Volume 17 Number 4 1989 Nucleic Acids Research

Remarkably AT-rich genomic DNA from the anaerobic fungus *Neocallimastix*

Alan G.Brownlee

CSIRO, Division of Animal Production, PO Box 239, Blacktown, NSW 2148, Australia

Received November 22, 1988; Revised and Accepted January 23, 1989

ABSTRACT

The genomic DNA of an anaerobic rumen phycomycete of the genus *Neocallimastix* has been purified and characterized. The non-repetitive fraction of the DNA has a G.C content of only 13%. The ribosomal RNA genes are highly reiterated, making up about 30% of the total DNA, and are evident as a more G.C-rich satellite with a repeating unit of about 9.4 kilobases (Kb). A.T-rich regions of DNA are highly dispersed and possess some sequence complexity. Chemical analysis of the DNA constituents reveals no evidence of modified bases. The genome of this anaerobic fungus has the highest A.T content of any organism so far described.

INTRODUCTION

Fungi of the genus *Neocallimastix* are common in the digestive tract of herbivorous animals fed on a fibrous diet. These fungi have a simple life-cycle alternating between a motile multi-flagellate zoospore and an extensive vegetative stage usually found associated with particulate plant material in the alimentary tract(1). They have been assigned to a new family, the Neocallimasticaceae, in the order Spizellomycetales of the Chytridiomycetes(2) but their exact taxonomic affinities are uncertain(3,4). Members of this genus have several unique features, in addition to the production of multi-flagellate zoospores, that distinguish them clearly from non-rumen chytrids and fungi in general. *Neocallimastix* lacks observable mitochondria and described isolates are obligate anaerobes. Anaerobic fermentation products include ethanol and H_2, production of the latter being associated with particular organelles termed hydrogenosomes(5). *N. frontalis* appears to possess an anaerobic pathway for monoenoic acid biosynthesis, in marked contrast to facultatively anaerobic fungi(6).

These fungi are efficient at degrading plant derived cellulose and hemicellulose and can be significant contributors to fibre digestion in the rumen(7). Various isolates of *Neocallimastix* spp. are under study at our laboratory(8). As a preliminary to assessing DNA relatedness amongst these isolates and their possible potential for improvement by genetic manipulation, DNA from a sheep rumen isolate of *Neocallimastix* sp. LM-2(8) was extracted and characterized. Some of the unusual properties of this DNA are reported here. *Neocallimastix* has the most A.T-rich genome of any organism, prokaryote or eukaryote, described to date.

MATERIALS AND METHODS

Fungal and Bacterial Strains and Growth Conditions

Neocallimastix sp. LM-2 was isolated from sheep ruminal digesta sampled at Prospect (New South Wales, Australia)(8). Larger quantities of fungal biomass were grown in 1

litre screw-capped bottles (Schott Wiesbaden, Fed. Rep. Germany) containing 500 ml of anaerobic medium 10 with 0.5% glucose as carbon source(8). Inocula were grown in gas-tight Hungate tubes containing 5 ml of the same medium supplemented with 0.05% agar. Some fungal cultures were grown in 50 ml of medium 10 in 100 ml sealed bottles. All cultures were static at 39°C in the dark for 4–6 days.

Escherichia coli strain JM103 was grown for M13 phage propagation as described(9).

DNA Analyses

Extraction Total LM-2 DNA was extracted by a gentle extraction method(10) through to the ethanol precipitation step. Copurifying flocculent polysaccharide was partially removed by centrifuging the redissolved sample at 100,000 ×g for 60 min at 4°C. The DNA was then purified by preparative CsCl density gradient centrifugation in the presence of Hoechst 33258(11). Mainband DNA and a prominent, more dense satellite band were removed from the gradient together, or in some preparations separately, by side puncture of the centrifuge tubes. Hoechst 33258 was removed by extraction with water-saturated n-butanol, the sample was diluted with water and the DNA precipitated with ethanol. DNA was redissolved in, and dialysed extensively against, 0.1×SSC (SSC is 0.15 M NaCl, 0.015 M sodium citrate, pH 7.0). More recent DNA preparations, on a smaller scale, employed a miniprep method devised for these polysaccharide-rich fungi(12).

