Volume 11 Number 2 1983 Nucleic Acids Research

The independent evolution of two closely related satellite DNA elements in rats *(Rattus)*

Frank R. Witney and Anthony V. Furano

Laboratory of Biochemical Pharmacology, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Building 4, Room 116, Bethesda, MD 20205, USA

Received 5 October 1982; Revised and Accepted 3 December 1982

ABSTRACT

We have identified and determined the sequence and organization of a new rat satellite DNA in Rattus rattus, the roof rat. This satellite DNA, which we call R. rattus satellite I', consists of tandem arrays of a 185 base pair (bp) repeat unit that we call a'. a' is 86% homologous to a 185 bp portion of the 370 bp repeat unit of the previously described rat satellite [Pech et al. (1979) Nucleic Acids Res. 7, 417-432] present in the common laboratory rat species, R. norvegicus. We can thereby distinguish two 185 bp portions of the satellite I 370 bp repeat unit: "a" (homologous to a') and "b". Satellite I has the structure $(a,b)_n$, and satellite I' has the structure $(a')_n$. Like a, a' is only about 60% homologous to b and fails to hybridize to b. Although R. norvegicus and R. rattus contain about the same total concentration of satellite sequences, R. norvegicus contains essentially only the a,b type (satellite I), whereas R. rattus contains a' type (satellite I') and lesser amounts of the a,b type (satellite I). The a,b type (satellite I) in R. rattus is very similar to that in R. norvegicus as judged both by hybridization and by the presence of all but one of the major restriction enzyme sites that characterize the R. norvegicus satellite I. In R. rattus the a' and a,b repeat units are not detectably present in the same tandem array. All of the sequence differences between a' (R. rattus) and a (R. norvegicus) can be accounted for by simple base substitutions, and the implication of this and other features of rat satellite DNA structure for satellite DNA evolution are discussed.

INTRODUCTION

A common feature of the genomes of most eukaryotes is the presence of families of highly repeated DNA sequences. A characteristic feature of one class of highly repetitive DNA is the organization of a basic repeat unit into long tandem arrays. Although this type of repetitive DNA was first detected in some organisms as a satellite band to the main DNA band after isopycnic centrifugation, any tandemly arranged, highly repeated DNA element is now referred to as satellite DNA. [For reviews, see (1, 2).] The function of satellite DNA is not known, and proposals include roles in chromosomal pairing (3, 4), reorganization (5, 6), or recombination (7). Alternatively, it has been suggested that, as postulated for other repeated DNA sequences, they have no function other than to propagate themselves (8, 9). The mechanisms by which

0305-1048/83/1102-0291$2.00/0

these sequences evolve are also not understood. Specific satellite sequences are usually confined to one or a group of closely related species and are thought, therefore, to arise de novo in a relatively short period of time. On the other hand, it has been reported that certain rodent satellite sequences are held in common between widely separated species and therefore appear to have been conserved for a long time but amplified to different extents in different animals (10).

The rat genome does not contain satellites that can be isolated by density gradient centrifugation. However, treatment of R. norvegicus DNA with the restriction enzyme Sau3A converts about 1% to 3% of the DNA into a set of high molecular weight fragments [about 10 kilobase pair (kbp)], while the remainder of the DNA is digested into much smaller fragments (11). The DNA in the high molecular weight component is greatly enriched for a tandemly arranged, highly repeated family called rat satellite I. The consensus nucleotide sequence of the 370 bp repeating unit of this satellite was determined by sequencing uncloned restriction fragments derived from the satellite. The repeat unit was found to be composed of four alternating, nonidentical 92 and 93 bp segments. This conclusion was later confirmed by base sequence studies of cloned 92 and 93 bp segments (12).

In order to gain insight into the mechanisms of repeat DNA evolution, we compared the genome of R. norvegicus to a closely related (but not interbreeding) species, R. rattus, for the presence of conserved repetitive DNA elements. We report here the identification and nucleotide sequence of a satellite present in R. rattus, but apparently only in very low concentration in R. norvegicus. This satellite is composed of tandemly repeated 185 bp elements clearly related, by sequence, to a portion of the R. norvegicus satellite I. In addition, by hybridization with cloned segments of the R. norvegicus satellite I, we also detected in R. rattus tandem arrays of a 370 bp repeat unit that is closely related to the R. norvegicus satellite I.

MATERIALS AND METHODS

DNA Sources and Purification

DNA was isolated from the livers of: (i) a laboratory-bred R. norvegicus (Sprague-Dawley); (ii) an R. norvegicus trapped in the wild; and (iii) an R. rattus trapped in the wild. Livers from the latter two rats were kindly provided by Dr. Stephen C. Frantz, New York State Department of Health, Troy, New York. High molecular weight DNA was purified after dry-ice pulverization of the livers (13), essentially as described earlier (14). DNA was also purified

by similar methods from cultured Jensen sarcoma cells, an R. norvegicus cell line [ATCC, CCL45 (15)]. A frozen starter culture of cells was obtained from the American Type Culture Collection. The cells were grown according to methods suggested by the supplier, including culturing in the absence of antibiotics.

