Long-range, high-throughput haplotype determination via haplotype-fusion PCR and ligation haplotyping

Daniel J. Turner*, Chris Tyler-Smith and Matthew E. Hurles

Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK

ABSTRACT

Ligation Haplotyping is a robust, novel method for experimental determination of haplotypes over long distances, which can be applied to assaying both sequence and structural variation. The simplicity and efficacy of the method for genotyping large chromosomal rearrangements and haplotyping SNPs over long distances make it a valuable and powerful addition to the methodological repertoire, which will be beneficial to studies of population genetics and evolution, disease association and inheritance, and genomic variation. We illustrate the versatility of the method both by genotyping a Yp paracentric inversion, found in ~60% of Northwest European males, that strongly influences the germline rate of infertility-causing XY translocations and by haplotyping two autosomal SNPs that lie 16.4 kb apart on chromosome 7, and which influence an individual's susceptibility to systemic lupus erythematosus.

INTRODUCTION

For a diploid organism, empirical haplotype data is difficult to obtain. There is increasing evidence that gene expression is regulated over long distances, and thus the phase of SNPs is of considerable functional relevance (4). Statistical methods of assigning haplotypes to genotypes, such as PHASE (1), are relied upon heavily, and are generally reliable over short ranges (2), with switch errors occurring infrequently (3), but these approaches are only highly reliable in regions of the genome where linkage disequilibrium (LD) is high. In many parts of the genome with lower LD, statistical methods would be inaccurate, and reliable haplotype status could only be obtained from experimental data (5). However, robust experimental methods for determining haplotypes over long distances are lacking.

Previously, we developed Haplotype Fusion PCR (HF-PCR), a method to juxtapose DNA sequences on the same single molecule, thus retaining phase information (6). Here we demonstrate a ligation haplotyping assay that can be applied to the condensed haplotypes produced by HF-PCR, which enables high-throughput long-range haplotyping of SNPs, as well as genotyping of chromosomal inversions. Both of these forms of genetic variation have hitherto proved refractory to high-throughput experimental analyses.

Previously published approaches to determining haplotypes have relied on allele-specific PCR (7), long range PCR (8) or a combination of both (9). None of these approaches is suitable for long-range determination of haplotypes: allele-specific PCR is prone to mispriming and longer amplicons carry an increased risk of jumping PCR artifacts, resulting in the creation of artificial recombinants by template switching errors. Additionally, the efficiency of long PCR decreases with increasing amplicon length (6).

In HF-PCR, a fusion PCR (10) is performed to bring the regions of interest into apposition (Figure 1A). The reaction is carried out in an emulsion, which divides a single PCR into millions of independent reactions (11,12). By controlling the amount of DNA template in an emulsion, we can ensure that paralogous sequence variants (PSVs) or SNPs from single molecules are fused together in the overwhelming majority of cases, resulting in a condensed haplotype. As with long PCR, successful HF-PCR requires the targeted nucleotides to be present on the same template strand, but whereas for long PCR, low molecular weight templates will increase the incidence of template switching, this is not possible in Ligation Haplotyping, being essentially a single molecule amplification technique. Furthermore, HF-PCR does not suffer from a decrease in amplification efficiency over long distances because in the first few rounds, amplification proceeds independently at the two loci, and it is only when these amplicons have reached a sufficiently high concentration that the fusion reaction occurs (6).

Published emulsification methods typically require the use of large volumes and are costly because of the quantity of reagents used, particularly polymerase (6,11,13).

