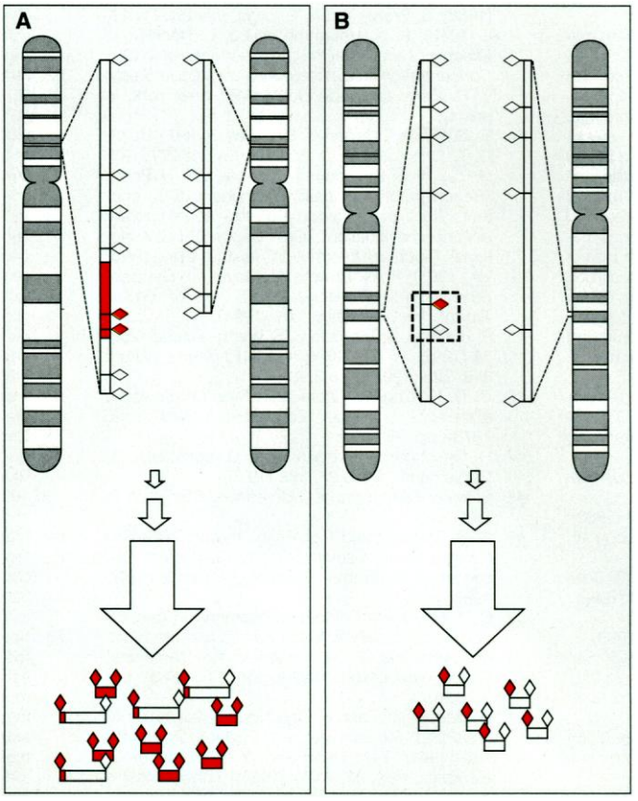
The Pluses of Subtraction

Richard M. Myers

A general problem in genetics is the identification of specific DNA sequence differences underlying phenotypic variation. These differences can be foreign DNA segments from an infecting pathogen that integrate into a host genome, spontaneously arising DNA rearrangements or point mutations that cause cancer in somatic tissues, or a wide variety of induced and naturally occurring mutations that are transmitted from generation to generation, often resulting in genetic diseases in humans. Each of these situations requires a somewhat different approach to identify the particular DNA difference in the genomes of the unaffected and the affected cell or organism that causes the phenotype. But, as a rule for all of these cases, the larger the genome or the smaller the DNA difference, the more difficult it is to find. In this issue of *Science*, Lisitsyn and co-workers (*1*) describe a new method, based on the principles of subtractive hybridization, for cloning even the smallest DNA differences between two or more complex genomes that shows great promise for speeding up this rate-limiting step in genetic analysis.

Subtractive hybridization was first used in 1966 to identify sequences from a small deletion in the bacteriophage T4 genome (*2*). Subtraction approaches usually involve repeated rounds of hybridization of excess DNA from the source lacking a desired sequence with DNA from the source containing the desired sequence, followed by removal of the undesired sequences at each round. Numerous such methods have been developed for isolating differentially expressed mRNA sequences (*3*), but there have been few examples of this approach being used to isolate differences in genomic DNA segments. One genomic subtraction approach (*4*) yielded a several hundredfold enrichment of desired sequences from a small (5–kilobase pair) deletion in yeast after several rounds of subtraction, making it possible to identify DNA fragments from the deletion among a small number of fragments that survived the procedure.



Two general applications of RDA. (**A**) Cloning DNA fragments present in an insertion. The left and right chromosomes are identical except for a small DNA insertion (red segment). Restriction enzyme cleavage sites are indicated by diamonds. Three rounds of RDA, depicted by arrows of increasing size, selectively generate DNA fragments from the insertion. The same procedure can be used to clone DNA fragments that are removed by a small deletion; in this case, the chromosome on the left would be "normal" and that on the right would harbor the deletion. (**B**) Cloning DNA fragments differing by RFLPs. The left and right chromosomes differ by a number of restriction enzyme cleavage sites (one of which is shown as a red diamond). Three rounds of RNA selectively generate DNA fragments that are unique to the left chromosome because of the polymorphic restriction sites. [Drawing courtesy of L. Stuvé]

The situation has been much more difficult with mammalian genomic DNA, however. Several DNA probes were isolated from the Y chromosome as the DNA sequence difference product between male and female genomes by a subtractive method (*5*). Kunkel and co-workers (*6*) combined subtraction with a method called PERT (phenol emulsion reassociation technique) (*7*) for increasing rehybridization rates to isolate several DNA probes from a deletion of a few million base pairs that turned out to be key reagents in cloning the Duchenne muscular dystrophy gene. Other groups (*8*) have successfully applied this approach to isolate difference products from similar-sized deletions in other regions of human chromosomes.

