Supplementary data need to be kept in public repositories

SIR — The reality of the genomics age is that there are many very large data sets that are most usefully saved and manipulated in electronic form. Many journals add online 'supplementary material' to articles as a service to authors wishing to publish volumes of such data that cannot be accommodated within the body of an article.

Supplementary-material collections maintained by publishers serve as archival repositories directly connected with the peer-reviewed scientific literature, often competing with or substituting for the deposition of data in public repositories.

To assess the use of these, we investigated supplementary-data archives for gene-expression profiling data, a widely used experimental protocol for which international standards for data representation have been developed.

We anticipated that such archives might be a useful source of data. But to our dismay, it was impossible to systematically analyse our sample, taken from 10,128 papers in 139 journals. No standards for organizing supplementary-data collections have been adopted either across journals or even for supplementary-data collections associated with articles in the same journal.

Data are represented in an enormous range of different file formats, from raw data files (such as Affymetrix.cel files) to spreadsheets (xls file extensions), documents (doc and pdf) and text files (txt and cvs). Within documents there are no standards for data organization: different documents provide different numbers of columns, contain both differential and absolute expression values, and often have few details about the signal processing applied to obtain data. We also encountered a significant number of typographic errors in gene names, database accession numbers and data-set identifiers.

There are public repositories for gene-expression profile data (Stanford MicroArray Database, the US National Center for Biotechnology Information Gene Expression Omnibus and the ArrayExpress repository at the European Bioinformatics Institute). We compared the accessibility of gene-expression profile data in public repositories with accessibility of data in supplementary-data archives. The public repositories provide numerous search and retrieval tools, including unique accession numbers and the ability to search by specimen, platform and profile data. Publishers' supplementary-materials archives provided none of these features. As a result, relevant data are far harder to locate than in public repositories.

These findings are not limited to gene-expression data. Even within the same journal, there is no consistency in reporting or format among bioinformatics resources. File extensions for documents, figures and movies include xls, doc, eps, jpg, tif, gif, pdf, ppt, qt, asf, wma and wmv. They may or may not include long lists of links, be compressed into zip files or offer the option of including the supplementary material as part of the downloadable document containing the printed version of the article.

Supplementary data often represent the raw experimental values and are especially important for researchers in the same field. Among the advantages of storing these data in public repositories are the integration of information with the community knowledge resources and the ability to track and maintain computer-readable associations between data sets.

On the basis of our analysis, we recommend that scientific journals adopt a policy, similar to *Nature*'s (see www.nature.com/nature/authors/policy/index.html#a7.2), of requiring that authors submit data to public repositories, if relevant repositories exist, and that the journal version should contain accession numbers, URLs and other appropriate specific indicators to the data source in the repositories.

Journals' supplementary-data archives should be restricted to idiosyncratic and nonstandard data types for which no public repository exists. Only then can community standards emerge.

Carlos Santos*, Judith Blake†, David J. States*‡
*Bioinformatics Program, University of Michigan, Ann Arbor, Michigan, USA
†The Jackson Laboratory, Bar Harbor, Maine, USA
‡Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA

Turkish science needs more than membership of the EU

SIR — Your Editorial "Turkey's evolution" (*Nature* **438**, 1–2; 2005), about the country's efforts to join the European Union (EU), states that "the opening of negotiations for EU membership offers the best hope for the continuing development of science in Turkey". This view is common in Europe, but I believe the assumptions behind it lack solid support.

First, you assume that EU policies adopted by Turkey during membership negotiations will lead to more economic investment in Turkish science. Such investment is needed if Turkey is to close the gap with more developed countries. But the increase in the science budget, to US$300 million in a country of 70 million, is inadequate. The €250 million (US$292 million) that Turkey contributed towards the EU's Sixth Framework programme is not expected to be recouped. And even though policies prescribed by the International Monetary Fund (IMF) have reduced investment in the country's educational infrastructure (E. Voydova and E. Yeldan *Comp. Econ. Stud.* **47**, 41–79; 2005), keeping to an IMF programme is a condition for Turkey's acceptance into the EU.

Second, although international scientific collaboration is crucial for scientific development in any country, the extent to which knowledge sharing and cooperation depends upon international economic and political relations is less clear. Some countries, such as Cuba, India and China, have achieved scientific progress in relatively independent economic or political circumstances. Political and cultural relations among countries at dissimilar levels of development might even impede progress on the weaker side — for example, through a 'brain drain' effect.

Last, I fear that entrusting all hope of development to the ambiguous political process of EU membership may undermine Turkey's existing — albeit weak — resolution to advance science.

The country needs a firm political resolution to implement long-term public investments in education and science, regardless of EU membership negotiations.

