PERSPECTIVES

Source–sink dynamics of virulence evolution

Evgeni V. Sokurenko, Richard Gomulkiewicz and Daniel E. Dykhuizen

Abstract | To understand the evolution of genetic diversity within species — bacterial and others — we must dissect the first steps of genetic adaptation to novel habitats, particularly habitats that are suboptimal for sustained growth where there is strong selection for adaptive changes. Here, we present the view that bacterial human pathogens represent an excellent model for understanding the molecular mechanisms of the adaptation of a species to alternative habitats. In particular, bacterial pathogens allow us to develop analytical methods to detect genetic adaptation using an evolutionary 'source–sink' model, with which the evolution of bacterial pathogens can be seen from the angle of continuous switching between permanent (source) and transient (sink) habitats. The source–sink model provides a conceptual framework for understanding the population dynamics and molecular mechanisms of virulence evolution.

Emerging and opportunistic diseases are caused by a microorganism invading a new habitat, either a new host species (as in the case of SARS and tuberculosis) or a new tissue compartment within the same host species (as in the case of *Escherichia coli*-mediated urinary tract infection and bacterial meningitis). Often, the pathogen can live for many generations in the new habitat and, as natural selection favours mutations that provide an immediate advantage (that is, evolution is short-sighted), mutations that are advantageous in the new habitat will be selected even if they are disadvantageous in the original habitat. This selection should leave a 'signature' or pattern of sequence variation in the genes targeted by natural selection, with which the genes and the type of selection can be recognized.

In bacteria, there are two sources of genetic variation that can facilitate adaptation to new habitats — horizontal acquisition of new genes and alteration of existing genes. Horizontal transfer between different bacterial clones, or even species, leads to the incorporation of new plasmids or multigene chromosomal insertions (islands) into the genome, and to gene replacement or modification by recombination[1]. The second type of variation includes point mutations in pre-existing genes or regulatory regions, and gene shuffling, duplication, inversion and deletion[2]. Traditionally, particularly for genomic islands (BOX 1), the signature looked for is the significantly higher frequency of a specific island or gene in strains isolated from the alternative habitat. However, this procedure is difficult, particularly when the selected changes are unknown.

In this paper, we describe a 'source–sink' model for adaptive evolution, which provides us with another method to find genetic changes that have been selected for in the alternative habitat. The source–sink model is derived from population ecology, where the 'sink' population is only maintained by immigration from the 'source' population. The evolutionary source–sink model presented here allows the population in the sink to maintain itself, but only transiently.

We emphasize here that the source–sink model of evolution can be applied to most bacterial pathogens and suggest a phylogeny-based method to search for genes undergoing source–sink evolution. It has previously been proposed that a shift from a commensal to a pathogenic lifestyle is likely to result in bacterial pathogens evolving adaptively within the host, during the course of an infection[3,4]. It has also been postulated, however, that the emergence of a new professional pathogen (that is, an organism adapted for circulation within the host population as a pathogen) from a non-pathogen is expected to be relatively rare, because it is likely to require the selection of multiple adaptive changes[5]. The source–sink model is applied to pathogens at early or intermediate stages of adaptation (before the pathogenic lifestyle might become self-sustainable), which we believe is the most common outcome.

The source–sink model

Source–sink ecological models[6] were developed, and are currently applied, in the population ecology of animals and plants. They relate to species that are distributed across source habitats, where populations are self-sustaining, and sink habitats, where populations can be maintained continuously only by immigration from established source habitats. In sinks, deaths eventually exceed births whereas the opposite is true in sources. For example, populations of an annual phlox plant in loose sandy loam are self-sustaining, and provide migrants into adjacent populations in denser soil that produce fewer seeds than are required for replacement[7]. Another example is the habitat dynamics of the three-spine stickleback fish, *Gasterosteus aculeatus*, which has a large ocean population that functions as a self-sustainable source habitat from which colonization of post-glacial freshwater lakes can occur[8]. These lakes, however, are relatively unstable sink habitats as they will disappear during the next ice age. Therefore, the sink population can maintain itself without constant immigration, but only transiently, as the population remains relatively small and has a high probability of extinction either through stochastic effects or habitat destruction.

Whether a habitat is a source or a sink depends on the organism, and the abiotic and biotic components of the environment[9]. Therefore, one organism's sink could be another organism's source. Also, environmental change could convert a sink into a
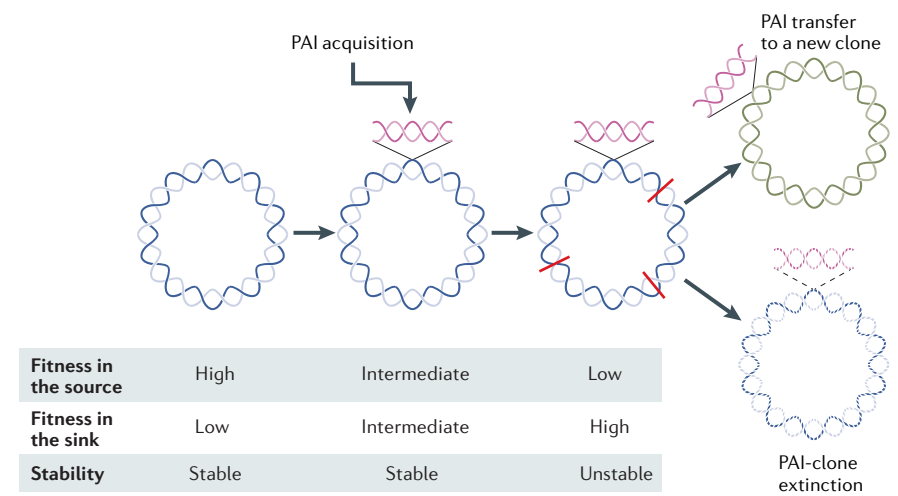
source and *vice versa*. Finally, a sink habitat can change into a source habitat through adaptive evolution of the organism.

A body of theory has developed over the past decade that describes how species could adaptively evolve in sink habitats[10]. These mathematical models of source–sink evolutionary dynamics have considered a broad range of general population and habitat scenarios, including 'closed' sinks (FIG. 1a), in which a population that is isolated in a sink habitat with no access to or from other habitats must adapt to avoid extinction[11–13]. Closed-sink models also apply to an isolated population that occupies a habitat that is initially a source habitat but becomes a sink owing to a gradual or abrupt change for the worse in the habitat conditions. Another scenario, called the 'black-hole sink' model (FIG. 1b), involves the recurrent, one-way influx of migrants from the source habitat to the sink habitat[14–18]. Other models include the reciprocal sink, where there is a constant exchange of migrants between source and sink habitats (FIG. 1c); reciprocal migration can occur either between distinct sources and sinks[19–22] or along continuous environments that gradually change between source and sink habitats[9,23].

A general conclusion of these theoretical considerations is that the evolution of sink populations is rare when the sink populations are small and short-lived owing to the harshness of the habitat and low immigration rates. This indicates that the sink population size will decrease rapidly, making it unlikely that the genetic changes required for adaptation will be fixed before extinction of the population. However, the theory does indicate that factors such as milder sinks, larger sink populations, intermediate migration rates and, in harsh sinks, environmental fluctuations, should boost the rate of adaptation to the sink habitat by increasing the number of generations that survive in the sink before the population is likely to go extinct. Once some adaptation takes place, the time to extinction is longer, allowing more time for further adaptation[5].

