

Fig. 3 CAT enzyme activity from lysates of transfected BALB/c 3T3 cells. *a*, Autoradiograph of CAT enzyme assays from cells transfected with plasmids as follows: lane 1, 15 µg of pPHSH3-cat with no addition; lane 2, 15 µg of pPHSH3-cat with 1 µg of pSV2-cmcy; lane 3, 15 µg of pDHSP-cat with no addition; lane 4, 15 µg of pDHSP-cat with 1 µg of pSV2-cmcy; lane 5, pPHSH3-cat with 1 µg of pfsmyc-20. The autoradiograph is from the 60-min time point of the enzyme assays. Ac-Cm, acetylated ¹⁴C-chloramphenicol; Cm, ¹⁴C-chloramphenicol. *b*, Results of enzyme assays of cells transfected with 15 µg of pPHSH3-cat alone (no addition), or with 1 µg of either pSV2-cmcy or pfsmyc-20. Results are plotted as per cent of the total ¹⁴C-chloramphenicol acetylated at each time point as indicated. *c*, Results of enzyme assays of cells transfected with 15 µg of pDHSP-cat alone (no addition) or with 1 µg of pSV2-cmcy. These results were generated with lysates from the same transfection experiment. Other identical experiments give 8-, 15-, 6-, 10- and 7-fold stimulation of CAT expression from pPHSH3-cat when co-transfected with pSV2-cmcy.

Methods: Transfections were performed using the CaPO₄ coprecipitation procedure²⁹. Cells were washed with Dulbecco's minimal essential medium (DMEM) and refed DMEM with 10% calf serum 16 h after addition of precipitates. Lysates were prepared ~48 h later by the procedure of Gorman *et al.*¹². 200 µg of protein were assayed from each time point shown. The percentage of acetylated chloramphenicol was determined by cutting out the acetylated and unacetylated ¹⁴C-chloramphenicol regions. Radioactivity was determined by scintillation counting.

cannot conclude that the stimulation observed is at the level of RNA synthesis. The observation that the sequences required for regulation lie more than 200 bases upstream of the normal *hsp70* start site is, however, consistent with this hypothesis. There is evidence for both structural¹⁹ and functional similarities between the *myc* gene product and products of the adenovirus E1a region. Both genes are capable of immortalizing primary cells and of complementing the ability of the c-Ha-ras gene to transform primary cells^{7,19,20}. The E1a region has been shown to stimulate transcription of a wide variety of cellular and viral promoters, including the mammalian *hsp70* gene^{6,14,21-27}. Stimulation of the adenovirus E2 promoter and the human

β-globin promoter by the E1a region does not depend on the presence of a specific regulatory sequence^{14,26}. These results contrast with the experiments described above for *myc* stimulation of the heat shock promoter region, and suggest that the E1a and *myc* gene products may differ in their specificities.

Regardless of mechanism, the observation that the *myc* gene can stimulate gene expression by acting through a specific sequence raises several interesting questions. Does the *myc* gene product bind directly to this sequence, or does it interact with or induce a cellular function that recognizes these sequences? The observation that the *myc* protein is localized to the nucleus is consistent with its ability to bind to DNA sequences^{4,5}. That the mouse *myc* gene can apparently act through *Drosophila* sequences suggests either that the mechanism of stimulation is strongly conserved over evolution or that sequences capable of responding to *myc* protein are fortuitously contained on this particular promoter. This observation then raises the possibility that *myc* can regulate gene expression through similar sequences in mammalian cells, perhaps allowing coordinated regulation of a set of genes.

We thank P. DiNocera, I. Dawid, C. Gorman, M. Gilman, H. Land and L. Parada for donation of plasmid DNAs, R. A. Weinberg for stimulating discussions, M. Gilman for comments on the manuscript and R.E.S.D. Myers for assistance in making plasmid pfsmyc-20. R.E.K. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research, A.S.B. is a fellow of the NIH. This work was supported by NSF grant PCM-8200309, NIH grants R01-GM32467 and PPG P01-CA26717 to P.A.S. and partially by NIH core grant P01-CA14051.

