

## Research



**Cite this article:** Gibson B, Wilson DJ, Feil E, Eyre-Walker A. 2018 The distribution of bacterial doubling times in the wild. *Proc. R. Soc. B* **285**: 20180789. <http://dx.doi.org/10.1098/rspb.2018.0789>

Received: 9 April 2018

Accepted: 18 May 2018

**Subject Category:**

Genetics and genomics

**Subject Areas:**

microbiology, genetics, evolution

**Keywords:**

generation time, mutation rates, bacteria

**Author for correspondence:**

Adam Eyre-Walker

e-mail: [a.c.eyre-walker@sussex.ac.uk](mailto:a.c.eyre-walker@sussex.ac.uk)

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.4114037>.

# The distribution of bacterial doubling times in the wild

Beth Gibson<sup>1</sup>, Daniel J. Wilson<sup>2</sup>, Edward Feil<sup>3</sup> and Adam Eyre-Walker<sup>1</sup>

<sup>1</sup>School of Life Sciences, University of Sussex, Brighton BN1 9QG, UK

<sup>2</sup>Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

<sup>3</sup>The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, UK

BG, 0000-0001-7984-6188; DJW, 0000-0002-0940-3311; EF, 0000-0003-1446-6744; AE-W, 0000-0001-5527-8729

Generation time varies widely across organisms and is an important factor in the life cycle, life history and evolution of organisms. Although the doubling time (DT) has been estimated for many bacteria in the laboratory, it is nearly impossible to directly measure it in the natural environment. However, an estimate can be obtained by measuring the rate at which bacteria accumulate mutations per year in the wild and the rate at which they mutate per generation in the laboratory. If we assume the mutation rate per generation is the same in the wild and in the laboratory, and that all mutations in the wild are neutral, an assumption that we show is not very important, then an estimate of the DT can be obtained by dividing the latter by the former. We estimate the DT for five species of bacteria for which we have both an accumulation and a mutation rate estimate. We also infer the distribution of DTs across all bacteria from the distribution of the accumulation and mutation rates. Both analyses suggest that DTs for bacteria in the wild are substantially greater than those in the laboratory, that they vary by orders of magnitude between different species of bacteria and that a substantial fraction of bacteria double very slowly in the wild.

## 1. Introduction

The bacterium *Escherichia coli* can divide every 20 min in the laboratory under aerobic, nutrient-rich conditions. But how often does it divide in its natural environment in the gut, under anaerobic conditions where it probably spends much of its time close to starvation? And what do we make of a bacterium, such as *Syntrophobacter fumaroxidans*, which only doubles in the laboratory every 140 h [1]. Does this reflect a slow doubling time (DT) in the wild, or our inability to provide the conditions under which it can replicate rapidly?

Estimating the generation time is difficult for most bacteria in their natural environment and very few estimates are available. The DT for intestinal bacteria has been estimated in several mammals by assaying the quantity of bacteria in the gut and faeces. Assuming no cell death Gibbons & Kapsimalis [2] estimate the DT for all bacteria in the gut to be 48, 17 and 5.8 h in hamster, guinea pig and mouse, respectively. More recently Yang *et al.* [3] have shown that the DT of *Pseudomonas aeruginosa* is correlated to cellular ribosomal content *in vitro* and have used this to estimate the DT *in vivo* in a cystic fibrosis (CF) patient to be between 1.9 and 2.4 h.

Although there are very few estimates of the generation time in bacteria, this quantity is important for understanding bacterial population dynamics. Here we use an indirect method to estimate the DT that uses two sources of information. First, we can measure the rate at which a bacterial species accumulates mutations in its natural environment through time using temporarily sampled data [4], or concurrent samples from a population with a known date of origin. We refer to this quantity as the accumulation rate, to differentiate

**Table 1.** Doubling time estimates (hours) for those species for which we have both an estimate of the accumulation and mutation rate. Accumulation rate (AR) references—(1) [5]; (2) [6,7]; (3) [8–12]; (4) [13–23]; (5) [8,24]. Mutation rate (MR) references—(6) [25]; (7) [26]; (8) [27]; (9) [28]; (10) [29].

species	accumulation rate per site per year	mutation rate per site per generation	DT (h) (s.e.)	laboratory DT (h)	ratio	AR ref.	MR ref.
<i>Escherichia coli</i>	$1.44 \times 10^{-7}$	$2.54 \times 10^{-10}$	15 (7.7)	0.33	45	1	6
<i>Pseudomonas aeruginosa</i>	$3.03 \times 10^{-7}$	$7.92 \times 10^{-11}$	2.3 (0.77)	0.5	4.6	2	7
<i>Salmonella enterica</i>	$2.50 \times 10^{-7}$	$7.00 \times 10^{-10}$	25 (7.9)	0.5	50	3	8
<i>Staphylococcus aureus</i>	$2.05 \times 10^{-6}$	$4.38 \times 10^{-10}$	1.87 (0.98)	0.4	4.7	4	9
<i>Vibrio cholerae</i>	$8.30 \times 10^{-7}$	$1.07 \times 10^{-10}$	1.1 (0.26)	0.66	1.7	5	10

it from the mutation rate, and the substitution rate, the rate at which mutations spread through a species to fixation. If we assume that all mutations in the wild are neutral, an assumption that we show to be relatively unimportant, then the accumulation rate is an estimate of the mutation rate per year,  $u_y$ . Second, we can estimate the rate of mutation per generation,  $u_g$ , in the laboratory using a mutation accumulation experiment and whole-genome sequencing, or through fluctuation tests. If we assume that the mutation rate per generation is the same in the wild and in the laboratory, an assumption we discuss further below, then if we divide the accumulation rate per year in the wild by the mutation rate per generation in the laboratory, we can estimate the number of generations that the bacterium goes through in the wild and hence the DT ( $DT = 8760 \times u_g/u_y$  where 8760 is the number of hours per year).

