6-1993

Leptospira Genomes Are Modified at 5'-GTAC

David Ralph
California Institute of Biological Research

Quideng Que
University of Nebraska-Lincoln

James L. Van Etten
University of Nebraska-Lincoln, jvanetten1@unl.edu

Michael McClelland
California Institute of Biological Research

JOURNAL OF BACTERIOLOGY, June 1993, p. 3913–3915
0021-9193/93/123913-03$02.00/0

Vol. 175, No. 12

Leptospira Genomes Are Modified at 5′-GTAC

DAVID RALPH,[1] QUIDENG QUE,[2] JAMES L. VAN ETTEN,[2] AND MICHAEL McCLELLAND[1]*

California Institute of Biological Research, 11099 N. Torrey Pines Road, La Jolla, California 92037,[1] and Department of Plant Pathology, University of Nebraska, Lincoln, Nebraska 68583-0722[2]

Received 21 January 1993/Accepted 12 April 1993

Genomic DNAs of 14 strains from seven species of the spirochete *Leptospira* were resistant to cleavage by the restriction endonuclease *Rsa*I (5′-GTAC). A modified base comigrating with m^4C was detected by chromatography. Genomic DNAs from other spirochetes, *Borrelia* group VS461, and *Serpulina* strains were not resistant to *Rsa*I digestion. Modification at 5′-GTAm^4C may occur in most or all strains of all species of *Leptospira* but not in all genera of spirochetes. Genus-wide DNA modification has rarely been observed in bacteria.

Genomic DNAs of 14 strains from seven species in the spirochete genus *Leptospira* were cleaved with the restriction enzymes *Ava*II (5′-GGWCC), *Hin*dIII (5′-AAGCTT), and *Eco*RI (5′-GAATTC), but not with *Rsa*I (5′-GTAC) (for example, see Fig. 1). When bacteriophage λ DNA was mixed with *Leptospira* genomic DNAs and *Rsa*I, only the λ DNA was cleaved. Thus, *Leptospira* genomic DNA is modified against restriction at the sequence 5′-GTAC or at a shorter sequence that encompasses 5′-GTAC. Previous publications indicate that *Leptospira* DNAs are cleaved by *Asc*I, *Bam*HI, *Bgl*II, *Cla*I, *Eco*RI, *Eco*RV, *Hin*dIII, *Hpa*I, *Nar*I, *Not*I, *Nru*I, *Nde*I, *Pst*I, *Pvu*II, *Sac*I, *Sgr*AI, *Sma*I, *Sse*I, *Srf*I, *Stu*I, *Xba*I, *Xho*I, and *Xmn*I digestion (4, 10, 12, 18, 20, 21, 27, 30–32), indicating that most DNA sequences in *Leptospira* genomes are not modified.

Genomic DNAs of the type strain of *Borrelia* group VS461 (2, 28) and six *Serpulina* strains (T. B. Stanton, U.S. Department of Agriculture, Ames, Iowa) were susceptible to *Rsa*I cleavage (data not shown). Thus, the modification of *Rsa*I sites occurs in at least one genus but not in all genera of spirochetes.

DNA modification by bacterium-encoded enzymes has been reported at 6*N*-methyladenine (m^6A), 5*C*-methylcytosine (m^5C), and 4*N*-methylcytosine (m^4C) (29). In addition, more exotic modifications have occasionally been observed in bacteriophage DNAs (e.g., glucosylated 5-hydroxymethylcytosine in *Escherichia coli* phage T4 [13, 24]). The ability of *Rsa*I to cleave some DNA modifications has been tested (reviewed in reference 17). These data indicate that *Rsa*I cleaves DNA modified at GTAm^5C but not GTm^6AC. The effect of GTAm^4C on *Rsa*I activity is unknown. In addition, it is possible that a previously unknown modification is responsible for preventing cleavage. Of the known modifications, only 5′-GTAm^5C was excluded by the observation that the *Leptospira* genomic DNA was resistant to cleavage by *Rsa*I.

