Molecular haplotyping by linking emulsion PCR: analysis of paraoxonase 1 haplotypes and phenotypes

James G. Wetmur1,2,* Madhu Kumar1 Li Zhang1 Caroline Palomeque3 Sylvan Wallenstein4 and Jia Chen3

1Department of Microbiology, 2Department of Human Genetics, 3Department of Community and Preventive Medicine and 4Department of Biomathematical Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA

ABSTRACT

Linking emulsion PCR (LE-PCR) enables formation of minichromosomes preserving phase information of two polymorphic loci, hence the haplotype. Emulsion PCR confines two amplicons of two linked polymorphic sites on a single template molecule to one aqueous-phase droplet. Linking PCR uses biotinylated, overlapping linking primers to connect these amplicons in the droplet. After LE-PCR, unlinked amplicons are removed on streptavidin-coated magnetic beads and single-stranded runoff products are capped by primer extension. Quantitative ASPCR can then be used to ascertain the haplotypes of the two polymorphic loci on the minichromosomes. Using LE-PCR, we determined the human paraoxonase-1 [PON1] molecular haplotypes at three loci (−909g>c, L55M, Q192R) in women who were compound heterozygotes for −909g>c/L55M (n = 89), −909g>c/Q192R (n = 77) and L55M/Q192R (n = 68). We observed a strong association between PON1 substrate specificity (paraoxon/phenylacetate substrate activity ratios) and −909g>c/Q192R haplotype. We have demonstrated here a powerful molecular haplotyping technology that can be applied in population studies.

INTRODUCTION

Haplotypes may be superior to genotypes for association studies. For example, when a promoter polymorphism affects transcription efficiency and a missense polymorphism in the same gene affects protein function, individuals heterozygous for these two polymorphisms have identical genotypes but may have different phenotypes (1). Currently, haplotypes are ascertained primarily by statistical inference using various algorithms including Expectation–Maximization (EM) [e.g. TagSNPs (2)] and Bayesian [e.g. PHASE (3)]. The development of technology that allows molecular determination of haplotypes is limited; most available methods are not suitable for population studies.

First, there are methods based on genotyping isolated chromosomes or clones, such as single-chromosome typing of individual sperm (4). More recent examples are cloning and genotyping fosmid/cosmid pools (5) and typing somatic hybrids (6). These methods are not easily applied to large population studies. A second method involves long-range PCR based on allele-specific amplification at one polymorphism (7), a method that recently has been combined with intramolecular ligation (8). Methods of this type are limited by the ability to carry out robust long-range PCR, especially in the context of allele specificity. A third proposed method would read haplotypes directly on single molecules by coincidence counting fluorescent oligonucleotides hybridized to polymorphic sites (9). This method would require specialized instrumentation and needs substantial probe development. Finally, there are methods based on single molecule PCR. The oldest approach was based on limiting dilution (10), a method still in use (11). Limiting dilution PCR requires a large number of reactions to obtain a single haplotype, both because of the dilution requirement and the inefficiency of PCR amplification from a single template molecule. A newer approach uses a gel to separate templates, permitting ‘polonies’ to be genotyped to determine haplotypes (12). This method is effective, but again requires specialized instrumentation. Another approach would be emulsion PCR (13,14). Emulsion PCR has been combined with beads and fluorescence activated cell sorter (FACS) measurements to permit genotyping and could theoretically be extended to haplotyping (15). In this study, we have developed LE-PCR, which combines linking PCR (16) and emulsion PCR to produce a haplotyping technology applicable to long DNA molecules that does not require any special instrumentation beyond real-time PCR.
METHODS

Study subjects
The study population is from an on-going study at the Mount Sinai Children’s Environmental Health Center to assess, prospectively, infant growth and neurodevelopment associated with pesticide exposure in urban New York City. The study protocol was approved by the Institutional Review Board. The study consists of pregnant women of multi-ethnic origin (Caucasian, African-American, and Hispanic of Caribbean origin) at 26–30 weeks of gestational age. A detailed description of the population has previously been published (17). Leukocyte DNA and plasma were isolated from blood as previously described (18).

