

Adaptive evolution of highly mutable loci in pathogenic bacteria

E. Richard Moxon*, Paul B. Rainey†, Martin A. Nowak‡ and Richard E. Lenski§

*Molecular Infectious Diseases Group, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. †Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK.

‡Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. §Center for Microbial Ecology, Michigan State University, East Lansing MI 48824, USA.

Bacteria have specific loci that are highly mutable. We argue that the coexistence within bacterial genomes of such 'contingency' genes with high mutation rates, and 'housekeeping' genes with low mutation rates, is the result of adaptive evolution, and facilitates the efficient exploration of phenotypic solutions to unpredictable aspects of the host environment while minimizing deleterious effects on fitness.

Introduction

All organisms are faced with the perpetual challenge of maintaining their fitness in diverse and changing environments. To meet this challenge, populations of organisms must possess mechanisms and strategies for responding to changes in their environment. These include phenotypic acclimation, by which an individual organism modifies some aspect of its behaviour, morphology or metabolism in response to environmental change, and genetic adaptation, whereby the genetic composition of a population may change as a result of natural selection.

Natural selection has produced a range of genetic mechanisms that facilitate acclimation to a wide variety of external stimuli. In bacteria, these range from feedback mechanisms, such as catabolite repression of transcription, to sophisticated 'two-component' sensory systems, in which a signal from the external environment is transduced through histidine protein kinases [1], ultimately regulating gene expression. These and similar mechanisms enable bacteria to modulate the activity of their genes in response to particular external conditions, thereby maintaining their fitness in changing environments. Indeed, the strong phylogenetic conservation of these mechanisms is testimony to their general and continuing utility; specific responses presumably reflect the probability of bacteria encountering particular environmental situations.

Provided that environmental factors (such as nutrients, temperature, osmolarity or acidity) remain within certain limits, then changes in the external environment may be accommodated by coordinated regulation of gene expression. Given, however, the diversity and unpredictability of environmental changes, these stereotypic responses are unlikely to contribute more than a

limited subset of the phenotypic states necessary for long-term evolutionary success. Confronted with a persisting unfavourable environment in which classical regulation of gene expression cannot provide an adequate response, a population of bacteria may face extinction unless it can adapt genetically by natural selection.

Pathogenic bacteria face especially stringent tests of their adaptive potential, due to the characteristic diversity and polymorphic nature of their hosts' immune responses. This is because, typically, bacterial infections occur within a matter of hours, during which time pathogenic organisms are transmitted between genetically distinct hosts, colonize epithelia and disseminate through a host to produce invasive disease. The capacity of bacteria to negotiate the differing environments in the host, including both intracellular and extracellular locations, is remarkable, especially as infections usually involve the clonal expansion of a single strain of the pathogen [2,3]. This conflict between the 'pathogenic personality' [4] of bacteria and the antagonistic response of the host provides a driving force for, and is shaped by, co-evolutionary processes that have been described by colourful metaphors such as gene-for-gene arms races [5-7] and the Red Queen hypothesis [8].

Given their relatively large population sizes and short generation times, pathogenic bacteria would seem to have considerable advantages over their hosts in adaptability and evolutionary flexibility. These apparent advantages to the pathogen may be offset by the immune systems that enable the host to generate an extensive repertoire of immunologically competent cells [9,10]. Such immunity represents a phenotypic response, in the sense that it is not inherited, but the ability to respond in this manner has a genetic basis that is presumably the result of evolutionary adaptation

Correspondence to: E.R. Moxon.

Table 1. Mechanisms of altering gene expression.

	Deterministic	Stochastic
Discriminate	Classical gene regulation	Gene conversion Site-specific recombination Oligonucleotide repeats Homopolymeric tracts Site-specific methylation
Indiscriminate	—	Transitions Transversions

by natural selection. Bacteria may similarly adapt genetically in ways that affect not only their ability to respond phenotypically to environmental variations, but also their propensity to undergo further genetic adaptation.

Biologists have long been fascinated by the evolution of those aspects of an organism's physiology, biochemistry and reproductive biology that affect its rates of genetic recombination and mutation, and hence determine the amount of heritable variation that is available for genetic adaptation by natural selection. For example, it has been argued that sexual reproduction provides important strategic advantages for hosts because of its role in generating the heritable variation that genetic adaptation requires. Indeed, it has been suggested that the selective pressures imposed by pathogenic microbes may have been responsible for the evolutionary origins and maintenance of sexual reproduction [10–14].

In this article, we shall review evidence indicating that pathogenic bacteria have evolved mechanisms for increasing the frequency of random variations in those genes that are involved in critical interactions with their hosts. Having elevated mutation rates in a specific subset of genes may be highly advantageous, allowing

certain phenotypic traits to respond rapidly, by natural selection, to unpredictable changes in the environment, while also ensuring the conservation of essential functions encoded by other genes. This hypothesis accords well with the co-existence in many pathogenic bacteria of highly mutable loci ('contingency' genes) and loci with much lower mutation rates ('housekeeping' genes).

Phenotypic variation generated by highly mutable genes

Phenotypic variation within populations can be generated by alterations to the sequence or conformation of DNA. Such genetic changes can result from inter-genomic events, such as recombination, or intra-genomic events, such as mutations. The contribution of intergenomic mechanisms (reviewed in [15–18]) to the generation of phenotypic diversity in clonal populations during an acute infection is minimal, and will not be considered further here. Instead, we shall focus on intragenomic mechanisms, which recent results show play an important part in enabling pathogenic bacteria to counter the complex and variable environment of their hosts.

Intragenomic mechanisms of altering gene expression can be classified as either deterministic or stochastic (Table 1). Deterministic processes encompass classical mechanisms of gene regulation, whereas stochastic processes comprise the various mutational mechanisms. Although mutational processes are stochastic, they are not necessarily uniformly distributed across the genome. We define as discriminate those mutational mechanisms that occur at certain loci at unusually high frequency. Classical mutations — transitions and transversions — by contrast are indiscriminate and occur at a relatively low frequency at all loci. The mechanisms that generate discriminate mutations have properties in

Table 2. Examples of genetic mechanisms generating random phenotypic variation at high frequency in specific loci*.