Genomic DNAs from three species of *Leptospira* were enzymatically hydrolyzed to nucleosides (11). The nucleosides were chromatographed on a reverse-phase RP-8 column (Brownlee Labs Inc., Santa Clara, Calif.) with the buffer systems of Gehrke et al. (11). Deoxynucleoside standards A, C, G, T, m^6A, and m^5C were purchased from Sigma (St. Louis, Mo.). m^4C was obtained from hydrolysates of *Bam*HI polylinker after methylation with *Bam*HI methyltransferase (New England Biolabs, Beverly, Mass.). All three strains contained no detectable m^6A or m^5C but did contain a small amount of a nucleoside that coeluted with m^4C. The levels of this base (expressed as a percentage of cytosines) were estimated to be 0.6% in *L. santarosai* shermani 1342 K (type strain), 1% in *L. borgpetersenii* ballum Mus 127, and 1.2% in *L. noguchii* louisiana LSU 1945. The error associated with these measures is at least 50%.

Statistical calculations were performed to estimate the number of *Rsa*I sites in *Leptospira* genomic DNA and thus the level of DNA modification necessary to block all *Rsa*I sites. *Rsa*I sites would occur once every 256 bp $[(1/4)^4]$ in a completely random DNA sequence with all 4 bases represented equally. However, *Leptospira* genomes are actually A+T-rich and vary in G+C content from about 35 to 40% (6). In a genome of 40% G+C, the sequence 5′-GTAC should occur about once every 278 bp $[1/(0.2)^2(0.3)^2]$ if the bases are distributed randomly. Such a frequency of *Rsa*I sites would require that 1.8% of cytosines be modified. The observed level of m^4C would therefore seem barely sufficient or insufficient to block all *Rsa*I sites. An alternative and potentially more accurate strategy to estimate the frequency of *Rsa*I sites was to observe the number of sites in sequences from various *Leptospira* genes available in the GenBank v70 data base plus more recently available sequences from Ralph et al. (23). The sample contained 19,227 bp after removal of closely homologous genes. This sample had a G+C content of 41.8%, slightly high because of the rRNA genes. *Rsa*I sites were present in this sample only 33 times, an average of once every 582 bp (Table 1), which was significantly less than the once every 278 bp predicted from the base composition of the sample. Thus, on the basis of the observed frequency of 5′-GTAC sequences in *Leptospira* DNA, modification at all *Rsa*I sites might require that only 0.86% of the cytosines (1/0.2 × 1/582) be modified. The level of modification detected by high-performance liquid chromatography could then more easily explain the resistance of the genomes to *Rsa*I. These calculations also make it unlikely that the modification involves a much shorter sequence, such as 5′-TAC, because one would then expect a higher percentage of cytosines to be modified.

However, analysis of the *Leptospira* genomic DNA sample raised another question. Why was 5′-GTAC underrepresented in the sample of *Leptospira* genomic DNAs? Statistical analysis was performed with the Markov chain (15) and improvements (8) in an attempt to understand this observation. The methods take into account the frequency of shorter oligonucleotide sequences when calculating the expected

* Corresponding author.

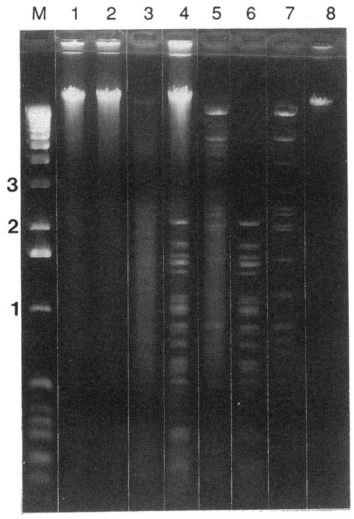

FIG. 1. Restriction digests of *L. borgpetersenii* DNA. Reactions were performed with 2 μg of genomic DNA and/or 1 μg of λ DNA and 15 U of restriction enzyme in recommended buffer for 2 h at 37°C. Genomic DNA, prepared as described by Perolat et al. (21), was a gift of Phillipe Perolat, Institute Pasteur de Nouméa, Nouméa, New Caledonia. All enzymes were purchased from Stratagene, La Jolla, Calif. λ DNA was purchased from New England Biolabs. Lanes: M, molecular weight marker (1-kb ladder purchased from GIBCO BRL, Gaithersburg, Md.); 1, undigested DNA from *L. borgpetersenii* serovar balcanica strain 1627 Burgas; 2, *Leptospira* DNA plus *Rsa*I; 3, *Leptospira* DNA plus *Xho*II; 4, *Leptospira* and λ phage DNAs plus *Rsa*I; 5, *Leptospira* and λ phage DNAs plus *Xho*II; 6, λ phage DNA plus *Rsa*I; 7, λ phage DNA plus *Xho*II; 8, undigested λ phage DNA. Numbers on the left are kilobases.

