

Common Adaptive Strategies Underlie Within-Host Evolution of Bacterial Pathogens

Yair E. Gatt¹ and Hanah Margalit^{*1}

¹Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

***Corresponding author:** E-mail: hanahm@ekmd.huji.ac.il.

Abstract

Within-host adaptation is a hallmark of chronic bacterial infections, involving substantial genomic changes. Recent large-scale genomic data from prolonged infections allow the examination of adaptive strategies employed by different pathogens and open the door to investigate whether they converge toward similar strategies. Here, we compiled extensive data of whole-genome sequences of bacterial isolates belonging to miscellaneous species sampled at sequential time points during clinical infections. Analysis of these data revealed that different species share some common adaptive strategies, achieved by mutating various genes. Although the same genes were often mutated in several strains within a species, different genes related to the same pathway, structure, or function were changed in other species utilizing the same adaptive strategy (e.g., mutating flagellar genes). Strategies exploited by various bacterial species were often predicted to be driven by the host immune system, a powerful selective pressure that is not species specific. Remarkably, we find adaptive strategies identified previously within single species to be ubiquitous. Two striking examples are shifts from siderophore-based to heme-based iron scavenging (previously shown for *Pseudomonas aeruginosa*) and changes in glycerol-phosphate metabolism (previously shown to decrease sensitivity to antibiotics in *Mycobacterium tuberculosis*). Virulence factors were often adaptively affected in different species, indicating shifts from acute to chronic virulence and virulence attenuation during infection. Our study presents a global view on common within-host adaptive strategies employed by different bacterial species and provides a rich resource for further studying these processes.

Key words: microbiology, within-host adaptation, within-host evolution, infectious diseases, microbial genetics.

Introduction

Despite the advent of antibiotics, bacterial infections continue to cause millions of deaths world-wide (Lozano et al. 2012). These infections can range from acute, lasting a few days, to chronic, enduring over dozens of years. The host environment constitutes a unique ecological niche, introducing pathogens to a wide array of challenges, with antibiotic treatment and the host immune system being the most pronounced. Also, although the host metabolic environment is rich in nutrients, it exposes the bacteria to fierce competition with the local microbiota (Frydenlund Michelsen et al. 2016). Bacteria have been shown to evolve to meet these challenges over the course of the infection, developing mechanisms of immune evasion and even different forms of antibiotic resistance via genetic changes. Recent developments in high-throughput genomic technologies allow monitoring of genomic changes in small time frames during chronic bacterial infections (Didelot et al. 2016), enabling the tracing of genomic and genetic changes. Most studies in this field have focused on the evolution of opportunistic pathogens that shift between vastly different environments (Winstanley et al. 2016; Young et al. 2017; Chung et al. 2017). However, genetic

changes are also prominent in other bacteria as they shift from initial to chronic infection (Marzel et al. 2016; Ley et al. 2019) and as they move between different subniches within the human host (Wylie et al. 2019).

Previous studies have demonstrated many mechanisms utilized by specific bacterial species to overcome the pressures and environmental changes they encounter (Alamro et al. 2014; Marvig, Sommer, et al. 2015; Didelot et al. 2016). These include loss of nonessential metabolic functions (Winstanley et al. 2016; Viberg et al. 2017), as well as tailoring of the expression and activity of metabolic pathways (Didelot et al. 2016; La Rosa et al. 2018). For example, Marvig et al. (2014) identified mutations in *Pseudomonas aeruginosa* that lead to reduced production of siderophores along with strong upregulation of systems for iron scavenging from heme. This phenomenon has been interpreted in the context of cheating behavior and subsequent upregulation of noncooperative iron uptake systems in bacterial populations (Andersen et al. 2015, 2018). It has the additional beneficial effect of downregulating the generally immunogenic siderophores (Wandersman and Delepelaire 2004).

Although previous studies have characterized genes that underwent loss of function or diversification during infection

for specific pathogens and specific clinical contexts (Lieberman et al. 2011; Marvig, Sommer, et al. 2015; Brodrick et al. 2017), it is intriguing to find out whether different pathogens have converged toward similar strategies for their adaptation within the host environment. Here, we take advantage of the ample sequencing data of bacterial pathogens derived from prolonged infections to systematically address this question in large scale, investigating convergent evolution both within species and across species. Using a statistical framework, we trace the evolution of different pathogens within their hosts and assess the relationship between genes that underwent adaptive changes in the various pathogens. Although in many cases there are no orthologous relationships between the genes identified as undergoing adaptive changes in the various species, it is remarkable that often they are involved in parallel cellular processes, suggesting that different bacterial species exploit common strategies for their within-host adaptation.

Results

Constructing the Database

We compiled data from studies that conducted genome sequencing of an infecting pathogen in at least two time points during chronic infection or carriage (table 1). Carriage is defined here as colonization by a pathogen with no apparent symptoms over the entire period. The database includes 1,421 patients infected with 29 bacterial species encompassing 7,197 isolates (fig. 1a and supplementary fig. 1a and b, Supplementary Material online). We downloaded the sequencing reads of all isolates and assembled them into genomes in a uniform manner (Materials and Methods). To avoid biases due to species infecting only few patients, we focused on 11 bacterial pathogens for which ample data were available: *Pseudomonas aeruginosa* (PA), *Mycobacterium tuberculosis* (TB), *Staphylococcus aureus* (SA), *Klebsiella pneumoniae* (KP), *Salmonella enterica* subsp. *enterica* serovar *Typhimurium* (ST), *Escherichia coli* (EC), *Enterococcus faecium* (EFci), *Haemophilus influenzae* (HI), *Campylobacter jejuni* (CJ), *Clostridioides difficile* (CD), and *Acinetobacter baumannii* (AB). The species we chose present a varied set of pathogens, including both opportunistic pathogens (PA, SA, KP, EC, EFci, CD, nontypable HI, and AB), and strict pathogens (TB, ST, CJ, and encapsulated HI), infecting the patients from a variety of sources including patient-to-patient transmission (TB, HI, CD, and some PA), contaminated food (CJ, ST), the hospital environment (PA, AB, and KP), and the patient's own microbiome (SA, EC, EFci, and KP).

Although the data include patients presenting one of two clinical scenarios, infection or carriage, for most species nearly all patients presented the same clinical scenario (either infection or carriage, fig. 1b and supplementary table 1, Supplementary Material online). For three species, SA, KP, and EFci, the data were derived from both patients with infection and patients with carriage.

Tracing the Phylogenetic Relationships between Different Isolates

During infection, the infecting bacteria present a genetically diversified population. Due to this within-host diversity, the isolates from each patient are in fact samples from a diverse population. This may introduce substantial sampling noise to the data: two isolates sampled at sequential time points do not necessarily have a progenitor–progeny relation, making the assessment of within-host evolution challenging. To reduce this sampling noise and trace the in-patient progenitor–progeny relationships of the isolates, we developed a two-level pipeline, termed TRACE (fig. 1c).

First, we assessed whether the isolates from each patient were derived from a monoclonal or polyclonal infection (Materials and Methods) and found that for most of the patients, all isolates were found on a single clade (supplementary fig. 1c, Supplementary Material online), hinting at a monoclonal infection. To prevent bias due to studies that did not choose the samples randomly, we repeated the above analysis excluding studies suspected to have such biases, reconfirming our previous observations regarding the dominance of monoclonality (supplementary fig. 1d, Supplementary Material online). We refer to all isolates from a single patient as representative of the same “strain” and to the ones that were found on the same clade as representative of the same “clone.”

Second, we developed an algorithm to construct the phylogenetic relationships between the isolates. Phylogenetic trees produced by previous construction methods include internal nodes representing inferred common ancestors, whereas in our data the ancestors were empirically sampled at previous time points. Accordingly, our algorithm for the construction of phylogenetic trees is parsimony-based and utilizes both inferred and sampled internal nodes for the inference of progenitor–progeny pairs (Materials and Methods). We assessed the accuracy of TRACE using simulated bacterial evolution (Worby and Read 2015) and discovered that it has high accuracy that is associated with the number of sampled isolates (supplementary fig. 2, Supplementary Material online). The time intervals between the different isolates were variable, being in the magnitude of days for some species and of years for others (supplementary fig. 3, Supplementary Material online).

Mutation Rates Are Higher at Earlier Stages of the Infection

We used the breseq pipeline (Barrick et al. 2014; Deatherage and Barrick 2014) to detect differences between the sequencing reads of each progenitor–progeny pair and annotated the different types of changes and their potential impact using SnpEff (Cingolani et al. 2012), while filtering out reads that were mapped exclusively to repeating sequences. We did not observe any biases in the composition of mutated nucleotides or amino acids for most species (supplementary figs. 4 and 5, Supplementary Material online). We noted a near ubiquitous decrease in mutation rates throughout the infection (supplementary fig. 6a, Supplementary Material online), indicating

Table 1. Studies Used for Database Construction.

Bacterial Species	Study
<i>Acinetobacter baumannii</i>	Hornsey et al. (2011), Lesho et al. (2013), Snitkin et al. (2013), Wright et al. (2016), Hua et al. (2017), Hawkey et al. (2018), Kim et al. (2018), and Gerson et al. (2019)
<i>Burkholderia dolosa</i>	Lieberman et al. (2011)
<i>Burkholderia cenocepacia</i>	Hassan et al. (2017)
<i>Burkholderia multivorans</i>	Silva et al. (2016) and Diaz Caballero et al. (2018)
<i>Burkholderia pseudomallei</i>	Hayden et al. (2012), Price et al. (2013, 2015), Viberg et al. (2017), Webb et al. (2017), and Sarovich et al. (2018)
<i>Campylobacter jejuni</i>	Thomas et al. (2014) and Crofts et al. (2018)
<i>Clostridioides difficile</i>	Sim et al. (2017) and Sachsenheimer et al. (2018)
<i>Enterococcus faecalis</i>	Arias et al. (2011)
<i>Enterococcus faecium</i>	Brodrick et al. (2016), Moradigaravand et al. (2017), and Raven et al. (2018)
<i>Escherichia coli</i>	Li et al. (2016), Brodrick et al. (2017), Gumpert et al. (2017), Porse et al. (2017), Snesrud et al. (2017), and Ghalayini et al. (2018)
<i>Fusobacterium nucleatum</i>	Bullman et al. (2017)
<i>Haemophilus influenzae</i>	Garmendia et al. (2014), Molerres et al. (2018), and Pettigrew et al. (2018)
<i>Helicobacter pylori</i>	Kennemann et al. (2011)
<i>Klebsiella aerogenes</i>	Philippe et al. (2015)
<i>Klebsiella pneumoniae</i>	Cannatelli et al. (2014), Nielsen et al. (2014), Mathers et al. (2015, 2019), Lynch et al. (2016), Martin et al. (2016), Gaibani et al. (2018), and Simner et al. (2018)
<i>Klebsiella quasipneumoniae</i>	Mathers et al. (2019)
<i>Mycobacterium abscessus</i>	Bryant et al. (2013) and Kreutzfeldt et al. (2013)
<i>Mycobacterium avium</i>	Kannan et al. (2019)
<i>Mycobacterium leprae</i>	Stefani et al. (2017)
<i>Mycobacterium tuberculosis</i>	Comas et al. (2012), Saunders et al. (2011), Sun et al. (2012), Clark et al. (2013), Farhat et al. (2013), Merker et al. (2013), Eldholm et al. (2014), Pérez-Lago et al. (2014), Liu et al. (2015), Meumann et al. (2015), Datta et al. (2016), Korhonen et al. (2016), Lillebaek et al. (2016), Herranz et al. (2017), Trauner et al. (2017), Wollenberg et al. (2017), and Xu et al. (2018)
<i>Neisseria meningitidis</i>	Omer et al. (2011)
<i>Pandoraea apista</i>	Greninger et al. (2017)
<i>Pseudomonas aeruginosa</i>	Cramer et al. (2011), Marvig, Jochumsen, et al. (2013), Marvig, Johansen, et al. (2013), Feliziani et al. (2014), Markussen et al. (2014), Singh et al. (2014), Stewart et al. (2014), Bianconi et al. (2015, 2019), Diaz Caballero et al. (2015), Marvig, Dolce, et al. (2015), Marvig, Sommer, et al. (2015), Sommer et al. (2016), Spilker and LiPuma (2016), van Mansfeld et al. (2016), Beaume et al. (2017), López-Causapé et al. (2017, 2018), Sherrard et al. (2017), Wang et al. (2017), Klockgether et al. (2018), and Williams et al. (2018)
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Enteritidis</i>	Klemm et al. (2016)
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i>	Marzel et al. (2016) and Octavia et al. (2017)
<i>Serratia marcescens</i>	Mathers et al. (2019)
<i>Staphylococcus aureus</i>	Mwangi et al. (2007), Howden et al. (2008, 2011), Bloemendaal et al. (2010), Gao et al. (2010), McAdam et al. (2011), Cameron et al. (2012, 2015), Passalacqua et al. (2012), Peleg et al. (2012), Young et al. (2012, 2017), Van Hal et al. (2014); Chen et al. (2015), Gladman et al. (2015), Tong et al. (2015), Azarian et al. (2016), Schwartbeck et al. (2016), Trouillet-Assant et al. (2016), Ankrum and Hall (2017), Coll et al. (2017), Price et al. (2017), Sabat et al. (2017), and Giulieri et al. (2018)
<i>Stenotrophomonas maltophilia</i>	Pak et al. (2015) and Esposito et al. (2017)
<i>Streptococcus pneumoniae</i>	Hiller et al. (2010) and Gladstone et al. (2015)

that most genomic changes occur at early stages of the infection in accordance with previous findings for PA (Yang et al. 2011; Bartell et al. 2019). These results were consistent even when excluding the large fractions of hypermutator clones found among the different strains in our database (supplementary fig. 6b and c, Supplementary Material online).

