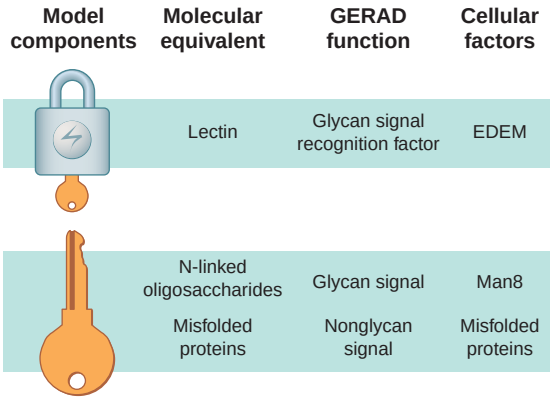


The key to quality control. During protein quality control, misfolded glycoproteins are retained in the ER and modified with Man8 glycans formed by the enzyme ER mannosidase I. This generates a GERAD signal resulting in transport of the modified aberrant glycoproteins out of the ER and their degradation by the proteasome (*5*). The Man8 glycans can be thought of as the shaft of a key, and the misfolded polypeptides of the aberrant glycoproteins as the key handle. Both the key handle and shaft must be present to initiate GERAD. A noncatalytic homolog of ER mannosidase I called EDEM promotes release of misfolded glycoproteins from calnexin and boosts their degradation (*1*, *2*). In this model, EDEM is the lock.

dase I, assisting this enzyme in the removal of noncompliant glycoproteins. Modification of glycoproteins by ER mannosidase I can be likened to a key and lock in which the Man8 glycans form the key shaft and EDEM the lock (see the figure). Questions that remain to be explored include whether, or how, EDEM contributes to the recognition of misfolded protein structure (the key handle). Regardless of the model, ER mannosidase I and EDEM are partners in the partitioning of newly synthesized glycoproteins between the folding and disposal pathways.

EDEM's contribution is particularly interesting because it binds to the carboxyl-terminal tail of calnexin. As Oda *et al.* demonstrate, the interaction of EDEM with calnexin promotes the release of misfolded α_1-antitrypsin from calnexin and accelerates its degradation (*2*). This observation raises the possibility that calnexin, by interacting with EDEM, actively cooperates in the degradation of defective glycoproteins at least in some cell types (*5*). Alternatively, EDEM's association with calnexin might provide newly synthesized glycoproteins with a salvage folding pathway that helps them to acquire their correct conformation. Because mammalian EDEM is part of the unfolded glycoprotein response (*8*), it is fun to speculate that its interaction with calnexin might be part of the normal stress response. Production of EDEM in response to stress could promote cell recovery by boosting degradation of terminally misfolded glycoproteins. In support of this notion, expression of EDEM has been shown to be necessary for degradation of misfolded glycoproteins in mammalian cells (*11*).

The new findings have broad implications because most physiological systems are, at their core, protein-driven processes. It is currently popular to investigate latent cellular responses to the accumulation of undegraded aberrant proteins during disease pathogenesis. However, it should be noted that degradation is the initial cellular response to protein misfolding. Whether, and how, either EDEM or calnexin might contribute to the broad spectrum of severity observed in certain diseases will be an exciting avenue for future study.

References

1. M. Molinari *et al.*, *Science* **299**, 1397 (2003).
2. Y. Oda *et al.*, *Science* **299**, 1394 (2003).
3. R. Sifers *et al.*, *Mol. Biol. Med.* **6**, 127 (1989).
4. L. Yang *et al.*, *Nature Med.* **9**, 3 (2003).
5. C. Cabral *et al.*, *Trends Biochem. Sci.* **26**, 619 (2001).
6. T. Sommer, D. Wolf, *FASEB J.* **11**, 1227 (1997).
7. A. Helenius, M. Aebi, *Science* **291**, 2364 (2001).
8. N. Hosokawa *et al.*, *EMBO Rep.* **2**, 415 (2001).
9. Y. Yoshida *et al.*, *Nature* **418**, 438 (2002).
10. C. Jacob *et al.*, *EMBO Rep.* **2**, 423 (2001).
11. H. Yoshida *et al.*, *Dev. Cell* **4**, 265 (2003).

