Bacterium genome sequence

SIR — Stephen Oliver in his Progress article champions a systematic approach to the discovery of gene function so as to make best use of the copious data being produced by large-scale genome sequencing projects[1]. He gives cogent arguments for using yeast for this purpose, discusses the problem of functional redundancy and suggests ways of analysing open reading frames revealed by sequencing but having no matches with sequences in databases. Even open reading frames assigned on the basis of similarity to already-known genes can be problematic because of the paucity of functional information in the databases. Of relevance to this are our genomic studies of the photosynthetic bacterium *Rhodobacter capsulatus*, described here. These studies take advantage of this organism's unique properties, which include a simple system for determining the functions of sequenced regions.

R. capsulatus can be grown heterotrophically or phototropically. It metabolizes a wide range of organic compounds and has two independent systems for fixing nitrogen. The genome consists of a single chromosome of 3.7 megabases and a 134-kilobase (kb) plasmid[2]. A high-resolution physical map with more than 3,000 sites for four restriction enzymes contains 250 mapped genes. An overlapping set of 192 cosmids covers the genome completely. Most importantly, it harbours a defective transducing phage, called the gene transfer agent (GTA), which packages 4.6-kb DNA fragments representing all parts of the DNA present in a GTA-producing strain[3]. The GTA can be used to construct deletions efficiently and systematically, an important feature noted by Oliver for yeast.

For *Rhodobacter*, the fragment to be deleted can be replaced with a cartridge encoding resistance to kanamycin or gentamicin, or both (Fig. 1). This method was originally used to delete the entire *nifHDK* region[4]. We have exploited it more recently in attempts to delete whole-cosmid-sized regions of the chromosome. In these first experiments, we used 26 cosmids covering 400 kb of the *Rhodobacter* chromosome. All 26 of the GTA lysates produced Km^r/Gm^r transductants, but with a wide range of morphologies and growth rates on complete medium. About half the cosmids yielded strains that had wild-type growth rates on complete medium. Figure 2 shows two of these deletions. The Southern blots indicate that, for cosmids 2D11 and 2D12, nearly all the chromosomal DNA contained in each cosmid is replaced by the antibiotic-resistance cassette. To date, we have confirmed the complete deletion of 55 kb covered by three overlapping cosmids (2D11, 2D12 and 2E1) and of 25 kb each, covered by 2E11 and 2H3. Maps of these cosmids are shown in ref. 5.

