



RNA fingerprinting by arbitrarily primed PCR.

M McClelland and J Welsh

Genome Res. 1994 4: S66-S81

References This article cites 29 articles, 8 of which can be accessed free at:
<http://genome.cshlp.org/content/4/1/S66.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © Cold Spring Harbor Laboratory Press

RNA Fingerprinting by Arbitrarily Primed PCR

**Michael McClelland and
John Welsh**

California Institute of Biological
Research, La Jolla, California 92037

One of the great strengths of the family of arbitrarily primed PCR methods is their simplicity.^(1,2) The extension of arbitrarily primed PCR fingerprinting to RNA^(3,4) has resulted in a tool with exciting potential for detecting differential gene expression. It is now possible to obtain a partly abundance-normalized sample of cDNAs produced in a single tube in a few hours. Fragments of differentially expressed genes can be cloned directly from PCR-amplified products isolated from the gel.

The method can provide a complex phenotype reflecting changes in the abundance of hundreds of RNAs under various conditions. Comparison of RNA fingerprints from different treatment groups allows one to draw inferences regarding gene regulation. Hypotheses regarding signal transduction can be tested and new hypotheses generated using this information. In this brief methods-oriented review we will justify our current preferred protocol for RNA fingerprinting. An accompanying article will discuss DNA fingerprinting.⁽⁵⁾

The first published variation of arbitrarily primed PCR applied to RNA uses a clamped 3' primer such as oligo(dT)CA for reverse transcription and an arbitrary primer for priming the second-strand cDNA.⁽³⁾ At these priming sites, the 3' 7 or 8 nucleotides of the primer usually match well with the template, but the nucleotides toward the 5' end of the primer also influence which sequences amplify. The term "differential display" has been coined for this variant and could be applied to all variants of RNA fingerprinting using arbitrary primers.

In one of the many alternatives for RNA fingerprinting with arbitrary primers, first-strand cDNA synthesis is primed using oligo(dT). The first strand cDNA can then be fingerprinted using a single arbitrary primer and two initial low-stringency cycles, incorporating the arbitrary primer at both ends for subsequent PCR amplification.⁽⁶⁾

In another variant of RNA fingerprinting using arbitrarily primed PCR, synthesis of first-strand cDNA is initiated from an arbitrary primer at those sites in the RNA that best match the primer. Second-strand synthesis is achieved by arbitrary priming using a thermostable DNA polymerase, at sites where the primer finds the best matches. Poorer matches at one end of the amplified sequence can be compensated for by very good matches at the other end. These two steps result in a collection of molecules that are flanked at their 3' and 5' ends by the exact sequence (and complement) of the arbitrary primer. These serve as templates for high stringency PCR amplification, resulting in fingerprints similar in appearance to those generated from genomic DNA. This variation on the arbitrarily primed PCR (AP-PCR) strategy (1) has been called RNA arbitrarily primed PCR (RAP-PCR).⁽⁴⁾ Because both primers are internal to the transcript, open reading frames are found in ~30% of products using this variation of RNA fingerprinting.

The ratio of the intensity of RAP-PCR products between samples correlates with the ratio of abundances of the corresponding RNA.^(4,7) Whereas the intensities of different bands within the same fingerprint vary independently of each other, the intensity of a band between fingerprints appears to be proportional to the concentration of its corresponding template sequence.^(4,7) Extraordinary as this may seem at first glance, it is not difficult to rationalize. Each fingerprint is dominated by many products that do not vary in quantity or intensity between samples, whereas each differentially expressed product constitutes only a tiny fraction of the total mass of DNA synthesized. In a carefully performed experiment, the total mass of DNA synthesized from sample to sample is constant. These factors ensure that the amplification efficiency for any particular product is identical between samples. Furthermore, the ratios of intensities for a

particular product reflect the single remaining variable, the template RNA ratios. The semiquantitative nature of arbitrarily primed PCR has also been observed in DNA fingerprinting at heterozygous loci in the F1 progeny of diploids.⁽⁸⁾ This property of arbitrarily primed PCR has allowed the ploidy of chromosomes in tumor cells to be determined by comparison with fingerprints from the normal diploid genome of the same individual.⁽⁹⁾

ABUNDANCE NORMALIZATION

The effect of RNA abundance on the efficiency of sampling by RNA fingerprinting is an important issue. Experiments that used 10-base-long primers perfectly matched at both ends for a rare message did not lead to the targeted cDNA dominating the fingerprint.⁽³⁾ Instead, other products still effectively competed to produce a fingerprint indicating the lack of abundance normalization, a problem that is typical of most arbitrarily primed PCR methods. Therefore, two RNAs with similar arbitrary priming and amplification efficiencies that differ by 100-fold in abundance will have fingerprint products displaying a 100-fold ratio of intensities. How can this problem be overcome? One obvious possibility is to increase the stringency of the arbitrary priming steps, thereby sampling only those transcripts with perfect or near-perfect matches. By increasing the number of bases that match in the arbitrary priming steps, one can increase the normalization. In an RNA sample of 10^8 -base complexity, a perfect match of ~ 8.5 bases at each of the two initial priming events would theoretically produce an average of only one PCR product. This would be a normalized sampling because the match would entirely control the probability of occurrence. This strategy, however, would be a very slow method to yield information on gene expression and also appears to be difficult to execute. Increasing the stringency in the first arbitrary steps has little effect on the fingerprint until a stringency is reached where the fingerprint becomes unreliable. In order that many RNA molecules be sampled simultaneously, mismatches must be tolerated. Whereas the melting temperatures for primer–template interactions drop precipitously in response to increasing mismatches, the first derivative of the melting curve tends to become smaller because cooperative melting is compromised. Then, a small increase in stringency (temperature) has only a modest effect on selectivity. The selectivity imparted by a single additional base pair can be easily overcome by the abundance factor. Even when a better match is thermodynamically favored, over the course of a reaction poor matches can initiate significant synthesis. Thus, it is difficult to completely abundance-normalize under low-stringency conditions (i.e., conditions that permit mismatches). To circumvent this phenomenon and improve normalization, we have used a nested strategy.

Nested RAP–PCR is a method designed to partially normalize the fingerprint with respect to mRNA abundance.^(10,11) The strategy is very similar to standard nested PCR methods, except that we do not know a priori the internal sequences of the amplified products. In this method the fingerprinting protocol is applied to the RNA, as described above. Then, a small aliquot of the first RAP–PCR fingerprint is further amplified at high stringency using a second nested primer having one, two, or three additional arbitrarily chosen nucleotides at the 3' end of the first primer sequence. Because sampling subsequent to the first low-stringency fingerprinting is at high stringency, and the polymerase has a strong preference for a perfectly matched 3' primer end, the arguments pertaining to the shape of the melting curve discussed above do not apply.

