performed the SBE reactions for the polony haplotyping and genotyping experiments with fluorescent deoxynucleotides. To do so, the acrylamide gel was covered with a frame seal chamber (MJ Research, Cambridge, MA) and annealing mix (0.25 μM SBE sequencing primer, 6× SSPE [standard saline phosphate/EDTA (0.15 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA/0.01% Triton X-100)] was added over the gel. The slides were heated at 94°C for 2 min, then at 56°C for 15 min. We removed unannealed primer by washing the slides 2 × 4 min in wash buffer 1 and then equilibrated the slides in 1× Klenow buffer (10 mM Tris-HCl, pH 7.5/10 mM MgCl₂). Next, we covered the gel with 40 μl of extension mix (1× Klenow buffer/0.12 units/ μl Klenow exo-polymerase/33 ng/ μl *Escherichia coli* single-stranded binding protein/1 μM Cy3 and Cy5 labeled deoxynucleotide) for 2 min at room temperature and then washed the slides in wash buffer 1. The slides were scanned on a scanning confocal microscope designed for microarrays (Scanarray 5000, Perkin-Elmer).

We performed SBE reactions with dideoxynucleotides as follows: we covered the acrylamide gel with a frame seal chamber (MJ Research) and then added annealing mix (0.25 μM sequencing primer/6× SSPE/0.01% Triton X-100) over the gel. We heated the slides at 94°C for 2 min, then at 56°C for 15 min. We removed unannealed primer by washing the slides 2 × 4 min in wash buffer 1. We equilibrated the slides in 1× Amplitaq FS buffer (10 mM Tris-HCl, pH 8.0/50 mM KCl/1.5 mM MgCl₂) and covered the gel with 40 μl of extension mix (1× Amplitaq FS buffer/1 μM TAMRA-12-ddUTP (Perkin-Elmer)/1 μM ROX-ddCTP (Perkin-Elmer)/1 μM Cy5-ddATP/1 μM FITC-ddGTP/Amplitaq FS 0.1 units/ μl /33 ng/ μl *E. coli* single-stranded binding protein). The slides were heated to 55° for 4 min and then washed and scanned as above.

Image Analysis. We acquired images of polony gels in .TIF format. We filtered these images with a Wiener filter and a median filter to remove speckle and noise and subtracted background fluorescence. We identified polonies computationally using the IMAGEQUANT (Amersham Biosciences, Piscataway, NJ) software package. This package quantified the fluorescent intensity of each polony and output the data as a text file. We then identified overlapping polonies using a MATLAB (Mathworks, Natick, MA) script, HAPCALL. For the polony genotyping experiment in which the relative abundance of two alleles is measured, we smoothed the images, identified the polonies, and called their genotypes using the MATLAB script polony_call.m. All MATLAB scripts are available on request.

Calculating the Logarithm of Odds (LOD) Score. The details of the LOD score calculations can be found in *Supporting Text*, which is published as supporting information on the PNAS web site, www.pnas.org.

