letters to nature

1. Wang, J. C. Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. *Q. Rev. Biophys.* **31**, 107–144 (1998).
2. Liu, L. F., Liu, C. C. & Alberts, B. M. Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-stranded break. *Cell* **19**, 697–707 (1980).
3. Hsieh, T. Knotting of the circular duplex DNA by type II DNA topoisomerase from *D. melanogaster*. *J. Biol. Chem.* **258**, 8413–8420 (1983).
4. Roca, J. & Wang, J. C. The capture of a DNA double helix by an ATP-dependent protein clamp: a key step in DNA transport by type II DNA topoisomerase. *Cell* **71**, 833–840 (1992).
5. Roca, J., Berger, J. M. & Harrison, S. C. & Wang, J. C. DNA transport by a type II topoisomerase: Direct evidence for a two-gate mechanism. *Proc. Natl Acad. Sci. USA* **93**, 4057–4062 (1996).
6. Rybenkov, V. V., Ullsperger, C. Vologodskii, A. V. & Cozzarelli, N. R. Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science* **277**, 690–693 (1997).
7. Osheroff, N., Shelton, E. R. & Brutlag, D. L. DNA topoisomerase II from *D. melanogaster*: relaxation of supercoiled DNA. *J. Biol. Chem.* **258**, 9536–9543 (1983).
8. Uemura, T. & Yanagida, M. Mitotic spindle pulls but fails to separate chromosomes in type II DNA topoisomerase mutants: uncoordinated mitosis. *EMBO J.* **5**, 1003–1010 (1986).
9. Ishida, R. *et al.* Inhibition of DNA topoisomerase II by ICRF-193 induces polyploidization by uncoupling chromosome dynamics from other cell cycle events. *J. Cell Biol.* **126**, 1341–1351 (1994).
10. Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A. & Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science* **271**, 1835–1837 (1996).
11. Strick, T. R., Allemand, J. F., Bensimon, D. & Croquette, V. The behavior of super-coiled DNA. *Biophys. J.* **74**, 2016–2028 (1998).
12. Strick, T. R., Croquette, V. & Bensimon, D. Homologous pairing in stretched supercoiled DNA. *Proc. Natl Acad. Sci. USA* **95**, 10579–10583 (1998).
13. Allemand, J. F., Bensimon, D., Lavery, R. & Croquette, V. Stretched and overwound DNA forms a Pauling-like structure with exposed bases. *Proc. Natl Acad. Sci. USA* **95**, 14152–14157 (1998).
14. Landau, L. & Lifchitz, E. *Theory of Elasticity* (Mir Editions, Moscow, 1967).
15. Moroz, J. D. & Nelson, P. Torsional directed walks, entropic elasticity and DNA twist stiffness. *Proc. Natl Acad. Sci. USA* **94**, 14418–14422 (1998).
16. Bouchiat, C. & Mézard, M. Elasticity model of a supercoiled DNA molecule. *Phys. Rev. Lett.* **80**, 1556–1559 (1998).
17. Brown, P. O. & Cozzarelli, N. R. A sign inversion mechanism for enzymatic supercoiling of DNA. *Science* **206**, 1081–1083 (1979).
18. Hua, W., Young, E. C., Fleming, M. L. & Gelles, J. Coupling of kinesin steps to ATP hydrolysis. *Nature* **388**, 390–393 (1997).
19. Harkins, T. T. & Lindsley, J. E. Pre-steady-state analysis of ATP hydrolysis by *Saccharomyces cerevisiae* DNA topoisomerase II. 1. A DNA-dependent burst in ATP hydrolysis. *Biochemistry* **37**, 7292–7298 (1998).
20. Harkins, T. T., Lewis, T. J. & Lindsley, J. E. Pre-steady-state analysis of ATP hydrolysis by *Saccharomyces cerevisiae* DNA topoisomerase II. 2. Kinetic mechanism for the sequential hydrolysis of two ATP. *Biochemistry* **37**, 7299–7312 (1998).
21. Wang, M. D. *et al.* Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902–907 (1998).
22. Visscher, K., Schnitzer, M. J. & Block, S. M. Single kinesin molecules studied with a molecular force clamp. *Nature* **400**, 184–189 (1999).
23. Zechiedrich, E. L. & Osheroff, N. Eukaryotic topoisomerases recognize nucleic acid topology by preferentially interacting with DNA crossovers. *EMBO J.* **9**, 4555–4562 (1990).
24. Roca, J., Berger, J. M. & Wang, J. C. On the simultaneous binding of eukaryotic DNA topoisomerase II to a pair of double-stranded DNA helices. *J. Biol. Chem.* **268**, 14250–14255 (1993).
25. Froelich-Ammon, S. J. & Osheroff, N. Topoisomerase poisons: harnessing the dark side of enzyme mechanism. *J. Biol. Chem.* **270**, 21429–21432 (1995).

errata

Uptake of apoptotic cells drives the growth of a pathogenic trypanosome in macrophages

Célio G. Freire-de-Lima, Danielle O. Nascimento, Milena B. P. Soares, Patricia T. Bozza, Hugo C. Castro-Faria-Neto, Fernando G. de Mello, George A. DosReis & Marcela F. Lopes

In Fig. 4b and c the label beneath the fourth column should have read 'Apo-1 + indomethacin' rather than 'Anti-α_v + indomethacin'. ☐

Formation of molecular gas in the tidal debris of violent galaxy–galaxy interactions

Jonathan Braine, Ute Lisenfeld, Pierre-Alain Duc & Stéphane Leon

The name of the third author, Pierre-Alain Duc, was inadvertently mis-spelled as Due. In Fig. 3, the right-hand vertical axis label should have been the same as that for the left-hand vertical axis: 'CO($1 \to 0$) temperature (mK)'. ☐

corrections

The DNA sequence of human chromosome 22

I. Dunham, N. Shimizu, B. A. Roe, S. Chissoe *et al.*

The names of the following authors were omitted from the complete list at the end of the Article: P. Wilkinson (The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK); A. Bodenteich, K. Hartman, X. Hu, A. S. Khan, L. Lane, Y. Tilahun & H. Wright (Department of Chemistry and Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 311, Norman, Oklahoma 73019, USA). ☐

Tribosphenic mammal from the North American Early Cretaceous

Richard L. Cifelli

The binomen *Montanalestes keebleri* was established for a tribosphenic mammal of probable eutherian affinities from the Lower Cretaceous (Aptian–Albian) Cloverly Formation, Montana, USA. The trivial name was designated for the Keebler family, of Billings, Montana. As pointed out to me by T. Harrison, this designation is implicitly plural, hence, the proper suffix is -*orum*[1]. The species name is hereby corrected to *Montanalestes keeblerorum*, as provided by articles 19 and 32 of the International Code of Zoological Nomenclature[1]. ☐

1. International Commission on Zoological Nomenclature. *International Code of Zoological Nomenclature* 4th edn (International Trust for Zoological Nomenclature, The Natural History Museum, London, 1999).

14. Suessbrich, H. Specific blockade of slowly activating I_{sK} channels by chromanols—impact on the role of I_{sK} channels in epithelia. *FEBS Lett.* **396**, 271–275 (1996).

15. Abbott, G. W. *et al.* MiRP1 forms I_{Kr} potassium channels with HERG and is associated with cardiac arrhythmia. *Cell* **97**, 175–187 (1999).

16. Busch, A. E. *et al.* The role of the IsK protein in the specific pharmacological properties of the I_{Ks} channel complex. *Br. J. Pharmacol.* **122**, 187–189 (1997).

17. Devor, D. C., Singh, A. K., Gerlach, A. C., Frizzell, R. A. & Bridges, R. J. Inhibition of intestinal Cl⁻ secretion by clotrimazole: direct effect on basolateral membrane K⁺ channels. *Am. J. Physiol.* **273**, C531–C540 (1997).

18. Rufo, P. A. *et al.* The antifungal antibiotic, clotrimazole, inhibits chloride secretion by human intestinal T84 cells via blockade of distinct basolateral K⁺ conductances. Demonstration of efficacy in intact rabbit colon and in an *in vivo* mouse model of cholera. *J. Clin. Invest.* **100**, 3111–3120 (1997).

19. Yang, W. P. *et al.* KvLQT1, a voltage-gated potassium channel responsible for human cardiac arrhythmias. *Proc. Natl Acad. Sci. USA* **94**, 4017–4021 (1997).

20. Schroeder, B. C., Kubisch, C., Stein, V. & Jentsch, T. J. Moderate loss of function of cyclic-AMP-modulated KCN2Q2/KCNQ3 K⁺ channels causes epilepsy. *Nature* **396**, 687–690 (1998).

21. Zerangue, N., Schwappach, B., Jan, Y. N. & Jan, L. Y. A new ER trafficking signal regulates the subunit stoichiometry of plasma membrane K_{ATP} channels. *Neuron* **22**, 537–548 (1999).

22. McDonald, T. V. *et al.* A minK-HERG complex regulates the cardiac potassium current I_{Kr}. *Nature* **388**, 289–292 (1997).

23. Curran, M. E. *et al.* A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. *Cell* **80**, 795–803 (1995).

24. Lohrmann, E. *et al.* A new class of inhibitors of cAMP-mediated Cl⁻ secretion in rabbit colon, acting by the reduction of cAMP-activated Cl⁻ conductance. *Pflügers Arch.* **429**, 517–530 (1995).

25. Devor, D. C., Singh, A. K., Frizzell, R. A. & Bridges, R. J. Modulation of Cl⁻ secretion by benzimidazolones. I. Direct activation of a Ca^{2+}-dependent K⁺ channel. *Am. J. Physiol.* **271**, L775–L784 (1996).

26. Mall, M. *et al.* Cholinergic ion secretion in human colon requires coactivation by cAMP. *Am. J. Physiol.* **275**, G1274–G1281 (1998).

27. MacVinish, L. J., Hickman, M. E., Mufti, D. A., Durrington, H. J. & Cuthbert, A. W. Importance of basolateral K⁺ conductance in maintaining Cl⁻ secretion in murine nasal and colonic epithelia. *J. Physiol. (Lond.)* **510**, 237–247 (1998).

28. Warth, R. *et al.* The cAMP-regulated and 293B-inhibited K⁺ conductance of rat colonic crypt base cells. *Pflügers Arch.* **432**, 81–88 (1996).

29. Diener, M., Hug, F., Strabel, D. & Scharrer, E. Cyclic AMP-dependent regulation of K⁺ transport in the rat distal colon. *Br. J. Pharmacol.* **118**, 1477–1487 (1996).

30. Trezise, A. E. & Buchwald, M. *In vivo* cell-specific expression of the cystic fibrosis transmembrane conductance regulator. *Nature* **353**, 434–437 (1991).

Uptake of apoptotic cells drives the growth of a pathogenic trypanosome in macrophages

Célio G. Freire-de-Lima*, Danielle O. Nascimento*, Milena B. P. Soares†, Patricia T. Bozza‡, Hugo C. Castro-Faria-Neto‡, Fernando G. de Mello*, George A. DosReis* & Marcela F. Lopes*

* *Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, 21944-970, Rio de Janeiro, RJ, Brazil*
† *Centro de Pesquisas Gonçalo Moniz, FIOCRUZ, 40295-001, Salvador, BA, Brazil*
‡ *Instituto Oswaldo Cruz, FIOCRUZ, 21045-900, Rio de Janeiro, RJ, Brazil*

After apoptosis, phagocytes prevent inflammation and tissue damage by the uptake and removal of dead cells[1]. In addition, apoptotic cells evoke an anti-inflammatory response through macrophages[2,3]. We have previously shown that there is intense lymphocyte apoptosis in an experimental model of Chagas' disease[4], a debilitating cardiac illness caused by the protozoan *Trypanosoma cruzi*. Here we show that the interaction of apoptotic, but not necrotic T lymphocytes with macrophages infected with *T. cruzi* fuels parasite growth in a manner dependent on prostaglandins, transforming growth factor-β (TGF-β) and polyamine biosynthesis. We show that the vitronectin receptor is critical, in both apoptotic-cell cytoadherence and the induction of prostaglandin E_2/TGF-β release and ornithine decarboxylase activity in macrophages. A single injection of apoptotic cells in infected mice increases parasitaemia, whereas treatment with cyclooxygenase inhibitors almost completely ablates it *in vivo*. These results suggest that continual lymphocyte apoptosis and phagocytosis of apoptotic cells by macrophages have a role in parasite persistence in the host, and that cyclooxygenase inhibitors have potential therapeutic application in the control of parasite replication and spread in Chagas' disease.

We have already shown that the onset of activation-induced cell death in CD4⁺ T cells exacerbates parasite replication in co-cultured macrophages infected with *T. cruzi*[5]. To investigate whether the clearance of apoptotic cells predisposes macrophages to *T. cruzi* infection, murine resident peritoneal macrophages were exposed to apoptotic, necrotic or viable splenic T cells first, and then washed and infected. Apoptotic, but not necrotic or living T cells increased *T. cruzi* growth in macrophage cultures (Fig. 1a). Similar results were obtained when apoptotic or necrotic cells were added after *T. cruzi* infection (data not shown). Nevertheless, treatment of lymphocytes with the caspase-inhibitor zVAD-fmk peptide before apoptosis induction, rescued T cells from death (data not shown) and prevented the increase in parasite replication (Fig. 1b) in a dose-dependent manner. In another model, peritoneal macrophages from mice infected with *T. cruzi* were incubated with apoptotic or necrotic cells. Apoptotic, but not necrotic cells also exacerbated endogenous *T. cruzi* growth in these *in vivo* infected macrophages (Fig. 1c). In agreement with *in vitro* results, a single *in vivo* injection of apoptotic, but not necrotic cells in *T. cruzi*-infected mice resulted in a sudden rise in parasitaemia (Fig. 1d).

Previous studies pointed to a role for an integrin in the recognition and ingestion of apoptotic cells by macrophages[6,7]. We observed that RGDS, but not RGES peptide blocked apoptotic cell binding and reduced *T. cruzi* growth within macrophages (data not shown).
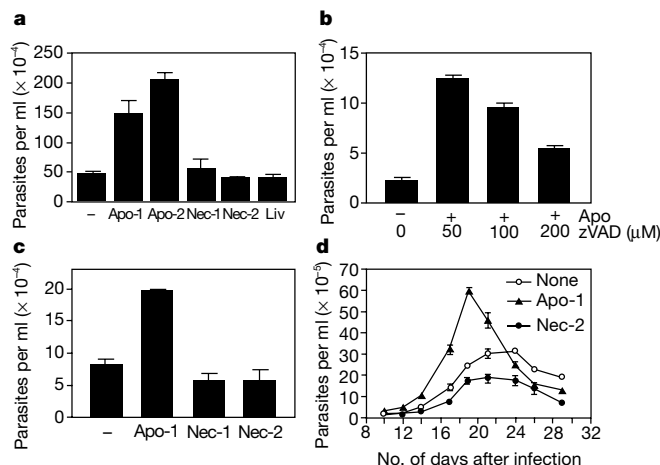

Figure 1 Apoptotic cells exacerbate parasite growth in *T. cruzi* infection. **a**, Peritoneal resident macrophages were either untreated (–) or exposed to Apo-1, Apo-2, Nec-1, Nec-2 or to living cells (Liv). After 5 days, cells were removed and macrophages were infected with *T. cruzi*; parasites were counted 10 days later. **b**, T cells (Apo) were treated with the indicated doses of zVAD-fmk (zVAD) before apoptosis induction (by heating) and added to macrophages infected *in vitro*. Parasites were counted 7 days later. Maximal cell death (100%) and increased parasite replication still occurred with zVAD-fmk at 50 μM. **c**, Macrophages from infected mice were exposed to Apo-1, Nec-1 or Nec-2 throughout culture, and trypomastigotes counted 10 days later. **d**, Mice were injected 7 days after infection with Apo-1 or Nec-2, and parasitaemia was monitored.

The vitronectin receptor (VnR) integrin ($\alpha_v\beta_3$, CD51/CD61) is involved in phagocytosis of apoptotic cells by macrophages[6,7]. Both VnR and CD36 take part in the macrophage receptor that bridges thrombospondin exposed on apoptotic cells[7]. Flow cytometry analysis revealed that both α_v- and β_3-integrin chains were absent from the surface of control resident macrophages, but their expression was upregulated in the course of *T. cruzi* infection (Fig. 2a, b). In addition, $\alpha_v\beta_3$ expression was induced in control macrophages after a 24-h culture (not shown). We therefore investigated the role of VnR in the binding of apoptotic cells by *in vivo* infected macrophages. After a 3-h incubation, nearly all macrophages were rosetted by apoptotic cells (mean $93.5 \pm 2.5\%$ positive macro-

Figure 2 Effects of apoptotic cells on macrophages infected with *T. cruzi* are mediated by the vitronectin receptor. α_v (**a**) and β_3 (**b**) VnR chains are expressed by macrophages from infected (lower panels), but not uninfected (upper panels), mice. Mac-1$^+$ cells were gated and analysed for α_v (**a**) or β_3 (**b**) expression. Bold lines, anti-α_v (**a**) or anti-β_3 (**b**) staining; dotted lines; cells stained after treatment with unlabelled anti-α_v (**a**) or anti-β_3 (**b**) monoclonal antibodies; dashed lines, unstained cells. **c**, Anti-VnR monoclonal antibodies and Fab inhibit apoptotic cell binding to macrophages. Macrophages infected *in vivo* were treated with anti-α_v Fab or control anti-CD69 Fab-1, anti-α_v, anti-β_3 or control hamster IgG monoclonal antibodies (mAb) (20 min), and then exposed to Apo-1. Adhesion was determined after 3 h. **d**, **e**, Anti-α_v or anti-β_3, but not control monoclonal antibody drives *T. cruzi* replication within macrophages. Macrophages infected *in vivo* were treated with monoclonal antibodies (**d**) or were treated first with anti-α_v Fab, Fab-1 or control hamster IgG Fab-2 (50 μg ml^{-1}) for 20 min, washed and treated with goat anti-hamster IgG antibody (GαH) at 10 μg ml^{-1} (**e**). Parasites were counted as indicated (**d**) or after 10 days (**e**). **f**, Anti-α_v Fab inhibits the apoptotic cell effects on *T. cruzi* replication. Macrophages infected *in vivo* were treated first with Fab fragments (20 min) and then with Apo-1. Parasites were counted 10 days later.

phages). Both intact monoclonal antibodies against α_v- or β_3-VnR chains and anti-α_v fragment of antigen binding (Fab) inhibited apoptotic cell binding by 40–50% (Fig. 2c).

We then investigated VnR involvement in parasite replication. Engagement of VnR by anti-α_v or anti-β_3 monoclonal antibodies, either soluble (Fig. 2d) or immobilized on plates (Fig. 3c), markedly increased *T. cruzi* growth in macrophages in the absence of apoptotic cells. Soluble anti-α_v Fab fragments failed to enhance parasite replication, but promoted *T. cruzi* growth upon crosslinkage by a secondary goat anti-hamster IgG (Fig. 2e). Most notably, soluble anti-α_v Fab completely blocked the apoptotic cell effect on parasite replication in macrophages obtained from infected mice (Fig. 2f). A control hamster anti-CD69 Fab had no effect (Fab-1 in Fig. 2f), even though it completely inhibited CD69 staining in the macrophages used in these experiments (data not shown). Overall, these results show that blockade of VnR was sufficient to ablate parasite replication, whereas VnR crosslinkage mimicked the effects of apoptotic cells on *T. cruzi* growth in macrophages.

Interaction with apoptotic cells leads macrophages to secrete TGF-β, which in turn suppresses their pro-inflammatory cytokine response[3]. In addition, TGF-β renders both phagocytic and non-phagocytic cells permissive to *T. cruzi* infection[8,9] and antagonizes interferon (IFN)-γ-induced nitric oxide (NO) production and trypanocidal activity of macrophages[10]. We investigated the role of TGF-β in *T. cruzi* infection in macrophages treated with apoptotic cells. Both uninfected and infected macrophages produced similar levels of TGF-β in response to apoptotic, but not necrotic, cells (Fig. 3a). We also detected marked secretion of TGF-β triggered by anti-α_v monoclonal antibody (Fig. 3b) in the absence of *T. cruzi* infection and apoptotic cells. Anti-α_v Fab promoted TGF-β production only in the presence of a secondary goat anti-hamster IgG (Fig. 3b), but completely ablated TGF-β secretion in response to apoptotic cells (data not shown). We treated infected macrophages with anti-TGF-β1 neutralizing antibody. Neutralization of TGF-β inhibited parasite replication induced by either apoptotic cells or immobilized anti-β_3 monoclonal antibody (Fig. 3c), whereas a control IgY antibody had no effect. Although the involvement of other co-receptors or co-factors[2] cannot be ruled out, the present data indicate that engagement of VnR by apoptotic cells may result in a TGF-β-dependent increase in parasite replication in macrophages. We also investigated whether apoptotic cells and TGF-β interfere with NO-dependent trypanocidal activity of macrophages. Apoptotic cells blocked NO production by IFN-γ/LPS-activated macrophages (Fig. 3d, left) and led to vigorous *T. cruzi* replication (Fig. 3d, right). Inhibition of NO production by apoptotic cells could be reversed by anti-TGF-β antibodies or anti-α_v Fab (Fig. 3e). Therefore, the uptake of apoptotic cells renders macrophages refractory to inflammatory cytokines, allowing parasite survival and growth even in the face of immune response.

