Microbiology Comment provides a platform for readers of Microbiology to communicate their personal observations and opinions in a more informal way than through the submission of papers.

Most of us feel, from time to time, that other authors have not acknowledged the work of our own or other groups or have omitted to interpret important aspects of their own data. Perhaps we have observations that, although not sufficient to merit a full paper, add a further dimension to one published by others. In other instances we may have a useful piece of methodology that we would like to share.

The Editors hope that readers will take full advantage of this section and use it to raise matters that hitherto have been confined to a limited audience.

Christopher M. Thomas, Editor-in-chief

Notes on designing a partial genomic database: The PfSBW25 Encyclopaedia, a sequence database for Pseudomonas fluorescens SBW25

Managing and analysing DNA sequence is a major problem for those research programmes that generate large amounts of short-run noncontiguous (SRNC) sequence data as is the case for many genomic-level projects. Data must be accurately collated and archived, it must be accessible and preferably in a form that enables it to be data-mined. The problem of dealing with SRNC data is one that our laboratory has had to contend with as a result of our analysis of the genome of Pseudomonas fluorescens strain SBW5. To find a way forward, we designed an SNRC database using a novel set of Perl-CGI scripts that enable storing and mining of sequence data. The use of Perl scripting means that the database is user-friendly and fully accessible from the Web, which promotes communication between collaborators. It also promotes the publication of both raw and processed data, and means that SNRC data and information derived from bioinformatic analysis can be hyperlinked with remote databases (e.g. GenBank, SWISS-PROT, PubMed etc.).

Microbial DNA analysis has become highly focused in the last five years on the manipulation and analysis of whole genome (WG) DNA sequences. The first of these was the 1.8 Mbp Haemophilus influenzae Rd genome completed in 1995 and since then, 27 additional microbial genomes have been completed. The value of these WG databases is tremendous, and there is substantial support for additional whole-genome sequencing efforts. Less obvious however, is the effort being put into sequencing projects where the complete genomic sequence is neither the primary nor the long-term goal of the research effort.

Many projects aim to obtain biologically relevant sequences from specific experimental strategies. One example is IVET (in vitro expression technology) screening, where genes expressed only under certain environmental circumstances are recovered (2). A second is direct fluorescence induction, where individual bacteria expressing green fluorescent protein fusions are isolated by cell sorting (4). A third approach is signature-tagged mutagenesis, in which sequences are obtained from transposons inserted into virulence genes (1). These approaches are of importance in understanding how bacteria survive and persist in different environments and have immediate application in understanding aspects of bacterial virulence and colonization. In some instances, IVET screenings are directly complemented by WG sequences, allowing rapid determination of genes or operons that are required for specific stages of bacterial pathogenesis. However, IVET screening of bacteria without WG sequences available is also being undertaken in a variety of laboratories.

The data generated by IVET screening differs from the sequence data generated during a WG sequencing project in two fundamental ways. Firstly, each IVET sequence tends to be short (200–700 bp), is the result of a single sequence run, and is unlikely to be incorporated into larger, contiguous consensus sequences (contigs). In contrast, WG project sequence data are incorporated into a growing contig that eventually covers the entire genome. Secondly, IVET and WG sequence data differ in the relative importance of each piece of sequence information: the WG sequence is an intermediate step in the generation of the final contiguous sequence, whereas the IVET sequence is of immediate experimental interest. IVET is not the only type of sequencing project that places value on SRNC sequence data. For example, environmental sequencing projects designed to extract DNA (or cDNA) sequences from communities of microbes (3) and projects sampling DNA from naturally occurring plasmids and plasmid families make use of SRNC DNA sequences for phylogenetic comparison, estimations of community genetic diversity and genomic capacity.

The very nature of SRNC sequencing projects means that these DNA sequences are often difficult to store, organize and manipulate as a group. SRNC data sets are comprised of many small sequences and it is often necessary to sort specialized information along with each entry (e.g. IVET isolation strategy and experimental conditions; BAC clone and specific end sequence and map location). Further, it is desirable not only to store and organize SRNC data, but also to be able to perform batch analysis and to do extensive homology (BLAST) searching on the

GUIDELINES

Communications should be in the form of letters and should be brief and to the point. A single small Table or Figure may be included, as may a limited number of references (cited in the text by numbers, and listed in alphabetical order at the end of the letter). A short title (fewer than 50 characters) should be provided.