Adaptive Changes and Loss of Function of Genes Are Ubiquitous during Infections

We recorded variations likely leading to loss of function (complete deletion of a gene, loss or gain of stop codon, loss of start codon, and frameshift mutation), and variations leading to nonsynonymous genetic changes of unknown significance

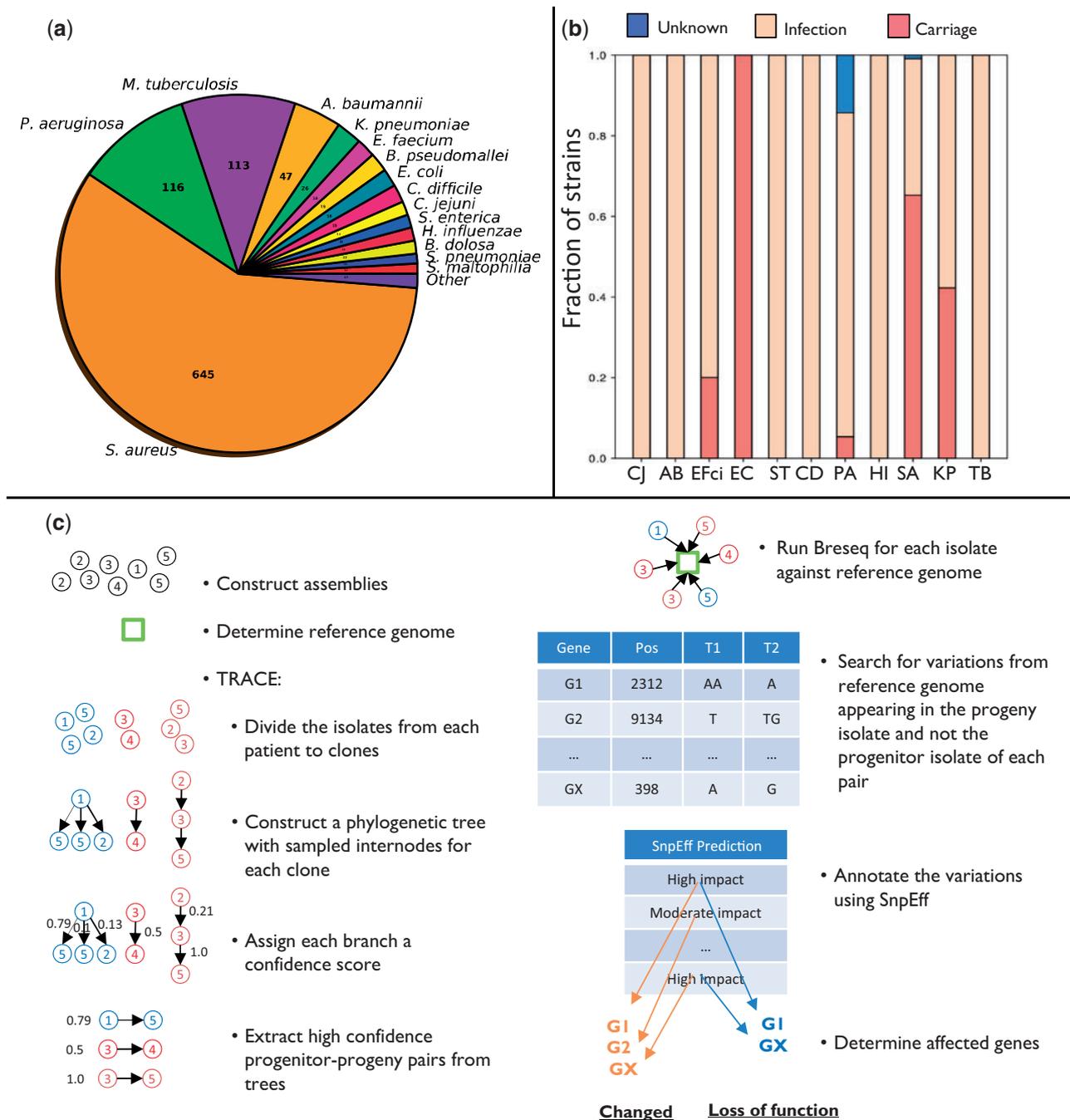


FIG. 1. Overview of the study. (a) Number of strains of each bacterial species in the database. (b) Fraction of strains in each species isolated from patients undergoing different clinical scenarios (infection/carriage). PA, *Pseudomonas aeruginosa*; TB, *Mycobacterium tuberculosis*; SA, *Staphylococcus aureus*; KP, *Klebsiella pneumoniae*; ST, *Salmonella enterica subsp. enterica serovar Typhimurium*; EC, *Escherichia coli*; EFci, *Enterococcus faecium*; HI, *Haemophilus influenzae*; CJ, *Campylobacter jejuni*; CD, *Clostridioides difficile*; AB, *Acinetobacter baumannii*. (c) Analysis pipeline for extracting genes undergoing either loss of function or any change (including loss of function) in a strain. We constructed assemblies for the different isolates using their respective sequencing reads and determined the reference genome closest to the isolates of each strain using QUAST. We utilized TRACE to assess the phylogenetic relations between the different isolates to extract progenitor–progeny isolate pairs (Materials and Methods). Breseq was then used to compare the sequencing reads of each isolate with the reference genome, and positions that differed from the reference genome and varied between the progeny and progenitor were determined. These variations were analyzed by SnpEff to define: 1) genes that underwent loss of function (based on high-impact variations appearing in the progeny isolate and not the progenitor isolate, leading, e.g., to deletion of a gene or to a stop codon); 2) changed genes (based on variations of moderate or high impact appearing in the progeny isolate and not the progenitor isolate, where moderate impact refers, e.g., to amino acid substitutions). The genes determined as undergoing loss of function make up a subgroup of the genes determined as changed.

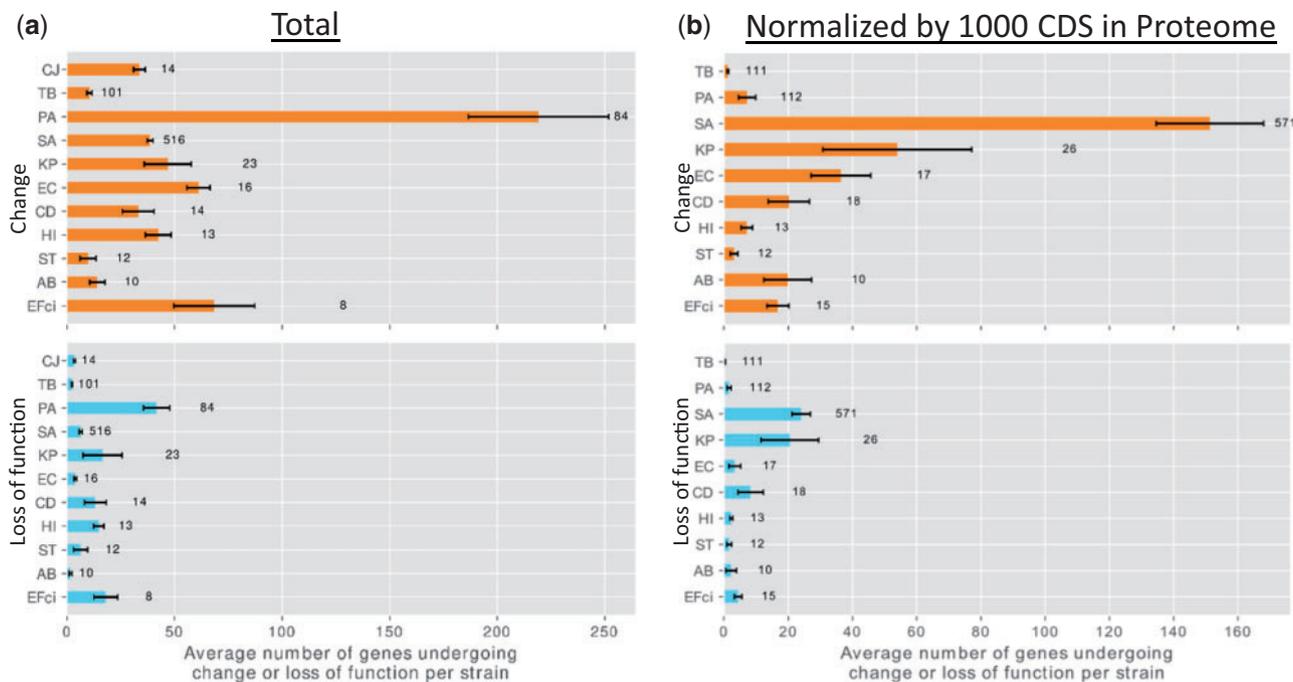


Fig. 2. Different species vary in the number of genes that underwent change or loss of function per strain. (a) Upper panel: Number of changed genes per strain in the data for each bacterial species. Lower panel: Number of genes that underwent loss of function per strain in the data for each bacterial species. Error bars represent standard error. Note that the group of genes that underwent loss of function is a subgroup of the changed genes. (b) The same, with the result for each species normalized by the number of proteins in its reference proteome, scaled to 1,000 coding sequences. Names of bacterial species are as in figure 1b. The number to the right of the bar represents the number of strains included for that bacterial species. CJ was excluded from (b) to maintain scale.

(e.g., amino acid substitutions), based on the annotations obtained by SnpEff (Cingolani et al. 2012). Many of the analyses described below were performed separately for two groups of genes: 1) all genes containing any genetic change (changed genes) and 2) the subgroup of genes containing variations likely leading to loss of function (genes undergoing loss of function). Changed genes, including the genes that underwent loss of function, were ubiquitous in our data and varied in their degree among strains of different species (fig. 2). The mutation burden of a strain during infection depends on its mutation rate, genome size, and the duration of the infection in generations (Drake 1991). In accordance with this, a high number of genes underwent mutations, including loss of function, in PA strains, which have large genomes (Stover et al. 2000), long durations of infection, and a high fraction of hypermutator clones (supplementary fig. 6c, Supplementary Material online). Only a small number of genes underwent genetic changes or loss of function in TB and ST strains: In the former, probably due to its extremely low mutation rate (Sherman and Gagneux 2011) and long generation time; in the latter, due to its shorter duration of infection, up to 120 days for all but one patient in our data (supplementary table 1, Supplementary Material online).

Genetic changes identified in our analysis are not necessarily the result of adaptive processes but may have occurred randomly. To pinpoint genes that were possibly affected by adaptive pressures, we searched for genes that repeatedly

underwent loss of function or repeatedly underwent changes in different strains independently. We developed a statistical framework to detect those genes that underwent loss of function or any change above random expectation for each species (Materials and Methods). For a gene that underwent loss of function (any change) in k strains, we simulated the probability for genes to undergo loss of function (any change) in at least k strains by chance alone, based on the fraction of genes that underwent loss of function (any change) in each strain during the infection (supplementary fig. 7, Supplementary Material online). A gene was defined as undergoing adaptive loss of function (change) if the computed probability was below a pre-defined statistical threshold (Materials and Methods). An important distinction is that although genes undergoing loss of function are a subset of the changed genes, genes undergoing adaptive loss of function are not necessarily a subset of adaptively changed genes, as they may pass the statistical threshold only in the analysis of the subgroup of genes that underwent loss of function but not in the analysis of all changed genes.

We found that genes undergoing adaptive change made up a substantial fraction of the genes with any genetic change, whereas genes undergoing adaptive loss of function made up a much smaller fraction of this subset of genes (fig. 3 and supplementary table 2, Supplementary Material online). The mean fractions of genes undergoing adaptive change and loss of function per species were correlated with the number of

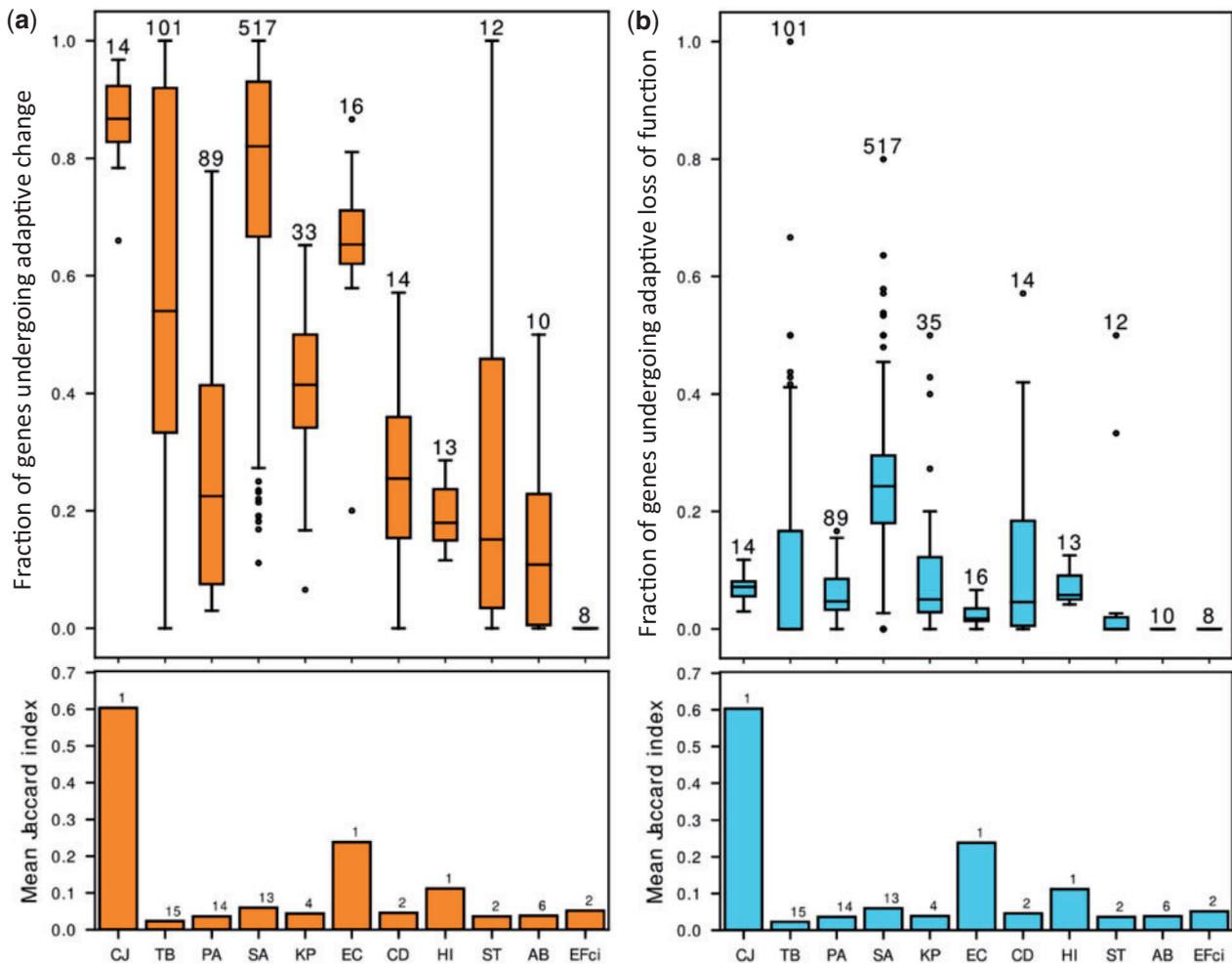


Fig. 3. The same genes of a bacterial species repeatedly undergo change or loss of function during infection by different strains. Our statistical framework defines genes that undergo change or loss of function repeatedly in a statically significant number of strains as undergoing adaptive change or loss of function, respectively. (a) Upper panel: Boxplots of the fractions of adaptively changed genes out of all changed genes in the different strains of each species. The number above the bar represents the number of strains included for that bacterial species. Lower panel: The mean similarity between the lists of genes undergoing change or loss of function in different strains of each species as measured by the Mean Jaccard Index (Materials and Methods). The number above the bar represents the number of studies included for that bacterial species. (b) Upper panel and lower panel as in (a), for genes undergoing loss of function. For most strains, there is no relationship between the fraction of genes undergoing adaptive change/loss of function and the similarity between strains of a species. Names of bacterial species are as in [figure 1b](#).

strains in the species in our data (Spearman correlation, $r = 0.66$ and $r = 0.63$, respectively), suggesting there may be additional genes undergoing adaptive change or loss of function that we lack the statistical power to detect. CJ and EC had very high fractions of adaptively changed genes despite low numbers of strains. For these two species, all or nearly all strains originated from a single study and likely had similar genetic background and were exposed to similar selective pressures. This is reflected in high Jaccard indices between the lists of genes undergoing change in different strains ([fig. 3a](#), Materials and Methods). This indicates that genes undergoing adaptive change or loss of function in species with a small number of studies may include false positives related to the low variability between the different strains in those studies. Interestingly, the fractions of adaptively changed genes and of genes undergoing adaptive loss of

function did not seem to greatly differ between opportunistic and strict pathogens or between pathogens with different routes of transmission.

Close inspection of the results revealed that transcription factors were prominent among genes undergoing adaptive change and among genes undergoing adaptive loss of function ([supplementary table 2](#), [Supplementary Material](#) online), possibly indicating considerable transcriptional changes during infections. Additionally, genes related to antibiotic resistance were prominent among all adaptive lists, highlighting the selective pressures exerted by these substances. We noticed many important virulence factors that underwent loss of function among many strains, including superantigen-like proteins, fibrinogen-binding proteins, staphcoagulases, and immune evasion protein Sbi of SA. In TB, many genes

undergoing adaptive change or loss of function belonged to the PE/PPE families, suspected of being involved in antigenic variation.

Genes Undergoing Adaptive Change Are Generally under Positive Selection and Are Not Associated with Lower Evolutionary Conservation, Specific Genomic Regions, or Clinical Scenarios

To study the selection acting on adaptively changed genes, we compared their dN/dS ratios with the rest of the genes in the genome. We found that nonadaptively changed genes are generally under purifying selection ($dN/dS < 1.0$), and that dN/dS values for adaptively changed genes are statistically significantly higher by Mann–Whitney U test (one tailed, P value ≤ 0.05). Many of the adaptively changed genes are under positive selection ($dN/dS > 1.0$), but there is a large variance between the different genes (supplementary fig. 8, Supplementary Material online). Most interestingly, this does not hold for PA where nearly the entire proteome seems to be under positive selection.

