Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 6831–6836, June 1997
Genetics

Selective amplification via biotin- and restriction-mediated enrichment (SABRE), a novel selective amplification procedure for detection of differentially expressed mRNAs

(circadian expression/competitive hybridization/coumarin 7-hydroxylase/magnetic separation/streptavidin-biotin affinity)

DANIEL J. LAVERY, LUIS LOPEZ-MOLINA, FABIENNE FLEURY-OLELA, AND UELI SCHIBLER*

Department of Molecular Biology, University of Geneva, 30, quai Ernest-Ansermet, CH1211 Geneva 4, Switzerland

Communicated by Walter J. Gehring, University of Basel, Basel, Switzerland, April 18, 1997 (received for review March 17, 1997)

ABSTRACT **We present a novel subtractive enrichment protocol for the identification of differentially expressed mRNA species. This procedure, called SABRE (selective amplification via biotin- and restriction-mediated enrichment), uses selective streptavidin-biotin affinity and restriction enzyme site reconstitution to enrich for cDNA species more abundant in one population than in another. Analysis of liver cDNA from a mouse strain expressing the neomycin resistance gene demonstrated that this procedure is capable of identifying species present in one population but absent from another. Furthermore, experiments to identify genes with circadian expression patterns in mouse liver demonstrated that SABRE is capable of detecting even modest 2- to 10-fold differences in accumulation of moderately rare mRNA species, representing as little as 0.03% of total mRNA. These experiments identified the gene encoding coumarin 7-hydroxylase as displaying circadian expression in mouse liver.**

In many biological systems, significant changes in phenotype can result from only modest changes in the expression of certain genes (see, for example, refs. 1 and 2). However, identifying these genes in a complex mRNA population may be difficult, especially if their level of expression is relatively low. One such example is the class of genes displaying circadian expression in the rodent liver. Gene products with circadian expression patterns are likely to accumulate to low levels, given their relatively short half-lives (3), and often show only modest differences (10-fold or less) in their highest and lowest levels of accumulation (4, 5). As more than half of all mRNA in the rat liver is encoded by fewer than 300 genes, with the remainder encoded by greater than 20,000 genes (6), the detection of genes with circadian expression in the liver can be technically challenging.

In recent years, several approaches have been used to identify differentially expressed genes. These include nonselective protocols such as differential display and serial analysis of gene expression, which identify differentially expressed species by gel analysis of PCR products or sequencing of cDNA 3′ ends, respectively (7, 8). However, these are often time- and labor-intensive and may not efficiently detect low-abundance species. In contrast, selective protocols such as representational difference analysis (9, 10) and enzymatic degrading subtraction (11), specifically increase the abundance of sequences overexpressed in one population relative to another. In these latter protocols, a mixture of two populations of double-stranded DNA are denatured and allowed to rehybridize. A DNA species overexpressed in one of the populations is more likely to form hybrids of which both DNA strands are derived from that population (homohybrids) than is a species equally expressed in both populations. Thus, specific purification of this subpopulation of homohybrids will result in the enrichment of differentially expressed species (9). While these protocols may be sufficient for the detection of relatively abundant transcripts with large differences in expression, they do not include quantitative assignment of their sensitivity limits with regard to mRNA abundance and amplitude of differential expression.

We present here a novel protocol that uses this selective enrichment principle but uses a novel combination of biotin-streptavidin affinity and restriction enzyme site reconstitution to purify the desired homohybrid subpopulation. The procedure, called SABRE (for selective amplification via biotin- and restriction-mediated enrichment), has allowed the identification of differentially expressed cDNA species whose mRNAs display low to moderate abundance (approximately 0.03% of mRNA) with differences in expression of 10-fold or less between two populations. Comparison of mouse liver cDNA from evening and morning using SABRE has identified the gene encoding coumarin 7-hydroxylase as displaying a circadian expression pattern.

MATERIALS AND METHODS

Construction and PCR Amplification of cDNA Libraries. For selection of cDNA species displaying circadian expression, total RNA was isolated as described (12) from livers of 3-month-old male BALB/c mice kept in darkness for 3 days and fasted for 24 hr, at hours of the day as indicated. For selection of cDNA species displaying absolute differences between two populations, a mouse strain was used that carries a mutant allele of the gene encoding the transcription factor DBP (L.L.-M., F. Conquet, M. Dubois-Dauphin, and U.S., unpublished work). This mutant allele was generated by insertion of a gene cassette expressing the bacterial neomycin resistance gene (*neo*) into the DBP locus. For these experiments, total RNA was isolated by the guanidine thiocyanate-organic extraction method (13) from livers of wild-type C57BL/6 × 129/Ola mice, or C57BL/6 × 129/Ola littermates homozygous for the disrupted DBP allele. Polyadenylylated RNA was twice selected by oligo(dT) chromatography (14). Double-stranded cDNA was synthesized according to Gubler (15), using RNaseH-minus Moloney murine leukemia virus reverse transcriptase (BRL). One microgram of cDNA was digested with restriction enzyme *Sau*3AI and ligated to a 20-fold molar excess of the double-stranded DNA oligonucleotide linker: 5′-GATCCCGAGTACAAGGATATCCG-3′ and 3′-GGCTCATGTTCCTATAGGCGGTGTG-5′.

Ligation products (150 to 1,500 base pairs in size) were separated from linkers by electrophoresis on a 2% low-melt agarose gel. DNA was recovered from excised gel pieces by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/946831-6$2.00/0

Abbreviations: SABRE, selective amplification via biotin- and restriction-mediated enrichment; FLC, ferritin light chain.

