BacFITBase: a database to assess the relevance of bacterial genes during host infection

Javier Macho Rendón [1,†], **Benjamin Lang** [2,†], **Gian Gaetano Tartaglia** [2,3,4,5,*] **and Marc Torrent Burgas** [1,*]

[1]Systems Biology of Infection Lab, Department of Biochemistry and Molecular Biology, Biosciences Faculty, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain, [2]Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain, [3]ICREA, 23 Passeig Lluis Companys 08010 and Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain, [4]Department of Biology 'Charles Darwin', Sapienza University of Rome, P.le A. Moro 5, Rome 00185, Italy and [5]Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genoa, Italy

ABSTRACT

Bacterial infections have been on the rise worldwide in recent years and have a considerable impact on human well-being in terms of attributable deaths and disability-adjusted life years. Yet many mechanisms underlying bacterial pathogenesis are still poorly understood. Here, we introduce the BacFITBase database for the systematic characterization of bacterial proteins relevant for host infection aimed to enable the identification of new antibiotic targets. BacFITBase is manually curated and contains more than 90 000 entries with information on the contribution of individual genes to bacterial fitness under *in vivo* infection conditions in a range of host species. The data were collected from 15 different studies in which transposon mutagenesis was performed, including top-priority pathogens such as *Acinetobacter baumannii* and *Campylobacter jejuni*, for both of which increasing antibiotic resistance has been reported. Overall, BacFITBase includes information on 15 pathogenic bacteria and 5 host vertebrates across 10 different tissues. It is freely available at www.tartaglialab.com/bacfitbase.

INTRODUCTION

The development of new antimicrobial therapies relies heavily on our understanding of the mechanisms of bacterial infection. Bacterial proteins are responsible for rewiring a myriad of biochemical processes essential for the efficient propagation of the pathogen (1,2). Recently, we showed that bacterial fitness *in vivo* does not correlate with data from *in vitro* studies (3). This is a major drawback for antimicrobial target discovery as many *in vitro* false negatives are disregarded for further testing. Therefore, it is crucial to understand how bacterial infection develops *in vivo* and which bacterial genes are required to infect a host.

BacFITBase is a manually curated database of bacterial genes that collates *in vivo* information on their relevance during host infection, as measured by transposon mutagenesis. Transposon mutagenesis experiments allow the measurement of fitness values for individual genes, allowing us to assess which genes are fundamental to infect a specific host organism (4). To address the contribution of a bacterial gene to infection, its fitness is measured through genome-wide transposon mutagenesis coupled with next-generation sequencing (Tn-seq) (5). Briefly, mutations targeting virtually all genes in the bacterial genome are generated by random insertion of transposons. Afterward, these mutants with randomly inactivated genes are grown in culture medium (input pool), inoculated in a host organism and finally recovered after infection (output pool). Genomic DNA from the input and output pool is extracted, and transposon insertion site junctions are amplified and quantified by next-generation sequencing.

BacFITBase provides a common framework and easy access to fitness data for individual bacterial genes during infection. It now covers the infection processes of 15 pathogenic bacteria in 5 model vertebrates across 10 different tissues (Table 1). A tutorial section provides a detailed step-by-step description of how to search and browse our data. This resource is available at www.tartaglialab.com/bacfitbase to help the research community in the systematic characterization of bacterial proteins involved in host infection.

---

*To whom correspondence should be addressed. Tel: +34 93 581 4147; Email: marc.torrent@uab.cat
Correspondence may also be addressed to Gian G. Tartaglia. Tel: +34 93 316 0116; Email: gian.tartaglia@crg.eu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Table 1. List of all studies included in the BacFITBase