Restriction Enzyme Analysis of High Molecular Weight DNAs

Restriction enzyme digests of the various DNA samples were carried out with the enzymes described in the figure legends, using conditions described by the suppliers. All enzymes were obtained from New England Biolabs, except EcoRI, which was from Bethesda Research Laboratories. Five µg of liver DNA or 2.5 µg of Jensen sarcoma DNA was digested at an enzyme : DNA ratio of 5 units/µg for 2 to 3 hours. The reactions were stopped by the addition of 0.1 volume of 50% (vol/vol) glycerol, 0.05% (wt/vol) bromphenol blue, 2.5% (wt/vol) sodium dodecylsulfate, and 150 mM sodium EDTA, pH 7.0. The samples were heated to 65°C for 5 minutes and subjected to electrophoresis on vertical slab gels (0.12 cm x 12 cm x 13 cm) of 8% polyacrylamide (acrylamide : bis-acrylamide, 29 : 1) and a buffer containing 0.09 M Tris-borate, pH 8.3, 2 mM EDTA (16). Electrophoresis was carried out at 190 volts until the bromphenol blue reached the bottom of the gel, usually about 1.5 hours. The DNA was visualized under UV light after the gel was stained for 15 minutes with a solution of 0.5 µg/ml of ethidium bromide in water.

Cloning of the 92/93 bp *R. rattus Eco*RI Satellite Fragments

Fifty µg of R. rattus DNA was digested with 25 units of EcoRI in a volume of 200 µl for 3 hours. Glycerol was added to a final concentration of 5% and the DNA was applied to a 2-ml Sepharose Cl-4B gel filtration column equilibrated with 50 mM Tris-HCl (pH 7.5) and 10 mM $MgCl_2$. The DNA was eluted from the column with the same buffer and 20 fractions containing 50 µl each were collected. A small portion of each fraction was analyzed on an 8% polyacrylamide gel, and the DNA contained in several fractions that eluted well after the bulk of the high molecular weight restriction fragments was judged to contain mainly 92/93 bp fragments and their multimers. One of these fractions, containing about 100 ng of DNA, was dephosphorylated by incubation with 1.5 units of calf intestinal phosphatase (Boehringer-Mannheim) at 37°C for 1 hour. The phosphatase was inactivated by heating the reaction mixture to 70°C for 3 hours. The dephosphorylated DNA was added to a ligation reaction containing 5 ng of EcoRI digested pBR325 DNA (17) and 0.5 unit of T4 DNA ligase (New England Nuclear). The ligation reaction was performed for 1 hour at 16°C using conditions described by the supplier. After the ligation, the recombinant DNA

molecules were directly used to transform Escherichia coli HB101 (18) as described by Morrison (19). Clear lysates of 20 tetracycline-resistant, chloramphenicol-sensitive transformants were prepared (20) and the plasmids screened for the presence of 92/93 bp EcoRI inserts. Eighteen of the 20 transformants had such an insert, and all of these were found to hybridize to R. rattus DNA, but not R. norvegicus DNA. Four of these recombinant plasmids were used in subsequent experiments. Restriction enzyme analysis showed that each plasmid contained a single 92 or 93 bp EcoRI fragment insert. Plasmid DNA was prepared from chloramphenicol-treated cultures by the clear lysate method of Clewell and Helinski (21), modified to contain a final concentration of 5.5% Triton-X as the only detergent.

Blot Hybridizations

Five µg of the rat DNA's were digested with the appropriate restriction enzymes, and the fragments were electrophoresed on 8% polyacrylamide gels as described above. The DNA fragments were then transferred to nitrocellulose filters (BA85, Schleicher and Schuell) as described by Southern (22) as modified by Wahl et al. (23). After incubation for at least 18 hours at 65°C in a solution containing 5X SSC (1X SCC is 0.15 M NaCl, 0.015 M Na citrate, pH 7.2), 0.1% polyvinylpyrrolidone, 0.1% Ficoll, 0.25% sodium dodecylsulfate, and 100 µg/ml sonicated salmon sperm DNA, the filters were incubated in the same solution with about 2 x 10^6 cpm of plasmid DNA made radioactive by nick translation (24). The specific activities of the plasmid DNAs are given in the legend to Fig. 2. The radioactive plasmids were boiled for 10 minutes in 10 mM sodium phosphate buffer, pH 7.0, before being added to the filters. Incubation was carried out for 24 hours at 65°C, after which the filters were washed at room temperature with 4 changes (200 ml each) of 2X SSC. After drying, the filters were exposed to x-ray film (Kodak XAR-5), at -70°C, using an intensifying screen (Dupont, Cronex Lightning-Plus).

