Cloning the Differences Between Two Complex Genomes
Author(s): Nikolai Lisitsyn, Natalya Lisitsyn, Michael Wigler
Source: *Science*, New Series, Vol. 259, No. 5097 (Feb. 12, 1993), pp. 946-951
Published by: American Association for the Advancement of Science
Stable URL: http://www.jstor.org/stable/2880614
Accessed: 08/07/2010 23:05

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=aaas.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

American Association for the Advancement of Science is collaborating with JSTOR to digitize, preserve and extend access to *Science.*

http://www.jstor.org

Cloning the Differences Between Two Complex Genomes

Nikolai Lisitsyn, Natalya Lisitsyn, Michael Wigler

The analysis of the differences between two complex genomes holds promise for the discovery of infectious agents and probes useful for genetic studies. A system was developed in which subtractive and kinetic enrichment was used to purify restriction endonuclease fragments present in one population of DNA fragments but not in another. Application of this method to DNA populations of reduced complexity ("representations") resulted in the isolation of probes to viral genomes present as single copies in human DNA, and probes that detect polymorphisms between two individuals. In principle, this system, called representational difference analysis (RDA), may also be used for isolating probes linked to sites of genomic rearrangements, whether occurring spontaneously and resulting in genetic disorders or cancer, or programmed during differentiation and development.

Genetic alterations underlie various biological processes. Programmed gene rearrangements occur in many contexts, including the generation of diversity in the immune system (*1*), mating type switching in yeasts (*2*), and antigenic variation in microbial organisms (*3*). Spontaneous losses or rearrangements of genetic material in somatic cells occur frequently during the development of cancers and leukemias. Similar events occurring in the germ line can result in inherited disorders. Even infectious disease can be viewed as the alteration of the genetic content of the infected organism. Various time-consuming methods have been applied to determining the nature of the genetic changes that occur in the above situations. A single method for defining the differences between two DNA populations could, in principle, be applied to each of the above problems and lead to the discovery of the genetic basis of many types of biological phenomena. We present here a general method (representational difference analysis, or RDA) for finding small differences between the sequences of two DNA populations. The method builds upon subtractive hybridization techniques that have been used in the past to find probes for large sequence differences between two genomes.

In 1984 Lamar and Palmer applied a subtractive hybridization technique to clone probes for the Y chromosome (*4*). They used an excess of sheared DNA from a female to drive hybridization of Sau 3A–cleaved DNA from a male. The Y chromosome–specific DNA was free to self-anneal and was subsequently cloned after ligation into an acceptor site of a plasmid. A similar method was applied by Kunkel *et al.* (*5*) and then by Nussbaum *et al.* (*6*) to clone probes for the Duchenne muscular dystrophy and the choroidermia loci, respectively. In both cases, large deletions on the X chromosome made it possible to use subtractive hybridization techniques with DNA from affected males as "driver" and DNA from normals as "tester" (*7*). In general, however, subtractive hybridization techniques do not achieve sufficient enrichment of the sequences that occur only in tester (the "target") partly because of the high complexity of the human genome, which prevents effectively complete hybridization (*8*). Even when subtractive steps are reiterated, "target" sequences are enriched only 100 to 1000 times (*9*). This enrichment is insufficient for more common situations in which the magnitude of enrichment required is 10^6.

A second means for DNA difference enrichment has been proposed on the basis of the second-order kinetics of self-reassociation (*9*). In theory, this method can be applied after an initial enrichment of target sequences has been achieved by subtraction. If a population of DNA fragments containing a target subpopulation enriched n times relative to unenriched fragments in tester is melted and reannealed so that only a small proportion of double-stranded tester DNA forms, double-stranded target DNA would be present n^2 times relative to the other sequences present as duplex DNA (*10*). We call this "kinetic" enrichment. To apply this method, it is necessary to use some kind of DNA amplification to purify the small quantities of double-stranded DNA that form away from all driver and all unreannealed tester sequences. We achieved this by ligating oligonucleotide adaptors to tester so that only double-stranded tester molecules were amplified during the polymerase chain reaction (PCR) (*11*). Subsequent reiterations of the method led to exponentially increasing enrichments of target.

In RDA, we lowered the DNA complexity of both tester and driver genomes by preparing a representative portion of each genome (a "representation"). The difference analysis of the two representations is based on simultaneous combination of subtractive and kinetic steps. The lowered complexity of the representations allowed us to achieve greater completeness during subtractive enrichment and, hence, a more effective kinetic enrichment. We now demonstrate the cloning of probes for single copy sequences present in (or absent from) one of the two genomes, and the cloning of probes for binary polymorphisms between two individuals. We discuss the applications of this approach to the discovery of probes for pathogenic organisms, for otherwise anonymous loci that have suffered genetic rearrangements, and for polymorphisms located near the genes affected by inherited disorders.

Subtractive and kinetic enrichments of PCR amplicons. We began by making representations of DNA populations that we call "amplicons"; DNA cleaved with relatively infrequent cutting restriction endonucleases was ligated to oligonucleotide adaptors, and amplified by PCR. The result was similar to a size fractionation, since after 20 rounds of amplification only low molecular size fragments, below 1 kb, were effectively amplified. The advantage of this method for representation compared to a physical fractionation is that only small amounts of starting material (less than 10 μg) are required. Only a subset of the whole genome was represented. However, when different restriction endonucleases were chosen, sets of amplicons that scan the genome could be made. In the following examples, we made amplicons of mammalian DNA after it underwent cleavage with Bam HI, Bgl II, and Hind III (*12*). We estimate the complexity of the resulting amplicons to be 55 times, 13 times, and 8 times less than the complexity of the starting genomic DNA, respectively (*13*).