*To whom correspondence should be addressed. Tel: +44 (0) 1223 494588; Fax: +44 (0) 1223 494919; Email: djt@sanger.ac.uk
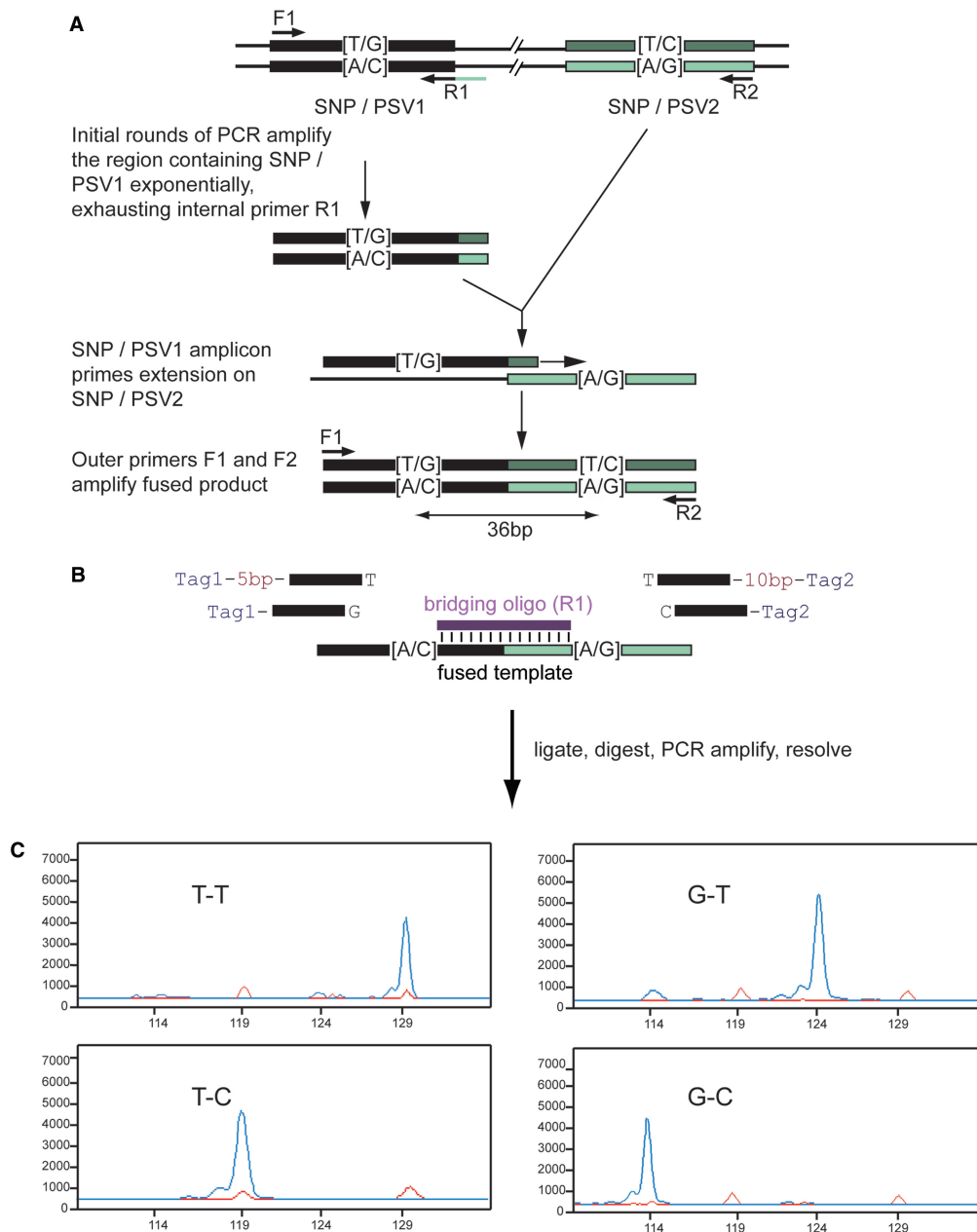
Figure 1. Ligation haplotyping. (**A**) Haplotype fusion PCR. The two nucleotides of interest lie several kilobases apart. Three primers, shown as black and black/green arrows, are used in a PCR reaction conducted in an oil : water emulsion. The internal primer R1 possesses a tail (green) that matches the sequence on the lower strand immediately to the left of SNP/PSV2. In the initial rounds of PCR, an amplicon containing SNP/PSV1 is amplified exponentially, whereas SNP/PSV2 is amplified linearly. Primer R1, initially present at a concentration $100\times$ lower than that of the outer primers, is used up. The top strand of the SNP/PSV1 amplicon is able to prime amplification on the SNP/PSV2 template and the fused molecule that is formed is amplified exponentially by the outer primers. (**B**) Ligation Haplotyping. A representation of the bottom strand of the HF-PCR product is shown, along with the oligonucleotides used in the Ligation Haplotyping reaction. Specific oligos of different lengths target each SNP and ligate to a common central 'bridging' oligo, generating a differently sized product for each of the four possible haplotypes. The specific oligos possess universal primer annealing sequences to facilitate simultaneous amplification of all ligation products. (**C**) Each haplotype produces a differently sized product during the ligation reaction, and following PCR using fluorescently labelled oligos, products are resolved by capillary electrophoresis.