Received 25 June; accepted 27 September 1984.

1. Hayward, W. S., Neel, B. G. & Astrin, S. M. *Nature* **290**, 475-479 (1981).
2. Shen-Ong, G. L. C., Keath, E. J., Piccoli, S. P. & Cole, M. D. *Cell* **31**, 443-450 (1982).
3. Leder, P. *et al.* *Science* **222**, 765-771 (1983).
4. Donner, P., Greiser-Wilke, I. & Moellering, K. *Nature* **296**, 262-266 (1982).
5. Abrams, H. D., Rohrschneider, L. R. & Eisenman, R. N. *Cell* **29**, 427-439 (1982).
6. Imperiale, M. J., Kao, H.-T., Feldman, L. T., Nevins, J. R. & Strickland, S. *Molec. cell. Biol.* **4**, 867-874 (1984).
7. Land, H., Parada, L. F. & Weinberg, R. A. *Nature* **304**, 596-601 (1983).
8. Holmgren, R., Livak, K., Morimoto, R., Freund, R. & Meselson, M. *Cell* **18**, 1359-1370 (1979).
9. Karch, F., Torok, I. & Tissieres, A. *J. molec. Biol.* **148**, 219-230 (1981).
10. Mason, P. J., Torok, I., Kiss, I., Karch, F. & Udvardy, A. *J. molec. Biol.* **156**, 21-35 (1982).
11. Kaufman, R. J. & Sharp, P. A. *Molec. cell. Biol.* **2**, 1304-1313 (1982).
12. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
13. Di Nocera, P. P. & Dawid, I. B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7095-7098 (1983).
14. Kingston, R. E., Kaufman, R. J. & Sharp, P. A. *Molec. cell. Biol.* **4**, 1970-1977 (1984).
15. Urlaub, G. & Chasin, L. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4216-4220 (1980).
16. Urlaub, G., Kas, E., Carothers, A. M. & Chasin, L. A. *Cell* **33**, 405-412 (1983).
17. Kelley, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603-610 (1983).
18. Hackett, R. W. & Lis, J. T. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6196-6200 (1981).
19. Ralston, R. & Bishop, J. M. *Nature* **306**, 803-806 (1983).
20. Ruley, H. E. *Nature* **304**, 602-607 (1983).
21. Nevins, J. R. *Cell* **29**, 913-919 (1982).
22. Berk, A. J., Lee, F., Harrison, T., Williams, J. & Sharp, P. A. *Cell* **17**, 935-944 (1979).
23. Bos, J. L. & ten Wolde-Kraamwinkel, H. C. *EMBO J.* **2**, 73-76 (1983).
24. Elkaim, R., Coding, C. & Kedinger, C. *Nucleic Acids Res.* **11**, 7105-7117 (1983).
25. Gaynor, R. B., Hillman, D. & Berk, A. J. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1193-1197 (1984).
26. Green, M. R., Treisman, R. J. & Maniatis, T. *Cell* **35**, 137-148 (1983).
27. Treisman, R. J., Green, M. R. & Maniatis, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7428-7432 (1983).
28. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
29. Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456-467 (1973).

DNA sequences from the quagga, an extinct member of the horse family

Russell Higuchi*, Barbara Bowman*, Mary Freiberger*, Oliver A. Ryder† & Allan C. Wilson*

* Department of Biochemistry, University of California, Berkeley, California 94720, USA

† Research Department, San Diego Zoo, San Diego, California 92103, USA

To determine whether DNA survives and can be recovered from the remains of extinct creatures, we have examined dried muscle from a museum specimen of the quagga, a zebra-like species (*Equus quagga*) that became extinct in 1883 (ref. 1). We report that DNA was extracted from this tissue in amounts approaching 1% of that expected from fresh muscle, and that the DNA was of

relatively low molecular weight. Among the many clones obtained from the quagga DNA, two containing pieces of mitochondrial DNA (mtDNA) were sequenced. These sequences, comprising 229 nucleotide pairs, differ by 12 base substitutions from the corresponding sequences of mtDNA from a mountain zebra, an extant member of the genus *Equus*. The number, nature and locations of the substitutions imply that there has been little or no postmortem modification of the quagga DNA sequences, and that the two species had a common ancestor 3–4 Myr ago, consistent with fossil evidence concerning the age of the genus *Equus*².