Despite various theoretical studies, direct empirical evidence for genetic diversification that can be attributed to source–sink evolution is sparse. Examples of sink-driven genetic diversification can be obtained only in a few macroorganisms where the sink populations exist for a long enough time for evolution to take place. For example, the three-spine stickleback fish that transiently colonizes (from an evolutionary perspective) glacial freshwater lakes undergoes rapid evolution for loss of armour and for changes

Box 1 | The source–sink dynamics of pathogenicity islands

Fitness in the source	High	Intermediate	Low
Fitness in the sink	Low	Intermediate	High
Stability	Stable	Stable	Unstable

Pathogenicity islands (PAIs) are horizontally acquired, large chromosomal insertions that are predominantly found in pathogenic species or clones[64]. Although some PAIs were shown to be adaptive for a commensal lifestyle[65], others encode multiple traits that are clearly adaptive for virulence, that is, colonization of the virulence habitat. However, their evolutionary origin and population dynamics are not well understood. We believe that the source–sink model presented in this article could provide insights into the dynamics of these elements and any horizontally acquired genes that could be adapted to shifting ecological habitats in bacterial species.

Consider a PAI that is specifically adapted to increase bacterial fitness in the virulence habitat. In contrast to pathoadaptive mutations, PAIs are not expected to be acquired by a recipient clone in the virulence habitat (which is normally sterile) but in reservoir habitats (which are commonly colonized by different species and clones), where the chances of encountering a PAI donor are significantly higher. Similar to the acquisition of pathoadaptive mutations, the acquisition of virulence-adapted PAIs is accompanied by a trade-off in the reservoir habitat, owing to possible interference with the function of the pre-existing genome, or at least owing to an increased genome size. However, the selection against PAI acquisition might not be as strong as the selection against pathoadaptive mutations that commonly knockout or alter the function of reservoir-adapted traits that are important in the source habitat. At the same time, the advantage that a multi-gene PAI provides in the virulence habitat is likely to be significantly greater than the advantage of pathoadaptive mutations in individual genes. Therefore, the trade-off effect of PAIs could be compensated by a significantly increased ability of PAI-carrying bacteria to spread into and colonize the virulence habitat that should, in turn, result in increased backflow of the PAI-carrying bacteria from the virulence habitat to the reservoir habitat (see figure; the PAI is shown in pink and pathoadaptive mutations are shown as red bars). Therefore, PAI-carrying clones might be able to sustain themselves indefinitely in nature by continuous shifting between the reservoir and virulence habitats.

However, the long-term stability of the PAI-carrying clone will be ensured only if dependence on the reservoir habitat is eliminated or significantly diminished; in other words, if this clone becomes a professional pathogen that can directly transmit and sustain itself between virulence habitats. In this case, PAIs could be fixed in a single clone and, eventually, lead to the emergence of new species or subspecies. When the reservoir habitat remains an important part of the ecology of the PAI-carrying clone, the overall stability of the clone will be increasingly jeopardized by the inevitable accumulation of pathoadaptive mutations in the course of repeated invasion of the virulence habitat. Pathoadaptive evolution will primarily target the pre-existing region of the genome adapted to the reservoir habitat and not the virulence habitat; pathoadaptive mutations will therefore eventually accumulate in PAI-carrying bacteria to a level where trade-off in the reservoir will not be compensated by the advantage in the virulence habitat and these PAI-carrying clone(s) will be eliminated. However, during this period, the PAI itself could be transferred to a source-adapted bacterium and form a new PAI-carrying clone that will undergo the evolutionary source–sink cycle again. In this way, the long-term stability of such transient PAIs will be ensured, despite the instability of PAI-carrying clones.

If the hypothesis of transient PAIs is correct, the recipient clones should always be of relatively recent origin and demonstrate progressive accumulation of pathoadaptive mutations. The transient PAI hypothesis allows PAI evolution to be modelled without assuming constant lateral transfer between species. Experimental evidence for the transient PAI hypothesis remains to be provided and the model itself requires further refinement. However, development of this hypothesis shows how the source–sink model could be applied to understand the evolution of bacterial pathogens by horizontal transfer and the acquisition of pathoadaptive mutations, and the population dynamics and genetic diversification of bacterial species that circulate in alternative habitats.
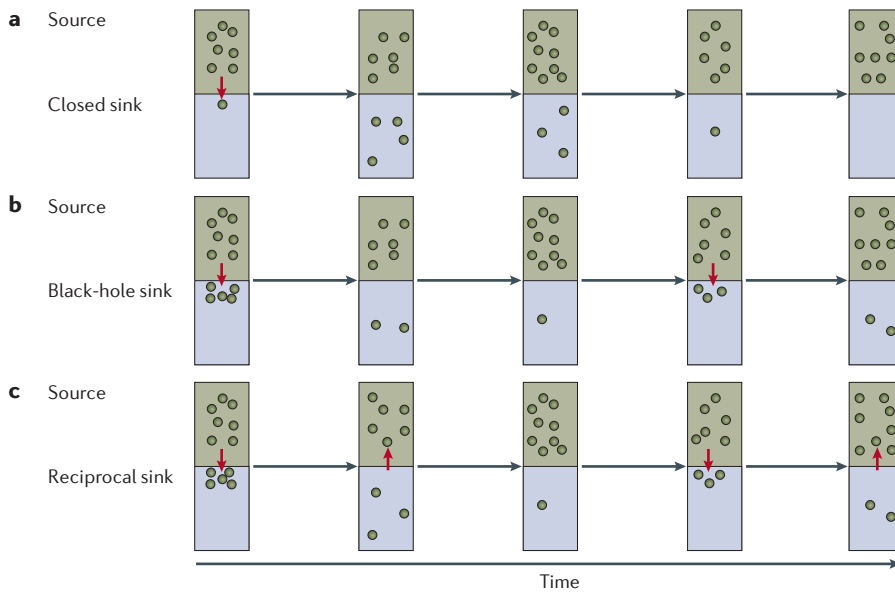
Figure 1 | **Different types of source–sink migration dynamics. a** | A closed-sink model, in which a population that is isolated in a sink habitat, with no access to or from other habitats, must adapt to avoid extinction[11]. **b** | A black-hole sink model, in which there is a recurrent, one-way influx of migrants from the source habitat to the sink habitat[14–18]. **c** | A reciprocal sink model, where there is a constant exchange of migrants between the source and sink habitats. Green box, source habitat; blue box, sink habitat; green dots, colonizing organisms; red arrows, inter-habitat migration.

in shape, size and behaviour compared with the population in the ocean (source habitat)[8,24,25]. Many of these changes are caused by recessive mutations, which can be held in the source population in low frequency until they become advantageous in the sink populations (M. A. Bell, personal communication).

