Within-host evolution of bacterial pathogens

Xavier Didelot[1], A. Sarah Walker[2], Tim E. Peto[2], Derrick W. Crook[2,3] and Daniel J. Wilson[2,4]

Abstract | Whole-genome sequencing has opened the way for investigating the dynamics and genomic evolution of bacterial pathogens during the colonization and infection of humans. The application of this technology to the longitudinal study of adaptation in an infected host — in particular, the evolution of drug resistance and host adaptation in patients who are chronically infected with opportunistic pathogens — has revealed remarkable patterns of convergent evolution, suggestive of an inherent repeatability of evolution. In this Review, we describe how these studies have advanced our understanding of the mechanisms and principles of within-host genome evolution, and we consider the consequences of findings such as a potent adaptive potential for pathogenicity. Finally, we discuss the possibility that genomics may be used in the future to predict the clinical progression of bacterial infections and to suggest the best option for treatment.

Evolutionary rates
The rates at which substitutions arise in a lineage (also known as molecular clock rates). Population genetics theory predicts a constant rate in a neutrally evolving population with a constant mutation rate, irrespective of changes in population size.

[1]*Department of Infectious Disease Epidemiology, Imperial College, London W2 1PG, UK.*
[2]*Nuffield Department of Medicine, Experimental Medicine Division, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.*
[3]*National Infection Service, Public Health England, London SE1 8UG, UK.*
[4]*Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.*
Correspondence to X.D. and D.J.W.
x.didelot@imperial.ac.uk; daniel.wilson@ndm.ox.ac.uk

doi:10.1038/nrmicro.2015.13
Published online 19 Jan 2016

Until recently, the extent of bacterial diversity associated with humans, and the ability of these bacteria to evolve in response to drug pressures, immune responses and the host environment, was presumed limited when compared with rapidly evolving viruses such as HIV[1,2]. Long-term estimates of bacterial evolutionary rates (over millions of years), calibrated using known dates of ecological and geological events, were of the order of 10^{-10} to 10^{-9} substitutions per site per year[3–5], which is low enough to suggest that recently diverged within-host populations should be completely monomorphic. However, recent studies have shown that short-term bacterial evolutionary rates (over months or years) are of the order of 10^{-7} to 10^{-5} substitutions per site per year and are thus much higher than the aforementioned long-term rates[6–9] — a discrepancy of great consequence that mirrors previous observations in viruses[10,11]. Although short-term bacterial evolutionary rates remain lower than those reported for RNA viruses (10^{-4} to 10^{-3}), bacteria have much larger genomes, which means that their per genome evolutionary rates can be comparable to that of RNA viruses or even higher[9,12,13].

Given these short-term per site evolutionary rates and the short timescale of within-host evolution of a few days to a few years, molecular techniques that sample only a small subset of the genome, such as multi-locus sequence typing (MLST)[14], are of limited use when applied to the study of within-host diversity. However, whereas higher resolution approaches had previously not been economically feasible for large population studies, the cost and time required to carry out whole-genome sequencing (WGS) of bacterial isolates has been substantially reduced over the past few years[15,16]. This has led to a flurry of WGS studies that have revealed previously unsuspected layers of within-host diversity that evolve measurably over time, to such an extent that bacteria can adapt to the specific conditions in a given host. Analysing the genome evolution of bacterial pathogens in human hosts in WGS studies requires a wide-range of specific methods, including the collection of clinical samples, the isolation and growth of pathogenic bacteria, DNA preparation, genome sequencing, sequence alignment, genome assembly, variant calling and comparative genome analysis (BOX 1).

In this Review, we discuss recent work that has advanced our understanding of genome dynamics in populations of pathogenic bacteria as they evolve within human hosts. After surveying the various evolutionary processes and forces that affect these populations, we describe how within-host diversity affects epidemiological studies of transmission between individuals. Finally, we consider the contribution of within-host evolution to antibiotic resistance and heightened or reduced virulence phenotypes, as well as to adaptation to the host environment.

Within-host evolutionary dynamics

Until the recent advances in WGS methods, little was known about genomic diversity in populations of pathogenic bacteria that infect human hosts. Previous genotyping methods such as pulsed-field gel eletrophoresis (PFGE), variable-number tandem repeats (VNTR), multi-locus enzyme electrophoresis (MLEE) and MLST

Box 1 | **Methodology for within-host studies**

Isolate collection and whole-genome sequencing
Several approaches can be used to capture the within-host diversity of bacteria. A first approach is to collect several clinical samples, either longitudinally or simultaneously, either at a single site or at several body sites. A second approach is to use a single clinical sample but sequence several separate genomes from the sample — for example, by culturing bacteria from the sample on a suitable medium, the subsequent selection of colonies and independent further sub-culturing[55]. A third approach is to sequence only a single sample, but to look for variation in the raw sequencing data before assembling these data into a genome[135]. However, it is important to note that no approach is guaranteed to fully sample the within-host pathogen population. Once each isolate has been grown for the duration of time necessary to yield sufficient DNA for whole-genome sequencing, the DNA is then extracted and purified. The most popular approach to sequencing is currently sequencing-by-synthesis[136], as implemented by the Illumina HiSeq and MiSeq sequencers. Sample preparation for these platforms entails fragmentation of the DNA and multiplexing, which enables several genomes to be sequenced at the same time by uniquely tagging every fragment from each genome; the typical output is paired-end reads corresponding to the two ends of the genomic fragments. In the past few years, read lengths for sequencing-by-synthesis have been increasing (up to 300 bp), as has the speed and affordability of sequencing protocols. This is especially true of benchtop sequencers, to the extent that real-time clinical applications of whole-genome sequencing are emerging[15,16,137,138].

Assembly and variant calling
The most common approach to assemble sequencing reads into a genome is to use a previously sequenced reference genome as a scaffold. Assuming that the reference genome and target genome are not too distantly related, each read can be mapped onto the reference genome[139], creating a so-called 'pileup'. The average number of reads mapping to each unique position of the reference genome is called the coverage or depth and represents an important measure of sequencing reliability. At each position in the reference genome, variant calling is carried out by determining whether the mapped reads of the target genome are identical or different to the reference genome. If the difference between reads is too high, or the site coverage is too low, the site may be left uncalled. If no closely related reference genome exists, lower throughput long-read technology can be used to generate one — a strategy used in the study of an outbreak of *Klebsiella pneumoniae*[140]. Alternatively, *de novo* assembly can be attempted, without the requirement of a reference genome[141], but a *de novo* assembly is typically able to reconstruct only segments of the genome — so called 'contigs'. In both reference-based and *de novo* approaches, repetitive regions of the genome are difficult to assemble[142], which is especially problematic when studying highly repetitive genes, such as the ones coding for surface proteins[143].

Comparative genome analysis
The simplest method for comparative genome analysis is to count the number of positions in which two genomes differ. When the two genomes are from the same host, this provides an initial estimate of the amount of within-host diversity[55]. When the two genomes are from different hosts, the number of differences reflects the likelihood of transmission between the two hosts[144], although it is difficult to define the threshold above which a transmission event can be definitively ruled out. Comparison of more than two genomes typically involves the reconstruction of their ancestry using phylogenetic methods. When the genomes have been sampled at different times, a popular approach is to use the Bayesian evolutionary analysis by sampling trees (BEAST) software tool to try to infer both the rate of evolution and a timescaled genealogy[9,145]. Alternatively, the evolutionary rate can be estimated using several pairs of sequentially sampled genomes from the same hosts[9]. Recombination can disrupt phylogenetic reconstructions, but new methods are being developed that can account for this in whole-genome datasets[146,147].

Multi-locus sequence typing
(MLST). A molecular epidemiology approach in which strains are typed by their nucleotide sequences at several loci, typically 400–500 bp fragments of seven housekeeping genes.

Genome assembly
A bioinformatics process in which overlapping sequencing reads are combined into longer, contiguous sequences known as 'contigs', ideally a single contig per chromosome but usually several.

Variant calling
A bioinformatics process that determines the nucleotide at a given genomic site based on sequencing reads.

Virulence
The quantifiable frequency or severity of disease.

Pulsed-field gel electrophoresis
(PFGE). A molecular epidemiology marker that enables strains to be typed by the lengths of the DNA molecules obtained after cutting the genome using a restriction enzyme.

Variable-number tandem repeats
(VNTR). A molecular epidemiology marker that enables strains to be typed by counting the number of copies of a specific repeat sequence, which may consist of one or more nucleotides and is known to occur at a given location in the genome.

Multi-locus enzyme electrophoresis
(MLEE). A molecular epidemiology approach in which strains are typed by the electrophoretic properties of several proteins.

Point mutations
Mutations that change a single nucleotide.

Mismatch repair systems
A mechanism found in all bacteria that repairs the mistakes introduced into the genome during DNA replication to enable clonal reproduction.

had comparatively low resolution[17–19], which could only identify the presence of multiple colonizing or infecting lineages rather than the evolution of each lineage. For example, PFGE[20] and VNTR[21,22] have shown that mixed colonization by several bacterial lineages is relatively common among *Staphylococcus aureus* carriers; such mixed colonization is most likely to be the result of two or more separate transmission events but, alternatively, the different lineages could have been transmitted simultaneously from a donor who was also colonized by several lineages. This would require a relatively loose transmission bottleneck, so that substantial diversity from the donor can be transmitted to the recipient, but little is known about the transmission bottleneck of bacterial pathogens[23]. A noteworthy exception in which the within-host evolution of a given lineage can be detected by low-resolution approaches (that is, sequencing only a few genes rather than whole genomes) is the stomach pathogen *Helicobacter pylori*, owing to its unique combination of a long infection time and very high evolutionary rate[24].