Mechanism	Organism	Determinant	Host	Reference
Gene conversion (homologous recombination)	<i>Borrelia hermsii</i>	Lipoprotein (<i>vmp</i>)	Human	[59]
	<i>Bacillus thuringiensis</i>	Toxin (<i>cryIA</i>)	Lepidoptera	[60]
	<i>Haemophilus influenzae</i>	Capsule (<i>capB</i>)	Human	[61]
Site-specific recombination	<i>Escherichia coli</i>	Fimbriae (<i>fimA</i>)	Human	[62,63]
	Incl1 Plasmid R721 (<i>E. coli</i>)	Pilin (<i>pilV</i>)	Human	[64]
	<i>Moraxella bovis</i>	Pilin (<i>tfp</i>)	Cattle	[65]
Oligonucleotide repeats	<i>Haemophilus influenzae</i>	Lipopolysaccharide (<i>lic</i> loci)	Human	[30]
		Fimbriae (<i>hifA</i> , <i>hifB</i>)	Human	[25]
	<i>Neisseria gonorrhoeae</i>	Opacity proteins (<i>opa</i>)	Human	[37,38]
	<i>Pseudomonas solanacearum</i>	EPS, virulence determinants (<i>phcA</i>)	Plants, especially solanaceous crops	[66]
Homopolymeric tracts	<i>Neisseria meningitidis</i>	Opacity proteins (<i>opc</i>)	Human	[27]
	<i>Mycoplasma hyorhinis</i>	Lipoprotein (<i>vlp</i>)	Swine	[28]
	<i>Bordetella pertussis</i>	Fimbriae	Human	[26]
Site-specific transposition	<i>Pseudomonas atlantica</i>	Extracellular polysaccharide	Not applicable	[67]
Site-specific methylation	<i>Escherichia coli</i>	Pilin (<i>papA</i>)	Human	[68]
Unknown	<i>Pseudomonas tolaasii</i>	Virulence, chemotaxis, attachment	Mushroom	[69,70]
	<i>Rhizobium phaseoli</i>	Symbiotic efficiency	Leguminous plants	[71]

*See [50,58] for further examples.

common with both classical gene regulation and classical point mutation: they are confined to specific loci, like classical gene regulation, but share with classical point mutation the stochastic timing of their occurrence (Table 1). Although we concentrate on alterations to gene expression caused by changes in nucleotide sequence, we would mention that they can also be caused by changes in DNA conformation, resulting for example from site-specific methylation.

We wish to emphasize the stochastic, discriminate mechanisms that have evolved to generate high frequency intragenomic changes in nucleotide sequence, or DNA conformation, at particular chromosomal loci (Table 1). These loci are characterized by the presence of certain nucleotide arrangements: tandem repeats, homopolymeric tracts or potential methylation sites (Table 2). Through mechanisms such as slipped-strand mispairing, polymerase slippage or recombination between homologous repeats, these nucleotide arrangements make the loci unstable, so that genetic variation that can affect transcription or translation is generated at a high frequency. The functional consequence is that these alterations in genotype can rapidly promote phenotypic heterogeneity, even in small, clonal populations of bacteria, such as the limiting inoculum that initiates an infection.

The term 'programmed DNA rearrangement' [19,20] has confusingly been applied to the evolution of unstable

nucleotide sequences in particular (non-random) loci. In a crucial sense, however, these mechanisms are not programmed; that is, specific mutations — and the phenotypic changes that they engender — do not occur more often when they are useful than when they are not, nor are they deterministic. Instead, they have apparently evolved at particular loci to promote random phenotypic variation, and hence adaptive potential, in the face of constantly changing host polymorphisms and host immune cells. Interestingly, host organisms also generate random diversity in B lymphocytes by similar mechanisms [21].

Contingency behaviour in a host-adapted pathogen

The pathogenic bacterium *Haemophilus influenzae*, a major cause of meningitis, uses the intragenomic mechanisms of slipped-strand mispairing and homologous recombination, which are *rec*-independent and *rec*-dependent, respectively, to generate high-frequency changes in the expression of genes encoding fimbrial, lipopolysaccharide and capsular polysaccharide cell-surface determinants (Fig. 1), which are important to its commensal and virulence behaviour. These mechanisms may be widely used by pathogenic bacteria to facilitate infection and counter host defence mechanisms.

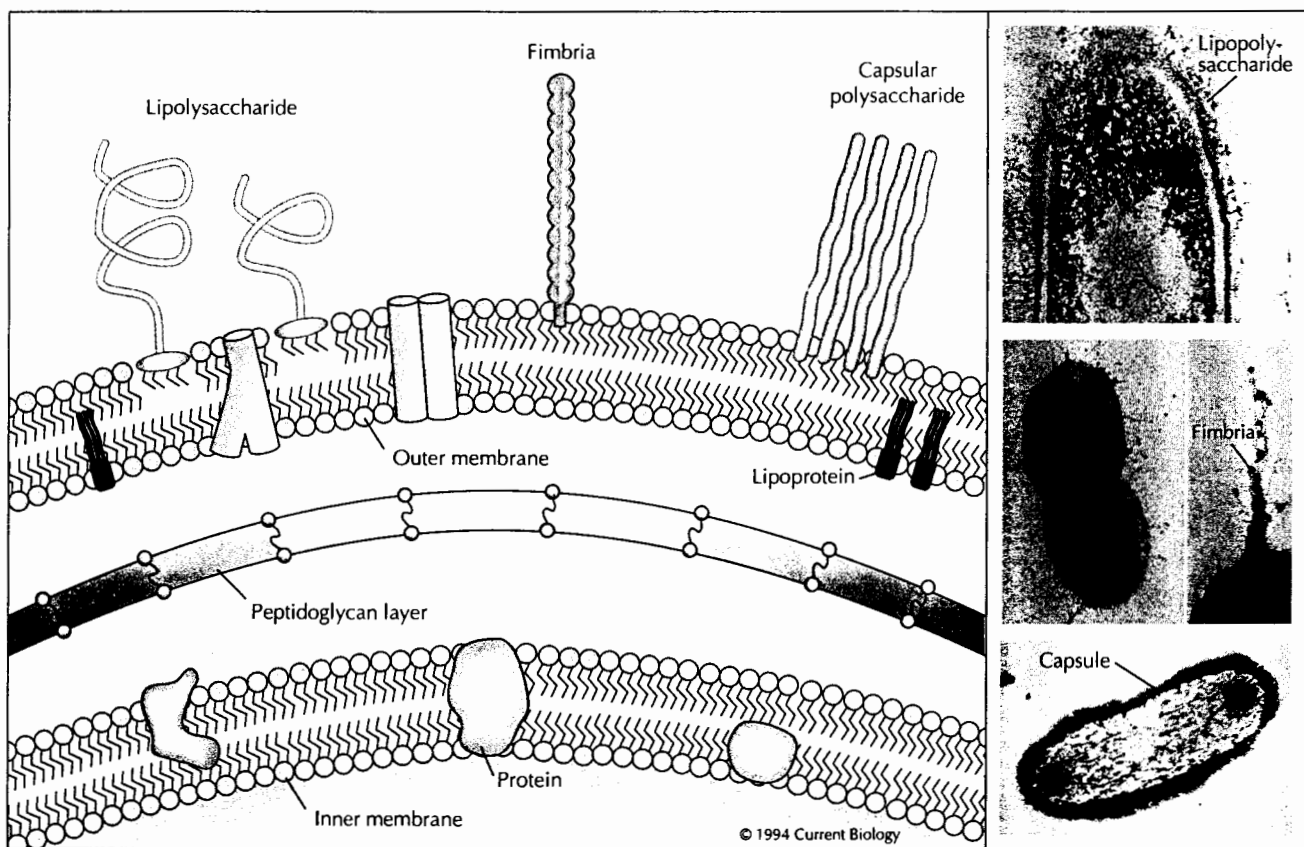


Fig. 1. The drawing represents a cross-section through the the inner and outer membranes of the *Haemophilus influenzae* cell wall. Surface exposed structures include lipopolysaccharide, fimbriae and capsular polysaccharide. The electron micrographs on the right show lipopolysaccharide (top; micrograph courtesy of Mark Kahn), fimbria (middle; micrograph courtesy of Loek van Alphen) and capsule (bottom).