Although the opportunities to study the evolutionary dynamics of source–sink models are limited for macroorganisms, we believe that bacterial species are ideally suited for the study of source–sink dynamics as they have short generation times and are easily disseminated from one habitat to another. Therefore, the key point in understanding whether the population dynamics of a bacterial (or any other) species can be understood by using source–sink models is to use species that can be found naturally in alternative habitats, because source and sink habitats are definably distinct environments[26]. Alternative habitats are especially clear for bacterial pathogens that invade different hosts and/or different compartments in the same host[3]. We will focus here on bacterial pathogens of humans as relatively large amounts of epidemiological data are available. Although many evolutionary models have been developed that analyse virulence from the perspective of pathogen fitness in the infected host, inter-host transmissibility or pathogen-induced host mortality, the source–sink model of virulence evolution

allows us to analyse multiple habitats of pathogens for which the infected host does not necessarily provide the main habitat.

The source–sink model for pathogens
Each pathogen has a reservoir habitat and a virulence habitat. The reservoir is defined as an environmental site, host organism or population or specific body compartment where the species can sustain itself continuously (by long-lasting colonization or inter-host circulation) and from where it can be transmitted to other habitats[27]. Because the reservoir is a continuous habitat for the pathogen, it can be defined as its source habitat[28]. The virulence habitat is a disease-susceptible host or specific compartment within the same host, in which growth of the pathogen causes clinical infection, that is, induces host damage either directly, by producing toxic compounds, or indirectly, by provoking self-damaging host responses[29,30]. Consequently, to understand whether the virulence habitat can be defined as a source or a sink habitat for a particular pathogen, it is crucial to determine to what extent the pathogen's reservoir habitat overlaps with, or depends on, its virulence habitat.

Regarding the relationship between virulence and reservoir habitats, human pathogens can be divided into three general categories. First, pathogens with reservoirs of non-human origin, where infection is

acquired by human contact with vertebrate animals or their products (zoonotic pathogens such as *Campylobacter jejuni* and *Salmonella enterica* serovar Enteritidis), arthropods, plants or environmental sites (for example, *Legionella pneumophila*) and where there is no transmission between humans or transmission is short lived. For these pathogens, their virulence habitat in humans can be defined as a non-self-sustainable sink habitat.

Second, 'professional' pathogens, where continuous circulation among humans depends on the ability of the microorganism to cause disease. Such pathogens can infect humans asymptomatically, but benign colonization usually follows disease (for example, by *Shigella* spp. or *Salmonella enterica* serovar Typhi) or does not provide the means for inter-host transmission (*Mycobacterium tuberculosis*)[31]. Also, inflammatory host responses during the course of disease development have been shown to be advantageous for professional pathogens such as *Shigella* spp.[32] Here, the virulence habitat is a source habitat that serves directly as the pathogen's reservoir or is required to establish reservoirs of the pathogen in a benign habitat. It should be noted, however, that in professional pathogens the source habitat could be a virulence niche in which a relatively mild infection is caused, and more severe or rare types of infection could represent sink habitats.

Lastly, opportunistic pathogens, where bacteria can circulate continuously in humans asymptomatically and cause clinical infections mainly when the host defences are at least partly compromised. For example, by underlying somatic illness, such as *Pseudomonas aeruginosa* in cystic fibrosis patients, *Serratia marcescens* in mechanically ventilated patients, and *E. coli* and *Streptococcus pneumoniae* in newborns. Also, an increase in opportunistic infections could be attributed to factors such as exposure to cold temperatures, viral infection or reduced exposure to daylight[33]. Clinical infections are generally not directly communicable (although the hospital environment presents a risk) and the continuous circulation of these microorganisms in human populations seems not to depend on their ability to cause disease[34,35]. Therefore, for these pathogens the virulence habitat — the tissues in which they cause disease — could be viewed as a transient, marginal habitat that, at least in the long term, is non-self-sustainable, in other words, a sink habitat.

The virulence habitat is an unstable, or sink, habitat for most bacterial pathogens in humans. However, the source versus sink distinction is based primarily on epidemiological data and is not usually clear-cut. For example, many species of putative opportunistic pathogens that circulate as non-pathogens in a large proportion of the human population have clonal groups that cause clinical infections relatively often and in seemingly healthy individuals (for example, uropathogenic *E. coli*). In addition, host heterogeneity in infection susceptibility is a significant factor in the occurrence of infectious disease, but is not readily recognized for many pathogens. Therefore, separation of source and sink habitats can be complicated, and molecular evolution analysis must be used to determine when the population dynamics of bacterial pathogens conform to the source–sink model. We propose that the way to determine which alternative habitat of a bacterial species is a source (stable) habitat and which is a sink (transient) habitat, is to identify a bacterial trait that is under different types of selection in alternative habitats, and compare its evolutionary stability through sequence analysis of the genes that respond to the differential selective pressures on the trait.

Expected source–sink dynamics

To determine the different types of micro-evolutionary events in bacterial pathogens that could be produced by source–sink dynamics, one can build evolutionary scenarios around a bacterial clone that is adapted to circulate in a self-sustainable manner as an asymptomatic colonizer of humans (that is, the commensal habitat is its reservoir or source), but can occasionally spread into a protected host compartment to cause disease owing to a co-incidental pre-adaptation of its commensal traits to function as virulence factors. We will also assume that this clone cannot be transmitted directly from one virulence habitat to another in a self-sustainable manner, in other words, the virulence habitat is its transient sink habitat.

It is possible that sink-adaptive changes in the clone would not occur during the course of disease (FIG. 2a), either because the clone only replicates for a few generations before host recovery or death, or because of a small population size of the invading bacteria. In this case, no sink-adaptive variants of any trait will emerge and so there will be no source–sink selective footprint.

It is also possible, however, that sink-adaptive mutations might be emerging, at low frequency, in the source habitat

and be selected and increase (sometimes dramatically) in the sink habitat, even when the sink habitat population is small and/or short-lived. If the sink habitats allow for a relatively large or transiently self-sustainable invading population, adaptive changes could emerge during the course of infection. Under natural conditions in both cases, a mixed population of bacterial clones would be established in the sink habitats, some carrying original source-adapted genes and others the sink-adaptive variants. We will

assume that the sink-adaptive variants will sweep through the bacterial population in the virulence habitat owing to, for example, improved ability to escape host defences or use available nutrients. On eventual host recovery or death, the fate of the sink-adaptive mutation depends on the type of sink habitat and the overall functional effect of the mutation.