Emulsion formation
The oil phase contained 4.5% Span 80 (#85548, Fluka), 0.4% Tween 80 (#8-0704, Sigma), 0.05% Triton X-100 (#T-9284, Sigma) made up to 100% with mineral oil (#M-3516, Sigma). The aqueous phase contained 1x Taq buffer (10 mM Tris–HCl, pH 8.0, 50 mM KCl), 300 µM each dNTP, 2.5 mM MgCl2, 50 µM Me2NCl, 1 µM each outside primer (o indicates outside primer), 0.1 µM each jumping primer, 100 mM/µl AmpliTaq Gold (ABI) and 1 ng/µl human genomic DNA. Emulsion primers were (* indicates reverse primer): 192/C0, AAAAAAAGGCC-TATTACAGTG-P/AAAAAAAAAGTGCTCAGGTCCCA-P; 192/C0, AAAAAACTGTGAGTGTTTTCTT-P/AAAAAAAAGTGCTCAGGTCCCA-P; 55/C0, AAAAAAGCCCC-ATATCAGTG-P; 55/C0, AAAAAAGCCCC-ATATCAGTG-P/AAAAAAAAAGTGCTCAGGTCCCA-P; 192/C0, AAAAAACTGTGAGTGTTTTCTT-P/AAAAAAAAGTGCTCAGGTCCCA-P; 55/C0, AAAAAAGCCCC-ATATCAGTG-P; 55/C0, AAAAAAGCCCC-ATATCAGTG-P/AAAAAAAAAGTGCTCAGGTCCCA-P.

PCR cleanup
Add 3 volumes of NX buffer (100 mM NaCl, 1% Triton X-100, 10 mM Tris–HCl, pH 7.5, 1 mM EDTA) to 5 volumes emulsion (oil plus aqueous phases). Vortex 20 s. Separate the phases in a microcentrifuge and remove most of the oil. Complete the cleanup with a Qiagen PCR purification kit and elute in 40 µl. The Qiagen PCR purification kit tolerates oil carryover.

Removal of unlinked amplicons
Wash 3 µl Dynabeads Myone streptavidin (Dynal Biotech) 3x in B&W buffer (10 mM Tris–HCl, pH 7.5, 1 mM EDTA, 2 M NaCl) and 1x in Taq buffer. Resuspend the beads in 40 µl eluate plus 4 µl 10x Taq buffer. Incubate at room temperature for 30 min, magnetize and save the supernatant.

Capping of runoff products
Each capping reaction used 5 µl of the supernatant in a 50 µl PCR reaction with 1x Taq buffer, 1.5 mM MgCl2, 200 µM each dNTP, 1 µM each capping oligonucleotide, 2.5 U Taq DNA polymerase (not hot start). Incubate at 55°C for 30 min. Capping oligonucleotides were: 192/C0, AAAAAAAGGCC-TATTACAGTG-P/AAAAAAAAAGTGCTCAGGTCCCA-P; 192/C0, AAAAAACTGTGAGTGTTTTCTT-P/AAAAAAAAGTGCTCAGGTCCCA-P; 55/C0, AAAAAAGCCCC-ATATCAGTG-P; 55/C0, AAAAAAGCCCC-ATATCAGTG-P/AAAAAAAAAGTGCTCAGGTCCCA-P.