Transforming growth factor-β shifts arginine metabolism in macrophages, decreasing NO and inducing ornithine production (by arginase) and subsequent polyamine biosynthesis[11]. Ornithine decarboxylase (ODC) catalyses putrescine synthesis from ornithine, and is considered as the limiting step in the polyamine biosynthetic pathway, leading to spermidine and spermine production[12]. Although ODC activity has not been identified in *T. cruzi*, intra- and extracellular forms of *T. cruzi* synthesize polyamines either through the alternative agmatine pathway[13] or by using exogenously added putrescine[14]. To investigate the possibility that macrophage putrescine synthesis is involved in the apoptotic cell effects that we observed, putrescine content and ODC activity were measured in macrophages treated with apoptotic cells. Macrophages treated with apoptotic cells markedly accumulated putrescine (Fig. 3f, left), and exogenous addition of putrescine alone increased parasite replication in macrophages (Fig. 3f, right). ODC activity was markedly upregulated in macrophages exposed to apoptotic cells or to immobilized anti-α_v monoclonal antibody, with delayed kinetics
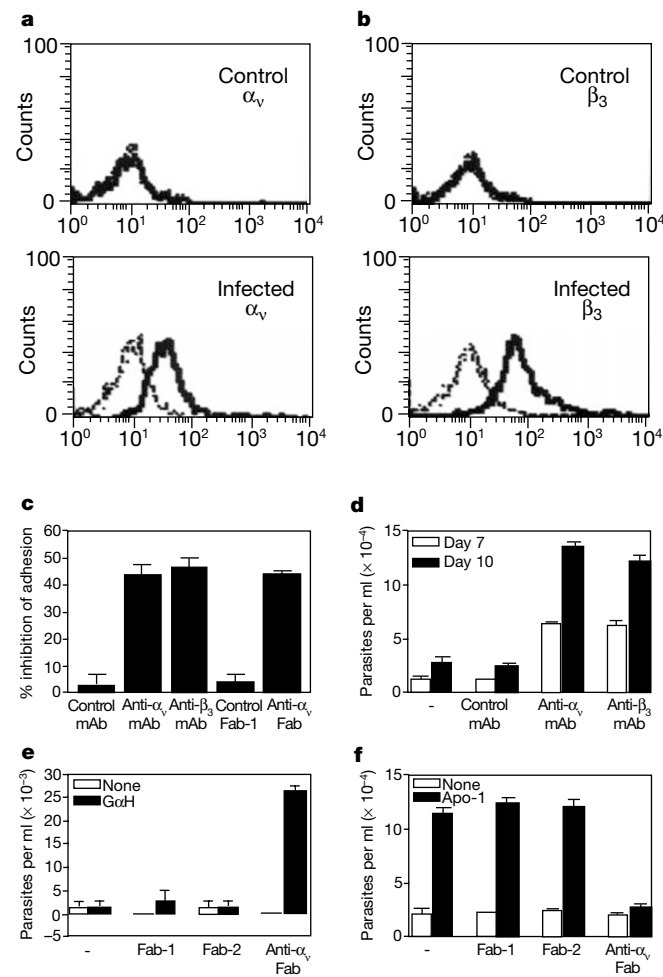
(Fig. 3g). Both anti-α_v Fab and anti-TGF-β antibodies suppressed apoptotic cell effects on ODC activity (Fig. 3h). A competitive ODC inhibitor, α-methylornithine[15], also blocked ODC activity induced by apoptotic cells (Fig. 3g). Addition of this inhibitor to macrophages treated with apoptotic cells or anti-α_v resulted in a dose-dependent decrease in parasite replication (Fig. 3i), but it had no inhibitory effect on basal *T. cruzi* growth in macrophages left without stimuli (Fig. 3i), and did not affect the uptake of apoptotic cells or TGF-β production (data not shown). These results indicate that TGF-β, produced upon engagement of VnR by apoptotic cells, may induce ODC activity in macrophages and promote polyamine-dependent parasite growth.

Searching therapeutic targets to prevent apoptotic cell effects, a role for prostaglandins on *T. cruzi* growth was investigated. Prostaglandin E_2 and PAF have been reported to induce TGF-β production by human macrophages exposed to apoptotic neutrophils[3]; prostaglandin E_2 increases arginase[16] and ODC[17] activities, and promoted *T. cruzi* growth (data not shown) in macrophages. We

therefore evaluated PGE_2 production by macrophages exposed to either apoptotic cells or immobilized anti-α_v monoclonal antibodies. Apoptotic cells induced PGE_2 production by macrophages, and this effect was prevented by anti-α_v Fab but not by control anti-CD69 Fab-1 (Fig. 4a). Vitronectin receptor ligation by immobilized anti-α_v monoclonal antibodies also increased PGE_2 levels, whereas anti-CD69 or anti-Mac-1 (α_M/β_2 integrin) control monoclonal antibodies had no effect (Fig. 4a). We tested three non-steroidal anti-inflammatory drugs (NSAIDs): aspirin, an inhibitor of both constitutive and inducible cyclooxygenases (COX); indomethacin, a preferential antagonist of constitutive COX[18]; and NS-398, an inducible COX inhibitor[19]. Indomethacin almost completely blocked anti-VnR induced PGE_2 (Fig. 4a) and TGF-β (data not shown) production. Indomethacin, aspirin and NS-398 significantly suppressed apoptotic cell effects on PGE_2 (Fig. 4b) and TGF-β (data not shown) secretion, on ODC activity *in vitro* (Fig. 4c) and *in vivo* (Fig. 4d), and on *T. cruzi* growth, without affecting basal parasite replication (Fig. 4e). Furthermore, COX
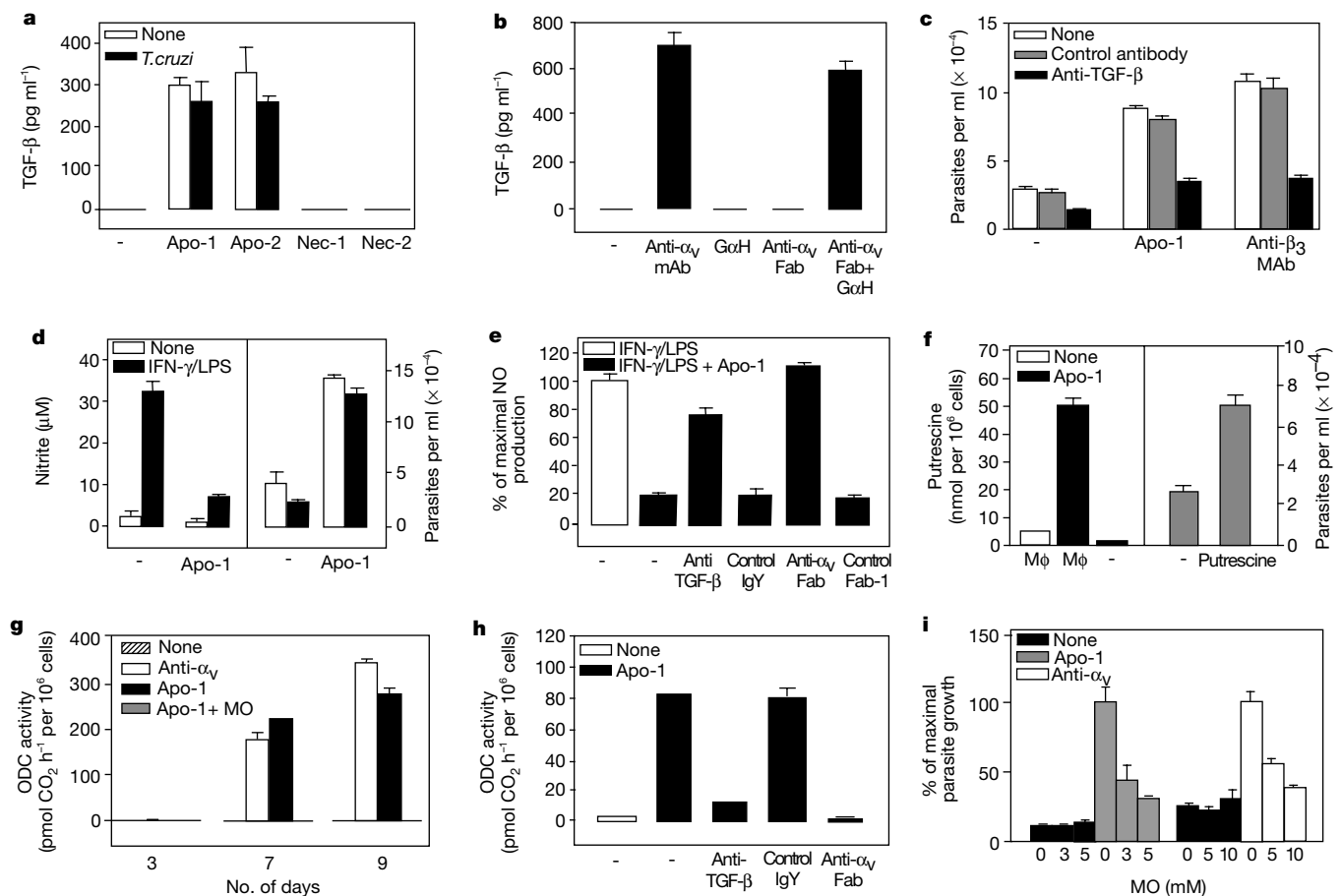

Figure 3 Effects of apoptotic cells on macrophages infected with *T. cruzi* depend on TGF-β. **a**, Apoptotic (Apo-1, -2), but not necrotic (Nec-1, -2), cells induce TGF-β production by uninfected (open bars) or *in vitro* infected (closed bars) macrophages TGF-β1 was evaluated in supernatants from 48-h cultures. **b**, VnR engagement induces TGF-β production by macrophages. Uninfected macrophages were exposed to anti-α_v monoclonal antibody (mAb) or Fab, followed or not by goat anti-hamster IgG (GαH), and TGF-β1 was evaluated in supernatants from 24-h cultures. **c**, Neutralization of TGF-β ablates the effect of apoptotic cells or anti-VnR mAb on macrophages infected with *T. cruzi*. *In vivo* infected macrophages were exposed to Apo-1 or to immobilized anti-β_3 mAb in the presence of either anti-hTGF-β1 (6 μg ml^{-1}) or control IgY antibodies. Trypomastigotes were counted after 10 days. **d**, **e**, Apoptotic cells suppress NO production by LPS/IFN-γ-activated macrophages. *In vivo* infected (**d**) or uninfected (**e**) macrophages were exposed to LPS plus IFN-γ and treated with Apo-1, anti-TGF-β or control IgY antibody, anti-α_v Fab or anti-CD69 Fab-1. Supernatants were assayed for

nitrite content after 48 h (**d**, left, **e**) and for parasites after 10 days (**d**, right). In **e**, results were expressed as a percentage of the maximal nitrite content in macrophages treated with LPS plus IFN-γ. **f**, Apoptotic cells (Apo-1) induce putrescine accumulation in uninfected macrophages (Mϕ) and exogenous putrescine (1 mM) increases parasite growth in macrophages infected *in vitro*. Putrescine (left) and parasite accumulation (right) were measured after 7 days. **g**, Kinetics of ODC induction. Macrophages were exposed to immobilized anti-α_v or to Apo-1. ODC inhibitor MO was added at 10 mM. ODC activity was measured at the indicated days. **h**, Apoptotic cell effects on ODC activity depend on TGF-β. Macrophages were exposed to Apo-1 in the presence of anti-α_v Fab, anti-TGF-β or control IgY antibodies. ODC activity was evaluated after 7 days. **i**, ODC activity is required for *T. cruzi* growth. *In vitro* infected macrophages were exposed to Apo-1 or to anti-α_v mAb, and MO was added at the indicated dosages. Trypomastigotes were counted after 8 (anti-α_v) or 10 (Apo-1) days in culture and are expressed as a percentage of the maximal parasite growth in the absence of MO.
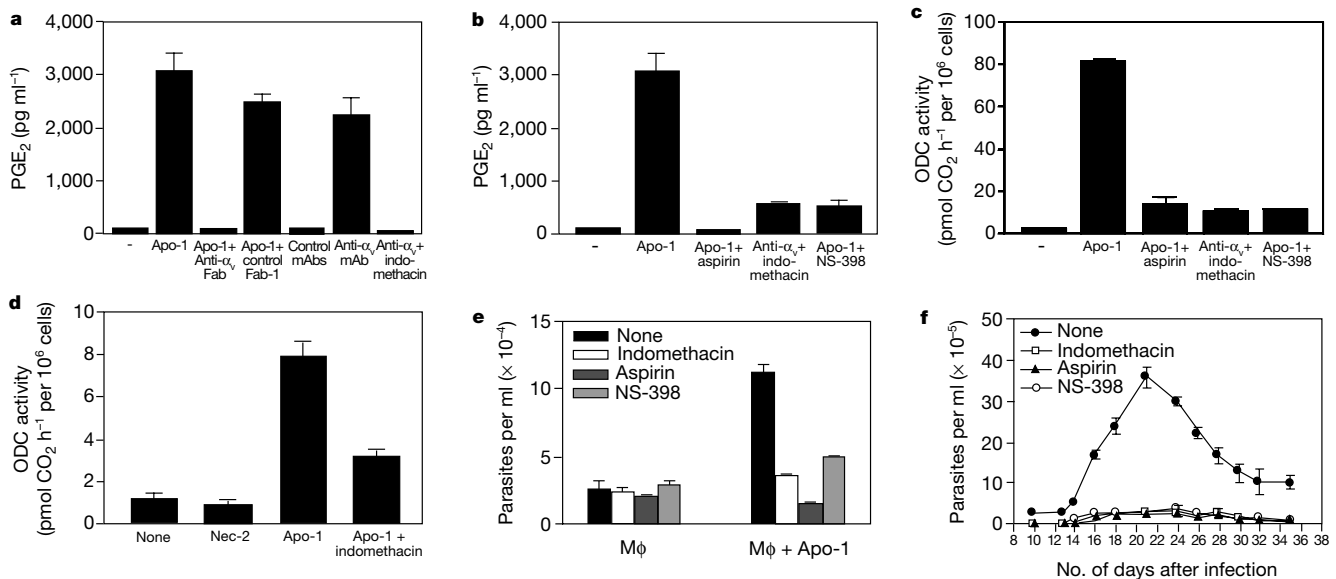
Figure 4 *In vitro* and *in vivo* effects of COX inhibitors on *T. cruzi* infection. **a**, Apoptotic cells or VnR ligation induce PGE$_2$ release by macrophages. Macrophages were treated with Apo-1, anti-α_v Fab or anti-CD69 Fab-1, immobilized anti-α_v, anti-CD69 or anti-Mac-1 mAbs, or anti-α_v plus indomethacin. PGE$_2$ was measured as described. For simplicity, results with anti-CD69 and anti-Mac-1 were combined in the figure (control mAbs). **b**, **c**, COX inhibitors block PGE$_2$ release and ODC activity *in vitro*. Macrophages were exposed to Apo-1 and treated with aspirin, indomethacin or NS-398. PGE$_2$ levels were measured after 24 h (**b**), and ODC activity was assayed after 7 days (**c**). **d**, Apoptotic cells

induce ODC activity *in vivo*. Mice ($n = 3$) were injected i.p. with Nec-2, Apo-1 or Apo-1 plus indomethacin. ODC activity was measured 10 days later in adherent PEC. **e**, COX inhibitors antagonize apoptotic cell effects on macrophages infected with *T. cruzi*. *In vitro* infected macrophages (Mϕ) were exposed to Apo-1 and treated with COX inhibitors as in **b**. Trypomastigotes were counted after 7 days. **f**, COX inhibitors suppress parasitaemia in *T. cruzi* infected mice. Mice ($n = 4$) were infected with *T. cruzi* and treated 7, 8 and 9 days after infection with aspirin, indomethacin or NS-398. Parasitaemia was monitored throughout acute infection.

inhibitors almost completely abolished parasitaemia in infected mice, after a brief 3-day treatment, one week after infection (Fig. 4f). These data indicate that the factors driving *T. cruzi* replication *in vivo* may be the same mediators as are involved in the parasite growth-promoting activity of apoptotic cells *in vitro*. However, these results should be taken with caution, as NSAIDs can have differential effects on macrophages depending on dosage[20]. More experiments are necessary to understand the discrepant effects of high-dose therapy with COX inhibitors seen in other reported studies on *T. cruzi* infection[21].

Our results show that VnR engagement by apoptotic cells triggers PGE$_2$ and TGF-β release by macrophages. The suppressive effect of PGE$_2$/TGF-β on pro-inflammatory cytokine expression[3] and NO production would create the appropriate environment for optimal *T. cruzi* growth within macrophages. Furthermore, the uptake of apoptotic cells reorientates macrophage metabolism towards putrescine production, helping parasite replication. We have shown that widely used drugs, such as aspirin and indomethacin, interfere with this pathway and are able to control parasitaemia in susceptible mice, making them potentially useful for therapy in the acute phase of Chagas' disease. □

Methods

Animals and infection

We infected BALB/c male mice (6 weeks of age) intraperitoneally (i.p.) with 10^4 metacyclic trypomastigote forms of *T. cruzi* clone Dm 28c (ref. 22).

Macrophages

Peritoneal exudate cells (PEC) removed from the peritoneal cavity of uninfected or acutely infected mice (3–4 weeks after infection) were cultured in complete medium (DMEM, 2 mM L-glutamine, 1 mM sodium pyruvate, 10 µg ml^{-1} gentamicin, MEM non-essential amino acids, 10 mM Hepes and 50 µM 2-ME) with 1% Nutridoma-SP (Boehringer Mannheim) or 5% FCS (experiments in Fig. 1a, c) at 3×10^5 cells ml^{-1} on 24-well plates (Corning). Macrophages from normal mice were infected with 1.5×10^6 *T. cruzi* metacyclic forms per well[5]. After 24 h, non internalized parasites were removed and macrophages cultured in complete medium (1 ml) at 37 °C, with 7% CO$_2$ for up to 10 days. Extracellular motile trypomastigotes were counted in culture supernatants[5].

Lymphocytes

Nylon-wool-filtered normal splenocytes were suspended in complete medium and heated (56 °C for 30 min) (Apo-1) or irradiated with 30 Gy (Apo-2) to obtain apoptotic cells as described[23]. Cells were also fixed with 1% paraformaldehyde for 20 min at room temperature (followed by extensive washing) (Nec-1) or frozen–thawed (Nec-2) to obtain necrotic cells[23]. Dead or viable cells (10^6 per well) were then added to macrophages.

Antibodies and reagents

Anti-β_3 integrin (CD61), PE-labelled anti-β_3 and anti-α_v (CD51), control hamster IgG, unlabelled and FITC- and PE-labelled anti-Mac-1, FITC-labelled anti-CD69, and anti-CD16/CD32 monoclonal antibodies were purchased from Pharmingen (San Diego). Purified chicken anti-hTGF-β1 IgY was from R&D Systems (Minneapolis) and control (chicken anti-canine IgG) IgY antibody (a gift from V. Laurentino) was prepared as described[24]. Anti-α_v mAb H9.2B8 (ref. 25), and anti-CD69 mAb H1.2F3 (ref. 26) were donated by E. Shevach. Fab fragments of control hamster IgG, anti-CD69 and anti-α_v monoclonal antibodies were produced with a commercial Fab preparation kit (Pierce). zVAD-fmk peptide was a gift from M. Lenardo. Other reagents were goat anti-hamster IgG antibody (Cappel-Organon Teknika Corporation), mrIFN-γ (Pharmingen), LPS (*Escherichia coli* O111:B4; Difco), α-methil-ornithine, MO (Marion Merrel Dow), Putrescine (Sigma), aspirin and indomethacin (Sigma), and NS-398 (Biomol).

Adhesion assay

Macrophages were cultured with Apo-1 and anti-α_v, anti-β_3 or control hamster IgG mAbs (10 µg ml^{-1}), anti-α_v Fab or control anti-CD69 Fab-1 (50 µg ml^{-1}) for 3 h, washed, fixed with 1% paraformaldehyde, and counted under phase-contrast microscopy. Adhesion was calculated as the percentage of rosetted macrophages out of a total of 100 macrophages counted per well.

TGF-β production

Macrophages were treated with apoptotic or necrotic cells, or with anti-α_v monoclonal antibody or Fab plus goat anti-hamster IgG (10 µg ml^{-1}). The content of TGF-β1 in supernatants was evaluated by sandwich ELISA as described[27].

