

UC Irvine

UC Irvine Previously Published Works

Title

Cross-hybridizing snake satellite, Drosophila, and mouse DNA sequences may have arisen independently.

Permalink

<https://escholarship.org/uc/item/8p02c2z5>

Journal

Molecular Biology and Evolution, 2(6)

ISSN

0737-4038

Authors

Levinson, G
Marsh, JL
Epplen, JT
[et al.](#)

Publication Date

1985-11-01

DOI

10.1093/oxfordjournals.molbev.a040374

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Cross-hybridizing Snake Satellite, *Drosophila*, and Mouse DNA Sequences May Have Arisen Independently¹

Gene Levinson,*† J. Lawrence Marsh,* Jörg T. Epplen,‡
and George A. Gutman†

*Department of Developmental and Cell Biology and Developmental Biology Center and

†Department of Microbiology and Molecular Genetics, University of California, Irvine;

‡Max-Planck-Institut für Immunbiologie, Freiburg, West Germany

Previous reports have interpreted hybridization between snake satellite DNA and DNA clones from a variety of distant taxonomic groups as evidence for evolutionary conservation, which implies common ancestry (homology) and/or convergence (analogy) to produce the cross-hybridizing sequences. We have isolated 11 clones from a genomic library of *Drosophila melanogaster*, using a cloned 2.5-kb snake satellite probe of known nucleotide sequence. We have also analysed published sequence data from snakes, mice, and *Drosophila*. These data show that (1) all of the cross-hybridization between the snake, fly, and mouse clones can be accounted for by the presence of either of two tandem repeats, [GATA]_n and [GACA]_n and (2) these tandem repeats are organized differently among the different species. We find no evidence that these sequences are homologous apart from the existence of the simple repeat itself, although their divergence from a common ancestral sequence cannot be ruled out. The sequences contain a variety of homogeneous clusters of tandem repeats of CATA, GA, TA, and CA, as well as GATA and GACA. We suggest that these motifs may have arisen by a self-accelerating process involving slipped-strand mispairing of DNA. Homogeneity of the clusters might simply be the result of a rate of accumulation of tandem repeats that exceeds that of other mutations.

Introduction

Hybridization between snake satellite DNA sequences and DNA clones from a variety of distant taxonomic groups, including flies (*Drosophila melanogaster*) and mice (*Mus musculus*), has been interpreted as evidence for evolutionary conservation, which implies common ancestry (homology) and/or convergence (analogy) to produce the cross-hybridizing sequences (Epplen et al. 1981, 1982a, 1982b, 1983; Jones and Singh 1981a, 1981b; Singh et al 1981, 1984). The snake sequences include a heterogeneous DNA fraction from the banded krait (designated Bkm; *Bungarus fasciatus*, family Elapidae) and a cloned satellite sequence from *Elaphe radiata* (family Colubridae) (designated pErs-5; described in table 1). Large numbers of cross-hybridizing snake and mouse sequences have been found on the W and Y chromosomes of snakes and mice, respectively, which has led to the suggestion that they may be involved in primary sex determination (cf. references cited above).

1. Key words: slipped-strand mispairing, simple repetitive sequences, tandem repeats, DNA sequence evolution, satellite DNA, evolutionary conservation. Abbreviations: GATA and GACA = both strands of duplex DNA (i.e., GATA/CTAT and GACA/CTGT, respectively), as do CATA, GA, CA, and TA. GAYA = GATA/CTAT or GACA/CTGT.

Address for correspondence and reprints: Gene Levinson, Department of Microbiology and Molecular Genetics, University of California, Irvine, California 92717.

Mol. Biol. Evol. 2(6):494-504, 1985.

© 1985 by The University of Chicago. All rights reserved.

0737-4038/85/0206-0166\$02.00

Downloaded from <https://academic.oup.com/mbe/article-abstract/2/6/494/981784>

by University of California, Irvine user

on 08 June 2018

Table 1
Description of Probes Used in Snake/Fly Cross-Hybridization Experiments

Probe	Source	Length	Description
pErs-5	W-chromosomal snake satellite DNA clone from <i>Elaphe radiata</i> (Epplen et al. 1982a)	2.5 kb	Contains tandem repeat clusters [GATA] ₁₅ [GACA] ₂ and [GATA] ₁₀ [GACA] ₁₀ flanked by 2.3 kb of snake DNA ^a
Probe 1	pErs-5	0.2 kb	<i>MnlI/DdeI</i> restriction fragment containing the GATA and GACA repeats and only 31 bp of flanking sequence
Probe 2	pErs-5	2.3 kb	Contains all of snake DNA flanking probe 1 and no GATA or GACA repeats
Probe 3	Sau 3A/ <i>HinfI</i> subclone from <i>Drosophila</i> λ clone 1	0.3 kb	Contains 44–46 tandem repeats of GATA and no tandem repeats of GACA
Probe 4	Probe 1 (labeled) + probe 3 (unlabeled)	0.2 kb	Contains tandem repeats of GACA (GATA repeats are blocked by 1,000-fold excess of unlabeled GATA)

^a For complete nucleotide sequence see Epplen et al. 1982a.