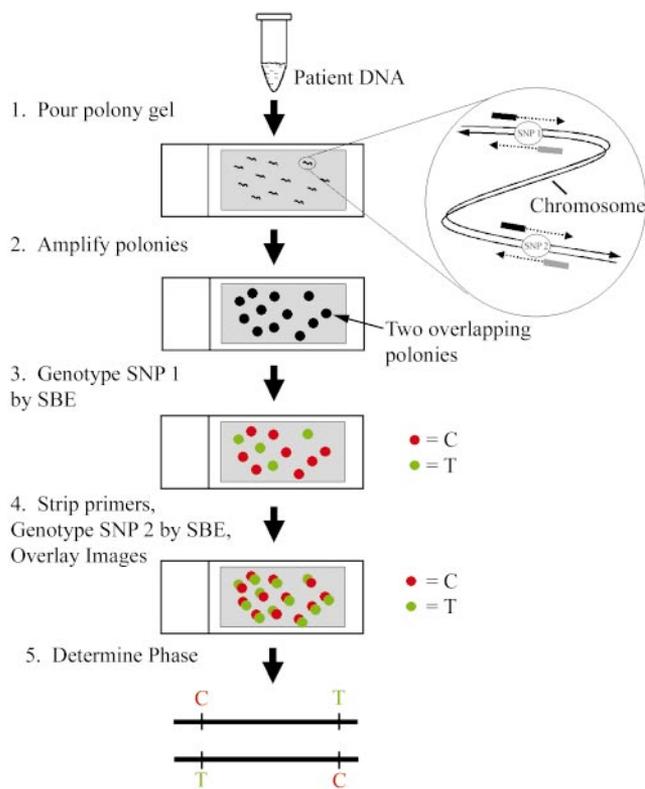


Fig. 1. Polony haplotyping. 1, Polymerize dilute sample of patient DNA into a polyacrylamide gel on a glass microscope slide. 2, Amplify single chromosomes by PCR with two primer pairs flanking the SNPs of interest (*Inset*). 3, Genotype the first SNP by performing an SBE reaction and scan the slide using a laser scanner. 4, Genotype the second SNP, scan the slide, and 5, merge the images to detect overlapping polonies. The haplotype is determined directly from the overlapping polonies.

Oligonucleotides and Patient DNA. Primer sequences and patient DNA used in this study are published as supporting information on the PNAS web site.

Results

In the polony haplotyping approach (Fig. 1), we use a small amount of patient DNA polymerized into an acrylamide gel containing all of the reagents necessary for PCR. Two pairs of PCR primers are included, each pair capable of amplifying one of the SNPs of interest. Because the concentration of patient DNA is so low, the chromosome fragments are well separated from each other on the surface of the slide. We then perform an in-gel PCR reaction by using a thermal cycler designed for slides to amplify two loci from a single DNA molecule. Because the acrylamide matrix restricts the diffusion of amplification products, double-stranded DNA accumulates around the chromosome, forming two overlapping polonies, each amplified from a different region on the same molecule of DNA. A key feature of this protocol is the use of modified primers in the PCR reaction that covalently attach one strand of the amplified DNA to the acrylamide matrix (19, 20). This feature allows us to remove the unattached strand from all polonies by heating and washing the slide, leaving single-stranded templates for subsequent SBE reactions (21–23). After genotyping all polonies at both loci by SBE, the phase of the SNPs is then determined by identifying overlapping polonies.

Amplification and Characterization of Multiple Polonies from Single DNA Molecules. To perform the haplotyping protocol, we needed to first establish that (*i*) multiple polonies could be amplified