A Set of Circularly Permuted Nested Primers

ZF-1	AACCCACCG GAGAGAA
ZF-2	ACCCACCGG AGAGAAAC
ZF-3	CCCCACCGGA GAGAAACC
ZF-4	CCCACCGGAG AGAAACCC
ZF-5	CCACCAGAGA GAAACCCA
ZF-6	CACCAGAGAG AAACCCAC
ZF-7	ACCAGAGAGA AACCCACC
ZF-8	CCAGAGAGAA ACCCACCA
ZF-9	CAGAGAGAAA CCCACCAG
ZF-10	AGAGAGAAAC CCACCAGA
ZF-11	GAGAGAAACC CACCAGAG
ZF-12	AGAGAAACCC ACCAGAGA
ZF-13	GAGAAACCCA CCAGAGAG
ZF-14	AGAAACCCAC CAGAGAGA
ZF-15	GAAACCCACC AGAGAGAA
ZF-1	AACCCACCG GAGAGAAA, etc.

Circular permutation of the primer sets to be used for this strategy allows any initial primer to be nested, for example, ZF-14 can be nested using ZF-15, ZF-1, and ZF-2.^(10,11)

This particular set of circularly permuted primers excludes thymine because this is the least discriminatory base when present at the 3' end of the primer.⁽¹²⁾

This strategy partially abundance-normalizes the sampling that occurs during RNA fingerprinting. Consider again the two messages that have equally good matches and equally good amplification efficiency but differ by 100-fold in abundance. The products derived from them will differ by 100-fold after RAP-PCR. Thus RAP-PCR fingerprinting produces a background of products that are not visible on the gel that includes products derived from low-abundance messages. A secondary round of amplification using a primer identical to the first except for an additional nucleotide at the 3' end of the molecule can be expected to selectively amplify those molecules in the background that, by chance, share this additional nucleotide. The additional nucleotide will occur in $1/16$ of the background molecules, accounting for both ends. There are many more molecules of low abundance in the RNA population than messages of high abundance, so most products produced by the high-stringency nesting step should derive from the low-abundance, high-complexity class.

Each additional nucleotide at the 3' end of the initial primer will contribute, in principle, a factor of $1/16$ to the selectivity. Two extra bases at each end lead to 256-fold selection, and 3 bases at each end lead to 4096-fold selection. Four bases at each end generally do not yield reliable products. In practice, the selectivity is probably somewhat less than simple statistics might suggest because, although *Taq* polymerase is severely biased against extending a mismatch at the last nucleotide, it is more tolerant of mismatches at the second or third positions. Additional selectivity might therefore be achieved by successive nesting experiments. Nonetheless, our initial experiments are consistent with the interpretation that considerable additional selectivity is achieved by this nested priming strategy.

Using total RNA as a starting material means that if nesting is successful, hnRNA sequences will inevitably be sampled even if a primer biased toward the poly(A) tail is used. So far, most of the ~50 differentially expressed genes we have isolated using unnested RAP-PCR from various systems have given strong, sharp signals in Northern analysis, which indicated that the

patterns derive predominately from messenger RNA. However, in nested primer experiments, where complexity becomes progressively more important than abundance, we anticipate that a greater proportion of the patterns will derive from hnRNA, which still contains introns. If this is considered a problem for a particular experimental design, then poly(A) selection or priming with oligo(dT) or oligo(dT)NN clamped primers (as in Ref. 3) might be preferred.

THE SIZE OF A RAP-PCR EXPERIMENT AND THE ISSUE OF COMPLETE SAMPLING

How thoroughly can we explore differential gene expression using RNA fingerprinting? This is perhaps the most important question when comparing RNA fingerprinting with other methods, such as subtractive hybridization. The number of fingerprinting reactions necessary to achieve various levels of coverage of the RNA complexity can be calculated. Neglecting for the moment problems of abundance normalization, we can determine the probability, $P(0)$, of missing a message as $P(0) = e^{-\mu}$. To have a 95% chance of detecting a particular message, $P(0) = 0.05$ and $\mu = 3$, three times the total complexity (i.e., number of different kinds of messages) must be sampled. Assuming two RNA concentrations (providing information on reproducibility), 50 bands per lane, and 100 lanes per gel, $(50 \times 100 / 2) = 2500$ messages can be displayed. Assuming a complexity of 15,000 messages and a factor of three to account for multiple sampling of the same message, the entire complexity of the cell could, in principle, be represented in $(15,000 / 2,500) \times 3 = 18$ gels at the 95% confidence level. In an experiment, involving more than one RNA, say n different RNA sources, $n \times 18$ gels must be run to achieve this level of coverage of all RNAs. This calculation ignores the effect of incomplete abundance normalization: rarer messages will be represented less frequently among the visible products. Thus, $n \times 18$ gels is a considerable underestimate of the effort needed to sample 95% of the RNAs. The method may indeed be incapable of sampling the rarest RNAs and is therefore not appropriate for finding a single particular RNA. Nevertheless, if there are several important molecular markers associated with a phenomenon, the chances of encountering some of them is excellent. Thus, RAP-PCR is ideally suited to situations where the number of genes differentially expressed is fairly high (i.e., >1 per 1000 transcripts).

Much of the technological development associated with sequencing, such as fluorescent-tagged primers, automated gel reading, and capillary electrophoresis, is readily adaptable to RAP-PCR. Therefore, a reasonable fraction of the genes expressed in many situations can be surveyed given existing technology, and this capability will be greatly enhanced by further technological developments.

A systematic way to divide up the pool of mRNAs (and a new acronym) has been suggested to survey the whole mRNA population with a series of other arbitrary primers in pairwise combinations.⁽¹³⁾ However, such calculations assume that the fingerprints are entirely normalized for abundance.

COMPARISON WITH OTHER METHODS

Genetic and reverse genetic methods for cloning interesting genes derive their power from the ability of the investigator to discern and follow a phenotype. Methods that rely on a biological assay of gene function have been very useful, particularly in the area of cancer research. However, appropriate bioassays for most genes are difficult to devise. When no biological assay can

be found or when genetics is inconvenient or impossible, other methods must be employed. Methods for detecting and cloning differentially expressed genes that do not rely on a biological assay of phenotype include subtractive hybridization and differential screening strategies, as well as RAP-PCR fingerprinting, which provides a molecular rather than a biological phenotype.