If there were no migration from the virulence habitat to the reservoir habitat, or migration is insignificant (a closed or
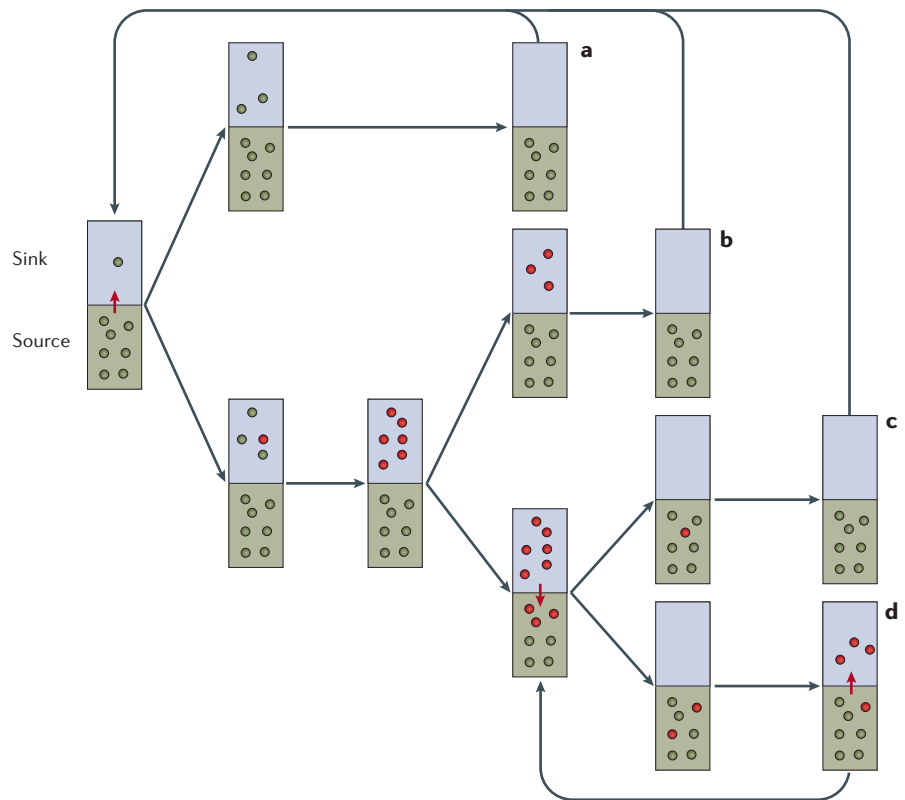

Figure 2 | **Different scenarios of sink-adaptive evolution of bacterial pathogens. a** | A scenario where there is no adaptive evolution during the course of infection; all new infections are caused by repeated invasion of the reservoir-adapted organisms and therefore there is no genetic signature. **b** | A scenario in which adaptive evolution occurs during the course of infection, but pathoadaptive mutants do not migrate back to the reservoir, or are eliminated very fast on migration, before reaching detectable numbers; virtually all new infections are caused by repeated invasion of the reservoir-adapted organisms. The genetic signature associated with this scenario is that the evolutionarily short-lived pathoadaptive mutations are found exclusively in the infecting organisms. **c** | A scenario in which pathoadaptive mutants migrate back to the reservoir and circulate there in detectable numbers, but are eventually eliminated from the reservoir and replenished with new migrants; new infections can be caused by re-invasion of the pathoadaptive organisms from the reservoir, but their frequency and migration back to the reservoir is not significant enough to continuously sustain a particular pathoadaptive population there. The genetic signature that is associated with this scenario is that the evolutionarily short-lived pathoadaptive mutations are found commonly but not exclusively in the infecting organisms. **d** | Shows the same scenario as (**c**) but the pathoadaptive organisms re-invade the virulence niche and migrate back to the reservoir frequently enough to continuously sustain the pathoadaptive population in the reservoir habitat. The genetic signature associated with this scenario is that the pathoadaptive mutations are recently emerged but show signs of evolutionary stability and are found in both infecting and reservoir organisms, with a predominance in infecting organisms. Green box, reservoir source habitat; blue box, virulence sink habitat; green dots, reservoir-adapted bacteria; red dots, patho-(sink-)adapted mutants; red arrows, inter-habitat migration.

'black-hole' sink, for example, cerebrospinal fluid and the bloodstream in meningititis and sepsis, respectively), the mutation would disappear when the virulence habitat collapses (FIG. 2b). In such situations, the sink-adaptive mutations would be found in isolates from the virulence habitat, and seldom or never in isolates from the reservoir habitat. These mutations would all have arisen independently in each individual affected by the disease and, unlike the source-adaptive non-mutated alleles, they will be of recent origin owing to the relatively transient existence of the virulence habitat from an evolutionary perspective.

When there is significant transmission from the virulence habitat back to the reservoir habitat (a reciprocal sink), the sink-adaptive mutations could spread in the source habitat. The ultimate fate of the mutation would depend on its effect on the fitness of the pathogen back in the reservoir habitat and its ability to re-invade the virulence habitat.

A gene mutation that changes the function or the expression level of an encoded protein usually results in a trade-off in the ability of the protein to perform the function for which it was originally adapted[36–38]. Therefore, although one cannot exclude the possibility that sink-adaptive mutations might be neutral or even beneficial in the reservoir habitat, it is most likely that they will be at least mildly deleterious and be selected against, unless revertant or compensatory mutations occur. Alternatively, a sink-adaptive mutation could lead to a greater level and/or more prolonged colonization of the virulence habitat and a higher rate of disease occurrence in the human population (although some of the sink-adaptive mutations could be adaptive at later stages of the infection but maladaptive for early invasion). In these cases, the higher number of organisms in the sink environment will lead to increased backflow of the sink-adaptive variants into the source habitat.

If the advantage of more frequent and/or increased spread from the virulence habitat does not compensate for selection against the mutation in the reservoir habitat, clones carrying the sink-adaptive mutation will be unstable and will eventually disappear from circulation (FIG. 2c). These sink-adaptive and source-deleterious alleles will be continuously emerging and continuously becoming extinct. There will not be enough time to accumulate selectively neutral variants associated with these sink-adaptive mutations before they become extinct and, therefore,

sink-adaptive and source-deleterious alleles will have a short-lived footprint compared with the original source-adapted alleles, which circulate continuously.

The deleterious effects of sink-adaptive mutations in the source habitat could be reduced or avoided if the genetic change is easily reversible. Such changes can occur in so-called simple contingency loci that contain repeats of single or few nucleotide sequences[39]. The hypermutability and reversibility of genetic changes at these loci result from *rec*-independent expansion or contraction of the number of repeat units, which leads to a switch in the translation reading frame or changes in the level of promoter activity.

However, if the deleterious effect of sink-adaptive mutations in the reservoir habitat is significant and non-reversible, it could still be balanced by increased migration of the mutation-carrying clone from the virulence habitat (FIG. 2d). In this case, the virulence habitat would stop being the 'sink' habitat for the mutant clone, because its circulation in the reservoir habitat will become dependent on its success in the virulence habitat — the clone will become more like a professional pathogen (for example, it is possible that some diarrhoeagenic *E. coli* belong to this category). Such clones might become stable to the point that neutral changes would start to accumulate in the sink-adaptive gene variants, that is, they will become similar to the source-adaptive variants.

In both the unstable and stable scenarios, sink-adaptive variants will probably predominate in the virulence habitat, but be sub-dominant in the reservoir habitat. However, in the unstable scenario, the adaptive mutations will be of a recent origin, whereas in the stable scenario the genes with the adaptive mutations will show signs of 'ageing'. To distinguish whether a particular pathogen undergoes stable or unstable dynamics, analytical methods of molecular evolution can be used to determine which particular habitats are primary, long-term habitats (sources) or marginal, transient habitats (sinks), and what genes are targeted by virulence-specific adaptive evolution.