Determination of T_m Thermal melting of LM-2 DNA was performed in 0.1×SSC with a Gilford Model 2600 spectrophotometer coupled to a Model 2527 Thermoprogrammer and a Hewlett Packard 7225A Plotter. T_m values were determined from first derivative plots of the thermal melting profile calculated automatically with a microprocessor at different levels of sensitivity (sampling frequency). Total genomic DNA as well as mainband DNA separated from the satellite band were examined individually.

Base Composition Analysis by HPLC Residual RNA in DNA samples was removed by alkaline hydrolysis(13). DNA was then ethanol precipitated and hydrolysed enzymatically to its constituent nucleosides(14). Commercially obtained phage Lambda DNA was treated identically and used as a reference. Best separating conditions for High Pressure Liquid Chromatography were found to be isocratic elution in 0.1% phosphoric acid from a Waters C_{18} μ Bondapak column (30 cm×3.9 mm)(15). Samples were analysed with a Waters Liquid Chromatograph with U6K injector and Model 481 variable wavelength monitor. Flow rate was 2.5 ml.min^{-1} and the eluate was monitored at 273 nm. Standard solutions of deoxyribonucleosides were used to measure retention times and to calibrate the column. Nucleoside concentrations were proportional to peak height within the range used. 5-Methyldeoxycytidine had a retention time of 3.2 min in this analysis and was clearly separated from the dC and dA peaks (see also 15). G.C content of lambda DNA was determined to be 48.7 mol % (49.8 mol % is the actual content). LM-2 DNA G.C content was determined relative to purified nucleosides as well as to lambda DNA as the standard; both results were fully consistent.

Restriction Digests, End-Labelling, Electrophoresis and Gel Transfer

Restriction digests of purified LM-2 DNA fractions and end-labelling by filling in overhanging 5′ termini with DNA polymerase I (Klenow fragment) and [^{32}P]-labelled deoxynucleotides were performed as described(16). Digests were electrophoresed in 0.8% or 1% agarose gels in Tris-acetate or Tris-borate buffers(16). End-labelled DNA fragments were detected by autoradiography after alkaline transfer from gels to Gene Screen Plus (Du Pont (Australia) Pty. Ltd.)(17). For analysis by hybridization restriction digests were transferred from gels to Gene Screen Plus as above. A probe was prepared from the plasmid

pKD-002, containing a portion of the *Neurospora crassa* ribosomal repeat(18), labelled with [^{32}P]dCTP(19). Hybridization was performed overnight at 58°C in 7% SDS, 10 mM EDTA, 0.5 M NaH_2PO_4 (pH 7.5). The membrane was washed in 1% SDS, 4×SSPE at 55°C (1×SSPE is 0.18 M NaCl, 10 mM sodium phosphate (pH 7.0); 10 mM EDTA) with a final wash in 2×SSPE, 0.5% SDS at 55°C. Autoradiography was at −70°C with one intensifying screen.

Cloning and Sequencing LM-2 DNA was digested with *Eco*RI and cloned into M13 mp10(9) also digested with *Eco*RI. Eight clear plaques were picked at random from plates containing X-Gal and IPTG. All were screened in a sequencing reaction with dideoxyT (20). No two were identical and all were assumed to be representative clones of LM-2 sequences. Three of these were randomly chosen for sequencing(20).

Other Analyses The buoyant density of LM-2 DNA was determined in CsCl at equilibrium in a Model E Ultracentrifuge with lambda DNA as the reference. A spectral analysis of the DNA was performed using a Varian DMS-100 UV Spectrophotometer (Varian Techtron Pty Ltd, Australia). Absorbance measurements were taken at 5 nm wavelength increments between 230 and 290 nm.