For the work presented here, we developed a method of emulsification by vortexing that allows 96 reactions to be prepared simultaneously in 96-well plate format using a small emulsion volume. There are several advantages of this approach: the volume of polymerase required is correspondingly small (2 units per reaction as opposed to >35 units for other published methods) (11,13) and

additionally, 96 emulsions can be prepared in 150 s reducing both labour and the length of time between preparing emulsions and thermocycling, and increasing emulsion uniformity.

Following HF-PCR, we perform a novel ligation-based haplotyping assay, giving a robust and high-throughput way of identifying haplotypes (Figure 1B). In contrast

to other allele-specific ligation assays, that rely upon the ability of ligases to be highly discriminating at the 3′ nucleotide at a ligation junction (e.g. ligase detection reaction (14), GoldenGate assay (15), SNPlex (16)), our assay exploits the highly specific base pairing requirements of the thermostable DNA ligase from *Thermus thermophilus* at both 5′ and 3′ sides of the ligation junction.

Inversions are an important form of structural variation in the human genome and are likely to contribute to both simple and complex disorders and to human evolution (6). However molecular analysis of inversions has proved difficult, because they are frequently too small to be detected by cytogenetic analysis, are not accompanied by changes in copy number or chromosome size, and inversion breakpoints typically fall within long inverted repeats. Given that the majority of inversions result from non-allelic homologous recombination (NAHR), which occurs at rates of 10^{-3} to 10^{-5} per generation, many inversions are likely to occur recurrently (6), precluding ascertainment of inversion status by genotyping adjacent or internal markers. Individual inversions are commonly assayed by interphase FISH or Pulsed Field Gel Electrophoresis and on a genome-wide scale from dense read-pair data from diverse individuals (17). However, none of these techniques is easily applied to population-scale genotyping. A small number of well-characterized inversions can be genotyped singly by long PCR or RFLP analysis using standard agarose gel electrophoresis (18,19), but this is only a realistic approach when the homologous sequences that give rise to the inversion are short, whereas inversions in the human genome tend to involve homologous recombination between long inverted repeats (20,21).

As with inversion assays, empirical methods for identifying SNP haplotypes have lagged behind statistical approaches, and typically require the inclusion of other family members' genotype data, which is costly, time consuming and not always practical (5).

Here we illustrate the efficacy of Ligation Haplotyping by genotyping a Yp paracentric inversion found in ~60% of Northwest European males, that strongly influences the germline rate of infertility-causing XY translocations (22), and which results from homologous recombination between ~300 kb long inverted repeats. In addition, we demonstrate the versatility of the method by haplotyping two autosomal interferon regulatory factor 5 (*IRF5*) SNPs that lie 16.4 kb apart on chromosome 7, the haplotypes of which are associated with risk of developing systemic lupus erythematosus (SLE).

MATERIALS AND METHODS

DNA samples

HapMap DNA samples were obtained from the Centre d'Etude du Polymorphisme Human. Other DNA samples were obtained as described previously (6).

Emulsification

When prepared singly by stirring, emulsions contain aqueous compartments with an average diameter of 15 μm (23)

and contain 5.7×10^5 aqueous compartments per μl. Increasing the speed of stirring reduces this average diameter. The speed of vortexing has a similar influence on aqueous compartment diameter. Reducing the compartment diameter increases the number of compartments and thus decreases the opportunity for any given compartment to contain more than one template molecule, thus reducing the background signal. However, the greater the number of aqueous compartments, the lower the number that contain template and polymerase, so the efficiency of the reaction decreases. We found that an aqueous compartment diameter of between 5 and 10 μm gave the optimal compromise between low background and reaction efficiency (Supplementary Figure 1).

Preparation of emulsions singly

For the Yp inversion, we prepared PCR aqueous phases in a total volume of 100 μl as described previously (6). For *IRF5* SNPs, we included 1 μM primers IRF5F1 (CGGGATGAAGACTGGAGTAGG) and IRF5R2 (GACAAGGAGGAGTAAGCAAGGAAC), and 10 nM primer IRF5Fus4 (GGGTGCCTACAGCAGGGTTCTGACCCTGGCAGGTCC). All primers were synthesised by Sigma. Thermocycling conditions for IRF5 were 98°C for 30 s; 33 cycles of 98°C for 10 s, 63°C for 30 s and 72°C for 15 s.