Pathogen	Host organism	Tissue	Reference
Salmonella enterica Serovar Typhimurium ST4/74	*Bos taurus (cow)*	Ileal mucosa	(6)
	Sus scrofa (pig)	Colonic mucosa	
	Gallus gallus (chicken)	Cecum	
Haemophilus influenzae Rd KW20	*Mus musculus (mouse)*	Lung	(7)
Streptococcus pyogenes M1 5448	*M. musculus (mouse)*	Skin	(8)
Porphyromonas gingivalis ATCC 33277	*M. musculus (mouse)*	Skin	(9)
Escherichia coli CFT073	*M. musculus (mouse)*	Spleen	(10)
Mycobacterium avium subsp. Paratuberculosis K10	*M. musculus (mouse)*	Spleen	(11)
Escherichia coli M12	*M. musculus (mouse)*	Spleen and mammary gland	(12)
E. coli O157:H7	*B. taurus (cow)*	Feces	(13)
Vibrio cholerae O1 biovar El Tor str. N16961	*Oryctolagus cuniculus (rabbit)*	Small intestine	(14)
Campylobacter jejuni subsp. jejuni 81–176	*M. musculus (mouse)*	Cecum	(15)
Klebsiella pneumoniae subsp. pneumoniae pneumoniae ATCC 43816 KPPR1	*M. musculus (mouse)*	Lung	(16)
Acinetobacter baumannii ATCC 17978	*M. musculus (mouse)*	Lung	(17)
Salmonella enterica Serovar Typhimurium SL1344	*M. musculus (mouse)*	Liver and spleen	(18)
Serratia marcescens Strain UMH9	*M. musculus (mouse)*	Spleen	(19)
Vibrio parahaemolyticus RIMD 2210633	*O. cuniculus (rabbit)*	Small intestine	(20)

MATERIALS AND METHODS

Fitness scores and z-scores

We collected data from publicly available transposon mutagenesis experiments containing either raw input/output read counts or fitness scores for all mutant genes available (Table 1) (6–20). The fitness score of a gene is calculated as the ratio of the normalized frequencies of input/output read counts. Reads for each transposon insertion site are normalized to the total number of reads obtained from a sample according to the following equation:

$$F^a = \frac{C_i^a / N_i}{C_0^a / N_0}$$

where F^a is the fitness score for gene a, C_0^a and C_i^a are the number of reads for a gene a before and after infection and N_0 and N_i are the number of total reads before and after infection, respectively. In order to normalize the fitness scores to allow comparison between different studies, the corresponding z-scores were calculated for each individual experiment (Figure 1). To assess the significance of a bacterial gene's fitness impact, P-values were calculated using a two-tailed one-sample Student's t-test on the distribution of fitness scores within each study provided that raw data were available. Otherwise, the P-values reported in the original study are shown.

Technical aspects

BacFITBase was built using PHP on an Apache web server with a MySQL database backend. BacFITBase stores no user data, except for the anonymous caching of BLAST search results for a given sequence in order to greatly speed up repeated searches. The open-source Bootstrap library was used to allow display on devices of any screen size, including mobile devices. Several icons were included from Font Awesome and the Noun Project, and a number of JavaScript libraries are used for table export and sorting. Please see the About section of the website for detailed attributions.
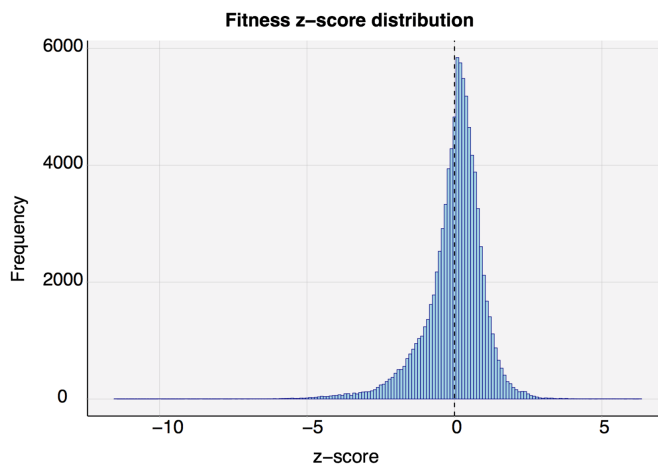

Figure 1. Histogram representing the z-score distribution among all studies included in the BacFITBase database. A fitness z-score < 0 indicates that a given mutation is more detrimental than the average mutation during infection, while a raw fitness score < 1 indicates that a given mutant is underrepresented in the output pool. As we can observe in the histogram, the extreme values in the fitness z-score distribution are shifted toward negative values. Under-represented instances give strong evidence that this particular gene is relevant in the infection process, as transposon insertion is typically associated with a loss of function phenotype.