RESULTS

Preparative CsCl density gradient centrifugation of LM-2 DNA in the presence of the fluorescent dye Hoechst 33258 consistently revealed a prominent satellite DNA banding at greater density (below) the mainband fraction. Separation of fungal ribosomal genes as a repeated satellite in this manner has been observed by others(11). First indications of the high A.T content of LM-2 DNA came from attempts to measure DNA concentrations fluorimetrically using 4′6-diamidino-2-phenylindole (DAPI) or Hoechst 33258(21). These results consistently indicated, using calf thymus DNA as a standard, that the concentrations of DNA solutions were 3 to 4-fold higher than those determined from the UV absorbance at 260 nm. These dyes are known to bind preferentially to A.T containing regions of DNA(21,22). A three-term spectral analysis of native DNA(23) in TE (10 mM Tris (Cl-, pH 8.0); 1 mM EDTA) yielded the following results: $A_{260/A280} = 2.00$, $A_{260/A230} = 3.66$ $\mu_1 = 441.086$, $\mu_2 = -7450.025$, $\mu_3 = 25318.198$ hence ϕ (fractional A.T content) = 0.83.

The DNA was further characterized by thermal melting and by HPLC analysis of enzymatically hydrolysed samples (Fig. 1). The thermal melting profile (Fig. 1a) clearly indicates two major components with T_m's (0.1×SSC) of 56.5°C and 64.6°C indicating a G.C content of 13.6 and 29.7 mol% respectively (24). These components are present in the approximate ratio 2:1 based on the areas under the first derivative plots (Fig. 1a). Separated mainband DNA from a CsCl density gradient had a T_m (0.1×SSC) of 55.0°C or a G.C content of 12.3%(24) (results not shown).

Chromatographic resolution of the constituent nucleosides after enzymatic hydrolysis was also used to estimate the G.C content (Fig. 1b). Nucleoside peaks were clearly resolved and no appreciable quantity of modified nucleosides was detected such as might produce anomalous melting behaviour or ultraviolet spectra. In common with other fungi(25) LM-2 does not contain 5-methylcytosine in detectable amounts (Fig. 1b). The G.C content estimated from the peak heights of eluted nucleosides is 18.0 mol%.

The buoyant density of LM-2 DNA in CsCl was determined to be 1.686 g.cm^{-3} in an analytical ultracentrifuge with lambda DNA as the reference (results not presented).