The complete nucleotide sequence of the 2.5-kb snake satellite clone pErs-5 was determined by Epplen et al. (1982a). It exhibits a pattern of hybridization to mouse DNA resembling that described for the heterogeneous Bkm satellite DNA fraction. A 422-bp subfragment from pErs-5, containing simple tandem repeats of [GATA]_n and [GACA]_n (as well as other sequences), was shown to account for this hybridization to mouse DNA, as well as to human DNA (Epplen et al. 1981, 1982a). Since that time, nucleotide sequences have been published for other cross-hybridizing clones from mice and *Drosophila*, revealing that these sequences all contain tandem repeats of GATA and/or GACA (Epplen et al. 1983; Singh et al. 1984).

Two central questions exist with regard to determining the evolutionary relationship of the cross-hybridizing sequences. (1) Are the sequences flanking these [GAYA]_n repeats also related? (2) Is the organization of the tandem repeats similar or different among these taxa? The rapid accumulation of DNA sequence data in recent years reveals that tandem reiterations of short repeat units are a widespread and common feature of the eukaryotic genome (for examples, see Slightom et al. 1980; Spritz et al. 1980; Jeang and Hayward 1983; Moore 1983; Rodakis et al. 1984; Tautz and Renz, 1984a, 1984b; [and references therein]; G. Levinson and G. A. Gutman, unpublished observations). It is possible that reiterations of particular simple motifs might arise independently in several taxa and therefore be analogous, not homologous. In such a case, the hybridization would not represent evolutionary conservation.

In the present article we demonstrate that all hybridization between the snake pErs-5 clone and *Drosophila* sequences can be accounted for by the presence of simple tandem repeats and that there is *no* evidence for hybridization of flanking nonrepetitive sequences. We also compared published nucleotide sequences of cross-hybridizing

snake, fly, and mouse clones. Both hybridization data and sequence comparisons suggest that the snake, fly, and mouse repetitive sequences are not conserved.

Material and Methods

Selection of *Drosophila* Clones

We screened a library prepared by Maniatis et al. (1978) containing randomly sheared embryonic DNA from *Drosophila melanogaster* inserted into a lambda bacteriophage vector. Our hybridization probe was nick-translated pErs-5, described in table 1. Ninety thousand plaques, representing nine haploid genome equivalents of *Drosophila* DNA, were screened by the procedure of Benton and Davis (1977). Hybridizations and washes were performed under low-stringency conditions, in order to recover imperfectly as well as perfectly matched clones. Hybridizations and washes were carried out at 50 C in $4 \times$ SSC (0.6 M sodium chloride, 0.06 M sodium citrate); this is ~ 40 C below average melting temperature for *Drosophila* DNA (see Burr and Schimke 1982; Beltz et al. 1983). Filters were prehybridized for 6 h without probe, then hybridized for 24 h, both times in $4 \times$ SSC containing 100 μ g/ml sheared, sonicated *Escherichia coli* carrier DNA. Filters were washed three times with $4 \times$ SSC (1.5 h per wash), dried, and autoradiographed with one intensifying screen for 16 h at -70 C.

Subcloning, Nick Translations, and Nucleotide Sequencing

Restriction fragments to be subcloned were isolated from agarose or acrylamide gels and ligated into the pUC8 vector or, for nucleotide sequencing, into the single-stranded bacteriophage vector M13 mp8/mp9 (Messing 1981).

Nick translations were performed according to the directions in a kit from Bethesda Research Labs, except that we increased our DNase concentration tenfold, to 40 ng/ml, and decreased the DNA concentration tenfold, to 2 mg/ml.

The structure of probe 3 (see below) was determined by sequencing one DNA strand by the M13 dideoxy method, as described by Sanger et al. (1980).

Analysis of Snake/Fly Cross-Hybridization

Nitrocellulose filter blots (shown in fig. 1) were prepared as follows: 30 ng of DNA from each of the eleven *Drosophila* lambda clones was digested with *Eco*R1, and another 30 ng with *Bam*H1. Each digest was electrophoresed in short 1% agarose gels. Four duplicate gels prepared in this fashion were blotted onto nitrocellulose filters by the transfer method of Southern (1975).