Fimbria phase variation

Fimbriae are proteinaceous appendages that confer on bacteria the ability to bind to mucous and to epithelial cells, through adherence to a sialic-acid containing lactosylceramide host receptor [22]. The fimbriae are composed of a repeating polypeptide subunit of molecular weight 22–27 kD. During a natural infection, strains isolated from the nasopharynx are fimbriated, reflecting the beneficial role of this phenotype in colonization [23]. In contrast, bacteria cultured from blood or cerebro-spinal fluid (CSF) are non-fimbriated, probably because the possession of fimbriae facilitates clearance by host cells [24]. The gene cluster required for *H. influenzae* fimbria synthesis includes *hifA*, which encodes the subunit polypeptide, and *hifB*, which is a homologue of several known periplasmic chaperones. Sequencing these genes showed that *hifA* and *hifB* are divergently transcribed (Fig. 2), and the effects of mutations of each gene confirm that both are essential for fimbria synthesis.

When *H. influenzae* variants were studied that either expressed or lacked fimbriae, it was found that switching between expression states is controlled at the transcriptional level. The key to the switching mechanism was recently found to lie in the presence of tandem repeats of the dinucleotide TA in the intergenic region between the *hifA* and *hifB* coding regions [25]. The number of TA repeats varies, usually through loss or gain of a single repeat. This change alters the spacing between the *hifA* and *hifB* promoter sequences, both of which overlap the TA repeat region (Fig. 1). The altered distance between the -35 and -10 sequences is known to be crucial for the interaction with RNA polymerase, as it determines the relative phasing of these binding elements with respect to the DNA helix. In keeping with this proposal, maximal expression of fimbriae correlated with the presence of ten TA repeats, giving a separation of 16 nucleotides between the -35 and -10 consensus sequences. Thus, rapid on-off

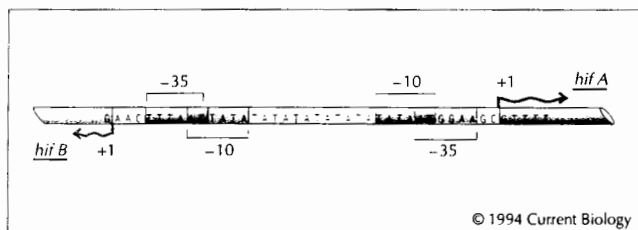


Fig. 2. Nucleotide sequence of the intergenic region between the *H. influenzae* piliation genes *hifA* and *hifB*. The positions of the transcriptional start points (+1) and putative promoter elements (-35 and -10) are indicated (in blue for *hifA* and red for *hifB*). The loss or gain of TA repeats, ten of which are shown, alters the spacing between the -10 and -35 consensus sequences, which are thought to be binding sites for RNA polymerase. This alters the efficiency with which RNA polymerase binds, thus affecting transcription. With ten TA repeats the spacing between the -10 and -35 elements is 16 nucleotides, which is predicted to allow efficient binding of RNA polymerase, whereas with nine or eleven repeats transcriptional efficiency is predicted to be reduced. These predictions correlate with the finding that fimbria expression is maximal with ten TA repeats, reduced with eleven TA repeats and absent with nine TA repeats [25].

switching of fimbriae expression is coordinated through transcriptional control of *hifA* and *hifB* [25].

The variable expression of the subunit of *Bordetella pertussis* fimbriae may also be controlled by a repeat sequence tract, in this case a tract of cytosine residues upstream of the fimbrial subunit gene [26], but the molecular details are less well understood in this case. Variable transcription of the gene encoding Opc, an outer membrane protein of *Neisseria meningitidis* that promotes adherence and invasion of epithelial and endothelial cells, is also modulated by a homopolymeric tract of cytosines [27]. In a similar manner, variable lipoproteins (Vlps) on the surface of the *Mycoplasma hyorhinis* membrane are subject to phase variation. Vlp expression is controlled at the transcriptional level by the number of adenine nucleotides in a poly(A) tract. The 'on' and 'off' expression states correspond to lengths of 17 and 18 residues, respectively, of the poly(A) tract and it is thought that changes in the length of the poly(A) tract may critically affect the spacing or secondary structure between the -10 and -35 or DR1-b structures [28,29].

Lipopolysaccharide antigenic variation

Lipopolysaccharide (LPS) is a macromolecule that is unique to Gram-negative bacteria, and it is the major component of the outer leaflet of the cell envelope. LPS consists of lipid A, which is anchored in the cell envelope, and core saccharide, which extends out from the cell surface. LPS is known to be a major virulence determinant of *H. influenzae*. Phenotypic variation in *H. influenzae* cell-surface LPS provides a further example of random variation generated by high frequency mutations. In this case, however, the variation is not due to a transcriptional mechanism like that responsible for fimbriae phase variation. Rather, the phenotypic expression of several LPS core saccharide structures can be reversibly lost or gained despite continuous synthesis of mRNAs for the enzymes of LPS biosynthesis. The switch is instead effected at the level of translation of these mRNAs, which is turned on and off by frame-shift mutations [30]. As the different core saccharide structures switch randomly and independently of each other, a population of *H. influenzae* founded by a single clone and residing within a single host can generate an extensive repertoire of variant LPS epitopes.

The molecular basis of the variable expression of one of these saccharide structures, gal α (1-4)gal β [31], has been studied in some detail. When colonies of *H. influenzae* were blotted onto nitrocellulose filters and allowed to react with a monoclonal antibody specific for the structure gal α (1-4)gal β , individual colonies showed either strong, intermediate or undetectable binding. Mutants have been isolated that do not express the digalactoside structure, and one of the mutations maps to a gene designated *lic2*, which is essential for synthesis of the digalactoside [32]. At the 5' end of the *lic2* gene there are multiple tandem repeats (about 16) of the sequence CAAT. However, the number of CAAT repeats varies; loss or gain of CAAT — usually a single

repeat — moves upstream translational initiation codons in and out of frame with the remainder of the coding sequence, thereby creating a translational switch [30,33] and resulting in variable synthesis of the digalactoside.