Once tester and driver amplicons were made, their adaptors were removed by cleavage, and only tester fragments were ligated to new adaptors at their 5′ ends. We then combined subtractive and kinetic steps into a single operation, the hybridization-amplification step (Fig. 1) (*14*). We abandoned all use of physical separation techniques, such as biotinylation with subsequent biotin-avidin chromatography (*9, 15*) or hydroxylapatite (*9*), to enrich for double-stranded tester because these methods are not completely reliable. Instead, DNA amplification was used for selective enrichment of double-stranded tester. After melting and reannealing in the presence of excess of driver amplicon, DNA molecules were treated with Taq DNA polymerase at

The authors are at Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724.

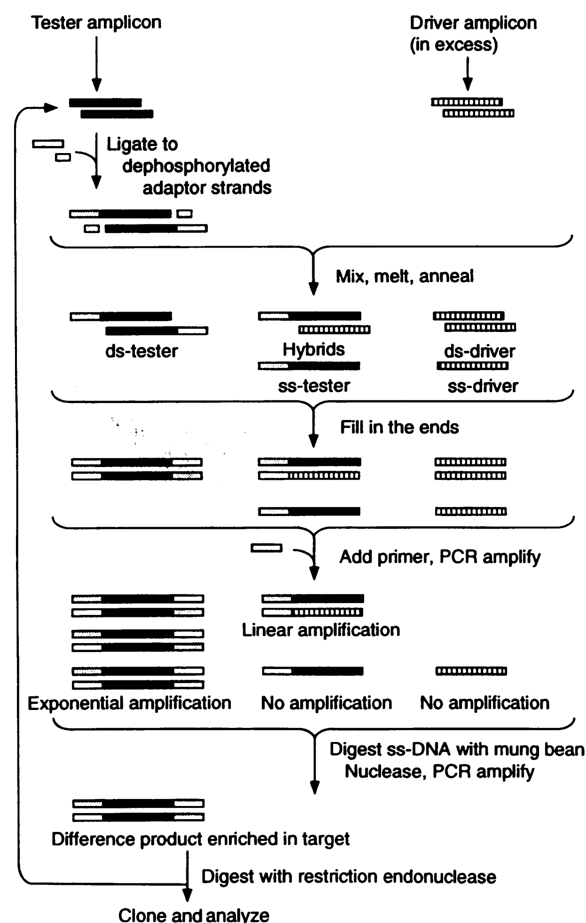
elevated temperatures in the presence of all four deoxynucleotide triphosphates. Only self-reannealed tester molecules had 5′ adaptors at each end of the duplex DNA and thus could be filled in at both 3′ ends. Therefore only self-reannealed tester could subsequently be amplified by PCR at an exponential rate. Driver DNA acted as a competitive inhibitor for the self-reannealing of those tester DNA fragments common to driver. Target DNA, which occurs only in tester, was thus enriched relative to other tester DNA after amplification. The PCR products were then either cloned, otherwise analyzed, or further enriched. In the last case, the PCR products were cleaved, and yet new PCR adaptors were added to their 5′ ends (*16*). This material was then diluted (*17*) with fresh driver amplicon and the hybridization-amplification step was repeated.

The first round of RDA was mainly subtractive. Subsequent rounds had a greatly increased kinetic component equal to the enrichment achieved on the previous step. For example, if target is equimolar with respect to tester (that is, single copy), and if driver amplicon is taken in n times excess to tester amplicon, target may be enriched e times after the first round, where $e < n$ (*8*). After the second round, target would be enriched yet another e times because of repetition of the subtractive component, multiplied by another e times because of the kinetic component. After the third round, total enrichment would be at least the square of that. If e is 50, at the end of the second round, target would be enriched by about 10^5, and at the end of the third round by more than 10^{10}.

Acquisition (or loss) of new sequences. In the first set of stringent tests of our protocols, we added, as single copies, adenovirus or bacteriophage λ DNA (or both) to a human DNA to create a model tester, and used the same DNA without viral DNA as driver. The viral DNA's were the target and we can view this model as testing the efficacy of the procedure in either of two situations: the acquisition of a pathogenic genome at a single copy of genomic DNA per infected cell; or the homozygous loss of DNA by spontaneous mutation. We prepared Bgl II amplicons from human DNA with adenovirus and λ DNA's as targets (Fig. 2A) or Hind III amplicons with λ DNA as target (Fig. 2B). If Bgl II amplicons were taken, small λ and adenovirus fragments were the major difference products even after two rounds (Fig. 2A, lane a). This represents an enrichment of more than 5×10^6 times from the starting material, and a probable enrichment of about 4×10^5 times from amplicons. There is a strong bias during amplification against fragments larger than 1.0 kbp, so that the larger Bgl II fragments of target (Fig. 2A, lanes c and d) are not enriched.

The enrichment from Hind III amplicons was not as effective. The λ Hind III fragment was greatly enriched after the third round (Fig. 2B, lane b), but other sequences were still present. Nevertheless, after the fourth round the expected target fragment was purified to near homogeneity (Fig. 2B, lane c). We attribute this outcome to the greater sequence complexity of the Hind III amplicon. In fact, we found that without some simplification of driver complexity, our method failed. Presumably, when the complexity of the driver was too high, subtractive and kinetic enrichments were diminished and competing processes dominated.