Approval for the publication rests with the Editor-in-Chief, who reserves the right to edit letters and/or to make a brief reply. Other interested persons may also be invited to reply. The Editors of Microbiology do not necessarily agree with the views expressed in Microbiology Comment.

Contributions should be addressed to the Editor-in-Chief via the Editorial Office.
**Table 1.** Overall organization of MINE and its SNRC database, the PfSBW25 Encyclopaedia

<table>
<thead>
<tr>
<th>Core MINE functions (scripts)*</th>
<th>Data input</th>
<th>Web forms for submitting and editing entries that are easily customized to accommodate any type or amount of information about each sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch BLAST analysis</td>
<td>Web form provides full access to unlimited numbers of local BLAST searches; BLAST reports are stored and displayed in a database log</td>
<td></td>
</tr>
<tr>
<td>MINE search engine</td>
<td>Web form allows retrieval of any subset of the data using sophisticated queries that don’t require the user to learn a query language</td>
<td></td>
</tr>
<tr>
<td>PfSBW25 Encyclopaedia-specific Web pages</td>
<td>Introduction</td>
<td>Background and details of the PfSBW25 SNRC sequencing project</td>
</tr>
<tr>
<td></td>
<td>Summary information</td>
<td>List of all database entries, predicted genes and putative functions assigned through homology; lists of sequences which currently have no identified homologies</td>
</tr>
<tr>
<td></td>
<td>Supplementary</td>
<td>Includes instructions for sequence submission, lab. protocols, information links to useful Web sites, lists of contributors (raw sequence data, Web pages and programming scripts), etc.</td>
</tr>
</tbody>
</table>

*In addition to the core scripts there are an additional eight simpler support scripts that allow users to i) view all files, ii) calculate basic statistics, iii) change the format of any entry (to GCG or FASTA format etc.), iv) summarize BLAST reports, v) make BLAST databases, vi) do batch analysis (e.g. calculate G+C content for each file), vii) log comments made by users of the database and viii) log visitors to the database. Also included in MINE (Version 1.0) are simple command line utility scripts (e.g. an FTP tool to automatically retrieve remote documents) and very basic, fully annotated Perl-CGI scripts that are made to demonstrate simple tasks for beginners who are interested in learning how Perl code (and MINE) works.

We have overcome the difficulties inherent in developing a SNRC database by designing and developing a novel set of simple Perl-CGI scripts that provide all these features. This collection of scripts for storing and mining sequence data is referred to as MINE (Molecular Information Explorer). The first SNRC database using MINE is the PfSBW25 Encyclopaedia developed to maintain and analyse sequences from the plant-colonizing bacterium *P. fluorescens* SBW25 and the associated plasmid pQBR103 [the database currently contains some 434 individual sequence entries comprising a total of 262 kbp (or ~3.2% of the *P. fluorescens* SBW25 genome and ~18% of the plasmid pQBR103)]. We have chosen to develop this database using Perl and CGI (Common Gateway Interface) scripting because we feel it is essential to make this database fully accessible from the Web and user-friendly. Web access promotes communication between collaborators, publication of both raw and processed data, and yields the advantage that SNRC data and information derived from bioinformatic analysis can be hyperlinked with remote databases (e.g. GenBank, SWISS-PROT, PubMed etc.). Perl is now one of the principal computer languages currently used to develop novel bioinformatic resources. It can easily and efficiently handle large amounts of text (e.g. raw sequence data) and co-operates well with a wide range of other bioinformatic tools and remote resources on the Internet (http://bio.perl.org/). Perl therefore offers an ideal foundation for writing a simple collection of code that provides a core set of functions and uses a user-friendly interface. The scripts of the MINE database can be accessed from any computer running a browser (e.g. Netscape).