The slipped-strand mispairing mechanism of the *lic2* frame shift is presumably similar to that underlying fimbria phase variation [34], a *rec*-independent mechanism capable of mediating high-frequency random changes in nucleotide sequence (Fig. 3). It should also be noted that the resulting mutation and its potential for fortuitously adaptive changes in phenotype can occur either during chromosomal replication or through mismatch repair; the latter mechanism is potentially important in the situation in which bacteria are under stress and possibly in a non-replicating state. It may be significant that DNA immediately up-stream of *lic2* is especially rich in AT nucleotides [32]. This would tend to facilitate strand separation and increase the likelihood of slipped-strand mispairing; such a region would be susceptible to the effects of altered

supercoiling, which is known to affect transcriptional efficiency. There is growing evidence in support of the view that changes in supercoiling have a role in the global regulation of gene expression [35].

It should be emphasized that the variation generated by the tandem CAAT repeats is quite distinct from another source of LPS antigenic variation, referred to as microheterogeneity [36]. LPS is a tertiary gene product, in that the genes involved in the synthesis of LPS encode mRNAs that encode the enzymes or regulatory proteins that ultimately determine the LPS structure. LPS synthesis is consequently subject to additional variation, beyond that due to genetic polymorphisms, as it depends in part on the interaction between the LPS biosynthetic enzymes and their cognate substrates, which is inherently a stochastic process. This stochasticity means that the LPS structures vary within a cell, and can contain different saccharide isomers and anomers, variable branching chains and variable alternative sugars in the basal structure. This diversity is 'blind' in that, for any bacterial cell, the prevalence of any particular structure can be predicted only within certain confidence limits.

The extent to which stochastic events influence the degree of microheterogeneity is not known, but given the non-linearity of the dynamics of LPS synthesis, even minor random perturbations in the biosynthetic pathway could have substantial consequences, as in deterministic chaos theory. In contrast to variation resulting from highly mutable tandem repeats, where sibling cells usually have the same number of repeats as the progenitor despite switching frequencies of about 0.001, the random diversity generated by microheterogeneity is not heritable. Nonetheless, microheterogeneity may be useful in producing many different surface configurations, increasing the likelihood that at least one cell gets past some barrier to infection, provided that once a cell gets past the barrier, any surface configuration will suffice. Thus, random phenotypic variation in LPS molecules, both that which is heritable and that which is not, may be important for the evolutionary success of *H. influenzae*, and natural selection can therefore be expected to maintain or enhance those molecular mechanisms that promote such variation.

A further example of antigenic variation generated by translational frame-shift mutations is that of the opacity (Opa) proteins of *Neisseria gonorrhoeae* [37,38]. Multiple copies of genes encoding gonococcal Opa proteins are located at a variety of loci on the chromosome. Each copy includes CTCTT repeats, in the region of the gene encoding the hydrophobic core of the leader peptide. Variation in the number of CTCTT repeats alters the translational reading frame, thereby determining whether or not the protein is translated. It has been suggested that the repeat region can adopt a triple-stranded H-DNA conformation, which would expose a stretch of single-stranded DNA that would be a potential target for single-strand specific nucleases [39].

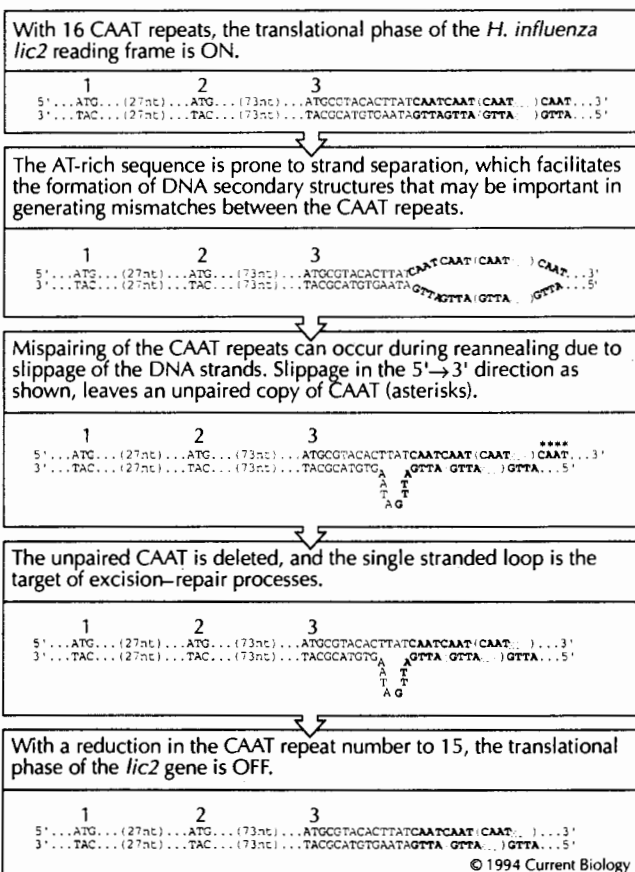


Fig. 3. Possible mutational mechanism responsible for LPS phase variation in *H. influenzae*, based on slipped-strand mispairing [30,34]. The DNA sequence of the 5' end of the *H. influenzae lic2* gene, required for the biosynthesis of a Gal α (1-4)Gal β disaccharide in the LPS core [32], is shown. Phase variation of the digalactoside is mediated through changes in the number of copies of tandem repeats of CAAT (bold); 16 copies of the tetranucleotide sequence are shown. Potential translational start codons (boxed ATG) for a long open reading frame are shown, and are positioned such that ATG 1 and 2 are in frame when *lic2* has 16 copies of CAAT, ATG 3 is in frame if *lic2* has 17 copies of CAAT and there is no ATG in-frame if there are 15 copies of CAAT.

Variable capsule production

Capsular polysaccharides of *H. influenzae* facilitate microbial survival by impeding host clearance mediated by complement and phagocytic killing. Capsule biogenesis depends on the chromosomal *cap* locus [40]. In type b organisms, the strains most often associated with invasive infections, *cap* consists of a duplicated ~18-kilobase (kb) sequence [41], an arrangement that is predictably unstable. Sequencing of *cap* has shown that it is a compound transposon, in which each copy of the capsulation genes is flanked by the insertion element *IS1016* [42]. This arrangement facilitates amplification of the capsulation genes by intragenomic homologous recombination — cells with up to five copies have been found [43]. In this way, gain or loss of the repeated segments modulates the amount of surface polysaccharide synthesized (Fig. 4) [44].

Interestingly, the reduction from two repeats to a single copy results in a capsule-deficient phenotype. This is because there is a small 1.2-kb deletion at the end of one of the repeats, leading to loss of most of that copy of *IS1016* and almost all of *bexA*, a gene essential for export of the polysaccharide. As a result, recombination between the segments of the near-tandem duplication leads to loss of a unique copy of *bexA* (Fig. 4). Thus, amplification of the capsulation genes is readily achieved, and the amplified genotype could have a selective advantage in countering host clearance mechanisms. Reduction to a single copy, on the other hand, switches off surface capsule production, which may favour microbial persistence through avoidance of anti-capsular antibodies [45] or through facilitation of attachment to, or invasion of, host epithelial cells [46].