Probes for polymorphic loci. In the experiments described above, the driver and tester were identical except for the



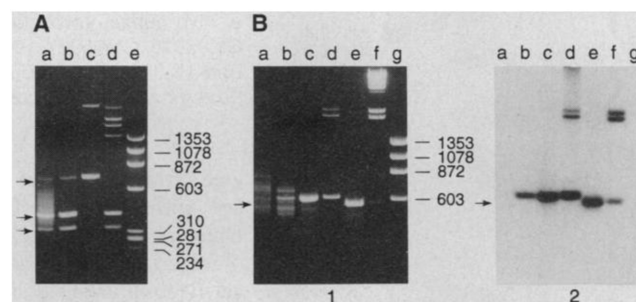


Fig. 1 (left). Schematic protocol for representational difference analysis, illustrating the hybridization and amplification steps after the preparation of amplicons. The representation stage is not illustrated. The details are described in (*12*). **Fig. 2 (above).** Representational difference analysis with viral DNA's added as targets. High molecular weight DNA (10 μg) purified from the lymphoid cell line DRL 484 was used for preparation of driver amplicons and 10 μg of the same DNA, containing equimolar amounts of target [120 pg of adenovirus-2 DNA or 160 pg of λ phage DNA (or both), both from New England Biolabs] was taken for preparation of tester amplicons (*12*). Difference products were prepared as described in (*14*). (**A**) Agarose gel electrophoresis of difference products of Bgl II amplicons, obtained after the second (lane a) and third (lane b) hybridization-amplification steps. λ phage (lane c) and adenovirus (lane d) amplicons were prepared as described (*12*) taking 1 to 5 ng of purified viral DNA ligated to adaptors (primer set 1; see Table 1). Hae III digest of φX174 RF DNA is in lane e. Sizes (bp) are indicated to the right. Arrows on the left indicate major difference products. (**B**) Agarose gel electrophoresis (B, 1) and Southern blot (B, 2) of difference products of Hind III amplicons. Difference products (0.5 μg) obtained after two, three, and four hybridization-amplification steps are shown in lanes a to c. The λ phage Hind III amplicons were prepared as described (*12*) taking 1 to 5 ng of purified viral DNA ligated to adaptor set 1 (lane d). The difference product after four hybridization-amplification steps was digested with Hind III (lane e). Hind III–digested λ phage DNA (lane f) and Hae III–digested φX174 DNA (lane g, sizes indicated to the right) are as shown. These DNA's were separated on a 2 percent agarose gel, transferred at reduced pressure with the LKB 2016 VacuGene Vacuum Blotting System (Pharmacia LKB Biotechnology) to GeneScreen Plus membranes (Du Pont) and hybridized to ^{32}P-labeled λ phage Hind III amplicon under conditions recommended by the membrane supplier. Arrow on left indicates the target λ fragment. The slight upward mobility shift of this fragment in lanes b, c, and d relative to lanes e and f is due to the presence of adaptors.