*To whom reprint requests should be addressed.

high-salt/phenol extraction, precipitated with ethanol, and resuspended in 1 ml of 10 mM Tris, pH 7.4/0.1 mM EDTA.

The cDNA fragment libraries were amplified by PCR using 5 μl of the cDNA library per reaction in a 50-μl reaction containing 10 mM Tris at pH 8.9, 50 mM KCl, 1.5 mM $MgCl_2$, 10% dimethyl sulfoxide, 0.4 mM dNTPs, 1 μCi [α-^{32}P]dATP, 2 μM oligonucleotide primer, and 2 units of *Taq* polymerase in a Stratagene Robocycler PCR machine, using a 60°C annealing temperature and 25–30 rounds of amplification. The following oligonucleotides were used as primers for PCR (*Eco*RI restriction sites are underlined): Primer T1, 5′-biotin-GTGTGGCGGAATTCCTTGTACTCGGGATC-3′; primer D1, 5′-GTGTGGCGGATATCCTTGTACTCGGGATC-3′; primer T2, 5′-biotin-CACACCGCGAATTCCTTGTACT-CGGGATC-3′; and primer D2, 5′-CACACCGCGATATC-CTTGTACTCGGGATC-3′.

Biotinylated oligonucleotides (Microsynth, Balgach, Switzerland) contain a biotin group coupled by a C9 linkage to the 5′ terminal base. For odd-numbered rounds of selection, oligonucleotides T1 and D1 were used to amplify the tester and driver DNA populations, respectively, while for even-numbered rounds, oligonucleotides T2 and D2 were used (see *Results*). After amplification, products were purified from free primers and nucleotides by organic extraction and precipitation with 0.65 vol of isopropyl alcohol in the presence of 0.3 M sodium acetate. DNA yields were calculated from incorporation of radionucleotides.

Hybridization and Selection. For each experiment, two hybridizations were performed. In the experimental hybridization, DNA amplified from the population of interest (tester) using the biotinylated, restriction site-containing PCR primer (T1 or T2) was combined with DNA amplified from the control population (driver) using the nonbiotinylated, mutant site-containing PCR primer (D1 or D2). In a control hybridization, DNA amplified from the control population using the biotinylated tester primer was combined with nonbiotinylated driver DNA also amplified from the control population. Products of this control hybridization were compared with experimental hybridization selection products and used as the source of driver material for the next round of selection. This is important, as small differences in amplification efficiency between two cDNA species in a population may result in large differences in accumulation after several rounds of PCR-based selection (see *Discussion*).

Driver DNA and tester DNA were combined at a ratio of 20:1 (typically, 20–40 μg of driver DNA and 1–2 μg of tester DNA), precipitated, and resuspended in 0.1 ml of 80% formamide in 10 mM Tris, pH 8.5. The samples were denatured by heating to 75°C for 10 min, diluted to 200 μl, and precipitated with ethanol. This denaturation at low DNA concentration is included to ensure complete reassortment of single strands. The samples were resuspended to 5 mg/ml DNA in 50% formamide/1 M NaCl/10 mM Tris, pH 8.5/5 mM EDTA. Samples were next transferred to 25-μl glass capillaries; these were sealed with a flame and incubated for 7 days at 52°C. Under such conditions, C_0t values approaching 4×10^3 M·sec can be attained (16).

After hybridization, samples were recovered from the capillaries by dilution in mung bean nuclease buffer (30 mM sodium acetate, pH 5.0/100 mM NaCl/1 mM zinc acetate/5% glycerol) to a final concentration of 0.1 mg/ml DNA. Ten micrograms of hybrids was digested with 1 unit/μg mung bean nuclease (Amersham) for 30 min at 30°C. After organic extraction and precipitation, the samples were resuspended in 100 μl of 10 mM Tris, pH 8.0/1 mM EDTA/0.1 M NaCl.

The nuclease-resistant hybrids were incubated with 50 μl of streptavidin-coupled paramagnetic beads (M-280, binding capacity approximately 400 pmol of cDNA/ml; Dynal, Oslo, Norway) in 0.4 ml 10 mM Tris, pH 8.0/1 mM EDTA/1 M NaCl/0.1% Triton X-100 (1 M TENT) by gentle rotation at room temperature for 2 hr. We have found that contamination due to nonspecific interactions with the paramagnetic beads can be reduced by the inclusion of heparin sulfate (1 mg/ml) in the incubation (data not shown). Streptavidin bead-DNA complexes were washed by magnetic concentration (Dynal magnetic concentrator; Dynal, Finland) and resuspension in three changes of 0.4 ml of 1 M TENT, followed by incubation twice in 0.4 ml of 100 mM NaCl/10 mM Tris, pH 8.0/1 mM EDTA/0.1% Triton X-100. Greater stringency can be obtained if the 100 mM NaCl washes are done instead with 50 mM NaCl for 10 min at 60°C (data not shown). The complexes were then incubated twice in 0.4 ml of 100 mM NaCl/50 mM Tris·HCl, pH 7.5/10 mM $MgCl_2$/1 mM DTT (*Eco*RI buffer) at 37°C for 30 min each.