Applying WGS to several isolates sampled from the same host has enabled the development of a much more complete picture of the various processes involved in within-host bacterial evolution (for example, for the hospital-associated pathogen *S. aureus*, see FIG. 1). Point mutations provide the raw material for this evolution, and studies comparing pairs of genomes sampled simultaneously or longitudinally from the same hosts have confirmed the high rate of within-host point mutation in *H. pylori*, with ~30 mutations per year per genome[25,26]. More modest rates of within-host point mutation have been reported for other bacterial pathogens, such as ~10 mutations per year per genome for *Klebsiella pneumoniae*[27], ~8 mutations per year per genome for *S. aureus*[28], ~2 mutations per year per genome for *Clostridium difficile*[6,29,30], ~1 mutation per year per genome for *Escherichia coli*[31] and ~0.5 mutations per year per genome for both *Mycobacterium tuberculosis*[32,33] and *Mycobacterium abscessus*[34]. These differences between species are partly due to differences in genome

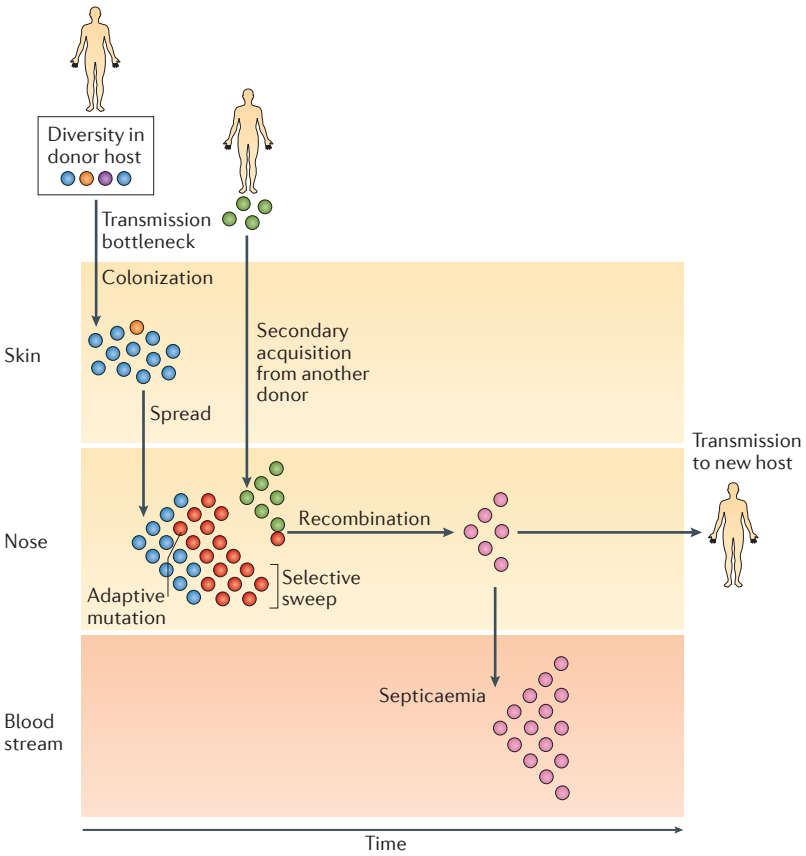


Figure 1 | **Within-host evolution of *Staphylococcus aureus*.** Infection with *Staphylococcus aureus* begins with the transmission of a *S. aureus* population (blue circles) from a donor host to a new host. Owing to the transmission bottleneck, only a subset of the within-host diversity in the donor host is transmitted to the newly infected host. Colonization begins at a given site — typically somewhere on the skin — and the population of the pathogen initially increases in size, with a new mutation arising occasionally (orange circle). The infection may also spread to other sites in the host, such as from the skin to the nose. When adaptive mutations (red circles) arise — that is, mutations that confer a selective advantage — they can quickly become fixed in the pathogen population through a selective sweep. When a second *S. aureus* population (green circles), transmitted by another donor, colonizes the same host, the two infecting populations may recombine to form a new genotype (purple circles). Transmission to other hosts may occur at any stage of colonization, as can dissemination to the bloodstream, where the pathogen can cause life-threatening septicaemia.

length but also to differences in per site mutation rates, which are modulated by varying levels of efficacy of the DNA mismatch repair systems (for example, *H. pylori* lacks several genes involved in this process that are found in most other bacteria)[35]. A much higher rate of point mutation, known as hypermutation, can also occur when the mismatch repair system becomes disrupted. Hypermutation has long been known to be important for adaptation to new ecological niches in the external environment[36,37] but has recently been shown to be important within the host as well (for example, in generating excess diversity in a *Burkholderia dolosa* infection in the lungs of a patient with cystic fibrosis)[38]. A related mechanism is that of phase variation, in which loci that are susceptible to high rates of mutation — for example, owing to the slippage of repetitive motifs — can rapidly modulate the expression of phenotypes over very short timescales, often displaying variability even within a single colony[39,40]. For example, during long-term carriage of *Neisseria meningitidis*, phase variation reduces the expression of certain genes that encode surface proteins to avoid detection by the host immune system[41].

Another factor that can lead to faster diversification of the genome than point mutation alone is the acquisition of genetic sequences from unrelated organisms through horizontal gene transfer[42]. One possible outcome is homologous recombination, in which a fragment of the chromosomal genome is replaced with a homologous sequence from another cell. This process has an important role in the genome evolution of many bacterial pathogens, such as *E. coli*[43], *Streptococcus pneumoniae*[44], *Salmonella enterica*[45] and *Campylobacter jejuni*[46]. Recombination is especially potent as a factor in within-host evolution when a mixed infection is present to provide genetic material for import. For example, in *H. pylori* it has been estimated that when multiple strains are present within a host then homologous recombination can accelerate evolution by up to 100-fold[25,26,47]. An alternative outcome is the gain of novel, non-homologous, genetic material, which is counterbalanced by the occasional loss of content through genome degradation[48]. For example, both the gain and loss of genes have been described during the early stages of adaptation of *Pseudomonas aeruginosa* to the lungs of patients with cystic fibrosis[49]. Genomic plasticity due to gene gain and loss has an important role in adaptation to changes in selective pressures, with antibiotic resistance genes in many pathogens being found on mobile genetic elements such as plasmids[50], transposons[51] and bacteriophages[52].

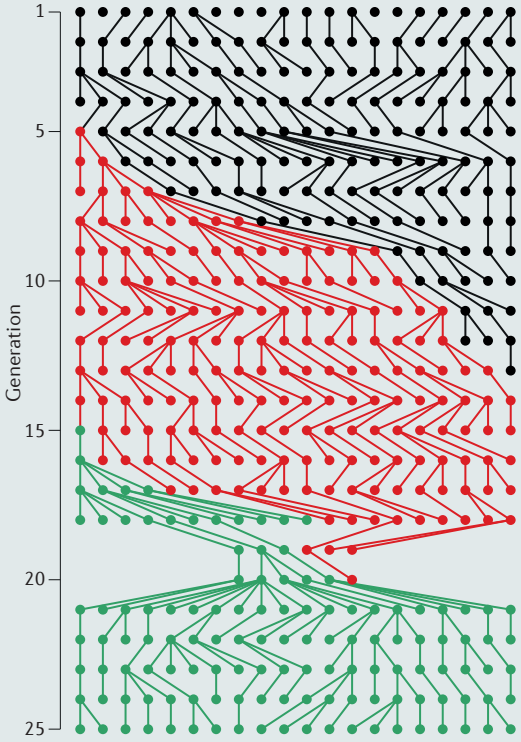
Within-host diversity is shaped by random genetic drift[53,54] (BOX 2), and by both purifying selection and diversifying selection. The role of genetic drift might seem counterintuitive as the total number of bacterial cells infecting a host can be very high, which dampens drift. However, drift is amplified by several factors, such as the isolation of distinct bacterial populations in different body parts and fluctuations in population size. Indeed, fluctuations in population size have been found to be an important factor for the within-host evolution of *S. aureus*[55]. The effect of selection on within-host diversity is non-random and can take many forms. At longer evolutionary scales, the effect of purifying selection dominates the evolutionary landscape, as evidenced by a paucity of non-synonymous polymorphisms relative to synonymous polymorphisms (measured by the dN/dS ratio). However, purifying selection is expected to be weaker in within-host populations than in other populations, owing to stronger genetic drift and the little time available to purge mutations that are only slightly deleterious[56]. This may be a factor in the seemingly higher rate of short-term evolution observed in within-host populations, as some mutations that persist over short time scales would be purged over longer time scales[8]. Diversifying selection, which favours the evolution of new variants, has been found to be important in several within-host studies[57–59]. For example, certain genes encoding outer membrane proteins of *H. pylori* have been found to evolve faster than the remainder of the genome,

Phase variation
A mechanism that bacteria use to enable the rapid evolution of a specific trait in which frequently occurring, reversible mutations control gene expression.

Horizontal gene transfer
The uptake of genetic material by a recipient cell using various mechanisms, such as transformation of naked DNA, bacteriophage-mediated transduction or plasmid-mediated conjugation.

Box 2 | **A primer on genetic drift**

Genetic drift is the process whereby allele frequencies fluctuate over time, owing to the birth and death of individuals in the population. To illustrate this concept, the evolution of a bacterial population was simulated under the assumptions of the Wright–Fisher model (see the figure). In this example, the population has an initial effective population size of 20 bacteria, and each generation (horizontal row) is formed by randomly selecting parents from the previous generation (parental relationships are shown by lines connecting individuals from one row to the next). In the fifth generation, the leftmost individual mutates (shown in red), and the new mutation is inherited by all of its descendants. This mutation slightly increases the fitness of its carriers, so that they have a slightly higher expected number of offspring when compared with the wild-type bacteria (shown in black). For that reason, even though its frequency was only 1 in 20 in the fifth generation, the mutant allele sweeps through the population, which means that its frequency progressively increases. By the fourteenth generation, the entire population carries the mutant allele, which means that it has become fixed and the non-mutated allele has become extinct. In the fifteenth generation, a new mutant allele appears (shown in green), which again has slightly increased fitness compared with the previous mutant allele (shown in red), and its frequency starts to increase progressively. However, in the nineteenth and twentieth generations, the bacteria undergo a bottleneck, so that the effective population size is reduced to only six individuals. Bottlenecks enhance the effect of genetic drift, so that sweeping, fixation and extinction occur at faster rates. By the end of the bottleneck, in the twenty-first generation, the newly mutated allele (shown in green) has swept through the population and the previously mutated allele (shown in red) has now become extinct. Similar loss of variation occurs when a new population emerges from a small number of individuals drawn from a larger population, a phenomenon known as the founder effect, which corresponds to the transmission bottleneck in the case of the transmission from a donor to a new host.



Homologous recombination
An evolutionary event in which a segment of the genome of a recipient cell is replaced with a homologous segment of the genome from a donor cell.

Random genetic drift
Variations in allele frequency in a population caused by the random genetic sampling that occurs during the birth and death of individuals.

Purifying selection
The tendency for an allele that incurs a survival or reproductive disadvantage to decrease in frequency and become lost. Deleterious alleles may nevertheless become fixed owing to random genetic drift.

Diversifying selection
A form of recurrent positive selection that favours the emergence of new alleles in a population; for example, the selective pressure of the host immune system on antigen evolution in pathogens.

dN/dS ratio
The ratio of the number of non-synonymous substitutions, which alter the protein sequence, to the number of synonymous substitutions, which do not alter the protein sequence, normalized by the ratio expected under neutrality. A dN/dS ratio below one indicates purifying selection and above one indicates positive selection.

Fixation
The point at which an allele replaces all alternative alleles of the same locus in a population. This coincides with loss of the other alleles.

Incomplete lineage sorting
A phenomenon whereby a gene tree is discordant with the population or species tree. This occurs when lineages that are ancestral to several different populations split before, and in a different order to, the splitting of the respective populations. For within-host populations, this causes discordance between phylogenies and transmission trees.

presumably because these proteins would otherwise be targeted by the host immune system[25,47,60]. In addition to the immune system, diversifying selection can also be driven by the use of antimicrobial drugs[61] and in either case may result in an 'arms race' between the bacterial pathogen and its human host.