Microtitre plate preparation of emulsions

We prepared PCRs as described above in a total volume of 25 μl and added each reaction to a different well of a 96-well plate along with 50 μl oil phase (40% DC5225C fluid (Corning), 30% DC749 fluid (Corning) and 30% AR20 silicone oil (Sigma) (13)) and a 3 mm tungsten carbide bead (Qiagen). The plate was sealed with a Microseal 'A' plate seal (Bio-Rad), inverted, centrifuged briefly and vortexed for 150 s, inverted, at speed 5 on a Vortex Genie 2, fitted with a microtitre plate adapter (both VWR). The plate was put the right way up and centrifuged extremely briefly before thermocycling as described above.

Ligation reactions and electrophoresis

We disrupted emulsions with diethyl ether (Aldrich) and incubated the recovered aqueous phases with 0.8 units Proteinase K (Sigma) for 56°C for 1 h to digest the polymerase and 95°C for 10 min to denature the Proteinase K, and made the volume up to 200 μl with water.

For IRF5 we phosphorylated 1 μm ligation primers IRF5-5T (IDT; TGAGCATCACCAATGTGACCTATCCTATCCGTGCCAGCAAGATCCAATCTAGA-2′UOMe-2′UOMe-2′UOMe-2′UOMe) and IRF5-5C (CGAGCATCACCAATGTGACCGTGCCAGCAAGATCCAATCTAGA-2′UOMe-2′UOMe-2′UOMe-2′UOMe) and 2 μM IRF5FusF4, and for Yp 2 μM primer InvFusion R1 and 3Way5T (Operon; TTACTGGGTGCTGGACATGCTCTGGTGCCAGCAAGATCCAATCTAGA-2′UOMe-2′UOMe-2′UOMe-2′UOMe 3′), and 0.5 μM 3Way5G (Operon; GTACTGGGTGCTGGACATGCTCTGCTGATGTCCGGTGCCAGCAAGATCCAATCTAGA-2′UOMe-2′UOMe-2′UOMe-2′UOMe 3′) at the 5′ terminus by incubating for 30 min at 37°C in 1 × T4 ligase buffer

(Roche; 66 mM Tris-HCl pH7.5, 5 mM MgCl$_2$, 5 mM DTT, 1 mM ATP) with 10 units of T4 polynucleotide kinase (Roche) in a total volume of 50 μl, after which we denatured the kinase by heating to 65°C for 20 min. For IRF5 we added 50 pmol of primers IRF5-3G (IDT; AmC6-GGGTTCCCTAAGGGGTTGGAGAAAGCGAGCTCG GGG) and IRF5-3T (IDT; AmC6-GGGTTCCCTAAG GGTTGGATTTCGGAAAGCGAGCTCGGGT), and for Yp 50 pmol of primers 3Way3A (Operon; AmC6-dT GGGTTCCCTAAGGGGTTGGACAGTGATTTCTGG GTAGCTACAATCATA 3′) and 3Way3G (Operon; Am C6-dT GGGTTCCCTAAGGGGTTGGATACTTCAGTG ATTTCTGGGTAGCTACAATCATG 3′), along with water to a total volume of 100 μl.

The optimal concentration of oligos in the ligation reaction was determined empirically, so as to provide peaks after capillary electrophoresis with similar heights: the concentration of central bridging oligo was kept constant, and concentrations of all allele-specific oligos were titrated until uniform results were obtained.

We set ligation reactions up in a total volume of 20 μl, containing 1 × *Tth* ligase buffer (ABGene; 20 mM Tris-HCl pH8.3, 50 mM KCl, 10 mM MgCl$_2$, 1 mM EDTA, 1 mM NAD$^+$, 10 mM DTT, 0.1% (v/v) Triton X-100), 2 μl of diluted fusion PCR product, 0.5 μl primer mix and 20 units *Tth* ligase (ABGene), and mixed these thoroughly before thermocycling (95°C for 2 min; 20 cycles of 95°C for 30 s, 64°C for 4 min; 4°C indefinitely). To digest the ligase, we added 1.6 units of Proteinase K (Sigma) to each reaction and incubated them as described above. Subsequently, we added 5 μl of Lambda exonuclease buffer (NEB; 67 mM glycine-KOH pH9.4, 2.5 mM MgCl$_2$, 50 μg ml^{-1} BSA) and 22.75 μl H$_2$O, 1 unit Lambda exonuclease (NEB) and 0.5 units *E. coli* exo I (USB), to each reaction and incubated them at 37°C for an hour, to digest unligated primers, followed by 65°C for 20 min to denature the exonucleases.