Mehmet Somel
Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany

Flu virus will not be sent in the regular US mail

SIR — The headline and photographs of your News story "Deadly flu virus can be sent through the mail" (*Nature* **438**, 134–135; 2005) are misleading with respect to the policy of the Centers for Disease Control and Prevention (CDC) regarding the transfer and use of the 1918 pandemic influenza virus. They could give the erroneous impression that the virus will be made widely available and sent through the regular US mail.

The CDC has not yet received any requests to work with the 1918 virus at a non-CDC facility and I have made it clear we currently have no plans to send the virus anywhere. Any requests we do receive will be considered on a case-by-case basis, taking into account scientific merit, biosafety and biosecurity concerns, as well as any additional standards deemed appropriate for this particular virus.

The CDC is the only agency that currently possesses this virus, and we have a special responsibility to balance the importance of scientific progress and collaboration with the moral and scientific imperatives of biosafety and biosecurity.

Julie Louise Gerberding
Centers for Disease Control and Prevention, 1600 Clifton Road Northeast, Atlanta, Georgia 30333, USA

BOOKS & ARTS

Physics ain't what it used to be

Science is venturing into areas where experimental verification simply isn't possible.

The Cosmic Landscape: String Theory and the Illusion of Intelligent Design
by Leonard Susskind
Little Brown: 2005. 416 pp. $24.95

George Ellis

Once upon a time, physics dealt with tangible objects — if you couldn't weigh them or smash them together, at least you could observe them. As times changed, physicists started to deal with more ethereal things: electromagnetic fields and space-time metrics, for example. You couldn't see them but you could measure their influence on particle trajectories and so justifiably claim evidence of their existence. Nowadays things have changed. A phalanx of heavyweight physicists and cosmologists are claiming to prove the existence of other expanding universe domains even though there is no chance of observing them, nor any possibility of testing their supposed nature except in the most tenuous, indirect way.

You gotta have faith: the idea of countless 'pocket universes' cannot be tested, so is it science?

How can this be a scientific proposal, when the core of science is testing theories against the evidence? In *The Cosmic Landscape*, Leonard Susskind argues that we should accept the reality of such universe domains on the basis of two theoretical elements that, taken together, could provide a solution to two major scientific conundrums. The first puzzle is the anthropic issue: the "apparent miracles of physics and cosmology" that make our existence possible. Many aspects of both physics and cosmology seem to be fine-tuned in such a way as to allow chemistry to function, planets to exist, and life to come into being. If they were substantially different, no life at all, and so no processes of darwinian evolution, would have occurred. Which particular aspect of this fine-tuning seems the most significant depends on one's discipline.

Susskind, a particle physicist, thinks the most important is the issue of the cosmological constant, relating to a universal repulsive force that acts on all matter. But this leads to the second conundrum: simple estimates suggest that this constant should be 120 orders of magnitude larger than recently observed. This is a major crisis for quantum field theory, which underlies these estimates. The link to the anthropic question is that if the constant were only twice as large, there would be no galaxies, stars, planets or life. The observed very small value of this constant, although contrary to our present theory of the quantum vacuum, is a necessary condition for our existence.

The first part of the proposed solution is the idea of a 'multiverse' — the existence of a huge number of 'pocket universes', like the vast expanding Universe domain we see around us, that are part of a much larger physical existence. These are supposed to arise through inflation, a process of extremely short-lived, very rapidly accelerating expansion that preceded the hot Big Bang era in the early Universe. 'Chaotic inflation' occurs if inflation is still occurring in distant domains around us today, forming overall a fractal-like structure of inflating domains and pocket universes.

The second part of the solution is the landscape of possibilities, a recent discovery in string theory, which is itself a proposed theory of fundamental physics that unites gravity with quantum physics. It has been suggested that the 'vacuum' of string theory is a structure of immensely complex possibilities, with each possible vacuum resulting in a different kind of local physics; for example, all possible values of the cosmological constant will occur in the different vacua of string theory. If we suppose that the pocket universes of chaotic inflation correspond to different vacua, then all possible kinds of local physics occur at different locations somewhere in the multiverse. If enough combinations of possibilities are realized in this way, then the incredibly special conditions for life to exist will inevitably occur somewhere in the multiverse. The apparent design of conditions favourable to life in our own universe domain can therefore be explained in a naturalistic way.

This is an intriguing picture that unites quite disparate elements of physics and cosmology in a synthesis that is satisfying in many ways. But the question here is whether it is a scientific proposal, as there is no chance whatsoever of observationally verifying its main prediction, the existence of numerous other expanding universe domains beyond our visual horizon. We might hope to base our prediction that the multiverse exists on the fact it is an inevitable outcome of well established physics, but the physics underlying the proposal is hypothetical, rather than established. String theory is neither well defined nor experimentally proven, despite the energy and enthusiasm of its proponents, and there are alternative theories. The inflation field has not been uniquely identified in physical terms, much less shown to have the properties supposed in chaotic inflation.