Host-to-host pathogen transmission

Using phylogenies to reconstruct transmission. WGS of bacterial pathogens has enabled the development of a new approach to track direct transmission between hosts, based on the simple idea that genomes sampled from the donor and the recipient following a direct transmission event are expected to be very similar[6,62–64]. Most studies of recent transmission reported so far have been based on single genomes sampled from each host and have therefore not taken within-host diversity into account. However, depending on the duration of host carriage compared with the rate of genomic evolution, within-host evolution can make direct transmission events difficult to reconstruct based on a single genome per host[23]. For example, consider single genomes sampled from three individual hosts, host A, host B and host C. The two genomes sampled from host B and host C may be more closely related to each other than to the genome sampled from host A; however, perhaps counterintuitively, this does not rule out the possibility that host A infected both host B and host C through a lineage that is different from the one that was sampled from host A (FIG. 2). In other words, the genetic variants are representative of the phylogenetic relationships between the sampled genomes, but these may not directly reflect the order and timing of transmission events from donor to recipient. This problem is similar to the relationship between species trees and gene trees caused by incomplete lineage sorting[65]. Formally establishing the relationship between phylogeny and transmission networks requires the explicit modelling of the transmission bottleneck and within-host genome dynamics, which increases the accuracy of the results but also causes much uncertainty regarding who infected whom if based on single genomes per host[2,66,67].

By capturing the diversity within each host — for example, by sequencing multiple genomes per host sampled either simultaneously or longitudinally — the profiling of within-host evolution can reconstruct transmission events with much greater certainty than with the sampling of single genomes. This approach greatly clarifies who infected whom; for example, in cases in which host

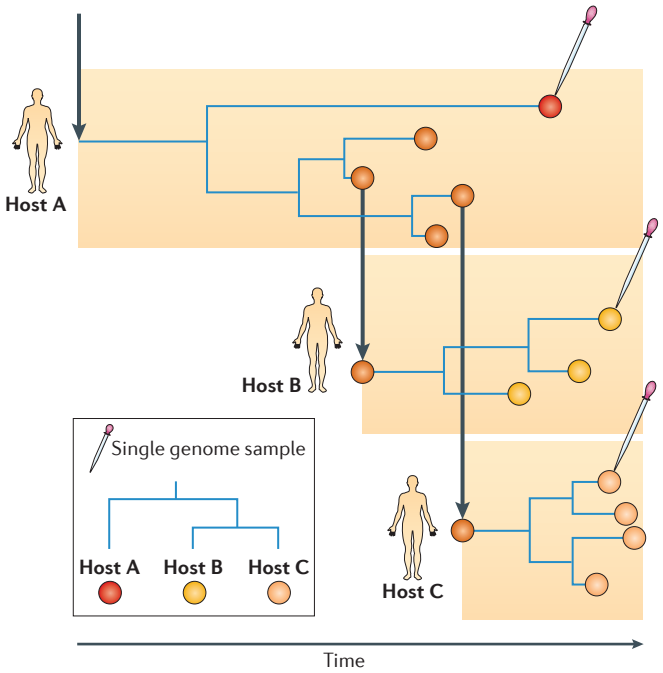


Figure 2 | **Effect of within-host evolution on the reconstruction of transmission events.** This example illustrates the benefit of sampling several genomes, rather than a single genome, from each host when reconstructing a transmission network. In this example, three hosts (host A, host B and host C) are infected with a bacterial pathogen, with host A having become infected first and having directly transmitted the pathogen to host B and then to host C, as indicated by the vertical arrows. If only a single genome is sampled from each host (see inset), their ancestry is only weakly informative about who infected whom. In the example shown, the ancestry of the genome sampled from host A is such that the genomes sampled from host B and host C are more closely related to one another than with the genome sampled from host A, which might be mistaken as evidence for host B having infected host C (or vice versa). This error can be avoided if within-host evolution is accounted for, but doing so with only a single genome sampled from each host results in substantial uncertainty in reconstructions of transmission. However, when several genomes are available for each host (coloured circles in the transmission tree), an ancestry analysis becomes much more informative about the transmission events. In the example shown, the three genomes sampled from host B form a clade that is closely related to a clade of genomes sampled from host A, as do the four genomes sampled from host C, correctly suggesting that host A infected both host B and host C. Different coloured circles represent genetic variation in the pathogen genome following within-host mutations.

A has infected host B, the diversity of pathogen genomes sampled from host B should be a clade contained within the diversity of pathogen genomes sampled from host A (FIG. 2), a principle that has been applied previously in other settings, such as in studies of HIV infection[68,69]. In a pioneering study that applied this approach to bacteria, an outbreak of *B. dolosa* among 14 patients with cystic fibrosis over 16 years was investigated by sequencing a total of 112 isolates, clearly identifying probable donors and recipients in transmission events[59]. A similar study used 168 consecutive isolates of *M. abscessus* from 31 patients of an adult cystic fibrosis centre in the United Kingdom, and found that 11 patients were likely to have infected one another during hospital visits[34]. Deep sequencing of *S. aureus* was crucial in revealing the likely role of a healthcare worker in spreading infections at a neonatal intensive care unit in the United Kingdom[70], tracing the source of an outbreak in an adult intensive care unit in Thailand to a single individual with high within-host diversity[71] and reconstructing the transmission network between animal patients and staff at a veterinary hospital in the United Kingdom[72].

Selective sweep
The rapid increase in frequency and fixation of an advantageous allele. Selective sweeps are caused by positive selection.

Identifying the cause of recurrent infections. A problem related to the reconstruction of transmission networks concerns the ability to distinguish reinfection from relapse in individuals who have subsequent episodes of disease. For example, in a study of African patients with recurrent invasive non-typhoidal *Salmonella* spp. infection, 25 out of 32 recurrent episodes were caused by identical or almost identical consecutive strains, suggesting relapse, whereas the remaining seven episodes showed many differences between consecutive strains, suggesting reinfection[73]. The same approach was applied to recurring cases of infection with *C. difficile*, and relapse was found to be about five times more frequent than reinfection in both the United Kingdom and Ireland[74,75]. The question of relapse versus reinfection is especially relevant for infections with *M. tuberculosis*, which are more frequent in individuals with HIV and can often recur after seemingly successful treatment. In a study focused on recurrent *M. tuberculosis* infection, mostly in patients who were HIV negative, 33 clear episodes of relapse were found compared with only three episodes of reinfection[76]. Another study on recurrent *M. tuberculosis* infection examined a cohort with a high prevalence of HIV-positive individuals and identified 55 episodes of relapse and 20 episodes of reinfection, with the latter more likely to be in individuals who were HIV positive[77]. This study also found that lineage 2 was more likely to cause reinfection whereas lineage 3 was more likely to cause relapse, echoing differences in adaptive strategies that have also been observed when comparing transmissibility between lineages of *M. tuberculosis*[78].

Evolution of antibiotic resistance

Natural selection within individual hosts. Following successful transmission, a pathogen must survive in the new host to enable subsequent transmission and continuation of its life cycle. Survival within the host poses many challenges — including physical barriers to colonization and infection, competition with the native microbiota, containment by the host immune system, basic sanitation and medical intervention — and the relative importance of these forces may differ from one host to the next. However, the ability of bacteria to rapidly adapt to selection pressures such as antibiotic treatment, even within a single host[25–34], is highlighted by the rapid rise in antibiotic resistance, exacerbated by the overuse and misuse of antibiotics[79].

The evolution of antibiotic resistance within individual hosts epitomizes the evolutionary principle of a selective sweep, as the survival advantage gained from the resistance-conferring gene (or mutation within a gene) in the presence of antibiotic treatment increases the frequency of the gene (or mutation) in the within-host population. Furthermore, horizontal gene transfer can accelerate the rate of spread of resistance-conferring genes that are carried by plasmids or other mobile elements.

Until the recent advances in WGS, the characterization of bacterial microevolution was the sole purview of experimental studies[80]. WGS captures this process in extraordinary detail and, as a result, has now provided valuable insights into many within-host examples of the evolution of antibiotic resistance. Consequently, the occurrence and spread of individual point mutations that increase antibiotic resistance and contribute to treatment failure can often be pinpointed, as observed in the evolution of vancomycin-resistant *S. aureus* in a patient with an initially vancomycin-susceptible bloodstream infection[81]. Rifampicin resistance-associated mutations in RNA polymerase β-subunit (*rpoB*), a β-lactam resistance-associated mutation in β-lactamase regulator protein (*blaR1*) and mutations in several genes associated with reduced vancomycin susceptibility were identified in bacteria isolated from the patient only after treatment with the respective antibiotic. Mutations progressively accumulated over time, concomitant with a stepwise decrease in vancomycin susceptibility[81].

The sheer potency of bacterial evolution was revealed by the evolution of extensively drug resistant (XDR) *M. tuberculosis* from a wholly drug-sensitive ancestor, in a single patient, over just 3.5 years[82,83]. Resistance evolved to seven drugs to which the patient was exposed: rifampicin (to which resistance was conferred by mutations in *rpoB*), isoniazid (mutations in catalase–peroxidase (*katG*)), ethambutol (mutations in arabinosyltransferase B (*embB*)), streptomycin (mutations in 16S rRNA (*rrs*) and rRNA small subunit methyltransferase (*gid*)), ofloxacin (mutations in DNA gyrase subunit B (*gyrB*)), ethionamide (mutations in D-inositol 3-phosphate glycosyltransferase (*mshA*)) and amikacin (mutations in *rrs*). In most cases, resistance-conferring mutations arose in one or more lineages in the within-patient population after only several months of treatment with the respective antibiotic. Ultimately, in an extreme example of clonal interference, a single multidrug-resistant lineage prevailed, displacing the others. The spread of advantageous resistance mutations increased the measured substitution rate 14-fold, and propelled the substitution of numerous synonymous mutations on the same genetic background as the advantageous mutations, in an example of a process known as hitchhiking.

Some naturally evolved resistance mutations seem to be pleiotropic, leading to multifaceted changes in bacterial phenotype, including resistance to drugs to which the population has not been exposed[81,84,85]. In five patients, *S. aureus* with reduced vancomycin susceptibility evolved from initially susceptible strains following fewer than 6 weeks of treatment with vancomycin. Non-synonymous mutations in the WalKR two-component regulator[84], which has a central role in controlling cell wall metabolism[86], were found in four of these patients, and in six of eight unrelated bacterial isolates with reduced vancomycin susceptibility. The majority of isolates with reduced vancomycin susceptibility also demonstrated reduced daptomycin susceptibility, despite not being exposed to this antibiotic[84]. Allelic exchange experiments showed that mutations in *walKR* led to morphological and transcriptional remodelling, including reduced autolytic activity, increased cell wall thickness, reduced biofilm formation and attenuated virulence in a wax moth model of infection[84]. This incidental pre-adaptation to other antibiotics, a phenomenon known as cross-resistance, presents a particular cause for concern by further limiting treatment options.