To amplify ligated oligos and to incorporate a fluorescent label we performed a final PCR reaction. As with the HF-PCR, we used Phusion DNA polymerase to avoid the addition of adenosine to the 3′ end of PCR products, to give single peaks after capillary electrophoresis. We carried out reactions in 1 × Phusion HF buffer, 200 μM dNTPs, 300 μM PCR primers MLPAF (Sigma; 5′ 6FAM GGGT TCCCTAAGGGGTTGGA 3′) and MLPAR (Sigma; 5′ GT GCCAGCAAGATCCAATCTAGA 3′), 2.5 μl LDR product and 1 unit Phusion DNA polymerase in a total volume of 50 μl, using the same thermal cycling conditions as above but with a 63°C annealing temperature.

To visualise the fluorescent PCR products we ran 1 μl of product on an ABI 3100 capillary electrophoresis sequencer along with 8 μl formamide (Fluka) and 1 μl GeneScan ROX 350 size standard (Applied Biosystems).

RESULTS AND DISCUSSION

In the Ligation Haplotyping reaction, allele-specific oligos for each SNP/PSV are brought together and are ligated simultaneously to a central bridging oligo. The assay is designed so that this bridging oligo is the same oligo that was used as the fusion primer in the prior reaction. Several

rounds of denaturation and annealing/ligation result in linear amplification of the product: the product does not serve as a template in subsequent rounds of ligation, which helps to maintain specificity (14). For each SNP/PSV the specific oligos differ in length, giving the ligation products a characteristic size depending upon the haplotypes present. Ligation primers possess universal sequences allowing PCR amplification of all ligation products simultaneously using fluorescently labelled primers, and PCR products are subsequently resolved by capillary electrophoresis (Figure 1C).

Yp inversion

We previously fine-mapped the breakpoint interval and identified flanking PSVs using four individual YACs containing the inversion breakpoints before and after the inversion (6) (Figure 2A). We performed Ligation Haplotyping on individual and pairwise mixtures of these YACs (data not shown), on genomic DNA from the same two individuals from which the YACs derive (Figure 2B) and demonstrated that Ligation Haplotyping identified only a single haplotype from each YAC, which corresponded precisely to the known PSV content in each YAC, and identified the correct pair of haplotypes in the individuals from whom the YACs were derived.

We then performed the assay on a panel of genomic DNA from thirteen individuals of diverse geographic ancestry, obtained from ECACC, to which the experimenter was blinded as to the inversion status of the individuals. In twelve individuals of this geographically diverse panel (92%) it was possible to assign either 'inverted' or 'reference' status as both breakpoint haplotypes were diagnostic of either orientation. In one individual, a haplotype was obtained that was a mixture of two breakpoint haplotypes (Figure 2B), one characteristic of the reference orientation and the other of the inverted orientation, presumably resulting from a gene conversion event that homogenises PSV2. Results were confirmed by bead haplotyping (6), showing that Ligation Haplotyping is capable of unambiguous inversion genotyping by assaying both inversion breakpoints: gene conversion produces aberrant pairs of breakpoint haplotypes, which are easily distinguishable from either of the inversion states.

The ability to sequence clones spanning inversion breakpoints in their entirety should enable the design of many such assays (17), allowing population-scale genotyping assays to be developed, which will facilitate our understanding of the impact of chromosomal inversions on complex traits (20).

IRF5 SNPs

The T allele of SNP rs2004640 creates a splice site in exon 1 of *IRF5* that is absent from the G allele, permitting expression of several unique IRF5 isoforms. The second SNP, rs2280714 is *cis*-acting, and the T allele of which is associated with elevated expression (Figure 3A). Individuals who possess T alleles at both SNP positions on the same chromosome have an elevated risk of developing SLE (24,25). We compared the results obtained by Ligation Haplotyping the 30 Centre d'Etude du
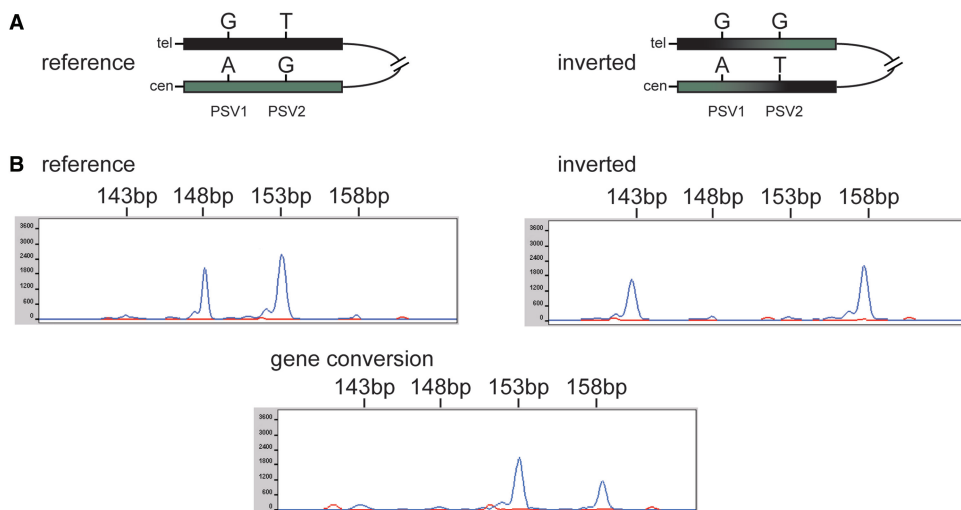
A

