MOLECULAR AND CELLULAR BIOLOGY, Aug. 1986, p. 3010–3013
0270-7306/86/083010-04$02.00/0

Vol. 6, No. 8

# The Ubiquitous Potential Z-Forming Sequence of Eucaryotes, $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$, Is Not Detectable in the Genomes of Eubacteria, Archaebacteria, or Mitochondria

DAVID S. GROSS†* AND WILLIAM T. GARRARD

*Division of Molecular Biology, Department of Biochemistry, The University of Texas Health Science Center at Dallas, Dallas, Texas 75235*

Received 9 January 1986/Accepted 22 April 1986

**The potential Z-forming sequence $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$ is an abundant, interspersed repeat element that is ubiquitous in eucaryotic nuclear genomes. We report that in contrast to eucaryotic nuclear DNA, the genomes of eubacteria, archaebacteria, and mitochondria lack this sequence, since even a single tract of ≥14 base pairs in length is not detectable through either hybridization or sequence analysis. Interestingly, the phylogenetic distribution of the $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$ repeat exhibits a striking parallel to that of $(dT\text{-}dC)_n \cdot (dG\text{-}dA)_n$, but not to other homocopolymeric sequences such as $(dC\text{-}dG)_n \cdot (dC\text{-}dG)_n$ or $(dT\text{-}dA)_n \cdot (dT\text{-}dA)_n$.**

The simple sequence $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$ is of unusual interest for at least two reasons: first, it is a highly reiterated, interspersed element that is ubiquitous in eucaryotic genomes (6, 18, 32, 36); and second, it is capable of converting to the left-handed, Z-DNA conformation when subjected in vitro to physiological levels of negative torsional stress (10, 20, 29). While the role of $d(CA)_n$ elements in eucaryotic nuclear genomes is unknown, several functions have been suggested. These include (i) acting as hot spots in homologous recombination (21, 30); (ii) facilitating exon shuffling (9) and proviral integration (31); (iii) promoting homogenization of repetitive gene arrays (11); (iv) preserving the ends of DNA molecules during DNA replication (28); (v) regulating gene transcription (8, 26); and (vi) structuring the nucleolus (37). We recently examined the chromatin structure of $d(CA)_n$ elements in cultured mammalian cells and obtained evidence that these sequences do not exist to a significant extent in the Z state in vivo; instead, they appear to quantitatively adopt a distinctive, "alternating-B" conformation on the nucleosomal surface (5).

To gain further insight into the possible function(s) of $d(CA)_n$ elements, we address the question of whether these sequences are also present in procaryotic genomes. Archaebacteria are of particular interest (2, 4, 39), since they possess a number of eucaryotic characteristics not found in eubacteria, including repeated sequence elements (27), eucaryote-like tRNA genes (13), mammal-like 7S RNA genes (19), and introns within tRNA (13), rRNA (14), and protein-coding (2) genes. Nonetheless, we demonstrate by sequence analysis and the use of a highly sensitive dot hybridization assay that even single $d(CA)_n$ tracts as short as 14 base pairs (bp) are absent from archaebacterial genomes, as well as from those of eubacteria and mitochondria.

**$d(CA)_n$ sequences are not detectable in non-nuclear genomes.** We have investigated whether $d(CA)_n$ sequences are present in *Saccharomyces cerevisiae* mitochondrial DNA or in the genomes of 13 phylogenetically diverse species of procaryotes, representative of both eubacteria and archaebacteria (2–4), by using a dot hybridization assay that allows the accurate quantitation of (12) and specific hybridization to (5) $d(CA)_n$ sequences. We used three hybridization stringencies that detect tracts of ≥52, ≥26, or ≥14 nucleotides and monitored the hybridization signals obtained from three different loads of each DNA sample (33, 100, and 333 ng) (Fig. 1 and 2). Sequential low-, moderate-, and high-stringency hybridization washes were conducted at 45, 50, and 65°C, respectively, with wash buffers containing either 0.207 M $Na^+$ (low stringency) or 0.032 M $Na^+$ (moderate and high stringency) as described previously (5). Mouse and *S. cerevisiae* nuclear genomes were included as positive controls, and pBR322, which by sequence analysis lacks $d(CA)_n$ blocks, was included as a negative control.