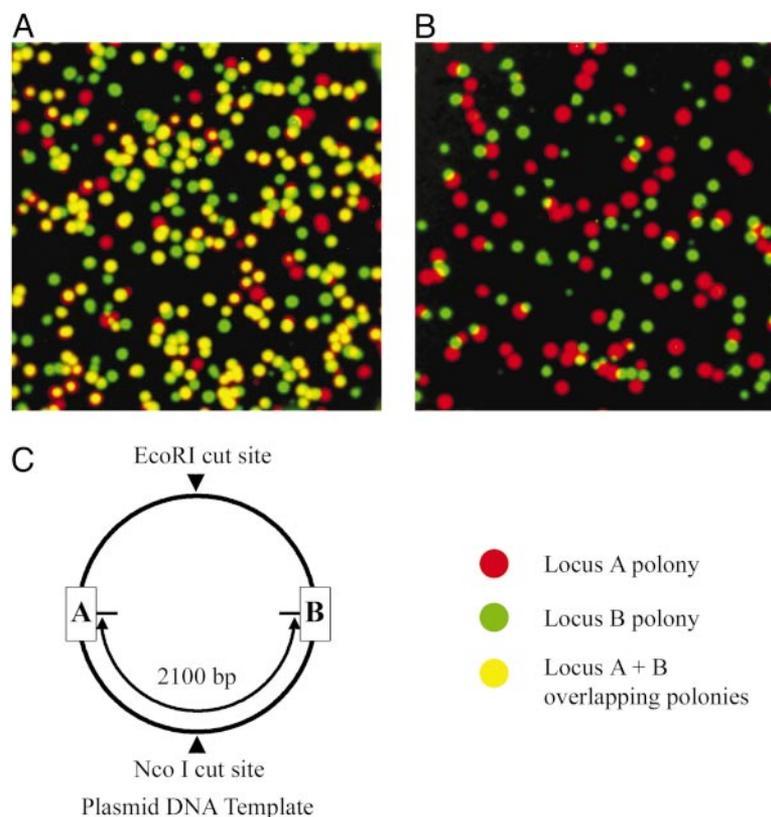


Fig. 2. Amplification of two polonies from a single DNA molecule. We cut plasmid DNA (C) with *EcoRI*, leaving locus A and locus B on the same template molecule, or with *EcoRI* and *NcoI*, separating the loci. We performed polony amplification with two pairs of primers designed to amplify loci A and B and hybridized cy5-labeled probe (red, locus A) and cy3-labeled probe (green, locus B) to the polonies. Eighty-six percent of locus A polonies overlapped a locus B polony (and 80% of locus B polonies overlapped a locus A polony) when the plasmid was singly cut (A), but <2% of locus A polonies significantly overlapped a locus B polony when the plasmid was doubly cut (B).

from a single molecule of DNA, and (ii) SBE reactions could be performed on polonies covalently attached to the acrylamide gel. To demonstrate that two polonies could be amplified from a single DNA molecule, we used two different primer pairs to amplify two separate loci from a linearized plasmid template. The polonies were then visualized by hybridizing locus specific primers labeled with fluorescent dyes to the acrylamide array (Fig. 2A). Numerous overlapping polonies could be identified after hybridization, as evident from the large number of yellow polonies. In a separate control reaction, we cleaved the plasmid with two restriction endonucleases, so that the two loci were no longer on the same DNA molecule (Fig. 2B). The doubly cut plasmid produced few polonies that overlapped. The polonies that did overlap did so only near their edges and were the result of two different DNA molecules falling near each other when the gel was poured. These results demonstrate that multiple polonies can be amplified from a single DNA molecule.

We next characterized the specificity of SBE on acrylamide-immobilized DNA. We used a single dye-labeled deoxynucleotide or dideoxynucleotide to extend primer/template duplexes by one base in a DNA polymerase catalyzed reaction. We performed four reactions for each nucleotide to determine the specificity of the SBE reaction for the correct base relative to all possible mismatches. The results are shown in Table 1. SBE reactions with both fluorescent deoxynucleotides and dideoxynucleotides incorporated only the correct base. We chose to use fluorescent deoxynucleotides in our SBE reactions because they perform better than fluorescent dideoxynucleotides, are not as expensive, and are compatible with our long-term goals of sequencing 30–100 bases from each polony (24).

Polony Genotyping. Quantifying the relative expression of two different alleles of the same gene is important for understanding the genetic basis of human variation (25) and understanding disease (26). We tested our ability to measure small changes in

Table 1. Specificity of SBE reaction with fluorescent (A) deoxynucleotides and (B) dideoxynucleotides

SBE with fluorescent deoxynucleotides				
Nucleotide added	Expected base			
	A	C	G	T
TMR-dATP	100 (10.8)	8.8 (0.8)	5.6 (0.8)	3.4 (0.5)
Cy3-dCTP	0.3 (0.1)	100 (5.3)	0.42 (0.1)	0.55 (0.1)
TMR-dGTP	2.4 (0.4)	0.8 (0.3)	100 (6.2)	1.5 (1.0)
Cy5-dUTP	0.18 (0.2)	0.58 (0.4)	0.59 (0.5)	100 (6.0)
SBE with fluorescent dideoxynucleotides				
Nucleotide added	Expected base			
	A	C	G	T
Cy5-ddATP	100 (16.1)	0.4 (0.1)	0.4 (0.1)	0.5 (0.2)
ROX-ddCTP	3.5 (0.1)	100 (15.2)	6.1 (1.0)	6.3 (1.1)
FITC-ddGTP	26.1 (4.8)	23.2 (6.6)	100 (6.0)	19.1 (2.3)
TAMRA-ddUTP	12.8 (5.9)	11.0 (2.5)	8.4 (2.8)	100 (5.0)

We quantified the incorporated fluorescence after SBE reactions with the correct base (bold) and the three possible mismatch bases. The fluorescence is normalized so the correct base incorporated an average of 100 relative fluorescent units. The standard deviation of the measurements is shown in parentheses ($n = 4$).