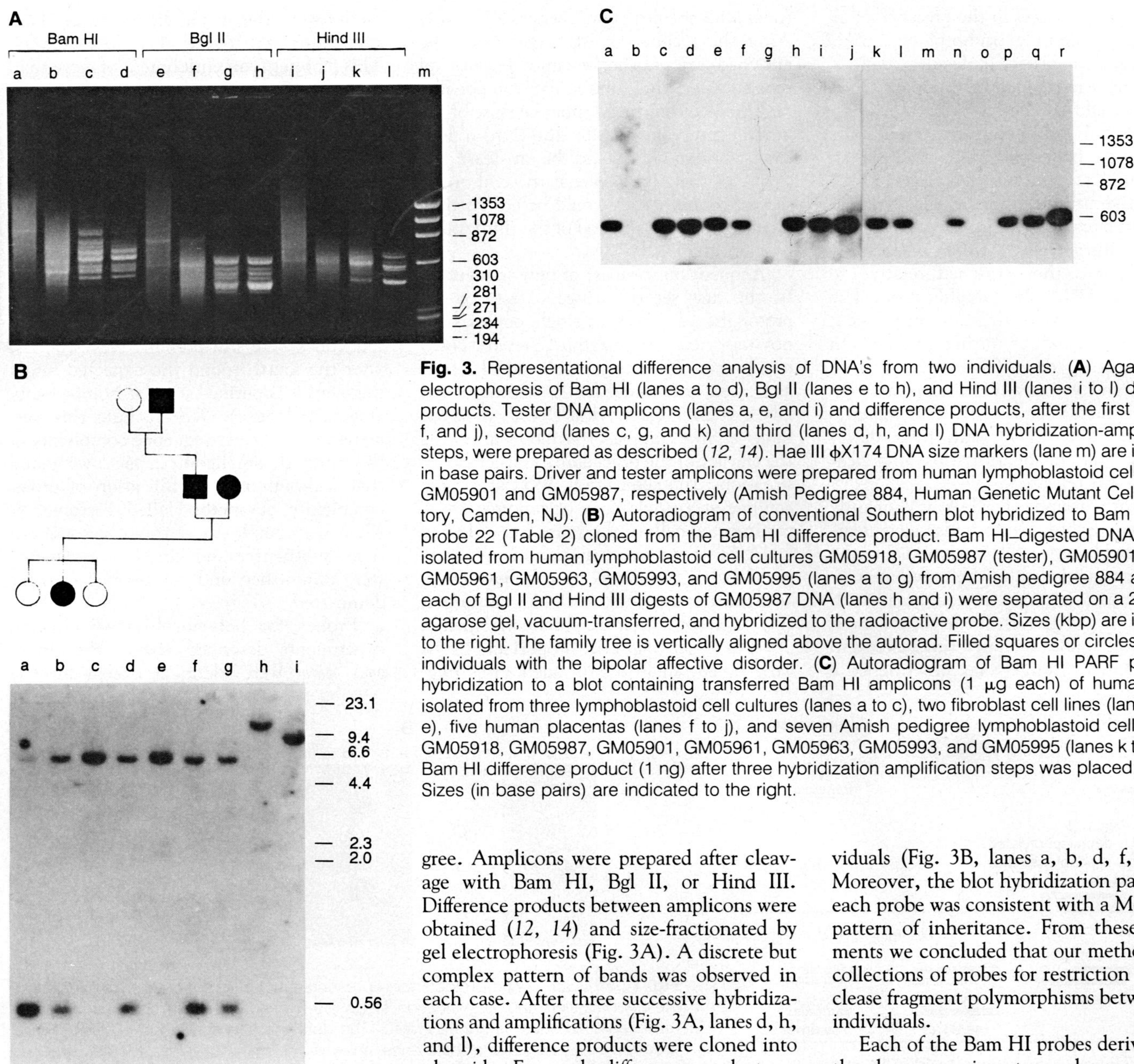


Fig. 3. Representational difference analysis of DNA's from two individuals. (**A**) Agarose gel electrophoresis of Bam HI (lanes a to d), Bgl II (lanes e to h), and Hind III (lanes i to l) difference products. Tester DNA amplicons (lanes a, e, and i) and difference products, after the first (lanes b, f, and j), second (lanes c, g, and k) and third (lanes d, h, and l) DNA hybridization-amplification steps, were prepared as described (*12, 14*). Hae III ϕX174 DNA size markers (lane m) are indicated in base pairs. Driver and tester amplicons were prepared from human lymphoblastoid cell cultures GM05901 and GM05987, respectively (Amish Pedigree 884, Human Genetic Mutant Cell Repository, Camden, NJ). (**B**) Autoradiogram of conventional Southern blot hybridized to Bam HI PARF probe 22 (Table 2) cloned from the Bam HI difference product. Bam HI–digested DNA's (2 μg) isolated from human lymphoblastoid cell cultures GM05918, GM05987 (tester), GM05901 (driver), GM05961, GM05963, GM05993, and GM05995 (lanes a to g) from Amish pedigree 884 and 2 μg each of Bgl II and Hind III digests of GM05987 DNA (lanes h and i) were separated on a 2 percent agarose gel, vacuum-transferred, and hybridized to the radioactive probe. Sizes (kbp) are indicated to the right. The family tree is vertically aligned above the autorad. Filled squares or circles indicate individuals with the bipolar affective disorder. (**C**) Autoradiogram of Bam HI PARF probe 22 hybridization to a blot containing transferred Bam HI amplicons (1 μg each) of human DNA's isolated from three lymphoblastoid cell cultures (lanes a to c), two fibroblast cell lines (lanes d and e), five human placentas (lanes f to j), and seven Amish pedigree lymphoblastoid cell cultures GM05918, GM05987, GM05901, GM05961, GM05963, GM05993, and GM05995 (lanes k to q). The Bam HI difference product (1 ng) after three hybridization amplification steps was placed in lane r. Sizes (in base pairs) are indicated to the right.

experimental perturbation introduced by the presence of exogenous DNA in tester. In order to assess the method for potential applications, we compared tester and driver amplicons from different individuals and expected to obtain a subset of restriction endonuclease fragments that were polymorphic between them. For example, if one of the two Bam HI sites flanking a short Bam HI fragment in tester was absent in both alleles from driver, leading to only large and poorly amplified Bam HI fragments in driver, the short Bam HI fragment of tester would be present in its Bam HI amplicon but absent in the Bam HI amplicon of the driver.

We were able to study two sisters from an Amish family with a well-established pedigree. Amplicons were prepared after cleavage with Bam HI, Bgl II, or Hind III. Difference products between amplicons were obtained (*12, 14*) and size-fractionated by gel electrophoresis (Fig. 3A). A discrete but complex pattern of bands was observed in each case. After three successive hybridizations and amplifications (Fig. 3A, lanes d, h, and l), difference products were cloned into plasmids. For each difference product we picked three probes for blot hybridization analysis, and found that all of them detected polymorphisms within the Amish family. We analyzed Bam HI difference products in the greatest detail (Table 2). Of 20 randomly picked clones, 12 different classes of clones remained after removing redundancies, and the inserts from nine of these were used as probes in DNA (Southern) blots of tester, driver, and five other members of the family. All probes detected two alleles, distinguished by a large and a small DNA fragment (Table 2, allele size). The small allele was always present in the tester (Fig. 3B, lane b, and Table 2, sample B) and always absent in the driver (Fig. 3B, lane c, and Table 2, sample A). The probes detected either one of these bands in presumed homozygotes (Fig. 3B, lanes c and e), or both bands in presumed heterozygous individuals (Fig. 3B, lanes a, b, d, f, and g). Moreover, the blot hybridization pattern for each probe was consistent with a Mendelian pattern of inheritance. From these experiments we concluded that our method yields collections of probes for restriction endonuclease fragment polymorphisms between two individuals.

Each of the Bam HI probes derived from the above experiment was also used in blot hybridizations to amplicons from the family and ten other unrelated human DNA's extracted from cell lines or placentas (Fig. 3C and Table 2). Such blots detect the presence or absence of small-size Bam HI fragments in the tested DNA, and thus are more readily performed than conventional Southern blots to total genomic DNA. We found that blotting amplicons was in complete concordance with blotting of total genomic DNA, as described above, in 63 out of 63 cases (nine probes times seven individuals). The results of the amplicon blots indicate that probes capable of detecting polymorphisms within the Amish family tend to detect polymorphisms in the human population at large. We call such polymorphisms PARF's, for polymorphic amplifiable restriction endonuclease fragments.