In a typical differential screening experiment, radioactive probes are made from cDNA from two RNA sources and used to screen a cDNA library prepared from one of the two. Occasionally, clones from the library hybridize to one or the other but not both probes. Unfortunately, low-abundance messages do not yield sufficient probe mass to allow favorable hybridization kinetics and detection levels. This problem is only partly alleviated by screening a library with a subtracted cDNA probe.⁽¹⁴⁾ Nevertheless, some significant discoveries have been made using this approach (e.g., Refs. 15 and 16). Improvements that allow the use of RNA from limited sources have been described (e.g., Refs. 17 and 18). Unlike RAP-PCR, which can be used to screen many RNA samples simultaneously, differential screening is designed for use in a sequential series of pairwise comparisons.

Subtractive hybridization (for review, see Ref. 19) is technically challenging, and has been used successfully by only a few laboratories. This method suffers from drawbacks similar to differential screening. The difficulties with subtractive hybridization methods derive mainly from the fact that rare messages have unfavorable hybridization kinetics, and many interesting genes are of this class. Abundant genes hybridize faster and more completely than low-abundance genes, making them more amenable to subtractive methods. A second problem is that subtraction by exhaustive hybridization of driver and target is employed to avoid sequences that are not differentially expressed. This has the undesirable effect of obscuring significant differences in gene expression that do not fall into the "all-or-nothing" category. Another shortcoming of the method is that, unlike RAP-PCR, it is most conveniently used for pairwise comparisons. Additional comparisons require Northern analysis or similar assays.

Subtractive hybridization and differential screening have proven useful to uncover differentially expressed genes between pairs of RNA populations. Nevertheless, these methods are rather cumbersome. It would be difficult to use these tools to provide a large number of genes with associated biological information from a large number of pairwise comparisons. RAP-PCR can be a powerful tool in this regard.

Theoretically representational difference analysis (RDA),⁽²⁰⁾ which uses an iterative PCR-based subtractive approach to enrich differences in genomic DNAs, could be adapted to isolate genes that are differentially expressed between RNA samples. At present, such an application has not been published. However, a possible combination between RDA and RAP-PCR can be envisioned.

One advantage of PCR-based RNA fingerprinting compared with subtractive hybridization and differential screening is that RAP-PCR yields results in a matter of a few days rather than after many steps that take weeks, a feature that RAP-PCR shares with two-dimensional gel electrophoresis of proteins.^(21,22) However, the main advantage of this technique compared with subtractive hybridization, differential screening, and two-dimensional protein electrophoresis is the fact that the method can compare more than two RNA samples simultaneously. Thus, the effect of a number of conditions can be assayed for any differentially amplified products that are observed. In a well designed experiment, one can select genes that fall into a relatively narrow category for further study. This has become very important because the advent of this technique means that finding differentially expressed genes is no longer rate limiting.

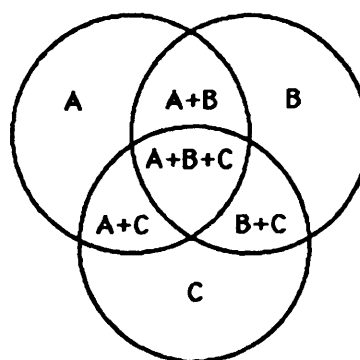
USING RAP-PCR TO ESTIMATE THE PROPORTION OF GENES THAT ACCOMPANY A PARTICULAR PHYSIOLOGICAL CHANGE

The essence of RAP-PCR is its ability to define genes in terms of regulatory sets. For example, consider only three RNA preparations, A, B, and C. Information can be obtained regarding the level of gene expression in 27 possible categories (3^3). Most genes will fall into the first of these categories:

	A	B	C		A	B	C
1	0	0	0	2	↑	0	0
3	↓	0	0	4	0	↑	0
5	0	↓	0	6	0	0	↑
7	0	0	↓	8	↑	↑	0
9	↑	↓	0	10	↑	0	↑
11	↑	0	↓	12	↓	↑	0
13	↓	↓	0	14	↓	0	↑
15	↓	0	↓	16	0	↑	↑
17	0	↑	↓	18	0	↓	↑
19	0	↓	↓	20	↑	↑	↑
21	↓	↓	↓	22	↓	↑	↑
23	↑	↓	↑	24	↑	↑	↓
25	↑	↓	↓	26	↓	↑	↓
27	↓	↓	↑				

(0 = no change; ↑ = increased expression; ↓ = decreased expression).

When the quantitative level of gene expression is also considered then the number of categories increases further. In addition, the treatments A, B, and C can be used in seven possible combinations, increasing even further the information about each sampled gene:



One tremendously useful resource that can be developed very easily using RAP-PCR is a collection of genes that fit any of a wide variety of regulatory categories. This could be achieved by comparing, in parallel, as many RNAs from an isogenic source as is practical. cDNA fragments for genes that fit into the various possible regulatory categories can be isolated and cloned. Concentrating on surveying only two RNA samples does not take advantage of the power of RAP-PCR. For this reason we would argue that the more treatments compared, the more is gained from using the RAP-PCR strategy.

The behavior of each transcript sampled in various experimentally manipulated regulatory scenarios can be determined. In a sense, what is obtained is a "byte" of information for each anonymous gene sampled, each "bit" of

which consists of the transcript abundance in a particular sample. One goal could be to detect and isolate fragments of genes that fit every observable byte or regulatory category. This information should be of interest because the promoters of such genes will also carry in cis functions associated with these regulatory categories. This could be of great utility in the study of promoter function at the level of sequence specificity. The promoters of such genes or targeted insertions into the genes could also be very useful for expressing heterologous genes in a particular subset of situations as may be necessary in some gene therapies, for instance.

In principle, there can be 6561 (i.e., 3^8) possible bytes of ternary bits for eight RNA samples (encoding increases, decreases, and no change), or more possible bytes if further information on the level of gene expression is included in each bit, such as separating increases in gene expression above and below a fourfold induction threshold. It is probable that some of the regulatory categories (i.e., bytes) will be very rare and only rarely sampled, but, with these exceptions, one could clone examples from each category.