These previous experiments with mammalian genomic DNA, while quite noble, are difficult to perform and, more important, yield enrichments of only between 10- and 100-fold. Thus additional work is required to sort through large numbers of undesired DNA probes, and the usefulness of the procedures is limited to cloning portions of the DNA present in fairly large deletions or insertions. What distinguishes the new method is its relative simplicity and its ability to enrich for the desired sequences 10^5- to 10^6-fold after a small number of rounds. This increases the sensitivity of the method so that it can be used to isolate very small insertions or deletions in a genome as complex as a mammal's. Furthermore, the sensitivity and experimental design of the method are unusual in that they provide a way to isolate at least a portion of the very small differences that result in restriction enzyme fragment length polymorphisms (RFLPs) between two complex genomes.

The method. Lisitsyn and coworkers demonstrate two general applications of their method, which they call RDA (representational difference analysis), for cloning DNA fragments that differ in size between two genomes. In the first experiment, they isolate restriction enzyme fragments from a viral DNA genome present as a single copy in a mammalian genome. In the second experiment, they isolate restriction enzyme fragments from human genomic DNA that are polymorphic between two closely related individuals. Notably, these difference fragments, or "targets," were the only DNA fragments isolated by both applications of the procedure. Thus, in the authors' experiments, no additional sifting through an enriched library was required with RDA.

In the initial step, the complexity of the two test genomes is decreased by amplifying a relatively small fraction (less than 10%) of each genome with a "whole genome" polymerase chain reaction (PCR) method (*9*). These "representations" of the genomes are used in a subsequent subtractive hybridization step, in which the amplified segments of the genome lacking the target are hybridized in great excess to those from the target-con-

The author is in the Departments of Physiology and Biochemistry/Biophysics at the University of California, San Francisco, CA 94143–0444.

taining genome. In this step, the hybridization is allowed to proceed to the point that a small percentage of the target segments reanneal, but a large percentage of the remaining DNA fragments in common between the two genomes reanneal because of the molar excess of the genome lacking the target. PCR amplification is then performed with primers that allow only the small (less than 1 kilobase pair) double-stranded target segments to amplify exponentially; the design is such that all other undesired DNA fragments either do not amplify or do so only linearly. Single-stranded DNA is then removed by a simple nuclease digestion step. Even after this single round of the method, the resulting double-stranded amplified fragments are greatly enriched in the desired target fragments. Additional rounds of the same treatment generate a few discrete DNA fragments, all of which appear to be difference products, that can be isolated in pure form from agarose gels and cloned.

Three key features distinguish RDA from other genomic subtraction methods and contribute to the technique's overall effectiveness:

First is the step in which the complexity of the two starting genomes is reduced by making representations. This step appears to be critical, at least when mammalian genomes are used, presumably because it allows the concentration of target fragments to be high enough that practical levels of reannealing can occur. Although a representation made with a single restriction enzyme limits the amount of the genome that is sampled, it is possible to obtain a higher fraction of the genome, although probably never 100%, as representations by using several different restriction enzymes in separate reactions. However, it is usually not critical that most of the genome be represented (*10*). For example, if RDA is being used to identify an unknown viral DNA insertion that is 10 kilobase pairs in length, the use of two or three restriction enzymes to generate representations has a very high chance of generating small target fragments that can be amplified by PCR.

Second is the "kinetic enrichment" step, which, in later rounds of the procedure, increases the amount of target that reanneals by a factor roughly the square of the amount of nontarget sequences that reanneal. This effect can be dramatic. For instance, if target sequences are present at 100 times the amount of nontarget sequences after a round or two of RDA, the kinetic enrichment in the next round can be as much as 10,000-fold. Although subtractive methods for normalizing complementary DNA libraries have used something akin to kinetic enrichment (*3*), it appears that none of the other genomic schemes were knowingly designed to take advantage of this phenomenon.

The third key feature is the attachment of linkers onto the ends of only the target-containing genome representation. This step provides a simple way (that is, PCR) to separate the difference products away from most of the undesirable DNA after each round of subtraction and kinetic enrichment, and it allows a much higher degree of purification than do commonly used physical separation methods, such as chromatography. It is also likely to be easier to perform.

Applications of RDA. RDA was designed as a versatile way to solve a particularly difficult step in genetic analysis, that of identifying DNA sequence differences between individuals. Thus, it is likely that it will be useful for studying a large number of biological problems in a wide variety of organisms. The method appears to be especially effective for isolating DNA fragments present in insertions or removed by deletion, which should lead to the identification of new viruses, transposable elements, and DNA segments at sites of chromosomal rearrangements that cause cancers and many hereditary diseases. Because it is so simple to confirm, by hybridization or PCR, that a putative target DNA is derived from an insertion or deletion, it seems likely that we will rapidly see many such successful applications of RDA.