## 2. Results

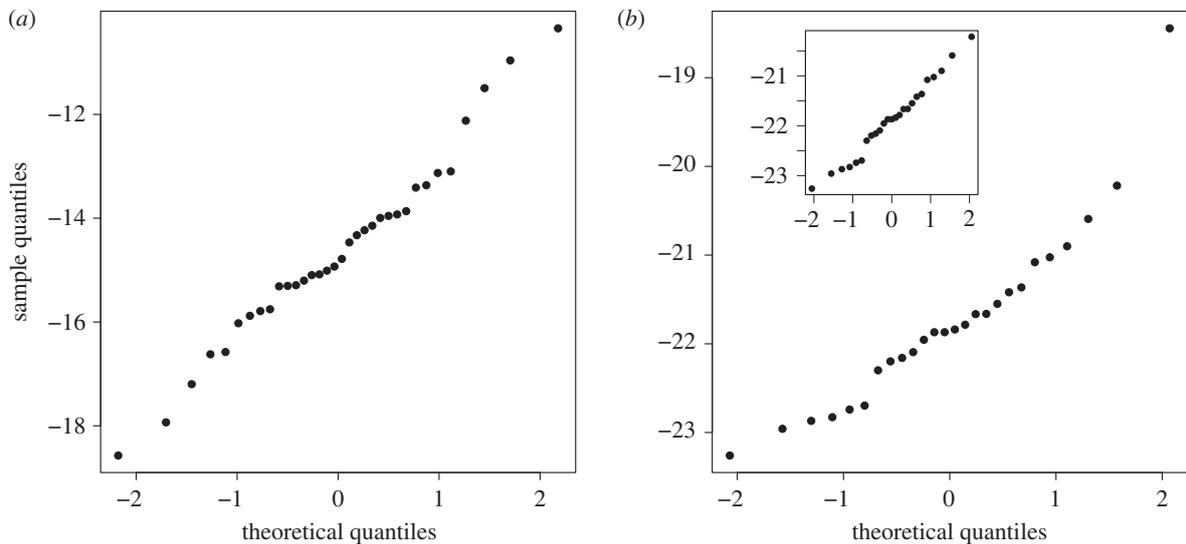
The accumulation rate in the wild and the mutation rate in the laboratory have been estimated for 34 and 26 bacterial species, respectively (electronic supplementary material, tables S1, S2); we only consider mutation rate estimates from mutation accumulation experiments, because estimates from fluctuation tests are subject to substantial sampling error and unknown bias, and we exclude estimates from hypermutable strains. For five species, *E. coli*, *P. aeruginosa*, *Salmonella enterica*, *Staphylococcus aureus* and *Vibrio cholerae*, we have both an accumulation and a mutation rate estimate and hence can estimate the DT. Among these five species we find our DT estimates vary from 1.1 h in *V. cholerae* to 25 h in *S. enterica* (table 1). In all cases the estimated DT in the wild is greater than that of the bacterium in the laboratory. For example, *E. coli* can double every 20 min in the laboratory but we estimate that it only doubles every 15 h in the wild.

In theory, it might be possible to estimate the DT in those bacteria for which we have either an accumulation or mutation rate estimate, but not both, by finding factors that correlate with either rate and using those factors to predict the rates. Unfortunately, we have been unable to find any factor that correlates sufficiently well to be usefully predictive. It has been suggested that the mutation rate is correlated to genome size in microbes [30] but, the current

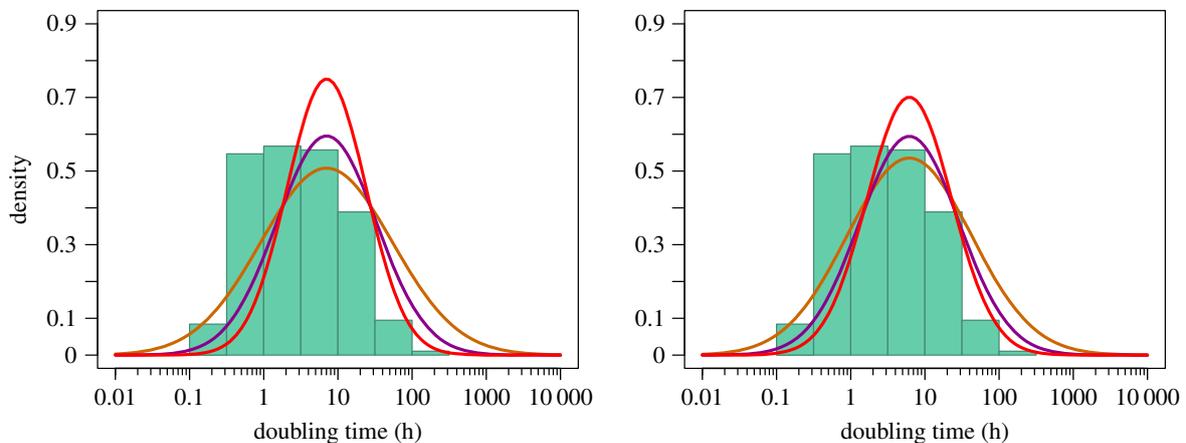
evidence for this correlation is very weak, and depends upon the estimate from *Mesoplasma florum* ( $r = -0.68$ ,  $p < 0.001$  with *M. florum* and  $r = -0.39$ ,  $p = 0.053$  without *M. florum*) (electronic supplementary material, figure S1) [31]. However, we can use the accumulation and mutation rate estimates to estimate the distribution of DTs across bacteria if we assume that there is no phylogenetic non-independence in the mutation and accumulation data, an assumption we address below. We can estimate the distribution of DTs by fitting distributions to the accumulation and mutation rate data, using maximum likelihood, and then dividing one distribution by the other. We assume that both variables are lognormally distributed, an assumption that is supported by  $Q-Q$  plots with the exception of the mutation rate per generation in *M. florum*, which is a clear outlier (figure 1). We repeated all our analyses with and without *M. florum*.

If the accumulation and mutation rate data are lognormally distributed then the distribution of DT is also lognormally distributed with a mean of  $\log_e(8760) + m_g - m_y$  and a variance of  $v_g + v_y - 2\text{Cov}(g,y)$ , where 8760 is the number of hours per year and  $m_g$ ,  $m_y$ ,  $v_g$  and  $v_y$  are the mean and variance of the lognormal distributions fitted to the mutation (subscript  $g$ ) and accumulation (subscript  $y$ ) rates.  $\text{Cov}(g,y)$  is the covariance between the accumulation and mutation rates. We might expect that species with higher mutation rates also have higher accumulation rates, because the accumulation rate is expected to depend on the mutation rate, but the correlation between the two will depend upon how variable the DT and other factors, such as the strength of selection, are between bacteria. The observed correlation between the log accumulation rate and log mutation rate is 0.077, but there are only five data points, so the 95% confidence intervals on this estimate encompass almost all possible values ( $-0.86$  to  $0.89$ ). We explore different levels of the correlation between the accumulation and mutation rates; it should be noted that  $\text{Cov}(g,y)$  can be expressed as  $\text{Sqrt}(v_g v_y) \text{Corr}(g,y)$  where  $\text{Corr}(g,y)$  is the correlation between the two variables.

The distribution of DTs in the wild inferred using our method is shown in figure 2. We infer the median DT to be 7.04 h, but there is considerable spread around this even when the accumulation and mutation rates are strongly correlated (figure 2a); as the correlation increases so the variance in DTs decreases, but the median remains unaffected. The



**Figure 1.** Normal  $Q-Q$  plots for the log of (a) accumulation and (b) mutation rate data. The main plot in (b) includes all 26 mutation rate estimates and the insert excludes *Mesoplasma florum* estimate.



**Figure 2.** The distribution of DTs among bacteria inferred assuming different levels of correlation between the accumulation and mutation rates—orange  $r = 0$ , purple  $r = 0.5$  and red  $r = 0.75$ . We also show the distribution of laboratory DTs (green histogram) from a compilation of over 200 species made by Vieira-Silva & Rocha [32]. In (a) we include all mutation rate estimates and in (b) we exclude the mutation rate estimate for *Mesoplasma florum*. (Online version in colour.)