DNA Sequence Determinations

Plasmid DNAs digested with EcoRI were radiolabeled either at the 5' end, with [γ-^{32}P]ATP using polynucleotide kinase (25), or at the 3' end, with [α-^{32}P]-dATP using the Klenow fragment of DNA polymerase I (26). The radioactive mixture containing both the vector and the 92 or 93 bp EcoRI fragment was subjected to either strand separation (25) or digestion with a second restriction enzyme. After purification on polyacrylamide gels, the fragments were subjected to the chemical sequencing reactions described by Maxam and Gilbert (25). Although more than one radioactive nucleotide was introduced at the ends of the fragments radiolabeled with DNA polymerase I, we were able to read

unambiguous sequencing ladders from these fragments. The graphic matrix computer program described by Maizel and Lenk (27) was used to compare homology between various DNA sequences.

RESULTS

Identification of a *Rattus rattus* Satellite I'

The organization of the tandemly arrayed rat satellite I from R. norvegicus, as determined by Pech et al. (11), is shown in Fig. 1A. The repeat unit of this satellite is a 370 bp sequence that can be subdivided into 92 and 93 bp segments by digestion with the EcoRI endonuclease [see (11) and below]. Therefore, complete digestion of R. norvegicus DNA with EcoRI results in the generation of a set of heterogeneous fragments that are 92 or 93 bp long, as well as a ladder of fragments that are multiples of 92/93. The ladder results from sequence alterations in the EcoRI site between neighboring EcoRI monomers. Since each EcoRI monomer is repeated only once in every four segments, digestion of R. norvegicus DNA with a restriction enzyme that has a recognition site in only one of the four segments, for example, HaeIII or HindIII, will generate a fragment 370 bp long (Fig. 1A). The tetramer is repeated about 100,000 times in the R. norvegicus genome and therefore constitutes about 2% (40,000 kbp) of the rat genome; tandem arrays of approximately 30 or so tetramers are located in the centromeric and teleomeric chromosomal regions of all the rat chromosomes (12).

In a study of rodent repeat DNA families, to be reported elsewhere, we found that only one of about 90 different middle and high repeat DNA families appears to be obviously different between R. norvegicus and R. rattus. Preliminary data suggested that this difference was due to the reduction in the R. rattus genome of sequences homologous to R. norvegicus satellite I DNA. In order to investigate this further, we compared restriction enzyme digests of R. norvegicus and R. rattus DNA (Fig. 1B). Three different sources of R. norvegicus DNA were used: a laboratory-bred R. norvegicus, a wild R. norvegicus trapped in New York State, and Jensen sarcoma cells (a slightly hyperploid cell line that was established in 1958 from a sarcoma which arose in R. norvegicus in 1907 and had been propagated since then). The R. rattus DNA was prepared from the liver of an animal trapped in New York State and thus was presumably of the Oceanian Rattus karyotype (n = 38) rather than the Asian Rattus karyotype (n = 42) or the R. norvegicus karyote (n = 42) (28).

All three R. norvegicus DNAs show indistinguishable ladders of multiples of a 92/93 bp EcoRI restriction fragment. Restriction of DNA from R. rattus also shows a similar pattern, except that the proportion of 92/93 bp fragments

is considerably enhanced relative to the higher multiples. Digestion of the R. norvegicus DNAs with HaeIII generates the expected 370 bp fragment that is easily seen in the HaeIII digest of Jensen sarcoma DNA (Fig. 1B), but difficult to see in the other R. norvegicus samples because of the higher concentrations