After magnetic concentration, the beads were resuspended in 30 μl of *Eco*RI buffer, and 10 units of restriction enzyme *Eco*RI (Amersham) were added. Digestions were incubated at 37°C for 1 hr, with gentle agitation every 10 min to avoid settling of the magnetic beads. The volume was then increased to 200 μl with *Eco*RI buffer, and two sequential magnetic separations were performed to eliminate the beads from the supernatant. The purified supernatant was extracted with phenol/chloroform and amplified by PCR (5 μl per 50-μl reaction) with an alternate set of primers for subsequent analysis (see *Results*). If required, the selection was repeated, using the products of the control hybridization (driver vs. driver) as the source of driver DNA. Radiolabeled products were displayed on 5% polyacrylamide/8 M urea gels to assess enrichment of specific species in the experimental hybridization relative to the control hybridization.

Cloning and Analysis of Candidate cDNA Fragments. The region of the sequencing gel containing the enriched band was excised, either while the gel was still wet or after drying. As a control, the same region of the control sample lane was also isolated. Material from these gel slices was extracted by boiling (in the case of dried gels; ref. 17) or elution in 10 mM Tris, pH 7.4/1 mM EDTA for 4 to 12 hr at room temperature (for wet gels). Extracted material was amplified by PCR as above, and the amplified material was digested with *Eco*RI, subcloned into a plasmid pBluescript-KS vector linearized with *Eco*RI, and used to transform competent *Escherichia coli*.

To identify plasmids containing the enriched cDNA fragments, diagnostic restriction digests with frequent cutting enzymes were used, when possible. Otherwise, bacterial clones were transferred in duplicate onto nitrocellulose filters, denatured *in situ* by alkaline-SDS lysis, and probed with PCR-radiolabeled DNA from the experimental or control selected populations (18). Clones that gave a greater signal with the experimental probe than with the control probe were sequenced, and their sequences were compared with sequence databases. For analysis of circadian expression, the cDNA fragments were used to generate riboprobes for use in RNase protection analyses (12) with total or poly$(A)^+$ mouse liver RNA. For coumarin 7-hydroxylase, the probe comprised cDNA sequences from nucleotide 1,198 (*Nco*I restriction site) to nucleotide 1,546 (*Sau*3AI restriction site), while for species A, a 180-bp *Sau*3AI-SpeI probe was used.

Enrichment of cDNA fragments was detected during selection rounds by Southern blot analysis of the starting libraries and the products of subsequent rounds of SABRE selection. One microgram of each sample was separated on a 5% sequencing gel, and electrotransferred to nylon membranes (Nytran, Schleicher & Schuell) essentially as described (19), except that transfers were conducted for 20 min rather than 2 hr. Probing of the blot with the coumarin 7-hydroxylase riboprobe was performed as described (20), except that hybridizations were done at 55°C. Signals were quantitated using a Bio-Rad GS-250 PhosphorImager system.

RESULTS

Experimental Design. Species more abundant in one double-stranded DNA population (the tester) than in another (the driver) can be enriched relative to species equally expressed in both by denaturing a mixture of both populations and allowing it to reassociate, and then isolating only double-stranded molecules of which both strands are derived from the tester population (tester homohybrids; Fig. 1). This population is enriched for species with greater abundance in the tester than in the driver because more complementary DNA strands are contributed by the tester population than the driver population, relative to those of species equally expressed in both populations (see refs. 9 and 21 for review).

The protocol for specific purification of tester homohybrids used in SABRE is outlined in Fig. 1. PCR amplification is used to generate driver and tester populations using PCR primers that are nearly identical, but have two significant differences: the tester PCR primer is biotinylated at the 5′ end and contains the recognition site for the restriction enzyme *Eco*RI, 5′-GAATTC-3′, near its 5′ end. In contrast, the driver PCR primer is nonbiotinylated and contains a mutation of the *Eco*RI site, 5′-GATATC-3′. After extensive hybridization of the driver and tester populations with driver DNA in excess, the hybrids are digested with mung bean nuclease to remove single-stranded molecules. Biotin-containing tester homohybrids and driver-tester heterohybrids are next captured by incubation with streptavidin-coated magnetic beads. The high avidity of the streptavidin–biotin interaction ($K_d = 1 \times 10^{-15}$ M) permits stringent washing of the bead-DNA complex to reduce contamination by nonbiotinylated driver DNA. Thus, using the stringent washing conditions described in *Materials and Methods*, contamination of tester DNA by driver DNA amounts to less than one part in 1×10^6 (data not shown).