MINE (Version 1.0) is built around three core scripts that provide a Web-based environment in which to i) input data; ii) do unlimited BLAST searches and store hyperlinked BLAST report files with their associated database entries; and iii) use the search engine to create sophisticated queries that can retrieve any subset of the data (e.g. all sequences greater than 400 nt that also have a G+C content greater than 60% and with at least one mononucleotide run of Gs longer than 7 nt). Search results can be displayed and saved in a variety of formats (as a list of file names, in FASTA format or report form etc.), which means that SNRC data can be exported into spreadsheet applications (e.g. Excel accepts the column-based report format) and a wide variety of specialized bioinformatic tools (e.g. GCG uses lists of files for sequence alignment and phylogenetic analysis, and requires FASTA-formatted sequences). Several small support scripts are also integrated with the three core scripts (Table 1). It is hoped that the structure of MINE (containing many independent, easy-to-read single-function scripts with full annotation) will allow accomplished Perl programmers to quickly customize and expand MINE to suit their own project needs, and provide a useful first step towards getting more biologists involved in basic bioinformatic programming.

Data in a MINE database can be viewed and accessed at three levels, progressing from the least processed data, to the most processed data outputs (Fig. 1). The ‘raw’ data can be viewed in the MINE database log that displays all files (sequence entries, BLAST reports etc.) as hyperlinked Web pages. A more biologically informative view of the data can be obtained by creatively using the MINE search engine to display subsets of the data according to user-defined criteria. Finally, it is hoped that the curators of each database will use MINE to produce summary pages augmented by other research data in the form of traditional Web (html) pages.

The PfSBW25 Encyclopaedia currently contains a collection of Web pages that summarizes the most biologically interesting features of this database’s sequence data. The most highly processed form of sequence data can be seen in the Web pages listing homologues between *P. fluorescens* or pQBR103 DNA sequences and sequences deposited in other databases. Each entry (generated by the database curator) is a paragraph listing the gene name, function and hyperlinks to original sequence entries, comments and references (i.e. GenBank, SWISS-PROT, PubMed etc.). In addition, each entry contains a comment about the origin of each of the PfSBW25 Encyclopaedia’s sequence submission, recording details such as the plasmid or cosmid name, any closely linked sequences (e.g. sequences from the other end of the cloned insert) or known mapping details. Finally, each entry is hyperlinked to the raw datafile in the MINE database file containing that sequence and its associated BLAST reports. This view of the data will be of the most use to researchers interested in the biology of *P. fluorescens* SBW25 and pQBR103.

Since all of the data in the MINE database is accessible through Web pages it is easy to integrate data from MINE with a wide range of supplementary information using hyperlinks and curator-generated Web pages. For example, the PfSBW25 Encyclopaedia’s Web pages currently contains ninety-eight curator-generated Web pages that provide background information and details about the various research projects and the methods that produce raw sequence data for this database. It is envisaged that MINE databases should be composed of two complementary
MINE provides a simple and efficient tool for rapidly publishing the results of non-genome-scale sequencing projects in a user-friendly Web-based SNRC database that combines storage of data with the ability to do basic data-mining. The welcome page of the PISBW25 Encyclopaedia illustrates, using a collection of novel and freely available elements the data found in the MINE database, a more meaningful way and viewed using the search engine. Lastly, information culled from MINE (e.g. summary table of entries, predicted genes and putative functions assigned through homology) is presented in the PISBW25 Encyclopaedia’s Web pages generated by the database curator.

Figure 1. Basic flow of information through MINE within the PISBW25 Encyclopaedia. Raw sequence data and all associated information are placed into the PISBW25 Encyclopaedia using MINE software. MINE forms the core of the database, allowing the storage, analysis and viewing of large amounts of SNRC data. This data can be viewed at three levels: all raw data files (DNA sequence entries, BLAST reports, analysis files) can be viewed in the database log as Web pages. This information can be processed in a more meaningful way and viewed using the search engine. Lastly, information culled from MINE (e.g. summary table of entries, predicted genes and putative functions assigned through homology) is presented in the PISBW25 Encyclopaedia’s Web pages generated by the database curator.

Acknowledgements

This work was supported in part by grants from the BBSRC. D.F. is supported as a Sloan-NSF postdoctoral Fellow in Molecular Evolution. We wish to thank Richard Moxon for his helpful criticism of the manuscript.