Revealing the adaptive potential of bacterial pathogens. The ability to study the evolution of antibiotic resistance in extensive detail using WGS has enabled the discovery of novel mutations and resistance mechanisms that may help predict treatment failure and aid in the development of new drugs[87–93]. These *in vivo* surveys of within-host evolution present a complementary approach to experimental studies. Although it is not possible for all variables to be controlled, observing natural evolution in within-host populations has the advantage of incorporating complexities such as strain differences, the host environment and realistic fitness trade-offs, which has highlighted the formidable adaptive potential of bacterial pathogens.

Certain genes have been identified as common sites of adaptive evolution in response to antibiotic treatment, with independently arising mutations often hitting different sites in the same gene. Common targets of selection include *rpoB* in response to treatment with rifampicin (which inhibits RNA polymerase) in several species[85,94–96], DNA gyrase subunit A (*gyrA*)[38,96–99] and DNA topoisomerase IV (*grlA*; also known as *parC*)[100] in response to fluoroquinolones (which target topoisomerases) in various species, and the cell wall metabolism genes *walKR* and *vraRS* in response to treatment with vancomycin (which inhibits cell wall synthesis) in *S. aureus*[81,84,101].

Reports of convergent evolution in independent patients who were exposed to the same drugs provide the strongest evidence for adaptation. A study investigated the evolution of resistance to a wide range of drugs in 123 strains of *M. tuberculosis* representing transmission clusters and epidemiologically unrelated cases. Resistance evolved independently up to 20 times to the first line drugs isoniazid, pyrazinamide, ethambutol and rifampicin through substitutions in *katG* and NADH-dependent enoyl-acyl carrier protein reductase (*inhA*; conferring resistance to isoniazid), *pncA* (conferring resistance to pyrazinamide), *embB* (conferring resistance to ethambutol) and *rpoB* (conferring resistance to rifampicin). Identical, independent substitutions were detected at two sites in *rpoB* and one in *embB*[87]. The observation that resistance to common antibiotics has evolved not once, but many times in parallel, highlights the adaptive potential and repeatability of bacterial evolution.

Given the rapidity with which bacteria can respond to selection pressures by within-host evolution, it may seem surprising that antibiotic resistance has not spread more quickly. It has been suggested that the fitness costs associated with resistance might explain this discrepancy because resistance-conferring substitutions in key enzymes may reduce the efficiency of replication and transcription, and specialist resistance-conferring proteins may be costly to produce[88,102]. However, compensatory mutations may arise that counteract the fitness costs associated with antibiotic resistance (FIG. 3). WGS enabled a joint investigation of naturally evolving rifampicin resistance-conferring

Clonal interference
An evolutionary dynamic in which selectively advantageous alleles at a given locus in one lineage outcompete advantageous alleles at other loci in other lineages, causing them to become extinct. In organisms with the capacity for genome recombination, this can be avoided by combining all advantageous mutations in the same genome.

Hitchhiking
The effect whereby an allele can increase in frequency even though it is not favoured by selection, only because it is found in the same genomes as other alleles of other loci that have a selective advantage.

Pleiotropic
The unexpected influence of one locus on multiple, apparently unrelated, phenotypes.

Pre-adaptation
A phenomenon whereby a previously existing trait confers an advantage in an environment to which it was not previously exposed (also known as exaptation).

Fitness trade-offs
The existence of some constraint, possibly mechanistic or genetic, that causes adaptations to one selection pressure to be disadvantageous with respect to another.

Convergent evolution
The occurrence of mutations resulting in the same phenotype in two or more independently evolving lineages; these often arise in the same gene and may even occur at the same site.

Compensatory mutations
Mutations that redress, possibly only partially, the fitness cost of mutations conferring adaptation to specific selection pressures, such as antibiotic resistance. Without compensatory mutations, adaptations that incur a fitness cost may be lost when the selection pressure is removed.

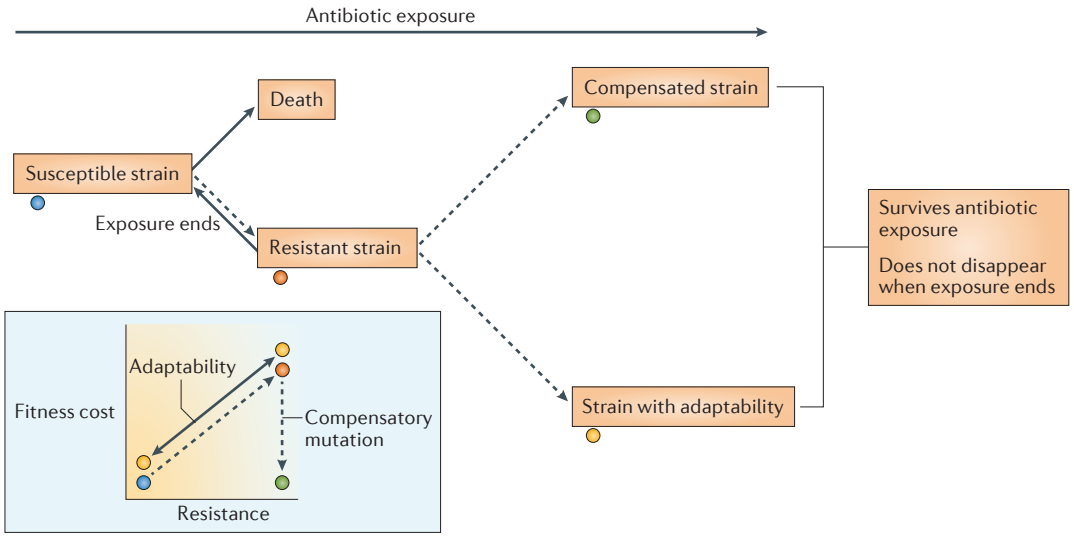


Figure 3 | **Within-host adaptive potential during exposure to antibiotics.** When exposed to an antibiotic, a susceptible bacterial strain (blue circle) is highly likely to be killed, but may occasionally survive by evolving into a resistant strain (orange circle). Resistance usually has a high fitness cost (inset), so that resistant strains usually disappear when not exposed to the antibiotic. However, resistant strains can evolve compensatory mutations (green circle) so that they remain resistant without the associated fitness cost. Such compensated strains pose a serious danger to public health, because they do not disappear simply as a result of antibiotic disuse. Alternatively, strains may evolve adaptability (yellow circle), enabling them to quickly switch resistance on or off and therefore avoid the associated fitness cost, presenting a similar risk to public health as that presented by compensated strains.

and compensatory mutations in tuberculosis[88]. Putative compensatory mutations were detected in 38 genes, with a particular enrichment of mutations affecting the interface between RNA polymerase α-subunit and β′-subunit. High *in vitro* fitness was demonstrated in strains bearing compensatory mutations subject to strong convergent evolution across lineages, demonstrating that a cost associated with resistance is not inevitable.

Even when fitness costs are inevitable, bacteria can evolve adaptability. A well-studied form of adaptability is the previously mentioned phase variation mechanism, in which frequent, reversible genetic changes determine the expression of particular genes, such as those encoding surface antigens. Rapid mutation enables the expression of these genes to be quickly adapted to changing selection pressures. Another form of adaptability that is increasingly being recognized is heteroresistance. For example, population analysis profiles (PAPs) reveal that, remarkably, most methicillin-resistant *S. aureus* (MRSA) strains are heteroresistant, meaning that the vast majority of the bacterial cells have only low-level or moderate-level resistance in the absence of exposure to methicillin, whereas cells exhibiting a several hundred-fold increase in resistance are present only at very low frequency[95]. On exposure to methicillin, these highly resistant bacterial cells can rapidly sweep through the population. Genome sequencing has found that high-level resistance can be conferred by numerous individual mutations within the *S. aureus* genome, many of which affect gene expression. Although this indicates a large target for selection, the over-representation of mutations in genes involved in transcription and the stringent response suggests a degree of convergent evolution[95].

Heteroresistance demonstrates the capacity for large bacterial populations in the host to mutate and harbour potentially advantageous mutations, facilitating adaptability. Observations of heteroresistance to a range of antibiotics in several pathogens, including *S. aureus*, *M. tuberculosis*, *S. pneumoniae* and enterococci[103], indicate that within-host evolution is not the exception, but the norm, in bacteria and may be a leading contributor to monotherapy treatment failure[100]. As the potential for adaptability seems to vary between bacterial strains, the capacity for within-host evolution might even explain differences in the prevalences of global lineages[104].

Adaptation to the host environment

In the absence of antibiotic selective pressure, bacterial pathogens must still overcome various challenges in the host, such as establishing a successful niche by acquiring nutrient sources, and evading killing by other microorganisms and the innate and adaptive immune system. Such challenges represent opportunities for adaptation if mutations arise that confer survival or reproductive advantages. WGS is beginning to uncover diverse signals of adaptation to the host environment, and this is shedding new light on the main forces that shape bacterial populations in the host.

Opportunities for adaptation are particularly abundant in infections of immunocompromised hosts caused by opportunistic pathogens that are usually adapted to living elsewhere. Infection of patients with cystic fibrosis by bacteria that are usually non-pathogenic in humans has been particularly well studied (FIG. 4). *In vitro* experiments have described an initial period of rapid adaptation to new environments that slows as the population improves

Adaptability
The ability to rapidly adapt to a change in selective pressure, such as antibiotic use.

Heteroresistance
Varying levels of antibiotic resistance within an extremely closely related population, such as an individual colony.

Stringent response
A stress response that diverts cellular resources towards survival during nutrient limitation by instigating widespread regulatory changes, including the upregulation of amino acid synthesis and protease production.