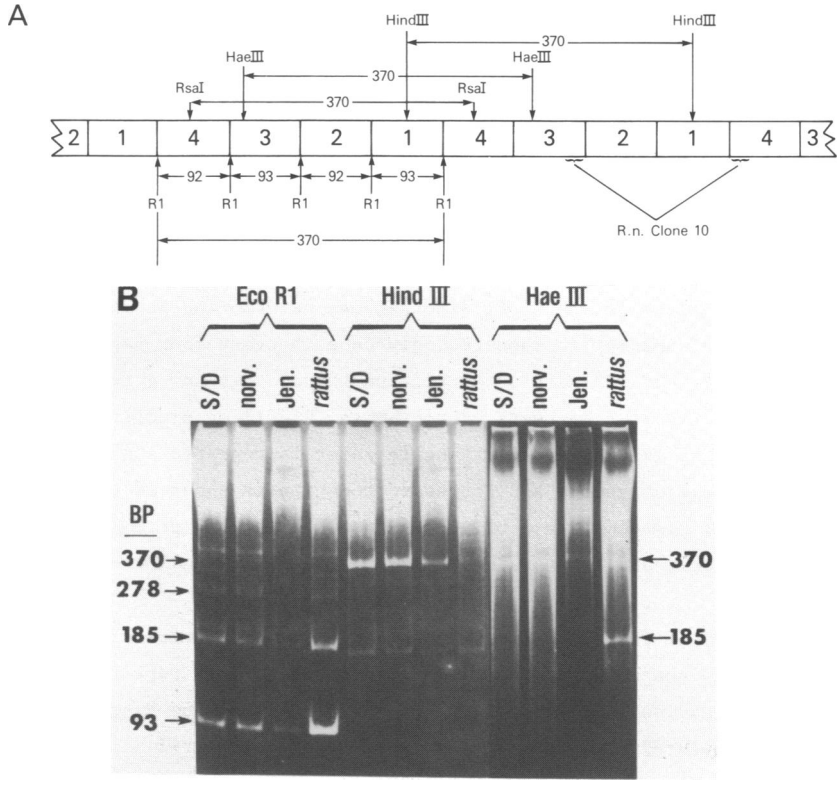

FIGURE 1. (A) Representation of the R. norvegicus satellite I. The 370 bp repeat unit of the satellite is composed of four nonidentical 92 and 93 bp segments, each of which contains an EcoRI (RI) site at its 5' end. The 370 bp repeat units are tandemly arranged, as described in (11) and Results. We have numbered the four segments as shown to facilitate comparison to our work. Characteristic restriction enzyme sites and the distance between them in bp are indicated. The clone R. n. 10 contains the R. norvegicus satellite I segments 1 and 2 and possibly short stretches (less than 20 bp) of segments 3 and 4 inserted into the Pst site in pBR322 by the "G-C" tailing procedure (F. R. Witney and A. V. Furano, unpublished data). (B) Restriction digests of rat DNAs. Five μg of the liver DNA samples or 2.5 μg of Jensen sarcoma DNA were incubated with the restriction enzymes EcoRI, HindIII, or HaeIII as described in Materials and Methods. The fragments were separated on an 8% (29:1 acrylamide:bis-acrylamide, wt/wt) polyacrylamide gel and stained with ethidium bromide. The sizes of fragments are given in base pairs. The letters at the top of the gel denote the source of the DNA: S/D, the Sprague-Dawley laboratory strain of R. norvegicus; norv., a feral R. norvegicus; Jen, Jensen sarcoma tissue culture cells; rattus, a feral R. rattus.

of DNA in these digests. In contrast to these results, HaeIII digestion of R. rattus DNA generates a 185 bp fragment, one-half the size of the satellite I repeat unit in R. norvegicus. These data suggested that R. rattus contains a satellite composed of tandem arrays of 92/93 bp EcoRI fragments, but that the EcoRI fragment which contains the HaeIII restriction site is repeated every 185 bp instead of every 370 bp, as it is in R. norvegicus.

To confirm this conclusion we isolated 92/93 bp EcoRI fragments from an EcoRI digest of R. rattus DNA and ligated them into the EcoRI site of the E. coli plasmid pBR325 as described in Materials and Methods. Twenty independent clones were selected, 18 of which contained a 92/93 bp fragment that hybridized to R. rattus DNA but not to R. norvegicus DNA. We chose four at random and found that two of the cloned inserts contained an identically located HaeIII restriction site, and that the other two contained identically placed sites for both RsaI and DdeI.

Fig. 2 shows that each class of the R. rattus clones can hybridize to R. rattus DNA but apparently not to R. norvegicus DNA. Both the EcoRI fragments and the 185 bp HaeIII fragment of R. rattus, hybridize to either class of R. rattus clones. Both classes of these clones also hybridize to the 185 bp fragment generated from R. rattus DNA by RsaI (data not shown). In other experiments (not shown here) we are able to detect in R. norvegicus a very low concentration of sequences that, in our hybridization conditions, must be > 85% homologous to the 92/93 bp fragments cloned from R. rattus. These were undetected with probes of the specific activity used in the hybridization in Fig. 2. From these data we conclude that the cloned inserts are EcoRI fragments present in high copy numbers in R. rattus DNA and may also be present but at much lower levels in R. norvegicus DNA. Furthermore, the R. rattus 92/93 bp fragments make up a 185 bp repeat unit in R. rattus consisting of a HaeIII-containing fragment and a RsaI-containing fragment arranged as tandem arrays in a typical satellite DNA pattern. As we shall show, this 185 bp repeat unit has about 85% base sequence homology to a 185 bp portion of the rat satellite I described by Pech et al. (11). Therefore, to remain consistent with their nomenclature, we shall refer to this R. rattus satellite as R. rattus satellite I'.

Rattus rattus Contains Two Related but Distinct Satellite DNAs

In addition to satellite I', we have also identified satellite I in the R. rattus genome. Fig. 2 shows that a clone containing R. norvegicus satellite I segments 1 and 2 (R. norvegicus clone 10; see Fig. 1A) hybridizes to a 370 bp HaeIII fragment of R. rattus DNA but not to the prominent 185 bp HaeIII frag-

ment, which is the satellite I' repeat unit. The faint signal seen slightly below the position of the 185 bp fragment most likely does not represent a hybrid since it has not been seen in other hybridization experiments (data not shown). Since HaeIII digests the R. norvegicus satellite I tandem arrays into 370 bp fragments, R. rattus DNA must contain tandem arrays of satellite I that are not only conserved enough in sequence to hybridize to the R. norvegicus satellite I, but must also have the tetramer organization characteristic of the R. norvegicus satellite I. We call this sequence R. rattus satellite I to distinguish it from its counterpart in R. norvegicus.