Andrew J. Spiers, 1 Dawn Field, 2 Mark Bailey3 and Paul B. Rainey1

1,2Department of Plant Sciences and Molecular Infectious Diseases Group, Institute of Molecular Medicine, 3University of Oxford, UK

Author for correspondence: A. J. Spiers. Tel: +44 1865 275 145. Fax: +44 1865 275 074. e-mail: andrew.spiers@plants.ox.ac.uk

Heterodimeric dihydroxyacetone kinase from a ptsI mutant of Escherichia coli

The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) mediates uptake of carbohydrates by group translocation and regulates metabolism in response to the availability of carbohydrates. Catabolite repression and inducer exclusion are just two manifestations of PTS-dependent regulation (8). With the objective to obtain a more comprehensive picture of the PTS-dependent regulatory network, the proteomes of a wild-type and an isogenic non-polar ptsI mutant were compared (1). ptsI encodes enzyme I, which is at the top of the divergent protein phosphorylation cascade of the PTS. It transfers phosphoryl groups from phosphoenolpyruvate to the general phosphoryl carrier protein HPr, hence they are distributed to the different carbohydrate transporters, some of which contain subunits of known regulatory activities. Escherichia coli cells growing exponentially in Luria broth were broken in a French pressure cell in the presence of 8 M urea and the soluble fraction, comprising cytoplasmic and membrane-associated proteins were separated by isoelectric focusing and gel electrophoresis (16 × 15 cm). Four proteins of which the spot intensity was strongly increased in the ptsI mutant (Fig. 1) were excised, electroblotted and identified by microsequencing as IAB\(^{\text{IIA}}\) (manX), sorbitol-6-phosphate 2-dehydrogenase (srld), and the gene products of ycgT and ycgS. A fifth spot of reduced intensity in the ptsI mutant was the periplasmic peptide cis-trans isomerase (kpA). IAB\(^{\text{IIA}}\) is the cytoplasmic subunit of the mannose transporter of the PTS. Sorbitol dehydrogenase is encoded by the sorbitol operon, which also encodes a PTS transporter for sorbitol (srld, srle and srlB) (9). YcgT and YcgS are 27% and 24% identical with the N-terminal and C-terminal half, respectively, of the dihydroxyacetone kinase (DHAK) of Citrobacter freundii. ycgT and ycgS are two cistrons of an operon which, in addition, contains ycgC. ycgC encodes a three-domain protein composed of an N-terminal domain of unknown function, an HPr-like domain and a C-terminal domain which is highly homologous to the N-terminal domain of enzyme I of the PTS (3). The association of the DHAK genes with a PTS gene, and PTS-dependent regulation of the operon strongly suggest that the DHAK operon belongs to the PTS regulon. ORFs with sequence homology to the N-terminal domain of YcgC occur in the genomes of Mycoplasma capricolum, Demococcus radiodurans, Selenomonas ruminantium, Streptomyces coelicolor and Staphylococcus epidermidis. They however are not fused with HPr- and EI-like domains.

Upregulation of PTS operons in response...
to inactivation of ptsI has been observed repeatedly (5, 6). Expression of the mannose operon, of which manX is the first cistron, of ptsI and of ptsG is repressed by Mlc, a transcriptional repressor of the PTS. It appears that the sorbitol (srl) and DHAK operons are subject to the same regulatory mechanism. According to a well supported model, inactivation of ptsI results in non-phosphorylation of IICB with consequent repression of McIP by non-phosphorylated IICB (5, 6) and as a consequence, upregulation of McIP-repressed operons (7).

Our observations agree with the experimental findings communicated by Paulsen et al. (4), who found the same genes (termed dbaK1, dbaK2 and dbaH) by analysis in silico of the region at 26 min of Escherichia coli (10) but now turns out to be a DHAK. Uprogelation of DHAK expression in a ptsI mutant comes as a surprise, because Jin & Lin’s (2) observation of compromised DHA metabolism in a ptsI mutant would have led us to expect the opposite. Paulsen et al. (4) suggest that YcgC (DhaH) in the non-phosphorylated state acts as an allosteric inhibitor of DHAK, or as an obligatory activator in the phosphorylated form. However, such allosteric regulation, if it occurred, had to be very tight to completely abolish DHAK activity in the induced state. The three proteins have been purified. All three are required for in vitro phosphorylation of dihydroxyacetone with phospho-HPr as immediate phosphoryl donor. YcgS, YcgT and YcgC are not an ATP-dependent kinase but, as originally proposed (2), a dihydroxyacetone specific enzyme II. This enzyme II analogue is soluble, however. It catalyses phosphorylation of a cytosolic substrate, in contrast to all known enzymes II of the PTS, which are membrane bound and catalyse transport concomitant with phosphorylation.