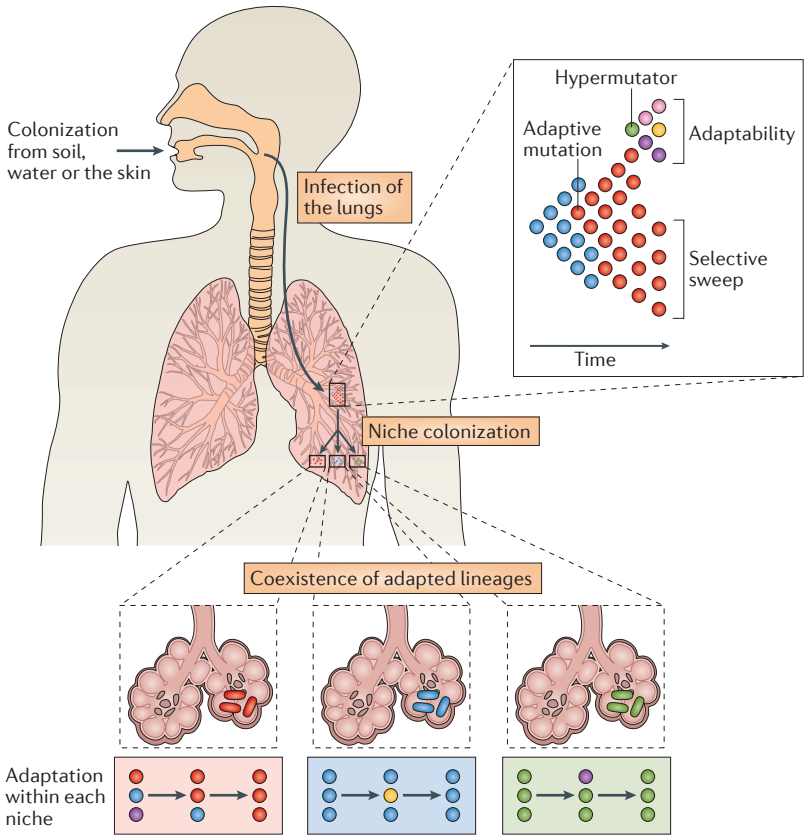


Figure 4 | **Within-host evolution of pathogens in the lungs of a patient with cystic fibrosis.** Infection in the lungs of a patient with cystic fibrosis begins with transmission from the environment or the skin of a donor, and progresses with a rapid increase in the size of the pathogen population. Mutations occasionally occur, some of which may be adaptive mutations that spread through the pathogen population in a selective sweep. Another important mechanism of genome evolution for these pathogens is hypermutation, whereby a strain loses the function of its mismatch repair machinery and thus becomes a hypermutator, with a mutation rate that is increased several fold, thus greatly increasing its evolutionary potential. As the infection progresses, the pathogen colonizes all ecological niches within the cystic fibrosis lungs, and separate adaptation to each niche leads to the coexistence of differentially adapted lineages.

in fitness[105]. Similar patterns of rapid adaptation have been detected in opportunistic infections in patients with cystic fibrosis. In an outbreak of *B. dolosa* among patients with cystic fibrosis, signatures of selection in the form of convergent evolution were detected in the genes responsible for O-antigen repeat expression in the lipopolysaccharide coat, oxygen-dependent regulation and resistance to ciprofloxacin[59]. O-antigens elicit strong immune responses and typically exhibit high variability that probably reflects selection for immune evasion. It is thought that changes in oxygen-dependent regulation may be crucial for the adaptation of free-living organisms to the cystic fibrosis lung, where oxygen availability is reduced owing to the build-up of mucus and the formation of biofilms[59].

Patterns of convergent evolution have been used to identify pathoadaptive mutations in *P. aeruginosa* infections in the lungs of patients with cystic fibrosis, highlighting the roles of genes encoding transcriptional regulators, a lipopolysaccharide biosynthetic protein, an outer membrane antigen and antibiotic resistance factors in pathogenicity[58,96,99]. In some cases, within-host adaptation may be facilitated by hypermutators[96,97,99,106], which accelerate the rate of adaptation by increasing the supply of genetic novelty, a greater proportion of which is likely to be beneficial in a novel environment[107]. Selection for changes to transcriptional regulators such as sigma factors and two-component regulatory systems indicates that transcriptional reprogramming may be a key element of adaptation to the host environment in opportunistic pathogens[96,99]. Such changes to gene expression can cause rapid phenotypic change with little evolution in terms of the number of genetic mutations. In general, examining the molecular mechanisms that drive signals of within-host adaptation is difficult because, unlike in the case of antibiotic resistance, the phenotypes are difficult to recreate *in vitro*, owing to the complexity of the within-host environment and the potential role of host–pathogen interactions. However, an experimental cystic fibrosis lung, which may negate some of these difficulties, has been used to study one candidate for adaptive transcriptional reprogramming, the iron-scavenging haem utilization system in *Pseudomonas* spp. The increased expression of this system was induced by mutations in promoter sequences and was demonstrated to be advantageous to the growth of *P. aeruginosa* in the presence of haemoglobin[108]. Although such experiments are challenging, they provide useful validation of the findings in observational studies. Within-host population structure is likely to be an important modifier of adaptation to the host environment. In homogeneous environments, drift and positive selection are expected to lead to the evolution of one prevailing lineage (BOX 2). However, within-host heterogeneity can facilitate the existence of several coexisting lineages, either through neutral processes such as the genetic isolation of subpopulations, or adaptive processes such as niche differentiation (FIG. 4). An investigation of the within-host population structure of *B. dolosa* infections in five patients with cystic fibrosis found stable coexistence of several lineages for more than five years, and evidence of parallel evolution even in individual patients, manifest as independent substitutions in genes with functions in antibiotic resistance, biosynthesis of the outer membrane, iron scavenging and oxygen sensing. These results uncovered a previously unsuspected population structure in the infected lung[38]. Sequencing of a *P. aeruginosa* infection sampled over a 32-year period revealed that the infecting strain had rapidly diverged into distinct sublineages, with unique functional signatures and a threefold variation in the rate of adaptation[97]. Differences in the frequency of these sublineages in the sinuses and the upper and lower respiratory tract, and their stable coexistence over several decades, suggested specialization of these sublineages to distinct within-host niches. The functional differences between sublineages were manifested in growth rates, mucoidy and the production of proteases, with one cluster of isolates displaying a phenotypic signature characteristic of chronic infections, namely longer doubling times, mucoidy and the loss of protease production[97].

Notwithstanding the numerous examples of within-host adaptation identified by WGS, most of the genome nevertheless remains subject to purifying selection. Even

Pathoadaptive
An adaptation that confers pathogenicity.

Hypermutators
An individual or lineage with increased mutation rate, usually as a result of a loss of functionality in DNA repair systems.

Positive selection
The tendency for an allele that confers a survival or reproductive advantage to increase in frequency and become fixed at a higher rate. Advantageous alleles may nevertheless become lost, owing to random genetic drift despite positive selection.

within individual hosts, the genome-wide dN/dS ratio typically lies below its neutral expectation of one and thus is indicative of purifying selection. In an outbreak of *B. dolosa* in patients with cystic fibrosis, singly mutated genes were shown to have a dN/dS ratio of 0.63, substantially below one[38], similar to the dN/dS ratios of 0.56 and 0.66 reported in *P. aeroginosa* outbreaks in patients with cystic fibrosis[96,97] and the dN/dS ratio of 0.55 reported for nasal populations of *S. aureus* colonizing individual hosts[55]. Taken together, these figures indicate that the prevailing effect of natural selection in the host is to conserve functionality for the majority of genes, with adaptation acting only on some genomic positions in a subset of genes.

Within-host adaptation might be less frequent in common bacterial pathogens of humans that have already had many generations over which to adapt to an infectious lifestyle. A study of asymptomatic nasal carriage of *S. aureus* in healthy adults concluded that adaptive evolution was rare, identifying only a weak signal of convergent evolution among 13 carriers. Notably, several non-synonymous or protein-truncating mutations were observed in the secreted staphylococcal enterotoxin type G (SEG; also known as EntG) and the surface anchored proteins Ebh and SasC[55]. Indeed, surface proteins and secreted proteins are common antigens that interact directly with the adaptive immune system causing diversifying selection to favour novel epitopes that can evade immune recognition. In another example, a longitudinal study of the asymptomatic carriage of *H. pylori* reported excess horizontal gene transfer in the Hop family of outer membrane proteins[25]. It may be that most within-host adaptation in common human pathogens occurs within antigenic loci, although well known examples of recurrent evolution in other genes, such as loss-of-function *agrC* and *lasR* mutations that knock out quorum sensing in *S. aureus* and *P. aeruginosa*, respectively, suggest this is not the full story[109].

Selection pressures within the host are complex and thus difficult to examine. Conflicts arise not only between bacteria and the host, owing to the need for nutrient acquisition, evasion of the host immune system and onward transmission, but also with other bacteria, owing to direct competition for resources. Complex selective forces may also be at work as a result of social dynamics within a bacterial community. For example, whereas loss-of-function mutations might in many cases signal host adaptation[110], the loss of siderophore production in *P. aeruginosa* during long-term infections in patients with cystic fibrosis can instead be driven by cheating behaviour[111]. In some patients, cheater cells within the *P. aeruginosa* population that no longer synthesize the siderophore pyoverdine still retain the pyoverdine receptor, enabling them to uptake iron bound to pyoverdine produced by cooperative bacteria. Only when the cooperators are lost from the population do the cheats lose the pyoverdine receptor.

Mucoidy
A bacterial phenotype describing the production of glycoproteins resembling mucus.

Quorum sensing
Mechanism by which a cell responds to changes in population size or density, classically by the secretion and detection of small peptides (also known as pheromones).

Evolution of disease severity

A key question that follows on from the discovery that bacteria rapidly evolve on timescales relevant to colonization and infection is what is the effect of this within-host bacterial evolution on disease manifestation? In other words, does within-host evolution increase bacterial virulence? Dramatic phenotypic changes and global transcriptional remodelling in the host support the feasibility of this notion[84,85,108,112,113]. For example, a single spontaneous mutation in *relA* in a persistent *S. aureus* infection was sufficient for permanent activation of the stringent response, which induced multifactorial phenotypic changes in the bacterial cells, including reduced growth, smaller colony size, reduced attachment to human cells and attenuated virulence in a wax moth model of infection[85]. WGS provides an unprecedented opportunity to reveal how within-host bacterial evolution is associated with the progression of disease in numerous bacterial pathogens.

Many clinically important bacterial pathogens are predominantly commensals in humans, including *S. aureus*, *N. meningitidis*, *S. pneumoniae*, *H. pylori* and *E. coli*. As carriage is common and invasive disease is not an obligate part of the lifecycle of such bacteria, most transmission may be asymptomatic, with disease being caused by previously carried bacteria rather than transmission. In the case of *S. aureus*, 82% of serious infections of deep tissue are self-infections, in the sense that the invasive strain matches the nasally carried strain[114]. Although the triggers for invasive disease are likely to be complex and multifactorial, and include the general health of the patient and host genetics, the rapidity of within-host evolution raises the important question of whether the carried bacteria can evolve to become more virulent.

In one long-term *S. aureus* nasal carrier who developed a severe bloodstream infection, WGS charted the genomic changes that accompanied the transition from asymptomatic carriage to invasive disease[28]. Over 13 months, several isolates were sequenced from each of a longitudinal series of samples, which identified 30 point mutations and four insertions or deletions. The population evolved at a steady rate except for a cluster of mutations preceding the transition to disease. Eight mutations differentiated the original nasal population from the bloodstream population, of which half were protein-truncating, including a mutation in *rsp*, a gene encoding an *AraC*-family transcriptional regulator (AFTR) previously implicated in pathogenicity[115]. AFTRs are regulators of carbon metabolism, stress responses and virulence that respond to changing environmental conditions, such as antibiotic use and oxidative stress[115]. In *N. meningitidis*, a loss-of-function mutation in the AFTR *mpeR* is associated with the hypervirulent ST-32 complex[116]. A statistically significant excess of protein-truncating mutations accompanied the progression of *S. aureus* to invasive disease in this patient. However, both for this patient and more generally, whether disease progression is driven by bacterial evolution (or whether, conversely, bacterial evolution is driven by the worsening health of the patient) remains an open question[28,117].