For purposes of calibration, we constructed three different mixtures of pJS5, a recombinant plasmid containing one $d(CA)_{33}$ block (25), with lambda DNA, which lacks $d(CA)_n$ sequences of >8 bp. The 33-, 100-, and 330-ng samples contained, respectively, 3, 10, and 30 copies of $d(CA)_{33}$ per *Escherichia coli* genome equivalent (Fig. 1 and 2, bottom row). While it was previously reported that $d(CA)_n$ sequences were not abundant in the *E. coli* genome (6, 35, 36), it is apparent that our assay is sensitive enough to readily detect one copy of $d(CA)_n$ of ≥14 nucleotides per eubacterial genome, and no reproducible hybridization signal above background is exhibited by any of the procaryotic or mitochondrial DNA samples (Fig. 1 and 2). In contrast, both yeast and mouse nuclear DNA samples possess readily detectable levels of these sequences, in agreement with previous reports (6, 35, 36). Therefore, $d(CA)_n$ sequences appear to be absent from nonnuclear genomes.

**Analysis of the GenBank database confirms that mitochondrial genomes lack $d(CA)_n$ sequences and supports the conclusion that these elements are absent from procaryotic genomes.** To complement the above results, we surveyed the GenBank database for the presence of $d(CA)_n$ elements ($n \geq 5$) in all available procaryotic and eucaryotic nucleic acid sequences. These results reveal that $d(CA)_n$ tracts of ≥14 nucleotides are absent from procaryotic sequences, the complete hu-

* Corresponding author.
† Present address: Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, Shreveport, LA 71130.

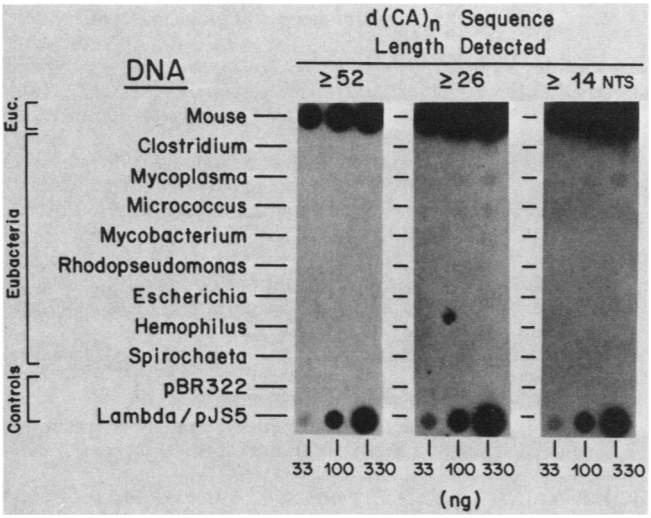


FIG. 1. d(CA)$_n$ sequences of ≥14 bp are not detected in eubacterial genomes. DNA samples were denatured, dot blotted onto Zeta-Probe nylon membrane (Bio-Rad) in the amounts indicated, and then subjected to hybridization with poly(dT-dG) · poly([$^{32}$P]dC-dA) essentially as described previously (5), except that sonicated lambda DNA (20 μg/ml) and 0.1% sodium dodecyl sulfate were included. Following hybridization, blots were subjected to three sequential, increasingly stringent washes, resulting in conditions approximating the solution $T_m$ of d(CA)$_7$, d(CA)$_{13}$, or d(CA)$_{26}$ (5). DNA samples selected for immobilization possessed single-strand lengths of ≥1 kilobase, as assayed by alkaline agarose gel electrophoresis (17) and were quantitated by the Hoechst 33258 fluorescence assay (16) with fluorescence enhancement corrected for variation in A+T content. Preparation of copy number control DNA (lambda/pJS5) is described in the text.