Each filter was hybridized to probe 1, 2, 3, or 4 (see table 1). Probe concentrations were <1 ng/ml at specific activities of 5×10^8 cpm/ μ g, and probe concentrations were adjusted so that each probe sequence was equally represented in the hybridization mixtures with respect to counts per minute. Prehybridization, hybridization, washing, and exposure conditions were identical among the four filters in all other respects. Prehybridizations, hybridizations, and washes were performed at 60 C in $4 \times$ SSC. Addition of several nucleic acid carriers was effective in preventing nonspecific hybridization, which we have otherwise observed under low-stringency conditions. Prehybridization and hybridization solutions contained the following carriers: yeast tRNA, polyriboadenylic acid, and sheared sonicated *E. coli* DNA, each at concentrations of 100 μ g/ml; plus DNase-nicked, denatured DNA from wild-type lambda, pUC8, and pBR322, each at concentrations of 1 μ g/ml. (Probe 4 also included 1 μ g/ml of unlabeled,

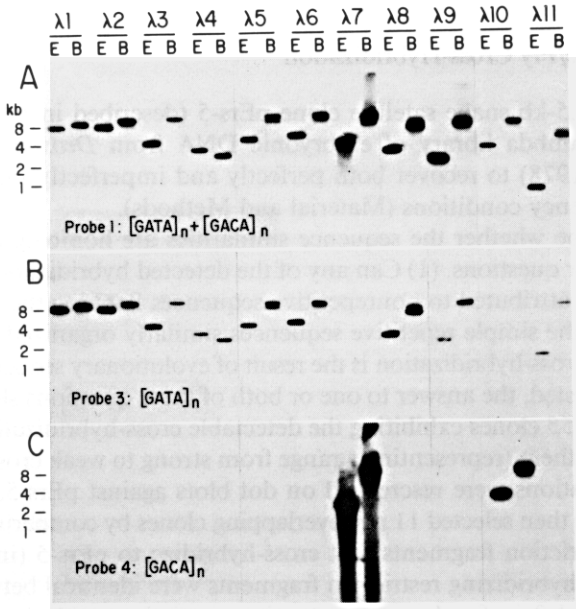


FIG. 1.—Hybridization of eleven nonoverlapping *Drosophila* lambda clones to simple repeats of GATA and GACA. DNA from each of the 11 clones was cut with *EcoRI* (*E*) or *BamHI* (*B*), and the fragments were separated on 1% agarose gels as described in the Material and Methods section. All gels were exposed for a standard length of time to allow comparison between signals and to permit detection of weak signals. Each of the probes is described in table 1. All detectable hybridization arises from tandem repeats of either GATA or GACA. Blots probed with sequences from pErs-5 flanking the tandem repeats (probe 2) showed no detectable hybridization and are therefore not shown. *A*, Probed with a 0.2-kb *MnII/DdeI* restriction fragment from snake clone pErs-5 (probe 1) containing tandem repeats of GATA and GACA (sequence shown in fig. 2*A*); *B*, probed with tandem repeats of GATA alone (probe 3); *C*, probed with tandem repeats of GACA alone (probe 4). Standard exposure led to overexposure of the background in lanes that contained strongly hybridizing fragments (lanes 7*E* and 7*B*).

DNase-nicked, denatured probe 3 DNA [1000-fold excess], to block hybridization of tandem repeats of GATA, as discussed in Results).

DNase-nicking was performed in 20 μ l at 15 C for 30 min, followed by boiling; DNase and DNA in the nicking mixture were present at concentrations of 1 μ g/ml and 50 μ g/ml, respectively.

Dot-Matrix Analysis

For dot-matrix analysis (as described by Hieter et al. 1980), the five sequences were compared in a pairwise fashion, using a program developed for the IBM PC by G. A. Gutman and B. Ward (unpublished data). The program parameters were set to score a match if 14 of 21 consecutive base pairs are identical between the two sequences. These parameters were chosen because they readily demonstrate patches of sequence similarity between genes of very distantly related organisms, while minimizing short, spurious sequence matches. In control experiments, these parameters revealed extensive alignments of matching base pairs (represented by linear arrays of dots) between (1) immunoglobulin-lambda-chain genes of chicken (Reynaud et al. 1983) and mouse (Bothwell et al. 1982) and (2) an α -tubulin gene of rat (Ginzburg et al. 1981) and a β -tubulin gene of yeast (Neff et al. 1983) while eliminating all spurious matches between randomly generated pseudo-DNA sequences, each 150 bp in length.

Results

Analysis of Snake/Fly Cross-Hybridization

Using the 2.5-kb snake satellite clone pErs-5 (described in table 1) as a probe, we screened a lambda library of embryonic DNA from *Drosophila melanogaster* (Maniatis et al. 1978) to recover both perfectly and imperfectly matched sequences under low-stringency conditions (Material and Methods).

To determine whether the sequence similarities are homologous, we set out to answer two major questions. (1) Can any of the detected hybridization between snake and fly clones be attributed to nonrepetitive sequences flanking the regions of simple repeats? (2) Are the simple repetitive sequences similarly organized in the snake and fly clones? If the cross-hybridization is the result of evolutionary sequence conservation, as has been suggested, the answer to one or both of these questions should be positive.