reference

inverted

B reference

inverted

gene conversion

Figure 2. Genotyping inversions using Ligation Haplotyping. (**A**) Homologous recombination of the ~300 kb long inverted Yp repeats results in inversion of the intervening DNA. The combination of PSVs at either side of the breakpoints, is diagnostic of 'reference' and 'inverted' status. (**B**) Ligation haplotyping on individuals produces two peaks, one for each breakpoint. In one individual a haplotype was obtained that was a mixture of 'inverted' and 'reference' breakpoint haplotypes, identifying a gene conversion event that has homogenised PSV2.
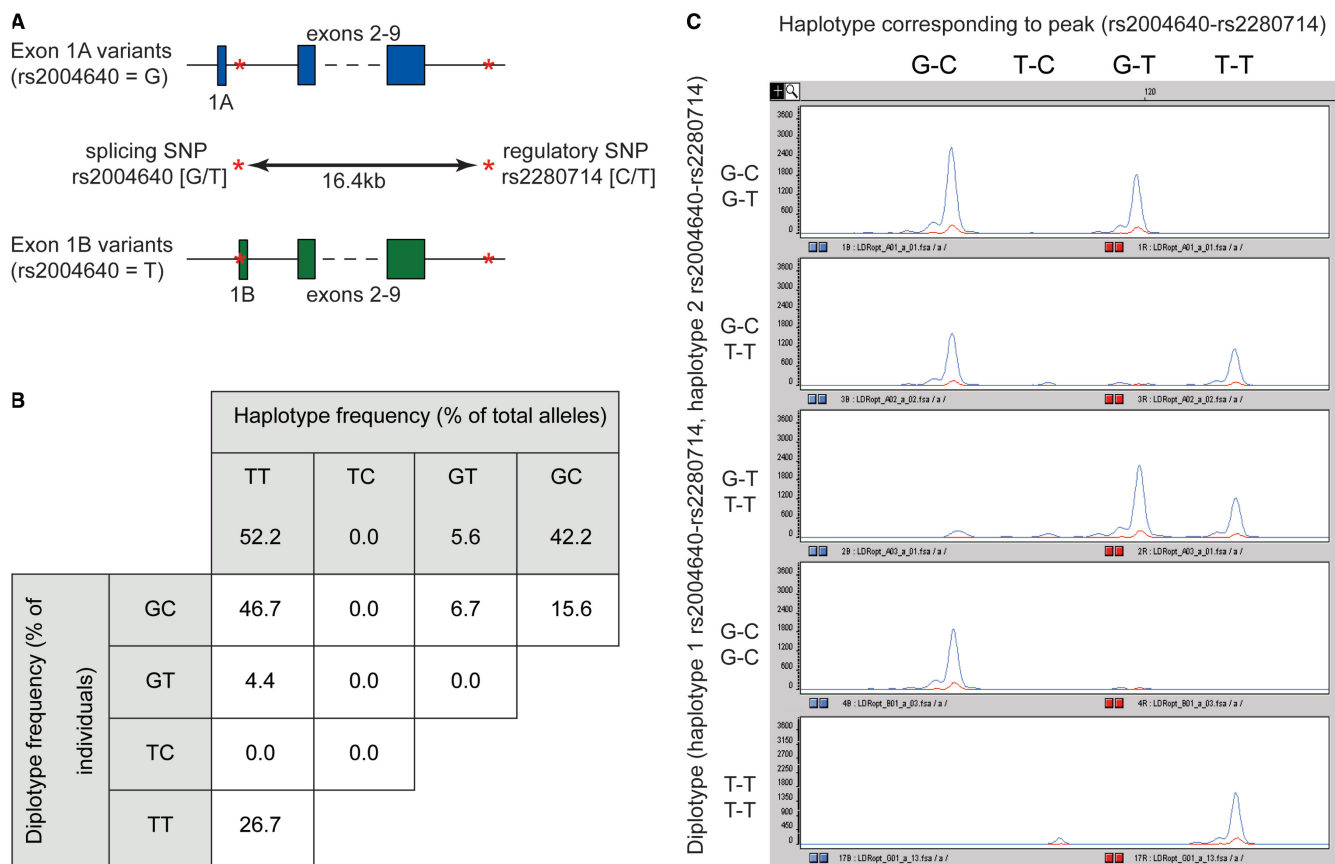
A