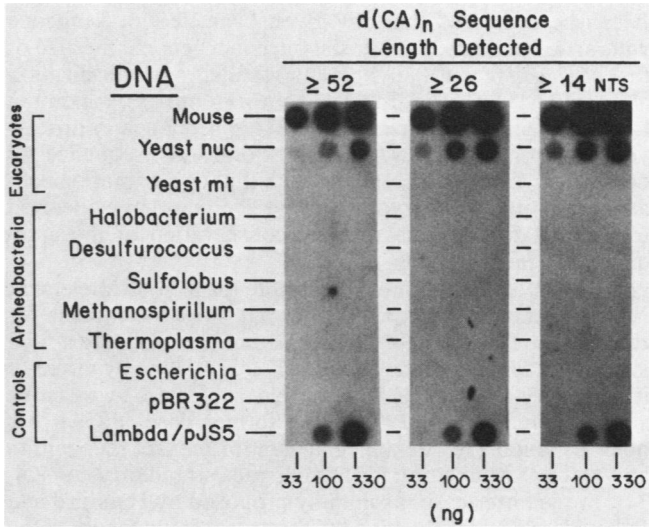


FIG. 2. d(CA)$_n$ sequences of ≥14 bp are not detected in archaebacterial or mitochondrial genomes. Dot hybridization and preparation of DNA samples were as described for Fig. 1. Densitometric analysis indicated that mouse and yeast haploid nuclear genomes, respectively, contain ca. 1.5 × 10$^5$ and ca. 10 d(CA)$_n$ elements of ≥52 bp. In addition, the yeast haploid genome contains approximately 50 d(CA)$_{13}$ and 100 d(CA)$_7$ tracts.

man, bovine, and mouse, mitochondrial genomes, selected yeast mitochondrial sequences (totalling over 50 kb), and available chloroplast sequences (ca. 40 kb) (Table 1). In fact, the longest d(CA)$_n$ tract found in a procaryotic sequence is exactly 10 bp in length, and only three of these have been reported from a total sequence pool of 5 × 10$^5$ bp (Table 1). In contrast, at least 77 examples of d(CA)$_n$ elements of ≥10 bp in length were found in eucaryotic nuclear DNA sequences, from a sequence pool that is only 4.5-fold larger. Perhaps more significantly, at least 20 examples of d(CA)$_n$ tracts of ≥26 bp and at least 5 examples of d(CA)$_n$ tracts of ≥52 bp have also been identified in these nuclear DNA sequences.

Of further interest, previous studies have indicated that d(CA)$_n$ elements tend to cluster with each other and with tracts of (dT-dC)$_n$ · (dG-dA)$_n$ in eucaryotic genomes (11, 25, 35). Therefore, the phylogenetic distribution of d(TC)$_n$ was similarly searched and was found to exhibit a striking parallel to that of d(CA)$_n$ with respect to its abundance in eucaryotic nuclear genomes and absence in procaryotic and organellar genomes (Table 1). In contrast, the other potential Z-forming homocopolymeric sequence, (dC-dG)$_n$ · (dC-dG)$_n$, appears to be absent from all genomes, both procaryotic and eucaryotic (Table 1). This finding, together with the hybridization artifacts associated with the detection of d(CG)$_n$ sequences (discussed in references 5 and 7), casts doubts on previous studies (6, 35) which purport to show the prevalent occurrence of d(CG)$_n$ in eucaryotic genomes.

The genomic distribution of a non-Z-forming homocopolymeric sequence, (dT-dA)$_n$ · (dT-dA)$_n$, was also examined (Table 1). Unlike d(CA)$_n$ or d(TC)$_n$, d(TA)$_n$ exists almost exclusively as short sequence blocks (<26 bp) in both nuclear and mitochondrial DNA, and its frequency appears to roughly parallel the A+T content of the genomes in which it is found.