Rau supplied us with a small piece of dried muscle and connective tissue attached to the salt-preserved skin of a quagga (*Equus quagga*) that died 140 yr ago and which had been stored in the Museum of Natural History at Mainz, West Germany³. After releasing the DNA from 0.7 g of this tissue with proteinase K and detergent⁴, we purified it by phenol extraction and ethanol precipitation, then fractionated it by gel-filtration and electrophoresis. Most of the DNA detectable by fluorescence in the presence of ethidium bromide was <500 base pairs (bp) long, if double-stranded. After transfer to nitrocellulose paper⁵, this DNA annealed with total genomic DNA of the mountain zebra (*Equus zebra*) that had been labelled with ³²P by nick-translation⁶. The intensity of the hybridization signal was roughly that expected if the ethidium fluorescence seen in the gel were due to DNA closely related to the probe; by contrast, when human DNA is probed with mountain zebra DNA, 1,000 times more human DNA is needed to give the same signal. These results imply that there is as much as 5 µg of low-molecular weight, zebra-like DNA per gram of the quagga tissue.

About 10 ng of quagga DNA were cloned into the λ gt10 vector, as detailed in Fig. 1 legend; 25,000 resultant plaques containing quagga DNA inserts were screened by hybridization to ³²P-labelled mtDNA from the mountain zebra. Plaques showing a positive reaction were picked off and purified, then subcloned into the EcoRI site of phage M13mp11, which provided templates for dideoxy sequencing⁷. (That nuclear DNA was represented among the quagga DNA clones was shown by screening the same plaque lifts with purified *E. zebra* satellite DNA as a probe.) These quagga mtDNA clones were then used

to identify *E. zebra* mtDNA clones containing homologous sequences.

Figure 1 compares the sequences of two quagga clones with the corresponding sequences from *E. zebra* mtDNA. By comparing these sequences with the known sequences of cow⁸ and human⁹ mtDNA, we established that they code for parts of two proteins: 117 bp originating from a gene (unidentified reading frame 1, URF1) coding for a protein of unknown function and 112 bp from the cytochrome oxidase I gene. At the 229 positions compared, the quagga and zebra sequences differ by 12 base substitutions, of which all are transitions and 10 do not cause amino acid replacements. There are no premature termination codons, deletions or additions relative to the zebra sequences. Such a pattern of divergence is close to that expected from previous mtDNA sequence comparisons of other species^{10,11}. A high ratio of transitions to transversions, and of silent changes to amino acid replacements, is typical of mtDNAs from closely related species; the absence of transversions between the equine sequences examined here is not surprising^{10,11}. The degree of sequence difference between the two species (~5%) is in approximate agreement with that estimated by George and Ryder for several pairs of living species of *Equus*, based on restriction maps of the whole mitochondrial genome¹².

A few of the differences between the quagga and zebra DNA sequences could be due to postmortem change (including cloning artefacts). At the 229 positions compared, the cow differs by three amino acid replacements from the zebra and by five from the quagga. Hence, the two amino acid differences (Fig. 1) between the zebra and quagga probably arose in the quagga lineage. A ratio of two changes on the quagga lineage to zero on the zebra lineage to three on the lineage linking the cow with the common ancestor of the quagga and zebra is a little surprising when considered in relation to the durations of these lineages, which are in the ratio of ~1:1:38 (see Fig. 2). Further sequencing of mitochondrial and nuclear DNA from quagga and zebra will permit a better assessment of the likelihood of postmortem change.