NO production

Macrophages were cultured with IFN-γ (40 U ml^{-1}) plus LPS (10 ng ml^{-1}), and exposed or not to Apo-1, anti-TGF-β or control IgY antibody (6 µg ml^{-1}), anti-α_v Fab or control anti-CD69 Fab-1. Supernatants were collected 48 h later and mixed with an equal volume of Griess reagent to determine nitrite content, as described[28].

PGE$_2$ release

Macrophages were exposed to immobilized anti-α_v or control monoclonal antibodies (10 µg ml^{-1}), Apo-1, Fab fragments, aspirin (10 µg ml^{-1}), indomethacin (1 µg ml^{-1}) or

NS-398 (1 μM). Culture supernatants (24 h) were collected and assayed for PGE$_2$ using a competitive ELISA kit (Cayman Chemicals).

ODC activity and putrescine content

Macrophages were cultured (10^6 per ml) with Apo-1 (3×10^6 per well), anti-α_v monoclonal antibody, MO or other reagents. Ornithine decarboxydase activity and putrescine content were evaluated as described[29].

Flow cytometry

Peritoneal exudate cells (PEC) were treated with F$_c$ block (anti-CD16/CD32) and stained with FITC-labelled anti-Mac-1 plus PE-labelled anti-α_v or anti-β_3 monoclonal antibodies. Unlabelled anti-α_v or anti-β_3 monoclonal antibodies were used to block staining specifically. Uninfected macrophages were also cultured for 24 h, detached and stained. Antibodies were used at 1 μg per 10^6 cells. 10^4 cells were acquired, and Mac-1-positive cells were gated and analysed for either α_v or β_3 expression on a B-D Xcalibur flow cytometer.

In vivo experiments

Infected mice ($n = 4$) were injected i.p. with 10^7 Apo-1 or Nec-2, 7 days after infection, or left without treatment. Parasitaemia was determined on blood samples from the tail. In other experiments, infected mice ($n = 4$) were injected i.p. at 7, 8 and 9 days after infection, with aspirin (50 mg kg^{-1}), indomethacin (1 mg kg^{-1}) or NS-398 (5 mg kg^{-1}), or left untreated, and parasitaemia was followed. For ex-vivo determination of ODC activity, mice ($n = 3$) were injected i.p. with 10^7 Apo-1 or Nec-2, or left without treatment. Mice injected with Apo-1 were untreated or treated with indomethacin (1 mg kg^{-1}) in the same day, and 4 days later. ODC activity was measured in adherent PEC (10^6 per well) 10 days later.

Presentation of results and statistics

Each experiment presented is representative of at least three independent experiments. Data are expressed as mean $\pm$ s.e. of duplicate determinations. For in vivo experiments, results are expressed as mean $\pm$ s.e. of individual animals. Significance was evaluated by Student's unpaired t-test, and all positive results mentioned were significant ($P < 0.05$ or <0.01) compared with controls.

1. Savill, J. Apoptosis: Phagocytic docking without shocking. *Nature* **392,** 442–443 (1998).
2. Voll, R. E., Herrmann, M., Roth, E. A., Stach, C. & Kalden, J. R. Immunosuppressive effects of apoptotic cells. *Nature* **390,** 350–351 (1997).
3. Fadok, V. A. *et al.* Macrophages that have ingested apoptotic cells in vitro inhibit proinflammatory cytokine production through autocrine/paracrine mechanisms involving TGF-β, PGE$_2$ and PAF. *J. Clin. Invest.* **101,** 890–898 (1998).
4. Lopes, M. F., Veiga, V. F., Santos, A. R., Fonseca, M. E. F. & DosReis, G. A. Activation-induced CD4$^+$ T cell death by apoptosis in experimental Chagas disease. *J. Immunol.* **154,** 744–752 (1995).
5. Nunes, M. P., Andrade, R. M., Lopes, M. F. & DosReis, G. A. Activation-induced T cell death exacerbates *Trypanosoma cruzi* replication in macrophages cocultured with CD4$^+$ T lymphocytes from infected hosts. *J. Immunol.* **160,** 1313–1319 (1998).
6. Savill, J., Dransfield, I., Hogg, N. & Haslett, C. Vitronectin receptor-mediated phagocytosis of cells undergoing apoptosis. *Nature* **343,** 170–173 (1990).
7. Savill, J., Hogg, N., Ren, Y. & Haslett, C. Thrombospondin cooperates with CD36 and the vitronectin receptor in macrophage recognition of neutrophils undergoing apoptosis. *J. Clin. Invest.* **90,** 1513–1522 (1992).
8. Silva, J. S., Twardzik, D. R. & Reed, S. G. Regulation of *Trypanosoma cruzi* infection in vitro and in vivo by transforming growth factor β (TGF-β). *J. Exp. Med.* **174,** 539–545 (1991).
9. Ming, M., Ewen, M. E. & Pereira, M. E. A. Trypanosome invasion of mammalian cells requires activation of the TGF-β signaling pathway. *Cell* **82,** 287–296 (1995).
10. Gazzinelli, R. T., Oswald, I. P., Hieny, S., James, S. L. & Sher, A. The microbicidal activity of interferon-γ-treated macrophages against *Trypanosoma cruzi* involves an L-arginine-dependent, nitrogen oxide-mediated mechanism inhibitable by interleukin-10 and transforming growth factor-β. *Eur. J. Immunol.* **22,** 2501–2506 (1992).
11. Boutard, V. *et al.* Transforming growth factor-β stimulates arginase activity in macrophages: implications for the regulation of macrophage cytoxocity. *J. Immunol.* **155,** 2077–2084 (1995).
12. Pegg, A. E. & McCann, P. P. Polyamine metabolism and function. *Am. J. Physiol.* **243,** C212–221 (1982).
13. Kierszembaum, F., Wirth, J. J., McCann, P. P. & Sjoerdsma, A. Arginine decarboxylase inhibitors reduce the capacity of *Trypanosoma cruzi* to infect and multiply in mammalian host cells. *Proc. Natl Acad. Sci. USA* **84,** 4278–4282 (1987).
14. Hunter, K. J., Le Quesne, S. A. & Fairlamb, A. H. Identification and biosynthesis of N^1,N^9-bis(glutathionyl)aminopropylcadaverine (homotrypanothione) in *Trypanosoma cruzi. Eur. J. Biochem.* **226,** 1019–1027 (1994).
15. Mamont, P. S. *et al.* α-Methyl ornithine, a potent competitive inhibitor of ornithine decarboxylase, blocks proliferation of rat hepatoma cells in culture. *Proc. Natl Acad. Sci. USA* **73,** 1626–1630 (1976).
16. Corraliza, I. M., Modolell, M., Ferber, E. & Soler, G. Involvement of protein kinase A in the induction of arginase in murine bone marrow-derived macrophages. *Biochim. Biophys. Acta* **1334,** 123–128 (1997).
17. Prosser, F. H. & Wahl, L. M. Involvement of the ornithine decarboxylase pathway in macrophage collagenase production. *Arch. Biochem. Biophys.* **260,** 218–225 (1998).
18. Meade, E. A., Smith, W. L. & DeWitt, D. L. Differential inhibition of prostaglandin endoperoxide synthase (cyclooxygenase) isozymes by aspirin and other non-steroidal anti-inflammatory drugs. *J. Biol. Chem.* **268,** 6610–6614 (1993).
19. Futaki, N. *et al.* Selective inhibition of NS-398 on prostanoid production in inflamed tissue in rat carrageenan-air-pouch inflammation. *J. Pharm. Pharmacol.* **45,** 753–755 (1993).
20. Jiang, C., Ting, A. T. & Seed, B. PPAR-γ agonists inhibit production of monocyte inflammatory cytokines. *Nature* **391,** 82–85 (1998).
21. Celentano, A. M. *et al.* PGE$_2$ involvement in experimental infection with *Trypanosoma cruzi* subpopulations. *Prostaglandins* **49,** 141–153 (1995).
22. Contreras, V. T., Salles, J. M., Thomas, N., Morel, C. M. & Goldenberg, S. In vitro differentiation of *Trypanosoma cruzi* under chemically defined conditions. *Mol. Biochem. Parasitol.* **16,** 315–327 (1985).
23. Griffith, T. S., Yu, X., Herndon, J. M., Green, D. R. & Ferguson, T. A. CD95-induced apoptosis of lymphocytes in an immune privileged site induces immunological tolerance. *Immunity* **5,** 7–16 (1996).
24. Sturmer, A. M., Driscoll, D. P. & Jackson-Matthews, D. E. A quantitative immunoassay using chicken antibodies for detection of native and recombinant α-amidating enzyme. *J. Immunol. Methods* **146,** 105–110 (1992).
25. Maxfield, S. R. *et al.* Murine T cells express a cell surface receptor for multiple extracellular matrix proteins. Identification and characterization with monoclonal antibodies. *J. Exp. Med.* **169,** 2173–2190 (1989).
26. Yokoyama, W. M. *et al.* Characterization of a cell surface-expressed disulfide-linked dimer involved in murine T cell activation. *J. Immunol.* **141,** 369–376 (1988).
27. Soares, M. B. P., David, J. R. & Titus, R. G. An in vitro model for infection with *Leishmania major* that mimics the immune response in mice. *Infect. Immun.* **65,** 2837–2845 (1997).
28. Green, L. C. *et al.* Analysis of nitrate, nitrite and [^{15}N] nitrate in biological fluids. *Anal. Biochem.* **126,** 131–138 (1982).
29. De Mello, F. G., Bachrach, U. & Nirenberg, M. Ornithine and glutamic acid decarboxylase activities in the developing chick retina. *J. Neurochem.* **27,** 847–851 (1976).

The major protein import receptor of plastids is essential for chloroplast biogenesis

Jörg Bauer*†, Kunhua Chen†‡, Andreas Hiltbunner*†, Ernst Wehrli§, Monika Eugster*, Danny Schnell‡ & Felix Kessler*

* Institute of Plant Sciences, Swiss Federal Institute of Technology, Universitätstrasse 2, CH-8092 Zürich, Switzerland
‡ Department of Biological Sciences, Rutgers, The State University of New Jersey, 101 Warren Street, Newark, New Jersey 07102, USA
§ Laboratory for Electronmicroscopy I, Institute of Biochemistry, Swiss Federal Institue of Technology, Schmelzbergstrasse 7, CH-8092 Zürich, Switzerland
† These authors contributed equally to this work

Light triggers the developmental programme in plants that leads to the production of photosynthetically active chloroplasts from non-photosynthetic proplastids[1]. During this chloroplast biogenesis, the photosynthetic apparatus is rapidly assembled, mostly from nuclear-encoded imported proteins[2–4], which are synthesized in the cytosol as precursors with cleavable amino-terminal targeting sequences called transit sequences. Protein translocon complexes at the outer (Toc complex)[5–7] and inner (Tic complex)[6,8,9] envelope membranes recognize these transit sequences, leading to the precursors being imported. The Toc complex in the pea consists of three major components, Toc75, Toc34 and Toc159 (formerly termed Toc86)[6,7,10,11]. Toc159, which is an integral membrane GTPase[12], functions as a transit-sequence receptor[5–7,13]. Here we show that *Arabidopsis thaliana* Toc159 (atToc159) is essential for the biogenesis of chloroplasts. In an *Arabidopsis* mutant (ppi2) that lacks atToc159, photosynthetic proteins that are normally abundant are transcriptionally repressed, and are found in much smaller amounts in the plastids, although ppi2 does not affect either the expression or the import

203

Formation of molecular gas in the tidal debris of violent galaxy–galaxy interactions

Jonathan Braine*, Ute Lisenfeld†, Pierre-Alain Duc‡ & Stéphane Leon§

* Observatoire de Bordeaux, UMR 5804, CNRS/INSU, BP89, F-33270 Floirac, France
† Institut de Radioastronomie Millimétrique, Avenida Divina Postora 7, NC18012 Granada, Spain
‡ Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK and CNRS and CEA/DSM/DAPNIA Service d'astrophysique, Saclay, 91191 Gif sur Yvette cedex, France
§ ASIAA, Academia Sinica, PO Box 1-87, Nanking, Taipei 115, Taiwan

In many gravitational interactions between galaxies, gas and stars that have been torn from the precursor galaxies can collect in tidal 'tails'. Star formation begins anew in some of these regions, producing tidal dwarf galaxies[1–4]. Observations of these new galaxies provides insight into processes relevant to galaxy formation more generally, because the timescale of the interaction is well defined. But tracking the star formation process has hitherto been difficult because the tidal dwarf galaxies with young stars showed no evidence of the molecular gas out of which those young stars formed[5–8]. Here we report the discovery of molecular hydrogen (traced by carbon monoxide emission) in two tidal dwarf galaxies. In both cases, the concentration of molecular gas peaks at the same location as the maximum in atomic-hydrogen density, unlike the situation in most gas-rich galaxies. We infer from this that the molecular gas formed from the atomic hydrogen, rather than being torn in molecular form from the interacting galaxies. Star formation in the tidal dwarf galaxies therefore appears to mimic the process in normal spiral galaxies like our own.

Tidal dwarf galaxies (TDGs) are gas-rich irregular galaxies made out of stellar and gaseous material pulled out by tidal forces from the disks of the colliding parent galaxies into the intergalactic medium[9–12]. They are found at the ends of long tidal tails, sometimes 100 kpc from the nuclei of their progenitors, and host active star-forming regions. TDGs contain two main stellar components: young stars recently formed by collapse of expelled atomic hydrogen (H I) clouds, and an older stellar population, at least 1 Gyr old, originally part of the disk of the parent galaxies. Their overall gaseous and stellar properties range between those of classical dwarf irregular galaxies and blue compact dwarf galaxies, with the exception of their metallicity which is higher—typical of the outer disk of a spiral galaxy[12]. Whether a large fraction of dwarf galaxies were formed through tidal encounters in the early Universe—when spiral galaxies were more gaseous and less metal-rich, and collisions more frequent—is an open question, and one of the reasons to study TDGs. One way to answer this question is the dark-matter content of dwarf galaxies. Observations of ordinary dwarf galaxies show that a lot of dark matter, or mass that is in some as yet invisible form, is necessary to account for their rotation velocities. Numerical simulations of gravitational interactions indicate that TDGs should have very little dark matter[13] if the dark matter is, as currently believed, in the form of a large halo and not in, say, a
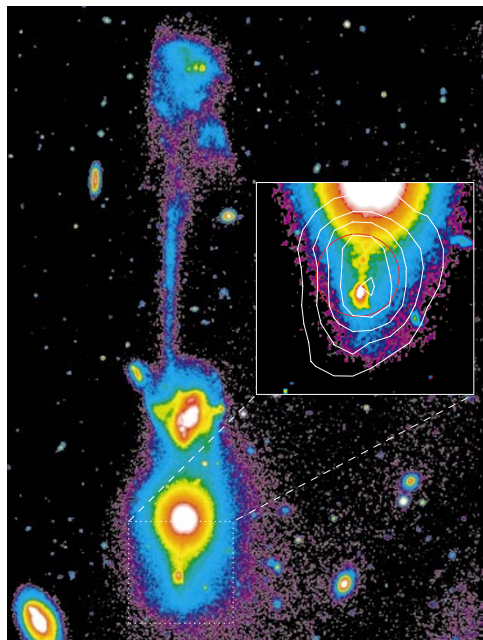

Figure 1 The southern Tidal Dwarf Galaxy (shown in the magnified view) in the interacting system Arp105. (The latter is also known as NGC3561 (ref. 24) and the ''Guitar''). H I emission contours are superposed on a magnified view of a V-band image of A105S (ref. 3). The frame is 4.4′ × 5.9′; north is up and east to the left. Red circle is (0,0) position of CO observations, and represents the full-width at half-maximum FWHM (22″) of the CO(1−0) beam. The Arp105 system[2,3] is an interaction between a spiral and an elliptical which has generated an H I-rich extended TDG at the end of the northern tidal tail and a more compact TDG at the tip of the southern tail from the spiral. A105S contains roughly $5 \times 10^8\,M_\odot$ of H I and strong Hα emission, corresponding to a star-formation rate of $\sim 0.2\,M_\odot\,yr^{-1}$. Nonetheless, stellar synthesis models[25] of A105S indicate that the stellar mass is dominated by the old spiral disk population while the luminosity comes mostly from stars formed *in situ*.
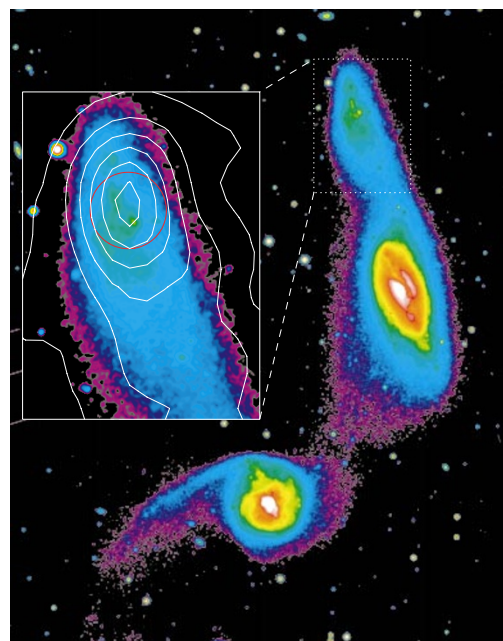

Figure 2 The tidal dwarf galaxy (shown in the magnified view) in the interacting system Arp245 (NGC2992/3[24]). H I emission contours are superposed on magnified view of a V-band image of A245N (ref. 14). The frame is 5.8′ × 7.4′; north is up and east to the left. Red circle is (0,0) position of CO observations and represents the FWHM (22″) of the CO(1−0) beam. Arp245 is an interaction between two spirals. The TDG, A245N, has been formed in the tidal tail which stems from NGC2992 and contains nearly twice as much H I as A105S but slightly weaker Hα emission. The old stellar population is more prominent in A245N than in A105S (ref. 14). The physical size and total H I mass of Arp245 are smaller than in Arp105.

867

Table 1 Properties of the Arp105 and Arp245 TDGs

	A105S	A245N
RA(J200)	11 h 11 min 13.5 s	0.9 h 45 min 44.1 s
Dec(J2000)	$+28°\,41'\,20''$	$-14°\,17'\,28''$
H I velocity (LSR)	$cz = 8,890\,\mathrm{km\,s^{-1}}$	$cz = 2,175\,\mathrm{km\,s^{-1}}$
Adopted distance	115 Mpc	31 Mpc
$L_{H\alpha}$	$(1-2) \times 10^{40}\,\mathrm{erg\,s^{-1}}$	$7 \times 10^{39}\,\mathrm{erg\,s^{-1}}$
M_B, $L_B/L_{B\odot}$	$-16.9,\ 9 \times 10^8$	$-17.25,\ 1.2 \times 10^9$
$B - V$	0.3	0.55
M_{HI}	$5 \times 10^8\,M_\odot$	$9 \times 10^8\,M_\odot$
M_{H_2}	$\geqslant 2.2 \times 10^8\,M_\odot$	$\geqslant 1.4 \times 10^8\,M_\odot$

Data from refs 2 and 3 for A105S, and ref. 14 for A245N. Position is (0,0) of CO map, and velocity is zero of spectra (Fig. 3). M_B and L_B include a correction for galactic absorption of 0.3 mag for A245N. A105S is at high Galactic latitude so no correction is applied. The molecular gas mass is estimated using a $N(H_2)/I_{CO}$ factor of $2 \times 10^{20}\,\mathrm{K\,km\,s^{-1}\,cm^{-2}}$ and probably represents a lower limit because weaker, undetected, CO emission may be present at other positions. We have included the mass of helium in the molecular clouds. Relative to the velocities of the TDGs, the spiral and elliptical in the Arp105 system have velocities of -130 and $-400\,\mathrm{km\,s^{-1}}$. In Arp245, the velocities of the spirals NGC 2992 and NGC 2993 are 155 and $245\,\mathrm{km\,s^{-1}}$ with respect to the TDG. LSR, local standard of rest; $L_{H\alpha}$, luminosity of the H_α line; L_B, luminosity in blue band; $L_{B\odot}$ luminosity in blue band of the Sun; B, blue band magnitude; V, visual band magnitude; M_B, blue band absolute magnitude; M_{HI}, mass of H I gas; M_{H_2}, mass of H_2 gas.