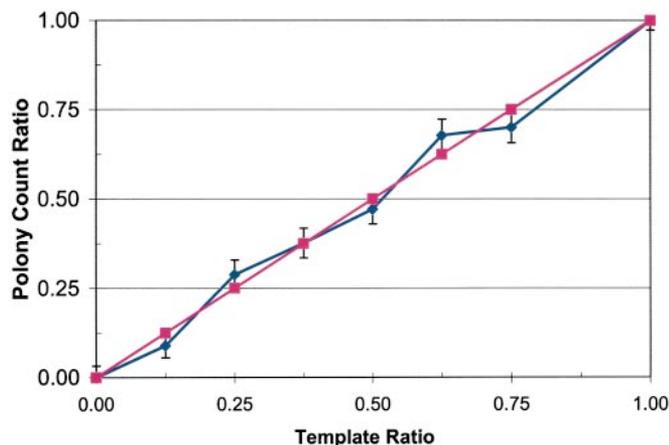


Fig. 3. Determining the allelic ratios using polony genotyping. We purified genomic DNA from two inbred strains of mouse with different alleles of the creatine kinase gene. We mixed the DNA at various ratios and performed polony genotyping. The observed allelic ratios correlate strongly to the predicted ratios ($r = 0.99$, error bars are shown at a 95% confidence level).

the ratio of two alleles by genotyping genomic DNA from two inbred strains of mouse with different alleles of the creatine kinase gene. We mixed DNA from each strain at various ratios and determined the relative abundance of each allele by performing polony amplification followed by SBE. Polonies were automatically identified and counted by using in-house software, and the results are shown in Fig. 3. The observed ratios (blue) were strongly correlated to the expected ratios (pink) with a correlation coefficient of 0.99. (Error bars represent 95% confidence level.) This demonstrates that precise quantitation of the relative abundance of two alleles can be determined by polony genotyping.

Polony Haplotyping on Patient DNA. To test whether we could use the polony technology to correctly determine both the genotype and haplotype from patient samples, we examined two SNPs that are 11.8 kb apart on chromosome 7. These SNPs, DK438, a T→C mutation in intron 4 of the *CFTR* gene, and DK445-2, a T→C mutation in intron 9 of the *CFTR* gene, have been previously characterized (D. Keen and D.E.H., unpublished work), and their phase is known because they are in strong linkage disequilibrium with one another. We used samples from two patients that were heterozygous at these alleles and amplified both loci in a polony reaction. Next, we performed an SBE reaction to genotype the SNP DK438 (Fig. 4A). In this example, green polonies correspond to the wild-type allele (T) and red polonies correspond to the mutant allele (C). Next, the primers were stripped from the slides, and we performed a second SBE reaction to genotype the SNP DK445-2 (Fig. 4B). Again, green polonies correspond to the wild-type allele (T) and red polonies correspond to the mutant allele (C). The images were merged, and overlapping polonies were identified computationally (*Materials and Methods*) and are circled in Fig. 4A and B. We found 22 overlapping polonies (Table 3, which is published as supporting information on the PNAS web site), all of which indicated the correct haplotype with T at DK438 in cis with T at DK445-2.

When polonies are amplified in a polony haplotyping reaction, there is a chance of observing overlapping polonies, not because the two polonies were amplified from the same molecule of DNA, but because two different DNA molecules happened to settle very close to one another when the gel was poured. Pairs of overlapping polonies that occur in this fashion do not provide information about the phase of the two SNPs being queried. To confirm that this phenomenon did not produce errors in our