On randomness and stress

The generation of phenotypic variations of the kind described above is a powerful strategy for responding to changes in the host environment, with its repertoire of genetic polymorphisms and immune mechanisms. Using just a few loci, a population of bacteria within a single host, derived from a single infecting clone, can generate substantial random variation which is

potentially useful for adapting, through natural selection, to a constantly changing host environment. This genetic and phenotypic variation can influence many aspects of bacterial behaviour, such as antigenicity, motility, chemotaxis, attachment to host cells, resistance to desiccation, acquisition of nutrients and sensitivity to antibiotics.

Variants generated by these specialized 'mutator' activities might appear to be programmed, or even directed, in so far as they are non-random (that is, discriminate) in their distribution across the genome. However, these mutational mechanisms are random (that is, stochastic) in the sense that specific mutations do not occur at higher rates when they are beneficial than when they are neutral or disadvantageous. Hence, such mechanisms do not challenge the fundamental tenet of neo-Darwinian theory, that mutations are random with respect to their selective utility [47,48]. Neither does the existence of these specialized mutators imply any novel mechanism that would require a reverse flow of information from the phenotype to the genotype, as has been suggested for so-called 'directed' mutations [49]. What is required instead is that natural selection has favoured the evolution of DNA sequences in certain genes ('contingency' genes), but not in others ('house-keeping' genes), that increase the probability of mutations occurring. Contingency and housekeeping genes differ in the extent to which their products interact with the unpredictable versus constant aspects of the environment, respectively.

It is possible, but not essential for our hypothesis, that the specialized mutator mechanisms have been further refined during evolution to give some degree of environmental responsiveness. That is, one might suppose that those mechanisms that cause higher mutation rates in contingency genes should be more active in cells that are under stress than those that are not. For example, the frequency with which variants are produced might be modulated by stress through the strategic placement of supercoiling-sensitive promoters adjacent to contingency genes [35].

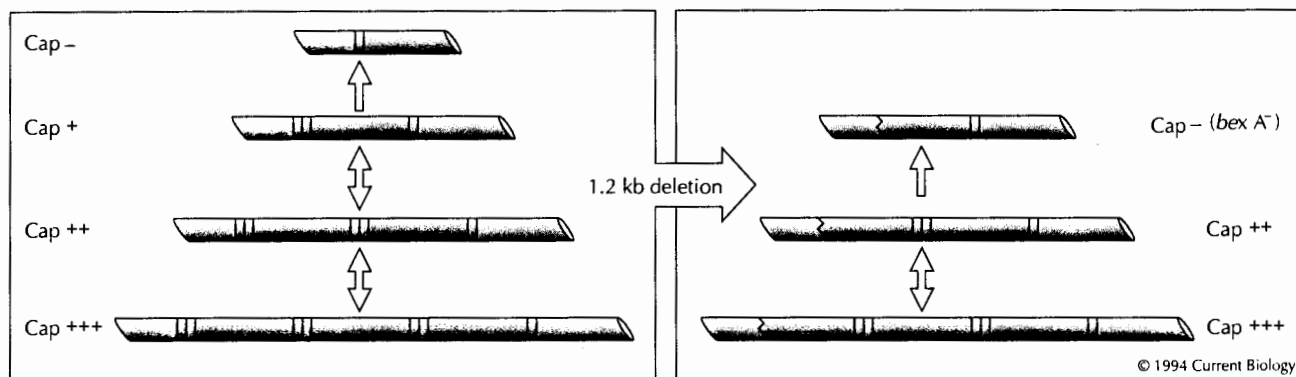


Fig. 4. The capsulation locus *cap* as a compound transposon [42]. Left, *cap* gene amplification and reduction by unequal homologous recombination between the direct repeats of *IS1016* (red) that flank the *cap* genes (blue). The *bexA* gene is also shown (green). Reversible and irreversible processes are indicated by double- and single-headed arrows, respectively. The level of capsule production (Cap) at each stage is indicated by number of + signs. Right, *cap* amplification and reduction typical of most *H. influenzae* type b strains. The truncated left-hand end of *cap* is shown, having arisen from a tandem duplication by deletion of 1.2 kb of DNA. Minimum configuration for capsulation is now the Cap++ state [37].

Recent studies on *Pseudomonas fluorescens* and *Pseudomonas tolaasii* by one of us (P.B.R, unpublished data) have shown that these bacteria generate substantial phenotypic variation when subjected to unfavourable environmental conditions. The occurrence of variant phenotypes is consistent with the hypothesis that polymorphisms within bacterial populations are triggered in response to stress [50], but this remains to be established, due to the inherent difficulties in separating the effects of intensified selection from those of an increased mutation rate. Nonetheless, it seems likely that the variants are generated in a random manner, because replicate populations (founded from a single bacterium) exhibit substantial differences in the genetic polymorphisms that arise. This genetic variation provides the raw material upon which natural selection can

operate and facilitates rapid adaptation to unpredictable environmental conditions.

Optimal mutation rates

In several population genetic models, the optimal mutation rate has been defined as that which maximizes a population's long-term geometric mean fitness in a fluctuating environment [51–54]. In particular, the optimal mutation rate must balance the genetic loads caused by deleterious mutations when the environment is constant, and by substitution of a favoured allele when the environment changes. This optimum may also be affected if there is some direct fitness cost associated with increased fidelity, for example as a consequence of producing more enzymes for DNA repair. However,

A simple model of specific versus general mutators

We present a simple mathematical model to illustrate the advantage of having higher mutation rates in those genes whose products interact with unpredictable and changing aspects of the environment. We assume an asexual haploid genome that is divided into housekeeping and contingency genes. By n and m we denote the total number of base pairs belonging to housekeeping and contingency parts of the genome, respectively. Mutation rates should be as low as possible in the housekeeping regions, but environmental fluctuations necessitate the occurrence of mutations in the contingency loci. Thus, we expect evolution to proceed towards a minimal mutation rate in the housekeeping regions and a certain optimal mutation rate in the contingency loci to provide genetic flexibility. If the processes responsible for the mutations do not distinguish between these two regions, then we expect a certain trade-off.