BLAST search

The NCBI BLAST suite version 2.9.0+ (March 2019) (21) is used to search by sequence similarity. The BLASTP program is used for amino acid sequences, and BLASTX for nucleic acid (coding) sequences. BLAST search results are cached for each unique sequence, which means that re-running a search using the same sequence will yield results nearly instantaneously. As on all other pages, results from the BLAST search page can be linked to and shared with other researchers using the '*Link to these results*' link at the bottom of the page. For sequences above a URL length of 2000 characters this link uses a sequence hash identifying the cached sequence, rather than the sequence itself.

Protein visualization (ProViz)

For UniProt proteins, a protein visualization (ProViz) automatically generated and displayed by ProViz (22). ProViz is an interactive exploration tool for investigating the structural, functional and evolutionary features of proteins and is likely to be particularly helpful for analyzing uncharacterized proteins.

USING BACFITBASE

BacFITBase consists of a text search function to find specific pathogenic bacterial genes, a BLAST search function to find bacterial genes similar to a protein or nucleic acid sequence of interest, a Browse function to quickly identify genes of high fitness impact during infection, and a Tutorial section to get started quickly by following a step-by-step guide. Please note that BacFITBase relies on JavaScript, therefore users will need to enable this in their web browser for full functionality.

Searching for a gene or protein

To search for a gene or protein, users just need to type its name or identifier in the Search tab and press the '*Search*' button. Any of the following options are available: gene symbols, gene locus identifiers, NCBI protein identifiers, UniProt protein accessions or a free-text search in the gene product's description. The Search function also supports smart partial matches, so free-text terms can be used (e.g. 'ribonuclease').

Searching within specific hosts and pathogens

Both pathogen and/or host species can be specified in the drop-down menus on the Search page. If no search term is given, this will result in a complete list of genes for these species, similar to the Browse view (described below).

Search results

After searching, the search results page will display a list of any bacterial genes matching the search term and species selected (Figure 2). The column matched by the search term is highlighted in green (if a search term was provided). For each pathogen species and gene, the search results page already shows a preview of the lowest fitness z-score across all available hosts, tissues and post-infection time points, and the corresponding *P*-value.

Tables on BacFITBase: sorting, downloading and linking to results

To sort any table on BacFITBase as desired, users can simply click on any of the column headers. All tables can be downloaded as a comma-separated CSV file for import into spreadsheet software such as Microsoft Excel or Apple Numbers using the '*Download Table*' button in the top right corner. An appropriate, readable file name is automatically generated. Any results pages can also be linked to and shared with other researchers by right-clicking and copying the '*Link to these results*' link at the bottom of a page or table.

Detailed view of infection fitness scores for a gene

After selecting a gene of interest, a view will open with all the infection fitness information available for this gene (Figure 3). The heading of this page provides information on the selected protein: protein and pathogen identifiers, sequence length, gene name and UniProt identifier.

In the table, all available experimental data are shown: Tissue name and ontology of the host organism, time after infection, transposon insertion site on the pathogen's main chromosome (if reported by the original study), fitness data during infection including the raw fitness score, normalized fitness z-score and *P*-value, and the reference to the original paper where the data were published. A brief description on the meaning of the raw score, normalized z-score and *P*-value is also available as mouse-over explanation on the column headers.

For any protein in UniProt, a ProViz is automatically provided by ProViz, an interactive exploration tool for investigating the structural, functional and evolutionary features of proteins, including Pfam domains (23) and transmembrane regions. This is particularly useful for uncharacterized proteins, which account for a large fraction of bacterial proteins (24). Alternatively, the protein's FASTA sequence can be displayed by pressing the '*Show protein sequence*' button, along with a '*Copy*' link in the top right corner to copy and paste the protein's sequence into other research tools, or into the BacFITBase BLAST Search to search for similar proteins. Similar proteins can be searched via BLAST using the '*Find similar proteins*' button. This allows rapid assessment of the fitness impact of a group of similar proteins across all pathogens in BacFITBase.