We isolated 55 clones exhibiting the detectable cross-hybridization and prepared DNA from 36 of them (representing a range from strong to weak cross-hybridization). These 36 preparations were rescreened on dot blots against pErs-5, and 34 of these were positive. We then selected 11 nonoverlapping clones by comparing sizes of *Eco*R1 and *Bam*H1 restriction fragments that cross-hybridize to pErs-5 (in one case where the sizes of cross-hybridizing restriction fragments were identical between two clones, i.e., lambda 1 and 2 in fig. 1, more extensive restriction analysis with a variety of enzymes showed that they were nonidentical).

We first considered whether GATA repeats or flanking regions of pErs-5 are responsible for the observed cross-hybridization to each *Drosophila* clone. We prepared multiple nitrocellulose filter blots of restriction digests from each of the 11 clones and probed two blots under low-stringency conditions with either the simple repeats (probe 1) or the flanking regions (probe 2) of pErs-5; these probes are described in table 1. Results of these experiments are shown in figure 1.

Probe 1 contains the GATA and GACA tandem repeats from pErs-5 and only 31 bp of nonrepeated flanking sequence. As shown in figure 1A, one or two restriction fragments from each fly clone cross-hybridized to probe 1 in each lane. These results demonstrate that each of the 11 *Drosophila* clones contains tandem repeats of $[GATA]_n$, $[GACA]_n$, or both. Each of the restriction fragments that cross-hybridized to the whole 2.5-kb snake clone pErs-5 (not shown) also cross-hybridized to probe 1.

Probe 2 contains *all* of the remaining 2.3 kb of snake DNA from pErs-5 that flanks probe 1 but contains *no* tandem repeats of GATA or GACA. The *Eco*R1 and *Bam*H1 restriction digests of each of the 11 *Drosophila* clones were probed under conditions identical to those used for probe 1 (see Material and Methods). We detected no hybridization to any fragment of any of the *Drosophila* clones, even after extended exposure of the autoradiograms (data not shown). These data indicate that hybridization between the snake and fly clones arises solely from the tandem repeats $[GATA]_n$ and/or $[GACA]_n$.

To determine whether $[GATA]_n$, $[GACA]_n$, or both were responsible for the observed cross-hybridization, we prepared a GATA-specific probe (probe 3) and a GACA-specific probe (probe 4); these probes are described in table 1. When the restriction digests of the 11 *Drosophila* clones were probed for tandem repeats of GATA (probe 3), nine of them (including the parent lambda 1) cross-hybridized, as shown in figure 1B. For each of the nine clones, the hybridizing fragment(s) was the same as that which hybridized to probe 1. These results show that each of these nine lambda clones contains tracts of $[GATA]_n$ of sufficient length for hybridization. Since the

remaining lambda clones (nos. 7 and 10) failed to hybridize to probe 3 (fig. 1B) but did hybridize to probe 1 (fig. 1A), their hybridization should arise solely from tandem repeats of GACA, as is confirmed below.

The above experiment does not rule out the possibility that tandem repeats of GACA might also hybridize to some of the nine GATA-containing *Drosophila* clones. To test whether any of these clones also contain tandem repeats of GACA and to confirm that clones 7 and 10 contain them, we used our GACA-specific probe 4. As shown in figure 1C, competition by unlabeled GATA tandem repeats completely blocked detectable hybridization to the nine *Drosophila* lambda clones that hybridized to probes 1 and 3. This shows that the blockage of GATA hybridization by unlabeled DNA was effective. Since blockage of GATA repeats did not interfere with hybridization to tandem repeats of GACA (see below), the results also show that tandem repeats of GATA are solely responsible for detectable hybridization to these nine clones. If tandem repeats of GACA are present in any of these clones, they are either too short (less than four or five repeats) or too highly mismatched to hybridize under low-stringency conditions.

The remaining two *Drosophila* clones (lambdas 7 and 10) did hybridize to probe 4. The hybridizing fragments are the same as those that hybridized to probe 1. This confirmed that these clones contain tandem repeats of GACA, as expected from the results above. Since these clones failed to hybridize to GATA repeats of probe 3 (fig. 1B), their hybridization to the snake clone can be entirely attributed to tandem repeats of GACA. If any tandem repeats of GATA are present in either of these clones, they are either too short or too highly mismatched to hybridize under low-stringency conditions.

In summary, the results of experiments with probes 1–4 show that all observed hybridization between the snake and fly clones arises from tandem repeats of either $[GATA]_n$ (9 of 11 clones) or $[GACA]_n$ (2 of 11 clones). The sequences flanking the tandem repeats of the snake clone show no evidence of hybridization to any of the 11 *Drosophila* clones. The results also show that the organization of tandem repeats in snake and mouse clones is different: unlike the situation in the snake satellite sequence, tracts of tandem repeats of GATA and GACA are not contiguous in any of these *Drosophila* clones.