Table 2 Molecular gas in TDGs

Source	Offset (δRA, δdec.)	I_{CO} (K km s^{-1})	Noise (mK)	Vel. (km s^{-1})	ΔV_{FWHM} (km s^{-1})
A105S	(0,0)	0.3 ± 0.05	2.5	19 ± 6	38 ± 10
		0.2 ± 0.05	3.5	16 ± 5	25 ± 10
A105S	(10,0)	0.1 ± 0.5	3	59 ± 6	19 ± 10
A105S	(−10,0)	0.15 ± 0.05	2.8	-4 ± 5	22 ± 13
A105S	Other	0.15 ± 0.05	2.5	15 ± 6	15 ± 8
A245N	(0,0)	1.3 ± 0.1	4.9	-32 ± 4	48 ± 10
		2.0 ± 0.5	18	-32 ± 10	66 ± 20
A245N	(3,14)	0.6 ± 0.15	7	-33 ± 10	47 ± 15
A245N	(0,−10)	0.9 ± 0.15	7.6	-27 ± 7	44 ± 14
A245N	(0,−20)	0.5 ± 0.12	9.5	-27 ± 12	32 ± 15

The offset is in arcsec with respect to the position given in Table 1 and the red circle in Figs 1 and 2. I_{CO} is the flux of the CO line expressed in K km s^{-1}. The lines without source and offset are the CO(2–1) observations of the preceding source and position. Noise levels (r.m.s.) are given for channel widths of 2 mHz in the CO(1−0) line and 2.5 MHz in the CO(2−1) line. The velocity of the line centre is with respect to the H I velocity given in Table 1. Note that the 'detections' of the off-centre A105S positions are uncertain, so the velocities and line widths may be meaningless. The spectra for each position were averaged, a continuum level was then subtracted such that the average flux outside the line window is zero, and resulting spectra were smoothed to yield the results presented in Fig. 3 and Table 2. No baselines other than the continuum level are subtracted from the data. The A105S 'other' position represents the average of the spectra for the (0,5) (0,−5) and (0,−10) positions. Taken individually, these points were not detected and yield 3σ limits of $I_{CO} \leqslant 0.3\,\mathrm{K\,km\,s^{-1}}$. The CO emission from A105S is consistent with a punctual source, much like for the optical and H I. The angular resolutions are respectively 22″ and 11″; beam efficiencies are 0.72 and 0.48.

rotating disk. Thus, if TDGs are found to possess the same dark-matter properties as other dwarf galaxies, then powerful constraints are placed on the form of dark matter. If, on the other hand, TDGs do not contain such dark matter, then tidal interactions cannot be the principal formation mechanism for these small galaxies, nor can dark matter be part of galactic disks.

The observations were performed with the 30-m telescope operated by the Institut de Radioastronomie Millimétrique (IRAM) on Pico Veleta, Spain, in June 1999. Carbon monoxide (CO) emission was detected in the southern TDG in Arp105 (Fig. 1; hereafter A105S) and main TDG in Arp245 (Fig. 2; hereafter A245N) in both the ground state $CO(J = 1-0)$ and the $CO(J = 2-1)$ transitions. Small maps were made of both sources to localize the CO emission with respect to the atomic hydrogen (H I), ionized gas (Hα), and optical continuum[3,14]. The central (0,0) $CO(1-0)$ spectra are shown in Fig. 3 along with the H I spectra at the same positions with a similar beamsize. The $CO(1-0)$ luminosities and derived H_2 mass estimates (see Table 1) of A105S and A245N are far greater than those of other dwarf galaxies[15]. Despite the different environments, the star-formation efficiency, defined as the rate of star formation per mass of molecular gas, is quite close to that observed in the Milky Way and other spiral galaxies[16].

Small CO maps have been made consisting of six positions towards A105S and four positions towards A245N (see Table 2). In both cases, the CO peaks at the H I column density maximum and
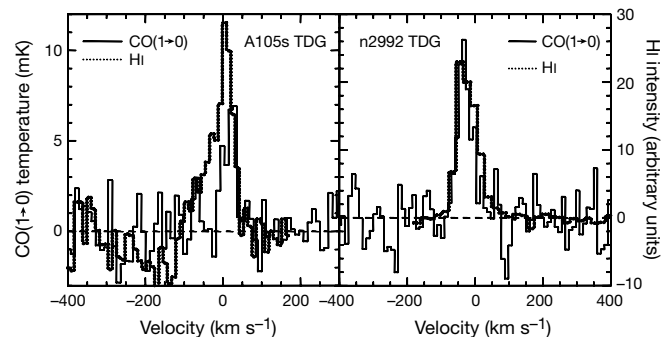

Figure 3 CO(1−0) and H I spectra of the (0,0) position of tidal dwarf galaxies A105S (left) and A245N (right). The velocities and line widths of the CO and H I emission are very similar. Towards the very compact TDG A105S, the CO emission is not resolved. The CO emission in A245N is extended with detections in at least 3 of 4 observed points. The Hα emission in A245N (ref. 14) decreases substantially towards the (0,−10) offset position while the H I (ref. 14) and CO (see Table 2) are still strong. In contrast, the Hα emission towards the (3,14) position is comparable to the centre whereas the CO and H I have decreased significantly. H I intensity is in arbitrary units.

the dynamics of the atomic and molecular components are virtually identical (Fig. 3). In spiral galaxies, on the other hand, H I and CO have very different distributions (see for example, refs 17 and 18), showing that the molecular gas that we have found in the TDGs has not simply been torn off the parent galaxies together with the H I, but rather has formed *in situ*. Although the calculations were performed for post-shock gas, an estimate of the molecule formation time is $t \approx n^{-1}$ Gyr (ref. 19) where n is the density of the atomic medium in particles per cm^3. Numerical simulations of Arp245 (ref. 14) yield an age of about 100 Myr for A245N; a rough age estimate for A105S can be obtained by dividing the projected distance to the spiral by the relative radial velocity, yielding about 200 Myr (ref. 3), sufficient for H_2 formation in standard atomic hydrogen clouds ($\bar{n} \approx 10\,\mathrm{cm^{-3}}$). The dust on which the H_2 forms is captured from the parent galaxies and is present in the atomic gas[20–22]. We conclude that the molecular gas has formed inside the H I clouds, and star formation is proceeding in a standard way from the molecular gas.

Our observations show that the molecular gas is an important component in the visible mass budget of TDGs, between ∼20% and $\geqslant 50\%$ of the atomic hydrogen mass (see Table 1). The fact that we detect large quantities of molecular gas, and that we have every reason to believe that this gas is the result of conversion from H I into H_2, indicates that the central regions of these objects should be gravitationally bound. If the H I were dense enough pre-encounter, then CO would form and be routinely detected beyond the optical radius in galactic disks, like H I—but it is not[17,20,21,23]. While it was clear that TDGs are kinematically decoupled from their parent galaxies, the evidence that TDGs were bound was (until now) morphological—the accumulation of matter at the tips of the tidal tails, and the presence of star-forming regions. Although measurements with higher angular resolution are necessary, our conclusion—that TDGs are bound—provides firmer ground for the calculation of the dynamical mass; this calculation relies on the assumption that the object is gravitationally bound and in equilibrium. A more certain dynamical mass will lead to a more certain determination of dark matter—which (by definition) is detected by a discrepancy between the velocities expected on the basis of the mass of what we see directly and those velocities in fact observed. □

1. Mirabel, I. F., Dottori, H. & Lutz, D. Genesis of a dwarf galaxy from the debris of the Antennae. *Astron. Astrophys.* **256**, L19–L23 (1992).
2. Duc, P.-A. & Mirabel, I. F. Recycled galaxies in the colliding system Arp 105. *Astron. Astrophys.* **289**, 83–93 (1994).

3. Duc, P.-A., Brinks, E., Wink, J. E. & Mirabel, I. F. Gas segregation in the interacting system Arp 105. *Astron. Astrophys.* **326**, 537–553 (1997).

4. Duc, P.-A. & Mirabel, I. G. Young tidal dwarf galaxies around the gas-rich disturbed lenticular NGC 5291. *Astron. Astrophys.* **333**, 813–826 (1998).

5. Brouillet, N., Henkel, C. & Baudry, A. Detection of an intergalactic molecular complex? *Astron. Astrophys.* **262**, L5–L8 (1992).

6. Walter, F. & Heithausen, A. The discovery of a molecular complex in the tidal arms near NGC 3077. *Astrophys. J.* **519**, L69–L72 (1999).

7. Smith, B. J. & Higdon, J. L. A search for CO(1–0) emission from the tidal structures of interacting and merging galaxies. *Astron. J.* **108**, 837–843 (1994).

8. Smith, B. J., Struck, C., Kenney, J. D. P. & Jogee, S. The molecule-rich tail of the peculiar galaxy NGC 2782 (Arp 215). *Astron. J.* **117**, 1237–1248 (1999).

9. Zwicky, F. Multiple galaxies. *Ergenisse Exakten Naturwissenschaften* **29**, 344–385 (1956).

10. *Structure and Properties of Nearby Galaxies* (eds Berkhuijsen, E. M. & Wielebinski, R.) 279–284 (Riedel, Dordrecht, 1978).

11. Hibbard, J. E. & van Gorkom, J. H. HI, HII, and R-Band observations of a galactic merger sequence. *Astron. J.* **111**, 655–695 (1996).

12. Duc, P.-A. & Mirabel, I. F. in *Galaxy Interactions at Low and High Redshift* (eds Barnes, J. & Sanders, D.) 61–70 (IAU Symp. 186, Kluwer, Dordrecht, 1997).

13. Barnes, J. E. & Hernquist, L. Formation of dwarf galaxies in tidal tails. *Nature* **360**, 715–717 (1992).

14. Duc, P.-A. *et al.* The interacting system NGC 2992/3 (Arp 245). *Astron. J.* (submitted).

15. Taylor, C. L., Kobulnicky, H. A. & Skillman, E. D. CO emission in low-luminosity, HI-rich galaxies. *Astron. J.* **116**, 2746–2756 (1998).

16. Kennicutt, R. C. Jr Star formation in galaxies along the Hubble sequence. *Annu. Rev. Astron. Astrophys.* **36**, 189–232 (1998).

17. Guélin, M. *et al.* 1.3 mm emission in the disk of NGC 891: Evidence of cold dust. *Astron. Astrophys.* **279**, L37–L40 (1993).

18. Braine, J., Combes, F. & Van Driel, W. NGC 4414: A flocculent galaxy with a high gas surface density. *Astron. Astrophys.* **280**, 451–467 (1993).

19. Hollenbach, D. & McKee, C. F. Molecule formation and infrared emission in fast interstellar shocks. III—results for J shocks in molecular clouds. *Astrophys. J.* **342**, 306–336 (1989).

20. Neininger, N., Guélin, M., Garcia-Burillo, S., Zylka, R. & Wielebinski, R. Cold dust and molecular line emission in NGC 4565. *Astron. Astrophys.* **310**, 725–736 (1996).

21. Dumke, M. *et al.* the interstellar medium in the edge-on galaxy NGC 5907. Cold dust and molecular line emission. *Astron. Astrohys.* **325**, 124–134 (1997).

22. Braine, J. *et al.* Gas and dust in the active spiral galaxy NGC 3079. *Astron. Astrophys.* **326**, 963–975 (1997).

23. Sage, L. J. The properties and origins of molecular gas in the lenticular galaxies NGC 404, 4710 and 5195. *Astron. Astrophys.* **239**, 125–136 (1990).

24. Arp, H. Atlas of peculiar galaxies. *Astrophys. J. Suppl. Ser.* **14**, 1–20 (1966).

25. Fritze-v.Alvensleben, U. & Duc, P.-A. in *The Magellanic Clouds and other Dwarf Galaxies* (eds Braun, J. M. & Richtler, T.) 141–145 (Proceedings of the Workshop of the Graduiertenkolleg Bonn-Bochum, Shaker, Aachen, 1998).

....

Geometric quantum computation using nuclear magnetic resonance

Jonathan A. Jones*†, **Vlatko Vedral***, **Artur Ekert***
& Giuseppe Castagnoli‡

* *Centre for Quantum Computation, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK*
† *Oxford Centre for Molecular Sciences, New Chemistry Laboratory, South Parks Road, Oxford OX1 3QT, UK*
‡ *Elsag, Via Puccini 2, 1615 Genova, Italy*

....

A significant development in computing has been the discovery[1] that the computational power of quantum computers exceeds that of Turing machines. Central to the experimental realization of quantum information processing is the construction of fault-tolerant quantum logic gates. Their operation requires conditional quantum dynamics, in which one sub-system undergoes a coherent evolution that depends on the quantum state of another sub-system[2]; in particular, the evolving sub-system may acquire a conditional phase shift. Although conventionally dynamic in origin, phase shifts can also be geometric[3,4]. Conditional geometric (or 'Berry') phases depend only on the geometry of the path executed, and are therefore resilient to certain types of errors; this suggests the possibility of an intrinsically fault-

tolerant way of performing quantum gate operations. Nuclear magnetic resonance techniques have already been used to demonstrate both simple quantum information processing[5–9] and geometric phase shifts[10–12]. Here we combine these ideas by performing a nuclear magnetic resonance experiment in which a conditional Berry phase is implemented, demonstrating a controlled phase shift gate.

Any quantum computation can be built out of simple operations involving only one or two quantum bits (qubits)[13]. A particularly simple two-qubit gate in many experimental implementations, such as nuclear magnetic resonance (NMR)[14], is the controlled phase shift. This may be achieved using a conditional Berry phase, and thus quantum geometrical phases can form the basis of quantum computation. We will use spin half nuclei as an example to demonstrate the feasibility of this approach, but the basic idea is general. In our experiments the state of one spin determines the Berry phase acquired by the other spin.

Suppose that a spin half nucleus undergoes a conical evolution with cone angle θ. Then the Berry phase is simply $\gamma = \pm\frac{1}{2}\Omega = \pm\pi(1 - \cos\theta)$ where the $\pm$ signs depend on whether the system is in the eigenstate aligned with or against the field, and Ω is the solid angle subtended by the conical circuit. We note that any deformation of the path of the spin which preserves this solid angle leaves the phase unchanged. Thus the phase is not affected by the speed with which the path is traversed; nor is it very sensitive to random fluctuations about the path.

Berry phases can be conveniently demonstrated in an NMR experiment[15] by working in a rotating frame. We consider an ensemble of spin half particles in a magnetic field, B_0, aligned along the z-axis; their precession frequency is then given by the Larmor frequency, ω_0. If the spins are irradiated by a circularly polarized radio-frequency field, B_1, at a frequency ω_{rf}, the total hamiltonian (neglecting relaxation) may be written in the rotating frame as $H = (\omega_0 - \omega_{rf})I_z + \omega_1 I_x$, where, following conventional NMR practice, the hamiltonian is described in product operator notation[16] and the field strengths are written in terms of their corresponding Larmor frequencies.

When $|\omega_1| \ll |\omega_0 - \omega_{rf}|$ the hamiltonian lies close to the z-axis, while when $|\omega_1| \gg |\omega_0 - \omega_{rf}|$, the hamiltonian lies close to the x-axis. If radio-frequency radiation is applied far from resonance, the system is effectively quantized along the z-axis, and if the radio frequency is swept towards resonance ($\omega_{rf} = \omega_0$), the effective hamiltonian rotates from the z-axis towards the x-axis. If the frequency sweep is applied adiabatically then the spin will follow the hamiltonian. Next, a circular motion can be imposed by adiabatically varying the phase of the radio frequency. When the hamiltonian returns to the x-axis the frequency sweep may be reversed, so that the spin returns to its original state, aligned along the z-axis. The Berry phase acquired in this cyclic process is $\pm\pi$. If the radio-frequency field is not swept all the way to resonance, but only to some final value ω_f, the hamiltonian ends at some angle to the z-axis, and so circuits with arbitrary cone angles can be implemented. A similar case occurs if the frequency sweep is replaced by an amplitude sweep, in which the radio frequency is always applied away from resonance, and its amplitude is raised smoothly from zero to some final value, ω_1.

This situation arises naturally in a system of two weakly coupled spins, I and S. For simplicity we consider a heteronuclear system, so that ω_I and ω_S are very different, and only one spin (say I) is close to resonance. The two transitions of I (corresponding to the two possible states of spin S) will be split by $\pm\pi J$, and so will have different resonance offsets. After an amplitude sweep the orientation of the effective hamiltonian depends on the resonance offset, and so θ (and hence the Berry phase acquired) will depend on the state of spin S. This permits a conditional Berry phase to be applied to spin I, where the size of the phase shift is controlled by spin S. If the radio frequency is applied at a frequency δ (measured in Hz)

869

The DNA sequence of human chromosome 22

I. Dunham, N. Shimizu, B. A. Roe, S. Chissoe *et al.*†

† *A full list of authors appears at the end of this paper*

Knowledge of the complete genomic DNA sequence of an organism allows a systematic approach to defining its genetic components. The genomic sequence provides access to the complete structures of all genes, including those without known function, their control elements, and, by inference, the proteins they encode, as well as all other biologically important sequences. Furthermore, the sequence is a rich and permanent source of information for the design of further biological studies of the organism and for the study of evolution through cross-species sequence comparison. The power of this approach has been amply demonstrated by the determination of the sequences of a number of microbial and model organisms. The next step is to obtain the complete sequence of the entire human genome. Here we report the sequence of the euchromatic part of human chromosome 22. The sequence obtained consists of 12 contiguous segments spanning 33.4 megabases, contains at least 545 genes and 134 pseudogenes, and provides the first view of the complex chromosomal landscapes that will be found in the rest of the genome.

Two alternative approaches have been proposed to determine the human genome sequence. In the clone by clone approach, a map of the genome is constructed using clones of a suitable size (for example, 100–200 kilobases (kb)), and then the sequence is determined for each of a representative set of clones that completely covers the map[1]. Alternatively, a whole genome shotgun[2] requires the sequencing of unmapped genomic clones, typically in a size range of 2–10 kb, followed by a monolithic assembly to produce the entire sequence. Although the merits of these two strategies continue to be debated[3], the public domain human genome sequencing project is following the clone by clone approach[4] because it is modular, allows efficient organization of distributed resources and sequencing capacities, avoids problems arising from distant repeats and results in early completion of significant units of the genome. Here we report the first sequencing landmark of the human genome project, the operationally complete sequence of the euchromatic portion of a human chromosome.

Chromosome 22 is the second smallest of the human autosomes, comprising 1.6–1.8% of the genomic DNA[5]. It is one of five human acrocentric chromosomes, each of which shares substantial sequence similarity in the short arm, which encodes the tandemly repeated ribosomal RNA genes and a series of other tandem repeat sequence arrays. There is no evidence to indicate the presence of any protein coding genes on the short arm of chromosome 22 (22p). In contrast, direct[6] and indirect[7,8] mapping methods suggest that the long arm of the chromosome (22q) is rich in genes compared with other chromosomes. The relatively small size and the existence of a high-resolution framework map of the chromosome[9] suggested to us that sequencing human chromosome 22 would provide an excellent opportunity to show the feasibility of completing the sequence of a substantial unit of the human genome. In addition, alteration of gene dosage on part of 22q is responsible for the aetiology of a number of human congenital anomaly disorders including cat eye syndrome (CES, Mendelian Inheritance in Man (MIM) 115470, http:// www.ncbi.nlm.nih.gov/omim/) and velocardiofacial/DiGeorge syndrome (VCFS, MIM 192430; DGS, MIM 188400). Other regions associated with human disease are the schizophrenia susceptibility locus[10,11], and the sequences involved in spinocerebellar ataxia 10 (SCA10)[12]. Making the sequence of human chromosome 22 freely available to the community early in the data collection phase has benefited studies of disease-related and other genes associated with this human chromosome[13–19].

Genomic sequencing

To identify genomic clones as the substrate for sequencing chromo-
some 22, extensive clone maps of the chromosome were constructed using cosmids, fosmids, bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs). Clones representing parts of chromosome 22 were identified by screening BAC and PAC libraries representing more than 20 genome equivalents using sequence tagged site (STS) markers known to be derived from the chromosome, or by using cosmid and fosmid libraries derived from flow-sorted DNA from chromosome 22. Overlapping clone contigs were assembled on the basis of restriction enzyme fingerprints and STS-content data, and ordered relative to each other using the established framework map of the chromosome[9]. The resulting nascent contigs were extended and joined by iterative cycles of chromosome walking using sequences from the end of each contig. In two places, yeast artificial chromosome (YAC) clones were used to join or extend contigs (AL049708, AL049760). The sequence-ready map covers 22q in 11 clone contigs with 10 gaps and stretches from sequences containing known chromosome 22 centromeric tandem repeats to the 22q telomere[20].

In the final sequence, one additional gap that was intractable to sequencing is found 234 kb from the centromere (see below). The gaps between the clone contigs are located at the two ends of the map, in the 4.3 Mb adjacent to the centromere and in 7.3 Mb at the telomeric end. These regions are separated by a central contig of 23 Mb. We have concluded that the gaps contain sequences that are unclonable with the available host-vector systems, as we were unable to detect clones containing the sequences in these gaps by screening more than 20 genome equivalents of bacterial clones using sequences adjacent to the contig ends.