frequency of a sequence and should eliminate most of the effects on expected frequency caused by codon usage or other constraints acting at the di- or trinucleotide level. This analysis predicted 45 5′-GTAC sites in the sample, whereas only 33 such sequences were observed. Furthermore, the 23S rRNA gene is different from all of the other *Leptospira* sequences in the data base in that it has 14 5′-GTAC sites in only 3,244 bp. When this one gene is excluded, the paucity of *Rsa*I sites is even more striking (19 observed, 29 expected) (Table 1). In contrast, a sample of 16,853 bp of *Borrelia* genes (23S rRNA sequence not included) had 34 5′-GTAC sites with a predicted frequency of 32 sites. A χ^2 test indicated that *Rsa*I sites may be significantly underrepresented in the *Leptospira* genomic DNA sample ($P < 0.1$) but not in a sample of genomic DNAs from another spirochete genus. Consistent with the observation that 5′-GTAC was underrepresented in the *Leptospira* genome, genomic digests and agarose gel electrophoresis of the fragments indicated that *Kpn*I sites (5′-GGTACC) were rarer than would otherwise be expected (data not shown). We speculate that the underrepresentation of 5′-GTAC in the *Leptospira* sample is related to the fact that this sequence is modified, as has been found for modifications in certain other species (e.g., references 5 and 9 and the references therein). For unknown reasons, 5′-CTAG is also exceptionally rare in the *Leptospira* DNA sample (Table 1) and in many other bacterial genomes (8, 16) but not in the *Borrelia* DNA sample. In contrast, CATG may be exceptionally rare in *Borrelia* sequences but not rare in the *Leptospira* DNA sample (Table 1) and is not rare in most other bacterial sequences (data not shown and reference 22). These observations indicate that there may be other unknown selective pressures that are working against these sequences in certain species.

Genomic DNA modifications that protect against restriction enzymes have been found several times in eubacterial genomes. So far, such sequence-specific modification systems are usually present in some, but not all, strains in any species. Such DNA modifications are usually part of sequence-specific restriction modification systems (e.g., see references 7 and 29). The possibility of a restriction endonuclease with a 5′-GTAC specificity present in all *Leptospira* genomes examined has not been excluded, although we failed to detect a 5′-GTAC-specific system (data not shown). If such a restriction system exists in all strains of all species in a genus, this distribution would be unprecedented.

DNA modifications not associated with a restriction endonuclease have been observed in bacteria, including the adenine-specific methylation (at 5′-GATC or a subset of such sites) which occurs in 3 of 22 strains of the spirochete *Borrelia burgdorferi* and in the 1 strain examined from five of six other species of *Borrelia* (19). However, such sequence-specific modification systems are similar to restriction modification in that they are usually present in some, but not all, strains in any species.

TABLE 1. Observed and expected occurrences for some tetranucleotides in a sample of *Leptospira* and *Borrelia* genes[a]

Tetranucleotide	Occurrence of tetranucleotide in:					
	Leptospira genome			*Borrelia* genome		
	Observed	Expected[b]	Observed/expected	Observed	Expected[b]	Observed/expected
GATC	81	96	0.85	50	48	1.04
TCGA	66	71	0.94	13	12	1.05
TGCA	55	50	1.11	92	80	1.15
ACGT	44	45	0.97	14	16	0.88
AGCT	40	39	1.01	132	135	0.98
CATG	33	34	0.98	21	52	0.63[c]
GTAC	19	29	0.66[c]	34	32	1.06
CTAG	18	27	0.68	49	51	0.96

[a] A total of 16,140 bases of *Leptospira* gene sequences and 16,853 bases of *Borrelia* gene sequences available in the GenBank v70 data base (excluding the 23S rRNA genes) were examined with MacVector software (IBI-Kodak, New Haven, Conn.) run on a Quadra 700.

[b] Predicted frequencies for tetranucleotide XYZW are calculated from the mono-, di-, tri-, and dinucleotide frequencies by the following equation (8):

$$\frac{[XYZW] \times [XY] \times [XNZ] \times [XNNW] \times [YZ] \times [YNW] \times [ZW]}{[XYZ] \times [XYNW] \times [XNZW] \times [YZW] \times [X] \times [Y] \times [Z] \times [W]}$$

[c] Significant at $P < 0.1$ (χ^2 test).