We found that both genes undergoing adaptive change and genes undergoing adaptive loss of function were only weakly associated with low evolutionary conservation (supplementary fig. 9, Supplementary Material online), demonstrating that the changes we discovered with our statistical framework are likely derived from adaptive processes within the host environment. We detected a small number of genomic regions that were enriched with genes undergoing adaptive change or loss of function (supplementary table 3, Supplementary Material online), these regions were mostly associated with operons and mobile genetic elements. Since different patients may present different clinical scenarios (carriage or infection), we assessed the results in view of this aspect. This analysis was irrelevant for most species, as nearly all patients presented a single common clinical scenario. EFCi was excluded from this analysis due to lack of genes passing the statistical threshold. For SA, there were minimal differences in the genes identified in the different clinical scenarios. For KP however, we noted that there was little overlap between genes that underwent adaptive change or loss of function in the different clinical scenarios (supplementary fig. 10, Supplementary Material online). This indicates that although genes that undergo host adaptation in SA are generalizable between different clinical scenarios, for KP they are scenario specific. This finding may be related to the small number of strains in each clinical scenario for KP. Increasing the number of strains in the different scenarios will allow us to detect additional genes as adaptively changing and may increase the overlap between the different scenarios.

To reinforce the results of our automatic pipeline we verified its performance using several approaches: 1) Simulations were used to validate the performance of the statistical framework, discovering it has extremely high specificity and high sensitivity (Supplementary Material online). 2) Comparison of the pipeline results to the original studies that sequenced the genomes demonstrated that our pipeline rediscovered 72% of the genes reported to undergo mutation in the original studies (Supplementary Material online). Notably, we only had

26% overlap when comparing our list of adaptively changed genes against a previously published list for PA (Marvig, Sommer, et al. 2015). Nearly all previously designated adaptively changed genes were identified by us as accumulating genetic changes in multiple strains within our data, but many of them did not pass our statistical threshold. This indicates the stringency of our statistical framework, which may hamper the detection of some genes that are under selective pressure to undergo mutation during the infection. 3) Assuming that the genes we detect as undergoing loss of function are indeed not functional in the progeny isolates, it is unlikely that the bacteria could survive if these genes included known essential genes. Indeed, we found that 96% of the essential proteins reported in a recent proteomic study of PA (Poulsen et al. 2019) were not identified as undergoing loss of function in any of the PA strains in our data (Supplementary Material online, see also supplementary fig. 11, Supplementary Material online). 4) Finally, as our pipeline is based on the mapping of sequencing reads against a reference genome, we compared its results with results obtained by an alternative approach based on de novo assembly and annotation of the different isolates using SPAdes (Bankevich et al. 2012) and prokka (Seemann 2014) (Supplementary Material online). We found the de novo assembly-based pipeline was extremely prone to incorrectly identifying genetic changes during the infections due to assembly errors. Accordingly, a consistent small fraction of the genes undergoing change or loss of function in the de novo assembly-based pipeline were rediscovered in our reference-based pipeline (supplementary fig. 12a and b, Supplementary Material online). In contrast, the fraction of genes undergoing change or loss of function by our pipeline that were rediscovered by the de novo assembly-based pipeline was quite variable between different bacterial species and strains, ranging between 0.0 and 0.2 for changed genes in CD to 0.5–1.0 for genes undergoing loss of function in PA (supplementary fig. 12c and d, Supplementary Material online). Importantly, the main conclusions derived by applying the two pipelines are consistent (see below and in Supplementary Material online).

Frameshift Mechanisms with Distinct Nucleotide Biases Underlie Adaptive Loss of Function

We used the algorithm SnpEff (Cingolani et al. 2012) to predict the molecular mechanisms underlying loss of function in our data (e.g., deletion, frameshift, or nonsense mutations that lead to loss of function). Comprehensive analysis of the distributions of different mechanisms leading to loss of function of genes in different bacterial strains and species demonstrated that frameshift mutations were the most common molecular mechanism of adaptive loss of function events in all bacterial species except CD and ST (fig. 4a). Frameshift mutations require only the change of a small number of nucleotides to induce loss of function and are easily reversible (Hu and Ng 2012), making them useful tools for evolution by loss of function. We noted that frameshifts were usually minimal, with only a single nucleotide inserted or deleted, and had distinct nucleotide biases in different species, diverging from the standard %GC of the respective organism

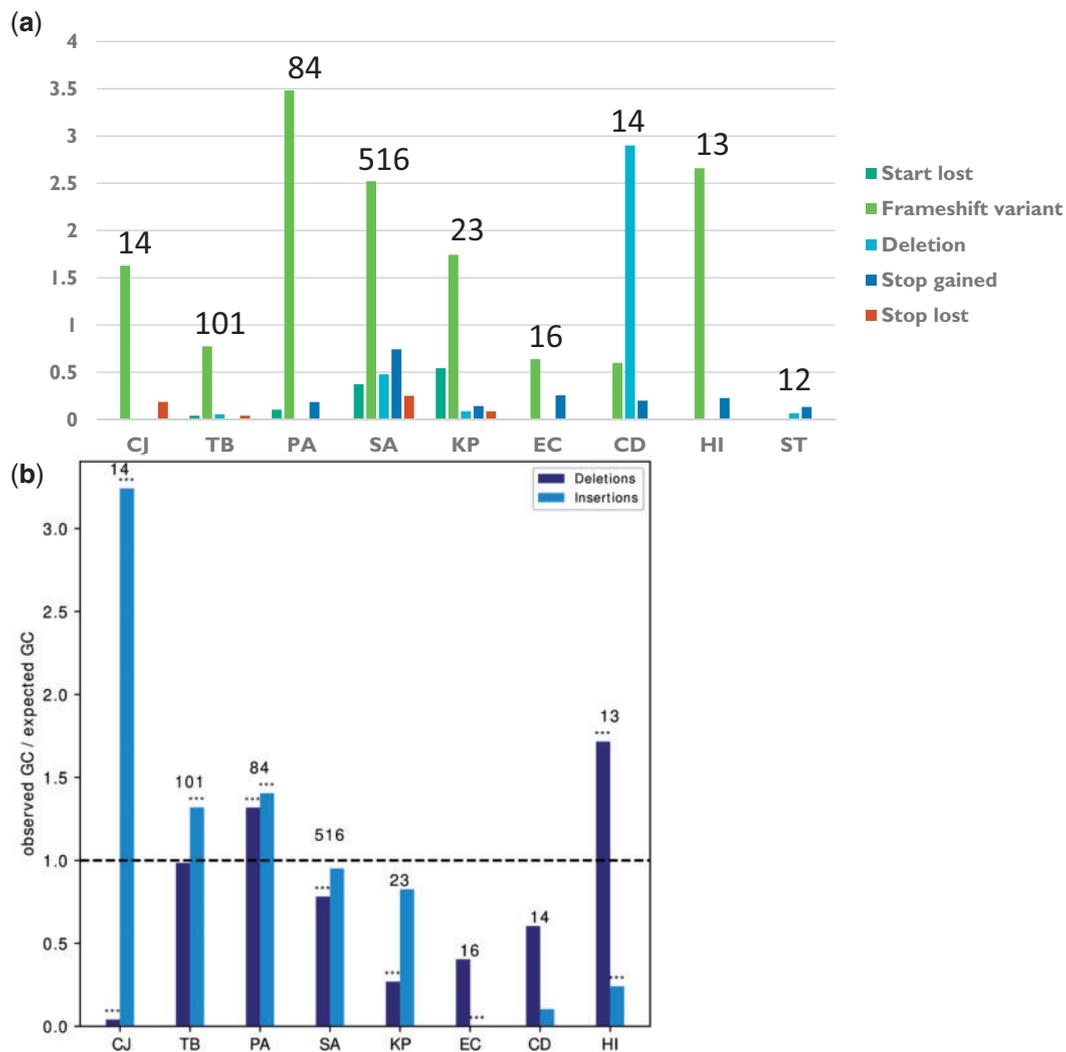


Fig. 4. Adaptive loss of function is mostly mediated by frameshift mutations. (a) The number of adaptive loss of function events due to different loss of function mechanisms, normalized by the number of progenitor–progeny pairs in each species. (b) %GC in insertions or deletions leading to frameshift events underlying adaptive loss of function, normalized by the %GC of the entire genome of each species. Asterisks denote %GC that statistically significantly differs from the %GC of the entire genome by binomial test. * $P \leq 0.05$, ** $P \leq 0.01$, and *** $P \leq 0.001$. Names of bacterial species are as in [figure 1b](#). The numbers above the bars represent the number of strains included for that bacterial species.

([fig. 4b](#)). Since frameshift mutations are related to insertion and deletion events (INDELs), we compared the predicted INDELs with the results of the de novo assembly-based pipeline and found that 85.8% of predicted INDELs were reaffirmed (Materials and Methods).

Flagellar Assembly, Lipooligosaccharide Sialylation, and Pore Formation Are Adaptively Changed in Different Species

It is intriguing to find out whether orthologs in different species have been identified as undergoing adaptive change or loss of function. We predicted orthologs for all genes within all species using reciprocal BlastP (Materials and Methods) and found three pairs of orthologous genes undergoing adaptive loss of function in multiple species. In addition, two triplets and 15 pairs of orthologous genes were identified as undergoing adaptive change ([supplementary table 4](#),

[Supplementary Material](#) online). For most genes undergoing adaptive change or loss of function during within-host adaptation of different species, we did not identify orthologous relationships, despite using extremely lenient thresholds for BlastP. To inspect whether these genes can be classified by similar functions or similar pathways they take part in, we assigned all genes undergoing adaptive change or loss of function their respective KEGG pathways ([Kanehisa and Goto 2000](#)) and gene ontology (GO) annotations ([Ashburner et al. 2000](#)). We established an additional statistical framework to assess those pathways and functions in which genes were adaptively affected in more species than expected at random (Materials and Methods). After correction for testing of multiple hypotheses, we found that genes related to the flagellum filament, to *N*-acylneuraminidase activity and to pore formation were adaptively changed in more species than expected at random ([supplementary table](#)

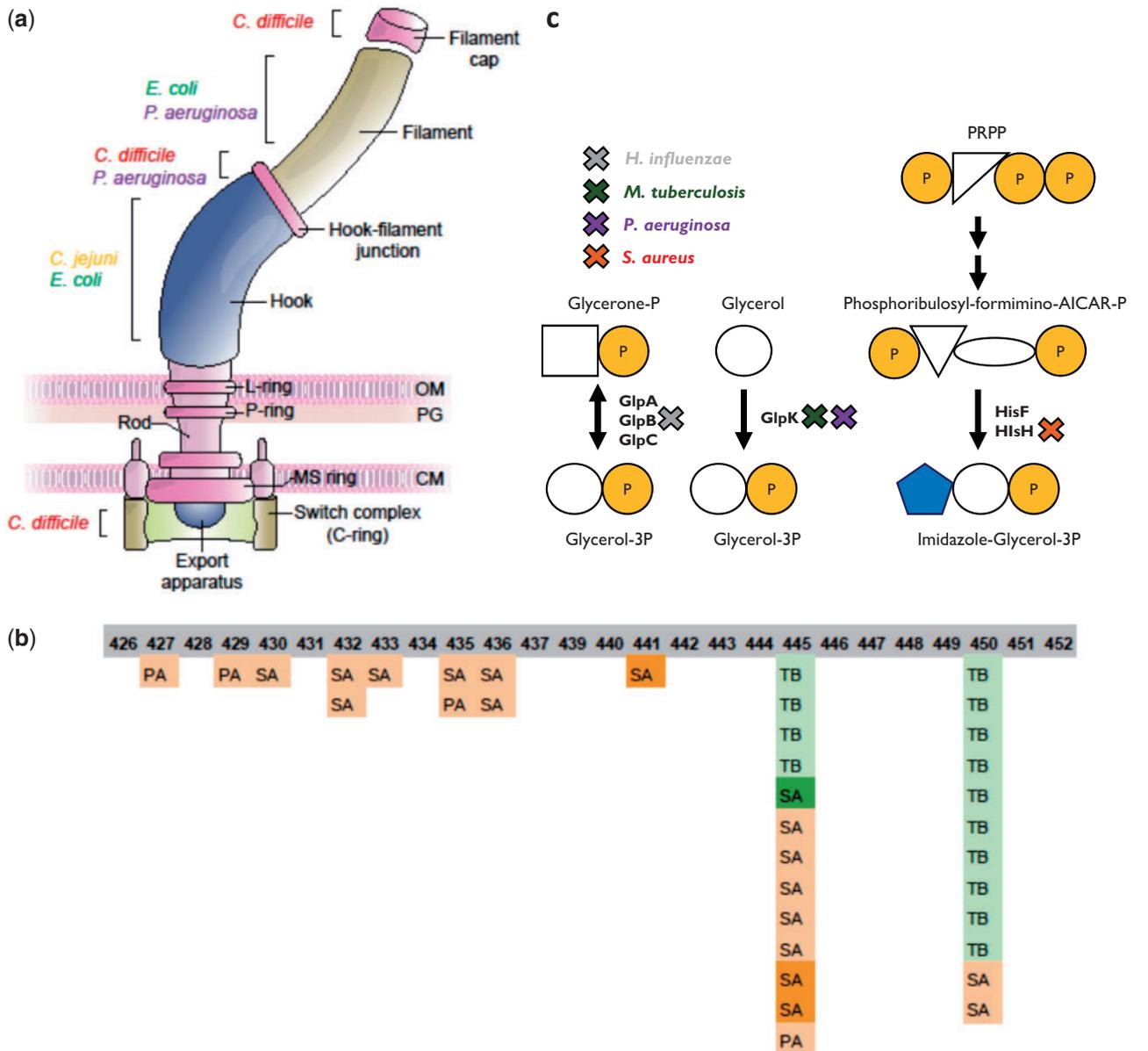


Fig. 5. Convergent adaptation of different pathogens is evident at the pathway and functional levels (a) Flagella pathway. A schematic illustration of the bacterial flagella. Parts that include genes that underwent adaptive change or loss of function in different bacterial species are denoted. (b) Rifampicin resistance. Locations of mutations in the RRDR of RNAP in TB, PA, and SA strains are denoted. The numbers denote the corresponding residues on the RpoB of TB, which were affected by the mutations. Color indicates rifampicin treatment: dark green—confirmed treatment by rifampicin; bright green—likely treated by rifampicin based on conventional medical indications for the use of the drug; light orange—likely not treated by rifampicin based on conventional medical indications for the use of the drug; dark orange—not treated by rifampicin. (c) Glycerol-3-phosphate synthesis. Genes involved in glycerol-3-phosphate synthesis that underwent change/loss of function in a large fraction of the strains in different bacterial species. X marks the gene encoding the relevant protein that underwent adaptive change or loss of function, with the color of X indicating the species. PRPP: phosphoribosyl pyrophosphate. Names of bacterial species are as in figure 1b. Shapes represent chemical compounds: orange circle—phosphate, blank circle—glycerol, blank square—glycerone, blue pentagon—imidazole, blank triangle—ribose or ribulose, and blank ellipse—formimino group + AICAR group.

5, Supplementary Material online). Genes related to flagellar assembly were in fact adaptively changed in all flagellated species except ST (fig. 5a) and in a remarkable 70–90% of strains. Mutations negatively affecting the function of these genes may cause the loss of the immunogenic flagella complex and a switch to nonmotile lifestyles, which has previously

been demonstrated for some of the species during host infection (Marvig Sommer, et al. 2015; Bloomfield et al. 2018; Aziz et al. 2019).