We might hope to detect the multiverse

indirectly by observing the remnants of the physical processes that underlie its existence; for example, the low value of the cosmological constant today could be such a hint. The problem here is that a multiverse proposal cannot in general be disproved this way, because if all possibilities exist somewhere in the multiverse, as some claim, then it can explain any observations, whatever they are. For example, no observations of anisotropy in the cosmic background radiation can disprove the multiverse hypothesis because all possible anisotropies will be generated in the different expanding universe domains; you just have to live in the right one.

The particular multiverse version proposed by Susskind, however, has the great virtue of being testable in one respect. It is supposed to have started out by quantum tunnelling, resulting in a spatially homogenous and isotropic universe with negative spatial curvature, and hence with a total density parameter $\Omega_0 < 1$. The best observationally determined value for this parameter, taking all the data into account, is $\Omega_0 = 1.02 \pm 0.02$. Taken at face value, this seems to contradict the proposed theory. But given the statistical uncertainties, the observations do not definitively exclude $\Omega_0 < 1$, so the theory survives; nevertheless, the observed value should be taken seriously in this era of 'precision cosmology'. These data are not discussed in the book — a symptom of some present-day cosmology, where faith in theory tends to trump evidence. Presumably the hope is that this observational result will go away as more evidence is collected.

The Cosmic Landscape is extremely well written, provides an excellent non-technical overview of the relevant physics, and tackles important questions in a lively way. However, it confuses the event horizon in the expanding universe with particle and visual horizons. In addition, like many multiverse writings, it uses the concept of infinity with gay abandon, when there is good reason — as pointed out by mathematician David Hilbert — to claim that it is not a good physical concept. The book also tries to justify the multiverse idea in terms of the 'many worlds' interpretation of quantum theory — an unproven and totally profligate viewpoint that many find difficult to take seriously.

As a philosophical proposal, the multiverse idea is interesting and has considerable merit. The challenge facing cosmologists now is how to put on a sound basis the attempts to push science beyond the boundary where verification is possible — and what label to attach to the resultant theories. Physicists indulging in this kind of speculation sometimes denigrate philosophers of science, but they themselves do not yet have rigorous criteria to offer for proof of physical existence. This is what is needed to make this area solid science, rather than speculation. Until then, the multiverse situation seems to fit St Paul's description: "Faith is the substance of things hoped for, the evidence of things not seen." In this case, it is faith that enormous extrapolations from tested physics are correct; hope that correct hints as to the way things really are have been identified from all the possibilities, and that the present marginal evidence to the contrary will go away. This book gives a great overview of this important terrain, as seen from an enthusiast's viewpoint. ■

George Ellis is in the Department of Mathematics, University of Cape Town, Cape Town 7701, South Africa.

Pet project

The Dog and Its Genome
edited by Elaine A. Ostrander, Urs Giger & Kerstin Lindblad-Toh
Cold Spring Harbor Laboratory Press: 2005. 584 pp. $135, £80

Stephen J. O'Brien

Genome technology has found its way into the living room with the completion of the whole-genome sequence of the domestic dog *Canis familiaris*, from a female boxer called Tasha. Finished just a year after its initiation in 2003, the remarkably complete sequence (representing an estimated 99% of the dog's 2.4 billion base pairs) achieves 7.5-fold coverage of the genome and is a major advance over the 1.5-fold sequence of a poodle published by Celera in 2003. The dog is now a front-line model for the discovery of disease genes, for gene annotation, and for probing the evolutionary roots of our mammalian origins. *The Dog and Its Genome*, edited by Elaine Ostrander, Urs Giger and Kerstin Lindblad-Toh, celebrates the completion of the dog sequence with 26 chapters on the genomic biology of man's best friend.

The book should appeal to dog fanciers, to genome biologists who wonder about the sequence's applications, and to students of comparative genomics. It presents well written and concise discussions of the history of dog breeds — there are generally estimated to be between 350 and 1,000, of which the American Kennel Club recognizes about 150 that do not exchange genes. As many as 20 breeds were developed by 1750, increasing to 76 by 1905. Yet the domestication of dogs can be traced back 14,000 years on the evidence of archaeological remains, maybe even 40,000 years based on molecular comparisons with wolves. Clearly, dogs are the oldest domesticated species, as detailed in two of the book's chapters, and the phylogenetic ancestry of dog breeds is described in three chapters.

Years from now, as dog genomics matures, this volume will be remembered as the starting point, with vivid pieces on the vast phenotypic variation described for dogs. The latest interpretation of dog genome status is presented for experts and aficionados alike. The remarkable history of inbreeding has led to a mosaic genome of alternating homozygous and heterozygous/polymorphic segments specific for each breed; these are particularly useful for linkage disequilibrium-based association mapping of complex or multifactorial traits.