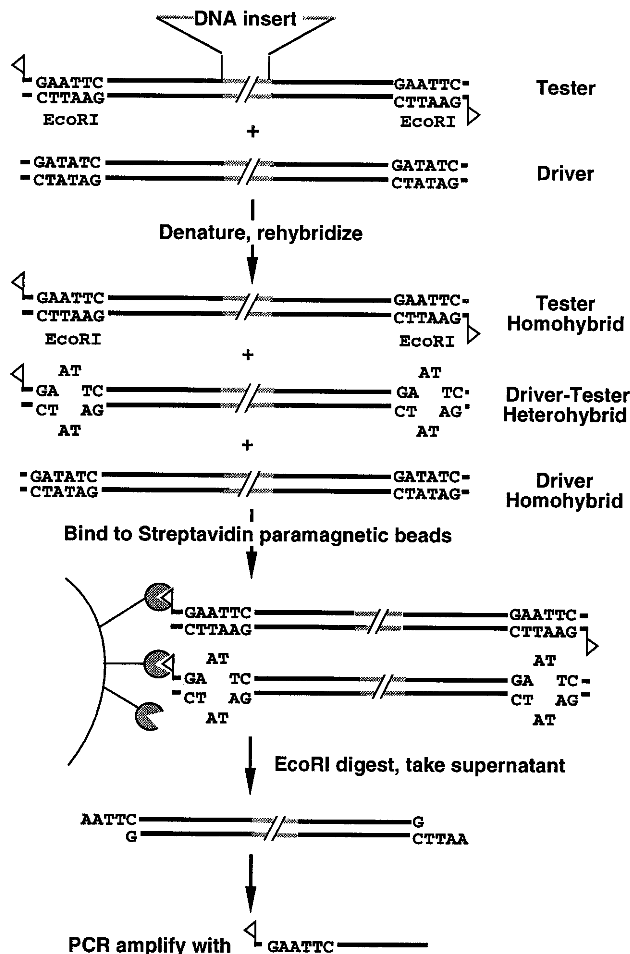

FIG. 1. Selection of tester homohybrid species by SABRE. Tester DNA, biotinylated at each 5′ end (open triangle) and with an *Eco*RI restriction site in each primer sequence, is hybridized with an excess of driver DNA, which is unbiotinylated and contains a mutated *Eco*RI site in the primer sequence. Three types of double-stranded molecules are formed: tester homohybrids, driver homohybrids, and driver–tester heterohybrids. Tester homohybrids and driver–tester heterohybrids are purified by capture of biotin-containing hybrids on streptavidin-coated paramagnetic beads. Next, digestion with *Eco*RI specifically releases the tester homohybrids from the beads. The released molecules are then reamplified by PCR for further analysis and selection.

The bead-DNA complexes are next incubated with the restriction enzyme *Eco*RI, which recognizes and cuts the double-stranded palindrome 5′-GAATTC-3′ present in the tester homohybrid primer sequences, releasing them from the beads. In contrast, the driver–tester heterohybrids are not cleaved by the enzyme, as their primer sequences form a heteroduplex of the wild-type and mutant *Eco*RI recognition sequences, which is not efficiently digested by the enzyme under the conditions used (22), and thus remain bound to the beads. Tester homohybrids released by digestion are subsequently amplified by PCR for another round of selection, or for isolation and subcloning of enriched species. For the subsequent selection round, alternate primers are used (for instance, T2 and D2 rather than T1 and D1), which differ from the previous pair of primers in the sequence 5′ to the *Eco*RI site. This is to avoid preferable amplification of contaminating species that would contain the entire PCR primer sequence, and thus would be more efficiently amplified than *Eco*RI-digested tester homohybrids.

The theoretical enrichment of a tester DNA fragment after one round of selection by this approach can be determined by calculating the probability of random association of two tester strands in a mixture of the driver and tester populations: In a hybridization reaction in which driver DNA is present at an R-fold excess over tester DNA, for a species X, which is expressed with a concentration of X in the tester but N-fold more in the tester population than in the driver, this probability is $X/(1 + R/N)$. For a species Y with a concentration Y equal in both driver and tester, this probability is $Y/(1 + R)$. Therefore, the ratio of X to Y, which was X/Y in the starting population, will change after one round of selection using an R-fold excess of driver to tester by the factor $(1 + R)/(1 + R/N)$ (D.L. and L.L.-M., unpublished work). This equation predicts that the degree to which the concentration of a tester-enriched species can be increased will depend on both the amplitude difference in the tester (N), and the ratio of driver to tester in the hybridization (R). It is important to note that in theory this procedure will enrich even modest differences in accumulation (N). In fact, the sensitivity of the procedure will be influenced by many factors, including the completeness of hybridization and the relative efficiency of PCR amplification of a species. If the C_ot value required for nearly complete annealing of a given species is not reached, this species will be lost during the selection procedure. Likewise, if a particular species is not amplified by PCR as efficiently as other species, it will be progressively depleted from the population during multiple rounds of selection.

Identification of *Sau*3AI cDNA Fragments Enriched in Liver cDNA of DBP Knockout Mice. To test whether SABRE could detect "all or nothing" differences in gene expression, analysis was performed using as the tester population liver cDNA from a mouse strain homozygous for a disrupted allele of the transcription factor DBP (referred to as the "knockout" strain; L.L.-M., F. Conquet, M. Dubois-Dauphin, and U.S., unpublished work), with liver cDNA from the wild-type parental strain (C57BL/6 × 129/Ola) as the driver population. Thus, mRNA from the bacterial neomycin resistance (*neo*) gene used to disrupt the DBP gene is present exclusively in the liver of the knockout strain. *Sau*3AI-digested cDNA libraries

were prepared from livers of knockout or wild-type mice, and used to generate tester and driver populations by PCR. Note that in these starting populations, no significant differences in the abundance of *Sau*3AI cDNA fragments can be detected (Fig. 2, lanes 1, 2). Experimental hybridizations were assembled with tester DNA amplified from the knockout library and driver DNA amplified from the wild-type library, while the control hybridizations were assembled with wild-type tester DNA and wild-type driver DNA. After four rounds of selection with a driver-to-tester ratio of 20:1, three prominent species were enriched in the experimental (knockout tester vs. wild-type driver) but not control (wild-type tester vs. wild-type driver) population (Fig. 2, lanes 3, 4). These species were isolated from polyacrylamide gels, subcloned into plasmid vectors, and analyzed by sequencing and characterization of their expression in both wild-type and knockout mouse liver RNA. Enriched species A was identified as a *Sau*3AI restriction fragment of the *neo* cDNA. Thus the method was capable of selecting a cDNA fragment present in one population but absent from another.