N-Acetylneuraminyl transferase activity is required in both HI and CJ for sialylation of lipooligosaccharide (LOS), a key process in the pathogenesis of both these species. Its

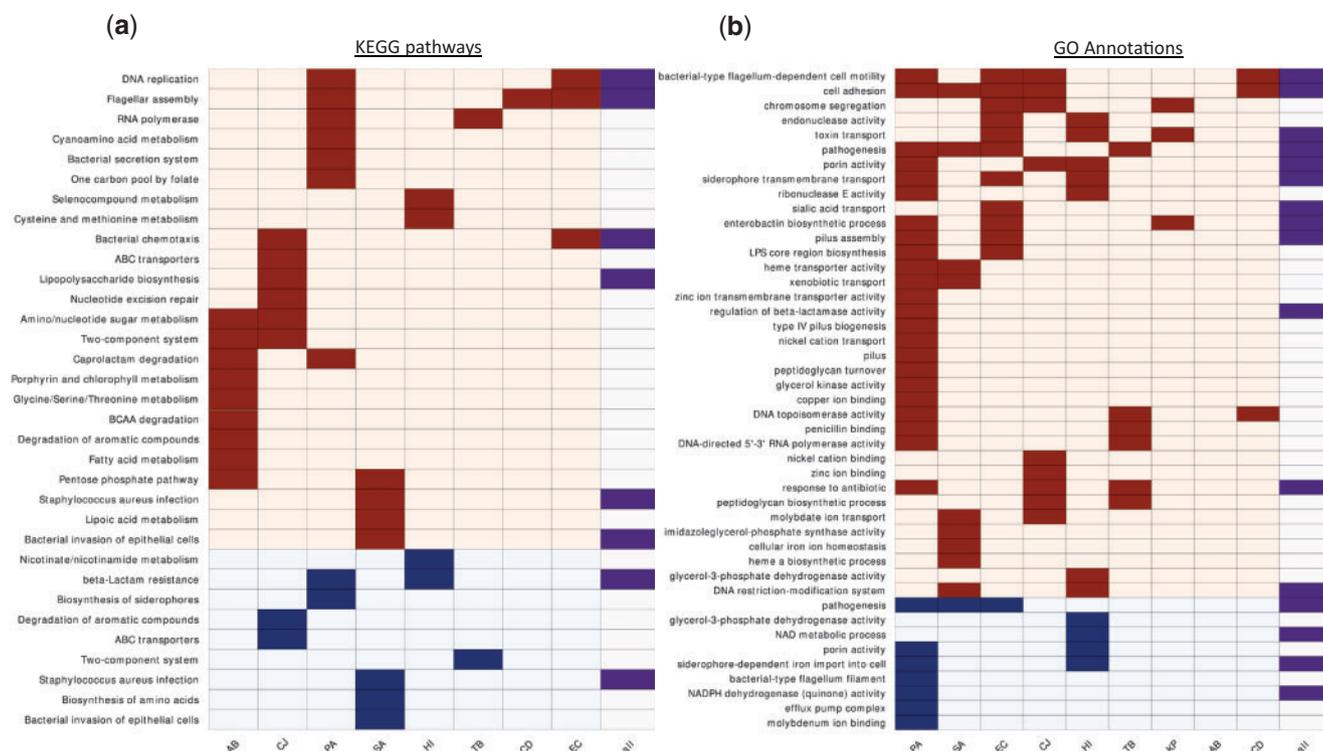


Fig. 6. Selected KEGG pathways and GO annotations that were overrepresented among genes undergoing change or loss of function in different bacterial species. (a) Each row represents a different pathway: Rows with light orange background—overrepresented pathways among changed genes; rows with light blue background—overrepresented pathways among genes undergoing loss of function. Dark boxes indicate that the pathway was overrepresented in the species appearing on the column. The rightmost column indicates whether the pathway was overrepresented when considering all strains of all species together. (b) As in (a) for GO annotations. Names of bacterial species are as in [figure 1b](#).

experimental loss *ex vivo* has been linked to increased immunogenicity and decreased resistance to mammalian serum ([Guerry et al. 2000](#)). *neuA* is the only gene related to this activity, and surprisingly it was mutated in most of the HI and CJ strains, most often in particular positions. *hddC*, a key component in the synthesis of an additional capsular polysaccharide, was also mutated in all CJ strains. This reaffirms and generalizes the results of [Pettigrew et al. \(2018\)](#), who noted sialyltransferase mutations in HI during infection of patients suffering from chronic obstructive pulmonary disease.

Genes related to pore formation underwent adaptive loss of function in most Gram-negative bacteria in our data, for which it is known that small hydrophilic drugs, such as β -lactams, use the pore-forming porins to gain access to the cell interior ([Ghai I and Ghai S 2018](#)). This generalizes results by previous studies, included in our data, which identified mutations in specific genes related to pore formation occurring during host adaptation in specific bacterial species ([Porse et al. 2017](#); [Sherrard et al. 2017](#); [Gaibani et al. 2018](#); [Pettigrew et al. 2018](#)). The consistent mutations in these genes across many species may be related to decreased susceptibility to these antibiotics.

Further Convergent Changes in KEGG Pathways Are Evident in Genes Undergoing Nonadaptive Changes or Loss of Function

The previous analyses were restricted to genes that we determined as undergoing adaptive change or loss of function in

the various species. In the next analysis, we assigned KEGG pathways to all genes that underwent change or loss of function in our data to assess whether we can detect convergent adaptation at the pathway level. We developed a statistical framework to assess which pathways were affected more than expected at random in the strains of each bacterial species and in all strains across all species jointly (Materials and Methods). This analysis was carried out to detect statistically significantly overrepresented pathways among genes undergoing change and among genes undergoing loss of function. Most overrepresented pathways were unique to single species, but some pathways stood out in multiple species ([fig. 6a](#) and [supplementary table 6, Supplementary Material online](#)). No pathway was overrepresented among genes undergoing change or loss of function at the whole-data set level alone. Most prominently, the *Flagellar assembly* and *Bacterial chemotaxis* pathways were overrepresented among changed genes in the same species as in the previous analysis, supporting a possible shift to nonmotility during infection. Pathways related to amino acid synthesis and metabolism were overrepresented among genes undergoing change and among genes undergoing loss of function in multiple species, a finding which could be related to adaptation to the amino acid-rich host environment.

Some of the other pathways that were overrepresented in multiple species are likely related to antibiotic pressures ([fig. 6a](#) and [supplementary table 6, Supplementary Material online](#)). For example, *Beta-lactam resistance* and *Two-*

component system, which are related to β -lactam antibiotics and colistin, respectively, with resistance to the latter emerging in AB via the PmrAB two-component system. Another remarkable example is the RNA polymerase (RNAP) complex. We detected adaptive changes in RNAP in both TB and PA strains (fig. 6a), as well as in the β -subunit of SA (supplementary table 2, Supplementary Material online). Close inspection of these mutations revealed many missense mutations within the rifampicin-resistance determining region (RRDR, supplementary fig. 13, Supplementary Material online) (Telenti et al. 1993; Andre et al. 2017), which has also been linked to decreased sensitivity to other antibiotics (Guérillot et al. 2018). Indeed, most PA and SA-infected patients with RRDR mutations in our data have no evidence of rifampicin exposure (fig. 5b), raising the possibility that these mutations might be selected due to treatment by other drugs. For in-depth analysis of RNAP mutations, see Supplementary Material online.

Common Changes in the Uptake of Iron and Other Metals

We performed a similar analysis to assess which GO annotations were statistically significantly overrepresented within the gene groups that underwent change or loss of function in the strains of each species and in all strains across all species jointly (fig. 6b and supplementary table 7, Supplementary Material online). We identified various GO annotations as overrepresented among genes undergoing change and among genes undergoing loss of function in multiple species, including GO annotations related to cell surface structures, such as flagella, pili, lipopolysaccharide (LPS), and peptidoglycans, alongside several GO annotations related to iron metabolism. Genes related to uptake or synthesis of siderophores were changed or underwent loss of function in most Gram-negative species, suggesting a possible shift from siderophore-based to heme-based iron scavenging or the development of cheating behavior, as has been shown for PA (Marvig et al. 2014; Andersen et al. 2015). Surprisingly, we found that the *phuR*, *hasR*, and *isdB* heme uptake systems were also mutated in some PA and SA strains. Since *phu* operon mutations in PA have been shown to lead to increased expression of the *phu* heme uptake system during chronic infections (Marvig et al. 2014), it is possible that these mutations do not impair heme uptake. Nevertheless, the *phu* system has been shown to be detrimental in high heme concentrations, so mutations impairing heme uptake may also confer an advantage in the host environment (Andersen et al. 2018). Heme biosynthesis was additionally overrepresented among changed genes in SA, supporting the exogenous uptake of this product. This result is also in line with overrepresentation among changed genes of the KEGG pathway *Porphyrin and chlorophyll metabolism* in AB (fig. 6a). Although iron metabolism was the most prominent, GO annotations related to the metabolism and uptake of other trace elements were also overrepresented among changed genes during the infection of multiple pathogens, including molybdenum, zinc, nickel, and copper, a finding which may suggest that modulation of metal scavenging may be a general adaptive mechanism during infections.

Many GO annotations were again linked to antibiotic resistance. Among those stood out the changes in glycerol-3-phosphate metabolism. A recent paper by Bellerose et al. (2019) revealed that changes in the TB glycerol kinase *glpK* reduce sensitivity to multiple antibiotics. In our data, we detected not only *glpK* as adaptively changed in TB but also several GO annotations related to glycerol-3-phosphate metabolism in three additional species. These results could indicate that changes in glycerol-phosphate metabolism may lead to reduced sensitivity to antibiotics in additional species, a possible global mechanism for reduced antibiotic sensitivity. Furthermore, all processes have similar products, possibly implicating the decrease in glycerol-3-phosphate itself as the adaptive process, as hypothesized by Bellerose et al. for TB (Bellerose et al. 2019) (fig. 5c). Mutations in the TB *glpK* during host adaptation were also noted by Trauner et al. (2017).

As noted above, many genes undergoing adaptive change or loss of function encode known virulence factors. Our analysis of GO annotations further supported this conclusion, with multiple annotations related to bacterial virulence overrepresented among genes undergoing change and among genes undergoing loss of function (fig. 6b). This may indicate decreased virulence during chronic infections or a shift from acute virulence to chronic virulence, as has been previously suggested for PA (Hauser et al. 2011; Winstanley et al. 2016). SA and *Burkholderia pseudomallei* have also been suggested to attenuate their virulence during certain chronic infections (Viberg et al. 2017; Suligoy et al. 2018).

To assess our conclusions, we applied our statistical framework to genes identified by the de novo assembly-based pipeline as undergoing change or loss of function, in order to identify among them genes that were adaptively changed or adaptively underwent loss of function. We found that many of our conclusions stay firm for this set of genes (Supplementary Material online).

Discussion

In this work, we took a global approach to study within-host adaptation of bacterial pathogens. We investigated 11 different taxa of bacterial pathogens, representing much of the variety of bacterial pathogens causing chronic disease in humans. We showed that these pathogens undergo genetic changes during infections, including loss of function mutations. Many identical changes were discovered independently in different strains belonging to the same bacterial species, infecting different patients, indicating adaptive evolution.

Although we found little orthologous relationship between genes undergoing adaptive change or loss of function in different species, analysis of meta-categories including KEGG pathways and GO annotations demonstrated that genes passing the statistical threshold for determining adaptive change or loss of function in different pathogens were often related to similar pathways, structures, and biological functions. Furthermore, when broadening the scope of our analysis beyond genes undergoing adaptive change or loss of function to genes which did not pass the statistical threshold, we unveiled pathways and functions that were consistently

affected in many strains of unrelated species via different genes. This demonstrates remarkable convergent evolution between different pathogens during host infection, despite distinct lifestyles and pressures.

Previous studies examined adaptation in the context of a single species and a single clinical presentation. In this study, we included data of multiple bacterial species, from different tissues, causing different clinical presentations, treated with different antimicrobial drugs, originating from different geographical regions and studied by variable laboratory techniques. DNA was extracted from single colonies, from multiple colonies or even directly from patient samples in the different studies. Additionally, the different studies employed different sequencing machines and approaches (long-read sequencing, single-end sequencing, paired-end sequencing, etc.). The joint analysis of these different data limits our ability to detect specific signals that may relate to specific clinical presentations associated with different symptoms or treatments. Rather, our approach allows us to detect global convergent adaptation, which is not limited to a specific scenario. Indeed, we detected substantial convergence despite these differences by using large numbers of patients and stringent statistical frameworks. Expansion of our framework with additional data could allow the detection of further convergent adaptive trends and overcoming nongeneralizable results for species with strains originating from a single study. Also, although we focused on protein-coding genes appearing in reference proteomes, applying similar methodologies to additional genomic elements/regions could lead to additional insights regarding within-host evolution (Khademi et al. 2019). It should be also possible to study subgroups of patients from our database to focus on specific treatments or clinical presentations associated with specific sets of symptoms.

Adaptive changes occurring in multiple species were frequently associated with common antigens recognized by the human immune apparatus, including flagellar filament, substrate of TLR5; LOS/LPS, substrates of TLR4; peptidoglycans, substrates of TLR2; iron acquisition systems, considered highly immunogenic (Wandersman and Delepelaire 2004). Many adaptive changes also relate to antibiotic persistence/resistance, especially evident in RNAP. Rifampicin is an antimicrobial agent acting by the inhibition of RNAP. Bacterial resistance to rifampicin is common and usually arises by mutations in the residues of the rifampicin binding site of RNAP (Campbell et al. 2001). The region where the mutations most commonly arise has been termed the RRDR (Telenti et al. 1993). We showed that RNAP accumulates mutations in PA, SA, and TB during infections, frequently located in the RRDR (supplementary fig. 13, Supplementary Material online). These mutations appear in many strains with no evidence of exposure to rifampicin, although we cannot rule out exposure to rifampicin administered for concurrent infections by other pathogens. Strong selective forces acting on genes linked to antibiotic resistance had clear mutational patterns, with mutations at the same positions recurring independently in different strains (fig. 5b). Similar patterns identified in genes that are not currently linked with antibiotic persistence/resistance, may hint at their involvement in these

processes (Supplementary Material online). This is supported by our independent detection of *glpK* frameshift mutations, which were recently linked to antibiotic resistance in TB (Bellerose et al. 2019) (fig. 5c). We suggest that similar mechanisms may be involved in antibiotic resistance in additional bacterial species. It is not surprising that the innate immune system and antibiotic treatment seem to be major drivers of convergent within-host adaptation in different species, as both exert powerful general selective pressures on bacterial pathogens that are usually not confined to specific species.

The adaptive mutations we detect arise repeatedly in independent infections, indicating that they may confer advantages within the host environment. It is fascinating to ponder why these mutations arise repeatedly in organisms that undergo patient-to-patient transmission rather than remain fixed across the species. Bacteria encounter stringent bottlenecks when infecting new patients (Moxon and Kussell 2017). We hypothesize that adaptive changes become disadvantageous in the context of these bottlenecks and are therefore subject to repeated cycles of negative and positive selection. The mutations might persist within small subpopulations or independently arise within each infective cycle as bacteria re-encounter their chronic niches. This theory is supported by genes undergoing adaptive change or loss of function that are of vital importance for acute virulence, including the ESX-1 secretion system of TB (Gröschel et al. 2016) as well as the flagella in CJ (Black et al. 1988).

Changes in virulence factors during infection have also been hypothesized to relate to virulence attenuation during chronic infections and to a shift from acute virulence to chronic virulence (Hauser et al. 2011; Winstanley et al. 2016). We believe these interpretations are not mutually exclusive, as bacteria employing acute infection strategies can be extremely virulent, leading to death within hours or days. The loss of factors required for acute virulence disarms the pathogen of these deadly weapons and underlies a shift toward lesser virulence (Furukawa et al. 2006). Indeed, pathogens isolated from established chronic infections are less cytotoxic than earlier isolates from the same patients (Gellatly and Hancock 2013). Therefore, the change or loss of function of acute virulence factors may underlie both an attenuation of virulence and a shift toward chronic virulence.