The size of the seven gaps in the telomeric region has been estimated by DNA fibre fluorescence *in situ* hybridization (FISH). No gap in this region is judged to be larger than ~150 kb. For three of these gaps, a number of BAC and PAC clones that contain STSs on either side of the gap were shown to be deleted for at least a minimal core region by DNA fibre FISH. As these clones come from multiple donor DNA sources, these results are unlikely to be due to deletion in the DNA used to make the libraries. Furthermore, the same result was observed for the gap at 32,600 kb from the centromeric end of the sequence, when the DNA fibre FISH experiments were performed on DNA from two different lymphoblastoid cell lines. One possible explanation for this observation is that DNA fragments containing the gap sequences are initially cloned in the BAC library but clones that delete these sequences have a significant selective advantage as the library is propagated. As the observed size range of the cloned inserts in the BAC libraries ranges from 100 kb to more than 230 kb (http://bacpac.med.buffalo.edu/), such deletion events

are not distinguishable on the basis of size from undeleted BACs. Additional analysis of the distribution of BAC end sequences from dbGSS (http://www.ncbi.nlm.nih.gov/dbGSS/index.html) suggests that the BAC coverage is sparser closer to the gaps and that this analysis did not identify any BACs spanning the gaps. The three remaining clone-map gaps in the proximal region of the long arm are in regions that may contain segments of previously characterized low-copy repeats[21]. These gaps could not be sized by DNA fibre FISH because of the extensive intra- and interchromosomal repeat sequences (see below) but were amenable to long-range restriction mapping. The gap between AP000529 and AP000530 was estimated to be shorter than 150 kb by comparison with a previously established long-range restriction map[22]. The gap closest to the centromere, which is less than 2 kb in size, could not be sequenced despite BAC clone coverage as it was unrepresented in plasmid or M13 libraries, and was intractable to all sequencing strategies applied. Detailed descriptions of several of the clone contigs have been published[21,23,24] or will be published elsewhere.

Each sequencing group took responsibility for completion of adjacent areas of the sequence as illustrated in Fig. 1. A set of minimally overlapping clones (the 'tile path') was chosen from the physical map and sequenced using a combination of a random shotgun assembly, followed by directed sequencing to close gaps and resolve ambiguities ('finishing'). The major problems encountered during completion of the sequence in the directed sequencing phase were CpG islands, tandem repeats and apparent cloning biases. Directed sequencing using oligonucleotide primers, very short insert plasmid libraries, or identification of bridging clones by screening high complexity plasmid or M13 libraries solved these problems.

The completed sequence covers 33.4 Mb of 22q with 11 gaps and has been estimated to be accurate to less than 1 error in 50,000 bases, by internal and external checking exercises[25]. The order and size of each of the contiguous pieces of sequence is detailed in Table 1. The largest contiguous segment stretches over 23 Mb. From our gap-size estimates, we calculate that we have completed 33,464 kb of a total region spanning 34,491 kb and that therefore the sequence is complete to 97% coverage of 22q. The complete sequence and analysis is available on the internet (http://www.sanger.ac.uk/HGP/Chr22 and http://www.genome.ou.edu/Chr22.html).

Sequence analysis and gene content

Analysis of the genomic sequence of the model organisms has made extensive use of predictive computational analysis to identify genes[26–28]. In human DNA, identification of genes by these methods is more difficult because of extensive splicing, lower density of exons and the high proportion of interspersed repetitive sequences. The accuracy of ab initio gene prediction on vertebrate genomic sequence has been difficult to determine because of the lack of sequence that has been completely annotated by experiment. To determine the degree of overprediction made by such algorithms, all genes within a region need to be experimentally identified and annotated, however it is virtually impossible to know when this job is complete. A 1.4-Mb region of human genomic sequence around the BRCA2 locus has been subjected to extensive experimental investigation, and it is believed that the 170 exons identified is close to the total number expressed in the region.

The most recent calibration of ab initio methods against this region (R.B.S.K. and T.H., manuscript in preparation) shows that with the best methods[29,30] more than 30% of exon predictions do not overlap any experimental exons, in other words, they are overpredictions. Furthermore, having now applied this analysis to larger amounts of data (more than 15 Mb from the Sanger Annotated Genome Sequence Repository which can be obtained as part of the Genesafe collection (http://www.hgmp.mrc.ac.uk/Genesafe/)), it is confirmed that prediction accuracy also varies considerably between different regions of sequence. It was hoped that these calibration efforts would lead to rules for reliable gene prediction

based on ab initio methods alone, perhaps on the basis of combining several different methods, GC content and so on. However, so far this has not been possible. The same analysis also shows that although ~95% of genes are at least partially predicted by ab initio methods, few gene structures are completely correct (none in BRCA2) and more than 20% of experimental exons are not predicted at all. The comparison of ab initio predictions and the annotated gene structures (see below) in the chromosome 22 sequence is consistent with this, with 94% of annotated genes at least partially detected by a Genscan gene prediction, but only 20% of annotated genes having all exons predicted exactly. Sixteen per cent of all the exons in annotated genes were not predicted at all, although this is only 10% for internal exons (that is, not 5′ and 3′ ends). As a result, we do not consider that ab initio gene prediction software can currently be used directly to reliably annotate genes in human sequence, although it is useful when combined with other evidence (see below), for example, to define splice-site boundaries, and as a starting point for experimental studies.

Fortunately, a vast resource of experimental data on human genes in the form of complementary DNA and protein sequences and expressed sequence tags (ESTs) is available which can be used to identify genes within genomic DNA. Furthermore about 60% of human genes have distinctive CpG island sequences at their 5′ ends[31] which can also be used to identify potential genes. Thus, the approach we have taken to annotating genes in the chromosome 22 sequence relies on a combination of similarity searches against all available DNA and protein databases, as well as a series of ab initio predictions. Upon completion of the sequence of each clone in the tile path, the sequence was subjected to extensive computational analysis using a suite of similarity searches and prediction tools. Briefly, the sequences were analysed for repetitive sequence content, and the repeats were masked using RepeatMasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html). Masked sequence was compared to public domain DNA and protein databases by similarity searches using the blast family of programs[32]. Unmasked sequence was analysed for C + G content and used to predict the presence of CpG islands, tandem repeat sequences, tRNA genes and exons. The completed analysis was assembled into contigs and visualized using implementations of ACEDB (http://www.sanger.ac.uk/Software/Acedb/). In addition, the contiguous masked sequence was analysed using gene prediction software[29,30].

Figure 1 The sequence of human chromosome 22. Coloured boxes depict the annotated features of the sequence of human chromosome 22, with the centromere to the left and the telomere to the right. Coordinates are in kilobases. Vertical yellow blocks indicate the positions of the gaps in the sequence and are proportional in size to the estimated size of each gap. From bottom to top the following features are displayed: positions of interspersed repetitive sequences including tandem repeats categorized by nucleotide repeat unit length (at this resolutrion Alu repeats are not visibly separated in some regions); the positions of the microsatellite markers in the genetic map of Dib et al.[36]; the tiling path of genomic clones used to determine the sequence labelled by their Genbank/EMBL/DDBJ accession number and coloured according to the source of the sequence; and the annotated gene, pseudogene and CpG island content of the sequence. Transcripts and pseudogenes oriented 5′ to 3′ on the DNA strand from centromere to telomere are designated '+', those on the opposite strand '−'. In the transcript rows, the annotated genes are subdivided by colour according to the criteria in the text. Annotated genes with approved gene symbols from the HUGO nomenclature committee are labelled. For details of all the genes with their positions in the reference sequence, see Supplementary Information, Table 1. In the case of the immunoglobulin variable region, the entire locus has been drawn as a single block; in reality, this is a complex of variable chain genes (see ref. 27). At the top is a graphical plot of the repeat density for the common interspersed repeats Alu and Line1, and the C + G base frequency across the sequence. Each is calculated as a percentage of the sequence using a sliding 100-kb window moved in 50-kb iterations. Since the production of Fig. 1, six accession codes have been updated. The new codes are AL050347 (for Z73987), AL096754 (for Z68686), AL049749 (for Z82197), Z75892 (for Z75891), AL078611 (for Z79997) and AL023733 (for AL023593).

Gene features were identified by a combination of human inspection and software procedures. Figure 1 shows the 679 gene sequences annotated across 22q. They were grouped according to the evidence that was used to identify them as follows: genes identical to known human gene or protein sequences, referred to as 'known genes' (247); genes homologous, or containing a region of similarity, to gene or protein sequences from human or other species, referred to as 'related genes' (150); sequences homologous to only ESTs, referred to as 'predicted genes' (148); and sequences homologous to a known gene or protein, but with a disrupted open reading frame, referred to 'pseudogenes' (134). (See Supplementary Information, Table 1, for details of these genes.) The *ab initio* gene prediction program, Genscan, predicted 817 genes (6,684 exons) in the contiguous sequence, of which 325 do not form part of the annotated genes categorized above. Given the calibration of *ab initio* prediction methods discussed above, we estimate that of the order of 100 of these will represent parts of 'real' genes for which there is currently no supporting evidence in any sequence database, and that the remainder are likely to be false positives.

The total length of the sequence occupied by the annotated genes, including their introns, is 13.0 Mb (39% of the total sequence). Of this, only 204 kb contain pseudogenes. About 3% of the total sequence is occupied by the exons of these annotated genes. This contrasts sharply with the 41.9% of the sequence that represents tandem and interspersed repeat sequences. There is no significant bias towards genes encoded on one strand at the 5% level ($\chi^2 = 3.83$).

A striking feature of the genes detected is their variety in terms of both identity and structure. There are several gene families that appear to have arisen by tandem duplication. The immunoglobulin λ locus is a well-known example, but there also are other immunoglobulin-related genes on the chromosome outside the immunoglobulin λ region. These include the three genes of the immunoglobulin λ-like (IGLL) family plus a fourth possible member of the family (AC007050.7). There are five clustered immunoglobulin κ variable region pseudogenes in AC006548, and an immunoglobulin variable-related sequence (VpreB3) in AP000348. Much further away from the λ genes is a variable region pseudogene, ~123 kb telomeric of IGLL3 in sequence AL008721 (coordinates ~9,420–9,530 kb from the centromeric

end of the sequence), and a cluster of two λ constant region pseudogenes and a variable region pseudogene in sequences AL008723/AL021937 (coordinates ~16,060–16,390 kb from the centromeric end).

Human chromosome 22 also contains other duplicated gene families that encode glutathione *S*-transferases, Ret-finger-like proteins[19], phorbolins or APOBECs, apolipoproteins and β-crystallins. In addition, there are families of genes that are interspersed among other genes and distributed over large chromosomal regions. The γ-glutamyl transferase genes represent a family that appears to have been duplicated in tandem along with other gene families, for instance the BCR-like genes, that span the 22q11 region and together form the well-known LCR22 (low-copy repeat 22) repeats (see below).

The size of individual genes encoded on this chromosome varies over a wide range. The analysis is incomplete as not all 5′ ends have been defined. However, the smallest complete genes are only of the order of 1 kb in length (for example, HMG1L10 is 1.13 kb), whereas the largest single gene (LARGE[15]) stretches over 583 kb. The mean genomic size of the genes is 19.2 kb (median 3.7 kb). Some complete gene structures appear to contain only single exons, whereas the largest number of exons in a gene (PIK4CA) is 54. The mean exon number is 5.4 (median 3). The mean exon size is 266 bp (median 135 bp). The smallest complete exon we have identified is 8 bp in the PITPNB gene. The largest single exon is 7.6 kb in the PKDREJ, which is an intron-less gene with a 6.7-kb open reading frame. In addition, two genes occur within the introns of other expressed genes. The 61-kb TIMP3 gene, which is involved in Sorsby fundus macular degeneration, lies within a 268-kb intron of the large SYN3 gene, and the 8.5 kb HCF2 gene lies within a 27.5-b intron of the PIK4CA gene. In each case, the genes within genes are oriented in the opposite transcriptional orientation to the outer gene. We also observe pseudogenes frequently lying within the introns of other functional genes.

Peptide sequences for the 482 annotated full-length and partial genes with an open reading frame of greater than or equal to 50 amino acids were analysed against the protein family (PFAM)[33], Prosite[34] and SWISS-PROT[35] databases. These data were processed and displayed in an implementation of ACEDB. Overall, 240 (50%) predicted proteins had matching domains in the PFAM database encompassing a total of 164 different PFAM domains. Of the residues making up these 482 proteins, 25% were part of a PFAM domain. This compares with PFAM's residue coverage of SWISS-PROT/TrEMBL, which is more than 45% and indicates that the human genome is enriched in new protein sequences. Sixty-two PFAM domains were found to match more than one protein, including ten predicted proteins containing the eukaryotic protein kinase domain (PF00069), nine matching the Src homology domain 3 (PF00018) and eight matching the RhoGAP domain (PF00620). Fourteen predicted proteins contain zinc-finger domains (See

Table 1 Sequence contigs on chromosome 22

Contig*	Size (kb)	
AP000522–AP000529	234	
gap		**1.9**
AP000530–AP000542	406	
gap		**~150**
AP000543–AC006285	1,394	
gap		**~150**
AC008101–AC007663	1,790	
gap		**~100**
AC007731–AL049708	23,006	
gap		**~50**
AL118498–AL022339	767	
gap†		**~50–100**
Z85994–AL049811	1,528	
gap		**~150**
AL049853–AL096853	2,485	
gap‡		**~50**
AL096843–AL078607	190	
gap†		**~100**
AL078613–AL117328	993	
gap		**~100**
AL080240–AL022328	291	
gap†		**~100**
AL096767–AC002055	380	
Total sequence length	33,464	
Total length of 22q	34,491	

* Contigs are indicated by the first and last sequence in the orientation centromere to telomere, and are named by their Genbank/EMBL/DDBJ accession numbers.
† These gaps are spanned by BAC and/or PAC clones with deletions.
‡ This gap shows a complex duplication of AL096853 in DNA fibre FISH.

Table 2 The interspersed repeat content of human chromosome 22

Repeat type	Total number	Coverage (bp)	Coverage (%)
Alu	20,188	5,621,998	16.80
HERV	255	160,697	0.48
Line1	8,043	3,256,913	9.73
Line2	6,381	1,273,571	3.81
LTR	848	256,412	0.77
MER	3,757	763,390	2.28
MIR	8,426	1,063,419	3.18
MLT	2,483	605,813	1.81
THE	304	93,159	0.28
Other	2,313	625,562	1.87
Dinucleotide	1,775	133,765	0.40
Trinucleotide	166	18,410	0.06
Quadranucleotide	404	47,691	0.14
Pentanucleotide	16	1,612	0.0048
Other tandem	305	102,245	0.31
Total	55,664	14,024,657	41.91

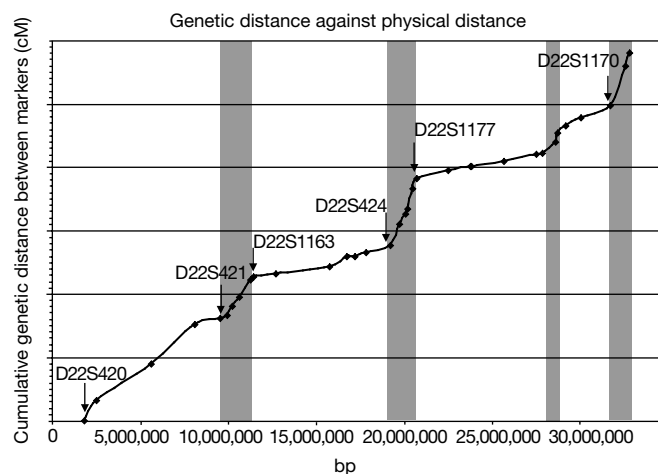**Figure 2** The relationship between physical and genetic distance. The sex-averaged genetic distances of Dib *et al.*[36] were obtained from ftp://ftp.genethon.fr/pub/Gmap/ Nature-1995/ and the cumulative intermarker distances for unambiguously ordered markers (in cM) were plotted against the positions of the microsatellite markers in the genomic sequence. It should be stressed that the *y* axis does not represent the true genetic distance between distant markers but the sum of the local intermarker distances. The positions of selected genetic markers are labelled. Grey regions are indicative of areas of relatively increased recombination per unit physical distance.

Supplementary Information, Table 2, for details of the PFAM domains identified in the predicted proteins).

Nineteen per cent of the coding sequences identified were designated as pseudogenes because they had significant similarity to known genes or proteins but had disrupted protein coding reading frames. Because 82% of the pseudogenes contained single blocks of homology and lacked the characteristic intron–exon structure of the putative parent gene, they probably are processed pseudogenes. Of the remaining spliced pseudogenes, most represent segments of duplicated gene families such as the immunoglobulin *κ* variable genes, the β-crystallins, CYP2D7 and CYP2D8, and the GGT and BCR genes. The pseudogenes are distributed over the entire sequence, interspersed with and sometimes occurring within the introns of annotated expressed genes. However, there also is a dense cluster of 26 pseudogenes in the 1.5-Mb region immediately adjacent to the centromere; the significance of this cluster is currently unclear.

Given that the sequence of 33.4 Mb of chromosome 22q represents 1.1% of the genome and encodes 679 genes, then, if the distribution of genes on the other chromosomes is similar, the minimum number of genes in the entire human genome would be at least 61,000. Previous work has suggested that chromosome 22 is gene rich[6] by a factor of 1.38 (http://www.ncbi.nlm.nih.gov/ genemap/page.cgi?F=GeneDistrib.html), which would reduce this estimate to ~45,000 genes. It is important, however, to recognize that the analysis described here only provides a minimum estimate for the gene content of chromosome 22q, and that further studies will probably reveal additional coding sequences that could not be identified with the current approaches.

Two lines of evidence point to the existence of additional genes that are not detected in this analysis. First, the 553 predicted CpG islands, which typically lie at the true 5′ ends of about 60% of human genes[31], are in excess of 60% of the number of genes identified (60% = 327, excluding pseudogenes); 282 of the genes identified have CpG islands at or close to the 5′ end (within 5-kb upstream of the first exon, or 1-kb downstream). Thus, there could be up to 271 additional genes associated with CpG islands undetected in the sequence. Second, there are 325 putative genes predicted by the *ab initio* gene prediction program, Genscan, that are not in regions already containing annotated transcripts. We
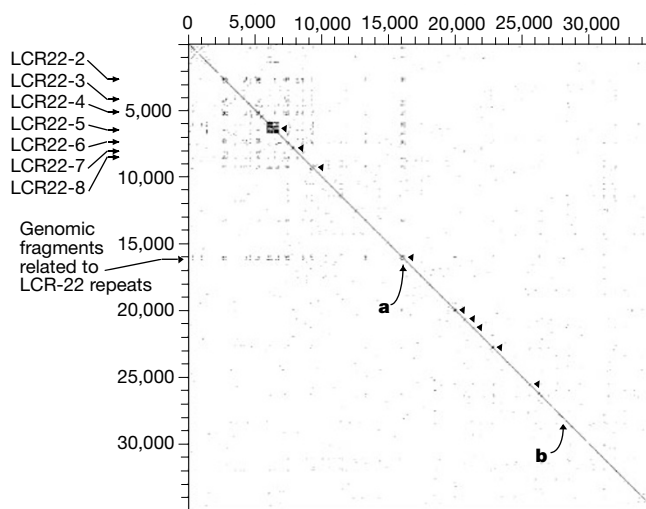

Figure 3 Intrachromosomal repeats on human chromosome 22. High- and medium-copy repeats and low complexity sequence were masked using RepeatMasker and Dust, and masked sequences were compared using Blastn. The results were filtered to identify regions of more than 50% identity to the query sequence, and were plotted in a 2D matrix with a line proportional to the size of the region of identity. Localized gene family repeats are indicated by arrowheads along the diagonal. From the top, these are the immunoglobulin λ locus, the glutathione *S*-transferase genes, the β-crystallin genes, the Ret-finger-protein-like genes, the apolipoprotein genes, the colony-stimulating factor receptor (CSF2RB) inverted partial duplication, the lectins LGALS1 and LGALS2, the APOBEC genes and the CYP2D genes. Two 60-kb regions of more than 90% homology are labelled 'a' (AL008723/AL021937) and 'b' (AL031595/AL022339). Seven low-copy repeat regions (LCR22) and a region containing related genomic fragments are indicated at the left margin.

estimate (see above) that roughly 100 of these will represent parts of real genes. Identifying additional genes will require further computational and experimental studies. These studies are continuing and entail testing candidate sequences for possible messenger RNA expression, implementing new gene prediction software able to detect the regions around or near CpG islands that currently have no identified transcript, and further analysis of sequences that are conserved between human and mouse. Furthermore, full-length cDNA sequences that accumulate in the sequence databases of human and other species will be used to refine the gene structures.