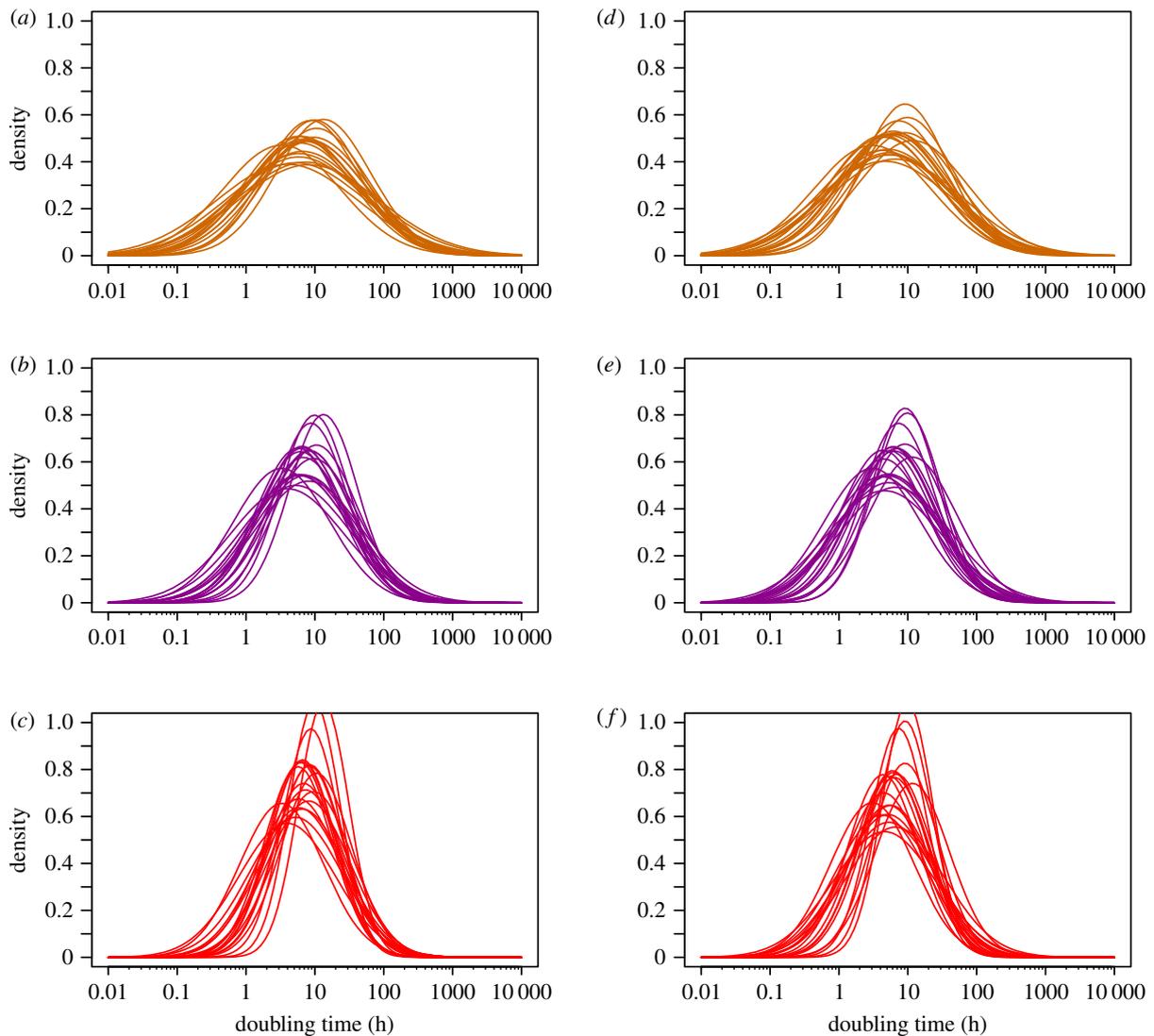
analysis suggests that most bacteria have DTs of between 1 and 100 h but there are substantial numbers with DTs beyond these limits. For example, even if we assume that the correlation between the accumulation and mutation rate is 0.5 we infer that 10% of bacteria have a DT of faster than 1 h in the wild and 4.2% have a DT slower than 100 h in the wild. If we remove the *M. florum* mutation rate estimate from the analysis the median doubling is slightly lower at 6.16 h, but there is almost as much variation as when this bacterium is included; at a correlation is 0.5 we infer that 12% of bacteria have a DT faster than 1 h in the wild and 3.5% have a DT slower than 100 h.

To investigate how robust these conclusions are to statistical sampling, we bootstrapped the accumulation and mutation rate estimates, refit the lognormal distributions and re-inferred the distribution of DT. The 95% confidence intervals for the median are quite broad at 3.4 to 14.2 h (3.1 to 11.3 h if we exclude *M. florum*). However, all bootstrapped distributions show substantial variation in the DT with a substantial fraction of bacteria with long DTs and also some with very short DTs (figure 3).

We have assumed that there is no phylogenetic inertia within the accumulation and mutation rate estimates. To

test whether this is the case we constructed a phylogenetic tree using 16S rRNA sequences and applied the tests of Pagel [33] and Blomberg *et al.* [34]. Both the accumulation and mutation rate data show some evidence of phylogenetic signal. For the accumulation data, Pagel's  $\lambda = 0.68$  ( $p = 0.001$ ) and Blomberg *et al.*'s  $K = 0.0005$  ( $p = 0.35$ ); and for the mutation rate data Pagel's  $\lambda = 0.88$  ( $p = 0.026$ ) and Blomberg *et al.*'s  $K = 0.5$  ( $p = 0.009$ ). We also find some evidence that the data depart from a Brownian motion model using Pagel's test (i.e.  $\lambda$  is significantly less than one) for the accumulation data ( $p < 0.001$ ) but not the mutation rate data ( $p = 0.094$ ); i.e. the accumulation rates are more different than we would expect from their phylogeny and a Brownian motion model. A visual inspection of the data suggests that the phylogenetic signal is largely contributed by species that are closely related, rather than deeper phylogenetic levels (figure 4a,b) and species for which we have accumulation and mutation rate estimates are interspersed with one another on the phylogenetic tree (figure 4c). It, therefore, seems unlikely that phylogenetic inertia will influence our results.

It is of interest to compare the distribution of DTs in the wild to the distribution of laboratory DTs (figure 1). The



**Figure 3.** DT distributions inferred by bootstrapping the accumulation and mutation rate data and refitting the lognormal distributions to both datasets. Each plot shows 20 bootstrap DT distributions assuming different levels of correlation between the accumulation and mutation rates—orange  $r = 0$ , purple  $r = 0.5$  and red  $r = 0.75$ . (a–c) Include all mutation rate estimates and (d–f) show the analysis after removal of the *Mesoplasma florum* mutation rate estimate. (Online version in colour.)

distributions are different in two respects. First, the median laboratory DT of 3 h is less than half the median wild DT of 7.04 h (6.16 h without *M. florum*); the two are significantly different ( $p = 0.012$  with *M. florum* and  $p = 0.016$  without *M. florum*, inferred by bootstrapping each dataset and recalculating the medians). Second, many more bacteria are inferred to have long DTs in the wild than in the laboratory.