Evidence for source–sink dynamics

On spreading from the reservoir habitat (source) into the virulence habitat (sink), adaptation to the sink is likely to occur not by horizontal gene transfer (clinical infections are usually monoclonal in nature) but by modification of existing genes, that is,

pathoadaptive loss-of-function or change-of-function mutations.

The loss-of-function mutations (gene knockouts or deletions) are expected to be selected when expression of the source-adapted gene is detrimental in the virulence niche, not simply by the 'use it or lose it' rule. For example, lysine decarboxylase interferes with the function of endotoxin in enterotoxigenic *E. coli* and *Shigella* spp. and the corresponding gene regions are deleted in these clones[40,41]. Also, a knockout of *hemB* (*hemB* encodes a protein involved in haemin biosynthesis) leads to the emergence of small-colony variants of *Staphylococcus aureus* that are better adapted for long-term persistence within host tissues than *S. aureus* with regular growth rates[42,43]. Sometimes, gene losses contribute to the emergence of a professional pathogen (for example, *Shigella* spp.). The tradeoff is then less important because the clone becomes less dependent on the original source habitat. For non-professional pathogens, however, where the original reservoir remains the primary habitat, the trade-off imposed by the pathoadaptive gene loss indicates that the mutant clones are likely to become unstable.

A classical example of pathoadaptive loss-of-function mutations is found in the *mucA* gene of *P. aeruginosa*. The primary habitats of *P. aeruginosa* are soil and freshwater, but it is also found as a transient asymptomatic colonizer of healthy humans[44,45]. *P. aeruginosa* is also an opportunistic pathogen that causes infections in patients with extensive burns, cystic fibrosis (CF) or indwelling devices such as urinary bladder catheters. Clinical infection is not associated with specific clonal lineages; the pathogens are generally acquired independently from the environment and are non-communicable (although transmission in a nosocomial setting is possible)[46]. Therefore, the virulence habitat for *P. aeruginosa* is a distinctly marginal sink habitat that does not contribute significantly to the overall natural circulation of the species.

MucA is a negative regulator of capsule production in *P. aeruginosa*, and deletion of *mucA* leads to overproduction of alginate polysaccharide on the bacterial surface[47]. The resultant 'mucoid' phenotype enables bacteria to resist phagocytosis and is highly adaptive in *P. aeruginosa* that cause chronic endobronchial and urinary tract infection in individuals with CF and a catheterized urinary bladder, respectively[48,49]. However, mucoid strains are not found in *P. aeruginosa* isolates that colonize humans asymptomatically or in environmental isolates, where

increased resistance to phagocytosis is not as important and cannot offset the energy cost and other limitations associated with capsule overproduction. Therefore, *mucA* knockouts occur and are selected for during the course of infections that are initiated by non-mucoid strains, but are eliminated quickly in the source population after shedding.

The distinctive source–sink dynamics of *P. aeruginosa* and the sink-adaptive, source-deleterious nature of the *mucA* mutations in *P. aeruginosa* allow us to examine the specific genetic footprint in this gene produced by the source–sink. The loss-of-function mutations in *mucA* are primarily frame-shift indels (insertions/deletions) or nonsense mutations and, sometimes, non-synonymous changes[48,49], and they can be located anywhere in the gene. A phylogenetic relationship between different MucA variants based on the *mucA* sequences from mucoid and non-mucoid isolates of *P. aeruginosa* from different CF patients and other sources[48,50] is depicted in FIG. 3a.

An important difference between the large nodes formed by the functional and non-functional MucA is the presence and absence of silent variation between genes encoding corresponding MucA variants, respectively. In other words, most functional MucA variants have identical amino acid sequences and form a single node on the tree, but are encoded by multiple alleles that differ from each other only by synonymous changes. This reflects their long-term stability within the species, which, over time, allows selectively neutral changes to accumulate. Same-node non-functional MucA variants are always encoded by one allele type, indicating their recent origin. This would be expected, considering that adaptation of the 'mucoid' variants in the sink habitat (lungs) is short-lived relative to the functional MucA that is found in the source habitat (the environment). Consequently, this type of phylogenetic analysis can be used to discover genes involved in the source–sink evolutionary dynamic.

In summary, the pattern of *mucA* variability created by the source–sink dynamics of *P. aeruginosa* is characterized by the existence of a pool of stable haplotypes (with accumulated silent variation) that are adapted to the source habitat and a series of recently derived haplotypes (without silent variation) that are adapted to the sink habitat, with the adaptive mutations being of various types that commonly occur in hot-spot positions. Obviously, the number of source- versus sink-adapted haplotypes on the tree will depend on the relative prevalence of alleles derived from source or sink habitats in the sample.

Using this methodology, we examined the pattern of allelic variability in another human pathogen, uropathogenic *E. coli*. *E. coli* primarily circulates in humans as a commensal organism of the large intestine (with faecal-oral transmission) but can cause intestinal and extraintestinal infections in seemingly healthy individuals[51]. Among the extraintestinal infections, urinary tract infections (UTIs) are most common, with about half of the women in the United States having at least one *E. coli*-related episode of cystitis in their lifetime[52]. These UTIs are mostly self-resolved within a few days, but up to 20% of affected women experience recurrent episodes of cystitis. The UTIs are caused by *E. coli* inhabiting the large intestine of the patient and are considered to be a non-communicable infection, although sexual activity is a predisposing factor for UTIs and sexual transmission has been recorded[53]. Because of the short duration of most UTIs and the lack of direct transmission, the urinary tract seems to be a marginal sink habitat for *E. coli*, with the intestinal tract being the source habitat. However, the source versus sink separation is not as clear here as it was for *P. aeruginosa*. Large numbers of bacteria are voided in the urine during the course of a UTI. Therefore, sink-adapted bacteria can return easily, and in large numbers, to faecal-oral circulation in the intestinal reservoir and are likely to be able to invade the urinary tract again as a pathogen. Also, certain clonal groups that include specific serotypes of *E. coli* seem to be genetically predisposed to cause UTIs[54]. Finally, *E. coli* can colonize the urinary tract asymptomatically for prolonged periods of time (months or even years) and at a relatively high level ($>10^3$ ml^{-1} in urine), indicating that urinary-tract-adapted strains exist.