Finally, we believe our results to be an important resource for screening candidate antivirulence targets. Antivirulence therapy is an emerging strategy of antimicrobial treatment employing agents targeting virulence factors to reduce the virulence of infecting pathogens rather than eliminating them, thus avoiding the exertion of selective pressures for resistance. A major limitation of this paradigm is that targeting virulence factors, which are beneficial in vivo, leads to reduced fitness of the pathogen and creates evolutionary constraints inducing resistance (Allen et al. 2014). Our results offer a rich resource of antivirulence targets that are non-beneficial in vivo and could therefore prove to be effective targets for disarming infecting bacteria. Conversely, targeting virulence factors that frequently accumulate genetic changes during infections may lead to development of antibiotic resistance, as evolution of those genes is possibly

less constrained within the host environment (Rezzoagli et al. 2018). The rich resource of data that we provide establishes the foundation to further study the basic principles of within-host adaptation and their exploitation in antibacterial therapy.

Materials and Methods

Data Organization

Compiling the Database

Whole-genome sequencing data of bacterial pathogens during infection were curated by searching NCBI's databases for relevant studies using several search terms, such as "within-host adaptation," "intra-host evolution," "infectious episode AND WGS," "longitudinal AND sequencing," "microevolution," "serial isolates," and "genomic adaptation." Relevant studies were defined as ones including whole-genome sequencing data of bacterial isolates sampled in at least two time points during a single infection of at least a single human host with a single bacterial pathogen (table 1, coinfections with other pathogens during the time period were allowed). NCBI Assembly accession numbers were collected when available, and otherwise NCBI SRA Run accession numbers were used. The database can be found in supplementary table 1, Supplementary Material online.

Most of the studies in our database included a single sequenced colony for each sample, and a single sample for each time point, with a subset of the studies including multiple colonies or samples for each time point. Another subset of the studies included data directly sequenced from patient samples, without the isolation of a clone beforehand. For simplicity, we refer to the sequencing results of a single colony as an isolate throughout the paper.

Downloading and Processing the Data

Data were downloaded using the SRA Toolkit (Leinonen et al. 2011) and assembled using the SPAdes Assembler v.3.10.1 (Bankevich et al. 2012). When assemblies were already available, they were directly downloaded. Reference or representative genomes were determined using NCBI's E-Dirct command line utilities v.5.80 (Kans 2019). When several reference genomes were available, the closest one to the assemblies was determined by a custom script exploiting QUAST v.4.5 (Gurevich et al. 2013) and Mash v.2.1 (Ondov et al. 2016). We chose the reference genomes where the highest fraction of the genome was aligned to contigs of the different assemblies and the average Mash distance to the assemblies was the smallest. Notably, for most species only a single reference genome was available, and for *E. coli* and for *A. baumannii* one reference genome was found to be the closest to all strains. Furthermore, for each strain the difference between the Mash distances of any two assemblies to the reference genome, was nearly always 0.0.

Determining the Phylogenetic Relationship between Isolates

We developed the TRACE algorithm to determine the phylogenetic relationship between the isolates from each patient as follows (fig. 1c):

- (1) Isolates from each patient are divided to clones based on phylogenetic trees: kSNP v.3.1 (Gardner and Hall 2013; Gardner et al. 2015) is used to construct core single-nucleotide polymorphism (SNP) matrices based on the assemblies of the isolates of each patient and up to 100 complete genomes for the relevant bacterial species from NCBI's Assembly database. PAUP v.4.0 (Swofford 2003) is then used to construct maximum likelihood trees out of the matrices, picking each consensus tree out of 1,000 bootstrapped samples. Finally, we inspect the location of the isolates on the constructed tree, and consider isolates found on a single clade to originate from the same clone. For additional validation, the MLST profile (Maiden et al. 1998) for each isolate is also predicted using *mlst* v.2.16 (Jolley and Maiden 2010; Seeman 2017) and isolates within the same clone, which differ in their profiles, are split to separate clones (Supplementary Material online).
- (2) In order to determine which isolates may share a direct progenitor–progeny phylogenetic link, TRACE then constructs the phylogenetic relationships between the isolates of each clone. kSNP v.3.1 (Gardner and Hall 2013; Gardner et al. 2015) and *breseq* v.0.32 (Barrick et al. 2014; Deatherage and Barrick 2014) are utilized to construct SNP matrices between the isolates and custom scripts are used to assess the least parsimonious path possible between the isolates in which each isolate is derived from an earlier time point isolate. In cases where two isolates of the same time point are available, an artificial common ancestor is constructed as the intersection of the SNP profiles of the two isolates, and trees including this isolate are additionally inspected. Clones with three or more isolates sampled at the same time point were excluded. Finally, TRACE defines pairs of progenitor and progeny isolates in each tree and assigns each pair a confidence level based on its frequency among the 1% highest scoring trees for that clone. We assessed the performance of the TRACE pipeline using both simulated and empirical data, verifying its high accuracy under a broad range of parameters (Supplementary Material online).

Analysis of Genetic Changes

Determining the Variation between Isolates

Breseq v.0.32 (Barrick et al. 2014; Deatherage and Barrick 2014) was used to call genetic and genomic differences between each isolate and the respective reference genome of the species by mapping the sequencing reads to the reference genome and identifying discrepancies between the aligned reads and the reference genome sequence. Sequence repeats were detected using *Vmatch* package (<http://www.vmatch.de>, last accessed November 16, 2020) and excluded from all analyses. The differences between the isolate and the reference genome were then compared within each progenitor–progeny pair using the COMPARE function of *gdttools*

(Deatherage and Barrick 2014) to determine genetic and genomic variations within the pair.

Defining Changed Genes and Genes Undergoing Loss of Function

We used SnpEff v.4.1 (Cingolani et al. 2012) to classify the impact of the above differences as low (e.g., synonymous substitutions), moderate (e.g., nonsynonymous substitutions), high (e.g., frameshift mutations), or modifier (e.g., mutations upstream to the coding sequence [CDS]). A gene with variations of moderate effect that were found in the progeny strain and not the progenitor strain was determined as changed within the progenitor–progeny pair. A gene with high-impact variations that were found in the progeny strain and not the progenitor strain was determined both as undergoing change and as undergoing loss of function within the progenitor–progeny pair. Genes with high-impact mutations in the progenitor strain were excluded as we considered them as putative pseudogenes.

A gene that underwent change or loss of function in at least one of progenitor–progeny isolate pairs of a strain was determined as a gene that underwent change or loss of function, respectively, in that strain. Additionally, each gene was assigned a score between 0.0 and 1.0 that reflects its loss of function across all progenitor–progeny isolate pairs of the strain. A similar score was determined for changed genes. The score is defined as the fraction of progenitor–progeny pairs in which the gene was affected out of all possible pairs. Possible pairs are defined as all pairs except the ones where the progenitor isolate is downstream in the tree (based on the results of TRACE) to a progenitor–progeny pair in which the gene was affected. For example, in case of two progenitor–progeny pairs $a \rightarrow b$ and $b \rightarrow c$, and a gene undergoing loss of function between a and b , the pair $b \rightarrow c$ should not be regarded as a possible pair for loss of function of this gene because the gene underwent loss of function upstream in the tree. Note that change or loss of function events that are predicted by TRACE to occur between a progenitor isolate and a common theoretical ancestor of two progeny isolates sampled at the same time point, are counted only once despite appearing in two progenitor–progeny pairs.

Statistical Framework to Assess Adaptive Change and Loss of Function

After assigning scores per strain to genes that underwent loss of function, we assigned gene scores at the species level. Each gene g in the reference genome of each species was assigned a score, L_g , which is the sum of its per-strain scores across all strains in which it underwent loss of function. This score reflects the number of strains in which the gene underwent loss of function, while taking into account the fraction of progenitor–progeny pairs in which these events occurred per strain. In further computations, we take the floor function of L_g , which is the greatest integer less than or equal to L_g ($\lfloor L_g \rfloor$), as the number of strains in which the gene underwent loss of function.

To evaluate if the repeated loss of function of a gene in several strains was statistically significant, we determined whether $\lfloor L_g \rfloor$ is greater than expected at random by random simulations. To create the random simulations, we first assigned each strain i a probability of loss P_i :

$$P_i = \frac{\text{sum of scores of genes undergoing loss of function in strain } i}{\text{number of genes in genome}}$$

Note that here also we do not simply count the number of genes that underwent loss of function in the strain, but take into account the fractions of progenitor–progeny pairs in which they underwent loss of function (their scores per strain). We then simulated loss of function in 10^6 genes, determining for each gene at random whether it did or did not undergo loss of function. The probability for each simulated gene to undergo loss of function in each strain is equal to the strain's loss probability, P_i . We counted the number of strains in which each gene underwent loss of function in the simulations and described the distribution of the number of genes that underwent loss of function in each number of strains (supplementary fig. 7, Supplementary Material online). Finally, based on this distribution, we treated the fraction of genes that underwent loss of function in at least $\lfloor L_g \rfloor$ strains based on the simulations as the P value for a gene g whose total score across all strains for loss of function was $\lfloor L_g \rfloor$. Genes with adjusted P value ≤ 0.1 after correcting for testing of multiple hypotheses by the Benjamini–Hochberg method (Benjamini and Hochberg 1995) were determined to undergo loss of function above random expectation and their loss of function was therefore considered as putatively adaptive. We conducted a similar statistical test for changed genes to detect adaptively changed genes, replacing L_g with C_g , a similar score for each gene that was determined as changed.

Computation of Mutation Rates

Mutation rates were calculated for strains with isolates sampled at three or more time points. For each isolate, we extracted the time point in which it was isolated and counted the number of SNPs that do not appear in isolates sampled at the very first time point for the strain. We then fitted a second-degree polynomial for each strain based on the time points and numbers of SNPs of the different isolates and extracted its first derivative (mutation rate) and second derivative (rate of change for the mutation rate).

Calculating dN/dS Ratios

dN/dS ratios were calculated using the raw counts of non-synonymous/synonymous mutations across all progenitor–progeny pairs of each strain.

Determining Hypermutator Clones

We defined hypermutator clones as clones with at least one isolate with a complete deletion or loss of function mutation in one of the genes of the DNA mismatch repair system, where mutations have been shown to confer a strong or a

very strong hypermutator effect in *E. coli*. We also included isolates with mutations in one of the genes of the GO system, where mutations have been shown to confer a weak to strong hypermutator effect in *E. coli* (Oliver and Mena 2010). For *M. tuberculosis*, we also included mutations in the *nucS* gene, which is required for mismatch repair in this organism (Castañeda-García et al. 2017). Orthologs in the different species were determined using BlastP v.2.4.0+.

Determining the Similarity in the Lists of Genes Undergoing Change or Loss of Function between the Different Isolates

To determine the similarity between different strains of each species, we compared their lists of changed genes. We computed the Jaccard index between the changed gene lists of each two strains in each species and finally calculated the mean of this measure across all strain pairs. A similar calculation was performed for the lists of genes that underwent loss of function.

Characterizing the Genomic Variations

Assessing Mechanisms of Adaptive Loss of Function

We calculated the relative frequency of different mechanisms underlying adaptive loss of function in the different species as determined by SnpEff v.4.1 (Cingolani et al. 2012). The number of events of genes undergoing adaptive loss of function caused by each mechanism was normalized by the number of progenitor–progeny pairs in the respective species.

Examining %GC in Frameshift Mutations

In order to characterize the nucleotides that were inserted or deleted in frameshift mutations, we compared the number of observed inserted/deleted G/C bases with the expected value based on the %GC of the entire genome by a binomial test.

Confirming INDELS by the De Novo Assembly-Based Pipeline

Frameshift mutations during the infection are caused by INDELS in the CDS of the gene. Since we found that frameshifts constitute the major mechanism underlying loss of function during the infection, we validated INDEL mutations predicted by our pipeline using the de novo assembly-based pipeline (Supplementary Material online). To test the presence of INDELS in the de novo assembly-based pipeline, we extracted the gene in which each INDEL was found and searched for CDS with the same annotation in the progenitor and progeny isolates for the respective pair in which the INDEL was determined. We chose cases where only a single CDS in both the progenitor and progeny isolates had the same annotation as the gene hosting the INDEL. We searched for matches between the progenitor and progeny sequences, where five nucleotides before and after the site of the predicted INDEL remained identical, and the INDEL site itself underwent the expected insertion or deletion. The result may be an underestimation due to SNPs or other variations in the preceding or succeeding nucleotides.

Detecting Genomic Regions That Are Enriched with Genes Undergoing Adaptive Change or Loss of Function

We performed hypergeometric tests for genomic regions of size 10,000 base pairs to detect regions that are enriched with genes undergoing adaptive change or loss of function in their respective genomes by considering the number of genes in the region, the number of genes undergoing adaptive change or loss of function, and the respective numbers for the entire genome. Regions with adjusted P value ≤ 0.1 after correction for testing of multiple hypotheses using the Benjamini–Hochberg method (Benjamini and Hochberg 1995) were considered statistically significantly enriched with genes that underwent adaptive change or loss of function.

Calling Orthologs Undergoing Adaptive Change or Loss of Function

Determination of orthologs between different species was based on reciprocal BLAST, defining orthologs as the best BLAST hits with an extremely lenient E -value cutoff of 100.0, with the aligned region covering at least 30% of the query and the BLAST score being at least 10% of the BLAST score of a perfect match for the query.

Analysis of Pathways and Functions Affected by the Mutations

We analyzed common pathways and functions affected by mutations in different species by two analyses, the first focused on genes undergoing adaptive change or loss of function and the second included all genes undergoing change or loss of function. For these analyses, we assigned each gene the pathways to which it belongs using KEGG (Kanehisa and Goto 2000) and its GO annotations using QuickGO (Ashburner et al. 2000; Binns et al. 2009; Gene Ontology Consortium 2019).

Assessing KEGG Pathways and GO Annotations for Genes That Underwent Adaptive Change or Loss of Function.

To assess whether genes related to similar pathways or functions underwent adaptive change or loss of function in different species, we counted in how many species each pathway or annotation included genes that underwent adaptive change or loss of function. As a control, we performed 100,000 simulations, where in each simulation we chose random genes as undergoing adaptive loss of function (change) in each species according to the actual number of genes that underwent adaptive loss of function (change) in that species, respectively. The P value for each pathway or annotation was the fraction of simulations where genes of the pathway underwent adaptive loss of function (change) in at least as many species as in the actual results. Finally, we corrected the P values for testing of multiple hypotheses using the Benjamini–Hochberg method (Benjamini and Hochberg 1995). The use of simulations controls for differences in the number of genes in different pathways.

Assessing KEGG Pathways and GO Annotations That Are Overrepresented among Genes Undergoing Change or Loss of Function. It is possible that there are genes in various strains

that underwent change or loss of function that were not determined as adaptive but are still involved in common pathways. To this end, we assessed the enrichment of specific KEGG pathways or GO annotations in the data of all genes undergoing change or loss of function in a strain by simulations. As a basis for the simulations, we used the number of genes N_i undergoing loss of function (change) in each strain i . We carried out 100,000 simulations of all our strains. In each simulation, we randomly chose N_i genes to undergo loss of function (change) in each strain i and counted in how many strains each KEGG pathway or GO annotation was affected. The P value for each pathway or annotation was the fraction of simulations where it underwent loss of function (change) in at least as many strains as in the actual results. Finally, we considered all pathways with adjusted P value ≤ 0.1 after correcting for testing of multiple hypotheses by the Benjamini–Hochberg method (Benjamini and Hochberg 1995) to be overrepresented among genes undergoing loss of function (change). This process was carried out for the strains of each bacterial species and for all the strains across all species jointly. We carried out a similar analysis for GO annotations.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We are grateful to Y. Altuvia and I. Rosenshine for fruitful discussions and invaluable advice. We thank M. Barsheshet and A. Bar for their support and comments and R. Hershberg, E. Bolotin, T. Pupko, U. Gophna, A. Stern, and R. Pascale for their helpful advice and assistance. We thank Y. Yadegari for the illustration in figure 5a. This work was supported by the Israel Science Foundation, administered by the Israeli Academy for Sciences and Humanities (grant 876/17). Y.E.G. is partially supported by fellowships of the Hoffman Program, the Foulkes Foundation, and the Data Sciences program of the Planning and Budgeting Committee of the Israel Council for Higher Education.