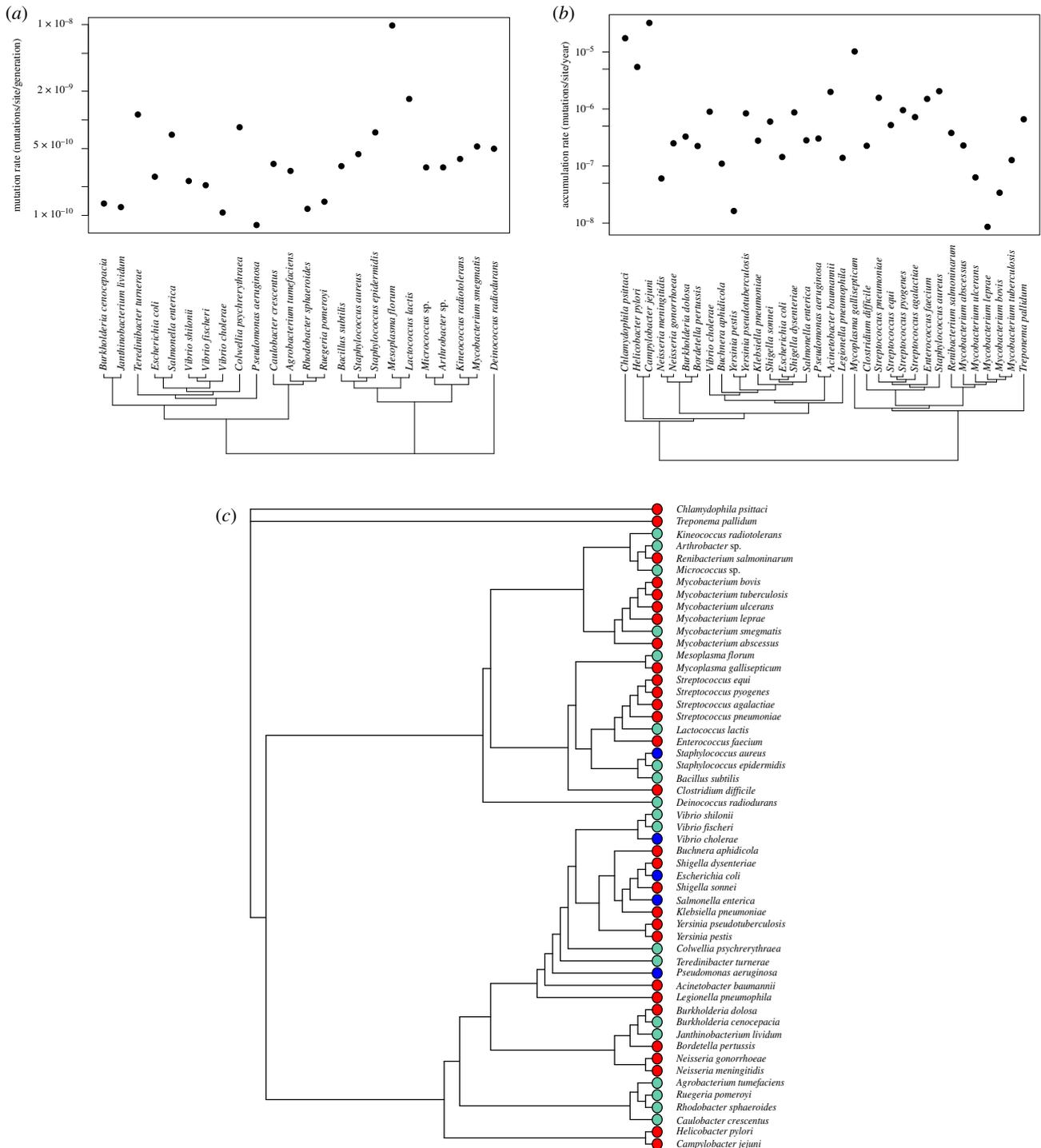
### 3. Discussion

The DT of most bacteria in their natural environment is not known. We have used estimates of the rate at which bacteria accumulate mutations in their natural environment and estimates of the rate at which they mutate in the laboratory to estimate the DT for several bacteria and infer the distribution of DTs across bacteria. We estimate that DT are generally longer in the wild than in the laboratory, but critically we also infer that DTs vary by several orders of magnitude between bacterial species and that many bacteria have very slow DTs in their natural environment.

The method, by which we have inferred the DT in the wild, makes three important assumptions. We assume that the

mutation rate per generation is the same in the laboratory and in the wild. However, it seems likely that bacteria in the wild will have a higher mutation rate per generation than those in the laboratory for two reasons. First, bacteria in the wild are likely to be stressed and this can be expected to elevate the mutation rate [35–39]. Second, if we assume that DTs are longer in the wild than the laboratory then we expect the mutation rate per generation to be higher in the wild than in the laboratory because some mutational processes do not depend upon DNA replication. The relative contribution of replication dependent and independent mutational mechanisms to the overall mutation rate is unknown. Rates of substitution are higher in Firmicutes that do not undergo sporulation suggesting that replication is a source of mutations in this group of bacteria [40]. However, rates of mutation accumulation seem to be similar in latent versus active infections of *Mycobacterium tuberculosis*, suggesting that replication independent mutations might dominate in this bacterium [41,42].

The second major assumption is that the rate at which mutations accumulate in the wild is equal to the mutation rate per year; in effect, we are assuming that all mutations are effectively neutral, at least over the time frame in which they are assayed (or that some are inviable, but the same



**Figure 4.** (a) 16S rRNA phylogeny and mutation rate estimates for 24 species of bacteria (two species are excluded because of erroneous positioning on the phylogeny—see electronic supplementary material, figure S2A,B for details). (b) 16S rRNA phylogeny and accumulation rate estimates for 34 species of bacteria. (c) 16S rRNA phylogeny combining species for which we have an estimate of the mutation rate and/or accumulation rate. Coloured dots indicate which kind of information each species provides—red, accumulation rate; green, mutation rate and blue, both a mutation rate and an accumulation rate. (Online version in colour.)

proportion are inviable in the wild and the laboratory). In those accumulation rate studies, in which they have been studied separately, non-synonymous mutations accumulate more slowly than synonymous mutations; relative rates vary from 0.13 to 0.8, with a mean of 0.57 (electronic supplementary material, table S3). There is no correlation between the time frame over which the estimate was made and the ratio of non-synonymous and synonymous accumulation rates ( $r = 0.2$ ,  $p = 0.53$ ). We did not attempt to control for selection because the relative rates of synonymous and non-synonymous accumulation are only available for a few

species, and the relative rates vary between species. However, we can estimate the degree to which more selection against deleterious non-synonymous accumulations in the wild causes the DT to be underestimated as follows. The observed rate at which mutations accumulate in a bacterial lineage is

$$\mu_{\text{obs}} = \alpha \mu_{\text{true}} \delta_i + (1 - \alpha)(1 - \beta) \mu_{\text{true}} \delta_s + (1 - \alpha) \beta \mu_{\text{true}} \delta_n, \quad (3.1)$$

where  $\alpha$  is the proportion of the genome that is non-coding and  $\beta$  is the proportion of mutations in protein coding

sequence that are non-synonymous.  $\delta_x$  is the proportion of mutations of class  $x$  ( $i$  is intergenic,  $s$  is synonymous and  $n$  is non-synonymous) that are effectively neutral.  $\alpha$  and  $\beta$  are approximately 0.15 and 0.7, respectively, in our dataset. Although there is selection on synonymous codon use in many bacteria [43], selection appears to be weak [44] we, therefore, assume that  $\delta_s = 1$ . This implies, from the rate at which non-synonymous mutations accumulate relative to synonymous mutations, that  $\delta_n = 0.6$ . A recent analysis of intergenic regions in several species of bacteria has concluded that selection is weaker in intergenic regions than at non-synonymous sites; we, therefore, assume that  $\delta_i = 0.8$  [45]. Using these estimates suggests that selection leads us to underestimate the true mutation rate per year in the wild by approximately 27%; this in turn means we have over-estimated the DT by approximately 37%, a relatively small effect. To investigate how sensitive this estimate is to the parameters in equation 1, we varied each of them in turn (electronic supplementary material, table S4). We find that the observed mutation rate is most sensitive to selection on synonymous codon use, because if there is selection on synonymous codon use this also affects our estimates of selection at non-synonymous sites and in intergenic. For example, if selection on synonymous codon use depressed the synonymous accumulation rate by 0.5 this would lead to an underestimate of the mutation rate of 63%, which would in turn have led to a 2.7-fold over-estimate of the DT.