One exception to the rule that sequence-specific modification systems are present in some, but not all, strains in any species is the *dam* methylation system (5′-Gm^{6}ATC), which occurs in all strains of several species in the family *Enterobacteriaceae* and related genera (3). Rather than being part of a restriction modification system, *dam* methylation is part of a daughter strand repair system (1, 14) and also regulates replication (25, 26). It is presumably because of a role in maintaining all cells that the *dam* modification is found throughout those species in which it occurs. In contrast, a restriction system in all cells of a species would be obsolete when the first resistant bacteriophage carrying the appropriate modification were propagated, which is presumably why strains within species vary in their restriction systems. The distribution of the DNA modification at 5′-GTAC in *Leptospira* genomes is more reminiscent of *dam* methylation than restriction modification. Although this distribution of modification at 5′-GTAC does not necessarily imply a role in functions identical to those of *dam*, it implies a similar species-wide role in fitness.

In summary, *Leptospira* genomes carry a sequence-specific DNA modification at 5′-GTAC that is probably m^{4}C specific. The distribution of this modification is not typical of a conventional restriction-modification system.

We particularly thank Guy Baranton, Isabelle Saint Giron, and Philippe Perolat for supplying strains and for their enthusiastic collaboration, without which this work would not have been possible. We thank Mike Nelson, Alan Greener, Bruno Sobral, Rhonda Honeycutt, Guy Baranton, Phillipe Perolat, and Cibra Stil Sinquing for critical reading of the manuscript. We thank John Welsh for many helpful discussions.

This work was supported in part by NIH grant R29AI32644 to John Welsh. D.R. was supported by set-up funds from CIBR. M.M. was supported in part by NIH grants 2RO1AI34829 and 1RO1HG00456.