Author Contributions

All authors contributed extensively to the work presented in this paper. H.M. and Y.E.G. conceived the study; Y.E.G. compiled the database, analyzed the data, and wrote the manuscript; H.M. supervised the work and wrote the manuscript.

Data Availability

Accession numbers for all relevant sequencing data are listed in supplementary table 1, Supplementary Material online. The scripts used for the downloading of the data, the processing of the data, the assessment of within-host variations, the TRACE pipeline, and the statistical framework are available through the following github repository: <https://github.com/YairGatt/WithinHostAdaptation> (last accessed November 16, 2020).

References

- Alamro M, Bidmos FA, Chan H, Oldfield NJ, Newton E, Bai X, Aidley J, Care R, Mattick C, Turner DPJ, et al. 2014. Phase variation mediates reductions in expression of surface proteins during persistent meningococcal carriage. *Infect Immun*. 82(6):2472–2484.
- Allen RC, Papat R, Diggle SP, Brown SP. 2014. Targeting virulence: can we make evolution-proof drugs? *Nat Rev Microbiol*. 12(4):300–308.
- Andersen SB, Ghoul M, Marvig RL, Lee Z-B, Molin S, Johansen HK, Griffin AS. 2018. Privatisation rescues function following loss of cooperation. *eLife* 7:e38594.
- Andersen SB, Marvig RL, Molin S, Krogh Johansen H, Griffin AS. 2015. Long-term social dynamics drive loss of function in pathogenic bacteria. *Proc Natl Acad Sci U S A*. 112(34):10756–10761.
- Andre E, Goeminne L, Cabibbe A, Beckert P, Kabamba Mukadi B, Mathys V, Gagneux S, Niemann S, Van Ingen J, Cambau E. 2017. Consensus numbering system for the rifampicin resistance-associated rpoB gene mutations in pathogenic mycobacteria. *Clin Microbiol Infect*. 23(3):167–172.
- Ankrum A, Hall BG. 2017. Population dynamics of *Staphylococcus aureus* in cystic fibrosis patients to determine transmission events by use of whole-genome sequencing. *J Clin Microbiol*. 55(7):2143–2152.
- Arias CA, Panesso D, McGrath DM, Qin X, Mojica MF, Miller C, Diaz L, Tran TT, Rincon S, Barbu EM, et al. 2011. Genetic basis for in vivo daptomycin resistance in enterococci. *N Engl J Med*. 365(10):892–900.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. *Nat Genet*. 25(1):25–29.
- Azarian T, Daum RS, Petty LA, Steinbeck JL, Yin Z, Nolan D, Boyle-Vavra S, Hanage WP, Salemi M, David MZ. 2016. Intrahost evolution of methicillin-resistant *Staphylococcus aureus* USA300 among individuals with reoccurring skin and soft-tissue infections. *J Infect Dis*. 214(6):895–905.
- Aziz A, Sarovich DS, Nosworthy E, Beissbarth J, Chang AB, Smith-Vaughan H, Price EP, Harris TM. 2019. Molecular signatures of non-typeable *Haemophilus influenzae* lung adaptation in pediatric chronic lung disease. *Front Microbiol*. 10:1622.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 19(5):455–477.
- Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ, Knoester DB, Reba A, Meyer AG. 2014. Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics*. 15(1):1039.
- Bartell JA, Sommer LM, Haagenen JAJ, Loch A, Espinosa R, Molin S, Johansen HK. 2019. Evolutionary highways to persistent bacterial infection. *Nat Commun*. 10(1):629.
- Beaume M, Köhler T, Greub G, Manuel O, Aubert JD, Baerlocher L, Farinelli L, Buckling A, Van Delden C; The Swiss Transplant Cohort Study. 2017. Rapid adaptation drives invasion of airway donor microbiota by *Pseudomonas* after lung transplantation. *Sci Rep*. 7(1):40309.
- Bellerose MM, Baek S-H, Huang C-C, Moss CE, Koh E-I, Proulx MK, Smith CM, Baker RE, Lee JS, Eum S, et al. 2019. Common variants in the glycerol kinase gene reduce tuberculosis drug efficacy. *mBio* 10(4):e00663.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 57(1):289–300.
- Bianconi I, D'Arcangelo S, Esposito A, Benedet M, Piffer E, Dinnella G, Gualdi P, Schinella M, Baldo E, Donati C, et al. 2019. Persistence and microevolution of *Pseudomonas aeruginosa* in the cystic fibrosis lung: a single-patient longitudinal genomic study. *Front Microbiol*. 9:3242.
- Bianconi I, Jeukens J, Freschi L, Alcalá-Franco B, Facchini M, Boyle B, Molinaro A, Kukavica-Ibrulj I, Tümmler B, Levesque RC, et al. 2015. Comparative genomics and biological characterization of

- sequential *Pseudomonas aeruginosa* isolates from persistent airways infection. *BMC Genomics*. 16(1):1105.
- Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. 2009. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25(22):3045–3046.
- Black RE, Levine MM, Clements ML, Hughes TP, Blaser MJ. 1988. Experimental *Campylobacter jejuni* infection in humans. *J Infect Dis*. 157(3):472–479.
- Bloemendaal ALA, Brouwer EC, Fluit AC. 2010. Methicillin resistance transfer from *Staphylococcus epidermidis* to methicillin-susceptible *Staphylococcus aureus* in a patient during antibiotic therapy. *PLoS One* 5(7):e11841.
- Bloomfield SJ, Midwinter AC, Biggs PJ, French NP, Marshall JC, Hayman DTS, Carter PE, Thornley C, Yap R, Benschop J. 2018. Long-term colonization by *Campylobacter jejuni* within a human host: evolution, antimicrobial resistance, and adaptation. *J Infect Dis*. 217(1):103–111.
- Brodrick HJ, Raven KE, Harrison EM, Blane B, Reuter S, Török ME, Parkhill J, Peacock SJ. 2016. Whole-genome sequencing reveals transmission of vancomycin-resistant *Enterococcus faecium* in a healthcare network. *Genome Med*. 8(1):4.
- Brodrick HJ, Raven KE, Kallonen T, Jamrozny D, Blane B, Brown NM, Martin V, Török ME, Parkhill J, Peacock SJ. 2017. Longitudinal genomic surveillance of multidrug-resistant *Escherichia coli* carriage in a long-term care facility in the United Kingdom. *Genome Med*. 9(1):70.
- Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, Reacher M, Haworth CS, Curran MD, Harris SR, et al. 2013. Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study. *Lancet* 381(9877):1551–1560.
- Bullman S, Peadarallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, Neuberger D, Huang K, Guevara F, Nelson T, et al. 2017. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. *Science* 358(6369):1443–1448.
- Cameron DR, Jiang JH, Abbott IJ, Spelman DW, Peleg AY. 2015. Draft genome sequences of clinical daptomycin-nonsusceptible methicillin-resistant *Staphylococcus aureus* strain APS211 and its daptomycin-susceptible progenitor APS210. *Genome Announc*. 3(3):e00568.
- Cameron DR, Ward DV, Kostoulias X, Howden BP, Moellering RC, Eliopoulos GM, Peleg AY. 2012. Serine/threonine phosphatase Stp1 contributes to reduced susceptibility to vancomycin and virulence in *Staphylococcus aureus*. *J Infect Dis*. 205(11):1677–1687.
- Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, Darst SA. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell* 104(6):901–912.
- Cannatelli A, Di Pilato V, Giani T, Arena F, Ambretti S, Gaibani P, D'Andrea MM, Rossolini GM. 2014. 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(8):4399–4403.
- Castañeda-García A, Prieto AI, Rodríguez-Beltrán J, Alonso N, Cantillon D, Costas C, Pérez-Lago L, Zegeye ED, Herranz M, Płociński P, et al. 2017. A non-canonical mismatch repair pathway in prokaryotes. *Nat Commun*. 8(1):14246.
- Chen C-J, Huang Y-C, Chiu C-H. 2015. Multiple pathways of cross-resistance to glycopeptides and daptomycin in persistent MRSA bacteraemia. *J Antimicrob Chemother*. 70(11):2965–2972.
- Chung H, Lieberman TD, Vargas SO, Flett KB, McAdam AJ, Priebe GP, Kishony R. 2017. Global and local selection acting on the pathogen *Stenotrophomonas maltophilia* in the human lung. *Nat Commun*. 8(1):14078.
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; Iso-2; Iso-3. *Fly* 6(2):80–92.
- Clark TG, Mallard K, Coll F, Preston M, Assefa S, Harris D, Ogowang S, Mumbowa F, Kirenga B, O'Sullivan DM, et al. 2013. Elucidating emergence and transmission of multidrug-resistant tuberculosis in treatment experienced patients by whole genome sequencing. *PLoS One* 8(12):e83012.
- Coll F, Harrison EM, Toleman MS, Reuter S, Raven KE, Blane B, Palmer B, Kappeler ARM, Brown NM, Török ME, et al. 2017. Longitudinal genomic surveillance of MRSA in the UK reveals transmission patterns in hospitals and the community. *Sci Transl Med*. 9(413):eaak9745.
- Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, Galagan J, Niemann S, Gagneux S. 2012. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat Genet*. 44(1):106–110.
- Cramer N, Klockgether J, Wrasman K, Schmidt M, Davenport CF, Tümmler B. 2011. Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol*. 13(7):1690–1704.
- Crofts AA, Poly FM, Ewing CP, Kuroiwa JM, Rimmer JE, Harro C, Sack D, Talaat KR, Porter CK, Gutierrez RL, et al. 2018. *Campylobacter jejuni* transcriptional and genetic adaptation during human infection. *Nat Microbiol*. 3(4):494–502.
- Datta G, Nieto LM, Davidson RM, Mehaffy C, Pederson C, Dobos KM, Strong M. 2016. Longitudinal whole genome analysis of pre and post drug treatment *Mycobacterium tuberculosis* isolates reveals progressive steps to drug resistance. *Tuberculosis* 98:50–55.
- Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods Mol Biol*. 1151:165–188.
- Diaz Caballero J, Clark ST, Coburn B, Zhang Y, Wang PW, Donaldson SL, Tullis DE, Yau YCW, Waters VJ, Hwang DM, et al. 2015. Selective sweeps and parallel pathoadaptation drive *Pseudomonas aeruginosa* evolution in the cystic fibrosis lung. *mBio* 6(5):e00981.
- Diaz Caballero J, Clark ST, Wang PW, Donaldson SL, Coburn B, Tullis DE, Yau YCW, Waters VJ, Hwang DM, Guttman DS. 2018. A genome-wide association analysis reveals a potential role for recombination in the evolution of antimicrobial resistance in *Burkholderia multivorans*. *PLoS Pathog*. 14(12):e1007453.
- Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. 2016. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol*. 14(3):150–162.
- Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A*. 88(16):7160–7164.
- Eldholm V, Norheim G, von der Lippe B, Kinander W, Dahle UR, Caugant DA, Mannsåker T, Mengsøel AT, Dyrhol-Riise AM, Balloux F. 2014. Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biol*. 15(11):490.
- Esposito A, Pompilio A, Bettua C, Crocetta V, Giacobazzi E, Fiscarelli E, Jousson O, Di Bonaventura G. 2017. Evolution of *Stenotrophomonas maltophilia* in cystic fibrosis lung over chronic infection: a genomic and phenotypic population study. *Front Microbiol*. 8:1590.
- Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, et al. 2013. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat Genet*. 45(10):1183–1189.
- Feliziani S, Marvig RL, Luján AM, Moyano AJ, Di Rienzo JA, Krogh Johansen H, Molin S, Smania AM. 2014. Coexistence and within-host evolution of diversified lineages of hypermutable *Pseudomonas aeruginosa* in long-term cystic fibrosis infections. *PLoS Genet*. 10(10):e1004651.
- Frydenlund Michelsen C, Hossein Khademi SM, Krogh Johansen H, Ingmer H, Dorrestein PC, Jelsbak L. 2016. Evolution of metabolic divergence in *Pseudomonas aeruginosa* during long-term infection facilitates a proto-cooperative interspecies interaction. *ISME J*. 10(6):1323–1336.
- Furukawa S, Kuchma SL, O'Toole GA. 2006. Keeping their options open: acute versus persistent infections. *J Bacteriol*. 188(4):1211–1217.
- Gaibani P, Campoli C, Lewis RE, Volpe SL, Scaltriti E, Giannella M, Pongolini S, Berlinger A, Cristini F, Bartoletti M, et al. 2018. In vivo