Attenuation of virulence as an evolutionary strategy. Rather than exacerbating virulence, there is evidence that bacterial evolution in the host can attenuate virulence[57,118]. *Burkholderia pseudomallei* is the causative agent of the

potentially life-threatening disease melioidosis, and is not considered a commensal. However, out of 707 survivors in a 23-year study in Darwin, Australia, one patient developed persistent asymptomatic carriage[57]. Two isolates from this patient sampled 11.5 years apart differed by 23 point mutations, including a protein-truncating mutation in the gene encoding the universal stress response sigma factor RpoS. Four deletions in chromosome 2 removed 221 genes during this time, including some genes involved in metabolism, survival outside of the host and pathogenesis. Both isolates showed loss-of-function mutations in the essential virulence factor *wcbR*, a component of the capsular polysaccharide I locus, suggesting that early attenuation of virulence may have promoted long-term persistence in this patient.

A recent study on the virulence of *H. pylori* sought to explain why two regions in Colombia, separated by only 200 km, and with similarly high prevalence of infection, showed a 25-fold difference in the incidence rates of gastric cancer[119]. In the region where the incidence of gastric cancer was low, most individuals were of African descent and colonized asymptomatically with African lineages of *H. pylori*, whereas in the region where the incidence of gastric cancer was high, African ancestry was low in the human population (~3%) but higher in the bacterial population (~20%). The highest risk of disease was found in individuals of non-African descent who were infected by *H. pylori* with substantial African ancestry. This suggests that, in Africa, *H. pylori* and its human host have co-evolved towards lower virulence, which could also explain why the high prevalence of *H. pylori* in Africa does not seem to correspond to similarly high incidence of disease[120]. According to the adaptive trade-off hypothesis, host–pathogen co-evolution can lead to the evolution of disease with reduced severity[121–123]. *H. pylori* is often transmitted in families or communities of closely related individuals[26,124,125], and the virulence optimum is expected to be particularly low for pathogens transmitted in this way because their long-term survival is linked with that of the host[126].

Strains that have evolved attenuated virulence may even provide a therapeutic tool. Although urinary tract infections (UTIs) caused by *E. coli* are severe and potentially life-threatening, *E. coli* can also be carried asymptomatically in the bladder in a state known as asymptomatic bacteriuria (ABU). Strains associated with ABU have been extensively used for therapeutic urinary bladder colonization in patients with chronic UTIs[127]. Previous work has shown that 50% of *E. coli* strains associated with long-term ABU evolved from uropathogenic strains in which genome reduction and inactivating mutations attenuated virulence[128]. By studying six patients colonized with a single prototypic ABU-associated strain, WGS showed that further evolution of the strain following therapeutic inoculation is suggestive of within-host adaptation specific to each individual host. In some patients who required repeated therapeutic inoculations with the prototypic strain, mutations were observed repeatedly in loci involved in iron uptake (*fecIR* promoter), the oxidative stress response (*frmR*) and osmoregulated periplasmic glucan synthesis (*mdoH*). The repeatability of evolution within individual hosts suggested the existence of a characteristic signature of adaptation to individual hosts sometimes known as host imprinting. Ongoing loss of gene function suggested that progressive evolution towards commensalism rather than virulence was favoured in these ABU-associated strains[127].

Conclusion

By enabling the study of microbial evolution inside our own bodies, WGS has revealed a remarkable degree of adaptability of bacteria in the human host. Even the notoriously slow-growing and slowly evolving[32] *M. tuberculosis* is capable of extraordinarily rapid adaptation in response to antibiotics, with one infection evolving resistance to seven antibiotics over a 42-month period[82]. The numerous examples in which resistance has evolved in response to antibiotic therapy (not only in *M. tuberculosis*[82,88] but also *S. aureus*[85,95], *E. coli*[129], *K. pneumoniae*[92,130] and *P. aeruginosa*[96,99]), together with the examples of convergent evolution in opportunistic pathogens such as *P. aeruginosa*[96,97] and *B. dolosa*[59], show that rapid within-host evolution is common across disparate species. This raises some important questions. First, what are the requirements for rapid within-host adaptation? Second, is within-host adaptation of bacterial pathogens sufficiently rapid to influence disease outcome? If so, which conditions favour increased virulence in the host, and which conditions favour attenuated virulence?

Rapid within-host adaptation depends on a range of factors, notably the rate at which mutations that confer a potential benefit occur, the effective population size and the fitness advantage of mutants[131]. The larger these factors are, the faster the rate of adaptation in the population as a whole. Recent studies in *M. tuberculosis* have shown that both the mutation rate and effective population size within the host are very small. The genome-wide mutation rate barely registers one mutation every two years[32,33,82], and diversity is so low that isolates sampled at the same time differ by an average of only 0.5 mutations[33]. Assuming, for illustrative purposes, a generation length of 24 hours[132], this is equivalent to an effective population size of fewer than 200 reproductively viable cells. Target sizes for selection, in terms of the number of possible beneficial mutations, differ between drugs but commonly involve around a dozen sites within a particular gene. To observe adaptation, in the presence of antibiotics, in a timescale of months, 1000-fold or greater daily replicative advantage of resistant over susceptible cells is required. This is plausible for antibiotic-induced selection, in which the advantage of resistance over susceptibility can in some cases be of this order of magnitude. However, these values suggest that *M. tuberculosis* is unlikely to evolve within-host adaptations during the course of a single infection unless there are many potentially beneficial mutations, or unless adaptation is a matter of life or death.

For within-host adaptation to influence the outcome of a single infection, selection must therefore be relatively strong, although other pathogens are likely to be more adaptable than *M. tuberculosis*. For example,

Melioidosis
An infectious disease caused by *Burkholderia pseudomallei*, endemic in South East Asia and Australia, which can lead to sepsis and pneumonia.

Adaptive trade-off hypothesis
The hypothesis that the long-term evolutionary success of a pathogen requires a balance between the duration of infection and virulence, based on the assumption that an increase in virulence decreases the average duration of infection.

Effective population size
The size of an idealized (neutrally evolving, homogeneous) population that is otherwise equivalent to an observed population. The effective population size is typically smaller than the number of individuals in the population, owing to population structure and variation in survival or reproductive viability. Effective population size is also known as N_e.

S. aureus exhibits approximately tenfold higher rates of mutation and within-host diversity than *M. tuberculosis*, at least during asymptomatic carriage[42], suggesting it has greater capacity for adaptation during the course of infection — not just to antibiotic treatment, but also to the host immune response. Although clear-cut examples of adaptation to the host have come mainly from opportunistic infections with *P. aeruginosa* and *B. dolosa*, several studies of *S. aureus* infection have shown that even single point mutations can lead to widespread changes in gene expression that may radically alter phenotypes, including virulence in model systems[84,85]. These studies strongly suggest that within-host bacterial adaptation has the potential to influence disease progression. However, separating cause from effect is crucial for drawing firm conclusions from such studies.

Whereas the adaptive potential of bacteria in the host is remarkable, it is less clear whether we should expect bacteria to evolve to become more or less harmful. Adaptive trade-off theory predicts that when the long-term survival of the pathogen depends on the well-being of the host, the pathogen will tend to evolve reduced virulence[121–123]. Several longitudinal studies of within-host evolution have reported the attenuation of virulence over time[85,127]. However, adaptive trade-off theory concerns selection acting on transmission at the population level, and not during the colonization or infection of individual hosts. When trade-offs exist between onward transmission and short-term survival, we should expect within-host evolution to favour immediate reward at the expense of long-term success.

The years ahead promise further insights as studies continue to investigate within-host evolution in an increasingly diverse array of pathogens. Fully capitalizing on the potential of WGS will require the development of new analysis methods for detecting recent transmission and adaptation, characterizing gene expression and elucidating phenotype-to-genotype relationships. For example, studies of transmission stand to gain in accuracy by sampling within-host diversity (FIG. 2), but existing analysis methods do not currently use this information to its full potential. Our current knowledge of the size of transmission bottlenecks is very limited, and so this is an area in which studies of diversity within donors and recipients could provide valuable information. Predicting bacterial phenotypes from genotypes is likely to grow in importance, particularly in translational settings for genome-based antibiotic-resistance prediction. Already, high levels of accuracy can be achieved[133,134], and the quality of prediction is set to increase as we unravel the genetic architecture of bacterial phenotypes. As sequencing technologies continue to improve, real-time WGS-based diagnostics will provide clinicians with greater insights into the pathogen population and its evolutionary potential to respond to different treatments, and therefore more information on which to base clinical decisions.

1. Fraser, C. *et al.* Virulence and pathogenesis of HIV-1 infection: an evolutionary perspective. *Science* **343**, 1243727 (2014).
2. Pybus, O. G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* **10**, 540–550 (2009).
3. Wilson, A., Ochman, H. & Prager, E. M. Molecular time scale for evolution. *Trends Genet.* **3**, 241–247 (1987).
4. Ochman, H., Elwyn, S. & Moran, N. A. Calibrating bacterial evolution. *Proc. Natl Acad. Sci. USA* **96**, 12638–12643 (1999).
5. Ochman, H. & Wilson, A. C. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26**, 74–86 (1987).
6. Didelot, X. *et al.* Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. *Genome Biol.* **13**, R118 (2012).
7. Wilson, D. J. *et al.* Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Mol. Biol. Evol.* **26**, 385–397 (2009).
8. Morelli, G. *et al.* Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. *PLoS Genet.* **6**, e1001036 (2010).
9. Biek, R., Pybus, O. G., Lloyd-Smith, J. O. & Didelot, X. Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol.* **30**, 306–313 (2015).
10. Ho, S. Y. W. *et al.* Time-dependent rates of molecular evolution. *Mol. Ecol.* **20**, 3087–3101 (2011).
11. Ho, S. Y. W. The changing face of the molecular evolutionary clock. *Trends Ecol. Evol.* **29**, 496–503 (2014).
12. Linz, B. *et al.* A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. *Nat. Commun.* **5**, 4165 (2014).
13. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).
14. Maiden, M. C. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl Acad. Sci. USA* **95**, 3140–3145 (1998).
15. Loman, N. J. *et al.* High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* **10**, 599–606 (2012).
16. Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012).
17. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A. & Crook, D. W. Transforming clinical microbiology with bacterial genome sequencing. *Nat. Rev. Genet.* **13**, 601–612 (2012).
18. Köser, C. U. *et al.* Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* **8**, e1002824 (2012).
19. Wilson, D. J. Insights from genomics into bacterial pathogen populations. *PLoS Pathog.* **8**, e1002874 (2012).
20. Cespedes, C. *et al.* The clonality of *Staphylococcus aureus* nasal carriage. *J. Infect. Dis.* **191**, 444–452 (2005).
21. Mongkolrattanothai, K. *et al.* Simultaneous carriage of multiple genotypes of *Staphylococcus aureus* in children. *J. Med. Microbiol.* **60**, 317–322 (2011).
22. Votintseva, A. A. *et al.* Multiple-strain colonization in nasal carriers of *Staphylococcus aureus*. *J. Clin. Microbiol.* **52**, 1192–1200 (2014).
23. Worby, C. J., Lipsitch, M. & Hanage, W. P. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLoS Comput. Biol.* **10**, e1003549 (2014).
24. Falush, D. *et al.* Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc. Natl Acad. Sci. USA* **98**, 15056–15061 (2001).
25. Kennemann, L. *et al. Helicobacter pylori* genome evolution during human infection. *Proc. Natl Acad. Sci. USA* **108**, 5033–5038 (2011).
This report details extensive mutation and recombination within individual hosts in five longitudinally sampled patients infected with *H. pylori*.
26. Didelot, X. *et al.* Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *Proc. Natl Acad. Sci. USA* **110**, 13880–13885 (2013).
27. Mathers, A. J. *et al. Klebsiella pneumoniae* carbapenemase (KPC) producing *K. pneumoniae* at a single institution: insights into endemicity from whole genome sequencing. *Antimicrob. Agents Chemother.* **59**, 656–1663 (2015).
28. Young, B. C. *et al.* Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc. Natl Acad. Sci. USA* **109**, 4550–4555 (2012).
This study charts the genetic changes associated with the transition from long-term asymptomatic carriage of *S. aureus* to invasive bloodstream infection in one patient, identifying an excess of loss-of-function mutations that separate carried from invasive isolates, including mutations in the transcriptional regulator *rsp*.
29. Eyre, D. W. *et al.* Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N. Engl. J. Med.* **369**, 1195–1205 (2013).
30. He, M. *et al.* Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat. Genet.* **45**, 109–113 (2013).
31. Reeves, P. R. *et al.* Rates of mutation and host transmission for an *Escherichia coli* Clone over 3 years. *PLoS ONE* **6**, e26907 (2011).
32. Ford, C. B. *et al.* Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat. Genet.* **43**, 482–486 (2011).
33. Walker, T. M. *et al.* Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.* **13**, 137–146 (2013).
34. Bryant, J. M. *et al.* Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *Lancet* **381**, 1551–1560 (2013).
35. Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547 (1997).
36. LeClerc, J., Li, B., Payne, W. & Cebula, T. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274**, 1208–1211 (1996).
37. Taddei, F. *et al.* Role of mutator alleles in adaptive evolution. *Nature* **387**, 700–702 (1997).
38. Lieberman, T. D. *et al.* Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat. Genet.* **46**, 82–87 (2014).

39. Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33 (1994).
40. Moxon, R., Bayliss, C. & Hood, D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* **40**, 307–333 (2006).
41. Alamro, M. *et al.* Phase variation mediates reductions in expression of surface proteins during persistent meningococcal carriage. *Infect. Immun.* **82**, 2472–2484 (2014).
42. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
43. Didelot, X., Méric, G., Falush, D. & Darling, A. E. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* **13**, 256 (2012).
44. Croucher, N. J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science* **331**, 430–434 (2011).
45. Didelot, X., Achtman, M., Parkhill, J., Thomson, N. R. & Falush, D. A bimodal pattern of relatedness between the *Salmonella* Paratyphi A and Typhi genomes: convergence or divergence by homologous recombination? *Genome Res.* **17**, 61–68 (2007).
46. Sheppard, S. K. *et al.* Progressive genome-wide introgression in agricultural *Campylobacter coli*. *Mol. Ecol.* **22**, 1051–1064 (2013).
47. Cao, Q. *et al.* Progressive genomic convergence of two *Helicobacter pylori* strains during mixed infection of a patient with chronic gastritis. *Gut* **64**, 554–561 (2015).
48. Andersson, J. O. & Andersson, S. G. Insights into the evolutionary process of genome degradation. *Curr. Opin. Genet. Dev.* **9**, 664–671 (1999).
49. Rau, M. H., Marvig, R. L., Ehrlich, G. D., Molin, S. & Jelsbak, L. Deletion and acquisition of genomic content during early stage adaptation of *Pseudomonas aeruginosa* to a human host environment. *Environ. Microbiol.* **14**, 2200–2211 (2012).
50. Rankin, D. J., Rocha, E. P. C. & Brown, S. P. What traits are carried on mobile genetic elements, and why? *Hered. (Edinb.).* **106**, 1–10 (2011).
51. Dingle, K. E. *et al.* Evolutionary history of the *Clostridium difficile* pathogenicity locus. *Genome Biol. Evol.* **6**, 36–52 (2014).
52. Stanczak-Mrozek, K. I. *et al.* Within-host diversity of MRSA antimicrobial resistances. *J. Antimicrob. Chemother.* **70**, 2191–2198 (2015).
53. Charlesworth, B. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
54. Kuo, C., Moran, N. & Ochman, H. The consequences of genetic drift for bacterial genome complexity. *Genome Res.* **19**, 1450–1454 (2009).
55. Golubchik, T. *et al.* Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage. *PLoS ONE* **8**, e61319 (2013).
56. Rocha, E. P. C. *et al.* Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J. Theor. Biol.* **239**, 226–235 (2006).
57. Price, E. P. *et al.* Within-host evolution of *Burkholderia pseudomallei* over a twelve-year chronic carriage infection. *mBio* **4**, e00388-13 (2013).
58. Marvig, R. L., Sommer, L. M., Molin, S. & Johansen, H. K. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat. Genet.* **47**, 57–65 (2015).
59. Lieberman, T. D. *et al.* Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat. Genet.* **43**, 1275–1280 (2011).
A study of an outbreak of *B. dolosa* in patients with cystic fibrosis, which revealed evidence for adaptation to the host in the form of convergent evolution across several patients of genes with functions in antibiotic resistance and bacterial membrane composition.
60. Krebes, J., Didelot, X., Kennemann, L. & Suerbaum, S. Bidirectional genomic exchange between *Helicobacter pylori* strains from a family in Coventry, United Kingdom. *Int. J. Med. Microbiol.* **304**, 1135–1146 (2014).
61. Palmer, A. C. & Kishony, R. Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat. Rev. Genet.* **14**, 243–248 (2013).
62. Gardy, J. L. *et al.* Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N. Engl. J. Med.* **364**, 730–739 (2011).
63. Snitkin, E. S. *et al.* Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci. Transl. Med.* **4**, 148ra116 (2012).
64. Croucher, N. J. & Didelot, X. The application of genomics to tracing bacterial pathogen transmission. *Curr. Opin. Microbiol.* **23**, 62–67 (2015).
65. Maddison, W. P. & Knowles, L. L. Inferring phylogeny despite incomplete lineage sorting. *Syst. Biol.* **55**, 21–30 (2006).
66. Ypma, R., van Ballegooijen, W. M. & Wallinga, J. Relating phylogenetic trees to transmission trees of infectious disease outbreaks. *Genetics* **195**, 1055–1062 (2013).
67. Didelot, X., Gardy, J. & Colijn, C. Bayesian inference of infectious disease transmission from whole genome sequence data. *Mol. Biol. Evol.* **31**, 1869–1879 (2014).
68. Ou, C. Y. *et al.* Molecular epidemiology of HIV transmission in a dental practice. *Science* **256**, 1165–1171 (1992).
69. Metzker, M. L. *et al.* Molecular evidence of HIV-1 transmission in a criminal case. *Proc. Natl Acad. Sci. USA* **99**, 14292–14297 (2002).
70. Harris, S. R. *et al.* Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. *Lancet Infect. Dis.* **13**, 130–136 (2013).
71. Tong, S. Y. C. *et al.* Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. *Genome Res.* **25**, 111–118 (2015).
In this study, all patients from two intensive care units in a hospital in Thailand were repeatedly screened for carriage of MRSA over a period of three months. Whole-genome sequencing of patients and staff enabled the reconstruction of transmission events within and between wards.
72. Paterson, G. K. *et al.* Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nat. Commun.* **6**, 6560 (2015).
73. Okoro, C. K. *et al.* High-resolution single nucleotide polymorphism analysis distinguishes recrudescence and reinfection in recurrent invasive nontyphoidal *Salmonella* Typhimurium disease. *Clin. Infect. Dis.* **54**, 955–963 (2012).
74. Eyre, D. W. *et al.* Whole-genome sequencing demonstrates that fidaxomicin is superior to vancomycin for preventing reinfection and relapse of infection with *Clostridium difficile*. *J. Infect. Dis.* **209**, 1446–1451 (2014).
75. Mac Aogáin, M. *et al.* Whole-genome sequencing improves discrimination of relapse from reinfection and identifies transmission events among patients with recurrent *Clostridium difficile* infections. *J. Hosp. Infect.* **90**, 108–116 (2015).
76. Bryant, J. M. *et al.* Whole-genome sequencing to establish relapse or re-infection with *Mycobacterium tuberculosis*: a retrospective observational study. *Lancet Respir. Med.* **1**, 786–792 (2013).
77. Guerra-Assunção, J. A. *et al.* Relapse or reinfection with tuberculosis: a whole genome sequencing approach in a large population-based cohort with high HIV prevalence and active follow-up. *J. Infect. Dis.* **211**, 1154–1163 (2015).
78. Guerra-Assunção, J. *et al.* Large scale population-based whole genome sequencing of *Mycobacterium tuberculosis* provides insights into transmission in a high prevalence area. *eLife* **4**, e05166 (2015).
79. World Health Organization. *Antimicrobial resistance global report on surveillance 2014.* (WHO, 2014).
80. Elena, S. F. & Lenski, R. E. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* **4**, 457–469 (2003).
81. Mwangi, M. M. *et al.* Tracking the *in vivo* evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **104**, 9451–9456 (2007).
82. Eldholm, V. *et al.* Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biol.* **15**, 490 (2014).
The first documented case in which an XDR strain of *M. tuberculosis* evolved from a drug susceptible ancestor within a single patient. Resistance for most drugs evolved several times, with a single lineage ultimately prevailing.
83. Koch, A. & Wilkinson, R. J. The road to drug resistance in *Mycobacterium tuberculosis*. *Genome Biol.* **15**, 520 (2014).
84. Howden, B. P. *et al.* Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR. *PLoS Pathog.* **7**, e1002359 (2011).
85. Gao, W. *et al.* Two novel point mutations in clinical *Staphylococcus aureus* reduce linezolid susceptibility and switch on the stringent response to promote persistent infection. *PLoS Pathog.* **6**, e1000944 (2010).
86. Delaunē, A. *et al.* The WalKR system controls major staphylococcal virulence genes and is involved in triggering the host inflammatory response. *Infect. Immun.* **80**, 3438–3453 (2012).
87. Farhat, M. R. *et al.* Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* **45**, 1183–1189 (2013).
88. Comas, I. *et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110 (2011).
Using several approaches, this study identified high-confidence compensatory mutations associated with rifampicin resistance-conferring mutations in *M. tuberculosis*. The authors noted an enrichment of these mutations in *rpoA* and *rpoC*, which encode subunits of RNA polymerase.
89. Peleg, A. Y. *et al.* Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of *Staphylococcus aureus*. *PLoS ONE* **7**, e28316 (2012).
90. Arias, C. a *et al.* Genetic basis for *in vivo* daptomycin resistance in enterococci. *N. Engl. J. Med.* **365**, 892–900 (2011).
91. Sydenham, T. V., Sōki, J., Hasman, H., Wang, M. & Justesen, U. S. Identification of antimicrobial resistance genes in multidrug-resistant clinical *Bacteroides fragilis* isolates by whole genome shotgun sequencing. *Anaerobe* **31**, 59–64 (2014).
92. Cannatelli, A. *et al. In vivo* evolution to colistin resistance by PmrB sensor kinase mutation in KPC-producing *Klebsiella pneumoniae* is associated with low-dosage colistin treatment. *Antimicrob. Agents Chemother.* **58**, 4399–4403 (2014).
93. Ba, X. *et al.* Novel mutations in penicillin-binding protein genes in clinical *Staphylococcus aureus* isolates that are methicillin resistant on susceptibility testing, but lack the mec gene. *J. Antimicrob. Chemother.* **69**, 594–597 (2014).
94. Saunders, N. J. *et al.* Deep resequencing of serial sputum isolates of *Mycobacterium tuberculosis* during therapeutic failure due to poor compliance reveals stepwise mutation of key resistance genes on an otherwise stable genetic background. *J. Infect.* **62**, 212–217 (2011).
95. Dordel, J. *et al.* Novel determinants of antibiotic resistance: identification of mutated loci in highly methicillin-resistant subpopulations of methicillin-resistant *Staphylococcus aureus*. *mBio* **5**, e01000 (2014).
This study reported that most MRSA populations exhibit heteroresistance; the majority of isolates are methicillin sensitive, but low-frequency mutants possess several-hundred-fold higher resistance. This heteroresistance enables rapid population adaptation upon antibiotic exposure, while avoiding constitutive expression of resistance genes.
96. Marvig, R. L., Johansen, H. K., Molin, S. & Jelsbak, L. Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet.* **9**, e1003741 (2013).
An evolutionary analysis of the *P. aeruginosa* DK2 lineage over 38 years identified pathoadaptive mutations — in genes relating to antibiotic resistance, the cell envelope and regulatory functions — occurring independently in several patients.
97. Markussen, T. *et al.* Environmental heterogeneity drives within-host diversification and evolution of *Pseudomonas aeruginosa*. *mBio* **5**, e01592-14 (2014).
This study details the investigation of a *P. aeruginosa* DK1 infection that had persisted for 32 years, which showed diversification and co-existence of sublineages with distinct functional and genomic signatures, and different rates of evolution. These sublineages may occupy different niches within the airways of patients with cystic fibrosis.
98. Wong, A. & Kassen, R. Parallel evolution and local differentiation in quinolone resistance in *Pseudomonas aeruginosa*. *Microbiology* **157**, 937–944 (2011).