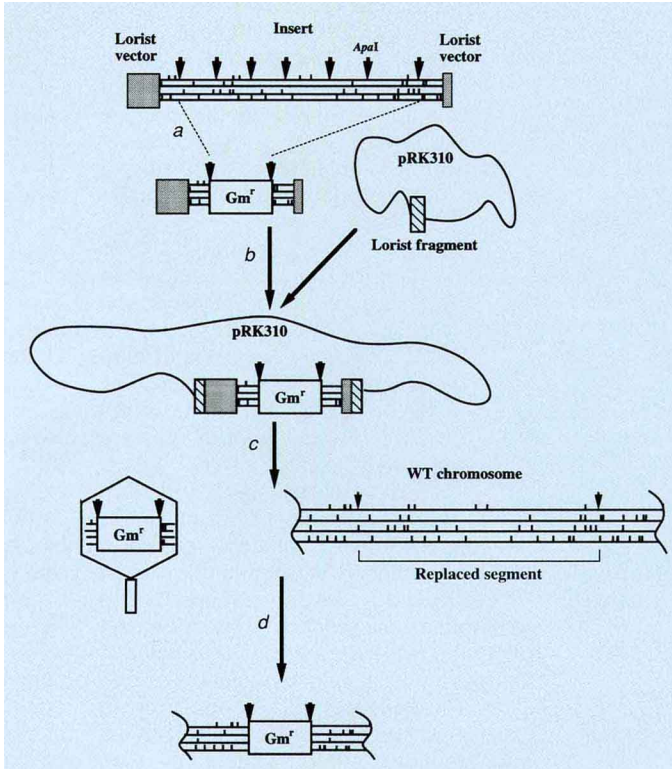

FIG. 1 Construction of deletion strains with the GTA. Cosmids containing 35–40-kb *R. capsulatus* DNA in the vector Lorist 6 are restricted with *Apa*I, which cuts many times in *Rhodobacter* DNA but never in the vector (*a*). The cassette carrying resistance to both gentamicin and kanamycin is ligated into the cut cosmid and used to transform an *Escherichia coli* strain carrying pRK310, which in turn contains a fragment of the Lorist vector, allowing co-integrants to be formed (*b*). This construct is transferred to a GTA-producing strain of *Rhodobacter* by conjugation. Culture supernatants contain some GTA particles that have packaged the resistance cassette flanked by *Rhodobacter* DNA. Adding such supernatants to wild-type *Rhodobacter* (*c*) and plating on antibiotics gives transductants (usually) in which the target sequence is replaced by the resistance cassette (*d*). Selection on two antibiotics reduces the background of spontaneous resistant mutants.

Another group of transductants showed more complex hybridization patterns, to be expected where essential genes are located. In some cases, merozygotes were selected, in which the cassette was inserted in the target sequence but some of the target sequence was maintained. A blot showing this phenomenon for cosmid 1A3 is included in Fig. 2. The complete sequence of this cosmid has been determined (V.K., M.F. and R.H., manuscript in preparation).

Another aspect of analysis considered desirable by Oliver is the study of gene expression, preferably in a way that permits global detection of the consequences of mutation in a regulatory gene. This type of analysis has been demonstrated for *R. capsulatus*, again exploiting the set of overlapping cosmids. The complete cosmid set was digested with *Eco*RV and the resulting fragments from each cosmid, after separation by electrophoresis, were transferred to a small number of blots. Effectively, the entire chromosome was displayed in 192 lanes in gene- or operon-sized pieces, according to their position on the chromosome. This high-resolution hybridization template was then used to detect, for example, the changes in transcription that accompany heat shock[5]. The same template can be used quantitatively to detect changes in transcription in deletion mutants.

Over the past four months, we have completed 250 kb of sequence (GeneBank accession number U57682 for cosmids 143–147). The project is being carried out cosmid by cosmid, so progress can be monitored in terms of gene assignments as the work continues. So far, about half the open reading frames can be assigned with high probability. Another 10–15% contain a recognized functional domain but no clear indication of the full protein's function. The remaining 35–40% open reading frames have no matches in any public database. Sequencing of the *R. capsulatus* chromosome is being carried out without massive financial support. A grant from the US Department of Energy program in Basic Energy Sciences has funded all the mapping and cosmid library construction to date. Part of the sequencing is being done in Prague by V. Paces and C. Vlcek at the Academy

1. Oliver, S. G. *Nature* **379,** 597–600 (1996).
2. Fonstein, M. & Haselkorn, R. *Proc. natn. Acad. Sci. U.S.A.* **90,** 2522–2526 (1993).
3. Yen, H. C. *et al. J. molec. Biol.* **131,** 157–168 (1979).
4. Scolnik, P. A. & Haselkorn, R. *Nature* **307,** 289–292 (1984).
5. Fonstein, M. *et al. EMBO J.* **14,** 1827–1841 (1995).