To be specific, we shall assume that the optimal mutation rate in the contingency loci requires, on average, k mutations per round of replication. The parameter k depends on the ecological scenario and the detailed dynamics of the environmental fluctuations. First, we consider a general mutator, defined by a constant (average) replication fidelity, q , over the whole genome. The mutation rate is defined as $1-q$. (We do not assume that the mutation rate is constant over the whole genome, only that the average is the same in contingency and housekeeping regions.) We want to maximize the probability of producing k mutations in the contingency regions and no mutation in the housekeeping regions. This probability is given by:

$$P(q) = \binom{m}{k} q^{n+m-k} (1-q)^k \quad (1)$$

The optimal q that maximises $P(q)$ is given by:

$$q = 1 - \frac{k}{n+m} \quad (2)$$

Next, we consider a specific mutator that has reduced its replication fidelity in the contingency region by a factor a (again on average). The probability of producing k mutations in the contingency regions and zero in the housekeeping is now:

$$P(a, q) = \binom{m}{k} (aq)^{m-k} (1-aq)^k q^n \quad (3)$$

For a given value q , the optimal a is:

$$a = \left(1 - \frac{k}{m}\right) / q \quad (4)$$

For a given value a , the optimal q is:

$$q = \left(1 - \frac{k}{n+m}\right) / a \quad (5)$$

The global optimum is achieved by having q as close as possible to 1, and a as given by equation (4).

Let P_g and P_s denote the maximum probabilities of producing cells with k mutations in their contingency loci for general and specific mutators, respectively. From equations (1) and (2), we obtain:

$$P_g = \binom{m}{k} \left(1 - \frac{k}{n+m}\right)^{n+m-k} \left(\frac{k}{n+m}\right)^k \quad (6)$$

Equation (3) and (4) together with $q = 1$ lead to:

$$P_s = \binom{m}{k} \left(1 - \frac{k}{m}\right)^{m-k} \left(\frac{k}{m}\right)^k \quad (7)$$

Comparing the ratio P_s/P_g gives an estimate of the advantage a specific mutator mechanism can provide. Assuming that k is much smaller than m , then:

$$P_s/P_g = \left(1 + \frac{n}{m}\right)^k \quad (8)$$

We see at once that specific mutators are strongly favoured if they are able to confine the contingency regions to small parts of the genome, that is, if m is much smaller than n . If, for example, the housekeeping part consists of a total of 10^6 base pairs, while the contingency regions make up a total of 10^3 base pairs, and $k = 1$, then an ideally designed specific mutator produces the optimal number of mutations in the contingency loci (without any in housekeeping genes) 1 000 times more frequently than an ideally designed general mutator.

Notice also that, for our analysis, it makes no difference whether the organism needs some particular mutations within the contingency regions to cope with some changed aspect of the environment, or only needs some arbitrary mutations within these regions (and, in both cases, no mutation in housekeeping regions). The first scenario may apply when a pathogen's success depends on a specific phenotypic change in order to colonize a new host genotype or to cross some intra-host barrier to infection. The second scenario may be most relevant when a pathogen's avoidance of host immunity depends on its having one (or more) of many possible changes in the structure or expression of some antigen. The equations, as they stand, now apply to the second scenario. For the first scenario, the binomial coefficients in equations (1), (3), (6) and (7) must be omitted; everything else remains unchanged.

there is some disagreement about the likelihood that mutation rates are optimized on a locus-by-locus basis, as opposed to over the entire genome [53,54].

One key issue concerns the linkage between the loci that cause differences in mutation rate and the loci at which the relevant mutations occur. If there is much recombination, then alleles that affect mutation rates may become dissociated from the adverse or beneficial effects that they have. If, however, there is tight linkage, then "selection holds mutators fully responsible, as it were, for both the good and the bad they cause" [53]. This linkage should be extremely tight in the systems that we have described, for two reasons. First, the elevated mutation rates at contingency loci are caused by the properties of the DNA sequences at those very same loci. Second, even if the mutation rates of contingency genes are also affected by other loci, the effective linkage is still likely to be very strong because of the clonal expansion of pathogenic bacteria within a host.

Another key issue concerns the relative likelihood of beneficial and deleterious mutations. In the special case of a locus that switches between two alleles, each adapted to a different environment, the optimal mutation rate is approximately equal to the frequency with which the environment alternates between these two states [54]. For example, if the environment changes every 100 generations, then the optimal rate of mutational switching is about 0.01. Exactly how this result might be generalized when there are more than two alleles and environments is not yet clear; nonetheless, it shows that the optimal mutation rate may be very high indeed if the environment changes frequently and if the potential for deleterious mutations is not too great. As we have emphasized, the host environment for pathogenic bacteria is expected to change very rapidly. And we have also emphasized that molecular mechanisms exist that allow deleterious mutations to be substantially reduced (but not eliminated) by having contingency genes that are much more mutable than housekeeping genes. A simple model illustrating the potential evolutionary advantage to an organism that has compartmentalised its genome into regions of high (contingency) and low (housekeeping) mutation rates is explained in the box.

Summary

We conclude by restating our thesis as a general hypothesis: mutation rates vary among sites in a genome, and this variation is adaptive because it promotes evolutionary flexibility in the face of environmental change, without necessarily increasing the overall load of deleterious mutations. In particular, we expect mutation rates to be higher in genes whose products interact with the environment in unpredictable ways. Such unpredictability implies that the relevant aspects of the environment change frequently and, moreover, that these changes cannot be tracked by simply varying the dosage of a limited repertoire of gene products. We refer to those genes whose products interact with the environment in such unpredictable

ways as contingency genes, in contrast to housekeeping genes whose products interact with the environment in a more predictable fashion. Although mutations are non-random (that is, not uniformly distributed) across sites in the genome, they are presumed to be random with respect to their immediate selective value to the organism; that is, a specific mutation does not occur at a higher rate in an environment where it confers an advantage than in an otherwise similar environment where it is neutral or disadvantageous.

The hypothesis that variation in mutation rates among loci is adaptive can be further evaluated using either comparative [55] or experimental [56] approaches. One could compare mutation rates between different loci within a particular organism (provided that genes can be identified *a priori* as serving either housekeeping or contingency functions); or one could compare mutation rates at a particular locus for several different organisms, where that gene has a housekeeping function in some species but serves a contingency function in others. In the experimental approach, populations of bacteria could be subjected (over a period of hundreds of generations) to frequent changes in the selection pressures acting on a particular gene product. The prediction is that the mutation rate at the target locus would increase relative to its ancestral value, whereas housekeeping genes should not exhibit any increase in their mutation rates. (If mutation rates also increased in housekeeping genes, that would support a general elevation in mutation rates in unpredictable environments, but it would not provide support for the hypothesis that variation among loci in mutation rates is adaptive.)

We have emphasized the application of our hypothesis to pathogenic bacteria, which are well studied and may have taken special advantage of the opportunities afforded by local variation in mutation rates because of the severe challenges imposed by their hosts' defensive capacities. The distinction between contingency and housekeeping genes may be especially clear for pathogens. But while we have focused on the applicability of our hypothesis to pathogenic bacteria, we would like to emphasize, in closing, that the potential advantages of intergenic variation in mutation rates may apply to their hosts, or to any other organism whose genome can be crudely divided into contingency and housekeeping functions. In this regard, it is interesting to note that, almost half a century ago, Haldane [57] suggested "it is an advantage to a (host) species to be biochemically diverse, and even to be mutable as regards genes concerned in disease resistance".