To summarize: R. rattus, contains two kinds of satellite DNA. The repeat unit of one is very closely related in size (370 bp) and sequence to the repeat unit of R. norvegicus satellite I. We refer to the R. rattus version as R. rattus satellite I. The other satellite in R. rattus is made up of a 185 bp repeat unit that is similar to a 185 bp portion of satellite I but that is

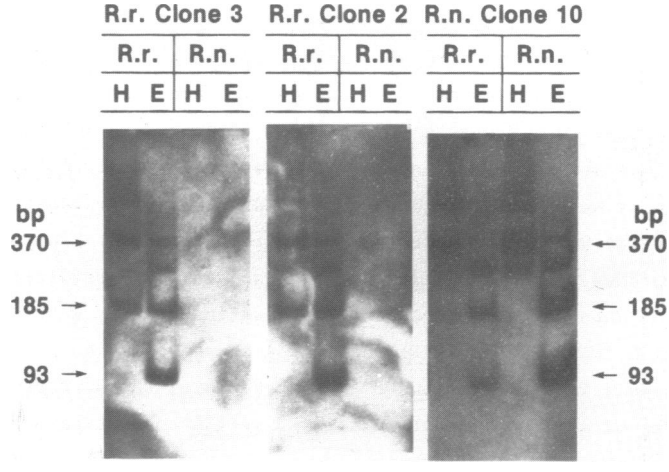

FIGURE 2. Hybridization of cloned satellite DNAs to rat DNAs. Five μg of R. rattus (R. r.) or R. norvegicus (R. n.) DNA was digested with either HaeIII (H) or EcoRI (E) and electrophoresed on an 8% polyacrylamide gel. The DNA fragments were then transferred to nitrocellulose filters and hybridized with ^{32}P-labeled plasmids as described in Materials and Methods. R. r. clone 3 contains a 93 bp R. rattus EcoRI fragment that has a HaeIII restriction site; R. r. clone 2 contains a 92 bp R. rattus EcoRI fragment that has RsaI and DdeI restriction sites; R. n. clone 10 contains R. norvegicus satellite I segments 1 and 2 (see Fig. 1). The R. r. clones were labeled to a specific activity of about 5 x 10^6 cpm/μg, whereas the R. n. clone 10 was labeled to a specific activity of about 3 x 10^7 cpm/μg. 2 x 10^6 cpm of radioactive plasmid was added to each hybridization. Hybrids were detected by autoradiography. Although the transfer of fragments from acrylamide gels is not nearly as efficient as from agarose gels, we find that transfer efficiencies are independent of length, at least up to about 0.4 kbp.

clearly different from it. We call this R. rattus satellite I'. A very small concentration of sequences closely related to R. rattus satellite I' may be present in R. norvegicus (data not shown), and we call this R. norvegicus satellite I'.

Mapping of the major restriction enzyme sites in R. rattus satellite I has demonstrated that its structure is very similar to the R. norvegicus satellite I except that the HindIII site (present in segment 1, Fig. 1A) is missing (data not shown). Fig. 2 shows that the R. norvegicus satellite I fragment present in R. n. clone 10 hybridizes less well to R. rattus DNA than to R. norvegicus DNA. From other experiments of this type, as well as from the relative hybridization of R. norvegicus and R. rattus DNA to 10 other R. norvegicus satellite I-containing clones, we have concluded that R. norvegicus contains about 5 times as much satellite I as R. rattus. However, comparison of the amount of the 185 bp HaeIII fragment derived from R. rattus satellite I' with the amount of the 370 bp HindIII rat satellite I fragment recovered from R. norvegicus DNA (Fig. 1B) suggests that lack of satellite I in R. rattus is compensated for by satellite I'. That is, the sum of I and I' in R. rattus appears to be equal to the amount of I in R. norvegicus.

Nucleotide Sequences of the 92/93 bp Cloned Fragments from *R. rattus*

The nucleotide sequences of 2 HaeIII-containing and 2 RsaI-containing cloned R. rattus EcoRI fragments were determined. Both HaeIII-containing EcoRI fragments had identical sequences, as did both RsaI-containing fragments. By contrast, cloned EcoRI fragments of a given type in R. norvegicus generally have 5% to 10% sequence divergence (12). This suggests that R. rattus satellite I' may be more homogeneous than R. norvegicus satellite I. The HaeIII-type fragment is about 60% homologous to the RsaI-type fragment. The four types of R. norvegicus EcoRI fragments have a similar degree of relatedness to each other (11).

The R. rattus satellite I' sequence is presented in Fig. 3A as a 185 bp composite of the 92 bp (RsaI type) and the 93 bp (HaeIII type) fragments. The first 20 bp of the 185 bp composite sequence is the last 20 bp of the HaeIII type fragment. Next is the entire RsaI type fragment and then the first 73 bp of the HaeIII type fragment (see Fig. 3C). Doing this maximizes the homology between the two satellite sequences, which was determined by using the two-dimensional dot matrix program developed by Maizel and Lenk (27) (Fig. 3B).