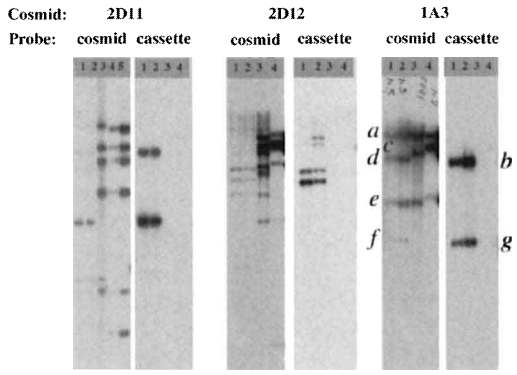

FIG. 2 GTA-mediated deletions in *R. capsulatus*. Each panel contains the following DNA samples restricted with *Eco*RV: lanes 1, 2, total DNA from two independent transductants; lane 3, total DNA from wild-type *R. capsulatus*; lane 4 (and 5 in the first panel), cosmid DNA. For each pair of panels, the hybridization probes were the total cosmid (left) or the antibiotic-resistance cassette (right), which contains a single *Eco*RV site. Discrepancies between fragment sizes in lanes 3 and 4 are due to vector sequences attached to *Rhodobacter* DNA in lane 4. For cosmids 2D11 and 2D12, most of the DNA present in the cosmids has been deleted from the transductants. For cosmid 1A3, fragments *a* (20 kb) and *e* (2 kb) cannot be deleted. Fragment *c* (11.5 kb) is disrupted by the cassette in the transductants, yielding fragments *d* and *f*; these fragments correspond in the right-hand panel to fragments *b* and *g*, which hybridize with the cassette probe, confirming the location of the disruption. The sum of the sizes of fragments *d* and *f* is 11–12 kb. Because the cassette size is 2 kb, the deletion in these transductants is only 1–2 kb. Fragment *c* is close to one end of cosmid 1A3. We believe that the recombination event creating these deletions began in the region of fragment *c* flanking the cassette and that the second crossover occurred in a region of partial homology located nearby, rather than at the far end of the cosmid where homology with the GTA-carried DNA is perfect.

of Sciences of the Czech Republic, with a grant from the NSF for international collaboration.

Vivek Kumar, Michael Fonstein
Robert Haselkorn
Department of Molecular Genetics & Cell Biology,
The University of Chicago,
Chicago, Illinois 60637, USA

OLIVER REPLIES — Kumar and colleagues demonstrate that techniques now exist in a wide range of organisms which will permit the systematic analysis of gene function. Such an analysis is most meaningful in organisms such as *Saccharomyces cerevisiae*, whose genome has been completely sequenced, and it is encouraging to learn that an international effort to sequence the *Rhodobacter capsulatus* genome has begun.

The techniques outlined by Kumar *et al.* for *R. capsulatus* have quite exact analogues in other systems. In *S. cerevisiae*, the complete genome sequence, together with the organism's high efficiency and accuracy of mitotic recombination, is being exploited in a gene replacement strategy involving the polymerase chain reaction (PCR)[1], and resulting in the efficient deletion of individual open reading frames (ORFs).

Following the successes of our European Yeast Genome Sequencing Network[2], we have formed another scientific network (EUROFAN) committed to elucidating the function of novel yeast genes. In EUROFAN, we are exploiting a gene-by-gene deletion strategy, using PCR-mediated replacement and (as with *R. capsulatus*) Gm^r as a replacement marker[3,4]. Some laboratories are also performing more extensive chromosomal deletions using a combinatorial approach[5], which we term "mass murder".

A parallel European network is working on functional analysis in the Gram-positive bacterium *Bacillus subtilis*. This group is exploiting the operon organization common for bacterial genes by using a single plasmid insertion event to (simultaneously) disrupt an ORF, create a *lacZ* fusion with the upstream ORF, and place the downstream ORF under the control of a regulatable promoter (S. D. Ehrlich, personal communication).

As with more conventional approaches to defining gene function, different model organisms will be used for the systematic analysis of particular biological systems. Thus, *R. capsulatus* may be used for an exhaustive genetic analysis of nitrogen fixation or photosynthesis, while *S. cerevisiae* may be used to study functions that are peculiar to eukaryotes and thus not accessible to experimentation with a bacterium.