Analysis of Published Sequence Data

We analysed pErs-5 (described in table 1) plus four other published nucleotide sequences of clones that cross-hybridize to snake satellite DNA: two from the mouse (*Mus musculus*, Epplen et al. 1983; Singh et al. 1984) and two from *Drosophila melanogaster* (Singh et al. 1981, 1984).

Figure 2 displays the GATA- and GACA-containing portions of these sequences. There is no evidence for conservation of the GATA or GACA repeat units between any of these clones: the number of repeat units in each cluster varies widely among the sequences; the tandem repeats are disrupted in a haphazard fashion by a variety of other motifs, including degenerate GATA and GACA motifs; there is no similarity between the clones with respect to either contiguous clusters of GATA and GACA or to sequences flanking the GATA and GACA repeats.

We next compared the regions flanking the GATA and GACA repeats by dot-matrix analysis. Since the sequences contain a variety of simple repetitive sequences, we limited our search to similarity over segments 21 nucleotides long. We compared the entire published sequence of each of the five clones in a pairwise fashion and

A

... TCAG GATA TAAGTCTGGAAGC [GATA]₁₅ [GACA]₂ GACT AAAAGC TATA [GATA]₁₀
 [GACA]₁₀ ...

B

... CACA CGCA GACA GATA [CATA]₅-TATA CATA-[CA]₅ GACT GATA GAT [GATA]₂ CATA
 [GATA]₂ GAT GATA GGTA GATA GAT GAGT A [GACA]₂ TAT GATA GGTA GAT GATA TA
 GACA GCATAT GATA GGTA GATG GATG ATA [GACA]₂ TAT GATA GGTA [GATA]₄ [GACA]₂
 GATA GATG AAA [GACA]₂ TAT [GATA]₆ CATA ([GATA]₄-[GACA]₃ TAT) ([GATA]₄
 [GACA]₃ TAT) GATA GACA ([GATA]₂ [GACA]₃ T) ([GATA]₂ [GACA]₃ T) [GATA]₁₄ GAT
 [GATA]₅ GAT GATA [GACA]₂ TAT [GATA]₁₂ TA [GACA]₃ TAT GATA GACA [GATA]₂
 [GACA]₃ TAT [GATA]₇ TA GATA GAT [GATA]₂ [GACA]₂ TAT [GATA]₅-[GA]₉ AA [GA]₃
 [GATA]₆ TGTT ...

C

... TTCTA GATA AAT [GATA]₃ GACA ACAGAA GATA CATA [GATA]₈ GAT [GATA]₂ GAT
 [GATA]₄ AATTAACACGTAGGGAGGTA GATA GACT TA GACA GAGC ...

D

... [GTTT]₂ ATTGCGAGT [GAT]₂ GAA [TA]₃ ATATAGTA GATA TATA GATA
 GATGTATAGATGGTTA [GATA]₂ GGTA [GATA]₅ GATG TAA [GATA]₇ GATG GTTA [GATA]₁₃
 TATAGATTATA GACA GCTA [GATA]₅ TATA GATA [TA]₅ [GATA]₂ GGTA GATA TATA
 [GATA]₂ TGTA GATA TATA GATA [TA]₅-GATT GATA-GGTA [GATA]₄ GATG [GATA]₂ GGTA
 AATA [GATA]₂ TATA [GATA]₆ GACA GATA AATA (GATA TAAA GATA)(GATA TAAA GATA)
 [GATA]₂ GATCAACCAATA GATA GATG TTA [GATA]₃ GTTT GATT ...

E

[TA]₅-[GATA]₄ GAGA [GATA]₁₅-[GA]₃-[GATA]₄ CATA [GATA]₁₀ GATG GATA AATA
 [GATA]₃ GATT ...

FIG. 2.—Portions of published sequences containing GATA and GACA repeats. Sequence spacing emphasizes a variety of short tandem repeat units, both perfect and imperfect, that occur in these regions. Some longer tandem duplications in *B* and *D* are enclosed in parentheses. Contiguous clusters of tandem repeats differing by a single base are underlined; clusters differing by a single base that are separated by only a few base pairs are designated by broken underlines (see Discussion). *A*, From 2.5-kb snake clone pERS-5 (Epplen et al. 1982a); *B*, inverse complement from 2.5-kb mouse clone pMC14 (Epplen et al. 1983); *C*, from mouse subclone M 3.1 (Singh et al. 1984); *D*, From *Drosophila* subclone CS314 2(8) (Singh et al. 1984); *E*, from *Drosophila* subclone CS319 (Singh et al. 1984).

found only a few scattered matches between the clones, matches similar in length and number to those we found when comparing different parts of each clone with itself. We did not find any visible alignments characteristic of conservation.