Exon 1A variants
(rs2004640 = G)

exons 2-9

1A

splicing SNP
rs2004640 [G/T]

16.4kb

regulatory SNP
rs2280714 [C/T]

Exon 1B variants
(rs2004640 = T)

1B   exons 2-9

B

	Haplotype frequency (% of total alleles)			
		TT	TC	GT	GC
		52.2	0.0	5.6	42.2
Diplotype frequency (% of individuals)	GC	46.7	0.0	6.7	15.6
	GT	4.4	0.0	0.0	
	TC	0.0	0.0		
	TT	26.7			

C

Haplotype corresponding to peak (rs2004640-rs2280714)

G-C   T-C   G-T   T-T

G-C
G-T

G-C
T-T

G-T
T-T

G-C
G-C

T-T
T-T

Diplotype (haplotype 1 rs2004640-rs2280714, haplotype 2 rs2004640-rs2280714)

Figure 3. Ligation haplotyping on genomic DNA. (**A**) The rs2004640 G allele results in IRF5 isoforms initiated at exon 1A, whereas the T allele results in exon 1B splice variants. The second SNP, rs2280714 is *cis*- acting, and the rs2280714 T allele is associated with elevated expression. Individuals with the T-T haplotype are at elevated risk of developing systemic lupus erythematosus (Figure adapted from (24,25)). (**B**) Observed diplotype frequencies obey the Hardy–Weinberg equilibrium. The TC allele was not found in any of the individuals used in the study. From the haplotype frequency (5.6%), the GTGT diplotype has a predicted frequency of 0.3% and was not found in these individuals. (**C**) Observed diplotypes produced by Ligation Haplotyping of genomic DNA from 90 CEPH individuals are shown. Being on Chromosome 7, individuals have two haplotypes, which for heterozygotes will result in two peaks following capillary electrophoresis.

Polymorphisme Human (CEPH) HapMap trios with haplotypes derived statistically (24,25) and obtained 100% concordance (Supplementary Table 1). Inheritance of the haplotypes was observed to be Mendelian in each case, with observed diplotype frequencies obeying the Hardy–Weinberg equilibrium (Figure 3B). Not all haplotypes were observed in the HapMap individuals: the GT haplotype occurs at a low frequency (5.6%) and no GTGT homozygotes were detected. Furthermore, no TC haplotypes were observed (Figure 3B and C).

In principle, the length of a haplotype that can be condensed is limited only by the size of the template DNA used in the emulsion. Emulsion preparation appears to have little shearing effect on genomic DNA 100–150 kb in length (6) and so HF-PCR and Ligation Haplotyping should be possible over a considerably longer range than long PCR.

