To b(gl) or not to b(gl): how cryptic are "cryptic" genes?

A striking feature that has emerged from the genome sequencing projects that have been completed so far is the extremely high coding percentage seen in all bacterial genomes. In the case of *Escherichia coli*, it is higher than 90% (in higher eukaryotes including humans, it can be as low as 2%). Despite running such a tight ship, many prokaryotic genomes carry genes that—it seems—are unable to function at all. Contrary to what one would expect on evolutionary grounds, these "cryptic genes" have nevertheless been retained as a part of the genetic repertoire of the organism (reviewed in Mukerji and Mahadevan 1997). Unlike the pseudogenes that are found in many eukaryotes, cryptic genes can be activated by a single mutational event, often the integration of an IS sequence within the promoter. This means that even though normally their promoters are silent, they have not been converted to pseudogenes by the accumulation of inactivating mutations over evolutionary time.

Several explanations have been put forward to account for their maintenance. One is that the genes are silent only under certain conditions (such as growth in the laboratory) but are expressed under specific "natural" conditions without the benefit of an activating mutation. A recent publication (Khan and Isaacson 1998) provides evidence in favour of this hypothesis in the case of the cryptic *bgl* operon of *E. coli* (which consists of a set of genes that participate in the uptake and catabolism of β-glucoside sugars such as salicin and arbutin).

Most wild type isolates of *E. coli* are unable to ferment arbutin or salicin because the *bgl* genes are silent. Khan and Isaacson (1998) cloned random fragments of *E. coli* genomic DNA upstream of the chloramphenicol acetyltransferase (CAT) reporter gene. A pathogenic strain of *E. coli* carrying such a library was then introduced into mice treated with chloramphenicol. They found that the bacterial cells that survived in the liver of the mouse carried DNA sequences from the *bgl* region, upstream of the CAT gene. The implication is that the *bgl* genes are expressed under pathogenic conditions in bacterial cells that infect the liver. This has been further substantiated by the amplification of *bgl*-specific RNA from the cells by RT-PCR.

These results support the idea that "cryptic" genes can in fact function under specific physiological conditions. The simplest explanation for their maintenance, then, is that positive selection acts on them when they are expressed because—at least in the environment of the mouse liver—they increase the fitness of the bacterial cells that harbour them. This explanation is consistent with a recent observation on yet another "cryptic" operon, the *cel* operon of *E. coli* (hitherto believed to have as its function the catabolism of cellobiose). It turns out that the *cel* operon is inducible in the presence of the disaccharide *N,N'*-diacetylchitobiose, a breakdown product of chitin, brought about by the action of bacterial chitinases (Keyhani and Roseman 1997). In other words, while cellobiose utilization requires mutational activation of the operon, chitobiose is a natural substrate for the system (likely to be found in the gut of insect-eating carnivores) and is catabolized by the products of the operon as it exists.

The results with the *bgl* genes are striking, but they also raise many questions. Firstly, it is not clear what the selective advantage *is* of expressing them during pathogenesis. As far as we can make out, the apparent function of the *bgl* system is to enable the cell to catabolize β-glucosides, and it is difficult to imagine how the bacterium might benefit by expressing the *bgl* genes during infection (there being no reason to suspect the existence of β-glucoside sugars in infected livers). One might hypothesise that the permease encoded by the *bglF* gene, normally responsible for transporting the glucoside substrate, is involved in transport of other (essential) molecules under conditions of stress. However, this explanation is not consistent with the observation of Khan and Isaacson (1998) that, when a strain of *E. coli* carrying a deletion of *bglF* was used in infection experiments, no appreciable difference in pathogenecity could be observed. The second question concerns the apparent breakdown of the tight regulation of the genes *in situ*. Optimal expression by *bgl* in laboratory cultures involves, in addition to activation of the silent promoter, regulation of anti-termination of transcription that is brought about by the presence of the β-glucoside substrate. How is the cell is able to overcome this regulatory hurdle? An unfortunate limitation

of the experiments of Khan and Isaacson (1998) is that under the conditions used by them the expression of the *bgl* genes cannot be quantitated. Therefore it is not possible to compare the levels of expression detected by them with the corresponding levels under fully induced expression in the laboratory (of a strain carrying an activated copy of the gene). Despite these caveats the observations are thought-provoking because they raise the possibility that "cryptic" genes are not, silent relics of an evolutionary past after all—whose retention in the genome is paradoxical—, but are instead an integral part of the genome.

The puzzle of cryptic genes in prokaryotes resembles that of genetic redundancy in higher organisms. Many genes in metazoans exhibit stage and tissue-specific patterns of expression during development but show no effect when mutationally inactivated—or, once again, so it seems (see Cooke *et al* 1997 for a review). The problem of maintaining a redundant gene against mutational pressure is akin to that of retaining a cryptic gene. A number of theories have been advanced to account for "redundancy. Among them is that cells maintain 'back-up' genes for important developmental functions. On the face of it, this does not make sense in terms of what we know about how evolution works. Another way of explaining redundancy would be to say that gene duplication is sometimes followed by the retention of both copies because one of them begins to acquire a new function without entirely losing the old one. Then, under certain conditions, the DNA sequence carrying the new function may be able to substitute for the original copy, thereby conveying an impression of redundancy. But if examined in the right environment, a null mutation at either locus ought to manifest a detectable (if weak) phenotype. Ponte *et al* (1998) have brilliantly substantiated this hypothesis in the case of the cell adhesion molecule csA in the haploid soil amoeba *Dictyostelium discoideum*. Under standard laboratory conditions of development on agar plates, csA-null mutants develop normally. But under conditions that partially mimic the natural 3-dimensional soil environment, not only is their development poor, their fitness is markedly lower than that of the wild-type. The moral is on the lines of Lewis Wolpert's suggestion, which was made in the context of putatively phenotypeless transgenic mice. Bacteria too should be taken out periodically from their culture flasks to a "test evening at the opera" in order to examine whether "cryptic" genes might be functional after all.

References

Cooke J, Nowak M, Boerlijst M and Maynard-Smith J 1997 Evolutionary origins and maintenance of redundant gene expression during metazoan development; *TIG* **13** 360–364

Keyhani N O and Roseman S 1997 Wild type *Escherichia coli* grows on the chitin disaccharide N,N'-diacetylchitobiose by expressing *cel* operon; *Proc. Natl. Acad. Sci. USA* **94** 14367–14371

Khan A M and Isaacson R E 1998 *In vivo* expression of the β-glucoside (*bgl*) operon of *Escherichia coli* occurs in mouse liver; *J. Bacteriol.* **180** 4746–4749

Mukerji M and Mahadevan S 1997 Cryptic genes: evolutionary puzzles; *J. Genet.* **76** 147–159

Ponte E, Bracco E, Faix J and Bozzaro S 1998 Detection of subtle phenotypes: The case of the cell adhesion molecule csA in *Dictyostelium*; *Proc. Natl. Acad. Sci. USA* **95** 9360–9365

S MAHADEVAN
Developmental Biology and Genetics Laboratory,
Department of Molecular Reproduction, Development and Genetics,
Indian Institute of Science,
Bangalore 560 012, India
(Email, mahi@serc.iisc.ernet.in)