The long-range chromosome landscape

Critical to the utility of the genomic sequence to genetic studies is the integration of established genetic maps. The positions of the commonly used microsatellite markers from the Genethon genetic map[36] are given in Fig. 1. The correlation of the order of markers between the genetic map and the sequence is good, within the limitations of genetic mapping. Only a single marker (D22S1175) is discrepant between the two data sets, and this lies in a sequence that is repeated twice on the chromosome (AL021937, see below). In the telomeric region, four of the Genethon markers must lie in our sequence gaps, and we were unable to identify clones from all libraries tested for these. Comparison of genetic distance against physical distance for all the microsatellites whose order is maintained between the datasets shows a mean value of 1.87 cM Mb⁻¹. However, the relationship between genetic and physical distance across the chromosome partitions into two types of region, areas of high and low recombination (Fig. 2). The areas of high recombination may represent recombinational hot spots, although we have not yet been able to identify any specific sequence characteristics common to these areas.

The mean G + C content of the sequence is 47.8%. This is significantly higher than the G + C content calculated for the
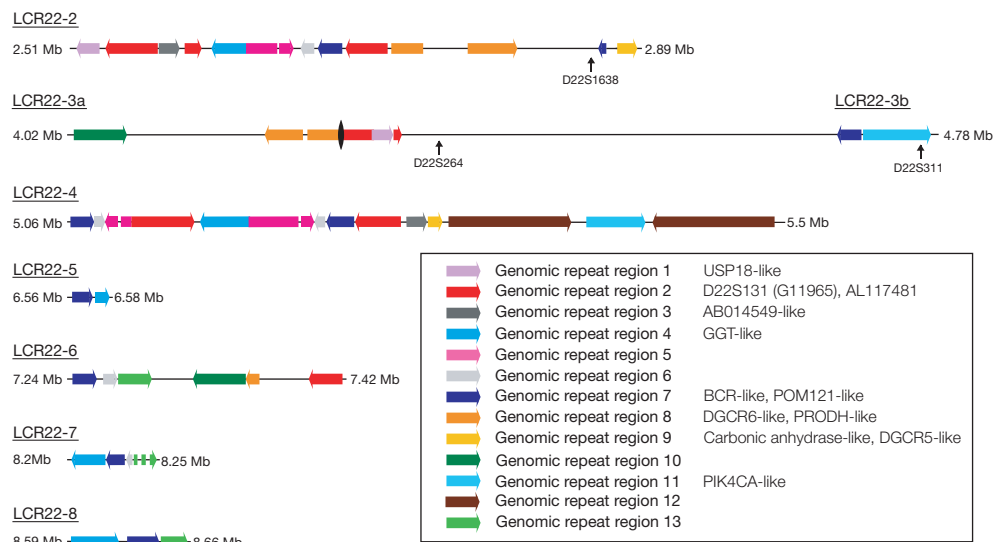
Figure 4 Sequence composition of the LCR22 repeats. Illustration of the sequence composition of seven LCR22 repeats. The span of each LCR22 region is shown in megabases from the centromere. Coloured arrows indicate the extent of one of the thirteen genomic repeat regions and the orientation of the repeat. The known gene and marker content of these genomic repeat regions is indicated in the key. The black oval indicates the position of the gap in the sequence in LCR22-3.

sum of all human genomic sequence determined so far (42%). Although this result was expected from previous indirect measurements of the G + C content of chromosome 22[7,8,37], the distribution is not uniform, but regionally segmented as illustrated in Fig. 1. There are clear fluctuations in the base content, resulting in areas that are relatively G + C rich and others that are relatively G + C poor. On chromosome 22 these regions stretch over several megabases. For example, the 2 Mb of sequence closest to the centromeric end of the sequence is relatively G + C poor, with the G + C content dropping below 40%. Similarly, the area between 16,000 and 18,800 kb from the centromeric end of the sequence is consistently below 45% G + C. The G + C rich regions often reach more than 55% G + C (for example, at 20,100–23,400 kb from the centromeric end of the sequence). This fluctuation appears to be consistent with previous observations that vertebrate genomes are segmented into 'isochores' of distinct G + C content[38] and is similar to the structure seen in the human major histocompatibility complex (MHC) sequence[39]. Isochores correlate with both genes and chromosome structure. The G + C rich isochores are rich in genes and *Alu* repeats, and are located in the G + C rich chromosomal R-bands, whereas the G + C poor isochores are relatively depleted in genes and *Alu* repeats, and are located in the G-bands[8,37,40]. The G + C poor regions of chromosome 22 are depleted in genes and relatively poor in *Alu* sequences. For example, the region between 16,000 and 18,800 kb from the centromeric end contains just three genes, two of which are greater than 400 kb in length. The G + C poor regions also are depleted in CpG islands, which are clustered in the gene-rich, G + C rich regions. Although it is tempting to correlate the sequence features that we see with the chromosome banding patterns, we believe that high-resolution mapping of the chromosome band boundaries will be required to assign definitively these to genomic sequence.

Over 41.9% of the chromosome 22 sequence comprises interspersed and tandem repeat family sequences (Table 2). The density of repeats across the sequence is plotted in Fig. 1. There is variation in the density of *Alu* repeats and some of the regions with low *Alu* density correlate with the G + C poor regions, for example, in the region 16,000–18,800 kb from the centromeric end, and these data support the relationship of isochores with *Alu* distribution. However, in other areas the relationship is less clear. We provide a World-Wide Web interface to the long-range analyses presented here and to further analysis of the many other repeat types and features of the sequence at http://www.sanger.ac.uk/cgi-bin/cwa/22cwa.pl. The 1-Mb region closest to the centromere contains several interesting repeat sequence features that may be typical of other pericentromeric regions. In addition to the density of pseudogenes described above, there is a large 120-kb block of tandemly repeated satellite sequence (D22Z3) centred 500 kb from the centromeric sequence start (not shown in Fig. 1, but evident from the absence of *Alu* and LINE1 sequences at this point). There is also a cluster of satellite II repeats 80-kb telomeric of the D22Z3 sequences. Isolated alphoid satellite repeats are found closer to the centromeric end of the sequence. Furthermore, this pericentromeric 1 Mb closest to the centromere contains many sequences that are shared with a number of different chromosomes, particularly chromosomes 2 and 14. During map construction, 33 out of 37 STSs designed from sequence that was free of high-copy repeats amplified from more than one chromosome in somatic cell hybrid panel analysis.

Low-copy repeats on chromosome 22

To detect intra- and interchromosomal repeats, we compared the entire sequence of chromosome 22 to itself, and also to all other existing human genomic DNA sequence using Blastn[32] after masking high and medium frequency repeats. The results of the intrachromosomal sequence analysis were plotted as a dot matrix (Fig. 3) and reveal a series of interesting features. Locally duplicated gene families lie close to the diagonal axis of the plot. The most striking is the immunoglobulin λ locus that comprises a cluster of 36 potentially functional V-λ gene segments, 56 V-λ pseudogenes, and 27 partial V-λ pseudogenes ('relics'), together with 7 each of the J and C λ segments[24]. Other duplicated gene families that are visible from the dot matrix plot include the clustered genes for glutathione *S*-transferases, β-crystallins, apolipoproteins, phorbolins or APO-BECs, the lectins LGALS1 and LGALS2 and the CYP2Ds. A partial inverted duplication of CSF2RB is also observed.

Much more striking are the long-range duplications, which are visible away from the diagonal axis. For example, a 60-kb segment of more than 90% similarity is seen between sequences AL008723/AL021937 (at ~16,060–16,390 kb from the centromeric end) and AL031595/AL022339 (at ~27,970–28,110 kb from the centromeric end) separated by almost 12 Mb. The 22q11 region is particularly rich in repeated clusters[41]. Previous work described a low-copy repeat family in 22q11 that might mediate recombination events leading to the chromosomal rearrangements seen in cat eye,

velocardiofacial and DiGeorge syndromes[21,42]. The availability of the entire DNA sequence allows detailed dissection of the molecular structure of these low-copy repeats (LCR22s). Edelmann *et al.* described eight LCR22 regions[21,42]. We were unable to find the LCR22 repeat closest to the centromere, but it may lie in the gap at 700 kb from the centromeric end of the sequence. The other LCR22 regions are distributed over 6.5 Mb of 22q11. Analysis of the sequence shows that each LCR22 contains a set of genes or pseudogenes (Fig. 4). For example, five of the LCR22s contain copies of the γ-glutamyl transferase genes and γ-glutamy-transfer-ase-related genes. There is also evidence that a more distant sequence at ~16,000 kb from the centromeric start of the genomic sequence shares certain sequences with the LCR22 repeats. This similarity involves related genomic fragments including parts of the Ret-finger-protein-like genes, and the IGLC and IGLV genes.

Regions of conserved synteny with the mouse

The genomic organization of different mammalian species is well known to be conserved[43]. Comparison of genetic and physical maps across species can aid in predicting gene locations in other species, identifying candidate disease genes[13], and revealing various other features relevant to the study of genome organization and evolution. For all the cross-species relationships, that between man and mouse has been most studied. We have examined the relationship of the human chromosome 22 genes to their mouse orthologues.

Of the 160 genes we identified in the human chromosome 22 sequence that have orthologues in mouse, 113 of the murine orthologues have known mouse chromosomal locations (data available at http://www.sanger.ac.uk/HGP/Chr22/Mouse/). Examination of these mouse chromosomal locations mapped onto the human chromosome 22 sequence confirms the conserved linkage groups corresponding to human chromosome 22 on mouse chro-
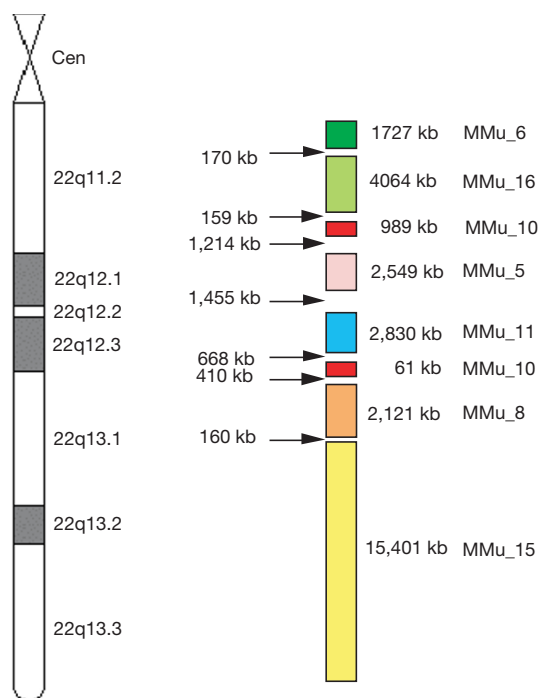
mosomes 6, 16, 10, 5, 11, 8 and 15[18,44–46] (Fig. 5). Furthermore, these studies allow placement of the sites of evolutionary rearrangements that have disrupted the conservation of synteny more accurately at the DNA sequence scale. For example, the breakdown of synteny between the mouse 8C1 block and the mouse 15E block occurs between the equivalents of the human HMOX1 and MB genes, which are separated by less than 160 kb that also contains a conspicuous 41-copy 18-nucleotide tandem repeat. A clear prediction from these data is that, for the most part, the unmapped murine orthologues of the human genes lie within these established linkage groups, along with the orthologues of the human genes that currently lack mouse counterparts. Exploitation of the chromosome 22 sequence may hasten the determination of the mouse genomic sequence in these regions.

Conclusions

We have shown that the clone by clone strategy is capable of generating long-range continuity sufficient to establish the operationally complete genomic sequence of a chromosome. In doing so, we have generated the largest contiguous segment of DNA sequence to our knowledge to date. The analysis of the sequence gives a foretaste of the information that will be revealed from the remaining chromosomes.

We were unable to obtain sequence over 11 small gaps using the available cloning systems. It may be possible that additional approaches such as using combinations of cloning systems with small insert sizes and low-copy number could reduce the size of these gaps. Direct cloning of restriction fragments that cross these gaps into small insert plasmid or M13 libraries, or direct sequencing approaches might eventually provide access to all the sequence in the gaps. However, closing these gaps is certain to require considerable time and effort, and might be considered as a specialist activity outside the core genome-sequencing efforts. It also is probable that the sequence features responsible for several of these gaps are unlikely to be specific to chromosome 22. In the best case, similar unclonable sequences might be restricted to the centromeric and telomeric regions of the other chromosomes and areas with large tandem repeats, and it will be possible to obtain large contiguous segments for the bulk of the euchromatic genome.

Over the course of the project, the emerging sequence of chromosome 22 has been made available in advance of its final completion through the internet sites of the consortium groups and the public sequence databases[47]. The benefits of this policy can be seen in both the regular requests received from investigators for materials and information that arise as the result of sequence homology searches, and the publications that have used the data[14–19]. The genome project will continue to pursue this data release policy as we move closer to the anticipated completed sequence of humans, mice and other complex genomes[47,48]. □

Figure 5 Regions of conserved synteny between human chromosome 22 and the mouse genome. Regions of mouse chromosomes with conserved synteny to human chromosome 22 are shown as adjacent coloured blocks, determined by the mouse map position of mouse orthologues to human chromosome 22 genes. The size of human chromosome 22 corresponding to each mouse chromosomal region is indicated in kb, as well as the size of the gap between the last orthologue in each conserved block. These data are available at http://www.sanger.ac.uk/Chr22/Mouse.

Methods

1. The Sanger Centre & The Genome Sequencing Centre. Toward a complete human genome sequence. *Genome Res.* **8,** 1097–1108 (1998).
2. Weber, J. L. & Myers, E. W. Human whole-genome shotgun sequencing. *Genome Res.* **7,** 401–409 (1997).
3. Green, P. Against a whole-genome shotgun. *Genome Res.* **7,** 410–417 (1997).
4. Collins, F. S. *et al.* New goals for the U.S. human genome project: 1998–2003. *Science* **282,** 682–689 (1998).
5. Morton, N. E. Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88,** 7474–7476 (1991).
6. Deloukas, P. *et al.* A physical map of 30,000 human genes. *Science* **282,** 744–746 (1998).
7. Craig, J. M. & Bickmore, W. A. The distribution of CpG islands in mammalian chromosomes. *Nature Genet.* **7,** 376–382 (1994).
8. Saccone, S., Caccio, S., Kusuda, J., Andreozzi, L. & Bernardi, G. Identification of the gene-richest bands in human chromosomes. *Gene* **174,** 85–94 (1996).

9. Collins, J. E. *et al*. A high-density YAC contig map of human chromosome 22. *Nature* **377,** 367–379 (1995).

10. Pulver, A. E. *et al*. Psychotic illness in patients diagnosed with velo-cardio-facial syndrome and their relatives. *J. Nerv. Ment. Dis.* **182,** 476–478 (1994).

11. Gill, M. *et al*. A combined analysis of D22S278 marker alleles in affected sib-pairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. Schizophrenia Collaborative Linkage Group (Chromosome 22). *Am. J. Med. Genet.* **67,** 40–45 (1996).

12. Zu, L., Figueroa, K. P., Grewal, R. & Pulst, S. M. Mapping of a new autosomal dominant spinocerebellar ataxia to chromosome 22. *Am. J. Hum. Genet.* **64,** 594–599 (1999).

13. Southard-Smith, E. M. *et al*. Comparative analyses of the dominant megacolon-SOX10 genomic interval in mouse and human. *Mamm. Genome* **10,** 744–749 (1999).

14. Nishino, I., Spinazzola, A. & Hirano, M. Thymidine phosphorylase gene mutations in MNGIE, a human mitochondrial disorder. *Science* **283,** 689–692 (1999).

15. Peyrard, M. *et al*. The human LARGE gene from 22q12.3-q13.1 is a new, distinct member of the glycosyltransferase gene family. *Proc. natl Acad. Sci. USA* **96,** 598–603 (1999).

16. Kao, H. T. *et al*. A third member of the synapsin gene family. *Proc. Natl Acad. Sci. USA* **95,** 4667–4672 (1998).

17. Mittman, S., Guo, J., Emerick, M. C. & Agnew, W. S. Structure and alternative splicing of the gene encoding alpha1I, a human brain T calcium channel alpha1 subunit. *Neurosci. Lett.* **269,** 121–124 (1999).

18. Seroussi, E. *et al*. TOM1 genes map to human chromosome 22q13.1 and mouse chromosome 8C1 and encode proteins similar to the endosomal proteins HGS and STAM. *Genomics* **57,** 380–388 (1999).

19. Seroussi, E. *et al*. Duplications on human chromosome 22 reveal a novel ret finger protein-like gene family with sense and endogenous antisense transcripts. *Genome Res.* **9,** 803–814 (1999).

20. Ning, Y., Rosenberg, M., Biesecker, L. G. & Ledbetter, D. H. Isolation of the human chromosome 22q telomere and its application to detection of cryptic chromosomal abnormalities. *Hum. Genet.* **97,** 765–769 (1996).

21. Edelmann, L., Pandita, R. K. & Morrow, B. E. Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. *Am. J. Hum. Genet.* **64,** 1076–1086 (1999).

22. McDermid, H. E. *et al*. Long-range mapping and construction of a YAC contig within the cat eye syndrome critical region. *Genome Res.* **6,** 1149–1159 (1996).

23. Johnson, A. *et al*. A 1.5-Mb contig within the cat eye syndrome critical region at human chromosome 22q11.2. *Genomics* **57,** 306–309 (1999).

24. Kawasaki, K. *et al*. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res.* **7,** 250–261 (1997).

25. Felsenfeld, A., Peterson, J., Schloss, J. & Guyer, M. Assessing the quality of the DNA sequence from the Human Genome Project. *Genome Res.* **9,** 1–4 (1999).

26. Mewes, H. W. *et al*. Overview of the yeast genome. *Nature* **387,** 7–65 (1997).

27. Blattner, F. R. *et al*. The complete genome sequence of Escherichia coli K-12. *Science* **277,** 1453–1474 (1997).

28. The C. elegans Sequencing Consortium. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282,** 2012–2018 (1998).

29. Solovyev, V. & Salamov, A. The Gene-Finder computer tools for analysis of human and model organisms genome sequences. *Ismb* **5,** 294–302 (1997).

30. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268,** 78–94 (1997).

31. Cross, S. H. & Bird, A. P. CpG islands and genes. *Curr. Opin. Genet. Dev.* **5,** 309–314 (1995).

32. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).

33. Bateman, A. *et al*. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27,** 260–262 (1999).

34. Hofmann, K., Bucher, P., Falquet, L. & Bairoch, A. The PROSITE database, its status in 1999. *Nucleic Acis Res.* **27,** 215–219 (1999).

35. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **27,** 49–54 (1999).

36. Dib, C. *et al*. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380,** 152–154 (1996).

37. Holmquist, G. P. Chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.* **51,** 17–37 (1992).

38. Bernardi, G. *et al*. The mosaic genome of warm-blooded vertebrates. *Science* **228,** 953–958 (1985).

39. The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401,** 921–923 (1999).

40. Bernardi, G. The isochore organization of the human genome. *Annu. Rev. Genet.* **23,** 637–661 (1989).

41. Collins, J. E., Mungall, A. J., Badcock, K. L., Fay, J. M. & Dunham, I. The organization of the gamma-glutamyl transferase genes and other low copy repeats in human chromosome 22q11. *Genome Res.* **7,** 522–531 (1997).

42. Edelman, L. *et al*. A common molecular basis for rearrangement disorders on chromosome 22q11. *Hum. Mol. Genet.* **8,** 1157–1167 (1999).

43. Eppig, J. T. & Nadeau, J. H. Comparative maps: the mammalian jigsaw puzzle. *Curr. Opin. Genet. Dev.* **5,** 709–716 (1995).

44. Bucan, M. *et al*. Comparative mapping of 9 human chromosome 22q loci in the laboratory mouse. *Hum. Mol. Genet.* **2,** 1245–1252 (1993).

45. Carver, E. A. & Stubbs, L. Zooming in on the human-mouse comparative map: genome conservation re-examined on a high-resolution scale. *Genome Res.* **7,** 1123–1137 (1997).

46. Puech, A. *et al*. Comparative mapping of the human 22q11 chromosomal region and the orthologous region in mice reveals complex changes in gene organization. *Proc. Natl Acad. Sci. USA* **94,** 14608–14613 (1997).

47. Bentley, D. R. Genomic sequence information should be released immediately and freely in the public domain. *Science* **274,** 533–534 (1996).

48. Guyer, M. Statement on the rapid release of genomic DNA sequence. *Genome Res.* **8,** 413 (1998).

49. Dunham, I., Dewar, K., Kim, U.-J. & Ross, M. T. in *Genome Analysis: A Laboratory Manual Series, Volume 3: Cloning Systems* (eds Birren, B. *et al*.) 1–86 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).