Discussion

We have isolated 11 groups of nonoverlapping *Drosophila* clones that cross-hybridize to the snake satellite clone pErs-5. By hybridization of representatives of each of these groups to defined probes, we have shown that all of the detectable cross-hybridization can be attributed to tandem repeats of either GATA or GACA and found no evidence for hybridization between the nonrepetitive sequences flanking these simple repeats. Since the flanking sequences did not appear to be conserved, we determined whether the organization of the simple repeats themselves is indicative of evolutionary conservation between snake and fly and found it to differ between these species: the snake clone contains long contiguous tandem repeats containing both GATA and GACA, while each of the *Drosophila* clones contains long tandem repeats of either GATA or GACA alone. These data suggest that these structures are not evolutionarily conserved between these taxa but cross-hybridize simply because they happen to contain tandem repeats of the same 4-bp motifs.

We have also compared published sequences from DNA clones from snakes, mice, and *Drosophila* that exhibit hybridization to snake satellite DNA and found that in each species the organization of the simple repeats differs from those of the others with respect to the number of clusters of tandem repeats, number of repeat units per cluster, occurrence of imperfect repeats, nature of the sequences that flank each of the clusters, and contiguity (or presence) of tandem repeats of GATA and GACA. Dot-matrix comparisons of the sequences flanking the simple repeats revealed no visible alignments suggesting conservation. These comparisons of sequence data corroborate our snake/fly cross-hybridization data, making the argument for evolutionary conservation of snake and fly and mouse sequences exceedingly weak. Of course, we have not ruled out the possibility that some of the sequences are derived from common ancestral seeds of tandem repeats; this is not implausible for the large clusters of GATA repeats localized on the W and Y chromosomes of snakes and mice, respectively (suggestions that these sequences may be involved in primary sex determination have been referenced in the introduction). Nevertheless, the only feature shared among the clones is the presence of long, homogeneous clusters of tandem repeats, some of which share simple motifs. These observations raise questions concerning both the mechanisms by which long iterations of similar simple motifs have evolved among these divergent taxa and the possible functions (if any) that they might serve.

As shown in figure 2, the snake, fly, and mouse clones each contain a different assortment of closely related motifs, including GATA, GACA, CATA, GA, TA, and CA. Of special interest are regions in which clusters of tandem repeats of two motifs that differ by a single base pair are contiguous (fig. 2, underlined sequences) or are separated by only a few base pairs (fig. 2, broken-underlined sequences). Examples include $[GATA]_n$ and $[GACA]_n$, $[CATA]_n$ and $[CA]_n$, $[GATA]_n$ and $[GA]_n$, and $[GATA]_n$ and $[TA]_n$. Each of these examples could be explained as being a result of base substitutions that have been propagated by multiple rounds of short tandem duplications. We therefore suggest that these tandem repeats in particular and simple repetitive sequences in general may be generated by mechanisms involving slipped-strand mispairing of DNA. Slipped-strand mispairing has been previously invoked to

explain a variety of changes in chromosomal DNA, including tandem reiterations (Kornberg et al. 1964; Jones and Kafatos 1982; Moore 1983; Rodakis et al. 1984; Tautz and Renz 1984*a*; 1984*b*), frameshift mutations (Streisinger et al. 1966; Farabaugh et al. 1978), duplications and deletions (Efstradiatis et al. 1980; Slightom et al. 1980; Flanagan et al. 1984), and illegitimate recombination and chromosomal integration (Hasson et al., 1984). Tandem duplications of the short repeat units could occur either by mispairing of the growing strand with the new strand during replication (Kornberg et al. 1964; Streisinger et al. 1966; Kornberg 1980) or repair of mispaired sequences in supercoiled DNA. A detailed discussion of the mechanisms by which slipped-strand mispairing could generate a variety of simple repetitive motifs, including extension of base-substituted motifs, will be presented elsewhere (G. Levinson and G. A. Gutman, unpublished observations).

A number of larger tandem duplications found in mouse and *Drosophila* sequences, each of which is made up of several smaller repeat units, are enclosed in parentheses in figure 2. These may have arisen from noncontiguous mispairing between the short simple repeat units. Another possibility is that they are the result of unequal crossing-over, as has been suggested for duplications and rearrangements of tandem repeats in silk fibroin genes (Rodakis et al. 1984).

Although the clusters of tandem repeats from snakes, mice, and *Drosophila* are organized differently, each sequence contains long tracts of perfect or near-perfect tandem repeats. Two possible reasons for this length and homogeneity are that (1) the rate of accumulation of the simple repeats may exceed that of other mutations in these regions or (2) these qualities may be favored by natural selection if the sequences have specialized functions, such as modification of the expression of nearby genes, as has been shown for genetically engineered constructs containing simple repeats (Hamada et al. 1984*a*, 1984*b*). These two possibilities are not mutually exclusive; they will be discussed in more detail elsewhere (G. Levinson and G. A. Gutman, unpublished observations).