The 185 bp R. rattus sequence and the related portion of the R. norvegicus satellite I have 159 out of 185 bp in common (86% homologous) and we have designated these homologous 185 bp elements as "a'" (satellite I') and "a" (satellite I). The R. rattus sequence is only 60% homologous (112 matches/

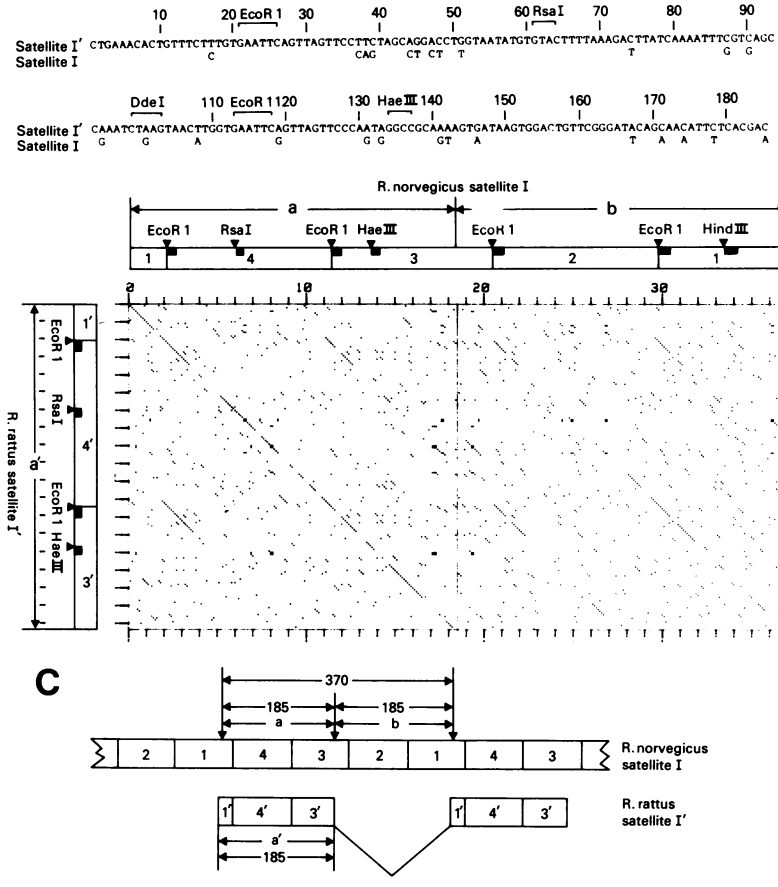

FIGURE 3. (A) Nucleotide sequence of the 185 bp R. rattus satellite repeat unit. The sequence was derived from EcoRI satellite fragments cloned from R. rattus. Two independently selected cloned segments of the HaeIII type, as well as two clones of the RsaI/DdeI type, were found to have identical nucleotide sequences (see text). Both strands of the HaeIII types were completely sequenced. About 70% of both strands of the RsaI/DdeI type were sequenced, and the remainder was sequenced 3 times from different restriction fragments. The first 20 bp of the 185 bp sequence shown is the last 20 bp of the HaeIII type fragment. Next is the entire sequence of RsaI/DdeI type fragment which is followed by the first 73 bp of the HaeIII-type fragment (see Fig. 3C). Only those bases that are mismatched between the two satellites are shown in the R. norvegicus sequence. (B) Dot matrix comparison between the repeat units of the R. rattus satellite I' and the R. norvegicus satellite I. The comparison was carried out between the consensus nucleotide sequence of the R. norvegicus satellite I sequence (11) (horizontal axis) and the R. rattus satellite I' sequence (vertical axis). Each dot represents the center of a three-base exact homology between the two sequences. Easily seen is the homology that exists at and near the EcoRI sites positioned every 92 to 93 bp in the repeat units of both satellites. For orientation, the various 92 and 93 bp EcoRI segments of the R. norvegicus satellite I are indicated.

The 185 bp portion of the R. norvegicus satellite I that is homologous to the R. rattus satellite I' sequence is called a, while the other 185 bp portion is called b. We call the R. rattus satellite I' sequence, a', to indicate its high degree of homology to a, and the various segments, 1', 4', and 3', to emphasize the relationships between a, a', and the previously published data on satellite I (11, 12). (C) Representation of the tandem arrays that comprise the R. norvegicus satellite I and the R. rattus satellite I'.

185 bp) to the other 185 bp portion (referred to as "b") of the R. norvegicus satellite I. Whereas satellite I has the structure $(a,b)_n$, satellite I' has the structure $(a')_n$.

DISCUSSION

We have identified a new, highly repetitive DNA family in the genus Rattus, which is present as the predominant satellite DNA in R. rattus. This satellite, which we call R. rattus satellite I', is organized as tandem repeats of a 185 bp element, a', that has considerable (86%) sequence homology with a 185 bp portion, a, of the R. norvegicus satellite I. Satellite I' is a polymer of a' elements, and satellite I is a polymer of alternating a,b elements (Fig. 3C). Whereas R. norvegicus contains essentially only the a,b type (satellite I), R. rattus contains both the a' arrays (satellite I') and the a,b arrays (satellite I). R. rattus contains about 4 times as much satellite I' as it does satellite I, and their combined concentration is about equal to the concentration of satellite I in R. norvegicus.

