CORRESPONDENCE

Reality of sequencing costs

SIR — In their interesting article on the *Caenorhabditis elegans* genome sequencing project, J. Sulston *et al.* (*Nature* **356**, 37–41; 1992) anticipate ". . . the cost, currently estimated at $1 per base with current methods. . ." of their strategy. How did they arrive at this estimate, and what is included in estimating the cost?

Let me do the following calculation for such a project in Germany. Two fluorescence sequencers (paid off over 5 years) cost about DM80,000 per year, 600 primers synthesized (length 15, DM5–8 per base) is more than DM45,000 for 120,000 base pairs (DM0.375 per bp). DNA preparations, chemicals, enzymes, gels and so on come to at least DM15,000 for 1,500 clones (DM0.12 per bp). If, say, one-third of the authors' time was spent on laboratory work such as cloning, DNA preparation, sequencing and editing (DM75,000 × 19/3) the cost for salaries would be DM475,000 per year. Rent and running cost for a laboratory of 120 people will be about DM30,000 per year. There will be additional costs for laboratory equipment, repair, laser-tube replacement, break-down time, overheads and so on, often neglected by scientists.

With such rather optimistic numbers I arrive at minimal costs of DM645,000 for 120,000 bp per year, or $3.00 per bp of final sequence. For 800,00 bp per year with full-time employment and 20 per cent overhead, the costs triple to DM2 million and will be close to $2.00 per bp.

I founded a sequencing company in 1990, in which about 600,000 bases were read for assembly during 6 months from two cosmids by using direct-blotting electrophoresis and colorimetric detection of digoxigenin. The actual costs were close to the above estimate. The costs do not include profits and the work was done by dedicated people. This is in accord with the present funding of the yeast project — requiring very high quality for the sequence data — by the European Commission with 2 ECU per bp, which is close to $3 per bp.

But it still appears to be a very long way before costs of $1 or $0.50 per base pair will be a reality. One should be very careful with such estimates, especially in a scientific paper, because they may have far-reaching consequences for biological research.

Fritz Pohl
Fakultät fur Biologie,
Universität Konstanz,
D-7750 Konstanz, 1 Germany

SULSTON REPLIES — I agree with Pohl's essential point that estimates of DNA sequencing costs must be all inclusive and made with care. Unfortunately, in an effort to shorten the paper and in response to a referee's opinion that cost issues would not be of broad interest, we removed the detailed discussion that was in our original manuscript.

A cost of $1 per base pair of finished sequence is the current production cost, not the actual cost of sequencing the three cosmids; the latter was indeed higher because of time spent in development. We have now settled down to a production routine, and are able to estimate our actual costs reasonably confidently. In so doing, we have allowed for all the hidden extras (such as fringe benefits, supervision and replacement of minor equipment) that, as Pohl points out, are so easily overlooked, and a 50 per cent overhead for rent and running costs of the laboratory space.

The discrepancy between our estimate and that of Pohl is due to: (1) heavier use of the ABI 373A — only two machines per megabase of finished sequencer per annum; (2) in-house synthesis of oligonucleotide primers, of which only 100 are now made per cosmid; and (3) efficient use of staff by job specialization, with appropriate levels of training for each operation.

The pace of sequencing is sustainable, because the members of the groups, though certainly dedicated and flexible in their activities, are not called upon to work excessively long hours for routine production. (Development work, and writing scientific papers, is another matter.)

John Sulston
MRC Laboratory of Molecular Biology,
Hills Rd,
Cambridge CB2 2QH, UK

Law of mass action

SIR — John Maddox (*Nature* **355**, 201; 1992) scolds molecular biologists for their naive reliance on qualitative observations, and calls for a resurrection of the law of mass action to explore crucial quantitative features of molecular regulatory networks.

We agree with the doctor's diagnosis and remedy, but would like to point out that the patient's plight is not as serious as might be inferred. In addition to the "army of people called molecular biologists" who uncover the qualitative structure of molecular control mechanisms, there are squads of theoretical biologists who wield the law of mass action to explore the quantitative implications of these mechanisms. As a small selection of recent studies that illustrate the usefulness of this approach, we draw your attention to models of cyclic AMP metabolism in slime moulds[1], molecular events in neurotransmitter release[2], quantitative triggers of cell division[3], genetic switching in early embryos[4] and interleukin-2 binding to T lymphocytes[5].

Modern science proceeds largely through the efforts of specialists, so that one cannot expect molecular biologists to deviate much from their spectacularly successful efforts. Other excellent biologists study integrative physiology, and a small group of theorists, from biology and other disciplines, labour to create sound and predictive connections between molecular and physiological levels. Communication among these specialists is essential. *Nature* could play a helpful role here by devoting more attention to existing quantitative and integrative research in molecular and cellular biology.

Lee Segel
Department of Applied Mathematics
and Computer Science,
Weizmann Institute of Science,
Rehovot,
76100 Israel
John J. Tyson
Department of Biology,
Virginia Polytechnic Institute
and State University,
Blacksburg, Virginia 24061, USA

1. Martiel, J-L. & Goldbeter, A. *Biophys. J.* **52**, 807–828 (1987).
2. Parnas, H. *et al. Int. Rev. Neurobiol.* **32**, 1–50 (1990).
3. Tyson, J. J. *Proc. natl. Acad. Sci. U.S.A.* **88**, 7328–7332 (1991).
4. Edgar, B. A. *et al. Devl. Genet.* **10**, 124–142 (1989).
5. Goldstein, B. *et al. Int. Immunol.* **4**, 23–32 (1992).

Artists' offspring

SIR — We read with interest the recent letter from R. A. Beck (*Nature* **356**, 189; 1992) suggesting the possibility that artists have more sons than daughters. We became sceptical after calculating that among the sample of artists' offspring 52.8 per cent were sons (1,834 of 3,474), compared to 51.9 per cent sons (2,046 of 4,002) among the sample of children of the general population in *Who's Who*, a very small difference indeed. A quick attempt to confirm the chi-square statistic produced a value of 2.072 (not using the Yates' continuity correction), corresponding to a two-sided P-value of 0.150. Thus an amusing hypothesis (associated, unfortunately, with a non-trivial amount of data compilation) must be put to rest. If one were to repeat this exercise with five other randomly chosen occupations, one would produce at least one chi-square value greater than 2.072, purely by chance, with probability 0.556.

L. V. Rubinstein
A. D. Koutsoukos
Biometric Research Branch,
National Cancer Institute,
Rockville, Maryland 20892, USA