Among the traits associated with the ability of *E. coli* to cause UTIs, the expression of type 1 fimbriae has been shown
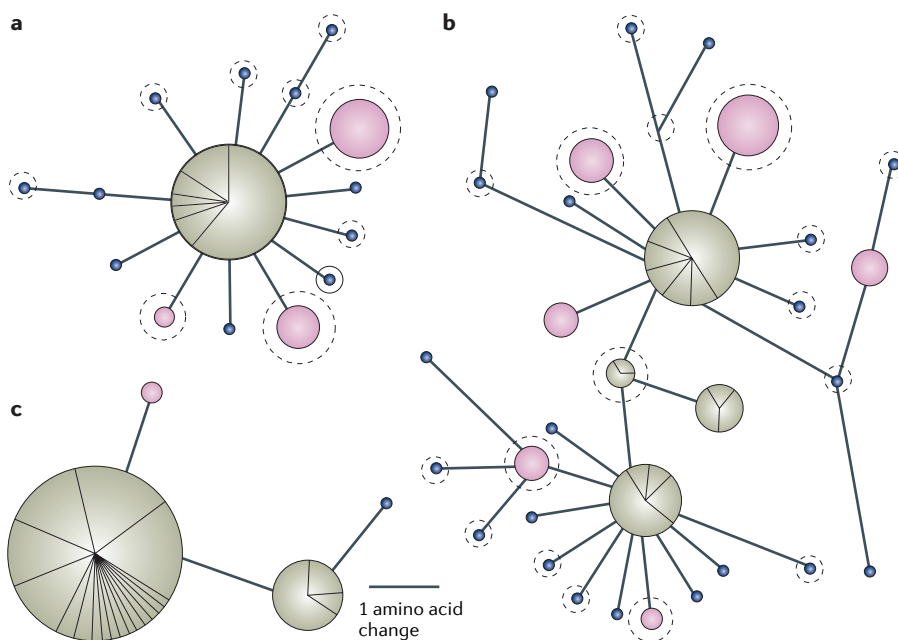

Figure 3 | **Phylogenetic analysis of source–sink dynamics.** The phylogenic relationships of structural variants of the MucA regulator (**a**); FimH adhesin (**b**) and adenylate cyclase (**c**) are shown. The protein tree is derived from the corresponding gene tree (built as a maximum likelihood unrooted phylogram) by collapsing branches with silent changes (for details see REF. 59). The length of the sequences used are 623 bp, 900 bp and 546 bp, respectively. The size of the node corresponds to the number of strains carrying the specific protein variant. Green circles represent nodes composed of a protein variant carried by multiple strains and encoded by multiple alleles, that is, differing in silent changes only (the proportional occurrence of individual alleles of the corresponding variant is represented as pie pieces). Pink circles represent nodes composed of protein variants found in multiple strains but encoded by one allele only. Blue dots represent nodes composed of a protein variant found in a single strain. Most branches connecting the pink or blue nodes with the green nodes did not contain silent changes on the original gene tree. No silent changes were found on branches between pink and blue nodes. Dashed circles show the nodes resulting from functional mutations (that is, a knockout of MucA or an increase of monomannose-binding in FimH).

to have a direct role in urovirulence by mediating bacterial binding to uroepithelial cells[55,56]. Type 1 fimbriae are thought to be crucial in faecal–oral transmission, by allowing *E. coli* to bind to, and therefore colonize, the oropharyngeal epithelial surface[57]. This transient colonization in the throat might increase the probability of successful passage of the clone through the stomach acid to become resident in the large intestine. As type 1 fimbriae function in physiologically distinct habitats, the urinary tract and the oropharynx, one might expect selection to favour differences in the adhesive properties of the fimbriae under these different conditions. In fact, it has been shown that type 1 fimbriae from uropathogenic *E. coli* tend to bind mono-mannose-containing glycoproteins with higher affinity than fimbriae from intestinal isolates[58]. This increased monomannose-binding capability results in an increased tropism of bacteria for uroepithelial cells and increased bladder colonization (as shown in the murine model of UTI), indicating that this property is adaptive for *E. coli* urovirulence[58,59].

Examining the gene that encodes the adhesive subunit of type 1 fimbriae, FimH, a 30 kDa lectin-like protein located on the fimbrial tip, revealed that the increased monomannose binding is caused by the presence of single amino-acid replacements in different regions of the FimH protein. Therefore, these point mutations in *fimH* are adaptive to the virulence habitat of uropathogenic *E. coli*, the urinary tract. At the same time, the increased affinity for monomannose residues is accompanied by an increased sensitivity of the mutant fimbriae to inhibition by soluble mannosylated glycoproteins, which are present in abundance in human saliva[59–61]. This inhibition presumably reduces the ability of type 1 fimbriae to mediate adhesion to the oropharyngeal epithelia and therefore decreases the chance of successful transmission of the clone through the stomach acid to the intestinal reservoir. Therefore, the monomannose-affinity-enhancing mutations in FimH are adaptive for urovirulence, but exhibit a functional trade-off with the original function of FimH in the reservoir habitat.

The phylogenic relationship of *fimH* adhesin alleles (from a set of isolates split between intestinal and UTI origin) was analysed in the same way as the *mucA* regulator alleles[62] (FIG. 3b). The phylogram structures are strikingly similar for the *P. aeruginosa* and *E. coli* proteins, with

nodes containing silent variation at the centre of the phylogram and a significant number of recently evolved variants forming singleton nodes or nodes without silent variation located on the external branches. Similar to the non-functional pathoadaptive MucA variants, the *E. coli* UTI-adaptive FimH variants (those carrying mutations that enhance monomannose binding) are found on the recently derived nodes that are more external. By contrast, similar to the fully functional MucA from environmental *P. aeruginosa*, FimH variants from which the external nodes are derived exhibit a low monomannose-binding capability (FimH that is adaptive for circulation in the intestinal habitat of *E. coli*) and form nodes that contain silent variation. This supports the hypothesis that the urinary tract is a sink habitat for *E. coli* bacteria that cannot support long-term circulation of the uro-adapted clone.

Unlike the *P. aeruginosa* MucA mutations, sink-adaptive FimH variants are found in intestinal isolates, although in a much smaller proportion than in UTI isolates[58,62]. Obviously, once the uropathogenic clones return to faecal–oral circulation, the uro-adaptive *fimH* alleles are not as strongly selected against in the reservoir habitat as the CF-adaptive or urocatheter-adaptive *mucA* alleles are. This is not surprising, because mutations in *fimH* do not abolish protein function, but result in an adaptive 'tune-up'. Therefore, clones carrying the mutant *fimH* return to faecal–oral circulation for at least some time and are likely to invade the urinary tract again, possibly at a higher rate than clones carrying the non-adapted *fimH*. However, the functional trade-off of *fimH* mutations is obviously significant enough to make the carrying clone unstable in the long term, especially considering that uropathogenic clones are likely to accumulate multiple sink-adaptive, but source-detrimental, mutations.

Importantly, the phylogenetic pattern demonstrated by *mucA* and *fimH* is not shown by haplotypes of *adk*, the gene encoding the housekeeping protein adenylate kinase (FIG. 3c), even though the sequences are from the same *E. coli* strains as the *fimH* alleles. Almost all *adk* alleles are within stable, multi-haplotype nodes, containing both strains isolated from the urinary tract and the intestinal tract. Therefore, there is no evidence of strong selection for uro-adaptive mutations in the *adk* gene. Apart from the adhesins, another trait that could be under selection in uropathogenic *E. coli* is the ability of fast

growth in urine[63]. Choosing the correct gene for the analysis is therefore crucially important when one wishes to determine whether or not the population dynamics of a bacterial pathogen, or any other species, conform to the source–sink model. Alternatively, when complete genome sequences for many isolates are available, one can use the molecular-phylogeny-based structure–function analysis described here to discover which genes are under the kind of diversifying selection expected in source and sink habitats.