Acknowledgements
This work was supported by Grants 31-45838.95 from the Swiss National Science Foundation and from ARPIDA AG Münchenstein, Switzerland.

R. Beutler, U. Kampfer, J. Schaller and B. Erni
Department of Chemistry and Biochemistry, University of Bern, CH-3012 Bern, Switzerland

Author for correspondence: B. Erni. Tel: +41 31 631 4346. Fax: +41 31 631 4887. e-mail: erni@ibc.unibe.ch


A common conserved amino acid motif module shared by bacterial and intercellular adhesins: bacterial adherence mimicking cell–cell recognition?

The whole genome of Neisseria meningitidis strain MC38 (serogroup B) has been recently sequenced and analysed (13), and considerable attention has been dedicated to explore the potential of the encoded outer-membrane proteins for the development of effective vaccines (10). The current vaccines, which are based on the group-specific capsular polysaccharides afford no protection against serogroup B strains because of their structural similarities with human glycoproteins. Therefore, the outer-membrane proteins are considered more promising candidates for new experimental vaccines. Of these, it has been shown that the putative adhesin encoded by the NMB0992 locus of MC38 is surface-exposed and a target of bactericidal antibodies (10). This molecule of approximately 56 kDa (591 aa) is closely related to Hsf (233 aa) and its allelic variant Hia (1098 aa), both adhesins of Haemophilus influenzae involved in the formation of type b surface fibrils (11). The overall amino acid sequence similarity of NMB0992 to the Hsf and Hia proteins is 57.9% and 50% identity, respectively. This suggests a role in adherence through a possible common mechanism.

The Hsf and Hia products share a well conserved C-terminal portion (grey boxes in Fig. 1a) containing the ATP-binding motif GINVS/GKT and a terminal sequence consensus typical of autotransporter polyprotein in Gram-negative bacteria (5). The final part of this region appears to be well conserved also in NMB0992 of N. meningitidis (9). In particular, the amino acid segment 423–591 of NMB0992 shows 42.3% identity with the 221–2353 region of Hsf and with the 961–1098 tract of Hia. Less clear is both the functional role and the features of the N-terminal portion of these proteins. As schematized in Fig. 1a, this tract is organized in three large repeats in Hsf corresponding to amino acid positions 174–608, 847–1291 and 1476–1914, and present in a single copy in Hia (221–658) and in NMB0992 (180–380), respectively (9, 11). This observation suggests that a modular structure could be a relevant factor in explaining the adhesin function and that the identification of the repetitive unit could be useful to gain some insight on the mechanisms of the adhesion events.
A short conserved region shared among eukaryotic and prokaryotic adhesins could be responsible for bacterial adhesion

The comparison of Hsf and Hia with NMB0992 reveals that the distribution of segments having local similarity is not uniform, but it seems to be distributed in blocks differently arranged in each protein (Fig. 1a). The 157–350 N-terminal tract of NMB0992 can be divided in two sub-regions, R1 and R2, sharing 28% aa identity and spanning positions 157–249 and 252–350, respectively. Both segments appear to be well conserved in Hsf and Hia. Their sequence alignment indicated the presence of a group of invariant positions (Fig. 1b) which served to identify a block about 100 aa long, present twice in NMB0992, three times in Hia and eight times in Hsf sequences (green arrows, Fig. 1a). The final portion of such a module appears to be better conserved and invariant in a battery of 22 pathogenic meningococcoserotype B strains (10). Around the invariant positions, a primary consensus can be identified in the motif G-W-X(2,5)-[K-R]-X(6,9)-[NDEQ]-X-[NDEQ]-X(5,8)-[NDE]-X(0,2)-[VI]. This pattern has been used to screen the SWISS-PROT (Release 39.6) and TrEMBL (Release 14.11) databases with the Scan Prosite tool available at the Expasy web server (http://www.expasy.org). The results, summarized in Fig. 1 (c), show that such a motif is present in two other adhesive molecules, namely HMW1 protein from H. influenzae and the neural adhesion protein NB-2 from human and rat. In addition, the scan retrieves three other apparently unrelated sequences (accession numbers P36698, P42967 and O29065) whose similarity to the adhesion pattern is, however, less stringent (Fig. 1c). HMW1 is a high-molecular-mass protein that mediates the attachment to the human epithelial cells by the non-typable acapsulate strains of H. influenzae (12). NB-2 belongs to the cell adhesion molecule (CAM) family (8). A large number of CAMs are members of the immunoglobulin superfamily, which comprises the N-CAM, L1 and contactin/F3 subgroups. These molecules have a number of immunoglobulin motifs and fibronectin-III repeats in their extracellular domains and mediate the cell–cell adhesion and signalling by homophilic binding resident in opposed membranes (7). Functional studies and crystal structures of some representative members of the CAM family show that the N-terminal extracellular domains are responsible for the intercellular adhesion, forming antiparallel dimers at the cell–cell interface (4).