The phylogenetic tree in Fig. 2 shows how the quagga mtDNA sequences are related to those of three other mammals. This

Unidentified reading frame 1

Quagga	C CCA ATC CTG CTC GCC GTC GCA TTC CTC ACA CTA GTT GAA CGA AAA GTC TTA GGC TAC ATA CAA CTT CGT AAA GGA CCC AAC ATC GTA GGC CCC TAT GGC CTA CTA CAA CCC ATT AC
ZebraT.....G.....T..... .C..... .G*

Cytochrome oxidase I

Quagga	A GGA GGA TTC GTT CAC TGA TTC CCT CTA TTC TCA GGA TAC ACA CTC AAC CAA ACC TGA GCA AAA ATT CAC TTT AGA ATT ATA TTC GTA GGG GTC AAC ATA ATT TTC TTC CCA
Zebra	G.....T.....G..... .C.....A.....T.....C*

Fig. 1 Sequences of the coding strands determined for two pieces of quagga mtDNA. The sequences are arranged in triplets corresponding to the amino acids that they encode. At 12 positions, the quagga sequences differ from those of mtDNA from a mountain zebra; only for these positions is the nature of the base specified for the zebra. The two asterisks identify triplets at which the zebra and quagga differ by an amino acid replacement.

Methods: Quagga DNA was cloned by the sequential blunt-ending and repair of the fragments with T4 DNA polymerase and *Escherichia coli* DNA polymerase¹⁹; addition of synthetic linker molecules containing the EcoRI restriction endonuclease recognition sequence²⁰; cleavage of the added linker with EcoRI to produce sticky ends; and ligation to the EcoRI-cleaved DNA of the λ phage cloning vector, λ gt10 (ref. 21) (this vector is suitable for the efficient cloning of small amounts of low-molecular weight DNA). To ~10 ng of quagga DNA were added 200 ng of EcoRI-cut λ gt10 and 1.5 µg of λ c1857 Sam7 DNA (New England Biolabs). These DNAs were precipitated with ethanol, dissolved in 4 µl of 10 mM Tris-HCl (pH 7.4), 10 mM MgCl₂, 50 mM NaCl, and kept at 67 °C for 5 min, then placed in a water bath at 50 °C for 15 min; 0.5 µl each of 10 mM ATP, 100 mM dithiothreitol and T4 DNA ligase (New England Biolabs) were added and the reaction incubated overnight at 4 °C. This DNA was then packaged *in vitro* into phage particles²² and allowed to infect *E. coli* BNN102 (ref. 23) having the *hfl150* genotype²⁴. (This host becomes a lysogen so frequently that intact or reassembled λ gt10 genomes form only minute plaques. Only genomes with inserts that interrupt the *cI* gene (*cI* repressor) will form large, clear plaques. Pre-annealing with λ c1857 Sam7 DNA favours the formation of multimeric genomes (monomeric genomes will not package into phage; ref. 25), which increases the efficiency of the cloning. Because this DNA requires a *supF* host for propagation, it does not form plaques on BNN102.) The resultant 'library' of non-lysogenic phage was screened by annealing ³²P-labelled, purified mountain zebra mtDNA to 'plaque lifts'²⁶ of the library. Plaques showing a positive reaction with this probe were picked off and phage from them purified and the DNAs isolated. The inserts in these phage genomes were removed with EcoRI and recloned into the EcoRI site of the replicative form of the genome of M13mp11 (Pharmacia P-L Biochemicals); this provided single-stranded M13 DNA templates containing quagga DNA sequences suitable for sequencing via the primed-synthesis, dideoxynucleoside chain-termination method of Sanger *et al.*⁷. Fragments of mountain zebra mtDNA that had been cut with *TaqI* and *Hpa*II were subcloned into the AccI site of M13mp11. Zebra clones containing DNAs related to the two quagga sequences were identified by probing with radioactively labelled quagga clones, then sequenced.