Determination of haplotypes by ASPCR
Each allele-specific PCR (ASPCR) reaction used 2 µl of capped product in a 20 µl PCR reaction. All four possible reactions were assembled with one ASPCR primer per SNP. PCR was carried out in 1x Taq buffer, 1.5 mM MgCl2, 200 µM each dNTP, 1 µM each ASPCR primer, 2.5 U AmpliTaq Gold DNA polymerase. 2% glycerol, 1x BSA (NEB), and 1x SYBR green in a LightCycler with cycling 1 min 55°C, 1 min 72°C, 30 s 94°C after an initial incubation at 95°C for 9 min to activate the polymerase. Ct values were determined by the LightCycler second derivative algorithm. ASPCR primers were: 192T, CAAATACATCTCCCAGGATT; 192C, CAAATACATCTCCCAGGATT; 55/192T, GAAGAAGAACAGAAAGTACAGC; 55/192C, GAAGAAGAACAGAAAGTACAGC. Determine the average Ct values for the two possible pairs. For example, consider average Ct for 192T/−909C + 192C/−909G and for 192T/−909G + 192C/−909C. Subtract to obtain ΔCt. If the absolute value of ΔCt > 1 and the individual Ct values of 192T/−909C and 192C/−909G are both greater than or less than the individual Ct values of 192T/−909G and 192C/−909C, call the haplotype based on the lowest Ct values. If a haplotype cannot be called, repeat the entire LE-PCR experiment. This calculation assumes all ASPCR primers are equally efficient when amplifying PCR products made without using an emulsion, as was the case for our four examples. If not, appropriate corrections must be employed.

PON1 phenotypes
PON1 activities with paraoxon and phenylacetate substrates were determined spectrophotometrically (18,19).

RESULTS AND DISCUSSION
To illustrate the power of our molecular haplotyping technology, we chose the paraoxonase-1 (PON1) gene depicted in Figure 1 (20). PON1 is an enzyme with multiple activities, including detoxification of organophosphates. It is believed to be important in preventing neurotoxic damage, and has also been implicated in vascular disease (21,22) although the mechanism in the latter is uncertain (23). Common polymorphisms in the 5’ regulatory region affect transcription (24). Although the −108 locus is most important for determining transcription level, the polymorphisms at the −909 and −108 loci are in strong linkage disequilibrium (LD) (18). PON1 also contains two common missense polymorphisms, L55M and...
**Figure 1.** Linking emulsion-PCR with human PON1. The exons (orange) of the 27 kb human PON1 gene (black) are drawn to scale. The location of the biotin (Bio) on the linking primer is indicated on each of the PCR amplicons containing either the promoter polymorphism –909g>c (yellow), the missense polymorphism L55M in exon 3 (red and blue) or the missense polymorphism Q192R in exon 6 (green). The arrows indicate the derivation of minichromosomes in the emulsion: –909g>c/L55M (yellow and red), –909g>c/Q192R (yellow and green) and L55M/Q192R (blue and green).

Q192R (25); the former has been shown to affect PON1 stability (26), and the latter the relative rate of hydrolysis of certain organophosphates compared with phenylacetate (27). We chose to examine the –909 polymorphism because it gave us the largest distance to Q192R. As illustrated below, haplotype inference would be a good predictor of molecular haplotypes for –909g>c and L55M and for L55M and Q192R, but would be an imprecise predictor for the most important functional alleles, –909g>c and Q192R. We had previously conducted a study of the relationship between PON1 phenotypes and genotypes in 378 pregnant women from multi-ethnic backgrounds (18). Although single nucleotide polymorphisms (SNPs) in PON1 were significantly associated with enzymatic activity, the use of phenylacetate as the substrate resulted in a considerable overlap of the PON1 activity across genotypes. In this study, we applied the newly developed LE-PCR molecular haplotyping method to this population to demonstrate that the method is robust in a population study setting; haplotype information from this method can better clarify the genetic influence on PON1 expression.

Figure 1 illustrates the overall design of LE-PCR for determining haplotypes for three common polymorphisms, –909g>c, L55M and Q192R in human PON1, a gene of 27 kb including the promoter region. PCR amplicons are produced at the two polymorphic loci –909g>c (yellow amplicon) and L55M (red amplicon) from a single template residing in an aqueous droplet of an emulsion. With the use of linking primers, the amplicons are fused into a –909g>c/L55M minichromosome. The same procedure can form minichromosomes containing –909g>c/Q192R or L55M/Q192R. Consequently, LE-PCR can produce two minichromosomes conserving phase information of two polymorphic loci on the two corresponding autosomes, i.e. haplotypes.