Acknowledgements: E.R.M. wishes to acknowledge scientists of the Molecular Infectious Diseases Group whose work has stimulated and formed the basis of parts of this review. This research is supported by programme grants from the Medical Research Council and the Wellcome Trust. P.B.R. is an AFRC Research Fellow. M.A.N. is a Wellcome Trust Senior Research Fellow and an E.P. Abraham Junior Research Fellow at Keble College, Oxford. R.E.L. is supported by the U.S. National Science Foundation (BIR-9120006 and DEB-9249916) and held a Guggenheim Foundation Fellowship while a Visiting Fellow at All Souls College, Oxford. R.E.L. thanks Paul Sniegowski for valuable comments.

References

1. STOCK EP, STOCK AM, MOTTONEN JM: Signal transduction in bacteria. *Nature* 1990, 344:395-400.
2. MOXON ER, MURPHY PA: *Haemophilus influenzae* bacteremia and meningitis resulting from survival of a single organism. *Proc Natl Acad Sci U S A* 1978, 75:1534-1536.
3. RUBIN LG: Bacterial colonisation and infection resulting from multiplication of a single organism. *Rev Infect Dis* 1987, 9:488-493.
4. FALKOW S: Molecular Koch's postulates applied to microbial pathogenicity. *Rev Infect Dis* 1988, 10 (suppl):274-276.
5. FLOR HH: The complementary genic systems in flax and flax rust. *Adv Genet* 1956, 8:29-54.
6. DAWKINS R, KREBS JR: Arms races within and between species. *Proc R Soc Lond [Biol]* 1979, 205:489-511.
7. LENSKI RE, LEVIN BR: Constraints on the coevolution of bacteria and virulent phage: A model, some experiments, and predictions for natural communities. *Am Nat* 1985, 125:585-602.
8. VAN VALEN L: A new evolutionary law. *Evol Theory* 1973, 1:1-30.
9. BURNET FM: *The Clonal Selection Theory of Immunity*. Nashville, Tennessee: Vanderbilt University Press; 1959.
10. GOLUB ES, GREEN DR: *Immunology: A Synthesis*, 2nd edn. Sunderland, Massachusetts: Sinauer Press; 1991.
11. JAENIKE J: An hypothesis to account for the maintenance of sex within populations. *Evol Theory* 1978, 3:191-194.
12. LIVELY CM: Evidence from a New Zealand snail for the maintenance of sex by parasitism. *Nature* 1987, 328:519-521.
13. SEGER J, HAMILTON WD: Parasites and sex. In *The Evolution of Sex*. Edited by Michod RE, Levin BR. Sunderland, Massachusetts: Sinauer Press; 1988:176-193.
14. HAMILTON WD, AXELROD R, TANESE R: Sexual reproduction as an adaptation to resist parasites (A Review). *Proc Natl Acad Sci U S A* 1990, 87:3566-3573.
15. DAVEY RP, REANNEY DC: Extrachromosomal genetic elements and the adaptive evolution of bacteria. *Evol Biol* 1980, 13:113-147.
16. LEVIN BR: The evolution of sex in bacteria. In *The Evolution of Sex*. Edited by Michod RE, Levin BR. Sunderland, Massachusetts: Sinauer; 1988:194-211.
17. MAYNARD SMITH J: The evolution of prokaryotes: does sex matter? *Annu Rev Ecol Syst* 1990, 21:1-12.
18. MAYNARD SMITH J, DOWSON CG, SPRATT BG: Localized sex in bacteria. *Nature* 1991, 349:29-31.
19. BORST P, GREAVES DR: Programmed gene rearrangements altering gene expression. *Science* 1987, 235:658-667.
20. PLASTERK RHA: Programmed DNA Rearrangements. *Trends Genet* 1992, 8 (Special Issue):403-463.
21. TONEGAWA S: Somatic generation of antibody diversity. *Nature* 1983, 302:575-581.
22. VAN ALPHEN L, GEELEN VAN DEN BROEK L, BLAAS L, VAN HAM M, DANKERT J: Blocking of fimbria-mediated adherence of *Haemophilus influenzae* by sialyl gangliosides. *Infect Immun* 1991, 59:4473-4477.
23. WEBER A, HARRIS K, LOHRKE S, FORNEY L, SMITH AL: Inability to express fimbriae results in impaired ability of *Haemophilus influenzae* b to colonize the nasopharynx. *Infect Immun* 1991, 59:4724-4728.
24. VAN ALPHEN L, POOLE J, OVERBEEKE M: The Anton Blood group antigen is the erythrocyte receptor for *Haemophilus influenzae*. *FEMS Microbiol Lett* 1986, 37:69-71.
25. VAN HAM M, VAN ALPHEN L, MOOI FR, VAN PUTTEN JPM: Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell* 1993, 73:1187-1196.
26. WILLEMS R, PAUL A, VAN DER HELDE HGJ, TER AVEST AR, MOOI FR: Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *EMBO J* 1990, 9:2803-2809.
27. OLYHOEK AJM, SARKARI J, BOPP M, MORELLI G, ACHTMAN M: Cloning and expression in *Escherichia coli* of *opc*, the gene for an unusual class 5 outer membrane protein from *Neisseria meningitidis*. *Microb Pathol* 1991, 11:249-257.
28. YOGEV D, ROSENGARTEN R, WATSON-MCKOWN R, WISE KS: Molecular basis of *Mycoplasma* surface variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5' regulatory sequences. *EMBO J* 1991, 10:4069-4079.
29. WISE KS, YOGEV D, ROSENGARTEN R: Antigenic variation in *Mycoplasmas*. In *Molecular Biology and Pathogenesis*. Edited by Maniloff J, McElhaney RN, Finch LR, Baseman JB. Washington, DC: American Society for Microbiology; 1992.
30. WEISER JN, LOVE JM, MOXON ER: The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. *Cell* 1989, 59:657-665.
31. VIRJI M, WEISER JN, LINDBERG AA, MOXON ER: Antigenic similarities in lipopolysaccharides of *Haemophilus* and *Neisseria* and expression of digalactoside structure also present on human cells. *Microb Pathog* 1990, 9:441-450.
32. HIGH NJ, DEADMAN ME, MOXON ER: The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the *Haemophilus influenzae* lipopolysaccharide epitope aGal(1-4)Gal. *Mol Microbiol*, in press.
33. WEISER JN, MASKELL DJ, BUTLER PD, LINDBERG AA, MOXON ER: Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. *J Bacteriol* 1990, 172:3304-3309.
34. LEVINSON G, GUTMAN GA: Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 1987, 4:203-221.
35. HIGGINS CF, DORMAN CJ: DNA supercoiling: a role in the regulation of gene expression and bacterial virulence. In *Bacterial Protein Toxins*. Edited by Rappuoli R et al. New York, Gustav Fischer; 1990:293-301.
36. MÄKELÄ PH, STOCKER BAD: Genetics of lipopolysaccharide. In *Handbook of Endotoxin Vol. 1 Chemistry of Endotoxin*. Edited by Riethel ET. Amsterdam: Elsevier Science Publishers; 1984:59-137.
37. STERN A, BROWN M, NICKEL P, MEYER TF: Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* 1986, 47:61-67.
38. STERN A, MEYER TF: Common mechanisms controlling phase and antigenic variation in pathogenic neisseriae. *Mol Microbiol* 1987, 1:5-12.
39. BELLAND RJ: H-DNA formation by the coding repeat elements of neisserial *opa* genes. *Mol Microbiol* 1991, 5:2351-2360.
40. KROLL JS, ZAMZE S, LOYNDS B, MOXON ER: Common organization of chromosomal loci for the production of different capsular polysaccharides in *Haemophilus influenzae*. *J Bacteriol* 1989, 171:3343-3347.
41. HOISETH SK, MOXON ER, SILVER RP: Genes involved in *Haemophilus influenzae* type b capsule expression are part of an 18-kb tandem duplication. *Proc Natl Acad Sci U S A* 1986, 83:1106-1110.
42. KROLL JS, LOYNDS BM, MOXON ER: The *Haemophilus influenzae* capsulation gene cluster: a compound transposon. *Mol Microbiol* 1991, 5:1549-1560.
43. HOISETH SK, CORN PG, ANDERS J: Amplification status of capsule genes in *Haemophilus influenzae* type b clinical isolates. *J Infect Dis* 1992, 165 (suppl 1):S114.
44. KROLL JS: The genetics of encapsulation in *Haemophilus influenzae*. *J Infect Dis* 1992, 165 (Suppl 1):S93-96.
45. HOISETH SK, CONNELLY CJ, MOXON ER: Genetics of spontaneous, high-frequency loss of b-capsule expression in *Haemophilus influenzae*. *Infect Immun* 1985, 49:389-395.
46. ST GEME JW III, FALKOW S: Loss of capsule expression by *Haemophilus influenzae* type b results in enhanced adherence to and invasion of human cells. *Infect Immun* 1991, 59:1325-1333.
47. LENSKI RE, MITTLER JE: The directed mutation controversy and neo-Darwinism. *Science* 1993, 259:188-194.
48. RAINEY PB, MOXON ER: Unusual mutational mechanisms and evolution. *Science* 1993, 260:1958.
49. CAIRNS J, OVERBAUGH J, MILLER S: The origin of mutants. *Nature* 1988, 335:142-145.
50. RAINEY PB, MOXON ER, THOMPSON IP: Intraclonal polymorphism in bacteria. *Adv Microb Ecol*, in press.
51. KIMURA M: On the evolutionary adjustment of spontaneous mutation rates. *Genet Res* 1967, 9:23-24.