Finally, although each study attempted to remove single nucleotide polymorphisms (SNPs) that had arisen by recombination, it is possible that some are still present in the data. Recombinant SNPs can have two effects. First, if they have recombined from outside the clade, they inflate the accumulation rate estimate and hence lead to an underestimate of the DT. Second, if there is recombination within a clade, they affect the phylogeny and potentially lead to the root of the tree being estimated as younger than it should be. This will lead to an overestimate of the DT.

It is important to appreciate that our method estimates an average DT within a particular environment that the bacteria were sampled from. The bacterium may go through periods of quiescence interspersed with periods of growth.

Despite the assumptions we have made in our method, our estimate of the DT of *Pseudomonas aruginosa* of 2.3 h in a CF patient is very similar to that independently estimated using the ribosomal content of cells of between 1.9 and 2.4 h [3]. There is also independent evidence that there are some bacteria that divide slowly in their natural environment. The aphid symbiont *Buchnera aphidicola* is estimated to double every 175–292 h in its host [46,47], and *Mycobacterium leprae* doubles every 300–600 h on mouse footpads [48–50], not its natural environment, but one that is probably similar to the human skin. Furthermore, in a recent selection experiment, Avrani *et al.* [51] found that several *E. coli* populations, which were starved of resources, accumulated mutations in the core RNA polymerase gene. These mutations caused these strains to divide more slowly than unmutated strains when resources were plentiful. Interestingly these same mutations are found at high frequency in unculturable bacteria, suggesting that there is a class of slow growing bacteria in the environment that are adapted to starvation.

Korem *et al.* [52] have recently proposed a general method by which the DT can be potentially estimated. They note that

actively replicating bacterial cells have two or more copies of the chromosome near the origin of replication but only one copy near the terminus, if cell division occurs rapidly after the completion of DNA replication. Using next-generation sequencing, they show that it is possible to assay this signal and that the ratio of sequencing depth near the origin and terminus is correlated to bacterial growth rates *in vivo*. Brown *et al.* [53] have extended the method to bacteria without a reference genome and/or those without a known origin and terminus of replication. In principle, these measures of cells performing DNA replication could be used to estimate the DT of bacteria in the wild. However, it is unclear how or whether the methods can be calibrated. Both Korem *et al.* (2015) and Brown *et al.* (2016) find that their replication measures have a median of approximately 1.3 across bacteria in the human gut. However, a value of 1.3 translates into different relative and absolute values of the DT in the two studies. Brown *et al.* [53] show that their measure of replication, iRep, is highly correlated to Korem *et al.*'s [52] measure, PTR, for data from *Lactobacillus gasseri*; the equation relating the two statistics is  $iRep = -0.75 + 2 PTR$ . Hence, when  $PTR = 1.3$ ,  $iRep = 1.85$  and when  $iRep = 1.3$ ,  $PTR = 1.03$ . The two methods are not consistent. They also yield very different estimates for the absolute DT. Korem *et al.* [52] show that PTR is highly correlated to the growth rate of *E. coli* grown in a chemostat. If we assume that the relationship between PTR and growth rate is the same across bacteria *in vivo* and *in vitro*, then this implies that the median DT for the human microbiome is approximately 2.5 h. By contrast, Brown *et al.* [53] estimate the growth rate of *Klebsiella oxytoca* to be 19.7 h in a new-born baby using faecal counts and find that this population has an iRep value of approximately 1.77. This value is greater than the vast majority of bacteria in the human microbiome and bacteria in the Candidate Phyla Radiation, suggesting that most bacteria in these two communities replicate very slowly. These discrepancies between the two methods suggest that it may not be easy to calibrate the PTR and iRep methods to yield estimates of the DT across bacteria.

Finally, how should we interpret our results for the five focal species in the context of what is known of their ecology? *Vibrio cholerae* displays the shortest DT of 1.1 h. *Vibrio* species are ubiquitous in estuarine and marine environments [54]. They are known to have very short generation times in culture, the shortest being *Vibrio natriegens* of just 9.8 min [55]. In the wild they can exploit a wide range of carbon and energy sources, and as such have been termed 'opportunotrophs' [56]. Natural *Vibrio* communities do not grow at an accelerated rate continuously, but can exist for long periods in a semi-dormant state punctuated by rapid pulses of high growth rates [57], or blooms [58], when conditions are favourable. It has also been argued that the unusual division of *Vibrio* genomes into two chromosomes facilitates more rapid growth [59]. By pointing to a very short DT in *V. cholerae*, our analysis is, therefore, consistent with what is known of the ecology of this species.

*Staphylococcus aureus* is predominantly found on animals and humans and inhabits various body parts, including the skin and upper respiratory tract [60]. It can cause infection of the skin and soft tissue as well as bacteraemia [61]. *Staphylococcus aureus* exhibits a range of modes of growth, some of which may allow it to survive stress and antimicrobials while in its host. For instance, small subpopulations can adopt a slow-growing, quasi-dormant lifestyle, either in a multicellular

biofilm or as small colony variants (SCVs) or persister cells [62]. Our short DT of 1.8 h suggests this is not the typical state for *S. aureus* in the wild, which is not surprising considering the incidence of SCVs in clinical samples is fairly low, between 1 and 30% [63].

*Pseudomonas aeruginosa* can inhabit a wide variety of environments, including inside soil, water, plants and animals. Like our other focal species, it is an opportunistic pathogen and can also infect humans, especially those with compromised immune systems, such as patients with CF. In this context infection is chronic. Parallel evolution, the differential regulation of genes which allow it to evade the host immune system and resist antibiotic treatment during infection [64], and evidence of positive selection [65] suggests *P. aeruginosa* can adapt to the lungs of individuals with CF for its long-term survival. It is known to actively grow in sputum [3], where it uses the available nutrition which supports its growth to high population densities [66]. Its ability to adapt and actively grow in the CF sputum is consistent with its relatively short DT of 2.3 h, especially considering this is the environment in which the accumulation rate was measured and matches that estimated by Yang *et al.* [3].

*Escherichia coli* and *S. enterica* primarily reside in the lower intestine of humans and animals, but can also survive in the environment. Although *E. coli* is commonly recovered from environmental samples, it is not thought able to grow or survive for prolonged periods outside of the guts of warm blooded animals, except in tropical regions where conditions are more favourable [67], although some phylogenetically distinct strains appear to reproduce and survive well in the environment [68]. In contrast, *Salmonella* is also an enteric colonizer of cold-blooded animals, in particular reptiles, is better adapted than *E. coli* at surviving and growing in environmental niches. For example, *Salmonella* can survive and grow for at least a year in soil [69], whereas *E. coli* can survive for only a few days [70]. Although these secondary niches may play a greater role in *Salmonella* than in *E. coli*, it remains the case the growth rates in the environment will be much lower than those in a gut. Therefore, the increased tenacity of *Salmonella* in non-host environments compared to *E. coli* might help to explain the slower DT in this species.