**Possible eucaryotic-specific function for d(CA)$_n$ sequences.**

TABLE 1. Abundance of simple sequences in procaryotic and eucaryotic genomes[a]

| Genome class (approx no. of unique sequences) | Nucleotides searched | No. of tracts of[b]: ≥14 nt | | | | ≥26 nt | | | | ≥52 nt[c] | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | CA | TC | CG | TA | CA | TC | CG | TA | CA | TC | CG | TA |
| Procaryotic[d] (450) | 5 × 10$^5$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Eucaryotic, nuclear (2,600) | 2.25 × 10$^6$ | 37 | 34 | 2 | 21 | 20 | 18 | 0 | 1 | 5 | 8 | 0 | 0 |
| Eucaryotic, organellar[e] (150) | 2.5 × 10$^5$ | 0 | 0 | 0 | 56 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

[a] Sequence data obtained from the National Institutes of Health GenBank database (February 1985 update) and references 1 and 33. Searches were performed on both upper and lower strands by using the sequence analysis program of Queen and Korn (23).

[b] nt, Nucleotides.

[c] Mismatches of ≤5% included.

[d] Procaryotic chromosomal DNA contains three examples of d(CA)$_5$, one example of d(TC)$_5$, six examples of d(CG)$_5$, and one example of d(TA)$_6$. These represent the longest stretches of each element thus far found.

[e] Includes both mitochondrial and chloroplast DNA sequences. The longest tracts of simple sequence identified in mitochondrial DNA are d(CA)$_5$, d(TC)$_4$, d(CG)$_4$, and d(TA)$_{15}$. For chloroplast DNA, the corresponding values are d(CA)$_3$, d(TC)$_4$, d(CG)$_3$, and d(TA)$_5$.

It can be estimated that any given 11-nucleotide sequence will, on a stochastic basis, exist once in a genome the size of that of *E. coli* (ca. $4 \times 10^6$ bp). Our results from hybridization and sequence analysis are consistent with this as a maximum frequency of occurrence of the $d(CA)_n$ sequence in procaryotic and organellar DNA. The lack of $d(CA)_n$ sequences in nonnuclear DNA therefore suggests that one or more mechanisms unique to eucaryotic nuclear genomes have operated to account for the evolutionary conservation of these sequence elements.

We believe that the most likely role for the d(CA) element is during meiosis, when it could act as a focal point for recombination, either through its potential to undergo strand slippage independently of torsional stress (11) or by virtue of its capacity to form Z-DNA when subjected to negative supercoiling (10, 20, 29). In the former case, $d(CA)_n$ sequences could provide a mechanism for inexact recognition and serve as a substrate for homologous strand invasion (24, 34). In the latter case, as initially proposed by Haniford and Pulleyblank (9), left-handed $d(CA)_n$ tracts could facilitate the correct pairing of homologous chromosomes during meiotic recombination. This idea is supported by recent work of Kmiec and Holloman (15), which indicates that the synaptic pairing reaction preferentially initiates at stretches of Z-DNA of homologous sequence. Indeed, it has recently been shown that the presence of $d(CA)_n$ blocks significantly increase the frequency of reciprocal meiotic exchange between homologous yeast chromosomes (38; D. Treco and N. Arnheim, personal communication). It is possible that $d(TC)_n$ elements, which exhibit a similar degree of reiteration (Table 1) and clustering (11, 25, 35) with $d(CA)_n$ tracts, also facilitate meiotic recombination, since these sequences have the potential to adopt a Z (or Z-like) conformation as well (22).