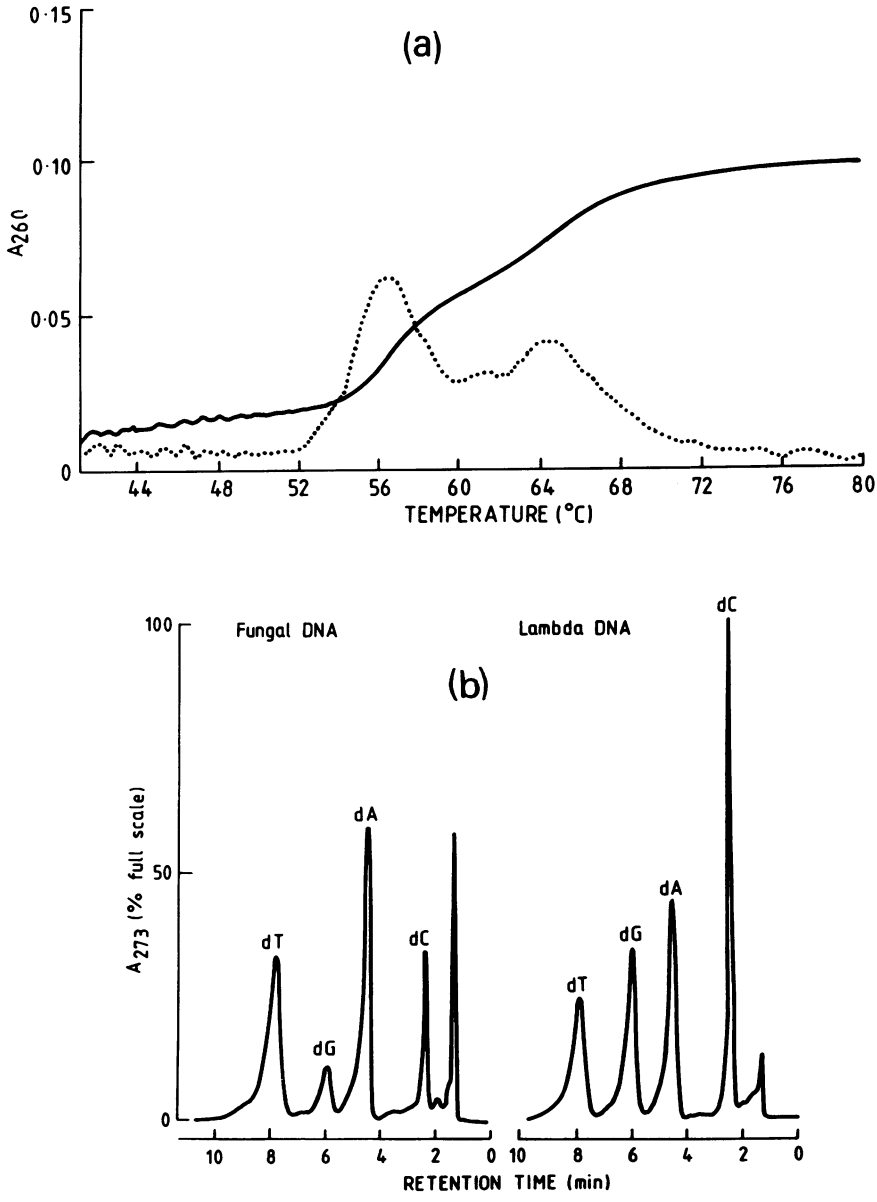

Fig. 1 Thermal melting and HPLC analysis of LM-2 DNA. (a). Thermal melting profile (unbroken line) and first derivative plot of total LM-2 DNA in 0.1×SSC determined at a temperature gradient of 1°C. min^{-1}. (b). Separation of hydrolysed nucleosides of LM-2 DNA by HPLC as described in Materials and Methods. Elution of reference (lambda) DNA is shown for comparison.

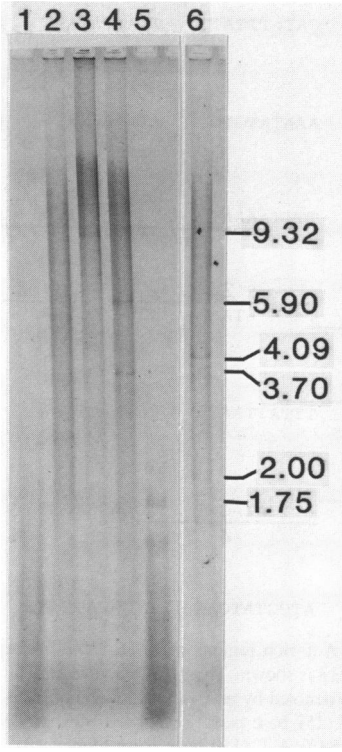

Fig. 2 Distribution of A.T-rich restriction sites in LM-2 DNA. 2−3 μg of one DNA preparation was digested overnight with the following restriction enzymes; a negative image of the ethidium bromide stained gel is presented. Lane 1, *Aha*III (TTTAAA); 2, *Asu*II (TTCGAA); 3, *Bam*HI (GGATCC); 4, *Xba*I (TCTAGA); 5, *Ssp*I (AATATT)6, *Eco*RI (GAATTC). Sizes (kb) of the visible satellite bands are marked alongside.