We draw the following conclusions from this study. First, we find no evidence for evolutionary conservation of the snake, fly, and mouse clones. Second, we suggest that these (and other) simple tandem repeats may have arisen by a process involving slipped-strand mispairing of the two strands of DNA. Third, since production of new repeat units by this mechanism should be a self-accelerating process, it is possible that the occurrence of long stretches of various perfect or near-perfect tandem repeats (including CATA, TA, CA, and GA, as well as GATA and GACA) is simply the result of a rate of accumulation that exceeds that of other mutations in those regions. Our data fail to support the existence of evolutionary conservation between these simple repetitive sequences (although we have not ruled it out). Therefore, these simple repeats may have arisen independently, and the homogeneity and/or length of particular tandem-repeat clusters could be the result of parallel or convergent evolution as a consequence of natural selection for presently unknown functions.

Acknowledgments

The authors thank A. Cellini for assistance with nucleotide sequencing. We also thank M. B. Frank for helpful discussions and criticisms concerning an early draft of the manuscript and G. L. G. Miklos and P. Simpson for stimulating discussions and correspondence during its preparation. This research was supported by U.S. Public

Health Service grants HD16519 (J.L.M.) and HAI14774 (G.A.G.). G. Levinson was supported by NIH Research Service award HD07029 and departmental stipend funds.

LITERATURE CITED

- BELTZ, G. A., K. A. JACOBS, T. H. EICKBUSH, P. T. CHERBAS, and F. C. KAFATOS. 1983. Isolation of multigene families and determination of homologies by filter hybridization methods. *Methods Enzymol.* **100**:266–284.
- BENTON, W. D., and R. W. DAVIS, 1977. Screening λ -gt recombinant clones by hybridization to single plaques *in situ*. *Science* **196**:180–182.
- BOTHWELL, A. L. M., M. PASKIND, T. IMANISHI-KARI, M. RETH, K. RAJEWSKY, and D. BALTIMORE. 1982. Somatic variants of immunoglobulin lambda light chains. *Nature* **298**:380–382.
- BURR, H. E., and R. T. SCHIMKE. 1982. Reduced-stringency DNA reassociation: sequence specific duplex formation. *Nucleic Acids Res.* **10**:719–733.
- EFSTRADIATIS, A., J. W. POSAKONY, T. MANIATIS, R. M. LAWN, C. O'CONNELL, R. A. SPRITZ, J. K. DERIEL, B. G. FORGET, S. W. WEISSMAN, J. L. SLIGHTOM, A. E. BLECHL, O. SMITHIES, F. E. BARALLE, C. C. SHOULDERS, and N. J. PROUDFOOT. 1980. The structure and evolution of the human β -globin gene family. *Cell* **21**:653–668.
- EPPLEN, J. T., A. CELLINI, S. ROMERO, and S. OHNO. 1983. An attempt to approach the molecular mechanisms of primary sex determination: W- and Y-chromosomal conserved simple repetitive DNA sequences and their differential expression in mRNA. *J. Exp. Zool.* **228**:305–312.
- EPPLEN, J. T., J. R. MCCARREY, S. SUTOU, and S. OHNO. 1982a. Base sequence of a cloned snake W-chromosome DNA fragment and identification of a male-specific putative mRNA in the mouse. *Proc. Natl. Acad. Sci. USA* **79**:3798–3802.
- . 1981. Sex-determining genes and gene regulation. Pp. 239–251 in M. J. NOVY and J. A. RESKO, eds. *Proceedings of the Oregon Regional Primate Research Center Symposium on Reproductive Biology. Vol. 1. Fetal endocrinology.* Academic Press, New York.
- . 1982b. Is sex-specifically arranged repetitive DNA involved in primary sex determination of vertebrates? Pp. 317–326 in B. BONNETAMIR, ed. *Human genetics. Part A: the unfolding genome.* Liss, New York.
- FARABAUGH, P. J., U. SCHMEISSNER, M. HOFER, and J. H. MILLER. 1978. Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J. Mol. Biol.* **126**:847–863.
- FLANAGAN, J. G., M.-P. LEFRANC, and T. H. RABBITS. 1984. Mechanisms of divergence and convergence of the human immunoglobulin $\alpha 1$ and $\alpha 2$ constant region gene sequences. *Cell* **36**:681–688.
- GINZBURG, I., L. BEHAR, D. GIVOL, and U. Z. LITTAUER. 1981. The nucleotide sequence of rat α -tubulin: 3'-end characteristics, and evolutionary conservation. *Nucleic Acids Res.* **9**: 2691–2697.
- HAMADA, H., M. G. PETRINO, T. KAKUNAGA, M. SEIDMAN, and B. D. STOLLAR. 1984a. Characterization of genomic poly(dT-dG)·poly(dC-dA) sequences: structure, organization, and conformation. *Mol. Cell. Biol.* **4**:2610–2621.
- HAMADA, H., M. SEIDMAN, B. H. HOWARD, and C. M. GORMAN. 1984b. Enhanced gene expression by the poly(dT-dG)·poly(dC-dA) sequence. *Mol. Cell. Biol.* **4**:2622–2630.
- HASSON, J.-F., E. MOUGNEAU, F. CUZIN, and M. YANIV. 1984. Simian virus 40 illegitimate recombination occurs near short direct repeats. *J. Mol. Biol.* **177**:53–68.
- HIETER, P. A., E. E. MAX, J. G. SEIDMAN, J. V., MAIZEL, JR., and P. LEDER. 1980. Cloned human and mouse kappa immunoglobulin constant and J region genes conserve homology in functional segments. *Cell* **22**:197–207.
- JEANG, K. T., and G. S. HAYWARD. 1983. A cytomegalovirus DNA sequence containing tracts