In summary, the availability of accumulation and mutation rate estimates allows us to infer the DT for bacteria in the wild, and the distribution of wild DTs across bacterial species. These DT estimates are likely to be underestimates because the mutation rate per generation is expected to be higher in the wild than in the laboratory, and some mutations are not generated by DNA replication. Our analysis, therefore, suggests that DTs in the wild are typically longer than those in the laboratory, that they vary considerably between bacterial species and that a substantial proportion of species have very long DTs in the wild.

## 4. Material and methods

We compiled estimates of the accumulation and mutation rate of bacteria from the literature. We only used mutation rate estimates that came from a mutation accumulation experiment with whole-genome sequencing. If we had multiple estimates of the mutation rate we summed the number of mutations across the mutation accumulation experiments and divided this by the product of the genome size and the number of generations that were

assayed. We averaged the accumulation rate estimates where we had multiple estimates from the same species. We recalculated the accumulation rates in two cases in which the number of accumulated mutations had been divided by an incorrect number of years: *E. coli* [5] and *Helicobacter pylori* [71]. For *E. coli*, we reestimated the accumulation rate using BEAST by constructing sequences of the SNPs reported in the paper and the isolation dates. For *Helicobacter pylori* the 3-year and 16-year strains appear to form a clade to the exclusion of the 0-year strain because they share some differences from the 0-year strain. If the number of substitutions that have accumulated between the common ancestor of the 3-year and 16-year strain and each of the two strains are  $S_3$  and  $S_{16}$ , respectively, then the rate of accumulation can be estimated as  $(S_{16} - S_3)/(13 \text{ years} \times \text{genome size})$ . For the isolates NQ1707 and NQ4060 we have estimated the accumulation rate to be  $5 \times 10^{-6}$  and for NQ1671 and NQ4191  $5.9 \times 10^{-6}$ . We excluded some accumulation rate estimates for a variety of reasons. We only considered accumulation rates sampled over an historical time frame of at most 1500 years. Most of our estimates of the accumulation rate are for all sites in the genome, so we excluded cases in which only the synonymous accumulation rate was given. We also excluded accumulation rates from hypermutable strains. Accumulation and mutation rate estimates used in the analysis are given in electronic supplementary material, tables S1 and S2, respectively.

The estimate of the standard error associated with our estimate of the DT was obtained using the standard formula for the variance of a ratio:  $V(x/y) = (M(x)/M(y))^2(V(x)/M(x)^2 + V(y)/M(y)^2)$  where  $M$  and  $V$  are the mean and variance of  $x$  and  $y$ . The variance for the accumulation rate was either the variance between multiple estimates of the accumulation rate if they were available, or the variance associated with the estimate if there was only a single estimate. The variance associated with the mutation rate was derived by assuming that the number of mutations was Poisson distributed.

We fit lognormal distributions to the accumulation and mutation rate data by taking the  $\log_e$  of the values and then fitting a normal distribution by maximum likelihood using the *FindDistributionParameters* in *Mathematica*. Normal Q-Q plots for the accumulation and mutation rate data were produced using the *qqnorm* function in R version 1.0.143. In fitting these distributions, we have not taken into account the sampling error associated with the accumulation and mutation rate estimates. However, these sampling errors are small compared to the variance between species: for the accumulation rates the variance between species is  $3.9 \times 10^{-11}$  and the average error variance is an order of magnitude smaller at  $3.6 \times 10^{-12}$ ; for the mutation rate data, the variance between species is  $7.5 \times 10^{-18}$  and the average variance associated with sampling is more than two orders of magnitude smaller at  $1.8 \times 10^{-20}$ . Note that we cannot perform these comparisons of variances on a log-scale because we do not have variance estimates for the log accumulation and mutation rates.

To estimate phylogenetic signal in the accumulation and mutation rates we generated phylogenetic trees for each set of species in the two datasets. 16S rRNA sequences were downloaded from the NCBI genome database (<https://www.ncbi.nlm.nih.gov/genome/>) and aligned using MUSCLE [72] performed in Geneious version 10.0.9 (<http://www.geneious.com>, Kearse *et al.* [73]). From these alignments, maximum-likelihood trees were constructed in RAxML [74] and integrated into the tests of Pagel [33] and Blomberg *et al.* [34] to the  $\log_{10}$  (accumulation rates) and  $\log_{10}$  (mutation rates) implemented in the *phylosig* function in the R package *phytools* v.0.6 [75]. For the mutation rate dataset two species were excluded because of erroneous positioning on the phylogeny. See electronic supplementary material, figure S2A,B for details.

**Data accessibility.** All relevant data for this work are available as electronic supplementary material accompanying this article.

**Authors' contributions.** All authors were involved in conceptualizing the work. B.G. and A.E.W. wrote the manuscript with support from D.J.W. and E.F. B.G. and A.E.W. carried out data collection and analysis. All authors provided critical feedback and shaped the research, analysis and manuscript.

**Competing interests.** The authors declare that they have no competing interests.

**Funding.** D.J.W. is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (grant no. 101237/Z/13/Z).

**Acknowledgements.** We are very grateful to Michael Lynch for sharing his mutation rate estimates prior to publication.