If these were the only uses for RDA, the technique would be quite valuable. However, the prospect of using the approach to identify at least some of the polymorphic variation due to single base differences expands the scope of RDA to a wholly different set of interesting applications. At the very least, it provides a way to isolate polymorphic DNA probes for standard meiotic mapping and genetic linkage analysis. Although there are already many highly polymorphic DNA markers available for the human and mouse genomes, a substantial number of the markers are not informative in any particular family study, so it is often necessary to search for new markers. In addition, there are many experimental organisms for which large numbers of polymorphic markers are not available, and RDA could provide a rapid way to make better genetic maps of these genomes. It may be possible to use RDA with groups of individuals to identify DNA markers in linkage disequilibrium with a disease or another phenotype caused by a founder mutation. Similarly, the method may be a useful adjunct for strategies for mapping regions of genomes that are identical-by-descent in related individuals affected with a genetic disease (*11*). Despite these potential applications, there are some important limitations in using RDA to identify single base changes. For example, the requirement for representations and the low chance that a specific single base mutation changes a restriction enzyme site in the required manner make it unlikely that RDA will be used for the initial identification of a mutation that causes a genetic disease.

Numerous other uses of both the deletion-insertion and single base polymorphism isolation schemes of RDA can be imagined. The classic closure problem of filling in gaps in physical maps of genomes (*12*) could benefit from applying RDA to DNA from somatic cell hybrids from mammalian systems and, perhaps, to cloned versus genomic DNA from less complex organisms. While RDA clearly has applications for studying cancer and genetic diseases in humans, the method could be a powerful tool for genetic analysis of other organisms, especially those with small genomes. In addition, there are many organisms that can be studied by classical genetics but that have been difficult to study at the molecular level because of the inability to do insertional mutagenesis. However, in many cases—for example, the zebrafish—it is straightforward to generate deletion mutations by irradiation (*13*). It may be possible to use RDA to simplify the cloning of DNA fragments from such deletions for the same types of gene searches that have been successful in *Drosophila* and other organisms.

References and Notes

1. N. Lisitsyn, N. Lisitsyn, M. Wigler, *Science* **259**, 946 (1993).
2. E. K. F. Bautz and E. Reilly, *ibid.* **151**, 328 (1966).
3. G. A. Galau, W. H. Klein, R. J. Britten, E. H. Davidson, *Arch. Biochem. Biophys.* **179**, 584 (1977); S. R. Patanjali, S. Parimoo, S. M. Weissman, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1943 (1991).
4. D. Straus and F. M. Ausubel, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1889 (1990).
5. E. E. Lamar and E. Palmer, *Cell* **37**, 171 (1984).
6. L. M. Kunkel, A. P. Monaco, W. Middlesworth, H. D. Ochs, S. A. Latt, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4778 (1985).
7. D. E. Kohne, S. A. Levinson, M. J. Byers, *Biochemistry* **16**, 5329 (1977).
8. R. L. Nussbaum, J. G. Lesko, R. A. Lewis, S. A. Ledbetter, D. H. Ledbetter, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6521 (1987); Y. Shiloh *et al.*, *Gene* **51**, 53 (1987); O. Mor *et al.*, *Nucleic Acids Res.* **19**, 117 (1991); Y. Nishi, K. Akiyama, B. R. Korf, *Mamm. Genome* **2**, 11 (1992).
9. K. W. Kinzler and B. Vogelstein, *Nucleic Acids Res.* **17**, 3645 (1989); H.-J. Lüdecke, G. Senger, U. Claussen, B. Horsthemke, *Nature* **338**, 348 (1989); P. E. Mueller and B. Wold, *Science* **246**, 780 (1989).
10. Nevertheless, one can imagine cases where it would be nice to be able to test an entire genome with RDA without having to rely on the use of multiple restriction enzymes. If indeed the reason for needing the representation step is simply that genome complexity must be reduced, it seems likely that this step could be eliminated and total genomic DNA could be used for studies involving genomes with complexities at least 10- to 20-fold lower than those of mammals.
11. E. S. Lander and D. Botstein, *Science* **236**, 1567 (1987).
12. M. V. Olson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7826 (1986); A. Coulson, J. Sulston, S. Brenner, J. Karn, *ibid.*, p. 7821; Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987).
13. C. Walker and G. Streisinger, *Genetics* **103**, 125 (1983).

14. I am grateful to D. Cox, A. Peterson, and R. John for useful discussions.