50. Asakawa, S. *et al*. Human BAC library: construction and rapid screening. *Gene* **191,** 69–79 (1997).

Contributors: I. Dunham[1], A. R. Hunt[1], J. E. Collins[1], R. Bruskiewich[1], D. M. Beare[1], M. Clamp[1], L. J. Smink[1], R. Ainscough[1], J. P. Almeida[1], A. Babbage[1], C. Bagguley[1], J. Bailey[1], K. Barlow[1], K. N. Bates[1], O. Beasley[1], C. P. Bird[1], S. Blakey[1], A. M. Bridgeman[1]*, D. Buck[1]*, J. Burgess[1]*, W. D. Burrill[1], J. Burton[1], C. Carder[1], N. P. Carter[1], Y. Chen[1], G. Clark[1], S. M. Clegg[1], V. Cobley[1], C. G. Cole[1], R. E. Collier[1], R. E. Connor[1]*, D. Conroy[1]*, N. Corby[1], G. J. Coville[1], A. V. Cox[1], J. Davis[1]*, E. Dawson[1], P. D. Dhami[1], C. Dockree[1], S. J. Dodsworth[1]*, R. M. Durbin[1], A. Ellington[1], K. L. Evans[1], J. M. Fey[1], K. Fleming[1], L. French[1], A. A. Garner[1], J. G. R. Gilbert[1], M. E. Goward[1], D. Grafham[1], M. N. Griffiths[1], C. Hall[1]*, R. Hall[1], E. Hall-Tamlyn[1], R. W. Heathcott[1]*, S. Ho[1]*, S. Holmes[1], S. E. Hunt[1], M. C. Jones[1], J. Kershaw[1], A. Kimberley[1], A. King[1], G. K. Laird[1], C. F. Langford[1], M. A. Leversha[1], C. Lloyd[1], D. M. Lloyd[1], I. D. Martyn[1], M. Mashreghi-Mohammadi[1], L. Matthews[1], O. T. McCann[1], J. McClay[1], S. McLaren[1], A. A. McMurray[1], S. A. Milne[1], B. J. Mortimore[1], C. N. Odell[1], R. Pavitt[1], A. V. Pearce[1], D. Pearson[1], B. J. Phillimore[1], S. H. Phillips[1], R. W. Plumb[1], H. Ramsay[1], Y. Ramsey[1], L. Rogers[1], M. T. Ross[1], C. E. Scott[1], H. K. Sehra[1], C. D. Skuce[1], S. Smalley[1], M. L. Smith[1], C. Soderlund[1], L. Spragon[1], C. A. Steward[1], J. E. Sulston[1], R. M. Swann[1], M. Vaudin[1]*, M. Wall[1], J. M. Wallis[1], M. N. Whiteley[1]*, D. Willey[1], L. Williams[1], S. Williams[1], H. Williamson[1]*, T. E. Wilmer[1], L. Wilming[1], C. L. Wright[1], T. Hubbard[1], D. R. Bentley[1], S. Beck[1], J. Rogers[1], N.Shimizu[2], S. Minoshima[2], K. Kawasaki[2], T. Sasaki[2], S. Asakawa[2], J. Kudoh[2], A. Shintani[2], K. Shibuya[2], Y. Yoshizaki[2], N. Aoki[2], S. Mitsuyama[2], B. A. Roe[3], F. Chen[3], L. Chu[3], J. Crabtree[3], S. Deschamps[3], A. Do[3], T. Do[3], A. Dorman[3], F. Fang[3], Y. Fu[3], P. Hu[3], A. Hua[3], S. Kenton[3], H. Lai[3], H. I. Lao[3], J. Lewis[3], S. Lewis[3], S.-P. Lin[3], P. Loh[3], E. Malaj[3], T. Nguyen[3], H. Pan[3], S. Phan[3], S. Qi[3], Y. Qian[3], L. Ray[3], Q. Ren[3], S. Shaull[3], D. Sloan[3], L. Song[3], Q. Wang[3], Y. Wang[3], Z. Wang[3], J. White[3], D. Willingham[3], H. Wu[3], Z. Yao[3], M. Zhan[3], G. Zhang[3], S. Chissoe[4], J. Murray[4], N. Miller[4], P. Minx[4], R. Fulton[4], D. Johnson[4], G. Bemis[4], D. Bentley[4], H. Bradshaw[4], S. Bourne[4], M. Cordes[4], Z. Du[4], L. Fulton[4], D. Goela[4], T. Graves[4], J. Hawkins[4], K. Hinds[4], K. Kemp[4], P. Latreille[4], D. Layman[4], P. Ozersky[4], T. Rohlfing[4], P. Scheet[4], C. Walker[4], A. Wamsley[4], P. Wohldmann[4], K. Pepin[4], J. Nelson[4], I. Korf[4], J. A. Bedell[4], L. Hillier[4], E. Mardis[4], R. Waterston[4], R. Wilson[4], B. S. Emanuel[5], T. Shaikh[5], H. Kurahashi[5], S. Saitta[5], M. L. Budarf[6], H. E. McDermid[6], A. Johnson[6], A. C. C. Wong[6], B. E. Morrow[7], L. Edelmann[7], U. J. Kim[8], H. Shizuya[8], M. I. Simon[8], J. P. Dumanski[9], M. Peyrard[9], D. Kedra[9], E. Seroussi[9], I. Fransson[9], I. Tapia[9], C. E. Bruder[9], K. P. O'Brien[9].

Addresses: 1, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 2, Department of Molecular Biology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan; 3, Department of Chemistry and Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 311, Norman, Oklahoma 73019, USA; 4, Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Blvd, St. Louis, Missouri 63108, USA 5, Division of Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia and the Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA; 6, Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada; 7, Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA; 8, California Institute of Technology, Division of Biology, Pasadena, California 91125, USA; 9, Department of Molecular Medicine, Clinical Genetics Unit, Karolinska Hospital, CMM bldg. L8:00, 17176 Stockholm, Sweden

* *Present addresses: Division of Virology, Department of Pathology, Tennis Court Road, Cambridge CB12 1QP, UK (A. M. Bridgeman); Lark Technologies, Radwinter Road, Saffron Walden Essex, CB11 3HY, UK (D. Buck, R. E. Connor, S. Ho); Australian Genome Research Facility, Gehrmann Laboratories, University of Queensland, St Lucia QLD 4072, Australia (J. Burgess, J. Davis); Incyte Europe Ltd, 214 Cambridge Science Park, Cambridge CB4 0WA, UK (D. Conroy); Tepnel Life Sciences, Innovation Centre, Scotscroft Building, Wilmslow Road, Didsbury, Manchester M20 8RY, UK (S. J. Dodsworth); Department of Brassica & Oilseeds Research, Cambridge Laboratory, John Innes Centre, Norwich, Norfolk, UK (C. Hall); Cancer Genetics Lab, Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin, New Zealand (R. W. Heathcott); Dept. of Zoology, University of Cambridge, Downing Street, CB2 3EJ, UK (M. N. Whiteley); Monsanto, Genomic Team, Mail Zone N3SA, 800 North Lindbergh Boulevard, St. Louis, Missouri 63167, USA (M. Vaudin); Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK (H. Williamson)*
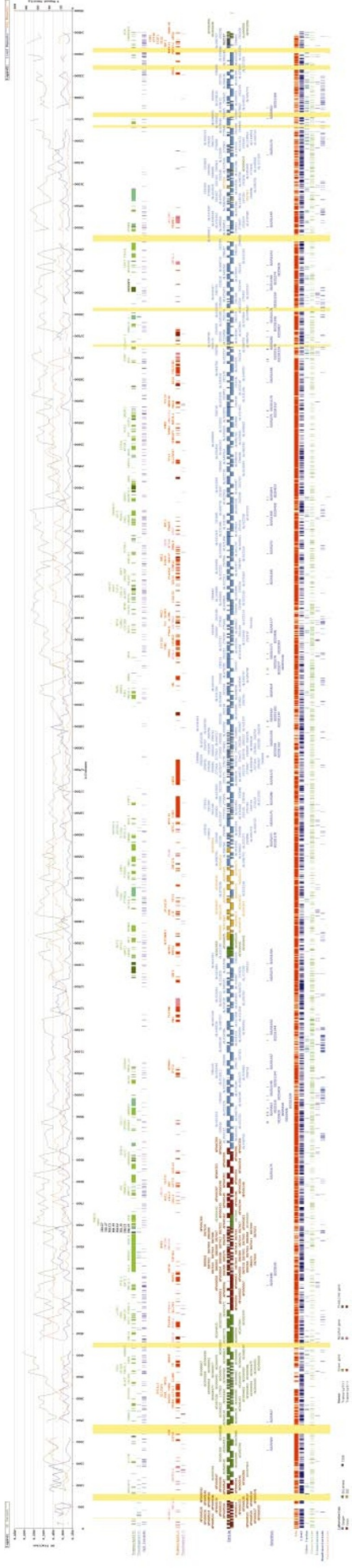
redistribution of atmospheric mass and non-adiabatic heating during a dipole mode event may result in such an event influencing the countries surrounding the Indian Ocean through atmospheric teleconnections, as evidenced by the northward shift of the Pacific Subtropical High off the Philippines which caused an unusually hot summer in the far east Asian countries in 1994 (refs 27, 28). Some atmospheric general circulation model experiments[29,30] have demonstrated the important effects of the Indian Ocean SST changes on east African rainfall variability. Further experiments with atmospheric and oceanic models, extended in scope, will be needed to understand the full implications of this hitherto unexplored mode of climate variability. The dipole mode event should also provide a testing ground for coupled ocean–atmosphere models. □

1. Neelin, J. D. et al. ENSO theory. J. Geophys. Res. C **103,** 14261–14290 (1998).
2. Zebiak, S. E. Air–sea interaction in the equatorial Atlantic region. J. Clim. **6,** 1567–1586 (1993).
3. Wallace, J. M. et al. On the structure and evolution of ENSO-related climate variability in the tropical Pacific. J. Geophys. Res. **103,** 14241–14260 (1998).
4. Flohn, H. East African rains of 1961/62 and the abrupt change of the White Nile discharge. Palaeoecol. Afr. **18,** 3–18 (1987).
5. Kapala, A., Born, K. & Flohn, H. in Proc. Int. Conf. on Monsoon Variability and Prediction 119–126 (Int. Centre for Theoretical Physics, Trieste, 1994).
6. Reverdin, G., Cadet, D. L. & Gutzler, D. Interannual displacements of convection and surface circulation over the equatorial Indian Ocean. Q. J. R. Meteorol. Soc. **112,** 43–67 (1986).
7. Rayner, N. A., Horton, E. B., Parker, D. E., Folland, C. K. & Hackett, R. B. Version 2.2 of the Global Sea-Ice and Sea Surface Temperature Data Set, 1903–1994 (Clim. Res. Tech. Note 74, UK Meteorological Office, Bracknell, 1996).
8. Kalnay, E. et al. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. **77,** 437–471 (1996).
9. Xie, P. & Arkin, P. A. Analyses of global monthly precipitation using gauge observations, satellite estimates and numerical model predictions. J. Clim. **9,** 840–858 (1996).
10. Barnett, T. P. Interaction of the Monsoon and Pacific Trade Wind System at interannual time scales. Part I: The equatorial zone. Mon. Weath. Rev. **111,** 756–773 (1983).
11. Wyrtki, K. An equatorial jet in the Indian Ocean. Science **181,** 262–264 (1973).
12. Yamagata, T., Mizuno, K. & Masumoto, Y. Seasonal variations in the equatorial Indian Ocean and their impact on the Lombok throughflow. J. Geophys. Res. **101,** 12465–12474 (1996).
13. Stout, J. E. & Young, J. A. Low-level monsoon dynamics derived from satellite winds. Mon. Weath. Res. **111,** 774–799 (1983).
14. Clarke, A. J. & Liu, X. Observations and dynamics of semiannual and annual sea levels near the eastern equatorial Indian Ocean boundary. J. Phys. Oceanogr. **23,** 386–399 (1993).
15. Reppin, J., Schott, F. A., Fischer, J. & Quadfasel, D. Equatorial currents and transports in the upper central Indian Ocean: Annual cycle and interannual variability. J. Geophys. Res. (in the press).
16. Vinayachandran, P. N., Saji, N. H. & Yamagata, T. Response of the equatorial Indian Ocean to an unusual wind event during 1994. Geophys. Res. Lett. **26,** 1613–1616 (1999).
17. Behera, S. K., Krishnan, S. & Yamagata, T. Anomalous air–sea coupling in the southern tropical Indian Ocean during the boreal summer of 1994. Geophys. Res. Lett. (in the press).
18. Murtugudde, R., McCreary, J. P. & Busalacchi, A. J. Oceanic processes associated with anomalous events in the Indian Ocean with relevance to 1997–1998. J. Geophys. Res. (in the press).
19. Potemra, J. & Lukas, R. Seasonal to interannual modes of sealevel variability in the western Pacific and eastern Indian Oceans. Geophys. Res. Lett. **26,** 365–368 (1999).
20. Yu, L. S. & Rienecker, M. M. Mechanisms for the Indian Ocean warming during the 1997–98 El Niño. Geophys. Res. Lett. **26,** 735–738 (1999).
21. Meyers, G. Variation of Indonesian throughflow and the El Niño–Southern Oscillation. J. Geophys. Res. **101,** 12255–12263 (1996).
22. Hastenrath, S., Nicklis, A. & Greischar, L. Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial Indian Ocean. J. Geophys. Res. **98,** 20219–20235 (1993).
23. Levitus, S., Boyer, P. & Antonov, J. World Ocean Atlas, Vol. 5, Interannual Variability of Upper Ocean Thermal Structure. (NOAA Atlas NESDIS, NODC/NOAA, Silver Spring, Maryland, USA, 1994).
24. Xie, S.-P. & Philander, S. G. H. A coupled ocean-atmosphere model of relevance to the ITCZ in the eastern Pacific. Tellus A **46,** 340–350 (1994).
25. Goswami, B. N. A multiscale interaction model for the origin of the tropospheric QBO. J. Clim. **8,** 524–534 (1995).
26. Goswami, B. N., Krishnamurthy, V. & Annamalai, H. A broad scale circulation index for the interannual variability of the Indian summer monsoon. Q. J. R. Meteorol. Soc. **125,** 611–633 (1999).
27. Kawamura, R., Sugi, M., Kayahara, T. & Sato, N. Recent extraordinary cool and hot summers in east Asia simulated by an ensemble climate experiment. J. Meteorol. Soc. Jpn. **76,** 597–617 (1998).
28. Nitta, T. Abnormal hot summer of 1994 and tropical convective activity. Weath. Serv. Bull./Jpn. Meteorol. Agency **63** (suppl.), 178–187 (1996) (in Japanese).
29. Goddard, L. & Graham, N. E. The importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa. J. Geophys. Res. (in the press).
30. Latif, M., Dommenget, D. & Dima, M. The role of Indian Ocean sea surface temperature in forcing east African climate anomalies. J. Clim. (in the press).

···

Tribosphenic mammal from the North American Early Cretaceous

Richard L. Cifelli

Oklahoma Museum of Natural History and Department of Zoology, University of Oklahoma, Norman, Oklahoma 73019, USA

···

The main groups of living mammals, marsupials and eutherians, are presumed to have diverged in the Early Cretaceous[1], but their early history and biogeography are poorly understood. Dental remains have suggested that the eutherians may have originated in Asia[2], spreading to North America in the Late Cretaceous, where an endemic radiation of marsupials was already well underway[3]. Here I describe a new tribosphenic mammal (a mammal with lower molar heels that are three-cusped and basined) from the Early Cretaceous of North America, based on an unusually complete specimen. The new taxon bears characteristics (molarized last premolar, reduction to three molars) otherwise known only for Eutheria among the tribosphenic mammals. Morphometric analysis and character comparisons show, however, that its molar structure is primitive (and thus phylogenetically uninformative), emphasizing the need for caution in interpretation of isolated teeth. The new mammal is approximately contemporaneous with the oldest known Eutheria from Asia. If it is a eutherian, as is indicated by the available evidence, then this group was far more widely distributed in the Early Cretaceous than previously appreciated. An early presence of Eutheria in North America offers a potential source for the continent's Late Cretaceous radiations, which have, in part, proven difficult to relate to contemporary taxa in Asia.

Class Mammalia
Infralegion Tribosphenida, Order and Family Uncertain
Montanalestes keebleri gen. et sp. nov.

Etymology. Montana: the state of origin; lestes (Greek): robber; *keebleri*: in honour of the Keebler family of Billings, Montana, in recognition of its support for field parties of the Oklahoma Museum of Natural History (OMNH).

Holotype. OMNH 60793, associated dentaries with right p3–5, m1–3 (terminology follows ref. 4), and alveoli for p2; and left p5, m1–3 (Fig. 1).

Locality and Horizon. OMNH V1156, basal unit VII (ref. 5), Cloverly Formation (Aptian–Albian; ~110 Myr); Carbon County, Montana, USA.

Diagnosis. Differs from primitive Tribosphenida (for example, *Kermackia*), in so far as is known, in having a semimolarized last premolar (lingually placed paraconid, incipient talonid basin, metaconid present) and only three molars; molars differ in having a broader, more fully basined talonid with three cusps consistently present. Differs from early marsupials and marsupial-like taxa[6] in postcanine tooth formula and in having semimolarized last premolar and a smaller size differential between m1–2; lower molars differ in lacking a labial postcingulid and in having a greater trigonid–talonid height differential, narrower talonid and a hypoconulid that is not lingually situated. Differs from *Slaughteria*[7] in having a semimolarized, less anteroposteriorly compressed last premolar; molars differ in having a smaller size differential between m1–2, no lingual knob on anterocingulid, more acute trigonid angle on m1 and taller paraconid relative to metaconid. Similar to early Eutheria in postcanine tooth formula (four or more premolars, three molars) and semimolarized last premolar; differs from most comparable taxon, *Prokennalestes*[8], in having salient lingual cingulids on the lower premolars and stronger paraconid and metaconid on p5, the former more lingually placed. Lower molars differ from those of *Prokennalestes* in having sequential length

363

increase from first to third, taller paraconid, and greater height differential between trigonid and talonid and between protoconid and metaconid, more distinct and anteriorly placed entoconid, and hypoconulid more projecting on m1–2.