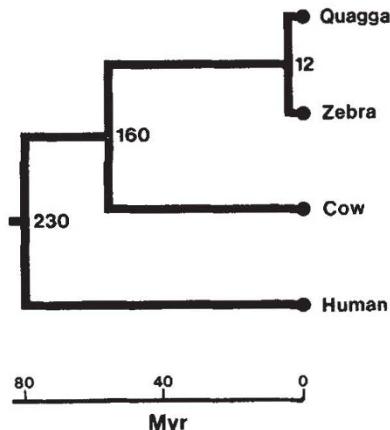


Fig. 2 A phylogenetic tree relating mtDNA sequences from the quagga to those of three other mammals. The branching order of the tree was determined by applying the parsimony method¹¹ to the amino acid sequences read from the nucleotide sequences; this method eliminates the phylogenetic noise created by multiple substitutions at silent sites²⁷. The positions of the nodes in the tree are in proportion to the corrected number of substitutions by which the nucleotide sequences differ. These numbers (12, 160 and 230) were obtained for distantly related sequences by multiplying the observed number of transversions by 10, based on recommendations by Brown *et al.*¹¹; no correction was needed for the quagga-zebra comparison because all the changes were transitions. The cow and human sequences were taken from Anderson *et al.*^{8,9}.

tree accounts for the observed sequence differences with fewer mutational events than trees with other branching orders. The biochemical tree matches that predicted by anatomical evidence, which associates the zebra and quagga most closely with each other, and more closely with the cow than with man. The time scale for the tree is based on the assumption that the time elapsed since the divergence of primate and ungulate lineages is 80 Myr¹¹ and that the rate of divergence among the sequences is constant. Divergence was estimated by correcting the observed sequence differences for multiple hits at the same nucleotide site, as discussed in Fig. 2 legend. The resulting estimates of 3–4 Myr for the quagga-zebra split and 55–60 Myr for the split between the odd-toed and even-toed ungulates are in satisfactory agreement with fossil evidence². This agreement supports the notion that evolution of mtDNA is approximately clock-like.

The present report seems to be the first demonstration that clonable DNA sequence information can be recovered from the remains of an extinct species. The only genetic macromolecules recovered previously from such remains were proteins¹³. By further sequencing of the mitochondrial and nuclear DNAs of the quagga and related extant species, we hope to obtain an evolutionary history of the genus *Equus* in which the quagga is accurately placed; the phylogenetic position of the quagga within this genus is disputed^{14–16}. In a preliminary report of some of the findings presented here¹⁷, we also described the recovery of DNA from a 40,000 yr-old, frozen Siberian mammoth and raised the possibility that DNA occurs in extracts of 25-Myr-old insects preserved in amber¹⁸. If the long-term survival of DNA proves to be a general phenomenon, several fields including palaeontology, evolutionary biology, archaeology and forensic science may benefit.

We thank R. Rau for the quagga tissue and M. George for highly purified mountain zebra DNA; B. Tong for help with the cloning; M. McClelland for help with computer sequence searches; R. Cann and K. Vincent for advice on DNA sequencing; R. T. White for advice on DNA extraction; T. Huynh and R. Davis for λ gt10 and C. Holcomb for advice on its use; and E. Prager for statistical help and editorial advice. NIH and NSF grants provided partial support for this work.

Received 25 June; accepted 6 September 1984.