HUMAN GENETICS

Primate Shadow Play

Richard A. Gibbs and David L. Nelson

Programs for the large-scale DNA sequencing of animal and plant genomes seem to be perpetually at a crossroads. With completion of the genome sequencing of human, mouse, rat, several fish and smaller model species, the question arises regarding which organisms should be analyzed next. Different characteristics (including experimental and economic relevance) make other creatures attractive candidates for genome sequencing, but even these criteria generate a rather short list. A more compelling argument is to distribute sequencing efforts around the tree of life in order to maximize the discovery of conserved coding sequences (exons) and regulatory elements. On page 1391 of this issue, Rubin and colleagues (*1*) present data from their sequencing of select genome regions of multiple primate species closely related to the human. They use these data in a method called "phylogenetic shadowing" that differs from previous cross-species genomic comparisons and works very effectively to reveal coding and regulatory regions in the human genome. Their work argues for prioritization of the genome sequencing of animals that are closely related to us.

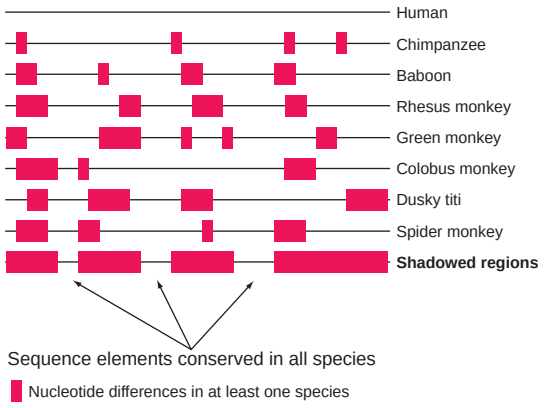
The premise of cross-species genomic discovery is that "what is important is conserved." The basic techniques of cross-species genomic comparison (pioneered long before genome-scale DNA sequencing was possible) and the ability to cross-hybridize DNA probes among species have been widely used to demonstrate the presence of coding regions in the human genome. The emergence of larger amounts of DNA sequence information from distant species has dramatically advanced the value of these techniques, because in silico analyses can define conserved regulatory elements in the genome with high base specificity. Key studies have shown that gene sequences conserved between human and mouse retained their capacity for tissue-specific expression when reconstructed in appropriate cell types (*2*). Ansari-Lari *et al.* (*3*) found new genes and exons with this approach but also observed large numbers of DNA sequence alignments between mouse and human that were noncoding and apparently nonfunctional. The mouse draft genome sequence (*4*) reveals that these alignments sum to a total of about 40% of the mouse genome, and their ubiquity has unfortunate practical consequences. Even with the excellent software now available (*5*), the signal from regulatory regions of the genome can be masked by the noise from sequences that are shared but are of no apparent importance. Recent studies from Eric Green's group show that further calibration of the phylogenetic distance of pairs of species can improve cross-species comparisons, but even multiple pairs spaced far apart do not completely overcome the caveats described above (*6*).

Rubin and co-workers (*1*) now offer a refreshing variation on the basic principle: "What is important is conserved." Their method of phylogenetic shadowing, which adds to the repertoire of methods for cross-species sequence comparisons, can be simply restated as "what is not critical can vary—at least some of the time." This inverted view requires a very different set of

The authors are in the Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA. E-mail: agibbs@bcm.tmc.edu

data where many closely related species are sampled, rather than pairs of evolutionarily distant species. Their work describes the phylogenetic shadowing of 17 primate species closely related to *Homo sapiens*, spanning 40 million years of evolution. In this method, sequences of closely related species are compared taking into account the phylogenetic relationships of the species analyzed.

Close examination of the sequence differences among these primate species revealed that although similarity is the rule, unerring conservation is the exception. Summing these exceptions reveals that the coding exons (as expected) as well as multiple regions smaller than typical exons (which may be regulatory elements) are highly conserved (see the figure). To aid the analysis of their primate sequence collection, Rubin's group developed a probabilistic model based on alternative assumptions of evolutionary rates. This model identified the boundaries of those sequences that are most conserved most of the time (see the figure). The authors experimentally analyzed several of these candidate regulatory regions with protein binding tests and gene reporter assays. They found binding of the predicted DNA regulatory sequences to nuclear proteins and enhanced transcription in reporter constructs.



Primates in shadowland. Phylogenetic shadowing enables multiple comparisons among DNA sequences from closely related primate species including human (*1*, *7*). In this way, the least variable regions of the genome, which should include exons and regulatory elements, can be identified.