52. LEVINS R: Theory of fitness in a heterogeneous environment. VI. The adaptive significance of mutation. *Genetics* 1967, 56:163-178.
53. LEIGH EG: The evolution of mutation rates. *Genetics* 1973, 73 (suppl):1-18.
54. ISHII K, MATSUDA H, IWASA Y, SASAKI A: Evolutionarily stable mutation rate in a periodically changing environment. *Genetics* 1989, 121:163-174.
55. HARVEY PH, PAGEL MD: *The Comparative Method in Evolutionary Biology*. Oxford: Oxford University Press; 1991.
56. LENSKI RE: Experimental evolution. In *Encyclopedia of Microbiology*, vol. 2. Edited by Lederberg J. San Diego, California: Academic Press; 1992:125-140.
57. HALDANE JBS: Disease and evolution. *La Ricerca Scientifica* 1949, 19 (suppl):3-11.
58. ROBERTSON BD, MEYER TF: Genetic variation in pathogenic bacteria. *Trends Genet* 1992, 8:422-427.
59. PLASTERK RHA, SIMON MI, BARBOUR AG: Transposition of structural genes to an expression sequence on a neoplasmid causes antigenic variation in the bacterium *Borrelia hermsii*. *Nature* 1985, 318:257-263.
60. CARAMORI T, ALBERTINI AM, GALIZZI A: *In vivo* generation of hybrids between two *Bacillus thuringiensis* insect-toxin-encoding genes. *Gene* 1991, 98:37-44.
61. KROLL JS, HOPKINS J, MOXON ER: Capsule loss in *Haemophilus influenzae* type b occurs by recombination-mediated disruption of a gene essential for polysaccharide export. *Cell* 1988, 53:347-356.
62. ABRAHAM JM, FREITAG CS, CLEMENTS JR, EISENSTEIN BI: An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc Natl Acad Sci USA* 1985, 82:5724-5727.
63. DORMAN CJ, HIGGINS CF: Fimbrial phase variation in *Escherichia coli*: dependence on integration host factor and homologies with other site-specific recombinases. *J Bacteriol* 1987, 169:3840-3843.
64. KIM S-R, KOMANO TJ: Nucleotide sequence of the R721 shufflon. *J Bacteriol* 1992, 174:7053-7058.
65. MARRS CF, RUEHL WR, SCHOOLNIK GK, FALKOW S: Pilin gene phase variation of *Moraxella bovis* is caused by an inversion of the pilin genes. *J Bacteriol* 1988, 170:3032-3039.
66. BRUMBLY SM, CARNEY BF, DENNY TP: Phenotype conversion in *Pseudomonas solanacearum* due to spontaneous inactivation of PhcA, a putative LysR transcriptional regulator. *J Bacteriol* 1993, 175:5477-5487.
67. BARTLETT DH, WRIGHT ME, SILVERMAN M: Variable expression of extracellular polysaccharide in the marine bacterium *Pseudomonas atlantica* is controlled by genome rearrangement. *Proc Natl Acad Sci USA* 1988, 85:3923-3927.
68. VAN DER WOUDE MW, BRAATEN BA, LOW DA: Evidence for global regulatory control of pilus expression in *Escherichia coli* by Lrp and DNA methylation: model building based on analysis of *pap*. *Mol Microbiol* 1992, 6:2429-2435.
69. GREWAL SIS, RAINEY PB: Phenotypic variation of *Pseudomonas putida* and *P. tolaasii* affects the chemotactic response to *Agaricus bisporus* mycelial exudate. *J Gen Microbiol* 1991, 137:2761-2768.
70. RAINEY PB: Phenotypic variation of *Pseudomonas putida* and *P. tolaasii* affects attachment to *Agaricus bisporus* mycelium. *Gen Microb* 1991, 137:2769-2779.
71. FLORES M, GONZALEZ V, PARDO MA, LEIJA, MARTINEZ E, ROMERO D, PINERO D, DAVILA G, PALACIOS R: Genomic instability in *Rhizobium phaseoli*. *J Bacteriol* 1988, 170:1191-1196.

Received: 23 September 1993; revised 3 November 1993.