Table 1. Sequences of primers used for representational difference analysis. Primer set 1 (R series) was used for preparing amplicon representations, and sets 2 (J series) and 3 (N series) were used for odd and even hybridization–amplifications, respectively. The OLIGO computer program (National Biosciences) was used to check the oligonucleotide design for the absence of strong secondary structure.

Primer set	Name	Sequence
1	R Bgl24	5′-AGCACTCTCCAGCCTCTCACCGCA-3′
	R Bgl12	5′-GATCTGCGGTGA-3′
2	J Bgl24	5′-ACCGACGTCGACTATCCATGAACA-3′
	J Bgl12	5′-GATCTGTTCATG-3′
3	N Bgl24	5′-AGGCAACTGTGCTATCCGAGGGAA-3′
	N Bgl12	5′-GATCTTCCCTCG-3′
1	R Bam24	5′-AGCACTCTCCAGCCTCTCACCGAG-3′
	R Bam12	5′-GATCCTCGGTGA-3′
2	J Bam24	5′-ACCGACGTCGACTATCCATGAACG-3′
	J Bam12	5′-GATCCGTTCATG-3′
3	N Bam24	5′-AGGCAACTGTGCTATCCGAGGGAG-3′
	N Bam12	5′-GATCCTCCCTCG-3′
1	R Hind24	Same as R Bgl24 (see above)
	R Hind12	5′-AGCTTGCGGTGA-3′
2	J Hind24	Same as J Bgl24 (see above)
	J Hind12	5′-AGCTTGTTCATG-3′
3	N Hind24	5′-AGGCAGCTGTGGTATCGAGGGAGA-3′
	N Hind12	5′-AGCTTCTCCCTC-3′

Table 2. Screening for presence of Bam HI PARF's in 17 human DNA samples.

Probe		DNA sample*																	Allele size (kbp)†	
No.	Percent‡	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	Large	Small
1	15.5	−	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	15	0.61, 0.67§
11	14.4	−	+	−	−	+	+	−	−	−	−	−	−	−	−	−	−	−	15	0.6
6	8.9	−	+	+	+	+	+	+	+	+	+	+	−	+	−	+	−	+	3.5	0.58
19	5.5	−	+	+	−	+	+	+	+	+	+	−	+	−	−	+	+	+	15	0.51
17	4.4	−	+	−	−	+	+	+	+	+	−	−	−	−	−	−	−	+	8	0.48
22	4.4	−	+	+	+	−	+	+	+	−	+	+	+	+	−	+	+	+	6.5	0.67
8	3.3	−	+	+	+	−	+	+	+	+	+	+	−	−	−	+	+	+	ND	0.62
24	3.3	−	+	−	+	+	+	+	+	−	+	+	−	+	+	−	+	−	>50	0.65
26	3.3	−	+	+	−	−	−	+	+	+	+	−	+	−	+	+	+	−	6,5‖	0.65
9	2.2	−	+	−	−	−	−	+	−	+	−	−	−	−	−	−	−	−	ND	0.47
65	2.2	−	+	+	+	+	+	+	+	+	+	+	+	+	−	+	+	+	4	0.74
3	1.1	−	+	+	+	−	+	+	+	+	+	+	−	+	−	+	+	+	ND	0.5

*Bam HI amplicons were prepared from DNA from seven Amish pedigree lymphoblastoid cell cultures, GM05901 (driver), GM05987 (tester), GM05918, GM05961, GM05963, GM05993, GM05995 (columns A to G), five different placentas (columns H to L), three lymphoblastoid cell lines established from the biopsies of leukemic patients (columns M, N, and O), and two fibroblast cell cultures, DRL 484 and DRL 569, established from the biopsies of DMD patients (columns P and Q), transferred to GeneScreen membrane, and hybridized to the indicated probes. The plus sign means that the small Bam HI PARF allele was present in the sample (that is, the probe hybridized to a band of the correct size in the amplicon); the minus sign means that the small allele was not detected (see, for example, Fig. 3C). †The sizes of the alleles hybridizing to PARF's are indicated, where known. ‡The percentage of clones in the Bam HI difference product (obtained after three hybridization-amplification steps) that hybridized to the indicated clone. §Two different small alleles were found in the human population. ‖Two different large alleles were found in the human population. ND, not determined.

The probes for PARF's are not equally abundant in the difference product. To obtain a measure of this uneveness, we hybridized each cloned Bam HI PARF to a grid of 90 individual randomly picked clones from the difference product of the two siblings, and its frequency in the collection was determined (see percentage value in Table 2). Of 90 randomly picked elements, only 20 distinct polymorphic probes were present, and at very different frequencies. We estimate that there should be of the order of 1000 Bam HI PARF's between two individuals. The uneven distribution of PARF's in our difference products probably reflects the specific conditions used in our protocol, which was designed for the detection of a small number of differences between two nearly identical genomes. We found that, where probes for polymorphic loci were deliberately sought, more representative difference products could be generated by diminishing the number of rounds of hybridization and amplification, increasing the complexity of the representation or decreasing the total number of PCR cycles.