BLAST search

The BLAST Search tab provides a search by sequence similarity. When the protein of interest is not in our database, the user may search for similar proteins using BLAST sequence alignment. Finding a similar protein with low z-score (and low *P*-value) is a strong indication that the query sequence may be relevant for infection.

To search for similar proteins in our database using BLAST, the user can paste in the protein or coding sequence of interest in FASTA format and then press the 'Search' button. Both protein and coding sequences can be used, but the proper format (protein or coding sequence) must be specified in the drop-down menu next to the 'Search' button (as illustrated by the examples provided on the BLAST search page).

BLAST results

When the BLAST alignment is ready (usually within a second or less), a search results page will open displaying alignment performance together with a complete description of the identified hits. This includes output columns from BLAST, such as the percentage of sequence identity between query and target in the successfully aligned region ('Identity'), the total number of amino acids that were successfully aligned between query and target ('Aligned'), and the required size of a sequence database in which the current

Search results for tolB

Please choose a gene below for details:

Download Table

| Pathogen	Locus	Protein	UniProt	Gene	Length	Product	p-Value	Fitness z-Score
Salmonella enterica Serovar Typhimurium SL1344	SL1344_0730	CBW16828.1	A0A0H3N961	tolB	430 aa	tolB protein precursor	3.3e-23	●●●●● -4.94
Vibrio parahaemolyticus RIMD 2210633	VP_1060	BAC59323.1	Q87QT9	tolB	450 aa	TolB protein	0.00085	●●●●○ -3.47

Retrieved 2 entries in 2 ms (Link to these results)

Figure 2. Search results. This page displays a list of any bacterial genes matching the search term and the host and pathogen species that were selected. The search results page displays the gene locus and NCBI protein identifier, the UniProt accessions code and the gene symbol (Please see BacFITBase description section for further information). This preview also shows the gene product description and its length, together with the fitness z-score and the corresponding *P*-value. In this example, we show the case of TolB, a periplasmic component of the Tol–Pal complex that plays a central role in the maintenance of cell wall integrity. Although this complex is not essential *in vitro*, mutants defective in the TolB-Pal complex have a reduced infectivity. The results in BacFITBase show that, upon deletion of TolB, the infection fitness of the mutant is strongly decreased.

tolB protein precursor

Salmonella enterica Serovar Typhimurium SL1344 ⬈

Length **430 aa**, Gene **tolB**, UniProt **A0A0H3N961** ⬈

Protein visualisation (by **ProViz**)
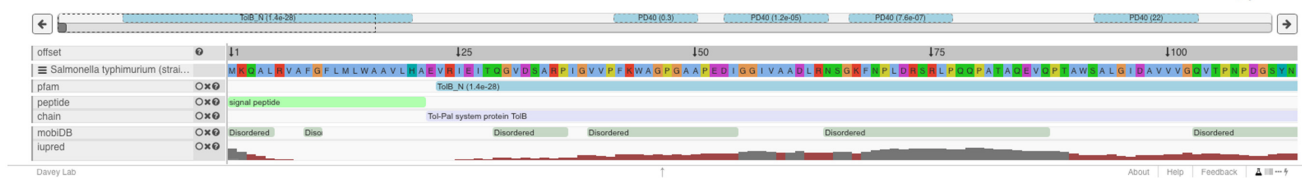
Show protein sequence A | Find similar proteins ☰ | Download Table

Host	Tissue	Time Post Infection	Transposon Insertion Site	Raw Fitness Score	p-Value	Fitness z-Score	Reference
Mouse (*Mus musculus gp91 phox*)	liver ⬈	48 h	815,461 ⬈	-6.68	3.3e-23	●●●●● -4.94	26787719 ⬈
Mouse (*Mus musculus gp91 phox*)	spleen ⬈	48 h	815,461 ⬈	-6.64	7.2e-19	●●●●● -4.91	26787719 ⬈
Mouse (*Mus musculus C57BL/6*)	liver ⬈	48 h	815,461 ⬈	-5.19	0.089	●●●●○ -3.83	26787719 ⬈
Mouse (*Mus musculus C57BL/6*)	spleen ⬈	48 h	815,461 ⬈	-4.91	0.26	●●●●○ -3.62	26787719 ⬈