These data validated their computational predictions and reinforced the underlying rationale for examining several close human relatives instead of just a few distant ones.

At least part of the reason for the success of phylogenetic shadowing is that the sum of the evolutionary distances spanned by several close relatives is as great as that between two distant species. But does this mean that we need to completely sequence the genomes of 17 different primates to gain all this knowledge? The eventual answer may be yes, if we are to get the full benefit of this approach. In the meantime, happily, much of the benefit comes from four to six close relatives of the human. With the readouts of chimpanzee DNA sequences accumulating and the complete sequencing of other primate genomes under discussion, we may be close to generating a basic data set that complements the genome sequencing of our more evolutionarily distant relatives.

Generating the data needed for human phylogenetic shadowing has other potential benefits. Sequencing a collection of closely related primate species could yield a better appreciation for the range of DNA sequence alterations that take place during speciation. These are expected to be a combination of regulatory, structural, and functional changes. Sorting out the contributions of each type of alteration will best be accomplished by both broad and deep comparisons of the genomes of numerous closely related species. Thus, Rubin and colleagues present us with a way forward for the large-scale sequencing projects that will enhance short-term goals to identify gene control elements, as well as long-term aims to understand overall differences among species.

References and Notes

1. D. Boffelli *et al.*, *Science* **299**, 1391 (2003).
2. R. C. Hardison, J. Oeltjen, W. Miller, *Genome Res.* **7**, 959 (1997).
3. M. A. Ansari-Lari *et al.*, *Genome Res.* **8**, 29 (1998).
4. Mouse Genome Sequencing Consortium, *Nature* **420**, 520 (2002).
5. K. A. Frazer *et al.*, *Genome Res.* **13**, 1 (2003).
6. E. Green, personal communication.
7. www.sciencemag.org/cgi/content/full/299/5611/1331/DC1.

ASTRONOMY

A Preposterous Universe

Alejandro Gangui

For centuries, astronomers have wondered how the galaxies and large-scale structures in our universe were formed. In the second half of the 20th century, cosmologists realized that these events had a witness: a hot bath of light, now cooled to a few kelvin above absolute zero, which is the afterglow of the big bang. The sky-pervading cosmic microwave background (CMB) radiation (*1*) was released just before matter began to get structured. About 10 years ago, tiny variations discovered in its effective temperature (*2*) provided information on the size of the primordial seeds that led to the nascent galaxies, after eons of gravitational evolution.

The author is in the Physics Department, University of Buenos Aires, and at the Institute for Astronomy and Space Physics/CONICET, Ciudad Universitaria, 1428 Buenos Aires, Argentina. E-mail: gangui@df.uba.ar

Now, another piece of evidence shows how these primordial seeds were moving some 400,000 years after the big bang. With a radio telescope at the South Pole, scientists from the DASI collaboration (*3*, *4*) have measured the minute level of orientation, or polarization, that these microwaves received when they emerged from the seething plasma—a signal that only the peculiar dynamics of the seeds present at that epoch can generate (*5*, *6*).

Most light around us is unpolarized. Its many individual waves oscillate in different planes as it propagates. But unpolarized light becomes polarized whenever it is scattered or reflected, as in sunglasses or in the surface of a lake. In these cases, most of the intensity of the scattered light is concentrated in one plane along the line of propagation, resulting in linearly polarized light.

Early on, when the universe was hot enough, matter was ionized and the free electron density was so high that photons could not propagate freely without colliding with electrons. But as the universe expanded and the ambient temperature decreased, the energetic collisions became less frequent. The relatively low-energy photons that ensued could not destroy the increasing number of neutral particles (essentially hydrogen and helium) that began to form through combination of protons, neutrons, and electrons. Soon after this "recombination" period, the CMB was released. According to theory, it is at this precise time, nearly 14 billion years ago, that the CMB became polarized.

CMB polarization was first proposed 35 years ago by Rees (*7*). However, there was no evidence of its existence until the DASI detection late last year. Polarization is an important probe for cosmological models and for the more recent history of our nearby universe. It arises from the interaction of the cosmic background radiation with free electrons; hence, CMB polarization can only be produced at the time of its last scattering, because afterwards no free electrons exist. Unlike temperature fluctuations, polarization is largely unaffected by inhomogeneities in the growing