This, however, was the median density of a heterogenous peak with considerable spread at equilibrium.

The restriction enzymes *Aha*III and *Ssp*I recognise the 6 base sequences TTTAAA and AATATT respectively. These sequences are present at high frequency in LM-2 genomic DNA (Fig. 2), in fact the bulk of the DNA is digested to less than 1.75 Kb by *Ssp*I (Fig. 2, lane 5). *Bam*HI (GGATCC)−33% AT content in its recognition sequence−cuts noticeably less frequently than *Eco*RI (GAATTC), *Asu*II (TTCGAA) or *Xba*I (TCTAGA), all 66% AT (Fig. 2). Together with the symmetrical shape of the melting curve (Fig. 1a) this indicates that the AT-rich sequences are dispersed throughout the genome. Direct evidence for this comes from actual sequence data (Fig. 3); randomly selected clones of LM-2 DNA in M13 reveal regions of AT-rich direct repeats and dyad symmetry.

The satellite DNA evident in preparative CsCl density gradients and in Fig. 1a is composed of a highly conserved sequence about 9.4 Kb in length (Fig. 4). The prominent restriction fragments evident in genomic DNA digests (Fig. 2) are common to the satellite and hybridize at moderate stringency to the recombinant plasmid pKD.002 (18) containing

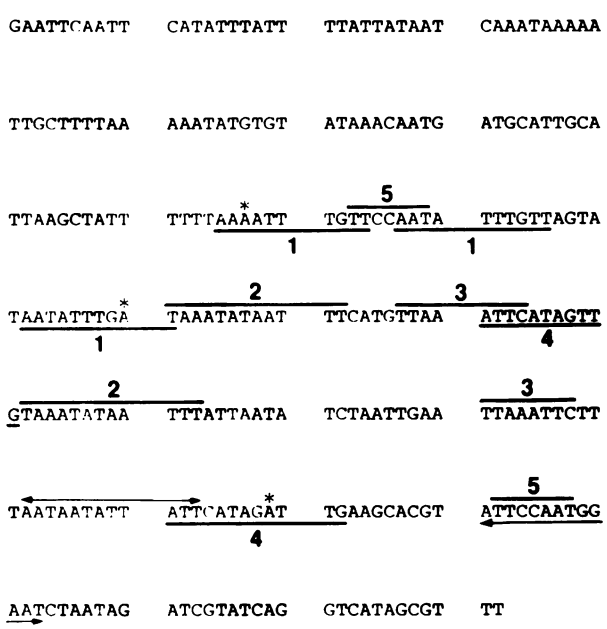

Fig. 3 Sequence complexity within A.T-rich regions of LM-2 DNA. A representative sequence of part of an *Eco*RI fragment of LM-2 DNA in M13 is shown. The A.T content is 80.9%; direct repeats are shown underlined or overlined (single base differences denoted by asterisks). Regions of dyad symmetry are marked with arrows. A second of the 3 clones sequenced (257 base pairs) comprised 80% A.T with a similar degree of complexity to that shown here. The third was 58% A.T (150 base pairs sequenced).