Even before such genes are cloned we can obtain a very interesting set of data from the fingerprints. In the 1960's and 1970's when the overlap in gene expression between tissues was estimated (e.g., Ref. 23), this information became a classic addition to textbooks despite the rather one-dimensional information that could be obtained. The reason these data were so interesting was that there was really no practical alternative to divine the scale of the similarities and differences in gene expression between tissues. RAP-PCR fingerprinting of RNA samples can be analyzed in a way that is orders of magnitude better than was possible at that time; the estimate of overlapping gene expression can be extrapolated from the distributions of many thousands of arbitrarily selected examples collected very quickly and easily. Thus, the scale of the intersection between RNA preparations can be estimated by simultaneous pairwise comparisons and examples of genes in different regulatory categories can be cloned if desired.

One obvious reservation about such data is the issue of abundance normalization. If nesting is not used, the observed products are biased toward highly expressed genes, which is a virtue if they are to be used, for example, as biomarkers or if the corresponding promoter is to be used to express heterologous sequences in a particular expression distribution. The issue of abundance normalization can be indirectly addressed, as needed, by comparing the distributions of genes in progressively greater nestings of the initial fingerprint. If there is a systematic difference in the tissue distribution for genes at each level of nesting, this can be attributed to differences in the properties of more common versus rarer messages. Such an observation would not be difficult to quantitate (e.g., by RNase protection of some examples), and the trend would be illuminating, just as were the measurements by hybridization kinetics.

In principle, using a 96-well plate and 96-well gel format one can perform PCR on two concentrations of eight RNA samples using four primers or primer pairs. Nesting each of these four primers to a depth of 3 bases yields at least 12 fingerprints in three further gels even before the possibilities for alternative nestings with different 3' ends are factored in. Given 50 bands per lane, this single set of experiments yields 600 genes sampled. Such information can be gathered for many thousands of anonymous genes, a fraction of which may be selected for cloning, using the byte criteria outlined above. Furthermore, the number of products that vary can be compared with the number that do not vary between different RNA samples. One can then extrapolate the overall level of the response that distinguishes the two RNA samples (within the limitations of abundance normalization, discussed earlier). As an example we present a synopsis of our first use of this idea below.

AN ESTIMATE OF THE PROPORTION OF GENES THAT RESPONDS TO TGF- β AND CYCLOHEXIMIDE

We have fingerprinted lung epithelial cells in the G₁ stage of the cell cycle under the following conditions: (1) TGF- β treatment, that halts progression of these cells into S phase; (2) cycloheximide, that halts protein synthesis; (3) cycloheximide plus TGF- β , that reveals primary response genes to TGF- β ; and (4) without treatment.

In a series of experiments with different primer pairs, we have surveyed at least 616 genes of which 28 genes respond to TGF- β or to cycloheximide treatment. An example of such a fingerprint is shown in Figure 1. We have distinguished between those genes that are secondary response genes to TGF- β and those that are primary response genes (the latter are also regulated

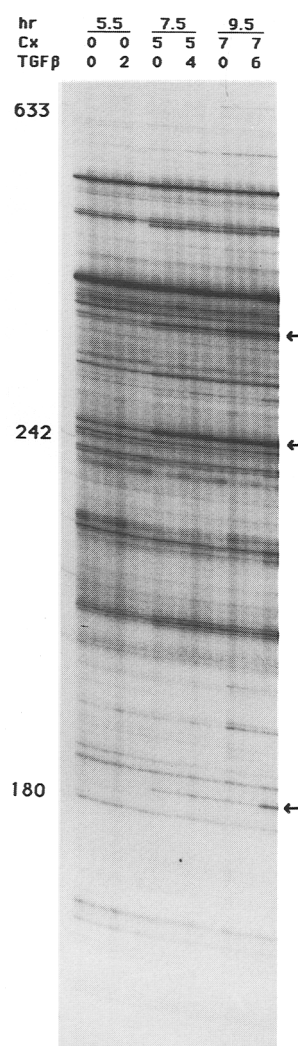


FIGURE 1 RAP-PCR fingerprint. Lung epithelial cells were released from confluence and treated with transforming growth factor- β (TGF- β) or cycloheximide (CX) or both for the times indicated and then harvested. RNA was prepared as described in the text. The primers used were Kinase A1 (5'-GAGGGTGCCTT) carrying a Ser/Thr kinase motif (manufactured by Genosys, Woodlands, TX) and S2 (5'-CCTCTGACTG) (Operon Technologies, Alameda, CA) Each RNA was fingerprinted at three RNA concentrations: 80 ng, 40 ng, and 20 ng per 20 μ l reverse transcription reaction. Separation was on a 5% polyacrylamide-50% urea sequencing gel with electrophoresis in 1 \times TBE at 50 W for 3 hr. The gel was dried, and autoradiography was performed with X-Omat film (Kodak) for 12 hr. Some differentially amplified cDNAs are indicated by arrows. The image shows only a portion of the data. Single-stranded DNA markers are indicated in bases.

by TGF- β even when cycloheximide is present). In addition, we observed an unexpected class of genes that are induced by cycloheximide but are then repressed by TGF- β .

We have also estimated the scale of each of these responses. For example, we determined that in the first hours of G₁, 0.8% (or less) of all transcripts sampled are up-regulated by TGF- β as a primary response. The abundances of 3% (or less) of all transcripts sampled are increased by CX, of which about one-third are down-regulated by TGF- β treatment. Extrapolation from these percentages to the rest of the RNA population is illuminating despite a systematic bias owing to abundance-based sampling. There is no reason to reject the possibility that all RNA abundance categories, including rarer RNAs that are less likely to be sampled in this protocol, may be affected to a similar degree as the sampled RNAs. Assuming that there are 15,000 expressed genes in the cell, then fragments from >3% of all the genes (>616/15,000) that are up-regulated by TGF- β and CX have been identified using only six primers or primer pairs. Cloning such genes constitutes a step in the direction of describing the complete response to these agents and is an essential prelude to determining the physiological role of each of these genes.

In a sense such calculations of the scale of differential gene expression are an extension of the hybridization kinetics work of the 1960's and 1970's (e.g., Ref. 23). However, as discussed above, the complex molecular phenotype generated by RAP-PCR reflects changes in abundances of hundreds of individual RNAs. Many different RNA samples can be compared simultaneously, and genes from assorted regulatory categories can be cloned directly. Hypotheses regarding the interactions of signal transduction pathways can be tested, and new hypotheses can be generated, for example, by comparing the effect of different combinations of hormones or drugs on differential gene expression.

TWO RAP-PCR PROTOCOLS

RNA is prepared using the guanadinium-CsCl procedure.⁽²⁴⁾ We currently perform RAP-PCR on total RNA or poly(A)-selected RNA. Two protocols for fingerprinting are presented.