Emulsification is a critical step. Oil-water emulsions were produced by naturally published methods (14). Emulsification involved vortexing one part aqueous phase and two parts oil phase (typical volume 150 µl) for 5 min. We examined the effect of this treatment on duplex DNA length and found that the majority of the DNA remained large enough to fall outside the separation range in a 1% agarose gel (data not shown). By limiting the human genomic DNA concentration to 1 ng/µl (~300 haploid genome equivalents/µl), the chance of more than one template being found in a 10 µm diameter droplet is <1%. This chance is even smaller in actual experiments as smaller droplets were routinely observed by light microscopy (data not shown). Haplotyping failures are characterized by the absence of difference between Ct values from ASPCR for the two possible haplotype pairs. We found that all such failures occurred in the emulsion PCR step and could not be rescued by repeating the purification and assay steps. At the beginning, haplotype failures occurred ~20–30% of the time, but after repeatedly forming the emulsions, the failure rate dropped to <5%.

Figure 2A illustrates linking primer design. Of the 42 nt in the 5′-biotinylated primers, the 32 nt at the 5′ end were complementary to one another. As illustrated in Figure 2B, the 26 nt at the 3′ end of each linking primer (red) that were complementary to the genomic template acted together with a corresponding outside primer (blue or green) to form an amplicon at each polymorphic site. The linkage between the polymorphic sites was only required before template denaturation. The 16 nt on the 5′ end of each linking primer contained sequences derived from the second amplicon. Figure 2C illustrates all of the LE-PCR products. The biotin-containing amplicons contained one strand extended from a linking primer and a complementary strand extended from an outside primer. These amplicons contained identical sequences at the biotin-containing ends. Self-priming led to the formation of the long minichromosomes which no longer contained biotins. For clarity, the sequences in the linking primers are depicted in red in the linked product. The unlinked biotinylated amplicons were removed on streptavidin-coated magnetic beads. Figure 2D illustrates how outside primer runoff products were capping with oligo(dT) by primer extension on 3′-phosphate-biotinylated nucleotides. The capping oligonucleotides were partially complementary to and overlapped the 3′ ends of the runoff products. The 3′ phosphate prevented the capping oligonucleotides from acting as primers. Thus, the runoff products are themselves extended using the capping oligonucleotide as a template. We found the capping step to be absolutely necessary to prevent randomization of alleles during subsequent ASPCR assays.

We determined haplotypes for the 89, 77 and 68 individuals in our total population of 378 women, in a previous study (17), who were compound heterozygotes at –909g>c/L55M, –909g>c/Q192R and L55M/Q192R, respectively. L55M/Q192R displayed the strongest LD, with only 1 of 68 compound heterozygotes with diploidy 55L/92Q and 55M/92R. –909g>c/L55M displayed strong but less complete LD, with only 9 of 89 compound heterozygotes with diploidy 90c/55L and –909g/55M. On the other hand, –909g>c/Q192R displayed much less LD, with 22 compound heterozygotes with diploidy 90c/192R and –909g/192Q versus 55 compound heterozygotes with diploidy 90c/192Q and –909g/192R.

Figure 3 illustrates the utility of this technology for determining haplotype–phenotype associations in our population study. We determined the PON1 activity in the plasma of subjects heterozygous for both –909g>c and Q192R using both phenylacetate and paraoxon as substrates. The results clearly demonstrate that individuals carrying higher expressing promoter (–909g) coupled with polymorphism with higher paraoxon-metabolizing activity (192R), labeled haplotype 1,
had higher paraoxon/phenylacetate activity ratios than individuals of haplotype 2 (P < 0.001 for both Mann–Whitney and t-tests).

In summary, we have developed a single-molecule haplotyping technology based on a combination of linking and emulsion PCR (LE-PCR). This technology is robust and feasible for application to population studies, as we have demonstrated by detailed analysis of the relationship between PON1 haplotypes and phenotypes.

ACKNOWLEDGEMENTS

We thank the National Institutes of Health for support of this work (NIEHS grants R21 ES011643, P01 ES09584 and P42 ES07384). Funding to pay the Open Access publication charges for this article was provided by National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES