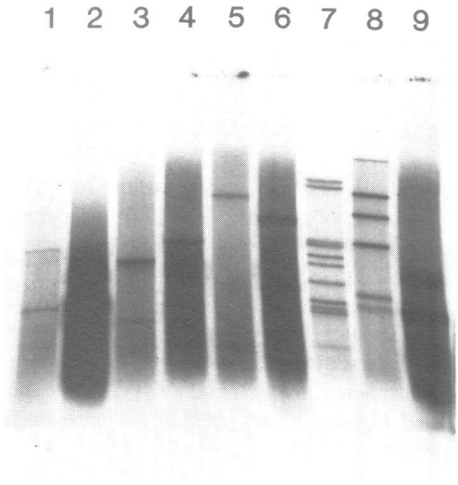

Fig. 4 End labelled restriction digests of satellite DNA. About 0.5 μg of LM-2 satellite DNA was digested with the indicated restriction enzymes, end labelled with [^{32}P] and separated on an agarose gel as described in Materials and Methods. Lane 1, *Eco*RI; 2, *Eco*RI and *Cla*I; 3, *Xba*I and *Cla*I; 4, *Cla*I; 5, *Bss*HII; 6, *Xba*I and *Bss*HII; 7, Lambda *Cla*I digest; 8, Lambda *Hind*III; 9, *Cla*I and *Bss*HII.

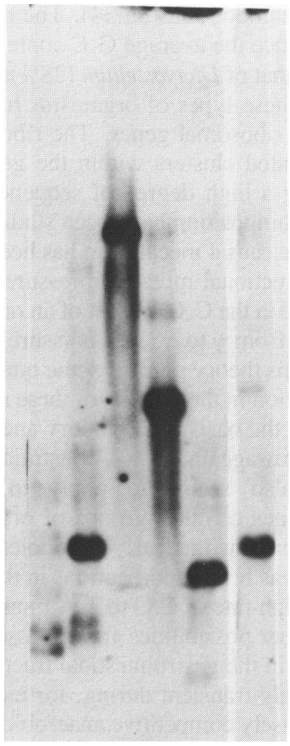

Fig. 5 Hybridization of LM-2 satellite DNA restriction fragments to *N. crassa* ribosomal gene clone pKD.002. DNA from the gel shown in Fig. 2 was blotted onto a nylon membrane and hybridized with pKD.002 as indicated in Materials and Methods.

the *Neurospora crassa* 17S, 5.8S and part of the 26S ribosomal RNA genes (Fig. 5). The satellite DNA therefore represents the rRNA genes present as a tandem repeat with a high degree of reiteration.

DISCUSSION

The remarkably high A.T content of LM-2 DNA puts it far outside the range of reported G.C values of other fungi(26). Of the related taxa within the class Chytridiomycetes(4) the G.C composition of a few species is known. Thus *Allomyces arbuscula* and *Blastocladiella emersonii* (order Blastocladiales) have G.C contents of 62.2%(26) and 66%(28) respectively, *Rhizophydium* sp. (order Chytridiales), 50.5% G.C(28) and *Rhizophlyctis (Karlingia) rosea* (order Spizellomycetales), 44% G.C(28). *Neocallimastix* is clearly distinct from all of these although related orders of lower and higher fungi can exhibit a wide range of G.C content(27).

Species of the malaria parasite *Plasmodium* have among the lowest recorded G.C contents of the protists: 18–20%(29,30) and other protozoa, notably certain holotrichs, have been documented with low G.C contents(31). The slime mould *Dictyostelium discoideum* (23% G.C)(32) has a distinctive organization of A.T-rich genomic regions(33). *Neocallimastix* sp. LM-2 presumably has a similar pattern of extremely A.T-rich intergenic regions flanking

coding regions with an extreme codon bias(33,34). The ribosomal repeat would seem to follow the same arrangement since the average G.C content of this portion of the genome (29.7%) is of the same order as that of *Dictyostelium* (28%)(35) and the holotrich protozoans (30−38%)(36) although both these types of organisms have extrachromosomal, usually linear, multiple copies of their ribosomal genes. The ribosomal genes of *Neocallimastix* are arranged as tandemly repeated clusters within the genome with a repeating unit of about 9.4−10.0 Kb containing a high degree of sequence homogeneity.