These results mean that this class of rodent satellites are quite similar in general form to the ubiquitous primate alpha satellites [(29-32); for review, see (1)]. Some alphoid satellites are tandemly arrayed dimers of related, but nonidentical 170 bp units; others are arrays of near identical 170 bp monomers. The two related subunits of the dimers are usually about 70% or so homologous to each other [for example, see (29, 30, 32)], which is similar to the relatedness between the a and b components of rat satellite I. This analogy between Rattus and primate alphoids is strengthened by the fact that R. rattus satellite I', being composed of tandem arrays of a single 185 bp repeat unit, has its counterpart in the African green monkey alpha satellite, which is also made up of tandem arrays of a single 172 bp repeat unit (31).

The 185 bp rat repeat units, however, differ from the primate alphoids in that the rat a, b, and a' sequences each retain remnants of their probable origin as a dimer of two related but now 30-40% divergent 92/93 bp elements which were first revealed in rat satellite I by Pech et al. (11). Most prominent is the 18 bp sequence that occupies position 19 through 36 in the a' sequence and is exactly repeated 92 bp away at position 111 through 128 (Fig.

3A). This 18 bp sequence, with at most two base substitutions, occurs at exactly the corresponding two positions in the a (Fig. 3A) and b [see (11)] repeat units of satellite I and is clearly revealed by the dot matrix of the a' versus the a,b dimer (Fig. 3B) or of the a' monomer versus itself (data not shown). By contrast, Wu and Manuelides (30) could find no internal periodicity in either of the 170 bp human alpha monomers nor could we in the 170 bp African green monkey alpha sequence, using the dot matrix analysis (data not shown).

Whether or not the 170 bp primate alpha sequences were originally assembled from shorter repeat units and then amplified, it is the dimers assembled from the two nonidentical alpha sequences that are now the units of evolution in primates (29, 30, 32). This is also true of the a and b rat elements which have been recently evolving as a dimer in R. norvegicus (11), and this dimeric structure is maintained in R. rattus even in the presence of a large excess of a' elements. In this genome (a',b) polymers must be rare since digestion of R. rattus with HinfI, which digests b elements (11) but not a' elements (Fig. 3A and unpublished data) into several small fragments, did not produce the size of a'-containing fragments expected from digestion of $(a',b)_n$ (Epstein, Witney, and Furano, unpublished data). Second, (a',a) polymers must also be rare because we can detect by hybridization most of the b sequences in fragments that are 370 bp (or integral multiples thereof) after digestion of R. rattus DNA with enzymes that have sites confined to the a element. We could not obtain this result if a significant number of a elements were sequestered in the (a',a) polymers, unless the a element in R. rattus is in excess over the b elements. This is not true in R. norvegicus since we do not detect 185 bp fragments after HaeIII digestion of R. norvegicus DNA (see Fig. 1B).

The formation and amplification of larger repeat units from divergent versions of a shorter sequence element as outlined above is a recurring motif in the evolution of a wide variety of satellite DNAs [see (1)]. The structure of the above satellites indicates that they could have evolved, or are now doing so, by multiple rounds of sequence duplication, divergence of the duplicated sequences, and subsequent amplification of dimers of the diverged sequences, as originally proposed by Southern (33) and verified by sequence determination (34) for the evolution of mouse satellite DNA. However, continuous progress along this evolutionary pathway is not inevitable. For example, the a' rat element which presumably is the product of at least a single round of the above steps now is evolving as a monomeric 185 bp unit like the African green monkey alpha sequence. Further evolution of the a' sequence might obliterate all traces of its dimeric origin, producing a 185 bp sequence

with no internal periodicities like the present day primate alpha sequences (30).

Sequence differences among the rat or primate repeat units can be accounted for by simple base substitutions [present work and (11, 29, 30, 32)]. In all cases certain corresponding regions of the primate alpha sequences or the rat a, b, or a' elements have not diverged. Therefore, tandem arrays of these satellites contain at least one 18-20 bp region of identical or near identical sequence, periodically repeated throughout the array, although the repeat interval varies with the satellite. In primates it is about 170 bp (29-32), but in rats, as mentioned above, it is 92/93 bp in both the $(a,b)_n$ and $(a')_n$ arrays. Although the significance of these periodicities is not clear (30, 35-37), the presence of corresponding conserved regions among the primate alphoids has been cited as evidence for the formation of present day alpha sequences by recombination between divergent ancestral alpha sequences (32). However, rounds of random base substitutions and amplification have also been proposed to explain the sequence divergence among the present day alphoids (30) as well as the generation of variants in the present day rat satellite I (11). If the latter is so, then one might suppose that the nondivergent regions of the above satellites are required for satellite amplification or function.