In addition, two other cDNA fragments were identified that were enriched in the knockout population: a *Sau*3AI fragment of the retinol-binding protein (RBP) cDNA (species B); and a *Sau*3AI fragment of the ferritin light chain (FLC) cDNA (species C). RNase protection analysis indicated that accumulation of RBP mRNA was higher in poly$(A)^+$ liver RNA of the knockout strain than of the wild type (1.5- to 2-fold; data not shown). This relatively small difference was detectable early in the selection process due to the relatively high abundance of the RBP cDNA.

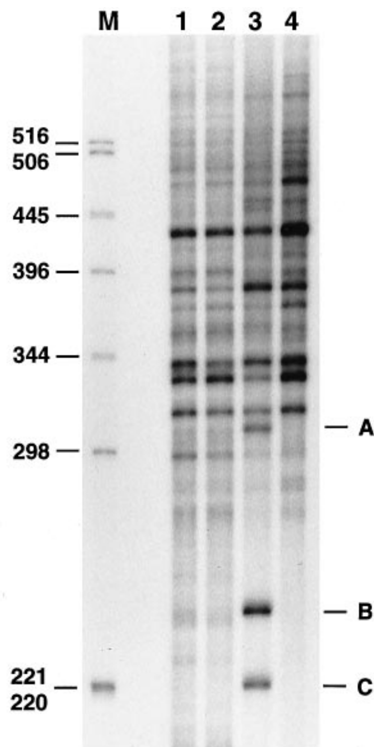

FIG. 2. Enrichment of cDNA fragments overexpressed in liver mRNA of DBP knockout mice. SABRE was performed on *Sau*3A-digested cDNA libraries from liver mRNA samples isolated at 8 p.m. from either homozygous DBP knockout (KO; lane 1) or wild-type mice (WT; lane 2), with KO cDNA as tester and WT cDNA as driver. Control hybridizations were performed with WT cDNA as tester and WT cDNA as driver. After four selection rounds, three prominent species enriched in the experimental KO vs. WT selection (lane 3) but not in the control WT vs. WT selection (lane 4) were identified as: A, *neo* gene cDNA fragment; B, retinol binding protein gene cDNA fragment; C, FLC gene cDNA fragment.

Analysis of the selected FLC *Sau*3AI cDNA fragment indicated that it represents a polymorphic allele of the FLC gene: a single base change has generated an additional *Sau*3AI restriction site in an FLC allele of the 129/Ola strain, used to generate the knockout mouse strain, which is not present in the C57BL/6 strain. RNase protection experiments demonstrated that the polymorphic allele was exclusively present in the pool of eight mice homozygous for the mutant DBP allele, and not in the wild-type littermates, suggesting that it is physically linked to the DBP locus (data not shown). Consistent with this, in humans both a FLC gene and the DBP gene have been localized to similar regions of chromosome 19 [19q13.1–19qter for FLC (23), 19q13.1–19q13.3 for DBP (24)]. As the driver cDNA population was derived from wild-type littermates of knockout mice resulting from crosses of 129/Ola × C57BL/6 F1 mice heterozygous for the DBP mutant allele, both of their FLC alleles were nonpolymorphic alleles derived from the C57BL/6 strain. Therefore, the additional *Sau*3AI FLC cDNA fragment was present in the tester cDNA population but not the driver cDNA population, leading to its enrichment during the selection procedure.

Identification of cDNA Species with Circadian Expression Patterns. A major interest in our laboratory is the study of molecular mechanisms controlling circadian gene expression (3). Therefore, the selection protocol was next applied to the identification of such genes in mouse liver. *Sau*3AI-digested cDNA libraries were generated from mRNA of BALB/c mouse livers harvested at 8 p.m. or 8 a.m., the peak and trough hours, respectively, of circadian cholesterol 7α-hydroxylase mRNA accumulation in mouse liver (L.L.-M., D.L., and U.S., unpublished work). Hybridizations were performed using 8 p.m. tester DNA and 8 p.m. driver DNA for experimental hybridizations, and 8 a.m. tester DNA and 8 a.m. driver DNA for control hybridizations. After four rounds of selection, several species that had been undetectable in the starting population were enriched in the experimental hybridization population (Fig. 3). Two species, A and B, were further analyzed by amplification from polyacrylamide gels, cloning into plasmids, and sequencing. Species A was subcloned and sequenced, but its sequence showed no significant homology with previous entries in sequence databases. In contrast, species B (Fig. 3*A*) was identified as a *Sau*3AI fragment of the cDNA encoding coumarin 7-hydroxylase, a liver-enriched member of the cytochrome P450 gene superfamily (refs. 25 and 26; nucleotides 1,095 to 1,445 in ref. 25). The enrichment of this cDNA fragment during the selection procedure is illustrated by Southern blot analysis (Fig. 3*B*) of the 8 a.m. and 8 p.m. starting libraries, as well as the products of each successive round of selection, either p.m. vs. a.m. (Exp) or a.m. vs. a.m. (Control). The coumarin 7-hydroxylase cDNA fragment, 4-fold more abundant in the 8 p.m. starting library than in the 8 a.m. starting library, was enriched by greater than 70-fold during the experimental selection (Exp), but not in the control selection (Control; Fig. 3*B*). The enrichment during round 4 was not as great as during previous rounds, perhaps because at this point the fragment had attained such a high concentration, representing a major species in the cDNA fragment population (Fig. 3), that its amplification by PCR was less efficient.