Boxer tricks: Tasha's genome will help researchers to understand human genetic diseases.

And dogs certainly have complex traits, notably the vast morphologic variation found in dog breeds as disparate as the chihuahua and the great dane. Dogs also have hard-wired behavioural acumen that allow them to herd livestock, locate missing persons and even sniff out human cancers at early stages. And of course they are loving companions like no other animals.

Generations of veterinary clinicians have identified nearly 500 human hereditary disease homologues in dogs, nearly all breed-specific; the 50 reviewed here have a confirmed genetic basis. Several have been treated successfully with futuristic gene-therapy protocols that should whet the appetite of the medical community. The book describes a cancer registry that documents the incidence and pathologies of a dozen neoplasms that account for 23% of deaths in the 65 million pet dogs in the United States. The challenge now will be to use the genome to detail the genetic bases of behaviours, morphological breed distinctiveness and the disposition of breed-specific cancers.

Researchers already have 'bibles' that define gene-based phenotypes suitable for interrogation by mouse, rat, fruitfly and human genetics. *The Dog and Its Genome* does the same for the canine genomics community. It should be consumed by researchers and their students quickly before forthcoming advances render it dated on their bookshelves. ■

Stephen J. O'Brien is in the Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland 21702-1201, USA.

On top of the world

Nanga Parbat in Pakistan, the ninth highest mountain in the world, is one of 90 or so mountains included in *Mountains From Space: Peaks and Ranges of the Seven Continents*. This striking book, which contains photos taken from space, is published by the German Aerospace Centre (DLR) and Harry N. Abrams, and is available in German and English editions.

Taking its name from the Sanskrit for 'sacred mountain', Nanga Parbat is 8,125 metres high and has cost many mountaineers their lives, including Günther Messner, whose brother and climbing partner Reinhold came up with the idea of the book. Reinhold Messner is among several authors who provide essays from their perspectives as mountaineers or scientists.

This picture was taken by the SPOT-5 Earth-observation satellite. A.A.

EXHIBITION

A close look at Darwin

Darwin
American Museum of Natural History, New York, until 29 May 2006.
www.amnh.org/exhibitions/darwin

Alan Packer

The American Museum of Natural History in New York bills its new exhibition, *Darwin*, as the most in-depth ever mounted on Charles Darwin's life and thought. It's also well timed, coming as it does in the midst of litigation over 'intelligent design' in Dover, Pennsylvania, and in the run-up to the bicentennial of Darwin's birth in 2009. All that aside, *Darwin* is splendid: evolutionary biologist Niles Eldredge's exhibition takes us on a fascinating tour through the life of a great thinker, in what is a superb example of the curator's art.

Visitors are greeted at the entrance by a live, and somehow mesmerizing, giant tortoise from the Galapagos (it can also be viewed remotely via webcam at the exhibition's website). They then encounter Darwin's magnifying glass. This serves as an iconic image — throughout the exhibition, a magnifying glass is positioned to allow the viewer a closer look at a specimen, symbolizing the overall theme of Darwin's lifelong devotion to close observation of nature. His theoretical conclusions, which are well explained in the exhibition, rest on a mountain of evidence that he saw with his own eyes. That evidence is presented here in abundance.

There are many noteworthy items on display, including original correspondence, specimens from Darwin's own collection, original notebooks, and pressed plants from the voyage of the *Beagle*. The writing box that belonged to his daughter Annie is included here (she died at the age of 10). So too is an amusingly exhaustive questionnaire that Darwin sent to 'gentleman farmers', enquiring about their experiences with artificial breeding. The atmosphere is congenial: sounds of ocean life are heard in the section devoted to Darwin's voyage on the *Beagle*; there is a 'condensed-time' video of the 'Sandwalk' footpath around Darwin's home and workplace, Down House in Kent, UK; and in the exit room, a voiceover of the final words from Darwin's book *On The Origin of Species* ushers you through a collection of orchids.

"Believing is easy, and knowing is hard, and it's knowing that matters most," wrote Neil Patterson in his introduction to cell biologist Christian de Duve's recent book *Singularities* (Cambridge University Press, 2005). Darwin removed the nebulous idea of belief from the discussion. In explaining what we know about the theory of evolution and its originator, given the limitations of what an exhibition can convey, *Darwin* could hardly be bettered.

When the *Darwin* exhibition closes at the American Museum of Natural History in New York on 29 May 2006, it will travel to the Museum of Science in Boston, The Field Museum in Chicago, the Royal Ontario Museum in Toronto and the Natural History Museum in London. ■

Among other exhibits, visitors to *Darwin* can see a cast of a glyptodont skeleton found by Darwin.

Alan Packer is senior editor at *Nature Genetics*.