Within a given organism, it will be important to construct strains which have an improved potential for elucidating the function of novel genes in particular areas of biological activity[6]. More generally, there is a need to develop global approaches to the analysis of gene function, such that data from these model organisms can be exploited rapidly for analysing species that are less genetically malleable, but which have great biological, medical, agricultural or industrial interest. In particular, it will be important to develop strategies to permit the 'functional mapping' of the genomes of such species onto those of the model organisms.

S. G. Oliver
Department of Biochemistry & Applied Molecular Biology,
UMIST, PO Box 88,
Manchester M60 1QD, UK

1. Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. *Nucleic Acids Res.* **21,** 3329–3330 (1993).
2. Vassarotti, A. *et al. J. Biotechnol.* **41,** 131–137 (1995).
3. Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. *Yeast* **10,** 1793–1808 (1994).
4. Wach, A. *Yeast* **12,** 259–265 (1996).
5. Fairhead, C., Llorente, B., Denis, F., Soler, M. & Dujon, B. *Yeast* (in the press).
6. Oliver, S. G. *Nature* **379,** 597–600 (1996).

Contamination of drinking water

SIR — Until now, the availability of organic carbon has been considered the key factor controlling microbial regrowth in drinking-water networks[1,2]. This availability is considered to be a potentially serious problem, especially in boreal regions (northern Europe, Russia and North America) where surface water, and sometimes ground waters, contain high amounts of organic matter[3]. Natural organic matter acts as a substrate for microbial growth and, when water is used for public supply, disinfectants such as chlorine, hypochlorite and ozone are used. In addition to killing microorganisms, these oxidizing agents break down large organic compounds to smaller ones[4,5]. These compounds can be easily decomposed, further increasing microbial growth in the distribution system and thus impairing water quality. Because of this, considerable effort and resources have been deployed to remove the organic contaminants from drinking water. In the United States, for example, current annual expenditure exceeds 5 billion dollars[6].

Finnish drinking waters contain high concentrations of organic carbon (75–640 $\mu g\ C\ l^{-1}$) easily available to microorganisms. But because the available organic carbon correlates negatively with microbial growth, we investigated the possibility that other factors regulate this growth. In experiments with ground and surface waters, addition of phosphate alone to water increases microbial growth (the number of culturable bacteria) to the same extent as does a mixture of added inorganic nutrients (see figure).

The concentration of phosphate needed to enhance microbial growth is extremely low. When we added phosphate ranging from zero to 50 $\mu g\ l^{-1}$ to the processed drinking water supplying five waterworks (three with surface water and two with ground water), we found a hyperbolic relationship between $[PO_4\text{-}P]$ and microbial growth, with a half-saturation value of only 2–3 $\mu g\ PO_4\text{-}P\ l^{-1}$. Investigation of samples from four distribution

1. LeChevallier, M. W., Babcock, T. M. & Lee, R. G. *Appl. envir. Microbiol.* 53, 2714-2724 (1987).
2. van der Kooij, D. *J. Am. Wat. Works Ass.* **84,** 57–66 (1992).
3. Cronan, C. S. in *Organic Acids in Aquatic Ecosystems* (eds Perdue, E. M. & Gjessing, E. T.) 245–260 (Wiley, Chichester, 1990).
4. van der Kooij, D., Hijnen, W. A. M. & Kruithof, J. C. *Ozone Sci. Engng* **11,** 297–311 (1989).
5. Daniel, P., Meyerhofer, P., Zafer, M. & Rice, E. in *Proc. 11th Ozone Wld Congr. Ozone in Water and Wastewater Treatment* Vol. 2 (Int. Ozone Ass., Stanford, 1993).
6. van den Leeden, F., Troise, F. L. & Todd, D. K. *The Water Encyclopedia* 2nd edn (Lewis, London, 1990).
7. American Public Health Association, American Water Works Association & Water Environmental Federation *Standard Methods for the Examination of Water and Waste Water* (eds Grenberg, A. E., Clesceri, L. S. & Eaton, A. D.) (Am. Publ. Hlth Ass., Baltimore, 1992).