If the a and a' rat sequences diverged from an ancestral sequence by random base substitutions that were fixed at about the same rate as base substitutions are fixed in single copy DNA [~ 10^{-8} substitutions per site per year (38, 39)], then the a and a' could have begun diverging about 15 million years ago. Since R. rattus and R. norvegicus diverged from a common ancestor about 2 million years ago (38), then by this scheme a and a' existed as divergent sequences in the ancestor of the present day Rattus species but have been amplified differently. This scenario agrees with the library hypothesis of Fry and Salser (10) mentioned in the Introduction, and its plausibility depends on the applicability of the single copy DNA data to estimate the age of the present day satellite sequence elements. At this point it is not possible to decide this since we do not know whether the mechanisms that spread variants through multigene families lead to a higher fixation rate of base substitutions in these DNAs than in single copy DNA (40). Furthermore, the ancestral satellite sequences may not have always been arranged in the genome (30) so as to be subject to the evolutionary events that might uniquely effect multigene families.

ACKNOWLEDGMENT

This investigation was supported in part by PHS Grant Number 5F32CA06864 awarded by the National Cancer Institute, DHHS, to F. R. Witney.

REFERENCES

1. Singer, M. F. (1982) Int. Rev. Cytol. 76, 67-112.
2. Brutlag, D. L. (1980) Annu. Rev. Genet. 14, 121-144.
3. Cooper, K. W. (1964) Proc. Natl. Acad. Sci. U.S.A. 52, 1248-1255.
4. Peacock, W. J., and Miklos, G. L. G. (1973) Adv. Genet. 17, 361-409.
5. Kurnit, D. M., Brown, F. L., and Maio, J. J. (1978) Cytogenet. Cell Genet. 21, 145-167.
6. Holmquist, G. P., and Dancis, B. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 4566-4570.
7. Miklos, G. L. G., and Nankivell, R. N. (1976) Chromosoma 56, 143-167.
8. Doolittle, W. F., and Sapienza, C. (1980) Nature 284, 601-603.
9. Orgel, L. E., and Crick, F. H. C. (1980) Nature 284, 604-607.
10. Fry, K., and Salser, W. (1977) Cell 12, 1069-1084.
11. Pech, M., Igo-Kemenes, T., and Zachau, H. G. (1979) Nucleic Acids Res. 7, 417-432.
12. Sealy, L., Hartley, J., Donelson, J., Chakley, R., Hutchison, N., and Hamkalo, B. (1981) J. Mol. Biol. 145, 291-318.
13. Britten, R. J., Graham, D. E., and Neufeld, B. R. (1975) Methods Enzymol. 29, 363-405.
14. Furano, A. V. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 3104-3108.
15. McCoy, T. A., Maxwell, M., and Kruse, P. F. (1959) Cancer Res. 19, 591-595.
16. Peacock, A. C., and Dingman, C. W. (1967) Biochemistry 6, 1818-1827.
17. Bolivar, F. (1978) Gene 4, 121-136.
18. Boyer, H. B., and Rowland-Dussoix, D. (1969) J. Mol. Biol. 41, 459-472.
19. Morrison, D. A. (1977) J. Bacteriol. 132, 349-351.
20. Ish-Horowitz, D., and Burke, J. F. (1981) Nucleic Acids Res. 9, 2989-2998.
21. Clewell, D. B., and Helinski, D. R. (1970) Biochemistry 9, 4428-4440.
22. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
23. Wahl, G. M., Stern, M., and Stark, G. R. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 3683-3687.
24. Maniatis, T., Jeffrey, A., and Kleid, D. G. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 1184-1188.
25. Maxam, A. M., and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
26. Shen, S., Slightom, J., and Smithies, O. (1981) Cell 26, 191-203.
27. Maizel, J. V., and Lenk, R. P. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 7665-7669.
28. Yosida, T. H., Kato, H., Tsuchiya, K., Sagai, T., and Moriwaki, K. (1974) Chromosoma 45, 99-109.
29. Rubin, C. M., Deininger, P. L., Houck, C. M., and Schmid, C. W. (1980) J. Mol. Biol. 136, 151-167.
30. Wu, J. C., and Manuelidis, L. (1980) J. Mol. Biol. 142, 363-386.
31. Rosenberg, H., Singer, M. F., and Rosenberg, M. (1978) Science 200, 394-402.
32. Donehower, L., Furlong, C., Gillespie, D., and Kurnit, D. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 2129-2133.
33. Southern, E. M. (1975) J. Mol. Biol. 94, 51-69.
34. Hörz, W., and Altenburger, W. (1981) Nucleic Acids Res. 9, 683-696.
35. Musich, P. R., Maio, J. J., and Brown, F. L. (1977) J. Mol. Biol. 117, 657-678.
36. Fittler, F., and Zachau, H. G. (1979) Nucleic Acids Res. 7, 1-13.
37. Singer, D. S. (1979) J. Biol. Chem. 254, 5506-5514.
38. Benveniste, R. E., Callahan, R., Sherr, C. J., Chapman, V., and Todaro, G. J. (1977) J. Virol. 21, 849-862.
39. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980) Cell 20, 555-566.
40. Dover, G. (1982) Nature 299, 111-117.