## References

- Harmsen HJM, van Kuijck BLM, Plugge CM, Akkermans ADL, De Vos WM, Stams AJM. 1998 *Syntrophobacter fumaroxidans* sp. nov., a syntrophic propionate-degrading sulfate-reducing bacterium. *Int. J. Syst. Bacteriol.* **48**, 1383–1388. (doi:10.1099/00207713-48-4-1383)
- Gibbons RJ, Kapsimalis B. 1967 Estimates of the overall rate of growth of the intestinal microflora of hamsters, guinea pigs, and mice. *J. Bacteriol.* **93**, 510–512.
- Yang L, Haagensen JAJ, Jelsbak L, Johansen HK, Sternberg C, Høiby N, Molin S. 2008 *In situ* growth rates and biofilm development of *Pseudomonas aeruginosa* populations in chronic lung infections. *J. Bacteriol.* **190**, 2767–2776. (doi:10.1128/JB.01581-07)
- Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG. 2003 Measurably evolving populations. *Trends Ecol. Evol.* **18**, 481–488. (doi:10.1016/S0169-5347(03)00216-7)
- Reeves PR *et al.* 2011 Rates of mutation and host transmission for an *Escherichia coli* clone over 3 years. *PLoS ONE* **6**, e26907. (doi:10.1371/journal.pone.0026907)
- Markussen T, Marvig RL. 2014 Environmental heterogeneity drives within-host diversification and evolution of *Pseudomonas aeruginosa*. *MBio* **5**, e01592-14. (doi:10.1128/mBio.01592-14.Editor)
- Marvig RL, Johansen HK, Molin S, Jelsbak L. 2013 Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet.* **9**, e1003741. (doi:10.1371/journal.pgen.1003741)
- Duchêne S, Holt KE, Weill F-X, Le Hello S, Hawkey J, Edwards DJ, Fourment M, Holmes EC. 2016 Genome-scale rates of evolutionary change in bacteria. *Microb. Genomics* **2**, e000094. (doi:10.1101/069492)
- Zhou Z, McCann A, Weill F-X, Blin C, Nair S, Wain J, Dougan G, Achtman M. 2014 Transient Darwinian selection in *Salmonella enterica* serovar Paratyphi A during 450 years of global spread of enteric fever. *Proc. Natl Acad. Sci. USA* **111**, 12 199–12 204. (doi:10.1073/pnas.1411012111)
- Okoro CK *et al.* 2012 Intracontinental spread of human invasive *Salmonella typhimurium* pathovariants in sub-Saharan Africa. *Nat. Genet.* **44**, 1215–1221. (doi:10.1038/ng.2423)
- Hawkey J, Edwards DJ, Dimovski K, Hiley L, Billman-Jacobe H, Hogg G, Holt KE. 2013 Evidence of microevolution of *Salmonella typhimurium* during a series of egg-associated outbreaks linked to a single chicken farm. *BMC Genomics* **14**, 800. (doi:10.1186/1471-2164-14-800)
- Mather AE. 2013 Distinguishable epidemics of multidrug-resistant *Salmonella typhimurium* DT104 in different hosts. *Science* **341**, 1514–1518.
- Stinear TP, Holt KE, Chua K, Stepanell J, Tuck KL, Coombs G, Harrison PF, Seemann T, Howden BP. 2014 Adaptive change inferred from genomic population analysis of the ST93 epidemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. *Genome Biol. Evol.* **6**, 366–378. (doi:10.1093/gbe/evu022)
- Baines SL *et al.* 2015 Convergent adaptation in the dominant global hospital clone ST239 of methicillin-resistant *Staphylococcus aureus*. *MBio* **6**, e00080-15. (doi:10.1128/mBio.00080-15)
- Ward MJ, Gibbons CL, McAdam PR, van Bunnik BAD, Girvan EK, Edwards GF, Fitzgerald JR, Woolhouse MEJ. 2014 Time-scaled evolutionary analysis of the transmission and antibiotic resistance dynamics of *Staphylococcus aureus* clonal complex 398. *Appl. Environ. Microbiol.* **80**, 7275–7282. (doi:10.1128/AEM.01777-14)
- Holden MTG *et al.* 2013 A genomic portrait of the emergences, evolution, and global spread of methicillin-resistant *Staphylococcus aureus*. *Genome Res.* **23**, 653–664. (doi:10.1101/gr.147710.112.Freely)
- Uhlemann A-C, Dordel J, Knox JR, Raven KE, Parkhill J, Holden MTG, Peacock SJ, Lowy FD. 2014 Molecular tracing of the emergence, diversification, and transmission of *S. aureus* sequence type 8 in a New York community. *Proc. Natl Acad. Sci. USA* **111**, 6738–6743. (doi:10.1073/pnas.1401006111)
- Alam MT, Read TD, Petit RA, Boyle-Vavra S, Miller LG, Eells SJ, Daum RS, David MZ. 2015 Transmission and microevolution of USA300 MRSA in U.S. households: evidence from whole-genome sequencing. *MBio* **6**, e00054-15. (doi:10.1128/mBio.00054-15)
- Smyth DS, McDougal LK, Gran FW, Manoharan A, Enright MC, Song JH, De Lencastre H, Robinson DA. 2010 Population structure of a hybrid clonal group of methicillin-resistant *Staphylococcus aureus*, ST239-MRSA-III. *PLoS ONE* **5**, e8582. (doi:10.1371/journal.pone.0008582)
- Harris SR *et al.* 2010 Evolution of MRSA during hospital transmission and intercontinental spread. *Science* **327**, 469–474. (doi:10.1126/science.1182395)
- Gray RR, Tatem AJ, Johnson JA, Alekseyenko AV, Pybus OG, Suchard MA, Salemi M. 2011 Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant *Staphylococcus aureus* ST239 genome-wide data within a Bayesian framework. *Mol. Biol. Evol.* **28**, 1593–1603. (doi:10.1093/molbev/msq319)
- Nübel U *et al.* 2010 A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog.* **6**, e1000855. (doi:10.1371/journal.ppat.1000855)
- Young BC *et al.* 2012 Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc. Natl Acad. Sci. USA* **109**, 4550–4555. (doi:10.1073/pnas.1113219109)
- Mutreja A *et al.* 2011 Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* **477**, 462–465. (doi:10.1038/nature10392)
- Lee H, Popodi E, Tang H, Foster PL. 2012 PNAS Plus: Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **109**, E2774–E2783. (doi:10.1073/pnas.1210309109)
- Dettman JR, Sztepanacz JL, Kassen R. 2016 The properties of spontaneous mutations in the opportunistic pathogen *Pseudomonas aeruginosa*. *BMC Genomics* **17**, 1–14. (doi:10.1186/s12864-015-2244-3)
- Lind PA, Andersson DI. 2008 Whole-genome mutational biases in bacteria. *Proc. Natl Acad. Sci. USA* **105**, 17 878–17 883. (doi:10.1073/pnas.0804445105)
- Long H *et al.* 2018 Evolutionary determinants of genome-wide nucleotide composition. *Nat. Ecol. Evol.* **2**, 1–4. (doi:10.1038/s41559-017-0425-y)
- Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012 Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl Acad. Sci. USA* **109**, 18 488–18 492. (doi:10.1073/pnas.1216223109)
- Drake JW. 1991 A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **88**, 7160–7164. (doi:10.1073/pnas.88.16.7160)
- Lynch M, Ackerman MS, Gout J, Long H, Sung W, Thomas WK, Foster PL. 2016 Genetic drift, selection and the evolution of the mutation rate. *Nat. Publ. Gr.* **17**, 704–714. (doi:10.1038/nrg.2016.104)
- Vieira-Silva S, Rocha EPC. 2010 The systemic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genet.* **6**, e1000808. (doi:10.1371/journal.pgen.1000808)
- Pagel M. 1999 Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884. (doi:10.1038/44766)