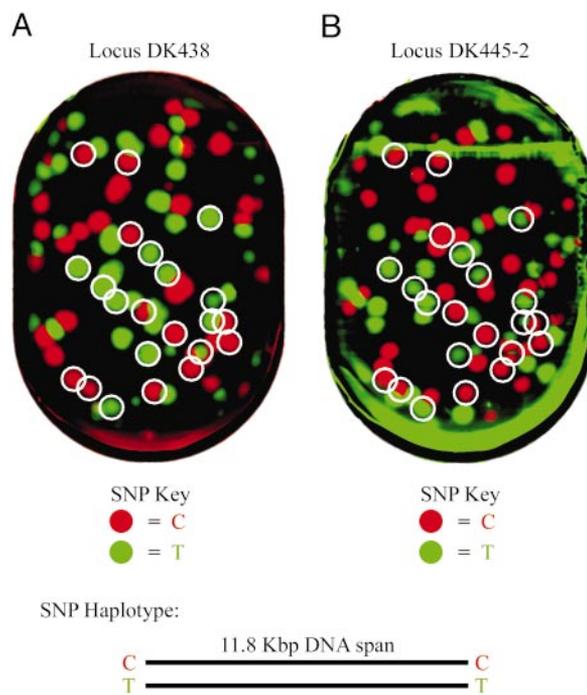


Fig. 4. Polony haplotyping of SNPs in the *CFTR* gene. We performed polony amplification on the SNPs DK438 and DK445-2. First, we genotyped DK438 by SBE (A), stripped the extended primers from the gel, and genotyped DK445-2 (B). Twenty-two overlapping polonies were identified computationally and have been circled in A and B. All 22 polonies indicated that the “c” variant at SNP DK438 was in cis with the “c” variant at DK445-2 (LOD = 26.2).

called haplotypes, we estimated the LOD score of the two possible haplotypes (see *Supporting Text*). For the two samples used in the polony haplotyping reaction, we found the LOD scores to be 23.8 and 26.2 (Table 3), indicating that we can have high confidence in the called haplotype.

One advantage of the polony haplotyping technology is that it requires very little patient DNA. In fact, a buccal swab should provide enough DNA for many haplotyping reactions. To test this hypothesis, we collected buccal swabs from five subjects and haplotyped two SNPs recently shown to be involved in hypolactasia (27). Because these SNPs, a G→A variant 22,018 bases upstream of the gene *MCM6* (designated G/A -22018), and a C→T variant 13,910 bases upstream of the gene *MCM6* (designated C/T -13910) are in strong linkage disequilibrium, we expected to find only one of the two possible haplotypes, G in cis with C, or G in cis with T. The results of these polony haplotyping reactions are summarized in Table 2. For all samples, the G variant at -22,018 was found to be on the same chromosome as the C variant at -13,910, consistent with the predicted linkage. In some samples, not every pair of overlapping polonies called the same haplotype. For example, in the patient PHS2, 32 pairs of overlapping polonies indicated the correct haplotype, but three pairs indicated the other haplotype. These overlapping polonies were most likely amplified from different template molecules that happened to overlap because the DNA was plated at a relatively high density. Despite these occasional overlapping pairs with a dissenting prediction, the calculated LOD scores (Table 2) indicate that we can have high confidence in the called haplotype.

The two pairs of SNPs that were haplotyped in the above examples were separated by 11.8 and 8.1 kb of genomic sequence, respectively. However, in principle, one should be able to phase two SNPs separated by any distance, as long as the sample is not sheared or degraded to the extent that most of the DNA

Table 2. Polony haplotyping at SNPs G/A –22018 and C/T –13910 in *MCM6* involved in lactose intolerance

Patient	C → T –13910 polonies	G → A –22018 polonies	Overlapping polonies	GC pairs	GT pairs	Called haplotype	LOD score
PHS1	325	328	49	46	3	GC	35.4
PHS2	295	334	35	32	3	GC	21.1
PHS3	232	234	38	38	0	GC	37.3
PHS4	286	290	25	21	4	GC	11.6
PHS5	274	274	11	10	1	GC	3.8

molecules in the sample are too short to contain both SNPs. To assess the degree of degradation in our samples, we performed agarose gel electrophoresis and chose the sample that contained the least amount of low molecular weight fragments. We estimated the average fragment size of this sample to be greater than 80 kb. Next, we performed polony haplotyping to phase two SNPs, DK331 and DK445-2, in the *CFTR* gene that are separated by 45 kb. There were 153 polonies amplified at the locus surrounding SNP DK331 (Fig. 5A) and 175 polonies amplified at the locus surrounding SNP DK445-2 (Fig. 5B). We identified 34 overlapping polonies from the merged scans, and 32 of these indicated the same haplotype. From this, we conclude that the A variant at SNP DK331 is on the same chromosome as the T variant at DK445-2 (LOD score 20.3).

