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 9DX, 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 modulates 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 catabolic 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 host's 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 advantage 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.
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 generic 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 selective advantages to 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 [10-16].

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 three genomic events, such as recombination, or intragenomic events, such as mutations. The contribution of intragenomic mechanisms (reviewed in 15-18) to the generation of phenotypic diversity in host 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 discriminative mutations have properties:

<table>
<thead>
<tr>
<th>Table 1. Examples of genetic mechanisms generating random phenotypic variation at high frequency in specific loci.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mechanism</strong></td>
</tr>
<tr>
<td>Gene conversion</td>
</tr>
<tr>
<td>Homologous recombination</td>
</tr>
<tr>
<td>Site-specific recombination</td>
</tr>
<tr>
<td>Cytotoxicity repeats</td>
</tr>
<tr>
<td>Site-specific transcription</td>
</tr>
<tr>
<td>Site-specific mutagenesis</td>
</tr>
<tr>
<td>Unknown</td>
</tr>
</tbody>
</table>

Note: (1) For further examples. © 1994 Cold Spring Harbor Laboratory.
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). Though 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 capsule 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](https://example.com/fig1.png)  
**Fig. 1.** The drawing represents a cross-section through 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).
Fimbriae phase variation

Fimbriae are proteaceous appendages that confer on bacteria the ability to adhere to and to epithelial cells, through adherence to a sialic-acid containing lactosylceramide host receptor [22]. The fimbriae are composed of a single, repeating 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 presence of fimbriae facilitates clearance by host cells [24]. The gene cluster required for H. influenzae fimbria synthesis includes hfaA, which encodes the subunit polypeptide, and hfaB, which is a homologue of several known periplasmic chaperones. Sequencing these genes showed that hfaA and hfaB 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 transcription 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 hfaA and hfaB 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 hfaA and hfaB polypeptide 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 transcription of RNA polymerase, as it determines the relative positioning of these 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 switching of fimbriae expression is coordinated through transcriptional control of hfaA and hfaB [25].

The variable expression of the subunits of Bordetella pertussis fimbriae may also be controlled by a repeat sequence tract. In this case a tract of -ctoine 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 Cpc, an outer membrane protein of Yersinia enterocolitica, is also modulated by a homopolymeric tract of -ctoine residues [27]. In a similar manner, variable transcription of cell surface -ctoine is modulated in the Mycoplasma hyorhinis membrane is subject to phase variation. Vip expression is controlled at the transcriptional level by a repeat region of nucleotides on the length of the A phase tract is not critical for cell surface expression of the -ctoine. However, it is thought that changes in the length of the A phase tract may affect the expression of a second structure between the -35 and -10 cis-DNA-h structures [26,27].

Lipopolysaccharide antigenic variation

Lipopolysaccharide (LPS) is a macromolecule that is unique to Gram-negative bacteria, and 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. Lipid A is known to be a major virulence determinant of H. influenzae. Phenotype 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 LPS by 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 [28]. As the different core saccharide structures switch randomly and independently of each other, a population of LPS variants is generated by a single clone and residing within a single host can generate an extensive repertoire of variant LPS antigens.

The molecular basis of the variable expression of one of these saccharide structures, galact-1,4-galp [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 saccharide galact-1,4-galp, individual colonies showed either strong, intermediate, or undetectable binding. Mutants have been isolated that do not express the digalactose structure, and one of the mutations maps to a gene designated hla-2 which is essential for synthesis of the digalactoside [32]. At the 5' end of the gal 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 ic2 frame shift is presumably similar to that underlying flhB phase variation [34], a re- 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 reversibly adapting to 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 ic2 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 coregulate 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 with 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 hentable. Nonetheless, microheterogeneity may be useful in producing many different surface configurations, increasing the likelihood that at least one cell gets past the 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 hentable and that which is not, may be important for the evolutionary success of K. pneumoniae, 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 CYTGR repeats, in the region of the gene encoding the hydrophobic core of the leader peptide. Variation in the numbers of CYTGR 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 could be a potential target for single-strand-specific nucleases [39].
Variable capsule production

Capsular polysaccharides of *H. influenzae* facilitate microfical survival by preventing host clearance mediated by complement and phagocytosis. Capsular 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 capsule genes is flanked by the inversion element IS1016 [42]. This arrangement facilitates amplification of the capsule gene 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 two of that copy of IS1016 and almost all of heed, 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 heed (Fig. 4). Thus, amplification of the capsule 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 favor 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 behavior, 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; 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 [46,47]. 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 [47]. What is required instead is that natural selection has favored the evolution of DNA sequences in certain genes (contingency genes), but not in others (housekeeping 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 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 super-coiling-sensitive promoters adjacent to contingency genes [49].