RAP-PCR Fingerprinting Using a Single 18-Base Primer

Equipment and Reagents

Thermal Cycler (Perkin-Elmer 9600 model)

PCR tubes (Microamp, Perkin-Elmer)

Arbitrary primer, e.g., M13 sequencing or reverse-sequencing primer

2 \times DNase I treatment Mixture: 20 mM Tris-HCl (pH 8.0), 20 mM MgCl₂, 200 U/ml of RNase-free DNase I (Boehringer Mannheim Biochemicals)

2 \times first-strand reaction mixture: 100 mM Tris-HCl (pH 8.3), 100 mM KCl, 8 mM MgCl₂, 20 mM DTT, 0.4 mM each dNTP, 20 μ M primer, and 1 U/ μ l MuLVRT (Stratagene Inc., La Jolla, CA)

2 \times second-strand reaction mixture: 10 mM Tris-HCl (pH 8.3), 25 mM KCl, 2 mM MgCl₂, 0.1 μ Ci/ μ l of [α -³²P]dCTP and 0.2 U/ μ l of *Taq* DNA polymerase (AmpliTaq; Perkin-Elmer Cetus). Addition of a second primer at this step (at \sim 20 μ M) is optional.

FIGURE 2 A RAP-PCR protocol for the use of pairwise sequential 10-mer primers. This protocol, written in MicroSoft Excel, allows the input of many variables. The example shown is set up for six RNA samples at four RNA concentrations, to be sampled by one initial primer and four subsequent primers. Cycling is performed according to the text.

Primers-Specify:

Primers are as follows:		Primer concentration:
First strand primers:	A	100 μ M
Second strand primers:	B, C, D, E	100 μ M

RNA:

RNA stock concentration:	200 ng/ μ l
RNA: Diluted by	1, 2, 4 fold

Experiment Parameters-Specify:

Number of first strand primers:	1	
Number of second strand primers, overall:	4	Should be 4 or more for
Number of RNAs:	6	convenient pipetting volumes.
Number of concentrations of RNA:	3	

First strand synthesis Reaction Mixture without primer.

Set up reactions on ice.		No. reactions:
	Per Reaction:	18.00
10 x RT (50,50,4)	1 μ l	18.00 μ l
100 mM DTT	1	18.00
5 mM dNTP	0.4	7.20
water	1.55	27.90
MuLV-RT (75 U/ μ l)	0.25	4.50
	4.20	75.60

10 x RT (50,50,4) is 500 mM Tris pH 8.3, 500 mM KCl, 40 mM MgCl₂.**Reaction Mixture with First Primers:**

Divide the Reaction Mix. into	1	aliquots:	75.60 μ l
Add 100 μ M primers:	A		14.40
final concentration will be:	16.00	μ M	90.00
Combine:	5.00	μ l with an equal volume of each RNA.	
Final reverse transcription reaction has a volume of:	10.00		
Synthesis: 5 min ramp to 37°C, Hold 60 min. Heat to 94°C for 2 min to kill RT.			
Bring final volume to:	40.00	μ l by adding:	30.00 μ l Water

Buffer concentrations at this point:

Tris	12.50	mM
KCl	12.50	mM
Mg	1.00	mM
Primer	4.00	μ M
dNTP	0.05	mM

Number of second primers used in this trial: 4

2 x AP-PCR reaction mix with Second Primer:

For second strand reaction of 20 μ l:		No. reactions:
10x Taq-Stoffel:	2 μ l	72
5 mM dNTP	0.8	144 μ l
water	5.9	58
Second Primer (100 μ M)	0.8	425
Hot dCTP	0.1	58 Add in next step
Stoffel (10 U/ μ l)	0.4	7
	9.2	29
		662.4

10x Taq-Stoffel buffer is 100 mM Tris pH 8.3, 100 mM KCl, 40 mM MgCl₂.

Add second primer:		B, C, D, E
Each second primer:	14.4	
Reaction mix:	165.6	
	180	

Add 10 μ l 2x AP-PCR Rx Mx with second primer to 10 μ l Rx. Mx. with 1^o primer:Final Reaction Volume (μ l): 20**Final Buffer components:**

Contributions from:	RT reaction	AP-PCR react. Mix.	Final
Tris (mM)	6.25	10.00	16.25
KCl (mM)	6.25	10.00	16.25
Mg (mM)	0.50	4.00	4.50
dNTP (mM)	0.03	0.20	0.23
1 ^o primer (μ M)	2.00	0.00	2.00
2 ^o primer (μ M)	0.00	4.00	4.00
RT (in Units/ μ l)	4.22	0.00	0.00
Stoffel (in Units/ μ l)	0.00	0.20	0.20

Cycling parameters:

Segment	Temp.	Time	
1	94	30 s	
2	40	1 m	
3	72	2 m	45 cycles

FIGURE 2 (See facing page for legend.)

Method

1. Prepare total RNA by guanidinium thiocyanate–cesium chloride centrifugation⁽²⁴⁾ or guanidinium thiocyanate–acid phenol–chloroform extraction.⁽²⁵⁾ Dissolve the final pellet in 100 μ l of water.

2. Add 100 μ l of 2 \times DNase I Treatment Mixture and incubate at 37°C for 30 min. Phenol–chloroform extract and ethanol precipitate. Measure RNA concentration spectrophotometrically.

3. Prepare treated RNA at two concentrations of about 200 ng/ μ l and 40 ng/ μ l by dilution in water. Template concentrations that yield reproducible fingerprints depend on RNA quality and to some extent on the choice and quality of the primers and must therefore be determined empirically.

4. Reactions are assembled on ice. Add 10 μ l of 2 \times First-Strand Reaction Mixture to 10 μ l of RNA at each concentration. The reaction is ramped to 37°C over 5 min in a 96-well format thermocycler (Perkin-Elmer, CT) and held at that temperature for an additional 10 min followed by 94°C for 2 min to inactivate the polymerase and finally cooling to 4°C.

5. Add 20 μ l of 2 \times Second-Strand Reaction Mixture to each 20 μ l of first-strand synthesis reaction. Cycle through one low-stringency step (94°C, 1 min; 40°C, 5 min; 72°C 5 min) followed by 35 high-stringency steps (94°C, 1 min; 60°C, 1 min; 72°C, 1 min).

6. Four microliters of each reaction is diluted in 18 μ l of 95% formamide, heated to 94°C for 2 min, then 1.5 μ l is loaded on a 5% polyacrylamide–50% urea sequencing gel and electrophoresed in 1 \times TBE at 50 W for 3 hr or until the Xylene cyanol reaches three-quarters of the way to the bottom of the gel. The gel is dried, and autoradiography is performed with X-Omat film (Kodak) for 12 hr. Fragments visualized generally range from ~50 bases to 1000 bases.