Discussion

As a prerequisite to determining haplotypes, it was necessary to demonstrate that polony technology could be used to determine genotypes. Polony genotyping is an inherently digital process; each DNA or RNA molecule is separately counted and provides one bit of information. The power of a digital genotyping has been demonstrated by Digital PCR (28), as well as other similar technologies (29, 30), finding applications in detecting loss of

heterozygosity (31), quantifying allelic imbalance (26), and detecting rare somatic mutations (30). The polony genotyping presented here should further extend the utility of the digital genotyping, because millions of polonies (18) can be genotyped in a single reaction.

The polony haplotyping reactions performed in this study used two pairs of primers to amplify two different polony products from the same molecule. This strategy was used due to the large distance (8–45 kb) between the pairs of SNPs haplotyped. For SNPs that are 1–4 kb apart, another feasible strategy is to use one pair of primers to amplify both SNPs in one single polony and then genotype both SNPs. In the strategy we used, two polonies must be amplified from a single DNA molecule; however, not every template molecule included in the reaction must give rise to two polonies to correctly determine the haplotype. In our experiments, the ratio of overlapping polony pairs to the total number of polonies amplified at a particular SNP locus was highly variable, ranging from 4% in one of the hypolactasia experiments (Table 2) to 86% in the plasmid experiment (Fig. 2). We define this ratio as the coamplification efficiency. Because 100–400 polonies were analyzed at each locus, accurate haplotypes were obtained in all cases; however, increasing the coamplification efficiency would decrease the number of polonies that need to be analyzed. Therefore, an important question is, what are the parameters that affect the coamplification efficiency?

We have found that the presence of ungelled acrylamide in the polony gel during thermal cycling decreases the coamplification efficiency. For this reason, we degassed all reagents used in the polony amplification and polymerized the gels under argon to minimize the amount of ungelled monomer present. In an alternate protocol, we poured the acrylamide gels on the slide, washed away any ungelled acrylamide, diffused in the templates and PCR reagents, and then performed the polony amplification (see *Supporting Text*).

DNA fragmentation or degradation can also result in low measurements of coamplification efficiency. If a number of DNA molecules in the polony amplification do not contain both loci, then there will be fewer overlapping polonies, and therefore the coamplification efficiency measurement will be lower than expected. This may explain why the DNA samples from buccal swabs (Table 2), which were purified by using a fast but relatively crude protocol, displayed a lower fraction of overlapping polonies (measured coamplification efficiencies of 4–15%) than the other DNA samples, which were purified by more standard methods (measured coamplification efficiencies of 15–34%). By using methods designed to purify intact chromosomes (32), we should further increase the coamplification efficiency.

For some applications, it is desirable to haplotype a large number (100–10,000) of SNPs along a chromosome. To achieve this, multiplexed reactions could be combined with polony arrays to obtain a large number of haplotypes per slide. PCR reactions can be multiplexed 30-fold or more on a routine basis (33). Multiplexing would not require any increase in the average coamplification efficiency, because phasing is transitive; if SNPs A, B, and C are heterozygous and we know the phase of SNP A and SNP B and the phase of SNP B and SNP C, then we know