- evolution of resistant subpopulations of KPC-producing *Klebsiella pneumoniae* during ceftazidime/avibactam treatment. *J Antimicrob Chemother.* 73(6):1525–1529.
- Gao W, Chua K, Davies JK, Newton HJ, Seemann T, Harrison PF, Holmes NE, Rhee H-W, Hong J-I, Hartland EL, et al. 2010. Two novel point mutations in clinical *Staphylococcus aureus* reduce linezolid susceptibility and switch on the stringent response to promote persistent infection. *PLoS Pathog.* 6(6):e1000944.
- Gardner SN, Hall BG. 2013. When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS One* 8(12):e81760.
- Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* 31(17):2877–2878.
- Garmendia J, Viadas C, Calatayud L, Mell JC, Martí-Literas P, Euba B, Llobet E, Gil C, Bengoechea JA, Redfield RJ, et al. 2014. Characterization of nontypable *Haemophilus influenzae* isolates recovered from adult patients with underlying chronic lung disease reveals genotypic and phenotypic traits associated with persistent infection. *PLoS One* 9(5):e97020.
- Gellatly SL, Hancock REW. 2013. *Pseudomonas aeruginosa*: new insights into pathogenesis and host defenses. *Pathog Dis.* 67(3):159–173.
- Gene Ontology Consortium. 2019. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47:D330–D338.
- Gerson S, Betts JW, Lucaßen K, Nodari CS, Wille J, Josten M, Göttig S, Nowak J, Stefanik D, Roca I, et al. 2019. Investigation of novel pmrB and eptA mutations in isogenic *Acinetobacter baumannii* isolates associated with colistin resistance and increased virulence in vivo. *Antimicrob Agents Chemother.* 63(3):e01586.
- Ghai I, Ghai S. 2018. Infection and drug resistance do depress understanding antibiotic resistance via outer membrane permeability. *Infect Drug Resist.* 11:523–530.
- Ghalayini M, Launay A, Bridier-Nahmias A, Clermont O, Denamur E, Lescat M, Tenailon O. 2018. Evolution of a dominant natural isolate of *Escherichia coli* in the human gut over the course of a year suggests a neutral evolution with reduced effective population size. *Appl Environ Microbiol.* 84(6):e02377.
- Giulieri SG, Baines SL, Guerillot R, Seemann T, Gonçalves da Silva A, Schultz M, Massey RC, Holmes NE, Stinear TP, Howden BP. 2018. Genomic exploration of sequential clinical isolates reveals a distinctive molecular signature of persistent *Staphylococcus aureus* bacteraemia. *Genome Med.* 10(1):65.
- Gladman SL, Seemann T, Gao W, Stinear TP, Howden BP, Monk IR, Tobias NJ. 2015. Large tandem chromosome expansions facilitate niche adaptation during persistent infection with drug-resistant *Staphylococcus aureus*. *Microbial Genomics.* 1(2):e000026.
- Gladstone RA, Gritzfeld JF, Coupland P, Gordon SB, Bentley SD. 2015. Genetic stability of pneumococcal isolates during 35 days of human experimental carriage. *Vaccine* 33(29):3342–3345.
- Greninger AL, Streithorst J, Golden JA, Chiu CY, Miller S. 2017. Complete genome sequence of sequential *Pandoraea apista* isolates from the same cystic fibrosis patient supports a model of chronic colonization with in vivo strain evolution over time. *Diagn Microbiol Infect Dis.* 87(1):1–6.
- Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R. 2016. ESX secretion systems: mycobacterial evolution to counter host immunity. *Nat Rev Microbiol.* 14(11):677–691.
- Guérillot R, Gonçalves da Silva A, Monk I, Giulieri S, Tomita T, Alison E, Porter J, Pidot S, Gao W, Peleg AY, et al. 2018. Convergent evolution driven by rifampin exacerbates the global burden of drug-resistant *Staphylococcus aureus*. *mSphere* 3(1):e00550.
- Guerry P, Ewing CP, Hickey TE, Prendergast MM, Moran AP. 2000. Sialylation of lipooligosaccharide cores affects immunogenicity and serum resistance of *Campylobacter jejuni*. *Infect Immun.* 68(12):6656–6662.
- Gumpert H, Kubicek-Sutherland JZ, Porse A, Karami N, Munck C, Linkevicius M, Adlerberth I, Wold AE, Andersson DI, Sommer MOA. 2017. Transfer and persistence of a multi-drug resistance plasmid in situ of the infant gut microbiota in the absence of antibiotic treatment. *Front Microbiol.* 8:1852.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8):1072–1075.
- Hassan AA, Maldonado RF, Dos Santos SC, Di Lorenzo F, Silipo A, Coutinho CP, Cooper VS, Molinaro A, Valvano MA, Sá-Correia I. 2017. Structure of O-antigen and hybrid biosynthetic locus in *Burkholderia cenocepacia* clonal variants recovered from a cystic fibrosis patient. *Front Microbiol.* 8:1027.
- Hauser AR, Jain M, Bar-Meir M, McColley SA. 2011. Clinical significance of microbial infection and adaptation in cystic fibrosis. *Clin Microb Rev.* 24(1):29–70.
- Hawkey J, Ascher DB, Judd LM, Wick RR, Kostoulis X, Cleland H, Spelman DW, Padiglione A, Peleg AY, Holt KE. 2018. Evolution of carbapenem resistance in *Acinetobacter baumannii* during a prolonged infection. *Microbial Genomics.* 4(3):e000165.
- Hayden HS, Lim R, Brittnacher MJ, Sims EH, Ramage ER, Fong C, Wu Z, Crist E, Chang J, Zhou Y, et al. 2012. Evolution of *Burkholderia pseudomallei* in recurrent melioidosis. *PLoS One* 7(5):e36507.
- Herranz M, Pole I, Ozere I, Chiner-Oms Á, Martínez-Lirola M, Pérez-García F, Gijón P, Serrano MJR, Romero LC, Cuevas O, et al. 2017. *Mycobacterium tuberculosis* acquires limited genetic diversity in prolonged infections, reactivations and transmissions involving multiple hosts. *Front Microbiol.* 8:2661.
- Hiller NL, Ahmed A, Powell E, Martin DP, Eutsey R, Earl J, Janto B, Boissy RJ, Hogg J, Barbadora K, et al. 2010. Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog.* 6(9):e1001108.
- Hornsey M, Loman N, Wareham DW, Ellington MJ, Pallen MJ, Turton JF, Underwood A, Gaulton T, Thomas CP, Doumith M, et al. 2011. Whole-genome comparison of two *Acinetobacter baumannii* isolates from a single patient, where resistance developed during tige-cycline therapy. *J Antimicrob Chemother.* 66(7):1499–1503.
- Howden BP, McEvoy CRE, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood V, Porter JL, et al. 2011. Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalkR. *PLoS Pathog.* 7(11):e1002359.
- Howden BP, Stinear TP, Allen DL, Johnson PDR, Ward PB, Davies JK. 2008. Genomic analysis reveals a point mutation in the two-component sensor gene graS that leads to intermediate vancomycin resistance in clinical *Staphylococcus aureus*. *Antimicrob Agents Chemother.* 52(10):3755–3762.
- Hu J, Ng PC. 2012. Predicting the effects of frameshifting indels. *Genome Biol.* 13(2):R9.
- Hua X, Zhou Z, Yang Q, Shi Q, Xu Q, Wang J, Shi K, Zhao F, Sun L, Ruan Z, et al. 2017. Evolution of *Acinetobacter baumannii* in vivo: international clone II, more resistance to ceftazidime, mutation in ptk. *Front Microbiol.* 8:1256.
- Jolley KA, Maiden MCJ. 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinf.* 11(1):595.
- Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28(1):27–30.
- Kannan N, Lai YP, Haug M, Lilleness MK, Bakke SS, Marstad A, Hov H, Naustdal T, Afset JE, Iøerger TR, et al. 2019. Genetic variation/evolution and differential host responses resulting from inpatient adaptation of *Mycobacterium avium*. *Infect Immun.* 87(4):e00323.
- Kans J. 2019. Entrez direct: e-utilities on the UNIX command line. Bethesda (MD): National Center for Biotechnology Information (US).
- Kennemann L, Didelot X, Aebischer T, Kuhn S, Drescher B, Droege M, Reinhardt R, Correa P, Meyer TF, Josenhans C, et al. 2011. *Helicobacter pylori* genome evolution during human infection. *Proc Natl Acad Sci U S A.* 108(12):5033–5038.
- Khademi SMH, Sazinas P, Jelsbak L. 2019. Within-host adaptation mediated by intergenic evolution in *Pseudomonas aeruginosa*. *Genome Biol Evol.* 11(5):1385–1397.

- Kim SJ, Kim Y-J, Ko KS. 2018. Genomic analysis of consecutive *Acinetobacter baumannii* strains from a single patient. *Front Microbiol.* 9:2840.
- Klemm EJ, Gkrania-Klotsas E, Hadfield J, Forbester JL, Harris SR, Hale C, Heath JN, Wileman T, Clare S, Kane L, et al. 2016. Emergence of host-adapted *Salmonella Enteritidis* through rapid evolution in an immunocompromised host. *Nat Microbiol.* 1(3):15023.
- Klockgether J, Cramer N, Fischer S, Wiehlmann L, Tümmler B. 2018. Long-term microevolution of *Pseudomonas aeruginosa* differs between mildly and severely affected cystic fibrosis lungs. *Am J Respir Cell Mol Biol.* 59(2):246–256.
- Korhonen V, Smit PW, Haanperä M, Casali N, Ruutu P, Vasankari T, Soini H2. 2016. Whole genome analysis of *Mycobacterium tuberculosis* isolates from recurrent episodes of tuberculosis, Finland, 1995–2013. *Clin Microbiol Infect.* 22(6):549–554.
- Kreutzfeldt KM, McAdam PR, Claxton P, Holmes A, Seagar AL, Laurenson IF, Fitzgerald JR. 2013. Molecular longitudinal tracking of *Mycobacterium abscessus* spp. during chronic infection of the human lung. *PLoS One* 8(5):e63237.
- La Rosa R, Johansen HK, Molin S. 2018. Convergent metabolic specialization through distinct evolutionary paths in *Pseudomonas aeruginosa*. *mBio* 9(2):e00269.
- Leinonen R, Sugawara H, Shumway M; International Nucleotide Sequence Database Collaboration. 2011. The sequence read archive. *Nucleic Acids Res.* 39(Database):D19–D21.
- Lesho E, Yoon E-J, McGann P, Snesrud E, Kwak Y, Milillo M, Onmus-Leone F, Preston L, St Clair K, Nikolich M, et al. 2013. Emergence of colistin-resistance in extremely drug-resistant *Acinetobacter baumannii* containing a novel pmrCAB operon during colistin therapy of wound infections. *J Infect Dis.* 208(7):1142–1151.
- Ley SD, de Vos M, Van Rie A, Warren RM. 2019. Deciphering within-host microevolution of *Mycobacterium tuberculosis* through whole-genome sequencing: the phenotypic impact and way forward. *Microbiol Mol Biol Rev.* 83(2):e00062.
- Li X, Mu X, Yang Y, Hua X, Yang Q, Wang N, Du X, Ruan Z, Shen X, Yu Y. 2016. Rapid emergence of high-level tigecycline resistance in *Escherichia coli* strains harbouring blaNDM-5 in vivo. *Int J Antimicrob Agents* 47(4):324–327.
- Lieberman TD, Michel J-B, Aingaran M, Potter-Bynoe G, Roux D, Davis MR, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, et al. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet.* 43(12):1275–1280.
- 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(7):580–585.
- Liu Q, Via LE, Luo T, Liang L, Liu X, Wu S, Shen Q, Wei W, Ruan X, Yuan X, et al. 2015. Within patient microevolution of *Mycobacterium tuberculosis* correlates with heterogeneous responses to treatment. *Sci Rep.* 5(1):17507.
- López-Causapé C, Rubio R, Cabot G, Oliver A. 2018. Evolution of the *Pseudomonas aeruginosa* aminoglycoside mutational resistome in vitro and in the cystic fibrosis setting. *Antimicrob Agents Chemother.* 62(4):e02583.
- López-Causapé C, Sommer LM, Cabot G, Rubio R, Ocampo-Sosa AA, Johansen HK, Figuerola J, Cantón R, Kidd TJ, Molin S, et al. 2017. Evolution of the *Pseudomonas aeruginosa* mutational resistome in an international cystic fibrosis clone. *Sci Rep.* 7(1):5555.
- Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, et al. 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380(9859):2095–2128.
- Lynch T, Chen L, Peirano G, Gregson DB, Church DL, Conly J, Kreiswirth BN, Pitout JD. 2016. Molecular evolution of a *Klebsiella pneumoniae* ST278 isolate harboring blaNDM-7 and involved in nosocomial transmission. *J Infect Dis.* 214(5):798–806.
- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A.* 95(6):3140–3145.
- Markussen T, Marvig RL, Gómez-Lozano M, Aanæs K, Burleigh AE, Høiby N, Johansen HK, Molin S, Jelsbak L. 2014. Environmental heterogeneity drives within-host diversification and evolution of *Pseudomonas aeruginosa*. *mBio* 5(5):e01592.
- Martin RM, Cao J, Brisse S, Passet V, Wu W, Zhao L, Malani PN, Rao K, Bachman MA. 2016. Molecular epidemiology of colonizing and infecting isolates of *Klebsiella pneumoniae*. *mSphere* 1(5):e00261.
- Marvig RL, Damkiær S, Khademi SMH, Markussen TM, Molin S, Jelsbak L. 2014. Within-host evolution of *Pseudomonas aeruginosa* reveals adaptation toward iron acquisition from hemoglobin. *mBio* 5(3):e00966.
- Marvig RL, Dolce D, Sommer LM, Petersen B, Ciofu O, Campana S, Molin S, Taccetti G, Johansen HK. 2015. Within-host microevolution of *Pseudomonas aeruginosa* in Italian cystic fibrosis patients. *BMC Microbiol.* 15(1):218.
- Marvig RL, Jochumsen N, Johansen HK, Høiby N, Molin S, Sommer MO, Jelsbak L, Folkesson A. 2013. Draft genome sequences of *Pseudomonas aeruginosa* B3 strains isolated from a cystic fibrosis patient undergoing antibiotic chemotherapy. *Genome Announc.* 1(5):e00804.
- 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(9):e1003741.
- Marvig RL, Sommer LM, Molin S, Johansen HK. 2015. Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat Genet.* 47(1):57–64.
- Marzel A, Desai PT, Goren A, Schorr YI, Nissan I, Porwollik S, Valinsky L, McClelland M, Rahav G, Gal-Mor O. 2016. Persistent infections by nontyphoidal *Salmonella* in humans: epidemiology and genetics. *Clin Infect Dis.* 62(7):879–886.
- Mathers AJ, Crook D, Vaughan A, Barry KE, Vegesana K, Stoesser N, Parikh HI, Sebra R, Kotay S, Walker AS, et al. 2019. *Klebsiella quasipneumoniae* provides a window into carbapenemase gene transfer, plasmid rearrangements, and patient interactions with the hospital environment. *Antimicrob Agents Chemother.* 63(6):e02513–e02518.
- Mathers AJ, Stoesser N, Sheppard AE, Pankhurst L, Giess A, Yeh AJ, Didelot X, Turner SD, Sebra R, Kasarskis A, et al. 2015. *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* at a single institution: insights into endemicity from whole-genome sequencing. *Antimicrob Agents Chemother.* 59(3):1656–1663.
- McAdam PR, Holmes A, Templeton KE, Fitzgerald JR. 2011. Adaptive evolution of *Staphylococcus aureus* during chronic endobronchial infection of a cystic fibrosis patient. *PLoS One* 6(9):e24301.
- Merker M, Kohl TA, Roetzer A, Truebe L, Richter E, Rüscher-Gerdes S, Fattorini L, Oggioni MR, Cox H, Varaine F, et al. 2013. Whole genome sequencing reveals complex evolution patterns of multidrug-resistant *Mycobacterium tuberculosis* Beijing strains in patients. *PLoS One* 8(12):e82551.
- Meumann EM, Globan M, Fyfe JAM, Leslie D, Porter JL, Seemann T, Denholm J, Steinar TP. 2015. Genome sequence comparisons of serial multi-drug-resistant *Mycobacterium tuberculosis* isolates over 21 years of infection in a single patient. *Microbial Genomics.* 1(5):e000037.
- Moleres J, Fernández-Calvet A, Ehrlich RL, Martí S, Pérez-Regidor L, Euba B, Rodríguez-Arce I, Balashov S, Cuevas E, Liñares J, et al. 2018. Antagonistic pleiotropy in the bifunctional surface protein FadL (OmpP1) during adaptation of *Haemophilus influenzae* to chronic lung infection associated with chronic obstructive pulmonary disease. *mBio* 9(5):e01176.
- Moradigaravand D, Gouliouris T, Blane B, Naydenova P, Ludden C, Crawley C, Brown NM, Török ME, Parkhill J, Peacock SJ. 2017. Within-host evolution of *Enterococcus faecium* during longitudinal carriage and transition to bloodstream infection in immunocompromised patients. *Genome Med.* 9(1):119.