Some primers work better than others. It is usually a good idea to screen several primers and use those that give the most qualitatively robust patterns for further work. Note that *Thermus thermophilus* (*Tth*) polymerase has reverse transcriptase activity and can substitute for both the RT and the DNA polymerase when a buffer containing some Mn^{2+} is used.

RNA Fingerprinting Using a Pair of 10-Base Primers Applied Sequentially

Our protocol for RNA fingerprinting with 10-base primers is similar to that described for longer primers; however, *Taq* Stoffel fragment is used instead of *Taq* polymerase.⁽²⁶⁾ This necessitates a dilution scheme that adjusts buffer components such that they are compatible with *Taq* Stoffel fragment.

2 \times First-Strand Reaction Mixture 100 mM Tris-HCl (pH 8.3), 100 mM KCl, 8 mM $MgCl_2$, 20 mM DTT, 0.4 mM each dNTP, 16 μ M primer, and 1 U/ μ l of MuLVRT, (Stratagene Inc., La Jolla, CA). Except for the primer concentration, this is the same as the first-strand reaction mixture for longer primers.

2 \times AP-PCR Reaction Mixture 20 mM Tris-HCl (pH 8.3), 20 mM KCl, 8 mM $MgCl_2$, 0.1 μ Ci/ μ l [α -³²P]dCTP, 0.2 U/ μ l *Taq* DNA polymerase (AmpliTaq; Perkin-Elmer Cetus), and 8 μ M secondary primer.

In this protocol, enough first-strand cDNA is prepared to allow for several second-strand reactions using different primers to be performed. The protocol has been designed using the spread-sheet Microsoft Excel, allowing automatic changes in the number of RNA preparations to be fingerprinted, RNA concentrations, and the number of first and second primers to be used, while maintaining the proper buffer components (Fig. 2) Here, we describe the case where one first-strand primer is used on six different RNA preparations at three concentrations, followed by four different secondary primers.

1. RNA is prepared as described above.
2. First-strand synthesis with a single primer is achieved as follows: Five microliters of total RNA at 200, 100, and 50 ng/ μ l is combined with 5 μ l First-Strand Reaction Mixture on ice, ramped to 37°C over 5 min, held at 37°C for 60 min, heated to 94°C for 2 min to inactivate the reverse transcriptase, and cooled to 4°C until the next step.
3. The reaction mixture is then diluted with water to four times its initial volume. In this case, the 10 μ l of reaction is diluted to 40 μ l by the addition of 30 μ l of water. This dilutes buffer components to 12.5 mM Tris (pH 8.3), 12.5 mM KCl, 1 mM MgCl₂, 50 μ M dNTP, and 4 μ M initial primer. This concentration of primer is sufficient to sustain PCR amplification in subsequent steps.
4. Ten microliters of the diluted first-strand reaction is combined with 10 μ l of 2 \times AP-PCR Reaction Mixture, which contains the secondary primer. At this point, the reaction components are as follows: 16.25 mM Tris (pH 8.3), 16.25 mM KCl, 4.5 mM MgCl₂, 0.23 mM dNTPs, 2 μ M first primer, 4 μ M second primer, and 0.2 U/ μ l *Taq* polymerase Stoffel fragment. The reactions are cycled through 45 cycles of the following temperature profile: 94°C for 30 sec, 40°C for 1 min, and 72°C for 2 min.
5. The reactions are analyzed on sequencing gels as described above.

There are at least two good reasons to use this latter protocol. First, the number of visible products in the fingerprints is increased relative to using an 18-mer or one 10-mer primer. Furthermore, those products that contain the first and second primer are likely to be oriented with the first primer at the 3' end and the second primer at the 5' end, thereby providing information about the sense strand of the original RNA. More than 50% of products in reactions using pairwise combinations of primers contain both primers and are distinct from those generated with either primer individually.⁽²⁷⁾

PREPARATION OF CLONES

Differentially amplified RAP-PCR products are cut from the gel and eluted into 50 μ l of TE for 2 hr at 65°C. A 5- μ l aliquot is then used to reamplify the RAP-PCR products employing the same oligonucleotide primers used to generate the fingerprint.⁽⁸⁾ *Taq* polymerase often adds an extra base to the 3' end of PCR products. Which base is added depends on the last base of the double-stranded sequence. For example, if the last base is an A, then an extra A is added. If the last base is a T, then it is removed and replaced with an A to produce an A:A mismatch. This phenomenon makes PCR products hard to blunt-end clone. We use the simple expedient of reamplifying polymorphic products using *Pfu* polymerase, which does not add extra bases to the 3' end of PCR products. *Vent* polymerase is similar to *Pfu* in this property.⁽²⁸⁾

The RAP-PCR products are cloned (e.g., into the pCR-Script vector, a modified form of the Bluescript vector, Stratagene Inc., San Diego, CA) using standard protocols. *EcoRV* or *SrfI* is used during the ligation to linearize vectors that have closed on themselves, thereby improving the efficiency. For each RAP-PCR product, six or more independently isolated single-stranded phagemids are sequenced using the Sequenase reagent kit v2 (U.S. Biochemical, Cleveland, OH) and [α -³²S]dATP (NEN Research Products, Boston, MA).

IDENTIFICATION OF THE CORRECT CLONE

A common difficulty with cloning arbitrarily primed PCR or RAP-PCR products is contamination of the isolated PCR product with unwanted comigrating species. In principle, a secondary gel purification step, for example, SSCP, could be included after the first PCR amplification to further enrich for the

correct product. However, after reamplification and cloning into the pBlue-Script plasmid, it is generally sufficient to sequence six clones in order to identify one that is represented twice or more. Alternatively, complementation tests (c-tests) can be performed on several independent clones. The most abundant product of the secondary PCR amplification is more likely than low-level background contamination to c-test positively. We have shown that confirmation of the correct product can be achieved by Southern transferring the RAP-PCR gel, or another gel produced by the same primer, and using the putative clone as a probe^(9,29) (Fig. 3). These Southern blots are much simpler than typical Southern blots. They have exposure times of only a few minutes because of the low complexity of the material on the gel. We have tested >20 products in this way. Only two clones that were prevalent in the sequencing proved to be incorrect when tested against Southern blots of the RAP-PCR gels.