Applications to pathogen discovery. In principle, RDA can be used to isolate probes for pathogens when DNA from infected tissue is compared to DNA from uninfected tissue from the same individual. We have demonstrated this in the model case of the acquisition of a large (30 to 50 kbp) viral genome even with a single copy per cell. Probes for smaller viral genomes should also be detectable, but might require several applications of this procedure with several different restriction endonucleases in order to find fragments from the viral DNA that are readily amplified. As it is now described, RDA cannot reasonably be expected to detect probes for pathogens that are present at less than one copy per ten cells in infected tissue. The kinetic enrichment component would select against target sequences considerably when they are present at concentrations lower than other tester sequences. However, procedures for "normalization" (*18*), that equalize the concentrations of all tester sequences, could be applied prior to subtractive and kinetic enrichment.

Application of RDA to the discovery of pathogens requires precise matching of the polymorphisms from the infected and uninfected DNA sources. Tester and driver DNA can derive from the same individual, if the individual is not a genetic mosaic. These DNA's cannot derive from unrelated individuals, as the abundant polymorphic differences in their DNA's would obscure the detection of the pathogen. However, the uninfected DNA source (driver) could, in principle, come from an identical twin, or be the pooled DNA from the parents of the infected individual, because virtually every DNA restriction fragment found in the genomic DNA of the infected individual can be expected to be present in at least one parental DNA (*19*). Pooled parental drivers might be sought if the pathogen is suspected to be ubiquitous in the infected individual.

Applications to detecting genetic abnormalities in cancer. RDA should enable the discovery of genomic alterations occurring in cancer cells. These could be of two distinct types: those that result in loss of restriction endonuclease fragments, such as might occur from deletions or gene conversions extending over heterozygous polymorphisms, and those that produce new restriction endonuclease fragments, such as might result from genomic rearrangements. In the former case, RDA could be applied without modifications with DNA from cancer cells as driver and normal DNA as tester. Unfortunately, the presence of normal stroma in a cancer biopsy would certainly interfere with the detection of loss of genetic information in the cancer cell. Hence, either cultures of cancer cells, xenographs of tumors, or highly purified cancer cells obtained by physical separation would be needed as the source for tester.

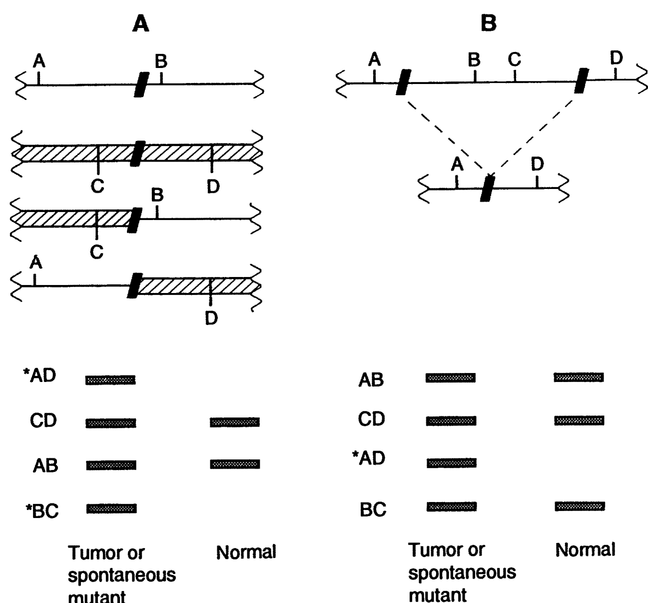


Fig. 4. Schematic representation of restriction endonuclease fragments of genomes that have suffered rearrangements. (**A**) Reciprocal translocation. (**B**) Deletion. Restriction endonuclease sites A, B, C, and D are as shown before and after rearrangements. Below is depicted the mobility of the new (*) and unrearranged fragments during gel electrophoresis. We assume that the normal chromosomes of the involved pair are present.

These restraints do not apply to the detection of genomic rearrangements. Genomic rearrangements, including translocations, insertions, inversions, and deletions, may result in the creation of new restriction endonuclease fragments "bridging" the site of the rearrangement (Fig. 4). Some of these bridging fragments may be amplifiable by PCR (fragment BC in Fig. 4A). Such bridging fragments would be discoverable by RDA when DNA from the tumor is used for preparation of tester amplicons and DNA from normal tissue of the same individual is used for preparation of driver amplicons.

The different-sized restriction endonuclease fragments created by genomic rearrangements can potentially be exploited another way. Fractionated size classes from tumor DNA digests may sometimes contain sequences that are not present in comparable size classes from normal DNA (Fig. 4). With the former as tester and the latter as driver, we can prepare amplicons after cleavage with a second restriction endonuclease and compare these in order to clone amplifiable restriction endonuclease fragments in proximity to the point of genetic rearrangement. The presence of normal cells among the tumor cells should not obscure the detection of probes for the rearrangement.

Application to genetic analysis. When RDA is applied to different individuals, and amplicons are used as the representation, probes for a special type of polymorphism that we call PARF's result. Other types of representations can yield different types of polymorphisms. In general, RDA may be useful for generating new sets of polymorphisms not only for species that have not previously undergone extensive molecular genetic characterization, but even for such well-studied species as humans and mice. Since PARF's are most often binary, they may be especially useful for creating a panel of probes that can be used with a standardized format for genetic typing.

RDA may find special uses when applied to pairwise comparisons of amplicons obtained from DNA's of individuals from specific groups. Straightforward application of RDA could yield probes for PARF's in the DNA of all individuals from a founder group affected by some autosomal dominant inherited disorder (the tester), but absent in the pooled DNA of all individuals from a normal group (the driver). Conversely, RDA could yield probes for PARF's present in the DNA of a normal individual (the tester), but absent in the DNA of all individuals from a founder group affected by a recessive inherited disorder (the driver). Combined with methods for coincidence cloning (*20*), such applications might accelerate the discovery of probes for rare PARF's in linkage disequilibrium with the dominant locus, or the absence of common PARF's in linkage disequilibrium with the recessive locus (*21*).