- of tandemly repeated CA dinucleotides hybridizes to highly repetitive dispersed elements in mammalian cell genomes. *Mol. Cell. Biol.* 3:1389-1402.
- JONES, C. W., and F. C. KAFATOS. 1982. Accepted mutations in a gene family: evolutionary diversification of duplicated DNA. *J. Mol. Evol.* 19:87-103.
- JONES, K. W., and L. SINGH. 1981*a*. Conserved repeated DNA sequences in vertebrate sex chromosomes. *Hum. Genet.* 58:46-53.
- . 1981*b*. Conserved sex-associated repeated DNA in vertebrates. Pp. 135-154 in G. DOVER and R. FLAVELL, eds. *Genome evolution*. Academic Press, London.
- KORNBERG, A. 1980. DNA replication. W. H. Freeman, San Francisco.
- KORNBERG, A., L. L. BERTSCH, J. F. JACKSON, and H. G. KHORANA. 1964. Enzymatic synthesis of deoxyribonucleic acid. XVI. Oligonucleotides as templates and the mechanism of their replication. *Proc. Natl. Acad. Sci. USA* 51:315-323.
- MANIATIS, T., R. C. HARDISON, E. LACY, J. LAUER, C. O'CONNELL, D. QUON, G. K. SIM, and A. EFSTRADIATIS. 1978. The isolation of structural genes from libraries of eukaryotic DNA. *Cell* 15: 687-701.
- MESSING, J. 1981. M13mp2 and derivatives: a molecular cloning system for DNA sequencing, strand-specific hybridization, and *in vitro* mutagenesis. Pp. 143-153 in A. WALTON, ed. *Third Cleveland symposium on macromolecules: recombinant DNA*. Elsevier, Amsterdam.
- MOORE, G. P. 1983. Slipped-strand mispairing and the evolution of introns. *Trends Biochem. Sci.* 8:411-414.
- NEFF, N. F., J. H. THOMAS, P. GRISAFI, and D. BOTSTEIN. 1983. Isolation of the β -tubulin gene from yeast and demonstration of its essential function *in vivo*. *Cell* 33:211-219.
- REYNAUD, C.-A., A. DAHAN, and J.-C. WEILL. 1983. Complete sequence of a chicken λ light chain immunoglobulin derived from the nucleotide sequence of its mRNA. *Proc. Natl. Acad. USA* 80:4099-4103.
- RODAKIS, G. C., R. LECANIDOU, and T. H. EICKBUSH. 1984. Diversity in a chorion multigene family created by tandem duplications and a putative gene-conversion event. *J. Mol. Evol.* 20:265-273.
- SANGER, F., A. R. COULSON, B. G. BARRELL, A. J. H. SMITH, and B. A. ROE. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* 143:161-178.
- SINGH, L., C. PHILLIPS, and K. W. JONES. 1984. The conserved nucleotide sequences of Bkm, which define Sxr in the mouse, are transcribed. *Cell* 36:111-120.
- SINGH, L., I. F. PURDOM, and K. W. JONES. 1981. Conserved sex-chromosome-associated nucleotide sequences in eukaryotes. *Cold Spring Harbor Symp. Quant. Biol.* 45:805-814.
- SLIGHTOM, J. L., A. E. BLECHL, and O. SMITHIES. 1980. Human fetal G γ - and A γ -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21:627-638.
- SOUTHERN, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
- SPRITZ, R. A., J. K. DERIEL, B. G. FORGET, and S. M. WEISSMAN. 1980. Complete nucleotide sequence of the human δ -globin gene. *Cell* 21:639-646.
- STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA, E. TERZHAGHI, and M. INOUE. 1966. Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* 31:77-84.
- TAUTZ, D., and M. RENZ. 1984*a*. Simple DNA sequences of *Drosophila virilis* isolated by screening with RNA. *J. Mol. Biol.* 172:229-235.
- . 1984*b*. Simple sequences are ubiquitous components of eukaryotic genomes. *Nucleic Acids Res.* 12:4127-4138.

WALTER M. FITCH, reviewing editor

Received January 31, 1985; revision received June 10, 1985.