RNase protection analysis using poly$(A)^+$ BALB/c mouse liver RNA isolated at 9 a.m. or 8 p.m. demonstrated that probes derived from both the coumarin 7-hydroxylase and species A cDNA fragments detected mRNA species of the expected sizes, which displayed greater abundance at 8 p.m. than at 9 a.m. (Fig. 3*C*), with differences in accumulation of 3-fold (species A) and 6-fold (coumarin 7-hydroxylase). Thus, these enriched cDNA fragments indeed represent differences in the accumulation of their cognate RNA species between the morning and evening populations.

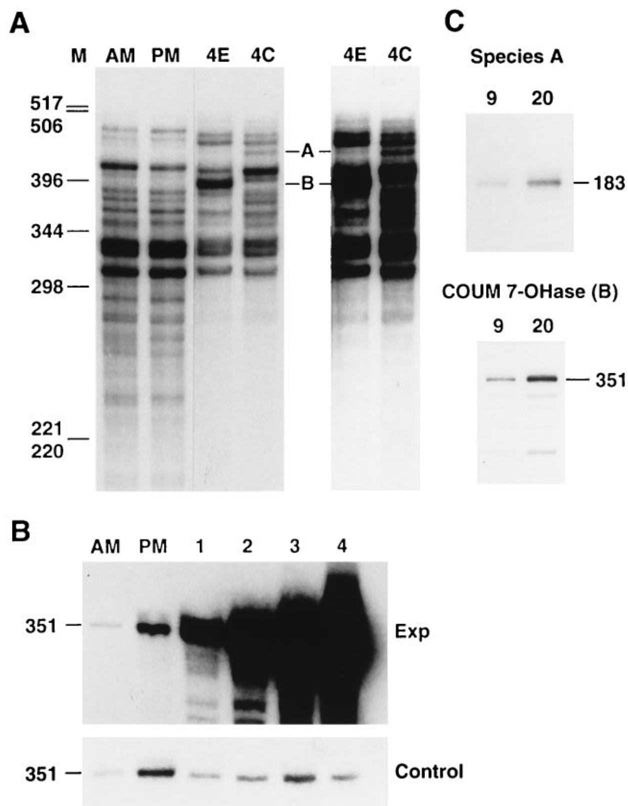

FIG. 3. Detection of evening-enriched mouse liver cDNA fragments. (*A*) *Sau3*A-digested cDNA libraries generated from mouse liver mRNA isolated at 8 a.m. (AM) or 8 p.m. (PM) were used for SABRE selection of evening-enriched species, using the PM tester vs. AM driver (4E), or the AM library as both tester and driver (4C). After four selection rounds, two species enriched in the experimental selection (4E) were isolated (A and B); A, cDNA fragment of unknown coding specificity; B, coumarin 7-hydroxylase gene cDNA fragment. Note that species A is distinct from the abundant species in 4C that migrates slightly faster than AM, size in bp of DNA molecular weight markers. (*Right*) Overexposure of lanes 4E and 4C to better detect species A. (*B*) Accumulation of coumarin 7-hydroxylase *Sau*3AI cDNA fragment in rounds of selection by SABRE. Southern blot analysis on PCR products of a.m. and p.m. libraries (AM, PM), and products of experimental (Exp) and control (Control) hybridization rounds (1 to 4, as indicated), using an RNA probe for the 351-bp coumarin 7-hydroxylase cDNA fragment. (*C*) Detection of coumarin 7-hydroxylase and species A mRNAs by RNase protection analysis of poly(A)$^+$ RNA from BALB/c mouse liver isolated at 9 a.m. (9) or 8 p.m. (20). Numbers at right refer to size in nucleotides of protected RNA fragments.

To determine whether coumarin 7-hydroxylase mRNA displayed circadian accumulation throughout the day, the coumarin 7-hydroxylase cDNA probe was used to detect its mRNA in total mouse liver RNA extracted at six time points around the clock. As demonstrated in Fig. 4 (left lanes), mRNA detected by the coumarin 7-hydroxylase cDNA probe demonstrated a circadian expression pattern in this mouse strain, with peak expression at approximately 10 p.m. (4 hr after the light-to-dark switch). In contrast, no significant change was found in accumulation of β-actin mRNA throughout the day (data not shown). In parallel, a standard curve titration was performed using known amounts of a synthetic coumarin 7-hydroxylase sense-strand "pseudo-RNA" to estimate coumarin 7-hydroxylase mRNA accumulation in mouse liver cells at different times of day. Comparison of the standard curve to the hybridization signal obtained with 10 μg of total mouse liver RNA indicated that peak levels of coumarin 7-hydroxylase mRNA represented approximately 300 copies per cell [0.03%, assuming 1×10^6 copies of mRNA and 50 pg of total RNA per liver cell (12)], with 6-fold greater accumulation in the evening than in the morning.