- Dolan, J. M. *Zooloz* **56**, 13–15 (1983).
- Savage, D. E. & Russell, D. E. *Mammalian Paleofaunas of the World*, 346–397 (Addison-Wesley, Reading, Massachusetts, 1983).
- Rau, R. *Ann. S. Afr. Mus.* **65**, 41–87 (1974); **77**, 27–45 (1978).
- Robbins, J. *et al.* *J. biol. Chem.* **254**, 6187–6195 (1979).
- Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Anderson, S. *et al.* *J. molec. Biol.* **156**, 683–717 (1982).
- Anderson, S. *et al.* *Nature* **290**, 457–465 (1981).
- Brown, G. G. & Simpson, M. V. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3246–3250 (1982).
- Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. *J. molec. Evol.* **18**, 225–239 (1982).
- George, M. & Ryder, O. A. *Genetics* **104**, s27 (1983).
- Prager, E. M., Wilson, A. C., Lowenstein, J. M. & Sarich, V. M. *Science* **209**, 287–289 (1980).
- Bennett, D. K. *Syst. Zool.* **29**, 272–287 (1980).
- Eisenmann, V. *Cah. Paléont.* **1**, 186 (1980).
- Groves, C. P. & Willoughby, D. P. *Mammalia* **45**, 321–354 (1981).
- Higuchi, R. & Wilson, A. C. *Fedn Proc.* **43**, 1557 (1984).
- Poinar, G. O. & Hess, R. *Science* **215**, 1241–1242 (1982).
- O'Farrell, P. *Focus* **3**, 1 (1981).
- Scheller, R. H., Dickerson, R. E., Boyer, H. W., Riggs, A. D. & Itakura, K. *Science* **196**, 177–180 (1977).
- Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning Techniques: A Practical Approach* (ed. Glover, D. M.) (IRL, Oxford, 1984).
- Hohla, B. *Meth. Enzym.* **68**, 299–309 (1979).
- Young, R. A. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1194–1198 (1983).
- Hoyt, M. A., Knight, D. M., Das, A., Miller, H. I. & Echols, H. *Cell* **31**, 565–573 (1982).
- Feiss, M. & Becker, A. in *Lambda II* (eds Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A.) 305–330 (Cold Spring Harbor Laboratory, New York, 1983).
- Woo, S. L. C. *Meth. Enzym.* **68**, 389–395 (1979).
- Doolittle, R. F. *Science* **214**, 149–159 (1981).

Nitrogen fixation by a methanogenic archaeabacterium

Patti A. Murray & Stephen H. Zinder

Department of Microbiology, New York State College of Agriculture and Life Sciences, Ithaca, New York 14853, USA

The ability to fix nitrogen (N_2) is found among a wide variety of the prokaryotic eubacteria, but not in eukaryotes¹. In addition to the prokaryotic eubacteria and eukaryotes, a third 'kingdom'—the archaeabacteria—has been defined based on the comparison of 16S ribosomal oligonucleotide sequence catalogues^{2,3}. Included in the archaeabacterial kingdom are certain obligate halophiles and thermoacidophiles, and the methanogens, strictly anaerobic, methane-producing bacteria⁴. Here we report diazotrophy by an archaeabacterium, the methanogen *Methanosarcina barkeri* strain 227. Because it has been proposed that the archaeabacteria, eubacteria and eukaryotes diverged at an early stage in evolution^{2,3}, the discovery of diazotrophy (N_2 fixation) in a member of the archaeabacterial group raises interesting evolutionary questions.¹

In the thermophilic *Methanosarcina* strain TM-1 (ref. 5), growth and methanogenesis can be limited by 1 mM NH_4^+ , as the sole source of fixed nitrogen in the growth medium⁶, as has been found in other methanogens^{7,8}. Under such nitrogen-limiting conditions, *Methanosarcina* strain TM-1 stores a glycogen-like polysaccharide in amounts up to 20 mg per g protein⁶. We initiated a similar comparative study using a mesophile, *Methanosarcina barkeri* strain 227 (ref. 9). However, this culture showed no signs of nitrogen limitation of growth, and could be transferred repeatedly in methanol growth medium from which NH_4^+ was omitted. The headspaces in these culture vials consisted of 70% N_2 /30% CO_2 . To test whether N_2 was the nitrogen source of these cultures, an N_2 -free argon headspace was used. Figure 1a shows that, in the absence of both NH_4^+ and N_2 , methanogenesis by *M. barkeri* strain 227 was significantly decreased. When 14% N_2 was added to the argon headspace, there was nearly the same final yield of methane as in the vials provided with 20 mM NH_4^+ . Cell growth, measured as dry weight, was only $15 \mu\text{g ml}^{-1}$ on day 16 in vials having only argon as the headspace, but was $95 \mu\text{g ml}^{-1}$ in vials with N_2 added.

This apparent ability of the *Methanosarcina* culture to fix N_2 could possibly be caused by an N_2 -fixing contaminant. However,