**Fig. 4.** The capsule locus cap in 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 break 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 Hi influenzae type b strains. The truncated left-hand end of cap is shown. **Young et al.** from a tandem duplication by deletion of 1.2 kb of DNA. Minimum configuration for capsule is now the Cap++ state [47].
Recent studies on *Pseudomonas fluorescens* and *Pseudomonas aeruginosa* 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 stresses that are not pre-existing, due to the inherent difficulties in separating the effects of unrestrained selection from those of an intrinsic mutation rate. Nonetheless, it seems likely that the variants are generated in a random manner, because replicating 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 load caused by deleterious mutations when the environment is constant, and by substitution of a favoured allele when the environment changes. This optimality may also be described 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 mutations

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 genotype that is divided into housekeeping and contingency genes. By a and n we denote the total number of base pairs belonging to housekeeping and contingency genes 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 a contingency locus. Thus, we expect evolution to proceed towards a minimal mutation rate in the housekeeping regions and a constant mutation rate in the contingency locus to provide genetic flexibility. If the process responsible for the mutations do not distinguish between their two regions, then we expect a certain average.

To be specific, we shall assume that the optimal mutation rate in the contingency locus maximizes, on average, k mutations per round of replication. The parameter $k$ depends on the ecological scenario and the dynamical changes of the environmental fluctuations. First, we consider a general mutation, defined by a constant averaged replication fidelity, q, over the whole genome. The mutation rate is defined as $q = 1 - q$. We do not assume that the mutation-rate is constant over the whole genome, that is, 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:

$$P_{k} = inom{n}{k} q^{n-k}(1-q)^{k}$$

(1)

The optimal $k$ that maximizes $P_{k}$ is given by:

$$k = \frac{n}{m}$$

(2)

Next, we consider a specific mutation that has reduced as replication fidelity to a factor $r$, in the contingency region by a factor $s$, again an average. The probability of producing $k$ mutations in the contingency regions and zero in the housekeeping's sum:

$$P_{k} = \binom{n}{k} (r)^{n-k}(1-s)^{k}$$

(3)

For a given value of $k$, the optimal $a$ is:

$$a = \left(\frac{1 - s}{1 - r}\right) k$$

(4)

For a given value of $k$, the optimal $a$ is:

$$a = \left(\frac{1 - s}{1 - r}\right) k$$

(4)

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

Let $P_{r}$ and $P_{s}$ denote the maximum probabilities of producing replicas with $k$ mutations in their contingency region for general and specific mutations, respectively, from equation (3) and (2), we obtain:

$$P_{r} = \binom{n}{k} (1 - s)^{k}$$

(5)

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

$$P_{s} = \binom{n}{k} (1 - r)^{k}$$

(6)

Comparing the ratio $P_{r}/P_{s}$ gives an estimate of the advantage of a specific mutation strategy that can provide. Assuming that $r$ is much smaller than $s$ or $m$, then:

$$P_{r}/P_{s} = \left(\frac{1 - s}{1 - r}\right)^{k}$$

(7)

We see at once that specific mutations are strictly favoured if they are able to confine the contingency regions to small parts of the genome, that is, if $m$ is much smaller than $n$. For example, the housekeeping part consists of a total of 10 base pairs, while the contingency regions make up a total of 100 base pairs. While the contingency regions make up a total of 100 base pairs, the specific region is much smaller than the general mutation. Notice also that, for our analysis, it makes no difference whether the organism needs some particular mutation within the contingency regions to cope with some changed aspect of the environment, or only needs some arbitrary mutation 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-prototype or to cross some intra-host barrier to infection. The second scenario may be more relevant when a pathogen's avoidance of host immunity depends on it 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), (5), (6) and (7) may 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 [55,54].

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 selection that affects mutation rate may become dissociated from the adverse or beneficial effects that they have. If, however, there is tight linkage, then selection holds mutator alleles 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 internalized varies 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 compartmentalized 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; the specific mutation does not occur in a higher rate in an environment where it confers an advantage than in an otherwise similar environment where it is neutral or deleterious.

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 as serving either housekeeping or contingency functions), or one could consider 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 loci 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 suggest 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. It is worth noting that we 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 ineradicable variation in mutation rates may apply to their hosts, or to any other organism whose genome can be credibly divided into contingency and housekeeping functions. In this regard, it is interesting to note that, about 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'.

Acknowledgments: E.R.M. wishes to acknowledge the scientists of the Molecular Infections Diseases Group whose work has stimulated and formed the basis of this review. This research is supported by programme grants from the Medical Research Council and the Wellcome Trust. P.B.J. is an AFB Research Fellow. M.A.N. is a Wellcome Trust Senior Research Fellow and an F.P. Alttamian Junior Research Fellow at Kelv College, Oxford. P.B.J. is supported by the U.S. National Science Foundation (IBN-9202946 and IBN-9216593) and holds a Guggenheim Foundation Fellowship. While a Visiting Fellow at All Souls College, Oxford. P.B.J. thanks Paul Stiegmann for valuable comments.
References