Many genetic diseases of progeny occur as a consequence of spontaneous germline genomic rearrangements in the gamete of one of the parents. The genome of the affected individual will in all probability have acquired restriction endonuclease fragments that are not present in the somatic cells of either parent. This situation is formally analogous to genetic rearrangements occurring in cancer cells (Fig. 4). In these special cases, RDA might be applied directly to the isolation of probes close to the sites of genetic abnormalities by taking DNA from the individual as tester and pooled DNA's from parents as driver.

Programmed rearrangements in somatic cells may also be discovered by RDA. Such events, for example, occur during the generation of diversity of the immune system (*1*), and have also been postulated during development of the nervous system (*22*). In principle, our method could be applied to the discovery of probes that detect these events. Finally, RDA may have special applications for the study of organisms that can be bred. Multiple backcrossing of a mutant strain to a polymorphic strain could lead to the identification of clonable restriction endonuclease polymorphic fragments tightly linked to the mutant locus.

Cautions. The data obtained with our protocol were highly reproducible; the same bands and even their relative intensities were obtained in multiple independent runs. Application of any new technique requires awareness of possible sources of artifact. One clear source can be PCR itself, because of the ease with which previous PCR products can contaminate reactions. Moreover, PCR products present after subtraction and enrichment do not necessarily reflect effective enrichment of target and may result from the stochastic nature of the process itself (9). Therefore each candidate difference product must be tested for its presence or absence in tester and driver amplicons. Another source of artifact can be anticipated during tissue sampling. Normal flora contaminating a specimen of tester will be readily enriched during difference analysis if that flora are not also present in driver. Other sources of artifact can be readily imagined. For example, we do not know to what extent we may find programmed genetic rearrangements specific for certain tissues or organs. We also do not know the extent of genetic mosaicism in somatic tissues, the result of spontaneous genetic events occurring during early development. Even so, application of RDA should provide an economical route to solving many types of genetic problems.

REFERENCES AND NOTES

1. O. Bernard, N. Hozumi, S. Tonegawa, *Cell* **15**, 1133 (1978).
2. J. N. Strathern, E. Spatola, C. McGill, J. B. Hicks, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2839 (1980); D. H. Beach, *Nature* **305**, 682 (1983).
3. E. Segal, E. Billyard, M. So, S. Storzbach, T. F. Meyer, *Cell* **40**, 293 (1985); L. H. T. Van der Ploeg, C. A. Cornelissen, P. A. M. Michels, P. Borst, *ibid.* **39**, 213 (1984).
4. E. E. Lamar and E. Palmer, *ibid.* **37**, 171 (1984).
5. L. M. Kunkel, A. P. Monaco, W. Middlesworth, H. D. Ochs, S. A. Latt, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4778 (1985).
6. R. L. Nussbaum, J. G. Lesko, R. A. Lewis, S. A. Ledbetter, D. H. Ledbetter, *ibid.* **84**, 6521 (1987).
7. Throughout this article, we use "target" to refer to sequences present in one population, the "tester," but absent in another, the "driver."
8. In general, a single cycle of subtraction can be expected to yield enrichments of target of the order of fN where N is the molar excess of driver to tester and f is the fraction of driver that reanneals.
9. I. Wieland, G. Bolger, G. Asouline, M. Wigler, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 2720 (1990).
10. To visualize this, consider viral sequences pres-

ent in excess (ten times more) relative to single-copy β-globin sequences. At early stages of self-reannealing, when 5.0 percent of the viral sequences are reannealed, only 0.5 percent of the β-globin sequences will be reannealed. The ratio of the viral sequences to the β-globin sequences in the double-stranded DNA will then be 5 percent of 10 to 0.5 percent of 1 (one hundred times more).