SOUTHERN BLOTS

DNA from RAP-PCR fingerprints are transferred to a membrane overnight (Duralon-UV, Stratagene, La Jolla, CA) by capillary action using 20× SSC buffer and UV cross-linked. Blots are prehybridized with 5× SSC, 0.5% blocking reagent (Boehringer Mannheim Biochemical), and 0.8% SDS at 65°C for 4 hr. Hybridization to blots is performed using conventional methods⁽²⁵⁾ and conditions recommended by the vendors. Radiolabeled probes can be synthesized by PCR from cloned RAP-PCR products using the T7-T3 oligonucleotide primer set (Stratagene Inc., San Diego, CA), and [α -³²P]dCTP.

When clones are then used on Northern blots, a minority give no signal because the message is too rare but give positive Southern blots, confirming probe quality. However, the subset that give positive Northern blots generally also displayed the expected differential gene expression.

FURTHER IMPROVEMENTS

The RAP-PCR method is amenable to various possible improvements and, given the previously demonstrated utility of RAP-PCR, it is worth developing

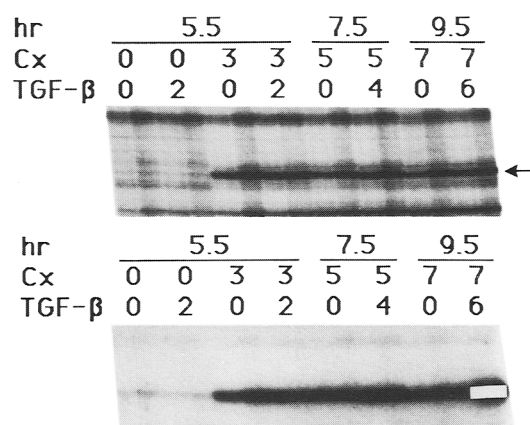


FIGURE 3 A product corresponding to a cycloheximide (CX)-inducible gene was excised from a RAP-PCR gel, reamplified, and cloned. The original RNA fingerprint was transferred to a nylon filter, and the filter was probed with the clone to verify that the correct band had been cloned. The *top* panel shows a close-up of the appropriate region from the RAP-PCR gel. The *bottom* panel shows the same region transferred to a membrane and probed with a clone from the differentially amplified band. The clone hybridized to the correct region and had the expected differential distribution between lanes, indicating that the clone represented the correct differentially amplified band.

some of these possibilities. For example, in principle, *Tth* DNA polymerase could be used as both the reverse transcriptase and the DNA polymerase,⁽³⁰⁾ reducing the number of steps. Conditions exist that are compatible with both properties of the enzyme.⁽³¹⁾ Alternatively, a viral reverse transcriptase and *Taq* DNA polymerase can be mixed from the beginning. These enzymes also have different optima, but it should be possible to find a compatible buffer system.

SAMPLING GENES THAT CARRY CONSERVED SEQUENCE MOTIFS

One possibility for further adaptation of RAP-PCR is to encode conserved motif sequences in the primers. Then, one could search for differentially expressed genes using motifs directed toward particular gene families of relevance to a particular biological phenomenon. Thus, one could attempt to bias in favor of zinc fingers, as we did with the primer ZF-1,⁽¹¹⁾ and protein kinase motif primers as shown in Figure 1. Such efforts are an extension of the work with motifs that have been tried for fingerprinting total genomic DNA, such as promoter and amino acid motifs⁽³²⁾, or hypervariable purine-pyrimidine repeats.⁽⁸⁾ Although the latter strategy does generate some products derived from RY repeat regions, the use of protein motifs is more difficult. RY repeats, such as GT repeats,⁽³³⁾ are much more common than protein sequence motifs. Rarer motifs must compete with a larger number of arbitrary priming events. It should be reiterated that increasing the stringency during the initial arbitrarily primed PCR generally reduces the reliability of the fingerprint. Furthermore, the consequence of wobble bases in the translation of a conserved amino acid motif is that a primer that has no redundant positions will match perfectly with only a small fraction of the nucleic acid sequences encoding the motif. Nevertheless, this strategy is well worth exploring, and it is very likely that a series of papers using primers that carry motifs will appear soon, each with their own acronym.

COMPARISON OF RNA FINGERPRINTING PROTOCOLS

What are the advantages and disadvantages of the RAP-PCR protocol⁽⁴⁾ compared with that of Liang and Pardee?⁽³⁾ In contrast to an arbitrary selection step in both directions,^(1,4) the protocol in Reference 3 uses an oligo(dT)-CA (or similar oligo(dT)-XM) primer for reverse transcription followed by an arbitrary primer in the other direction. The method samples 3' ends that are mostly noncoding. In contrast, arbitrary priming from both ends can sample open reading frames (ORFs) in ~30% of mRNA products^(4,7,11) because the authentic reading frame will often occur in arbitrarily selected stretches of a few hundred bases in a typical mRNA sequence. Thus, our strategy has the advantage that conserved ORFs between species or protein family members can occasionally be observed in data base searches (e.g., Ref. 11).

One variant of the protocol we use can be designed to arbitrarily prime only once after denaturation before switching to high-stringency annealing (e.g., 60°C for an 18-base primer). Any contaminating dsDNA is usually primed only once and is therefore not amplified during PCR.⁽⁴⁾ The protocols employing 10-mers or an anchored oligo(dT)NN primer use 35°C annealing steps throughout and can sample contaminating dsDNA in the classic manner of RAPD.⁽²⁾ Finally, the oligo(dT)-XM primers are as highly promiscuous as other primers, if not more so, and must result in many products that are sampled from hnRNA or inside mRNAs and other products in which the oligo(dT)-CA primer occurs at both ends. To ensure sampling of only mRNA and not hnRNA or residual genomic DNA, the RNA must first be poly(A) selected.

Each of these approaches has strengths and weaknesses. Although the pro-

tocol using an anchored oligo(dT)NN primer is excellent, arbitrary priming from both directions, followed by nesting, can have significant advantages, not least of which is the sampling of ORFs.

ACKNOWLEDGMENTS

We thank Rhonda Honeycutt for helpful comments on the manuscript. This work was supported in part by National Institutes of Health grants AI34829 and HG00456 to M.M. and AI32644 to J.W.