The sequence analysis provides further indications of a general analogy between NMB0992 and N-CAM. The secondary structure prediction shows that the 157–350 N-terminal tract of NMB0992 has a wide beta structure propensity. In addition, the length and presence of repeats in this region is fully compatible with a modular three-dimensional structure consisting of two immunoglobulin-like domains (data not shown). This folding, adopted by eukaryotic molecules involved in cell–cell adhesion, has been observed also in bacterial adhesins like invasin (Yersinia pseudotuberculosis) (3), intimin (enteropathogenic Escherichia coli) (1) and FimH (uro-pathogenic E. coli) (2), suggesting that this structure could be functionally relevant for the interaction with the host cell. The analogy between NMB0992 and N-CAMs can perhaps suggest a molecular function for the conserved motif G-W-X(2,5)-K-[TGN]-X(6,9)-[NDEQ]-X-[NDEQ]-X(5,8)-[NDE]-X(0,2)-
in addition to homophilic recognition, the extracellular domains of N-CAM are involved in heterophilic interactions like binding to heparin and heparan sulfate. The above-mentioned sequence motif can be in fact adapted to the sequence 1\textsuperscript{89}WKHKG RDFVLLKDYRF1\textsuperscript{146} from human N-CAM. This region, conserved between N-CAM molecules of several eukaryotic species, has been proposed to be a heparin binding site (6).

Our analysis identified proteins of N. meningitidis and H. influenzae with a sequence motif and a predicted fold common to N-CAMs. We suggest that the functional role of this similarity might enable the bacterial cells to promote adherence by mimicking the cell–cell recognition phenomena that occur at the neural level.

Maria Scarselli, Rino Rappuoli\textsuperscript{1} and Vincenzo Scarlato\textsuperscript{1,2}
\textsuperscript{1}Department of Molecular Biology, IRIS, Chiron S.p.A., Via Fiorentina 1, 53100 Siena, Italy
\textsuperscript{2}Department of Biology, University of Bologna, Via Selmi 3, 40126 Bologna, Italy

Author for correspondence: V. Scarlato. Tel: +39 0577 243565. Fax: +39 0577 243564. e-mail: enzo-scarlato@chiron.it


**Epigenesis – a request for information on loss of adaptive phenotypes**

Every microbiologist involved in environmental studies knows only too well that bacteria isolated for a phenotype linked to their role in the environment are liable to lose this phenotype upon cultivation under laboratory conditions. The lost phenotype is most often an adaptive phenotype, i.e. the capacity to synthesize a (non-essential) protein(s) in response to particular environmental conditions. This loss is essentially different to the loss of the strain itself, although both may be avoided by deep-freeze conservation.

**What causes adaptive phenotype loss?**

In the cases where it has been studied (and published) it has been attributed to the instability of a plasmid. However, many adaptive traits are not encoded by plasmids. Another common but untested hypothesis is that of (unknown) mutations. This hypothesis has been strengthened by the discovery of adaptive mutations arising during prolonged stationary phase. However, the very high rate of disappearance of adaptive phenotypes in strains recently isolated from the environment may not be in favour of explanations that invoke mutations. A less familiar explanation may be based on epigenetic modifications.