11. R. K. Saiki, *et al. Science* **239**, 487 (1988).
12. For preparation of amplicons both tester and driver DNA samples were digested with restriction endonuclease (New England Biolabs) and 1 μg of each DNA digest was mixed with 0.5 nmol of 24-bp and of 12-bp unphosphorylated oligonucleotides (Table 1, primer set 1) in 30 μl of T4 DNA ligase buffer (New England Biolabs). Oligonucleotides were annealed by cooling the mixture gradually from 50° to 10°C for 1 hour and then ligated to human DNA fragments by overnight incubation with 400 U of T4 DNA ligase at 16°C. After ligation, both tester and driver DNA samples were amplified. Each of ten tubes taken for preparation of driver amplicons and two tubes used for preparation of tester amplicons contained (in 400 μl) 67 mM tris-HCl, pH 8.8 at 25°C, 4 mM $MgCl_2$, 16 mM $(NH_4)_2SO_4$, 10 mM β-mercaptoethanol, bovine serum albumin (100 μg/ml), 300 μM (each) dATP, dGTP, dCTP, and dTTP, 1 μM 24-bp primer, and 80 ng of DNA with ligated adaptors. The tubes were incubated for 3 minutes at 72°C in a thermal cycler (Perkin-Elmer Cetus), 15 U of Taq polymerase (AmpliTaq, Perkin-Elmer Cetus) was added, the reactions were overlaid with mineral oil and incubated for 5 minutes to fill in 5′ protruding ends of ligated adaptors, and amplified for 20 cycles (each cycle including 1 minute incubation at 95°C and 3 minutes at 72°C, with the last cycle followed by an extension at 72°C for 10 minutes). After amplification both driver and tester amplicons were digested with the same restriction endonuclease (10 U/μg) to cleave away adaptors. For removal of adaptors, 10 μg of tester amplicon DNA digest was subjected to electrophoresis through 2 percent NuSieve agarose (low melting point, FMC BioProducts), and DNA fragments (150 to 1500 bp) were recovered after melting of the agarose slice and Qiagen-tip20 chromatography (Quiagen Inc.).
13. These estimates in degree of simplification were calculated from data on the mean restriction endonuclease fragment lengths and the proportion of fragments of <1 kbp [D. T. Bishop, J. A. Williamson, M. H. Skolnick, *Am. J. Hum. Genet.* **35**, 795 (1983)]. Our calculations of the number of fragments with an amplifiable size are not based on complete information. We do not know, for example, the exact mathematical dependence of amplification efficiency upon fragment length, nor the proportion of fragments of any length that amplify poorly. Moreover, the occurrence of restriction endonuclease cleavage sites is not truly random.
14. In preparation for the hybridization and amplification step, fragments of tester amplicons were ligated to new adaptors (Table 1, primer set 2, and Fig. 1). The tester amplicon (0.5 μg) ligated to adaptors and the driver amplicon (40 μg) DNA's were mixed, ethanol-precipitated, dissolved in 4 μl of 3 × EE buffer (*15*), overlaid with 30 μl of mineral oil (Perkin-Elmer Cetus), and denatured by heat; 1 μl of 5 M NaCl solution was added and DNA was hybridized for 20 hours at 67°C. At the end of hybridization, part (10 percent) of the resulting DNA was incubated with 15 U of Taq polymerase (5 minutes, 72°C) in 400 μl of PCR mixture without primer to fill in ends of the reannealed tester, and then amplified for 10 cycles (1 minute at 95°C, 3 minutes at 70°C, and held for 10 minutes more for the last round) after addition of the same 24-bp oligonucleotide to which the tester was ligated. Single-stranded DNA molecules present after amplification were degraded by a 30-minute incubation with 20 U of mung bean nuclease (New England Biolabs) in a volume 40 μl, diluted (1:5) in 50 mM tris-HCl (pH 8.9), and heated (95°C, 5 minutes) to inactivate the enzyme. A portion (40 μl) of the solution was amplified for 15 to 20 cycles under the same conditions as before the mung bean nuclease treatment. Amplified DNA (3 to 5 μg) was digested with the original restriction endonuclease, and 200 ng of the digest was ligated to the third adaptor (Table 1, primer set 3). This DNA (50 to 100 ng) was mixed with 40 μg of driver amplicon and the hybridization and amplification procedures were repeated as in the first cycle. A sample (200 ng) of the digest obtained after the second hybridization-amplification step (Table 1, primer set 2) was then ligated to the second set of adaptors, and 100 to 400 pg of this material together with 40 μg of driver amplicon was taken for the third round of hybridization, with a final amplification after mung bean nuclease digestion for 20 to 25 cycles. When a fourth hybridization-amplification step was performed, 5 pg of material from the third round was ligated to the third adaptors (Table 1, primer set 3) and mixed with 40 μg of driver amplicon before proceeding.
15. D. Straus and F. M. Ausubel, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1889 (1990).
16. PCR adaptors were alternated between rounds of hybridization and amplification to avoid the accumulation of PCR products that might interfere with subsequent amplifications. The adaptors are designed to reconstruct the original restriction endonuclease cleavage sites that defined the tester fragments, and are therefore removable by cleavage with that same enzyme.
17. If enriched tester is insufficiently diluted, further enrichment due to the subtractive component will be hampered because driver amplicon sequences may not then be in excess over amplicon sequences from tester. Further enrichment due to the kinetic component will also be diminished because of the excessive self-reannealing of non-target tester. If enriched tester is diluted too much, so little target will reanneal that it may no longer be detectable by PCR. The protocols described should serve as a practical guide.
18. S. R. Patanjali, S. Parimoo, S. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1943 (1991).
19. Homologous germline recombination between two polymorphic restriction endonuclease sites would create a new restriction fragment length polymorphism present only in the progeny, but this should be a rare event.
20. A. J. Brookes and D. J. Porteous, *Nucleic Acid. Res.* **19**, 2609 (1991).
21. We do not have enough experience with RDA to predict what would happen if DNA's pooled from individuals of one group were used as tester and DNA's pooled from individuals of another group were used as driver. We expect the difference product would yield probes to PARF's common to most of the first but absent in all of the second group.
22. H. Yokota and M. Oishi, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6398 (1990); J. J. Chun, D. G. Schatz, M. A. Oettinger, R. Jaenisch, D. Baltimore, *Cell* **1**, 189 (1991); M. Matsuoka *et al.*, *Science* **254**, 81 (1991).
23. We thank B. Stillman, T. Grodzicker, J. Watson, and R. Axel for constructive criticisms of the manuscript; K. Kidd for useful discussions, S. Teplin for preparative synthesis of oligonucleotides, L. Rodgers for growth of human cell cultures; T. C. Caskey for DNA purified from the lymphoid DR1 cell line 484; J. Duffy and M. Ockler for preparation of figures; and P. Bird for secretarial assistance. Supported by grants from the National Cancer Institute, and the American Cancer Society. M.W. is an American Cancer Society Research Professor. The first author is on leave from the Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov Square, Moscow, 123182 Russia. Part of this work was conceived there.

21 October 1992; accepted 6 January 1993