Conclusion

The source–sink model of population dynamics presented here provides a conceptual framework for understanding both short-term and long-term patterns in evolution. The assumption that virulence habitats are marginal sink habitats for at least some pathogens and that virulence-enhancing genetic adaptation is mostly transient in nature is supported by the molecular evolutionary footprint of genes under selection in sinks. In this paper we have focused on opportunistic pathogens adapting to pathogenic habitats because we have both a clear idea of what the source and sink habitats are and molecular data on the adaptation of particular genes are available, however, this analytical technique is not limited to analysis of the evolution of virulence. We could use the source–sink approach to determine the genes under source–sink adaptation for professional pathogens adapting to commensal habitats or we could use it to determine whether a bacterial species spans different habitats or is adapted to a single habitat. The genetic footprint proposed for the source–sink model could help in understanding the adaptive dynamics of other prokaryotic or eukaryotic species known or suspected to exist in alternative habitats which, eventually, could provide insights into the molecular genetic mechanisms of niche differentiation and the emergence of new species.

Evgeni V. Sokurenko is at the Department of Microbiology, University of Washington, Seattle, Washington 98105,USA.

Richard Gomulkiewicz is at the Department of Mathematics and School of Biological Sciences, Washington State University, Pullman, Washington 99164, USA.

Daniel E. Dykhuizen is at the Department of Biology, University of Louisville, Louisville, Kentucky 40292, USA.