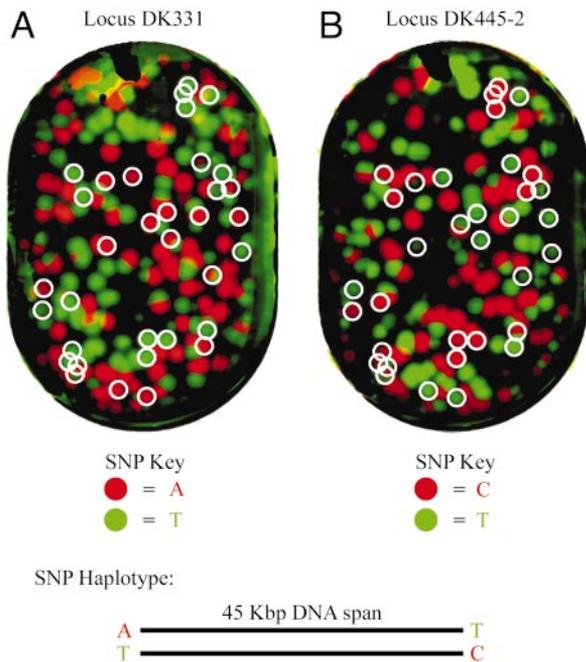


Fig. 5. Phasing of SNPs separated by a genomic distance of 45 kb. We amplified and genotyped SNPs DK331 and DK445-2 by SBE. One hundred fifty-three polonies were genotyped at the DK331 locus (A), and 175 polonies were genotyped at the DK445-2 locus (B). We identified 34 overlapping polonies (circled in the above figure), 32 of which indicated that the A variant at SNP DK 331 is in cis with the T variant of SNP DK445-2 (LOD = 20.3).

the phase of SNP A and SNP C. This means that, in a multiplexed haplotyping reaction with M primer pairs, it is not necessary that any one DNA molecule amplifies M overlapping polonies, only that enough pairs of overlapping polonies are amplified to infer the total haplotype for all SNPs. By using 50- μ m polonies (Fig. 6, which is published as supporting information on the PNAS web site) in a multiplexed protocol, we should be able to obtain the entire haplotype by using a single slide.

To our knowledge, no molecular haplotyping technology has previously demonstrated the phasing of two SNPs >20 kb apart (34) without separating the chromosomes by cloning or somatic cell fusion. Here we haplotype two SNPs 45 kb apart, and there are no apparent barriers to working with larger distances. Polony haplotyping also requires very small amounts of patient DNA, simplifying its collection. These features suggest polony haplotyping will become a valuable tool in understanding genetic

variation. Because this technique amplifies single DNA molecules, it also should be possible to determine the relative frequency of different haplotypes in pooled samples, a feature that may further reduce the cost of linkage disequilibrium studies. Performing multiple rounds of SBE with fluorescent deoxynucleotides would extend haplotyping closer to genomic sequencing. The notion of measuring what alleles are in cis in a piece of DNA is generalizable to determining what sequences are in cis on each single molecule of DNA or RNA.

We thank Diane Keen for the *CFTR* patient samples and Xiaohua Huang for advice on acrylamide and nucleic acid chemistry. We thank Aimee Dudley, Barak Cohen, Martin Steffen, Vasu Badarinarayana, and other members of the Church lab for helpful discussions and critical readings of the manuscript. This work was supported by the U.S. Department of Energy (DE-FG02-87-ER60565).