Organisms of extreme G.C composition have been studied for some time but no wholly satisfactory explanation as to the causal mechanism has been advanced. Sueoka introduced the notion of nonrandom or directional mutation pressure toward either A.T (α) or G.C (γ) pairs(37,38). Hence a change in the G.C content of an organism (Sueoka only considered bacteria) is caused by mutation from γ to α (A.T pressure) or from α to γ (G.C pressure) at rates u and v respectively. This theory permits some quantitative analysis of the relative effects of genetic drift and selection in the fixation of these mutations in the genome(39,40). Several studies have confirmed the basis of the theory and the resultant consequences for amino acid composition, codon usage and tRNA constraints in bacteria and mitochondria (37,38,40,41). The theory also seems to apply to the observed compositional compartmentation of the genomes of other groups of organisms(42). The nature of the mutation pressure, however, remains unclear. Ultraviolet radiation or natural alkylating agents have been suggested(43) as have modifications in the DNA replicating system(44). There is strong evidence for a high rate of C.G to T.A transitions induced by demethylation of 5-methylcytosine(45,46). Other possibilities are discussed by Elton(47). *Neocallimastix* has only been detected thus far in the gastrointestinal tract of herbivores. Exposure to the atmosphere is believed to be only transient during, for example, inoculation of newborn animals. This confined yet intensely competitive anaerobic environment has possibly been an important factor guiding the compositional changes that have caused *Neocallimastix* DNA to diverge so markedly.

Whatever the explanation for the extreme A.T content of *Neocallimastix* the organism clearly qualifies as a 'genetic code limit organism'(48) and, if its taxonomic placement withstands scrutiny, it should prove extremely informative to compare features such as codon usage and gene regulation with the related orders of aquatic fungi.

ACKNOWLEDGEMENTS

I gratefully acknowledge the assistance of Dr Peter Stuart and Ms Barbara Inglis with the thermal melting studies, Dr Greg Ralston for the use of Model E analytical ultracentrifuge facilities, Mr Bruce Wilson for aid in sequencing, Mr Leo Sharry for access to HPLC equipment and Dr Brett Tyler for providing the recombinant plasmid pKD.002.

REFERENCES

1. Bauchop T. (1981) Agric. Environm. *6*, 339−348.
2. Heath, I.B., Bauchop, T. and Skipp, R.A. (1983) Can. J. Bot. *61*, 295−307.
3. Heath, I.B. and Bauchop, T. (1985) Can. J. Bot. *63*, 1595−1604.
4. Barr, D.J.S. (1988) Bio Systems *21*, 351−356.
5. Yarlett, N., Orpin, C.G., Munn, E.A., Yarlett, N.C. and Greenwood, C.A. (1986) Biochem. J. *236*, 729−739.
6. Body, D.R. and Bauchop, T. (1985) Can. J. Microbiol. *31*, 463−466.
7. Akin, D.E., Gordon, G.L.R. and Hogan, J.P. (1983) Appl. Env. Microbiol. *46*, 738−748.
8. Phillips, M.W. and Gordon, G.L.R. (1988) BioSystems *21*, 377−383.
9. Messing, J. (1983) Meth. Enzymol. *101*, 20−78.