Description. The Meckelian groove (Fig. 2) and labial mandibular foramen, found in a number of Cretaceous mammals including presumed eutherians[8], are absent. A roughened depression at the junction of the ascending and horizontal rami may indicate the presence of the coronoid, a presumed primitive condition with a broad distribution among early mammals, including some eutherians[8]. The angular process is elongate, hook-like, projects posteriorly and is in-turned medially, as in primitive Tribosphenida
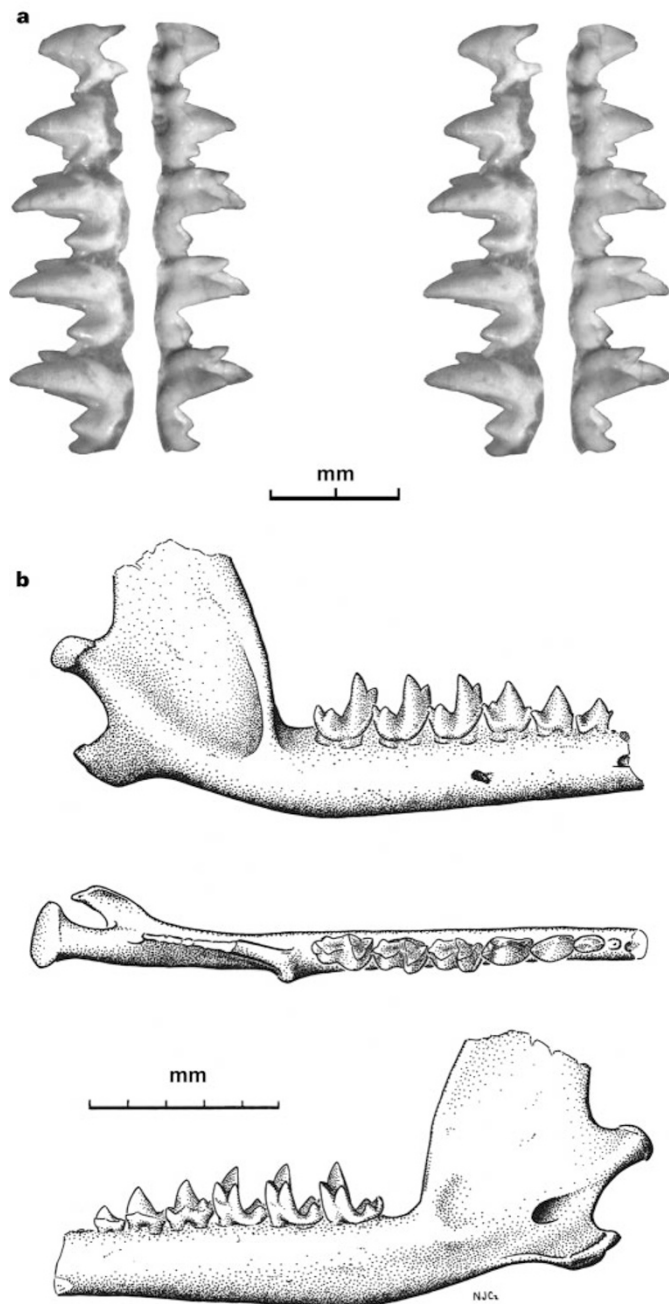

Figure 1 *Montanalestes keebleri* gen. et sp. nov., OMNH 60793. **a**, Stereopairs of p4−m3 (other parts of dentition and mandible omitted for clarity) in labial (left) and lingual (right) views. **b**, restored right mandible and dentition in labial (top), occlusal (middle) and lingual (bottom) views.
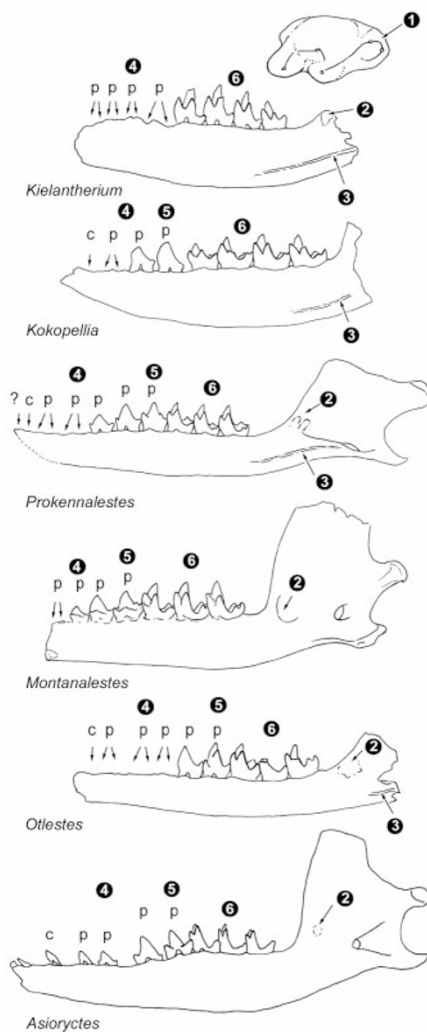

Figure 2 Jaws of Cretaceous tribosphenic mammals, showing distribution of selected tooth and jaw characters (some restored after refs given in Supplementary Information). *Kielantherium* is a plesiomorphic tribosphenidan of uncertain affinities[14]; *Kokopellia* is a marsupial or close relative[6]; and *Prokennalestes*[8], *Otlestes*[16] and *Asioryctes*[22] are Cretaceous eutherians. 1, Talonid. Primitively (*Kielantherium*), this is small, poorly basined and bears only two cusps; in remaining taxa it is broader and well basined, with three cusps[11]. 2, Coronoid bone. This element, or a facet indicating its presence, is a primitive feature presumably lost early in marsupial history and independently among eutherians, but retained in many Cretaceous mammals[6,8]. 3, Meckelian groove (uncertain in *Otlestes*[16] and *Kokopellia*[6]). This feature is another primitive character retained in some early eutherians and other Mesozoic mammals[8,14,23]. 4, Premolar count. The primitive counts for premolars and molars are uncertain, depending partly on varying interpretation and relevance of non-tribosphenic *Peramus*, and partly on disputed homologies[12,24]. At least four premolars are present in *Peramus* and *Kielantherium*, whereas deltatheroidans and marsupials have only three[11,13–15]. The widespread occurrence of at least four in early eutherians is assumed to be primitive. 5, Last premolar. This tooth is simple in *Kokopellia*, marsupials and most Mesozoic mammals; by contrast, the premolars and molars of eutherians generally form a more evenly graded series, with the last premolar tending toward a semimolariform appearance[8,12]. *Montanalestes* appears to be more advanced than the rather variable[10] *Prokennalestes*, but the talonid and metaconid are not as well developed as in somewhat younger *Otlestes*[16]. 6, Molar count. This is uncertain for many early tribosphenidans[7]. *Kielantherium*, like marsupials and deltatheroidans, has four; *Montanalestes* is similar to eutherians in having only three molars, presumably an advanced condition[11,12,15]. ?, alveolus for tooth of uncertain identity; p, premolar (or alveoli (arrows) documenting presence of a premolar).

generally[9]. At least four premolars are present; the relatively small size of p3 is similar to those in some Cretaceous eutherians that have five premolars[10]. The premolar–molar series intergrade: p5 is semimolariform, with a lingually placed paraconid; a metaconid that is clear but that is placed low on the crown and is not so lingually situated; and a small talonid basin enclosed lingually by the cingulum (although only one talonid cusp is present). The molars are virtually unworn and facets are poorly developed. They are advanced with respect to primitive tribosphenidans (Fig. 2) in having a well developed, basined, tricuspid talonid. The molars generally resemble those of pappotheriids and some early eutherians[11], although the latter tend to have a reduced paraconid[12]. Molars of *Montanalestes* are unusual in having a relatively strong paraconid and a sequential increase in length from m1–3.

Montanalestes is undoubtedly a tribosphenic mammal, as shown by the broad, basined, three-cusped molar talonids[13]; and it is advanced with respect to dentally primitive taxa such as *Kielantherium*[14] in various features (Fig. 2). *Montanalestes* is clearly not a marsupial: it retains more than three premolars, lacks specialized molar characteristics[6,12,15] and has a molarized distal premolar[11,12], among other features. Eutherian affinities are, however, suggested by dental characteristics. The presence of four

molars appears to be primitive for tribosphenic mammals[11,12,15]; *Montanalestes* shares the condition seen in eutherians, a presumed reduction to three molars. The premolars of most early mammals are simple and form a sharp break with the molar series, whereas eutherians are presumed to be derived in having premolar–molar series intergrade, with a tendency for the last premolar to be somewhat molariform[8,12]. *Montanalestes* is similar to eutherians in this respect: the last premolar bears a strong, lingually placed paraconid, a metaconid and a slightly basined talonid, all features commonly seen in premolars undergoing molarization. By comparison to *Prokennalestes* and other early Eutheria[2,8,11,16], *Montanalestes* presents a mixture of advanced (for example, metaconid on p5, loss of Meckelian groove) and primitive (for example, unreduced paraconid) features. Molar structure is of particular interest in this context because it has traditionally been emphasised in interpretation of early mammal radiations[11,17]. Morphometric analysis shows that marsupials and dentally advanced, Late Cretaceous eutherians can be recognized in the basis of lower molar shape factors; plesiomorphic tribosphenidans are also distinct (Fig. 3). However, certain early tribosphenidans, *Montanalestes* and some early eutherians are structurally appropriate antecedents to both major groups of tribosphenic mammals. Lower molars appear not to be particularly informative for critical early tribosphenidan taxa (as illustrated by the case of *Holoclemensia*, earlier suggested to be marsupial[17], but more closely resembling eutherians based on referred lower molars; Fig. 3), indicating that taxa known only from molar structure must be interpreted cautiously. The relationship of *Montanalestes* to other early eutherians remains uncertain because available data produce ambiguous results.

The earliest generally accepted eutherians are from the Early and early Late Cretaceous of Asia[8,10,16,18]. Previous reports of Early Cretaceous Eutheria in North America[17] are based on extremely fragmentary specimens and have been controversial because of differing interpretations of tooth formula and specimen association[7]. Uncontested eutherians do not appear in North America until the early Campanian (about 80 Myr)[19]. By contrast, marsupials and presumed relatives were well established on the continent by the medial Cretaceous, about 100 Myr[6,20]. These distributions have led to the suggestion that eutherians and marsupials arose in Eurasia and North America, respectively, and that North American eutherians are derivatives of Late Cretaceous immigration from Asia[3]. Except for ungulates[4], however, Late Cretaceous eutherians of Asia and North America represent distinct clades whose intercontinental relationships are problematic[11,19]. If *Montanalestes* is related to Eutheria, as the balance of evidence indicates, then the continent of origin for the group is called into question: evidently, Eutheria enjoyed a far broader Early Cretaceous distribution (echoing that of the triconodont mammal *Gobiconodon*[21]) and an earlier evolutionary radiation than previously envisaged. An early presence of Eutheria in North America indicates that some of the continent's distinctive Late Cretaceous taxa may have evolved *in situ*, rather than having a proximal Asiatic origin—a theory to be tested by recovery of further faunas from the medial Cretaceous, where the record so far has been notoriously poor[13]. □
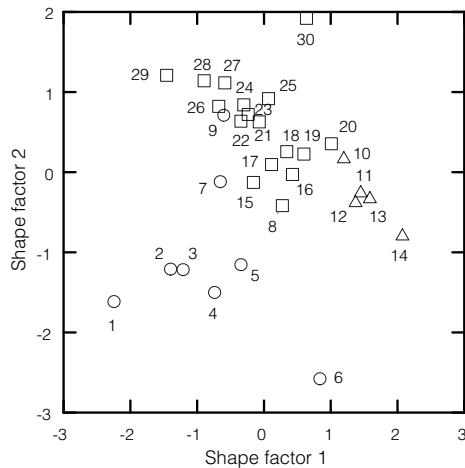
Figure 3 Comparison of lower molar shape in selected Cretaceous tribosphenic mammals and the eupantothere *Peramus*. Measurements and methodology follow ref. 11; data from this source has been standardized and variation due to size has been removed. Plot shows the first two axes of a principal components analysis. Positive scores (representing advanced conditions) on axis 1 reflect mainly increased length of the posthypocristid and width of the talonid, and relatively lower protoconid; this shape factor serves mainly to distinguish marsupials from eutherians and primitive tribosphenic/non-tribosphenic mammals. Positive scores on axis 2 reflect increased trigonid width, anteroposterior compression of the trigonid and lowered height differential between protoconid and metaconid. This shape factor generally distinguishes non-tribosphenic and some primitive tribosphenic mammals from the more advanced marsupials and, especially, eutherians. Taxa in the middle of the graph (for example, *Montanalestes* (8), *Pappotherium* (7), *Procerberus* (15)) have molars that, in terms of the two shape factors, are structurally antecedent to both marsupials and eutherians. Note placement and clustering of *Deltatheridium* (4) with primitive taxa (1–5), and placement of *Holoclemensia* (9) and *Kokopellia* (12) with respect to undoubted eutherians and marsupials. Circles, non-tribosphenic taxa and Theria of metatherian–eutherian grade; triangles, marsupials; squares, eutherians. 1, *Peramus*; 2, *Hypomylos*; 3, *Kielantherium*; 4, *Deltatheridium*; 5, *Kermackia*; 6, *Potamotelses*; 7, *Pappotherium*; 8, *Montanalestes*; 9, *Holoclemensia*; 10, *Didelphodon vorax*; 11, *Pediomys elegans*; 12, *Kokopellia*; 13, *Glasbius*; 14, *Alphadon marshi*; 15, *Procerberus*; 16, *Prokennalestes* spp.; 17, *Batodon*; 18, *Paranyctoides* spp.; 19, *Kennalestes*; 20, *Gallolestes pachymandibularis*; 21, *Gypsonictops hypoconus*; 22, *Cimolestes cereberoides*; 23, *Otlestes*; 24, *Asioryctes*; 25, *Purgatorius unio*; 26, *Cimolestes incisus*; 27, *C. magnus*; 28, *Protungulatum* spp.; 29, *Cimolestes propalaeoryctes*; 30, *Zalambdalestes lechei*.

1. Lillegraven, J. A., Thompson, S. D., McNab, B. K. & Patton, J. L. The origin of eutherian mammals. *Biol. J. Linn. Soc.* **32,** 281–336 (1987).
2. Butler, P. M. in *Structure, Function and Evolution of Teeth* (eds Smith, P. & Tchernov, E.) 125–138 (Freund, Tel Aviv, 1990).
3. Kielan-Jaworowska, Z. in *Palaeontology, Essential of Historical Geology* (ed. Gallitelli, E. M.) 367–383 (S.T.E.M. Mucci, Modena, 1982).
4. Nessov, L. A., Archibald, J. D. & Kielan-Jaworowska, Z. Ungulate-like mammals from the Late Cretaceous of Uzbekistan and a phylogenetic analysis of Ungulatomorpha. *Bull. Carnegie Mus. Nat. Hist.* **34,** 40–88 (1998).
5. Ostrom, J. H. Stratigraphy and paleontology of the Cloverly Formation (Lower Cretaceous) of the Bighorn Basin area, Montana and Wyoming. *Bull. Peabody Mus. Nat. Hist.* **35,** 1–234 (1970).
6. Cifelli, R. L. & de Muizon, C. Definition and jaw of *Kokopellia juddi*, a primitive marsupial or near marsupial from the medial Cretaceous of Utah. *J. Mamm. Evol.* **4,** 241–258 (1997).

7. Butler, P. M. A new interpretation of the mammalian teeth of triposphenic pattern from the Albian of Texas. *Breviora* **446**, 1–27 (1978).

8. Kielan-Jaworowska, Z. & Dashzeveg, D. Eutherian mammals from the Early Cretaceous of Mongolia. *Zool. Scripta* **18**, 347–355 (1989).

9. Novacek, M. J. The skull of leptictid insectivorans and the higher-level classification of eutherian mammals. *Bull. Am. Mus. Nat. Hist.* **183**, 1–112 (1986).

10. Sigogneau-Russell, D., Dashzeveg, D. & Russell, D. E. Further data on *Prokennalestes* (Mammalia, Eutheria inc. sed.) from the Early Cretaceous of Mongolia. *Zool. Scripta* **21**, 205–209 (1992).

11. Butler, P. M. Early trends in the evolution of tribosphenic molars. *Biol. Rev.* **65**, 529–552 (1990).

12. Clemens, W. A. Jr & Lillegraven, J. A. New Late Cretaceous, North American advanced therian mammals that fit neither the marsupial nor eutherian molds. *Contrib. Geol. Univ. Wyoming* **3**, 55–85 (1986).

13. Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A. Jr (eds) *Mesozoic Mammals—the First Two-thirds of Mammalian History* (Univ. of California Press, Berkeley, 1979).

14. Dashzeveg, D. & Kielan-Jaworowska, Z. The lower jaw of an aegialodontid mammal from the Early Cretaceous of Mongolia. *Zool. J. Linn. Soc.* **82**, 217–227 (1984).

15. Rougier, G. W., Wible, J. R. & Novacek, M. J. Implications of *Deltatheridium* specimens for early marsupial history. *Nature* **396**, 459–463 (1998).

16. Nessov, L. A., Sigogneau-Russell, D. & Russell, D. E. A survey of Cretaceous tribosphenic mammals from middle Asia (Uzbekistan, Kazakhstan and Tajikistan), of their geological setting, age and faunal environment. *Palaeovert.* **23**, 51–92 (1994).

17. Slaughter, B. H. Mid-Cretaceous (Albian) therians of the Butler Farm local fauna, Texas. *Zool. J. Linn. Soc.* **50** (suppl.), 131–143 (1971).

18. Wang, Y. Q., Hu, Y.-M., Chow, M.-C. & Li, C.-K. in *Sixth Symposium on Mesozoic Terrestrial Ecosystems* (eds Sun, A.-L. & Wang, Y.-Q.) 221–228 (China Ocean, Beijing, 1995).

19. Fox, R. C. *Paranyctoides maleficus* (new species), an early eutherian mammal from the Cretaceous of Alberta. *Spec. Publ. Carnegie Mus. Nat. Hist.* **9**, 9–20 (1984).

20. Cifelli, R. L. *et al.* in *Vertebrate Fossils of Utah* (ed. Gillette, D. D.) 219–242 (Utah Geological Survey, Salt Lake City, 1999).

21. Jenkins, F. A. Jr & Schaff, C. R. The Early Cretaceous mammal *Gobiconodon* (Mammalia, Tricono-donta) from the Cloverly Formation in Montana. *J. Vert. Paleontol.* **8**, 1–24 (1988).

22. Kielan-Jaworowska, Z. Evolution of the therian mammals in the Late Cretaceous of Asia. Part II. Postcranial skeleton in *Kennalestes* and *Asioryctes. Palaeontol. Pol.* **37**, 65–83 (1974).

23. Luo, Z. in *In the Shadow of the Dinosaurs—Early Mesozoic Tetrapods* (eds Fraser, N. C. & Sues, H.-D.) 98–128 (Cambridge Univ. Press, Cambridge, 1994).

24. Luckett, W. P. in *Mammal Phylogeny, volume 2—Mesozoic Differentiation, Multituberculates, Mono-tremes, Early Therians, and Marsupials* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 182–204 (Springer, New York, 1993).

....

Oldest playable musical instruments found at Jiahu early Neolithic site in China

Juzhong Zhang*†, **Garman Harbottle**‡, **Changsui Wang**† **& Zhaochen Kong**§

* *Institute of Cultural Relics and Archaeology of Henan Province, Zhengzhou, Henan, China 450000*

† *Archaeometry Laboratory, University of Science and Technology of China, Hefei, Anhui, China 230026*

‡ *Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973, USA*

§ *Paleobotany Laboratory, Institute of Botany, Academia Sinica, Beijing, China 100080*

....

Excavations at the early Neolithic site of Jiahu[1,2] **in Henan Province, China have produced what may be the earliest complete, playable, tightly-dated multinote musical instruments. Jiahu was occupied from 7000 BC to 5700 BC, considerably antedating the well known Peiligang culture**[3–5]**. Here we describe six exquisitely made complete flutes which were found in radiocarbon-dated excavation layers, along with fragments of perhaps 30 more. The**

flutes are made from the ulnae of the red-crowned crane (*Grus japonensis* Millen) and have 5, 6, 7 and 8 holes. The best preserved flute has been played and tonally analysed. In addition to early musical artefacts, the archaeological record at Jiahu[1,2] contains important information on the very foundations of Chinese society. We describe the archaeological characteristics of the Jiahu site, details concerning its dating, its place in the prehistory of the Chinese Neolithic, the ethnicity of its population and the results of a tonal analysis of a nearly 9,000-year-old musical instrument found there.

Jiahu lies in the 'Central Yellow River Valley'[3] in mid-Henan Province, east of Mount Funiu and bounded by the flood plains of the Ni river to the south and the Sha river to the north. It was discovered by Zhu Zhi in 1962 but its significance was not fully appreciated until other sites related to the Peiligang culture[3–5] had been excavated. Jiahu is an irregular ellipse, 275 m long (east–west) and 260 m wide (north–south). Below a modern agricultural layer of loess 0.3–1.0 m deep lies a layer of the Han Dynasty (206 BC–AD 220), and beneath this Neolithic layers which together are 0.5–0.9 m thick. Of Jiahu's 55,000 m², only 2,400 m² (4.4%) have been excavated, revealing 45 house foundations, 370 cellars, 9 pottery kilns, more than 300 graves and thousands of artefacts made out of bone, pottery, stone and other materials. There are three sub-periods at Jiahu; during the latest two the site shows distinct areas for residences, workshops and graves[1]. The differential furnishing of the graves provides evidence for early social stratification. Jiahu appears to represent an organized, well established sedentary village of long duration in the very early Chinese Neolithic.

The dating of Jiahu is crucial. So far we have determined twenty ¹⁴C dates: nine from charcoal (three in each of the three sub-periods mentioned above), five from plant ash, four from human bone, one from a whole fruit pit and one from a broken grain of carbonized rice. The conventional ¹⁴C ages range from 7,017 to 8,285 ¹⁴C years BP. After dendrochronological correction the corresponding range is 7,700 to 9,000 calendar years BP or nearly 5700 to 7000 cal. (calibrated) BC. The three sub-periods are dated as follows: the oldest is 6600–7000 cal. BC; the second oldest 6200–6600 cal. BC; and the third ~5700–6200 cal. BC. Thus, Jiahu spans some 1,300 years. If this range of dates is borne out by subsequent research, then Jiahu represents a Neolithic cultural phase distinctly earlier at its inception than the Peiligang culture, where the dates of many of the sites range from 5100 BC to 6300 BC[3–5]. On the basis of these dates, and considering the repertoire of ceramic shapes and stone tools[1], we would tentatively characterize Jiahu as a culture parallel to, overlapping in duration, and very possibly related to, the Peiligang. The same might apply to the site of Houli in the eastern Yellow River Valley (north-central Shandong Province), dated 5000 BC to 6200 BC[3]. For the ¹⁴C determinations, see Supplementary Information.

A careful study of the bones of 400 individuals removed from more than 300 graves indicates that the Jiahu ethnic group may be identified with the North Asian Mongolian group, and also with the Xiawanggang and Miaodigou groups in Henan Province and the Dawenkou, Yedian and Xixiahou groups that were later found in Shandong Province. The range of male heights was from 170 to 180 cm. In the late Palaeolithic Zhoukoudian Cave, 'unspecialized'

Table 1 Analysis of intervals of flute M282:20

Location of interval	Average value in cents[14]	Musical description
Between hole 1 and hole 2	284	Minor third
Between hole 2 and hole 3	244	Slightly larger than major second (a whole tone)
Between hole 7 and tube sound	260	Slightly smaller than minor third, slightly larger than a whole note.