34. Blomberg SP, Garland T, Ives AR. 2003 Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745. (doi:10.1111/j.0014-3820.2003.tb00285.x)
35. Bjedov I, Tenaillon O, Gerard B, Souza V, Denamur E, Radman M, Francois T, Matic I. 2003 Stress-induced mutagenesis in bacteria. *Science* **300**, 1404–1409.
36. Galhardo RS, Hastings PJ, Rosenberg SM. 2007 Mutation as a stress response and the regulation of evolvability. *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435. (doi:10.1080/10409230701648502)
37. Foster PL. 2007 Stress-induced mutagenesis in bacteria. *Crit. Rev. Biochem. Mol. Biol.* **42**, 373–397. (doi:10.1080/10409230701648494)
38. Maclean RC, Torres-barceló C, Moxon R. 2013 Evaluating evolutionary models of stress-induced mutagenesis in bacteria. *Nat. Rev. Genet.* **14**, 221–227. (doi:10.1038/nrg3415)
39. Shewaramani S, Finn TJ, Leahy SC, Kassen R, Rainey B, Moon CD. 2017 Anaerobically grown *Escherichia coli* has an enhanced mutation rate and distinct mutational spectra. *PLoS Genet.* **13**, 1–22. (doi:10.1371/journal.pgen.1006570)
40. Weller C, Wu M. 2015 A generation-time effect on the rate of molecular evolution in bacteria. *Evolution*. **69**, 643–652. (doi:10.1111/evo.12597)
41. Ford CB *et al.* 2011 Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat. Genet.* **43**, 482–486. (doi:10.1038/ng.811)
42. Lillebaek T, Norman A, Rasmussen EM, Marvig RL, Folkvardsen DB, Andersen ÅB, Jelsbak L. 2016 Substantial molecular evolution and mutation rates in prolonged latent *Mycobacterium tuberculosis* infection in humans. *Int. J. Med. Microbiol.* **306**, 580–585. (doi:10.1016/j.ijmm.2016.05.017)
43. Hershberg R, Petrov DA. 2008 Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299. (doi:10.1146/annurev.genet.42.110807.091442)
44. Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE. 2005 Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* **33**, 1141–1153. (doi:10.1093/nar/gki242)
45. Thorpe HA, Bayliss SC, Hurst LD, Feil EJ. 2017 Comparative analyses of selection operating on non-translated intergenic regions of diverse bacterial species. *Genetics* **206**, 363–376. (doi:10.1534/genetics.116.195784)
46. Ochman H, Elwyn S, Moran NA. 1999 Calibrating bacterial evolution. *Proc. Natl Acad. Sci. USA* **96**, 12 638–12 643. (doi:10.1073/pnas.96.22.12638)
47. Clark MA, Moran NA, Baumann P. 1999 Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol. Biol. Evol.* **16**, 1586–1598. (doi:10.1093/oxfordjournals.molbev.a026071)
48. Shepard CC. 1960 The experimental disease that follows the injection of human leprosy bacilli into foot-pads of mice. *J. Exp. Med.* **112**, 445–454. (doi:10.1016/j.infsof.2008.09.005)
49. Rees RJ. 1964 Limited multiplication of acid-fast bacilli in the foot-pads of mice inoculated with *Mycobacterium leprae*. *Br. J. Exp. Pathol.* **45**, 207–218.
50. Levy L. 1976 Bactericidal action of dapsone against *Mycobacterium leprae* in mice. *Antimicrob. Agents Chemother.* **9**, 614–617. (doi:10.1128/AAC.9.4.614)
51. Avrani S, Bolotin E, Katz S, Hershberg R. 2017 Rapid genetic adaptation during the first four months of survival under resource exhaustion. *Mol. Biol. Evol.* **34**, 1758–1769. (doi:10.1093/molbev/msx118)
52. Korem T *et al.* 2015 Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* **349**, 1101–1106. (doi:10.1126/science.aac4812)
53. Brown CT, Olm MR, Thomas BC, Banfield JF. 2016 Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256–1263. (doi:10.1038/nbt.3704)
54. Reidl J, Klose KE. 2002 *Vibrio cholerae* and cholera: out of the water and into the host. *FEMS Microbiol. Rev.* **26**, 125–139. (doi:10.1111/j.1574-6976.2002.tb00605.x)
55. G. ER. 1961 *Pseudomonas natriegens*, a marine bacterium with a generation time of less than 10 min. *J. Bacteriol.* **83**, 736–737.
56. Polz MF, Hunt DE, Preheim SP, Weinreich DM. 2006 Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Phil. Trans. R. Soc. B* **361**, 2009–2021. (doi:10.1098/rstb.2006.1928)
57. Blokesch M, Schoolnik GK. 2008 The extracellular nuclease Dns and its role in natural transformation of *Vibrio cholerae*. *J. Bacteriol.* **190**, 7232–7240. (doi:10.1128/JB.00959-08)
58. Takemura AF, Chien DM, Polz MF. 2014 Associations and dynamics of Vibrionaceae in the environment, from the genus to the population level. *Front. Microbiol.* **5**, 1–26. (doi:10.3389/fmicb.2014.00038)
59. Yamaichi Y, Iida T, Park KS, Yamamoto K, Honda T. 1999 Physical and genetic map of the genome of *Vibrio parahaemolyticus*: presence of two chromosomes in *Vibrio* species. *Mol. Microbiol.* **31**, 1513–1521. (doi:10.1046/j.1365-2958.1999.01296.x)
60. Schenck LP, Surette MG, Bowdish DME. 2016 Composition and immunological significance of the upper respiratory tract microbiota. *FEBS Lett.* **590**, 3705–3720. (doi:10.1002/1873-3468.12455)
61. John JF. 2004 Staphylococcal infection: emerging clinical syndromes. In *Staphylococcus aureus* (eds DAA Ala'Aldeen, K Hiramatsu), pp. 1–30. Chichester, UK: Horwood Publishing Limited.
62. Bui LMG, Conlon BP, Kidd SP. 2017 Antibiotic tolerance and the alternative lifestyles of *Staphylococcus aureus*. *Essays Biochem.* **61**, 71–79. (doi:10.1042/EBC20160061)
63. Proctor RA, von Eiff C, Kahl BC, Becker K, McNamara P, Herrmann M, Peters G. 2006 Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat. Rev. Microbiol.* **4**, 295–305. (doi:10.1038/nrmicro1384)
64. Huse HK, Kwon T, Zlosnik JEA, Speert DP, Marcotte EM, Whiteley M. 2010 Parallel evolution in *Pseudomonas aeruginosa* over 39 000 generations *in vivo*. *MBio* **1**, e00199-10. (doi:10.1128/mBio.00199-10.Editor)
65. Smith EE *et al.* 2006 Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl Acad. Sci. USA* **103**, 8487–8492. (doi:10.1073/pnas.0602138103)
66. Palmer KL, Mashburn LM, Singh PK, Whiteley M. 2005 Cystic fibrosis sputum supports growth and cues key aspects of *Pseudomonas aeruginosa* physiology. *J. Bacteriol.* **187**, 5267–5277. (doi:10.1128/JB.187.15.5267)
67. Winfield MD, Groisman EA. 2003 Role of nonhost environments in the lifestyles of *Salmonella* and *E. coli*. *Appl. Environ. Microbiol.* **69**, 3687–3694. (doi:10.1128/AEM.69.7.3687)
68. Oh S, Buddenborg S, Yoder-Himes DR, Tiedje JM, Konstantinidis KT. 2012 Genomic diversity of *Escherichia* isolates from diverse habitats. *PLoS ONE* **7**, e47005. (doi:10.1371/journal.pone.0047005)
69. Davies RH, Wray C. 1996 Seasonal variations in the isolation of *Salmonella typhimurium*, *Salmonella enteritidis*, *Bacillus cereus* and *Clostridium perfringens* from environmental samples. *Zentralbl. Veterinarmed. B* **43**, 119–127. (doi:10.1111/j.1439-0442.1996.tb00435.x)
70. Bogosian G, Sammons L. 1996 Death of the *Escherichia coli* K-12 strain W3110 in soil and water. *Appl. Environ. Microbiol.* **62**, 4114–4120.
71. Kennemann L *et al.* 2011 *Helicobacter pylori* genome evolution during human infection. *Proc. Natl Acad. Sci. USA* **108**, 5033–5038. (doi:10.1073/pnas.1018444108)
72. Edgar RC. 2004 MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)
73. Kearse M *et al.* 2012 Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. (doi:10.1093/bioinformatics/bts199)
74. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (doi:10.1093/bioinformatics/btu033)
75. Revell LJ. 2012 phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223. (doi:10.1111/j.2041-210X.2011.00169.x)