10. Specht, C.A., Di Russo, C.C., Novotny, C.P. and Ullrich, R.C. (1982) Anal. Biochem. *119*, 158–163.
11. Raeder, U. and Broda, P. (1985) Letts. Appl. Microbiol. *1*, 17–20.
12. Brownlee, A.G. (1988) Fungal Genetics Newsletter *35*, 8–9.
13. Eick, D., Fritz, H.-J. and Doerfler, W. (1983) Anal. Biochem. *135*, 165–171.
14. Gehrke, C.W., McCune, R.A., Gama-Sosa, M.A., Ehrlich, M. and Kuo, K.C. (1984) J. Chromatog. *301*, 199–219.
15. Wagner, I. and Capesius, I. (1981) Biochim. Biophys. Acta *654*, 52–56.
16. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual (Cold Spring Harbour Laboratory, New York).
17. Reed, K.C. and Mann, D.A. (1985) Nucl. Acids Res. *13*, 7207–7221.
18. Russell, P.J., Wagner, S., Rodland, K.D., Feinbaum, R.L., Russel, J.P., Bret-Harte, M.S., Free, S.J. and Metzenburg, R.L. (1984) Mol. Gen. Genet. *196*, 275–282.
19. Feinberg, A.P. and Vogelstein, B. (1984) Anal. Biochem. *137*, 266–267.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natn. Acad. Sci. U.S.A. *74*, 5463–5467.
21. Paul, J.H. and Myers, B. (1982) Appl. Env. Microbiol. *43*, 1393–1399.
22. Müller, W. and Gautier, F. (1975) Eur. J. Biochem. *54*, 385–394.
23. Felsenfeld, G. (1968) Meth. Enzymol. *12*, 247–253.
24. Mandel, M., Igambi, L., Bergendahl, J., Dodson, M.L. and Scheltgen, E. (1970) J. Bacteriol. *101*, 333–338.
25. Antequera, F., Tamame, M., Villanueva, J.R. & Santos, T.J. (1984) Biol. Chem. *259*, 8033–8036.
26. Kurtzman, C.P. (1984) Microbiol. Sci. *1*, 44–48.
27. Kurtzman, C.P. (1985) in Gene Manipulation in Fungi eds. J.W. Bennett and L.L. Lasure (Academic Press) pp. 35–63.
28. C.R.C. Handbook of Biochemistry and Molecular Biology (3rd. Edn.) Nucleic Acids Vol III. G.D. Fasman ed. (1976) C.R.C. Press, Cleveland Ohio.
29. Pollack, Y., Katzen, A.L., Spira, D.T. and Golenser, J. (1982) Nucl. Acids Res. *10*, 539–546.
30. McCutchan, T.F., Dame, J.B., Miller, L.H. and Barnwell, J. (1984) Science. *225*, 808–811.
31. Schildkraut, C.L., Mandel, M., Levisohn, S., Smith-Sonneborn, J.E. and Marmur, J. (1962) Nature *196*, 795–796.
32. Sussman, R. and Rayner, E.P. (1971) Arch. Biochem. Biophys. *144*, 127–137.
33. Kimmel, A. and Firtel, R.A. (1983) Nucl. Acids. Res. *11*, 541–552.
34. Weber, J.L. (1987) Gene *52*, 103–109.
35. Firtel, R.A. and Bonner, J. (1972) J. Mol. Biol. *66*, 339–361.
36. Blackburn, E.H. (1982) in The Cell Nucleus, Vol. X. Ribosomal DNA, part A. eds. H. Busch and L. Rothblum (Academic press) pp. 145–170.
37. Sueoka, N. (1962) Proc. Natl. Acad. Sci. USA *48*, 582–592.
38. Jukes, T.H. and Bhushan (1986) J. Mol. Evol. *24*, 39–44.
39. Sueoka, N. (1988) Proc. Natl. Acad. Sci. USA *85*, 2653–2657.
40. Muto, A. and Osawa, S. (1987) Proc. Natl. Acad. Sci. USA *84*, 166–169.
41. Osawa, S., Ohama, T., Yamao, F., Muto, A., Jukes, T.H., Ozeki, H. and Umesono, K. (1988) Proc. Natl. Acad. Sci. USA *85*, 1124–1128.
42. Bernardi, G. and Bernardi, G. (1985) J. Mol. Evol. *22*, 363–365.
43. Singer, C.E. and Ames, B.N. (1970) Science *170*, 822–826.
44. Speyer, J.F. (1976) Biochem. Biophys. Res. Commun. *21*, 6–10.
45. Selker, E.U. and Stevens, J.N. (1985) Proc. Natl. Acad. Sci. USA *82*, 8114–8118.
46. Bird, A.P. (1986) Nature *321*, 209–213.
47. Elton, R.A. (1973) J. Molec. Evol. *2*, 263–276.
48. Woese, C.R. and Bleyman, M.A. (1972) J. Molec. Evol. *1*, 223–229.