Correspondence to E.V.S
e-mail: evs@u.washington.edu

1. Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nature Rev. Microbiol.* **2**, 414–424 (2004).
2. Sokurenko, E. V., Hasty, D. L. & Dykhuizen, D. E. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends Microbiol.* **7**, 191–195 (1999).
3. Levin, B. R. & Bull, J. J. Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends Microbiol.* **2**, 76–81 (1994).
4. Meyers, L. A., Levin, B. R., Richardson, A. R. & Stojiljkovic, I. Epidemiology, hypermutation, within-host evolution and the virulence of *Neisseria meningitidis*. *Proc. Biol. Sci.* **270**, 1667–1677 (2003).
5. Antia, R., Regoes, R. R., Koella, J. C. & Burgstrom, C. T. The role of evolution in the emergence of infectious disease. *Nature* **426**, 658–661 (2003).
6. Pulliam, H. R. Sources, sinks, and population regulation. *Am. Nat.* **132**, 652–661 (1988).
7. Levin, D. & Clay, K. Dynamics of synthetic *Phlox drummondii* populations at the species margin. *Am. J. Bot.* **71**, 1040–1050 (1984).
8. Bell, M. A. Lateral plate evolution in the threespine stickleback: getting nowhere fast. *Genetica* **112–113**, 445–461 (2001).
9. Kirkpatrick, M. & Barton, N. H. Evolution of a species' range. *Am. Nat.* **150**, 1–23 (1997).
10. Holt, R. D., Barfield, M. & Gomulkiewicz, R. in *Species Invasions: Insights into Ecology, Evolution, and Biogeography* (eds Sax, D. F., Stachowicz, J. J. & Gaines, S. D.) 259–290 (Sinauer Associates Inc., Sunderland USA, 2004).
11. Lynch, M. & Lande, R. in *Biotic Interactions and Global Climate Change* (eds Kareiva, P. M., Kingsolver, J. G. & Huey, R. B.) 234–250 (Sinauer Associates Inc., Sunderland USA, 1993).
12. Bürger, R. & Lynch, M. Evolution and extinction in a changing environment: a quantitative-genetic analysis. *Evolution* **49**, 151–163 (1995).
13. Gomulkiewicz, R. & Holt, R. D. When does evolution by natural selection prevent extinction? *Evolution* **49**, 201–207 (1995).
14. Gomulkiewicz, R., Holt, R. D. & Barfield, M. The effects of density dependence and immigration on local adaptation and niche evolution in a black-hole sink environment. *Theor. Popul. Biol.* **55**, 283–296 (1999).
15. Holt, R. D., Barfield, M. & Gomulkiewicz, R. Temporal variation can facilitate niche evolution in harsh sink environments. *Am. Nat.* **164**, 187–200 (2004).
16. Holt, R. D. & Gaines, M. S. Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. *Evol. Ecol.* **6**, 433–447 (1992).
17. Holt, R. D. & Gomulkiewicz, R. How does immigration influence local adaptation? A re-examination of a familiar paradigm. *Am. Nat.* **149**, 563–572 (1997).
18. Holt, R. D., Gomulkiewicz, R. & Barfield, M. The phenomology of niche evolution via quantitive traits in a 'black-hole' sink. *Proc. Biol. Sci.* **270**, 215–224 (2003).
19. Kawecki, T. J. Demography of source–sink populations and the evolution of ecological niches. *Evol. Ecol.* **9**, 38–44 (1995).
20. Kawecki, T. J. Adaptation to marginal habitats: contrasting influence of the dispersal rate on the fate of alleles with small and large effects. *Proc. Biol. Sci.* **267**, 1315–1320 (2000).
21. Holt, R. D. Demographic constraints in evolution: towards unifying the evolutionary theories of senescence and niche conservatism. *Evol. Ecol.* **10**, 1–11 (1996).
22. Fry, J. D. The evolution of host specialization: are 'tradeoffs' overrated? *Am. Nat.* **148**, S84–S107 (1996).
23. Barton, N. H. in *Integrating Genetics and Ecology in a Spatial Context* (eds Antonovics, J. & Silvertown, J.) 365–392 (Blackwell, London, 2001).
24. Kimmel, C. B. *et al.* Evolution and development of facial bone morphology in threespine sticklebacks. *Proc. Natl Acad. Sci. USA* **102**, 5791–5796 (2005).
25. Cresko, W. A. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc. Natl Acad. Sci. USA* **101**, 6050–6055 (2004).
26. Watkinson, A. R. & Sutherland, W. J. Sources, sinks and pseudo-sinks. *J. Animal Ecol.* **64**, 126–130 (1995).
27. Haydon, D. T., Cleaveland, S., Taylor, L. H. & Laurenson, M. K. Identifying reservoirs of infection: a conceptual and practical challenge. *Emerg. Infect. Dis.* **8**, 1468–1473 (2002).
28. Holt, R. D. & Hochber, M. E. in *Adaptive Dynamics of Infectious Diseases: in Pursuit of Virulence Management* (eds Dieckmann, U., Metz, J. A. J., Sabelis, M. W. & Sigmund, K.) 104–119 (Cambridge University Press, UK, 2002).
29. Finlay, B. B. & Falkow, S. Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* **61**, 136–169 (1997).
30. Casadevall, A. & Pirofski, L. A. The damage-response framework of microbial pathogenesis. *Nature Rev. Microbiol.* **1**, 17–24 (2003).
31. Stewart, G. R., Robertson, B. D. & Young, D. B. Tuberculosis: a problem with persistence. *Nature Rev. Microbiol.* **1**, 97–105 (2003).
32. Sansonetti, P. J., Arondel, J., Huerre, M., Harada, A. & Matsushima, K. Interleukin-8 controls bacterial transepithelial translocation at the cost of epithelial destruction in experimental shigellosis. *Infect. Immun.* **67**, 1471–1480 (1999).
33. Dowell, S. F. Seasonal variation in host susceptibility and cycles of certain infectious diseases. *Emerg. Infect. Dis.* **7**, 369–374 (2001).
34. Blakebrough, I., Greenwood, B., Whittle, H., Bradley, A. & Gilles, H. The epidemiology of infections due to *Neisseria meningitidis* and *Neisseria lactamica* in a northern Nigerian community. *J. Infect. Dis.* **146**, 626–637 (1982).
35. Schneider, R. F. Bacterial pneumonia. *Semin. Respir. Infect.* **14**, 327–332 (1999).
36. Taper, M. L. & Case, T. J. Quantitative genetic models for the co-evolution of character displacement. *Ecology* **66**, 355–371 (1985).
37. Brown, J. S. & Pavlovic, N. B. Evolution in heterogeneous environments: effects of migration on habitat specialization. *Evol. Ecol.* **6**, 360–382 (1992).
38. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19**, 207–233 (1988).
39. Bayliss, C .D., Field, D. & Moxon, E. R. The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J. Clin. Invest.* **107**, 657–662 (2001).
40. Maurelli, A. T., Fernandez, R. E., Bloch, C. A., Rode, C. K. & Fasano, A. 'Black holes' and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 3943–3948 (1988).
41. Day, W. A. Jr, Fernandez, R. E. & Maurelli, A. T. Pathoadaptive mutations that enhance virulence: genetic organization of the *cadA* regions of *Shigella* spp. *Infect. Immun.* **69**, 7471–7480 (2001).
42. von Eiff, C. *et al.* A site-directed *Staphylococcus aureus hemB* mutant is a small-colony variant which persists intracellularly. *J. Bacteriol.* **179**, 4706–4712 (1997).
43. Proctor, R. A. *et al.* Persistent and relapsing infections associated with small-colony variants of *Staphylococcus aureus*. *Clin. Infect. Dis.* **20**, 95–102 (1995).
44. Ohara, T. & Itoh, K. Significance of *Pseudomonas aeruginosa* colonization of the gastrointestinal tract. *Intern. Med.* **42**, 1072–1076 (2003).
45. Romling, U., Kader, A., Sriramulu, D. D., Simm, R. & Kronvall, G. Worldwide distribution of *Pseudomonas aeruginosa* clone C strains in the aquatic environment and cystic fibrosis patients. *Environ. Microbiol.* **7**, 1029–1038 (2005).
46. Ruiz, L., Dominguez, M. A., Ruiz, N. & Vinas, M. Relationship between clinical and environmental isolates of *Pseudomonas aeruginosa* in a hospital setting. *Arch. Med. Res.* **35**, 251–257 (2004).
47. Martin, D. W. *et al.* Mechanism of conversion to mucoidy in *Pseudomonas aeruginosa* infecting cystic fibrosis patients. *Proc. Natl Acad. Sci. USA* **90**, 8377–8381 (1993).
48. Boucher, J. C., Yu, H., Mudd, M. H. & Deretic, V. Mucoid *Pseudomonas aeruginosa* in cystic fibrosis: characterization of *muc* mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection. *Infect. Immun.* **65**, 3838–3846 (1997).
49. Yu, H., Hanes, M., Chrisp, C. E., Boucher, J. C. & Deretic, V. Microbial pathogenesis in cystic fibrosis: pulmonary clearance of mucoid *Pseudomonas aeruginosa* and inflammation in a mouse model of repeated respiratory challenge. *Infect. Immun.* **66**, 280–288 (1998).
50. Anthony, M. *et al.* Genetic analysis of *Pseudomonas aeruginosa* isolates from the sputa of Australian adult cystic fibrosis patients. *J. Clin. Microbiol.* **40**, 2772–2778 (2002).
51. Johnson, J. R. & Russo, T. A. Extraintestinal pathogenic *Escherichia coli*: the other bad *E. coli*. *J. Lab. Clin. Med.* **139**, 155–162 (2002).
52. Hooton, T. M. & Stamm, W. E. Diagnosis and treatment of uncomplicated urinary tract infection. *Infect. Dis. Clin. North Am.* **11**, 551–581 (1997).
53. Brown, P.D. & Foxman, B. Pathogenesis of urinary tract infection: the role of sexual behavior and sexual transmission. *Curr. Infect. Dis. Rep.* **2**, 513–517 (2000).
54. Johnson, J. R., Brown, J. J., Carlino, U. B. & Russo, T. A. Colonization with and acquisition of uropathogenic *Escherichia coli* as revealed by polymerase chain reaction-based detection. *J. Infect. Dis.* **177**, 1120–1124 (1998).
55. Connell, I. *et al.* Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proc. Natl Acad. Sci. USA* **93**, 9827–9832 (1996).
56. Hung, C. S. *et al.* Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Mol. Microbiol.* **44**, 903–915 (2002).
57. Bloch, C., Stocker, B. & Orndorff, P. A key role for type 1 pili in enterobacterial communicability. *Mol. Microbiol.* **6**, 697–701 (1992).
58. Sokurenko, E. V., Courtney, H. S., Maslow, J., Siitonen, A. & Hasty, D. L. Quantitative differences in adhesiveness of type 1 fimbriated *Escherichia coli* due to structural differences in *fimH* genes. *J. Bacteriol.* **177**, 3680–3686 (1995).
59. Sokurenko, E. V. *et al.* Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc. Natl Acad. Sci. USA* **95**, 8922–8926 (1998).
60. Thomas, W. E., Nilsson, L. M., Forero, M., Sokurenko, E. V. & Vogel, V. Shear-dependent 'stick-and-roll' adhesion of type 1 fimbriated *Escherichia coli*. *Mol. Microbiol.* **53**, 1545–1557 (2004).
61. Thomas, W. E., Trintchina, E., Forero, M., Vogel, V. & Sokurenko, E. V. Bacterial adhesion to target cells enhanced by shear force. *Cell* **109**, 913–923 (2002).
62. Sokurenko, E. V. *et al.* Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in *Escherichia coli*. *Mol. Biol. Evol.* **21**, 1373–1383 (2004).
63. Gordon, D. M, & Riley, M. A. A theoretical and experimental analysis of bacterial growth in the bladder. *Mol. Microbiol.* **6**, 555–562 (1992).
64. Ziebuhr, W., Ohlsen, K., Karch, H., Korhonen, T. & Hacker, J. Evolution of bacterial pathogenicity. *Cell. Mol. Life Sci.* **56**, 719–728 (1999).
65. Nowrouzian, F. L., Wold, A. E. & Adlerberth, I. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J. Infect. Dis.* **191**, 1078–1083 (2005).

DATABASES
The following terms in this article are linked online to:
Entrez Genome Project: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj
Campylobacter jejuni | *Escherichia coli* | *Gasterosteus aculeatus* | *Legionella pneumophila* | *Mycobacterium tuberculosis* | *Pseudomonas aeruginosa* | *Salmonella enterica* | *Serratia marcescens* | *Staphylococcus aureus* | *Streptococcus pneumoniae*
UniProtKB: http://ca.expasy.org/sprot
FimH
Access to this links box is available online.