REFERENCES

1. Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **1**: 7213–7218.
2. Williams, J.G., M.K. Hanafey, J.A. Rafalski, and S.V. Tingey. 1993. Genetic analysis using random amplified polymorphic DNA markers. *Methods Enzymol.* **218**: 704–740.
3. Liang, P. and A. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**: 967–971.
4. Welsh, J., K. Chada, S.S. Dalal, D. Ralph, R. Cheng, and M. McClelland. 1992. Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res.* **20**: 4965–4970.
5. McClelland, M. and J. Welsh. 1994. DNA fingerprinting by arbitrarily primed PCR. *PCR Methods Applic.* (this issue)
6. McClelland, M. and J. Welsh. Arbitrarily primed polymerase chain reaction method for fingerprinting. Patent application filed October, 1990.
7. Wong, K.K., C.-H. Mok, J. Welsh, M. McClelland, S.-W. Tsao, and R.S. Berkowitz. 1993. Identification of differentially expressed RNA in human ovarian carcinoma cells by arbitrarily primed PCR fingerprinting of total RNAs. *Int. J. Oncol.* **3**: 13–17.
8. Welsh, J., C. Petersen, and M. McClelland. 1991. Polymorphisms generated by arbitrarily primed PCR in the mouse: Application to strain identification and genetic mapping. *Nucleic Acids Res.* **19**: 303–306.
9. Peinado, M.A., S. Malkhosyan, A. Velazquez, and M. Perucho. 1992. Isolation and characterization of allelic losses and gains in colorectal tumors by arbitrarily primed polymerase chain reaction. *Proc. Natl. Acad. Sci.* **89**: 10065–10069.
10. McClelland M., K. Chada, J. Welsh, and D. Ralph. 1993. Arbitrary primed PCR fingerprinting of RNA applied to mapping differentially expressed genes. In *DNA fingerprinting: State of the science. Symposium: November 9th (1992)* (ed. S.D. Pena, R. Charkraborty, J.T. Epplen, and A.J. Jeffereys). Birkhauser Verlag, Basel, Switzerland.
11. Ralph, D., J. Welsh, and M. McClelland. 1993. RNA fingerprinting using arbitrarily primed PCR identifies differentially regulated RNAs in Mink lung (Mv1Lu) cells growth arrested by TGF β . *Proc. Natl. Acad. Sci.* **90**: 10710–10714.
12. Kwok, S., D.E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson, and J.J. Sninsky. 1990. Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* **18**: 999–1005.
13. Bauer, D., H. Muller, J. Reich, H. Riedel, V. Ahrenkiel, P. Warthoe, and M. Strauss. 1993. Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Res.* **21**: 4272–4280.
14. Watson, J.B., E.F. Battenberg, K.K. Wong, F.E. Bloom, and J.G. Sutcliffe. 1990. Subtractive complementary DNA cloning of RC3, a rodent cortex-enriched mRNA encoding a novel 78 residue protein. *J. Neurosci. Res.* **26**: 397–408.
15. Kuo, C.H., K. Yamagata, R.K. Moyzis, M.W. Bitensky, and N. Miki. 1986. Multiple opsin mRNA species in bovine retina. *Brain Res.* **387**: 251–260.
16. An, G., T.H. Huang, J. Tesfaigzi, J. Garcia Heras, D.H. Ledbetter, D. M. Carlson, and R. Wu. 1992. An unusual expression of a squamous cell marker, small proline-rich protein gene, in tracheobronchial epithelium: Differential regulation and gene mapping. *Am. J. Respir. Cell Mol. Biol.* **7**: 104–111.
17. Brunet, J.F., E. Shapiro, S.A. Foster, E.R. Kandel, and Y. Iino. 1990. Identification of a peptide specific for *Aplysia* sensory neurons by PCR-based differential screening. *Science* **252**: 856–859.
18. Smith, D.E. and T. Grindley. 1992. Differential screening of a PCR-generated mouse embryo cDNA library: Glucose transporters are differentially expressed in early post implantation mouse embryos. *Development* **116**: 555–561.
19. Sargent, T.D. 1987. Isolation of differentially expressed genes. *Methods Enzymol.* **152**: 423–432.
20. Lisitsyn, N., N. Lisitsyn, and M. Wigler. 1993. Cloning the differences between two complex genomes. *Science* **259**: 946–951.

21. Celis, J.E., H.H. Rasmussen, P. Madsen, H. Leffers, B. Honore, K. Dejgaard, P. Gromov, and H.-J. Hoffmann. 1993. The human keratinocyte two-dimensional gel protein database (update 1993) *Electrophoresis* **14**: 1091–1198.
22. Rasmussen, H.H., J. van Damme, M. Puype, B. Gesser, J.E. Celis, and J. Vandekerckhove. 1992. Microsequences of 145 proteins recorded in the two-dimensional gel protein database of normal human epidermal keratinocytes. *Electrophoresis* **13**: 960–969.
23. Davidson, E.H. and R.J. Britten. 1973. Organization, transcription and regulation in the animal genome. *Quart. Rev. Biol.* **48**: 565–613.
24. Chirgwin, J., A. Przybyla, R. MacDonald, and W.J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**: 5294–5299.
25. Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
26. Sobral, B.W.S and R.J. Honeycutt. 1993. High output genetic mapping of polyploids using PCR-generated markers. *Theor. Appl. Genet.* **86**: 105–112.
27. Welsh, J. and M. McClelland. 1991. Genomic fingerprinting with AP-PCR using pairwise combinations of primers: Application to genetic mapping of the mouse. *Nucleic Acids Res.* **19**: 5275–5279.
28. Garrity, P.A. and B.J. Wold. 1992. Effects of different DNA polymerases in ligation-mediated PCR: Enhanced genomic sequencing and in vivo footprinting. *Proc. Natl. Acad. Sci.* **89**: 1021–1025.
29. Wong, K.K. and M. McClelland. 1992. A *BlnI* restriction map of *Salmonella typhimurium*. *J. Bacteriol.* **174**: 1656–1661.
30. Myers, T.W. and D.H. Gelfand. 1991. Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry* **30**: 7661–7666.
31. Young, K.K., R.M. Resnick, and T.W. Myers. 1993. Detection of hepatitis C virus RNA by a combined reverse transcription-polymerase chain reaction assay. *J. Clin. Microbiol.* **31**: 882–886.
32. Birkenmeier, E.H., U. Schneider, and S.J. Thurston. 1992. Fingerprinting genomes by use of PCR primers that encode protein motifs or contain sequences that regulate gene expression. *Mamm. Genome* **3**: 537–545.
33. Stallings, R.L., A.F. Ford, D. Nelson, D.C. Torney, C.E. Hildebrand, and R.K. Moyzis. 1991. Evolution and distribution of (GT)_n repetitive sequences in mammalian genomes. *Genomics* **10**: 807–815.