**Epigenetic modifications**

An epigenetic modification is a phenotypic modification that is transmitted over generations even when the conditions that gave rise to it are no longer present. Examples of epigenesis are known mainly in eukaryotes. However, epigenesis was shown in the lactose operon of Escherichia coli as long ago as 1957. Two independent studies of the same strain both revealed striking dissimilarities between subpopulations that could be maintained through many generations according to whether or not these subpopulations were previously induced by a lactose analogue. Induced and uninduced cells were transferred to media containing either a very low inducer concentration (4) or the simultaneous presence of lactose and glucose (1, 2, 3). The result was that previously induced cells continued to produce β-galactosidase whereas uninduced cells never did. Thus, two populations of genetically identical bacteria growing in identical conditions, could display two different phenotypes, depending on a remote event in the history of the culture. This is epigenesis (7). In both studies, the difference was due to the lactose permease, which was present in the membrane of the previously induced cells but not of the uninduced cells. This allowed a very low external concentration of inducer to accumulate in the cytoplasm up to a level sufficient to induce the lactose operon; in the absence of the permease, this level could not be attained at low levels of lactose or in the presence of glucose (Fig 1a).

Thus, the permease was able, once present, to facilitate its own synthesis (together with that of the β-galactosidase), by allowing uptake of the inducer: this is positive feedback. The experiments were pursued for a long time (150 generations in one case), which indicated that the cells had a stable phenotype. Finally, it should be noted that the concentration of external inducer necessary to induce expression of the lac operon in uninduced bacteria is much higher than the concentration below which induction is lost. This is hysteresis (Fig. 1b).

These two phenotypes exemplify what physicists call bistability (or multistationarity) in systems in which two permanent or steady states coexist. Multistationarity and hysteresis are well known properties of dynamic systems in which a positive feedback loop is active and where at least one of the differential equations that represent the system is non-linear (8). This is clearly the case of the lactose operon. It is also the case of bacteriophage lambda. Thomas & Van Ham (10) have shown that a positive feedback loop is responsible for the coexistence of two stable states, lysis or lysogeny, between which a phage infecting a bacterium has to choose. No genetic change is required and the two states coexist in the same population of bacteria, under the same environmental conditions.

To complete this overview of epigenetic modifications, it should be noted that a very short pulse of inducer (a fluctuation in the medium composition) is, in the case of the first example, sufficient to transform all the uninhibited induced cells. In this case, it is easy to switch from one steady state to the other. More complex methods are required to switch lambda from the temperate to the virulent state, and no methods of switching (or preventing switching) are known for some temperate phages. This means that in some cases of multistationarity, one of the steady states, once established, may be irreversibly maintained.
How might the loss of an adaptive phenotype be due in some instances to an epigenetic modification?

Suppose the bacterium may display two steady states, one of which corresponds to the ability to respond to a given environment. In this state, the adaptive phenotype is expressed in this environment. Cultivation in conditions remote from those of the environment (in the laboratory) generally prevents the expression of the phenotype, which can however be restored by a known signal. However, laboratory conditions might actually favour the switch to another steady state that corresponds to the inability to respond to that signal, which I call loss of the adaptive phenotype. This switch might be reversible but microbiologists generally throw the ‘lost’ culture away and do not search for reversal to the desired phenotype.

Is this hypothesis likely?

A positive feedback loop combined with a non-linear relationship is necessary for multistationarity, as first conjectured by biologists (9) and later demonstrated by mathematicians (5, 6). It is significant that adaptation to environmental changes is often controlled by a regulator that in some instances positively regulates its own synthesis, in addition to controlling the expression of the genes of the regulons involved in this response. Examples include the regulator GlnG, which controls dinitrogen metabolism in Klebsiella pneumoniae, and many of the regulators that belong to the so-called two-component family. This does not prove that there is epigenesis in each case but shows that the minimum requirements for it are satisfied.

Could epigenesis and multistationarity improve the fitness of strains in the environment?

Consider a strain that lives in an environment subject to many fluctuations. It would be advantageous for its phenotype to remain adapted to one type of conditions whilst those conditions prevail and for this phenotype to change only when other conditions prevail that require a different phenotype. In this way, temporal fluctuations in the intensity of a signal would not entail energetically expensive fluctuations in the bacterial phenotype; it would remain adapted to the average intensity. In other words, epigenesis could be a way of maintaining a stable phenotype. In the case of simultaneously conflicting demands for different phenotypes, multistationarity may allow different bacteria within the same population to have different phenotypes and hence allow the population to benefit from heterogeneity.