- Moxon R, Kussell E. 2017. The impact of bottlenecks on microbial survival, adaptation, and phenotypic switching in host–pathogen interactions. *Evolution* 71(12):2803–2816.
- Mwangi MM, Wu SW, Zhou Y, Sieradzki K, de Lencastre H, Richardson P, Bruce D, Rubin E, Myers E, Siggia ED, et al. 2007. Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. *Proc Natl Acad Sci U S A* 104(22):9451–9456.
- Nielsen LE, Snestrud EC, Onmus-Leone F, Kwak YI, Avilés R, Steele ED, Sutter DE, Waterman PE, Lesho EP. 2014. IS5 element integration, a novel mechanism for rapid in vivo emergence of tigecycline non-susceptibility in *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 58(10):6151–6156.
- Octavia S, Wang Q, Tanaka MM, Sintchenko V, Lan R. 2017. Genomic heterogeneity of *Salmonella enterica* serovar Typhimurium bacteriuria from chronic infection. *Infect Genet Evol* 51:17–20.
- Oliver A, Mena A. 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. *Clin Microbiol Infect* 16(7):798–808.
- Omer H, Rose G, Jolley KA, Frapy E, Zahar J-R, Maiden MCJ, Bentley SD, Tinsley CR, Nassif X, Bille E. 2011. Genotypic and phenotypic modifications of *Neisseria meningitidis* after an accidental human passage. *PLoS One* 6(2):e17145.
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 17(1):132.
- Pak TR, Altman DR, Attie O, Sebra R, Hamula CL, Lewis M, Deikus G, Newman LC, Fang G, Hand J, et al. 2015. Whole-genome sequencing identifies emergence of a quinolone resistance mutation in a case of *Stenotrophomonas maltophilia* bacteremia. *Antimicrob Agents Chemother* 59(11):7117–7120.
- Passalacqua KD, Satola SW, Crispell EK, Read TD. 2012. A mutation in the PP2C phosphatase gene in a *Staphylococcus aureus* USA300 clinical isolate with reduced susceptibility to vancomycin and daptomycin. *Antimicrob Agents Chemother* 56(10):5212–5223.
- Peleg AY, Miyakis S, Ward DV, Earl AM, Rubio A, Cameron DR, Pillai S, Moellering RC, Eliopoulos GM. 2012. Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of *Staphylococcus aureus*. *PLoS One* 7(1):e28316.
- Pérez-Lago L, Comas I, Navarro Y, González-Candelas F, Herranz M, Bouza E, García-de-Viedma D. 2014. Whole genome sequencing analysis of inpatient microevolution in *Mycobacterium tuberculosis*: potential impact on the inference of tuberculosis transmission. *J Infect Dis* 209(1):98–108.
- Pettigrew MM, Ahearn CP, Gent JF, Kong Y, Gallo MC, Munro JB, D’Mello A, Sethi S, Tettelin H, Murphy TF. 2018. *Haemophilus influenzae* genome evolution during persistence in the human airways in chronic obstructive pulmonary disease. *Proc Natl Acad Sci U S A* 115(14):E3256–E3265.
- Philippe N, Maigre L, Santini S, Pinet E, Claverie J-M, Davin-Régli A-V, Pagès J-M, Masi M. 2015. In vivo evolution of bacterial resistance in two cases of *Enterobacter aerogenes* infections during treatment with imipenem. *PLoS One* 10(9):e0138828.
- Porse A, Gumpert H, Kubicek-Sutherland JZ, Karami N, Adlerberth I, Wold AE, Andersson DI, Sommer MOA. 2017. Genome dynamics of *Escherichia coli* during antibiotic treatment: transfer, loss, and persistence of genetic elements in situ of the infant gut. *Front Cell Infect Microbiol* 7:126.
- Poulsen BE, Yang R, Clatworthy AE, White T, Osmulski SJ, Li L, Penaranda C, Lander ES, Shores N, Hung DT. 2019. Defining the core essential genome of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 116(20):10072–10080.
- Price EP, Sarovich DS, Mayo M, Tuanyok A, Drees KP, Kaestli M, Beckstrom-Sternberg SM, Babic-Sternberg JS, Kidd TJ, Bell SC, et al. 2013. Within-host evolution of *Burkholderia pseudomallei* over a twelve-year chronic carriage infection. *mBio* 4(4):e00388.
- Price EP, Sarovich DS, Viberg L, Mayo M, Kaestli M, Tuanyok A, Foster JT, Keim P, Pearson T, Currie BJ. 2015. Whole-genome sequencing of *Burkholderia pseudomallei* isolates from an unusual melioidosis case identifies a polyclonal infection with the same multilocus sequence type. *J Clin Microbiol* 53(1):282–286.
- Price JR, Cole K, Bexley A, Kostiou V, Eyre DW, Golubchik T, Wilson DJ, Crook DW, Walker AS, Peto TEA, et al. 2017. Transmission of *Staphylococcus aureus* between health-care workers, the environment, and patients in an intensive care unit: a longitudinal cohort study based on whole-genome sequencing. *Lancet Infect Dis* 17(2):207–214.
- Raven KE, Gouliouris T, Parkhill J, Peacock SJ. 2018. Genome-based analysis of *Enterococcus faecium* bacteremia associated with recurrent and mixed-strain infection. *J Clin Microbiol* 56(3):e01520.
- Rezzoagli C, Wilson D, Weigert M, Wyder S, Kümmerli R. 2018. Probing the evolutionary robustness of two repurposed drugs targeting iron uptake in *Pseudomonas aeruginosa*. *Evol Med Public Health* 2018:246–259.
- Sabat AJ, Hermelijn SM, Akkerboom V, Juliana A, Degener JE, Grundmann H, Friedrich AW. 2017. Complete-genome sequencing elucidates outbreak dynamics of CA-MRSA USA300 (ST8-spa t008) in an academic hospital of Paramaribo, Republic of Suriname. *Sci Rep* 7(1):41050.
- Sachsenheimer FE, Yang I, Zimmermann O, Wrede C, Müller LV, Gunka K, Groß U, Suerbaum S. 2018. Genomic and phenotypic diversity of *Clostridium difficile* during long-term sequential recurrences of infection. *Int J Med Microbiol* 308(3):364–377.
- Sarovich DS, Webb JR, Pitman MC, Viberg LT, Mayo M, Baird RW, Robson JM, Currie BJ, Price EP. 2018. Raising the stakes: loss of efflux pump regulation decreases meropenem susceptibility in *Burkholderia pseudomallei*. *Clin Infect Dis* 67(2):243–250.
- Saunders NJ, Trivedi UH, Thomson ML, Doig C, Laurenson IF, Blaxter ML. 2011. 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(3):212–217.
- Schwartzbeck B, Birtel J, Treffon J, Langhankl L, Mellmann A, Kale D, Kahl J, Hirschhausen N, Neumann C, Lee JC, et al. 2016. Dynamic in vivo mutations within the *ica* operon during persistence of *Staphylococcus aureus* in the airways of cystic fibrosis patients. *PLoS Pathog* 12(11):e1006024.
- Seeman T. 2017. MLST—scan contig files against PubMLST typing schemes. Available from: <https://github.com/tseemann/mlst>. Accessed November 16, 2020.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069.
- Sherman DR, Gagneux S. 2011. Estimating the mutation rate of *Mycobacterium tuberculosis* during infection. *Nat Genet* 43(5):400–401.
- Sherrard LJ, Tai AS, Wee BA, Ramsay KA, Kidd TJ, Ben Zakour NL, Whitley DM, Beatson SA, Bell SC. 2017. Within-host whole genome analysis of an antibiotic resistant *Pseudomonas aeruginosa* strain sub-type in cystic fibrosis. *PLoS One* 12(3):e0172179.
- Silva IN, Santos PM, Santos MR, Zlosnik JE, Speert DP, Buskirk SW, Bruger EL, Waters CM, Cooper VS, Moreira LM. 2016. Long-term evolution of *Burkholderia multivorans* during a chronic cystic fibrosis infection reveals shifting forces of selection. *mSystems* 1(3):e00029.
- Sim JHC, Truong C, Minot SS, Greenfield N, Budvytiene I, Lohith A, Anikst V, Pourmand N, Banaei N. 2017. Determining the cause of recurrent *Clostridium difficile* infection using whole genome sequencing. *Diagn Microbiol Infect Dis* 87(1):11–16.
- Simner PJ, Antar AAR, Hao S, Gurtowski J, Tamma PD, Rock C, Opene BNA, Tekle T, Carroll KC, Schatz MC, et al. 2018. Antibiotic pressure on the acquisition and loss of antibiotic resistance genes in *Klebsiella pneumoniae*. *J Antimicrob Chemother* 73(7):1796–1803.
- Singh G, Srinivasan R, Cheng J, Peng Z, Fujimura K, Baek MS, Panzer AR, Tringe SG, Chen F, Sorek R, et al. 2014. Rearrangement of a large novel *Pseudomonas aeruginosa* gene island in strains isolated from a patient developing ventilator-associated pneumonia. *J Clin Microbiol* 52(7):2430–2438.
- Snesrud E, Ong AC, Corey B, Kwak YI, Clifford R, Gleeson T, Wood S, Whitman TJ, Lesho EP, Hinkle M, et al. 2017. Analysis of serial isolates

- of mcr-1-positive *Escherichia coli* reveals a highly active ISAp1 transposon. *Antimicrob Agents Chemother.* 61(5):e00056.
- Snitkin ES, Zelazny AM, Gupta J; NISC Comparative Sequencing Program, Palmore TN, Murray PR, Segre JA. 2013. Genomic insights into the fate of colistin resistance and *Acinetobacter baumannii* during patient treatment. *Genome Res.* 23(7):1155–1162.
- Sommer LM, Alanin MC, Marvig RL, Nielsen KG, Høiby N, von Buchwald C, Molin S, Johansen HK. 2016. Bacterial evolution in PCD and CF patients follows the same mutational steps. *Sci Rep.* 6(1):28732.
- Spilker T, LiPuma JJ. 2016. Draft genome sequences of 63 *Pseudomonas aeruginosa* isolates recovered from cystic fibrosis sputum. *Genome Announc.* 4(2):e00231.
- Stefani MMA, Avanzi C, Bühner-Sékula S, Benjak A, Loiseau C, Singh P, Pontes MAA, Gonçalves HS, Hungria EM, Busso P, et al. 2017. Whole genome sequencing distinguishes between relapse and reinfection in recurrent leprosy cases. *PLoS Negl Trop Dis.* 11(6):e0005598.
- Stewart L, Ford A, Sangal V, Jeukens J, Boyle B, Kukavica-Ibrulj I, Caim S, Crossman L, Hoskisson PA, Levesque R, et al. 2014. Draft genomes of 12 host-adapted and environmental isolates of *Pseudomonas aeruginosa* and their positions in the core genome phylogeny. *Pathog Dis.* 71(1):20–25.
- Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FSL, Hufnagle WO, Kowalik DJ, Lagrou M, et al. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406(6799):959–964.
- Suligoy CM, Lattar SM, Noto Llana M, González CD, Alvarez LP, Robinson DA, Gómez MI, Buzzola FR, Sordelli DO. 2018. Mutation of Agr is associated with the adaptation of *Staphylococcus aureus* to the host during chronic osteomyelitis. *Front Cell Infect Microbiol.* 8:18.
- Sun G, Luo T, Yang C, Dong X, Li J, Zhu Y, Zheng H, Tian W, Wang S, Barry CE, et al. 2012. Dynamic population changes in *Mycobacterium tuberculosis* during acquisition and fixation of drug resistance in patients. *J Infect Dis.* 206(11):1724–1733.
- *Swofford DL. 2003. PAUP: phylogenetic analysis using parsimony (* and other methods) version 4. Sunderland (MA): Sinauer Associates.
- Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Colston MJ, Matter L, Schopfer K, Bodmer T. 1993. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet* 341(8846):647–650.
- Thomas DK, Lone AG, Selinger LB, Taboada EN, Uwiera RRE, Abbott DW, Inglis GD. 2014. Comparative variation within the genome of *Campylobacter jejuni* NCTC 11168 in human and murine hosts. *PLoS One* 9(2):e88229.
- Tong SYC, Holden MTG, Nickerson EK, Cooper BS, Köser CU, Cori A, Jombart T, Cauchemez S, Fraser C, Wuthiekanun V, et al. 2015. Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. *Genome Res.* 25(1):111–118.
- Trauner A, Liu Q, Via LE, Liu X, Ruan X, Liang L, Shi H, Chen Y, Wang Z, Liang R, et al. 2017. The within-host population dynamics of *Mycobacterium tuberculosis* vary with treatment efficacy. *Genome Biol.* 18(1):71.
- Trouillet-Assant S, Lelièvre L, Martins-Simões P, Gonzaga L, Tasse J, Valour F, Rasigade JP, Vandenesch F, Muniz Guedes RL, Ribeiro de Vasconcelos AT, et al. 2016. Adaptive processes of *Staphylococcus aureus* isolates during the progression from acute to chronic bone and joint infections in patients. *Cell Microbiol.* 18(10):1405–1414.
- Van Hal SJ, Steen JA, Espedido BA, Grimmond SM, Cooper MA, Holden MTG, Bentley SD, Gosbell IB, Jensen SO. 2014. 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(2):363–367.
- van Mansfeld R, de Been M, Paganelli F, Yang L, Bonten M, Willems R. 2016. Within-host evolution of the Dutch high-prevalent *Pseudomonas aeruginosa* clone ST406 during chronic colonization of a patient with cystic fibrosis. *PLoS One* 11(6):e0158106.
- Viberg LT, Sarovich DS, Kidd TJ, Geake JB, Bell SC, Currie BJ, Price EP. 2017. Within-host evolution of *Burkholderia pseudomallei* during chronic infection of seven Australasian cystic fibrosis patients. *mBio* 8(2):e00356.
- Wandersman C, Deleplaire P. 2004. Bacterial iron sources: from siderophores to hemophores. *Annu Rev Microbiol.* 58(1):611–647.
- Wang K, Chen Y-Q, Salido MM, Kohli GS, Kong J-L, Liang H-J, Yao Z-T, Xie Y-T, Wu H-Y, Cai S-Q, et al. 2017. The rapid in vivo evolution of *Pseudomonas aeruginosa* in ventilator-associated pneumonia patients leads to attenuated virulence. *Open Biol.* 7(9):170029.
- Webb JR, Price EP, Currie BJ, Sarovich DS. 2017. Loss of methyltransferase function and increased efflux activity leads to doxycycline resistance in *Burkholderia pseudomallei*. *Antimicrob Agents Chemother.* 61(6):e00268.
- Williams D, Fothergill JL, Evans B, Caples J, Haldenby S, Walshaw MJ, Brockhurst MA, Winstanley C, Paterson S. 2018. Transmission and lineage displacement drive rapid population genomic flux in cystic fibrosis airway infections of a *Pseudomonas aeruginosa* epidemic strain. *Microb Genomics.* 4(3):e000167.
- Winstanley C, O'Brien S, Brockhurst MA. 2016. *Pseudomonas aeruginosa* evolutionary adaptation and diversification in cystic fibrosis chronic lung infections. *Trends Microbiol.* 24(5):327–337.
- Wollenberg KR, Desjardins CA, Zalutskaya A, Slodovnikova V, Oler AJ, Quiñones M, Abeel T, Chapman SB, Tartakovsky M, Gabrielian A, et al. 2017. Whole-genome sequencing of *Mycobacterium tuberculosis* provides insight into the evolution and genetic composition of drug-resistant tuberculosis in Belarus. *J Clin Microbiol.* 55(2):457–469.
- Worby CJ, Read TD. 2015. 'SEEDY' (Simulation of Evolutionary and Epidemiological Dynamics): an R package to follow accumulation of within-host mutation in pathogens. *PLoS One* 10(6):e0129745.
- Wright MS, Iovleva A, Jacobs MR, Bonomo RA, Adams MD. 2016. Genome dynamics of multidrug-resistant *Acinetobacter baumannii* during infection and treatment. *Genome Med.* 8(1):26.
- Wylie KM, Wylie TN, Minx PJ, Rosen DA. 2019. Whole-genome sequencing of *Klebsiella pneumoniae* isolates to track strain progression in a single patient with recurrent urinary tract infection. *Front Cell Infect Microbiol.* 9:14.
- Xu Y, Liu F, Chen S, Wu J, Hu Y, Zhu B, Sun Z. 2018. In vivo evolution of drug-resistant *Mycobacterium tuberculosis* in patients during long-term treatment. *BMC Genomics.* 19(1):640.
- Yang L, Jelsbak L, Marvig RL, Damkiaer S, Workman CT, Rau MH, Hansen SK, Folkesson A, Johansen HK, Ciofu O, et al. 2011. Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci U S A.* 108(18):7481–7486.
- Young BC, Golubchik T, Batty EM, Fung R, Larmer-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG, et al. 2012. Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. *Proc Natl Acad Sci U S A.* 109(12):4550–4555.
- Young BC, Wu C-H, Gordon NC, Cole K, Price JR, Liu E, Sheppard AE, Perera S, Charlesworth J, Golubchik T, et al. 2017. Severe infections emerge from commensal bacteria by adaptive evolution. *eLife* 6:e30637.