REFERENCES

1. **Au, K. G., K. Welsh, and P. Modrich.** 1992. Initiation of methyl-directed mismatch repair. J. Biol. Chem. **267:**12142–12148.
2. **Baranton, G., D. Postic, I. Saint Girons, P. Boerlin, J.-C. Piffaretti, M. Assous, and P. A. D. Grimont.** 1992. Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS461 associated with Lyme borreliosis. Int. J. Syst. Bacteriol. **42:**378–383.
3. **Barbeyron, T., K. Kean, and P. Forterre.** 1984. DNA adenine methylation at GATC sequences appeared recently in *Escherichia coli* lineage. J. Bacteriol. **160:**586–590.
4. **Baril, C., J. L. Herrmann, C. Richaud, D. Margarita, and I. Saint Girons.** 1992. Scattering of the rRNA genes on the physical map of the circular chromosome of *Leptospira interrogans* serovar icterohaemorrhagiae. J. Bacteriol. **174:**7566–7571.
5. **Bhagwat, A. S., and M. McClelland.** 1992. DNA mismatch correction by Very Short Patch repair may have altered the abundance of oligonucleotides in the *E. coli* genome. Nucleic Acids Res. **20:**1663–1668.
6. **Brendle, J. J., M. Rogul, and A. D. Alexander.** 1974. Deoxyribonucleic acid hybridization among selected leptospiral serotypes. Int. J. Syst. Bacteriol. **24:**205–214.
7. **Brooks, J. E., and R. J. Roberts.** 1982. Modification profiles of bacterial genomes. Nucleic Acids Res. **10:**913–934.
8. **Burge, C., A. M. Campbell, and S. Karlin.** 1992. Over- and under-representation of short oligonucleotides in DNA sequences. Proc. Natl. Acad. Sci. USA **89:**1358–1362.
9. **Cooper, D. N., and M. Krawczak.** 1989. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. Hum. Genet. **83:**181–188.
10. **Fukunaga, M., and I. Mifuchi.** 1989. Unique organization of *Leptospira interrogans* rRNA genes. J. Bacteriol. **171:**5763–5767.
11. **Gehrke, C. W., R. A. McCune, M. A. Gama-Sosa, M. Ehrlich, and K. C. Kuo.** 1984. Quantitative reverse-phase high-performance liquid chromatography of major and modified nucleosides in DNA. J. Chromatogr. **301:**199–219.
12. **Herrmann, J. L., E. Bellenger, P. Perolat, G. Baranton, and I. Saint Girons.** 1992. Pulsed-field gel electrophoresis of *Not*I digests of leptospiral DNA: a new rapid method of serovar identification. J. Clin. Microbiol. **30:**1696–1702.
13. **Huang, L.-H., C. M. Farnet, K. C. Ehrlich, and M. Ehrlich.** 1982. Digestion of highly modified bacteriophage DNA by restriction endonucleases. Nucleic Acids Res. **10:**1579–1591.
14. **Marinus, M. G.** 1987. DNA methylation in *E. coli*. Annu. Rev. Genet. **21:**113–131.
15. **McClelland, M.** 1985. Selection against *dam* methylation sites in the genomes of DNA enterobacteriophage. J. Mol. Evol. **21:** 317–322.
16. **McClelland, M., R. Jones, Y. Patel, and M. Nelson.** 1987. Restriction endonucleases for pulsed field mapping of bacterial genomes. Nucleic Acids Res. **15:**5985–6005.
17. **McClelland, M., and M. Nelson.** 1992. Effect of site-specific methylation on DNA modification methyltransferases and restriction endonucleases. Nucleic Acids Res. **20:**2145–2157.
18. **Nielsen, J. N., C. H. Armstrong, and N. C. Nielsen.** 1989. Relationship among selected *Leptospira interrogans* serogroups as determined by nucleic acid hybridization. J. Clin. Microbiol. **27:**2724–2729.
19. **Norton Hughes, C. A., and R. C. Johnson.** 1990. Methylated DNA in *Borrelia* species. J. Bacteriol. **172:**6602–6604.
20. **Pacciarini, M. L., M. L. Savio, S. Tagliabue, and C. Rossi.** 1992. Repetitive sequences cloned from *Leptospira interrogans* serovar hardjo genotype hardjoprajitno and their application to serovar identification. J. Clin. Microbiol. **30:**1243–1249.
21. **Perolat, P., F. Grimont, B. Regnault, P. A. D. Grimont, E. Fournie, H. Thevenet, and G. Baranton.** 1990. rRNA gene restriction patterns of *Leptospira*: a molecular typing system. Res. Microbiol. **141:**159–171.
22. **Phillips, G. J., J. Arnold, and R. Ivarie.** 1987. Mono- through hexanucleotide composition of the *Escherichia coli* genome: a Markov chain analysis. Nucleic Acids Res. **15:**2611–2638.
23. **Ralph, D., M. McClelland, J. Welsh, G. Baranton, and P. Perolat.** 1993. *Leptospira* species categorized by arbitrarily primed polymerase chain reaction (PCR) and by mapped restriction polymorphisms in PCR-amplified rRNA genes. J. Bacteriol. **175:**973–981.
24. **Revel, H. R.** 1983. DNA modification: glucosylation, p. 156–165. *In* C. K. Mathews, E. M. Kutter, G. Mosig, and P. Berget (ed.), Bacteriophage T4. American Society for Microbiology, Washington, D.C.
25. **Russel, D. W., and K. Horiuchi.** 1991. The *mutH* gene regulates the replication and methylation of the pMB1 origin. J. Bacteriol. **173:**3209–3214.
26. **Russel, D. W., and N. D. Zinder.** 1987. Hemimethylation prevents DNA replication in *E. coli*. Cell **50:**1071–1079.
27. **Segers, R. P. A. M., J. A. van Gestel, G. J. J. M. van Eys, B. A. M. van der Zeijst, and W. Gaastra.** 1992. Presence of putative sphingomyelinase genes among members of the family *Leptospiraceae*. Infect. Immun. **60:**1707–1710.
28. **Welsh, J., C. Pretzman, D. Postic, I. Saint Girons, G. Baranton, and M. McClelland.** 1992. Genomic fingerprinting by arbitrarily primed PCR resolves *Borrelia burgdorferi* into three distinct phyletic groups. Int. J. Syst. Bacteriol. **42:**370–377.
29. **Wilson, G. G., and N. E. Murray.** 1992. Restriction modification systems, p. 585–628. *In* A. Cambell (ed.), Annual reviews of genetics. Academic Press, Palo Alto, Calif.
30. **Woodward, M. J., and G. J. Sullivan.** 1991. Nucleotide sequence of a repetitive element isolated from *Leptospira interrogans* serovar hardjo type hardjo-bovis. J. Gen. Microbiol. **137:**1101–1109.
31. **Zuerner, R. L.** 1991. Physical map of chromosomal and plasmid DNA comprising the genome of *Leptospira interrogans*. Nucleic Acids Res. **19:**4857–4860.
32. **Zuerner, R. L., and C. A. Bolin.** 1991. Nucleic acid probe characterizes *Leptospira interrogans* serovars by restriction fragment length polymorphisms. Vet. Microbiol. **24:**355–366.