Because HF-PCR only occurs in a small fraction of the aqueous compartments within an emulsion (<0.1%) this part of this assay should be readily amenable to multiplexing, using different sets of primers that do not compete for resources due to amplification proceeding in separate compartments. Additionally, as with the ligation detection reaction, SNPlex and GoldenGate assays, the Ligation Haplotyping part should also be compatible with high levels of multiplexing. The assay could easily be adapted for use on addressable arrays or with combination ultra high-throughput sequencing technologies, which will increase throughput and decrease cost further.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Stephens,M. and Donnelly,P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.*, **73**, 1162–1169.
2. Marchini,J., Cutler,D., Patterson,N., Stephens,M., Eskin,E., Halperin,E., Lin,S., Qin,Z.S., Munro,H.M. *et al.* (2006) A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.*, **78**, 437–450.
3. International HapMap Consortium. (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
4. Kleinjan,D.A. and van Heyningen,V. (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.*, **76**, 8–32.
5. Lin,S., Cutler,D.J., Zwick,M.E. and Chakravarti,A. (2002) Haplotype inference in random population samples. *Am. J. Hum. Genet.*, **71**, 1129–1137.
6. Turner,D.J., Shendure,J., Porreca,G., Church,G., Green,P., Tyler-Smith,C. and Hurles,M.E. (2006) Assaying chromosomal inversions by single-molecule haplotyping. *Nat. Methods*, **3**, 439–445.
7. Lo,Y.M., Patel,P., Newton,C.R., Markham,A.F., Fleming,K.A. and Wainscoat,J.S. (1991) Direct haplotype determination by double ARMS: specificity, sensitivity and genetic applications. *Nucleic Acids Res.*, **19**, 3561–3567.
8. McDonald,O.G., Krynetski,E.Y. and Evans,W.E. (2002) Molecular haplotyping of genomic DNA for multiple single-nucleotide polymorphisms located kilobases apart using long-range polymerase chain reaction and intramolecular ligation. *Pharmacogenetics*, **12**, 93–99.
9. Michalatos-Beloin,S., Tishkoff,S.A., Bentley,K.L., Kidd,K.K. and Ruano,G. (1996) Molecular haplotyping of genetic markers 10 kb apart by allele-specific long-range PCR. *Nucleic Acids Res.*, **24**, 4841–4843.
10. Yon,J. and Fried,M. (1989) Precise gene fusion by PCR. *Nucleic Acids Res.*, **17**, 4895.
11. Dressman,D., Yan,H., Traverso,G., Kinzler,K.W. and Vogelstein,B. (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl Acad. Sci. USA*, **100**, 8817–8822.
12. Tawfik,D.S. and Griffiths,A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.*, **16**, 652–656.
13. Margulies,M., Egholm,M., Altman,W.E., Attiya,S., Bader,J.S., Bemben,L.A., Berka,J., Braverman,M.S., Chen,Y.J. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.
14. Barany,F. (1991) Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proc. Natl Acad. Sci. USA*, **88**, 189–193.
15. Fan,J.B., Oliphant,A., Shen,R., Kermani,B.G., Garcia,F., Gunderson,K.L., Hansen,M., Steemers,F., Butler,S.L. *et al.* (2003) Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.*, **68**, 69–78.
16. Tobler,A.R., Short,S., Andersen,M.R., Paner,T.M., Briggs,J.C., Lambert,S.M., Wu,P.P., Wang,Y., Spoonde,A.Y. *et al.* (2005) The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *J. Biomol. Tech.*, **16**, 398–406.
17. Tuzun,E., Sharp,A.J., Bailey,J.A., Kaul,R., Morrison,V.A., Pertz,L.M., Haugen,E., Hayden,H., Albertson,D. *et al.* (2005) Fine-scale structural variation of the human genome. *Nat. Genet.*, **37**, 727–732.
18. Liu,Q., Nozari,G. and Sommer,S.S. (1998) Single-tube polymerase chain reaction for rapid diagnosis of the inversion hotspot of mutation in hemophilia A. *Blood*, **92**, 1458–1459.
19. Small,K., Iber,J. and Warren,S.T. (1997) Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nat. Genet.*, **16**, 96–99.
20. Stefansson,H., Helgason,A., Thorleifsson,G., Steinthorsdottir,V., Masson,G., Barnard,J., Baker,A., Jonasdottir,A., Ingason,A. *et al.* (2005) A common inversion under selection in Europeans. *Nat. Genet.*, **37**, 129–137.
21. Stankiewicz,P. and Lupski,J.R. (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet.*, **18**, 74–82.
22. Jobling,M.A., Williams,G.A., Schiebel,G.A., Pandya,G.A., McElreavey,G.A., Salas,G.A., Rappold,G.A., Affara,N.A. and Tyler-Smith,C. (1998) A selective difference between human Y-chromosomal DNA haplotypes. *Curr. Biol.*, **8**, 1391–1394.
23. Ghadessy,F.J., Ong,J.L. and Holliger,P. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl Acad. Sci. USA*, **98**, 4552–4557.
24. Graham,R.R., Kozyrev,S.V., Baechler,E.C., Reddy,M.V., Plenge,R.M., Bauer,J.W., Ortmann,W.A., Koeuth,T., Gonzalez Escribano,M.F. *et al.* (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.*, **38**, 550–555.
25. Sigurdsson,S., Nordmark,G., Goring,H.H., Lindroos,K., Wiman,A.C., Sturfelt,G., Jonsen,A., Rantapaa-Dahlqvist,S., Moller,B. *et al.* (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am. J. Hum. Genet.*, **76**, 528–537.