99. Yang, L. *et al.* Evolutionary dynamics of bacteria in a human host environment. *Proc. Natl Acad. Sci. USA* **108**, 7481–7486 (2011).
100. Kim, S., Lieberman, T. D. & Kishony, R. Alternating antibiotic treatments constrain evolutionary paths to multidrug resistance. *Proc. Natl Acad. Sci. USA* **111**, 14494–14499 (2014).
101. van Hal, S. J. *et al. In vivo* evolution of antimicrobial resistance in a series of *Staphylococcus aureus* patient isolates: the entire picture or a cautionary tale? *J. Antimicrob. Chemother.* **69**, 363–367 (2014).
102. Sun, G. *et al.* Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J. Infect. Dis.* **206**, 1724–1733 (2012).
103. Morand, B. & Mühlemann, K. Heteroresistance to penicillin in *Streptococcus pneumoniae*. *Proc. Natl Acad. Sci. USA* **104**, 14098–14103 (2007).
104. Ford, C. B. *et al. Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat. Genet.* **45**, 784–790 (2013).
105. Barrick, J. E. *et al.* Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247 (2009).
106. Feliziani, S. *et al.* Coexistence and within-host evolution of diversified lineages of hypermutable *Pseudomonas aeruginosa* in long-term cystic fibrosis infections. *PLoS Genet.* **10**, e1004651 (2014).
107. Montanari, S. *et al.* Biological cost of hypermutation in *Pseudomonas aeruginosa* strains from patients with cystic fibrosis. *Microbiology* **153**, 1445–1454 (2007).
108. Marvig, R. L. *et al.* Within-host evolution of *Pseudomonas aeruginosa* reveals adaptation toward iron acquisition from hemoglobin. *mBio* **5**, e00966-14 (2014).
109. Allen, R. C., Popat, R., Diggle, S. P. & Brown, S. P. Targeting virulence: can we make evolution-proof drugs? *Nat. Rev. Microbiol.* **12**, 300–308 (2014).
110. Weinert, L. A. *et al.* Genomic signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*. *Nat. Commun.* **6**, 6740 (2015).
111. Andersen, S. B., Marvig, R. L., Molin, S., Krogh Johansen, H. & Griffin, A. S. Long-term social dynamics drive loss of function in pathogenic bacteria. *Proc. Natl Acad. Sci. USA* **112**, 10756–10761 (2015).
112. Croucher, N. J. *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat. Genet.* **45**, 656–663 (2013).
113. Damkiær, S., Yang, L., Molin, S. & Jelsbak, L. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proc. Natl Acad. Sci. USA* **110**, 7766–7771 (2013).
114. von Eiff, C., Becker, K., Machka, K., Stammer, H. & Peters, G. Nasal carriage as a source of *Staphylococcus aureus* bacteremia. Study group. *N. Engl. J. Med.* **344**, 11–16 (2001).
115. Yang, J., Tauschek, M. & Robins-Browne, R. M. Control of bacterial virulence by AraC-like regulators that respond to chemical signals. *Trends Microbiol.* **19**, 128–135 (2011).
116. Fantappiè, L., Scarlato, V. & Delany, I. Identification of the *in vitro* target of an iron-responsive AraC-like protein from *Neisseria meningitidis* that is in a regulatory cascade with Fur. *Microbiology* **157**, 2235–2247 (2011).
117. Young, B. C. & Wilson, D. J. On the evolution of virulence during *Staphylococcus aureus* nasal carriage. *Virulence* **3**, 454–456 (2012).
118. Smith, E. E. *et al.* Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl Acad. Sci. USA* **103**, 8487–8492 (2006).
119. Kodaman, N. *et al.* Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proc. Natl Acad. Sci. USA* **111**, 1455–1460 (2014).
120. Campbell, D. I. *et al.* The African enigma: low prevalence of gastric atrophy, high prevalence of chronic inflammation in West African adults and children. *Helicobacter* **6**, 263–267 (2001).
121. Anderson, R. M. & May, R. M. Coevolution of hosts and parasites. *Parasitology* **85**, 411–426 (1982).
122. Ewald, P. W. Host-parasite relations, vectors, and the evolution of disease severity. *Annu. Rev. Ecol. Syst.* **14**, 465–485 (1983).
123. Alizon, S., Hurford, A., Mideo, N. & Van Baalen, M. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. Evol. Biol.* **22**, 245–259 (2009).
124. Suerbaum, S. & Josenhans, C. *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat. Rev. Microbiol.* **5**, 441–452 (2007).
125. Schwarz, S. *et al.* Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathog.* **4**, e1000180 (2008).
126. Agnew, P. & Koella, J. C. Virulence, parasite mode of transmission, and host fluctuating asymmetry. *Proc. Biol. Sci.* **264**, 9–15 (1997).
127. Zdziarski, J. *et al.* Host imprints on bacterial genomes-rapid, divergent evolution in individual patients. *PLoS Pathog.* **6**, 95–96 (2010).
128. Klemm, P., Roos, V., Ulett, G. C., Schembri, M. A. & Svanborg, C. Molecular characterization of the *Escherichia coli* asymptomatic bacteriuria strain 83972: the taming of a pathogen. *Infect. Immun.* **74**, 781–785 (2006).
129. Toprak, E. *et al.* Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44**, 101–105 (2011).
130. Espedido, B. A. *et al.* Whole genome sequence analysis of the first Australian OXA-48-producing outbreak-associated *Klebsiella pneumoniae* isolates: the resistome and *in vivo* evolution. *PLoS ONE* **8**, e59920 (2013).
131. Whitlock, M. C. Fixation probability and time in subdivided populations. *Genetics* **164**, 767–779 (2003).
132. Gill, W. P. *et al.* A replication clock for *Mycobacterium tuberculosis*. *Nat. Med.* **15**, 211–214 (2009).
133. Gordon, N. C. *et al.* Prediction of *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J. Clin. Microbiol.* **52**, 1182–1191 (2014).
134. Stoesser, N. *et al.* Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. *J. Antimicrob. Chemother.* **68**, 2234–2244 (2013).
135. Eyre, D. W. *et al.* Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in *Clostridium difficile* transmission. *PLoS Comput. Biol.* **9**, e1003059 (2013).
136. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
137. Eyre, D. W. *et al.* A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open* **2**, e001124 (2012).
138. Reuter, S. *et al.* Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. *JAMA Intern. Med.* **173**, 1397–1404 (2013).
139. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and variants calling using mapping quality scores. *Genome Res.* **18**, 1851–1858 (2008).
140. Stoesser, N. *et al.* Genome sequencing of an extended series of NDM-producing *Klebsiella pneumoniae* neonatal infections in a Nepali hospital characterizes the extent of community versus hospital-associated transmission in an endemic setting. *Antimicrob. Agents Chemother.* **58**, 7347–7357 (2014).
141. Nagarajan, N. & Pop, M. Sequence assembly demystified. *Nat. Rev. Genet.* **14**, 157–167 (2013).
142. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).
143. Foster, T. J., Geoghegan, J. A., Ganesh, V. K. & Höök, M. Adhesion, invasion and evasion: the many functions of the surface proteins of *Staphylococcus aureus*. *Nat. Rev. Microbiol.* **12**, 49–62 (2013).
144. Walker, T. M. *et al.* Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007–2012, with whole pathogen genome sequences: an observational study. *Lancet Respir. Med.* **2**, 285–292 (2014).
145. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
146. Didelot, X. & Wilson, D. J. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* **11**, e1004041 (2015).
147. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).

Acknowledgements

X.D. is funded by the UK Biotechnology and Biological Sciences Research Council (BBSRC; grant BB/L023458/1) and the UK National Institute for Health Research (NIHR) Health Protection Research Unit on Modelling Methodology (grant HPRU-2012-10080). T.E.P. and D.W.C. are NIHR senior investigators. D.J.W. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (grant 101237/Z/13/Z). This study was supported by the Health Innovation Challenge Fund, a parallel funding partnership between the Wellcome Trust (grant WT098615) and the UK Department of Health (grant HICF-T5-358), the NIHR Health Protection Research Unit in Healthcare Associated Infection and Antimicrobial Resistance (grant HPRU-2012-10041) and the NIHR Oxford Biomedical Research Centre. The views expressed in this publication are those of the authors and not necessarily those of the funders.

Competing interests statement

The authors declare no competing interests.