about global records of species-level changes.

J. David Archibald
Department of Biology, San Diego State University, San Diego CA 92182–4614, USA

References

1. J. D. Archibald, *Dinosaur Extinction and the End of an Era: What the Fossils Say* (Columbia Univ. Press, New York, 1996).

■

Human Genome Project: Data Quality

Steven E. Koonin's interdisciplinary perspective on the Human Genome Project (*Science*'s Compass, 2 Jan., p. 36) is intriguing, but I would like to comment on two key issues regarding data quality.

Koonin's proposal that quality monitoring be based on "gold standard" sequences does not meet the condition that the laboratory under review should be blind as to which of its sequences will be checked. Consequently, it is unlikely to yield representative results. A preferable procedure, which is already in use for sequencing centers funded by the National Institutes of Health (*1*), is to randomly select previously sequenced and submitted clones for resequencing by one or more competing laboratories, which would be motivated to find as many errors as possible.

The issue of setting cost-effective quality standards is more difficult, and Koonin is correct in observing that it is important to be quantitative. The Human Genome Project's target error rate of less than 1 in 10,000 at the per base level (*1*) is a pragmatic one, chosen because it is both attainable with current technology at a reasonable cost (currently about 50 cents per base pair) and adequate for essentially all known uses of the sequence. A higher error rate would tend to degrade our ability to use the sequence as a reference against which human genetic variation can be catalogued and would also tend to complicate gene-finding efforts.

Studying this issue using computer simulation, as Koonin proposes, is desirable in principle, but is impractical given our current limited knowledge. Biological features of unsuspected types likely exist even in well-studied sequences and cannot be modeled before they are discovered. Simulating the laboratory sequencing process, particularly those aspects that involve biological organisms (bacteria or humans), similarly appears beyond the state of the art.

The Human Genome Project has had a highly interdisciplinary character since its inception, and will continue to benefit from expert advice in order to meet the unprecedented challenges that still remain. Such input is most productive when biological and physical principles are effectively merged.

Phil Green
Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195, USA
E-mail: phg@u.washington.edu

References and Notes

1. Further comments on the full JASON study led by Koonin and reported on in his article may be found at http://bozeman.genome.washington.edu/ScienceLetter/JasonComments.html
2. "Report of NHGRI Workshop on DNA Sequence Validation" (National Human Genome Research Institute, 15 April 1996) (www.nhgri.nih.gov/HGP/Reports/dna_sequence_workshop.html).

Response: Green offers a number of thoughtful comments on the JASON report. I respond below to those contained in his letter:

1) The JASON study enumerates a number of nonexclusive quality-assessment strategies, including both competitive random resequencing and "gold standards." The report discusses the drawbacks of the latter that Green mentions, but still concludes that gold standards would be useful in that, be-

yond the sequence itself, the process by which it is obtained is important. The extent to which it is possible to operate in a "best behavior" mode will be instructive in assessing DNA sequencing performance. At the very least, such trials will establish a lower limit to the error rate expected. Given that gold standards would imply only a minor increment to the resequencing load, it would seem prudent to include them, even if, as some claim, "one already knows the answer."

2) Green's use of the imprecise phrase "tend to" in his third paragraph illustrates the quality standard point we wished to make. A quantitative study of how scientific utility and cost vary with sequence accuracy would seem essential to setting the specifications for the project and does not seem an impossible task.

3) Computer simulation has proven to be an extraordinarily useful tool to understand and optimize all manner of complex natural and artificial systems; the sequencing process should be no different. Useful simulations require that one understands the distribution and correlation of errors. Acquiring such understanding requires careful and targeted process experiments, which are quite distinct from production sequencing. The report advocates that the time and resources be expended to do these experiments, so that high-fidelity simulations can be constructed and exploited.

As Green notes, fusing the languages and approaches of different disciplines toward a common challenge is often a separate challenge of its own. This latter can be met only by a continuing dialog in which all parties are both students and teachers.

Steven E. Koonin
Division of Physics, Mathematics, and Astronomy,
California Institute of Technology,
Pasadena, CA 91125, USA
E-mail: koonin@caltech.edu

■

α-Synuclein Gene and Parkinson's Disease

Mihael H. Polymeropoulos *et al.* (Reports, 27 June 1997, p. 2045) (*1*) reported an Ala53Thr mutation in the α-synuclein gene, at the PD-1 locus on chromosome 4q21-q23 (*2*), in a large Italian family and three smaller Greek kindreds with autosomal dominant Parkinson's disease (PD). The phenotype in the Italian family, however, differs slightly from idiopathic Parkinson's disease, with earlier onset and a more rapid disease progression (*3*), and linkage to this locus has been excluded in most of the tested families (Technical Comments, 18 July, pp. 387 and 388; Letters, 14 Nov., p. 1212; corrections, 5 Dec., p. 1696) (*4*). These studies (*4*), however, provide only indirect evidence that the PD-1 locus might rarely be involved in familial Parkinson's disease. The frequency of α-synuclein gene mutations in dominant Parkinson's disease remains to be determined.

We have examined the entire coding sequence of the α-synuclein gene in 25 families with dominant Parkinsonism. Families were selected with the use of the following criteria: (i) presence of at least two affected first-degree relatives in two successive generations; (ii) definite Parkinson's disease characterized by two of four essential signs (bradykinesia, rigidity, resting tremor, or asymmetrical onset); (iii) marked improvement of symptoms after administration of levodopa, except for two de novo cases; and (iv) absence of the exclusion criteria (supranuclear ophthalmoplegia, cerebellar or pyramidal signs, apraxia, severe autonomic or postural disturbance, or dementia within 2 years of onset).

Twenty-four of these families were French, and one was Italian. There were 13 men and 12 women among the index cases. The mean age at onset of the index cases was 54 ± 16 years (range, 28 to 77 years). We did not observe severe dementia in any patient. After DNA extraction from lymphoblasts, the entire coding region of the longest form of the α-synuclein gene, which encodes a protein of 140 amino acids, was amplified by reverse transcriptase polymerase chain reaction and directly sequenced as previously described (*5*). We did not find any mutations or polymorphisms in the index cases of any family. Vaughn *et al.* (*6*) have reported not finding the Ala53Thr mutation in 230 cases of familial Parkinson's disease, but they searched only for this specific mutation. We have confirmed that the Ala53Thr mutation is rare in Parkinson's disease. We also found—in a large series of families with dominant Parkinson's disease—no other mutations anywhere in the coding sequence of the α-synuclein gene. Therefore, unless mutations are present in other, noncoding regions of the α-synuclein gene (that is, in promoter or regulatory sequences), PD-1 could only represent, at best, a minor locus for dominant Parkinson's disease in our study population.

The French Parkinson's Disease Genetics Study Group (7)
c/o Alexis Brice,
Institut National de la Santé et de la Recherche Médicale (INSERM) U289,
Hôpital de la Salpêtrière,
47 boulevard de l'Hôpital,
75651 Paris Cedex 13, France
E-mail: brice@moka.ccr.jussieu.fr