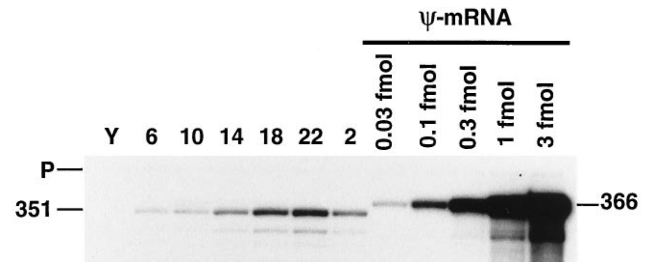

FIG. 4. Circadian accumulation of coumarin 7-hydroxylase mRNA. (*Left*) Ten micrograms of total liver mRNA isolated from C57BL/6 × 129/Ola mice at hours of the day indicated (6, 10, 14, 18, 22, 2) was analyzed by RNase protection for accumulation of coumarin 7-hydroxylase mRNA (351 nucleotides), using the same probe as in Fig. 3. Y, yeast RNA alone. (*Right*) In parallel, indicated amounts (3 to 0.03 fmol) of a synthetic RNA (ψ-RNA) corresponding to the coumarin 7-hydroxylase mRNA were analyzed by RNase protection with the same probe, generating a protected fragment of 366 nucleotides rather than 351, due to polylinker sequences. These signals were used to quantitate coumarin 7-hydroxylase mRNA accumulation in the mouse liver samples.

DISCUSSION

The SABRE subtractive cloning method is a sensitive and reliable method for the detection of even low-abundance DNA species whose expression demonstrates modest (2- to 10-fold) differences in accumulation between two populations. The efficiency of the two purification steps (biotin capture and release by restriction digestion) used to enrich the population for tester homohybrid molecules permits the purification of very small quantities of material with low levels of contamination from driver molecules (less than 1 part in 1 million). One drawback in the purification procedure is that digestion with *Eco*RI will result in the loss of cDNA fragments containing internal *Eco*RI sites. While this should occur rarely with libraries made from DNA digested with frequent-cutting enzymes, this problem can be avoided by using in the selection a restriction enzyme, such as *Bam*HI (GGATCC), whose recognition sequence is included in that of the enzyme used to generate the digested library, such as *Sau*3AI (GATC).

A further advantage of SABRE is that multiple rounds of selection can be readily performed, without extensive manipulation of the driver and tester material such as linker removal and addition. In selective procedures such as SABRE, representational difference analysis, and enzymatic degrading subtraction, multiple selection rounds are required to identify enriched species of low initial abundance, as even substantial enrichment after one selection round may not be sufficient to permit their identification and isolation away from more abundant species. In the SABRE protocol, selected material is simply reamplified by PCR and is used directly for the next selection round.

In addition, use of the products of the driver–driver self-subtraction as driver material for subsequent selection rounds reduces the incidence of false positive results, by suppressing the enrichment of species that are preferentially amplified by the PCR reaction. These changes occur because of intrinsic differences in the ability of any two DNA sequences to be amplified by the PCR reaction. Thus, as many sequential PCR rounds are often required, even relatively modest differences in PCR amplification efficiency can be magnified over the course of selection. Indeed, we have detected significant and reproducible changes in relative abundance of species after multiple rounds of PCR amplification alone (data not shown). However, the species thus enriched will also be enriched in the

driver vs. driver hybridization. Thus, comparison of the products from the experimental and control selections will permit the identification of these artifactually selected species.

Experiments with the DBP knockout mouse strain show that the SABRE selection protocol is capable of identifying cDNA fragments demonstrating quantitative (RBP) as well as qualitative (*neo*, FLC) differences in mRNA expression between the two samples. In addition, it has allowed the identification of cDNA restriction fragments whose mRNAs display greater accumulation in mouse liver in the evening than in the morning. One such mRNA encodes coumarin 7-hydroxylase, a member of the cytochrome P450 gene superfamily that catalyzes the hydroxylation of coumarin, a compound found in sweet clover and other plants, which is mildly toxic and capable of causing liver damage (25–27). Coumarin 7-hydroxylase mRNA accumulation displayed a circadian expression pattern, with peak expression (approximately 300 copies per cell) detected at 10 p.m. In an evening feeder such as the mouse, circadian expression of the coumarin 7-hydroxylase gene is perhaps advantageous in that it provides increased amounts of the enzyme when the animal is feeding, to metabolize coumarin ingested with the food.

Our results demonstrate that the SABRE procedure is capable of enriching differentially expressed species with relatively low accumulation, such as coumarin 7-hydroxylase. However, the present protocol was unable to increase the accumulation of a *Sau*3AI cDNA fragment from another gene with circadian expression, the mouse cholesterol 7α-hydroxylase gene. The mRNA accumulation from this gene displays a circadian pattern in rat and mouse liver, with peak expression at 8 p.m., approximately 8- to 10-fold higher than that at 8 a.m. (ref. 5; L.L.-M., D.L., and U.S., unpublished work). Despite this, a 124-bp *Sau*3AI fragment of the cholesterol 7α-hydroxylase cDNA was lost during the same selection procedure in which the coumarin 7-hydroxylase cDNA fragment was enriched (data not shown).

The loss of this species is likely due to one or both of the critical parameters discussed above: insufficient hybridization or preferential loss during PCR amplification. Peak accumulation of cholesterol 7α-hydroxylase is approximately 10-fold lower than that of coumarin 7-hydroxylase (approximately 20 to 30 copies per liver cell at 8 p.m., and approximately 2–4 copies per cell at 8 a.m.; data not shown). The hybridization conditions used, with driver material in excess, may not have been extensive enough to permit sufficient rehybridization of the cholesterol 7α-hydroxylase cDNA strands, preventing their enrichment during selection. In addition, a bias against smaller DNA fragments may exist in the hybridization reaction, owing to the complementary primer sequences present on each single-stranded molecule. These may result in the formation of single-stranded hairpin structures, which impede hybridization of the complementary DNA strand, and would be more significant for a shorter DNA strand than for a longer one. Finally, PCR bias might further deplete the cholesterol 7α-hydroxylase cDNA fragment from the selected population if it is less efficiently amplified in the PCR reaction. We have found that with our conditions, species less than 150 bp in length are less efficiently amplified than species of 150 to 500 bp in length (data not shown).