- Douglas, J. A., Boehnke, M., Gillanders, E., Trent, J. M. & Gruber, S. B. (2001) *Nat. Genet.* **28**, 361–364.
- Drysdale, C. M., McGraw, D. W., Stack, C. B., Stephens, J. C., Judson, R. S., Nandabalan, K., Arnold, K., Ruano, G. & Liggett, S. B. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 10483–10488.
- Hoehle, M. R., Kopke, K., Wendel, B., Rohde, K., Flachmeier, C., Kidd, K. K., Berrettini, W. H. & Church, G. M. (2000) *Hum. Mol. Genet.* **9**, 2895–2908.
- Davidson, S. (2000) *Nat. Biotechnol.* **18**, 1134–1135.
- Stephens, M., Smith, N. J. & Donnelly, P. (2001) *Am. J. Hum. Genet.* **68**, 978–989.
- Niu, T., Qin, Z. S., Xu, X. & Liu, J. S. (2002) *Am. J. Hum. Genet.* **70**, 157–169.
- Tishkoff, S. A., Pakstis, A. J., Ruano, G. & Kidd, K. K. (2000) *Am. J. Hum. Genet.* **67**, 518–522.
- Hawley, M. E. & Kidd, K. K. (1995) *J. Hered.* **86**, 409–411.
- Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. (2001) *Nat. Genet.* **29**, 229–232.
- Patil, N., Berno, A. J., Hinds, D. A., Barrett, W. A., Doshi, J. M., Hacker, C. R., Kautzer, C. R., Lee, D. H., Marjoribanks, C., McDonough, D. P., et al. (2001) *Science* **294**, 1719–1723.
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., et al. (2002) *Science* **296**, 2225–2229.
- Couzin, J. (2002) *Science* **296**, 1391–1393.
- Zhang, K., Deng, M., Chen, T., Waterman, M. S. & Sun, F. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 7335–7339.
- Clain, J., Fritsch, J., Lehmann-Che, J., Bali, M., Arous, N., Goossens, M., Edelman, A. & Fanen, P. (2001) *J. Biol. Chem.* **276**, 9045–9049.
- Michalatos-Beloin, S., Tishkoff, S. A., Bentley, K. L., Kidd, K. K. & Ruano, G. (1996) *Nucleic Acids Res.* **24**, 4841–4843.
- Ruano, G., Kidd, K. K. & Stephens, J. C. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6296–6300.
- Woolley, A. T., Guillemette, C., Li Cheung, C., Housman, D. E. & Lieber, C. M. (2000) *Nat. Biotechnol.* **18**, 760–763.
- Mitra, R. D. & Church, G. M. (1999) *Nucleic Acids Res.* **27**, e34.
- Rehman, F. N., Audeh, M., Abrams, E. S., Hammond, P. W., Kenney, M. & Boles, T. C. (1999) *Nucleic Acids Res.* **27**, 649–655.
- Vasiliskov, A. V., Timofeev, E. N., Surzhikov, S. A., Drobyshev, A. L., Shick, V. V. & Mirzabekov, A. D. (1999) *Biotechniques* **27**, 592–594, 596–598, 600.
- Pastinen, T., Partanen, J. & Syvanen, A. C. (1996) *Clin. Chem.* **42**, 1391–1397.
- Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. & Syvanen, A. C. (1997) *Genome Res.* **7**, 606–614.
- Dubiley, S., Kirillov, E. & Mirzabekov, A. (1999) *Nucleic Acids Res.* **27**, e19.
- Mitra, R. D. (2000) Ph.D. thesis (Massachusetts Institute of Technology, Cambridge, MA).
- Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B. & Kinzler, K. W. (2002) *Science* **297**, 1143.
- Yan, H., Dobbie, Z., Gruber, S. B., Markowitz, S., Romans, K., Giardiello, F. M., Kinzler, K. W. & Vogelstein, B. (2002) *Nat. Genet.* **30**, 25–26.
- Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L. & Jarvela, I. (2002) *Nat. Genet.* **30**, 233–237.
- Vogelstein, B. & Kinzler, K. W. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9236–9241.
- Chetverina, H. V., Samatov, T. R., Ugarov, V. I. & Chetverin, A. B. (2002) *Biotechniques* **33**, 150–152, 154, 156.
- Lizardi, P. M., Huang, X., Zhu, Z., Bray-Ward, P., Thomas, D. C. & Ward, D. C. (1998) *Nat. Genet.* **19**, 225–232.
- Zhou, W., Galizia, G., Lieto, E., Goodman, S. N., Romans, K. E., Kinzler, K. W., Vogelstein, B., Choti, M. A. & Montgomery, E. A. (2001) *Nat. Biotechnol.* **19**, 78–81.
- Kwan, H. S., Li, C. C., Chiu, S. W. & Cheng, S. C. (1991) *Nucleic Acids Res.* **19**, 1347.
- Kokoris, M., Dix, K., Moynihan, K., Mathis, J., Erwin, B., Grass, P., Hines, B. & Dueterhoeft, A. (2000) *Mol. Diagn.* **5**, 329–340.
- McDonald, O. G., Krynetski, E. Y. & Evans, W. E. (2002) *Pharmacogenetics* **12**, 93–99.