We thank the following for their gifts of procaryotic chromosomal DNA: M. Norgard, University of Texas Health Science Center, Dallas (*Mycobacterium chelonei, Hemophilus influenza*); G. Fox, University of Houston (*Halobacterium marismortui, Desulfuroccus mobilis*); C. R. Woese and B. P. Kaine, University of Illinois (*Sulfolobus solfataricus, Methanospirillum hungatei*); J. Baseman, University of Texas Health Science Center, San Antonio (*Mycoplasma pneumoniae*); E. P. Greenberg and B. Brahansha, Cornell University (*Spirochaeta aurantia*); S. Kaplan and T. Donahue, University of Illinois (*Rhodopseudomonas sphaeroides*); and D. Searcy, University of Massachusetts (*Thermoplasma acidophilum*). We also thank A. Gilliam and P. Tucker for their gift of recombinant clone pJS5; C. Szent-Gyorgyi for stimulating discussions; M. Norgard for helpful suggestions; E. Hernandez for technical assistance; and S. Gross and S. Alexander for assistance in preparing the manuscript.

D.S.G. was the recipient of a National Institutes of Health postdoctoral fellowship. This work was supported by Public Health Service grants GM22201, GM29935, GM25829, and GM31689 from the National Institutes of Health and grant I-823 from the Robert A. Welch Foundation.

## LITERATURE CITED

1. **Craik, C. S., Q.-L. Choo, G. H. Swift, C. Quinto, R. J. MacDonald, and W. J. Rutter.** 1984. Structure of two related rat pancreatic trypsin genes. J. Biol. Chem. **259**:14255–14264.
2. **Doolittle, W. F.** 1985. Archaebacteria coming of age. Trends Genet. **1**:268–269.
3. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Leuhrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209**:457–463.
4. **Garrett, R. A.** 1985. The uniqueness of archaebacteria. Nature (London) **318**:233–235.
5. **Gross, D. S., S.-Y. Huang, and W. T. Garrard.** 1985. Chromatin structure of the potential Z-forming sequence $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$: evidence for an "alternating-B" conformation. J. Mol. Biol. **183**:251–265.
6. **Hamada, H., M. G. Petrino, and T. Kakunaga.** 1982. A novel repeated element with Z-DNA forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA **79**:6465–6469.
7. **Hamada, H., M. G. Petrino, T. Kakunaga, M. Seidman, and B. D. Stollar.** 1984. Characterization of genomic poly(dT-dG) · poly(dC-dA) sequences: structure, organization, and conformation. Mol. Cell. Biol. **4**:2610–2621.
8. **Hamada, H., M. Seidman, B. H. Howard, and C. M. Gorman.** 1984. Enhanced gene expression by the poly(dT-dG) · poly(dC-dA) sequence. Mol. Cell. Biol. **4**:2622–2630.
9. **Haniford, D. B., and D. E. Pulleyblank.** 1983. The *in vivo* occurrence of Z-DNA. J. Biomol. Struct. Dynam. **1**:593–609.
10. **Haniford, D. B., and D. E. Pulleyblank.** 1983. Facile transition of poly[d(TG) · d(CA)] into a left-handed helix in physiological conditions. Nature (London) **302**:632–634.
11. **Hentschel, C. C.** 1982. Homocopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to $S_1$ nuclease. Nature (London) **295**:714–716.
12. **Kafatos, F. C., C. W. Jones, and A. Efstratiadis.** 1979. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res. **7**:1541–1552.
13. **Kaine, B. P., R. Gupta, and C. R. Woese.** 1983. Putative introns in tRNA genes of prokaryotes. Proc. Natl. Acad. Sci. USA **80**:3309–3312.
14. **Kjems, J., and R. A. Garrett.** 1985. An intron in the 23S ribosomal RNA gene of the archaebacterium *Desulfurococcus mobilis*. Nature (London) **318**:675–677.
15. **Kmiec, E. B., and W. K. Holloman.** 1986. Homologous pairing of DNA molecules by *Ustilago* rec1 protein is promoted by sequences of Z-DNA. Cell **44**:545–554.
16. **Labarca, C., and K. Paigen.** 1980. A simple, rapid, and sensitive DNA assay procedure. Anal. Biochem. **102**:344–352.
17. **McDonell, M. W., M. N. Simon, and F. W. Studier.** 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. **110**:119–146.
18. **Miesfeld, R., M. Krystal, and N. Arnheim.** 1981. A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ and β globin genes. Nucleic Acids Res. **9**:5931–5947.
19. **Moritz, A., and W. Goebel.** 1985. Characterization of the 7S RNA and its gene from halobacteria. Nucleic Acids Res. **13**:6969–6979.
20. **Nordheim, A., and A. Rich.** 1983. The sequence $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ forms left-handed Z-DNA in negatively supercoiled plasmids. Proc. Natl. Acad. Sci. USA **80**:1821–1825.
21. **Proudfoot, N. J., and T. Maniatis.** 1980. The structure of a human α-globin pseudogene and its relationship to α-globin gene duplication. Cell **21**:537–544.
22. **Pulleyblank, D. E., D. B. Haniford, and A. R. Morgan.** 1985. A structural basis for S1 nuclease sensitivity of double-stranded DNA. Cell **42**:271–280.
23. **Queen, C., and L. J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12**:581–599.
24. **Radding, C. M.** 1982. Homologous pairing and strand exchange in genetic recombination. Annu. Rev. Genet. **16**:405–437.
25. **Richards, J. E., A. C. Gilliam, A. Shen, and P. W. Tucker.** 1983. Unusual sequences in the murine immunoglobulin μ-δ heavy chain region. Nature (London) **306**:483–487.
26. **Santoro, C., F. Costanzo, and G. Ciliberto.** 1984. Inhibition of eukaryotic tRNA transcription by potential Z-DNA sequences. EMBO J. **3**:1553–1559.
27. **Sapienza, C., and W. F. Doolittle.** 1982. Repeated sequences in