The speed and sensitivity of the selection procedure can be improved by the use of protocols to accelerate DNA hybridization, such as a thermal cycler-based adaptation of the phenol emulsion reassociation technique (ref. 28; H. Kuebler and B. Frey, personal communication). Alternatively, the formation of intramolecular hairpin structures could be suppressed by using shorter PCR primers and higher hybridization temperatures, both of which would render these hairpins less stable.

PCR-induced changes in the selected population can be reduced by optimization of PCR parameters, such as buffer composition and choice of enzyme, to improve efficiency in PCR and to equalize amplification efficiency independent of species and length of PCR product. Recently, the Boehringer Mannheim Expand High Fidelity PCR system, using a combination of *Taq* and Pwo DNA polymerases, was found to give both more faithful amplification of a mixture of amplification products, and better suppression of nonselected, abundant species when used in combination with the thermal cycler phenol emulsion reassociation technique hybridization protocol during SABRE selection (H. Kuebler and B. Frey, personal communication).

We thank V. Ossipow, B. Frey, and H. Kuebler for critical reading of the manuscript and helpful discussions, B. Frey and H. Kuebler for communication of unpublished data, and N. Roggli for expert preparation of figures. This work was supported by the Swiss National Science Foundation, the Canton of Geneva, and Boehringer Mannheim, Penzberg, Germany.

1. Gorman, M. & Baker, B. S. (1994) *Trends Genet.* **10,** 376–380.
2. Struhl, G., Struhl, K. & Macdonald, P. M. (1989) *Cell* **57,** 1259–1273.
3. Wuarin, J., Falvey, E., Lavery, D., Talbot, D., Schmidt, E., Ossipow, V., Fonjallaz, P. & Schibler, U. (1992) *J. Cell Sci.* **16** Suppl., 123–127.
4. Ogawa, H., Pitot, H. C. & Fujioka, M. (1994) *Arch. Biochem. Biophys.* **308,** 285–291.
5. Noshiro, M., Nishimoto, M. & Okuda, K. (1990) *J. Biol. Chem.* **265,** 10036–10041.
6. Savage, M. J., Sala-Trepat, J. M. & Bonner, J. (1978) *Biochemistry* **17,** 462–467.
7. Liang, P. & Pardee, A. B. (1992) *Science* **257,** 967–971.
8. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995) *Science* **270,** 484–487.
9. Lisitsyn, N. (1995) *Trends Genet.* **11,** 303–307.
10. Hubank, M. & Schatz, D. G. (1994) *Nucleic Acids Res.* **22,** 5640–5648.
11. Zeng, J., Gorski, R. A. & Hamer, D. (1994) *Nucleic Acids Res.* **22,** 4381–4385.
12. Schmidt, E. E. & Schibler, U. (1995) *J. Cell Biol.* **128,** 467–483.
13. Chomczynski, P. & Sacchi, N. (1994) in *Current Protocols in Molecular Biology*, eds. Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (Wiley, New York), pp. 4.2.4–4.2.8.
14. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
15. Gubler, U. (1988) *Nucleic Acids Res.* **16,** 2726.
16. Britten, R. J., Graham, D. E. & Neufeld, B. R. (1974) *Methods Enzymol.* **29,** 363–406.
17. Raju, R., Hoppe, B. L., Navaneetham, D. & Conti-Fine, B. M. (1995) *BioTechniques* **18,** 33–35.
18. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 1.90–1.104.
19. Brown, T. (1994) in *Current Protocols in Molecular Biology*, eds. Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidnan, J. G., Smith, J. A. & Struhl, K. (Wiley, New York), pp. 2.9.9–2.9.10.
20. Fonjallaz, P., Ossipow, V., Wanner, G. & Schibler, U. (1996) *EMBO J.* **15,** 351–362.
21. Milner, J. J., Cecchini, E. & Dominy, P. J. (1995) *Nucleic Acids Res.* **23,** 176–187.
22. Thielking, V., Alves, J., Fliess, A., Maass, G. & Pingoud, A. (1990) *Biochemistry* **29,** 4682–4691.
23. McGill, J. R., Naylor, S. L., Sakaguchi, A. Y., Moore, C. M., Boyd, D., Barrett, K. J., Shows, T. B. & Drysdale, J. W. (1987) *Hum. Genet.* **76,** 66–72.
24. Stubbs, L., Carver, E., Ashworth, L. & Lopez-Molina, L. (1996) *Mamm. Genome* **7,** 65–67.
25. Squires, E. J. & Negishi, M. (1988) *J. Biol. Chem.* **263,** 4166–4171.
26. Negishi, M., Lindberg, R., Burkhart, B., Ichikawa, T., Honkakoski, P. & Lang, M. (1989) *Biochemistry* **28,** 4169–4172.
27. Fentem, J. H. & Fry, J. R. (1993) *Comp. Biochem. Physiol.* **104,** 1–8.
28. Miller, R. D. & Riblet, R. (1995) *Nucleic Acids Res.* **23,** 2339–2340.