the genomes of halobacteria. Zbl. Bakt. Hyg., I Abt. Orig. C 3:120–127.
28. **Shampay, J., J. W. Szostak, and E. H. Blackburn.** 1984. DNA sequences of telomeres maintained in yeast. Nature (London) 310:154–157.
29. **Singleton, C. K., M. W. Kilpatrick, and R. D. Wells.** 1984. S1 nuclease recognizes DNA conformational junctions between left-handed helical $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$ and contiguous right-handed sequences. J. Biol. Chem. 259:1963–1967.
30. **Slightom, J. L., A. E. Blechl, and O. Smithies.** 1980. Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21:627–638.
31. **Stringer, J. R.** 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. Nature (London) 296:363–366.
32. **Sun, L., K. E. Paulson, C. W. Schmid, L. Kadyk, and L. Leinwand.** 1984. Non-Alu family interspersed repeats in human DNA and their transcriptional activity. Nucleic Acids Res. 12:2669–2690.
33. **Swift, G. H., C. S. Craik, S. J. Stary, C. Quinto, R. G. Lahaie, W. J. Rutter, and R. J. MacDonald.** 1984. Structure of the two related elastase genes expressed in the rat pancreas. J. Biol. Chem. 259:14271–14278.
34. **Szostak, J. W., T. L. Orr-Weaver, and R. J. Rothstein.** 1983. The double-strand break repair model for recombination. Cell 33:25–35.
35. **Tautz, D., and M. Renz.** 1984. Simple DNA sequences of *Drosophila virilis* isolated by screening with RNA. J. Mol. Biol. 172:229–235.
36. **Tautz, D., and M. Renz.** 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. 12:4127–4138.
37. **Thomas, J. R., R. I. Bolla, J. S. Rumbyrt, and D. Schlessinger.** 1985. DNase I-resistant nontranscribed spacer segments of mouse ribosomal DNA contain poly(dG-dT) · poly(dA-dC). Proc. Natl. Acad. Sci. USA 82:7595–7598.
38. **Treco, D., B. Thomas, and N. Arnheim.** 1985. Recombination hot spot in the human β-globin gene cluster: meiotic recombination of human DNA fragments in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 5:2029–2037.
39. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA 74:5088–5090.