Retrieved 4 of 4 entries in 1.3 ms (Link to these results)

Figure 3. Detailed view of infection fitness scores for a gene. This page displays all the infection fitness information available for a bacterial gene, along with a ProViz visualization of the protein's sequence and structural features. The detailed view shows the fitness raw values and z-scores for each entry together with all the details of the experiment: hosts, tissues, transposon insertion sites, post-infection time points and the corresponding *P*-value. The reference to the original paper where the data were published is also included in the last column of the table with the corresponding PubMed link.

match could be found just by chance (the 'bit score'). The *E*-value is the expected number of false positive matches given the size of the search database used. Matches with 100% sequence identity in the aligned region are highlighted in green. The meaning of the Pathogen, Locus, Protein, Gene, Product, *P*-value and Fitness z-score columns can be found in the 'Browse Tab' section below, or via the mouse-over information symbols in the top row of any table.

Browsing the entire database (the Browse Tab)

The Browse tab provides an overview of all entries in the BacFITBase database. A pathogenic species of interest can be chosen in the selection element at the top. This table is sorted by significance and fitness z-score, which means that bacterial genes with a high and significant fitness impact during infection are listed first. Arrows next to each field link out to useful external databases, including the NCBI Taxonomy database, a comprehensive taxonomic database ('Pathogen'), the Ensembl Bacteria database ('Locus'), which provides genome annotation for many bacterial

species and the NCBI Protein database ('Protein'), which provides protein sequences and information. Additionally, the 'UniProt Accession' and 'Gene Symbol' columns link out to the UniProt Knowledgebase (25), which provides comprehensive protein annotation.

Downloading the entire database (the Download Tab)

The entire BacFITBase database is available for download on the Download page. Currently, BacFITBase v1 is available, and will be upgraded with new data as they become available.

DISCUSSION

Infectious diseases are caused by micro-organisms known as pathogens, which have the capacity to enter, colonize and grow within a host, causing infection and damage. The use of antibiotics to treat bacterial infections has undoubtedly been one of the most important advances in healthcare, sav-

ing millions of lives since their discovery and widespread use (26,27).

Antibiotic development has mainly focused on the identification of 'essential' genes and proteins in bacteria whose inhibition is lethal under *in vitro* conditions (i.e. bacterial growth in culture). However, pathogenic bacteria do not grow alone but in a complex host environment. Thus, we need to revise the definition of what 'essential' means so that it includes the biological context of infection, i.e. which genes are essential for the pathogen during host infection.

For a given bacterium, the *in vivo* fitness cost of deleting a single gene is correlated with the number of interactions with host proteins (3). Therefore, proteins with high impact on pathogen fitness during infection may cause extensive rewiring of the host interactome. These observations indicate that infectious diseases are only properly understood in the context of the host–pathogen interactions. Toward this end, a promising approach is to systematically characterize proteins involved in host infection. The possibility of accessing fitness data from disparate sources quickly and easily should accelerate the identification of new proteins involved in life-threatening infectious diseases.

In this context, BacFITBase constitutes a valuable resource to systematically classify bacterial proteins relevant for host cell invasion and infection. BacFITBase will facilitate the task of identifying target proteins and interspecies complexes that will help us to understand the mechanisms of infection, and the design of new antimicrobial molecules aimed to interfere with the formation of such complexes. In the next several years, we expect our database to grow continuously, becoming an even more comprehensive repository of bacterial proteins which could be important targets in the fight against infectious diseases.

REFERENCES

1. Ahmed,H., Howton,T.C., Sun,Y., Weinberger,N., Belkhadir,Y. and Mukhtar,M.S. (2018) Network biology discovers pathogen contact points in host protein-protein interactomes. *Nat. Commun.*, **9**, 1–13.
2. Pan,A., Lahiri,C., Rajendiran,A. and Shanmugham,B. (2016) Computational analysis of protein interaction networks for infectious diseases. *Brief. Bioinform.*, **17**, 517–526.
3. Crua Asensio,N., Munõz Giner,E., De Groot,N.S. and Torrent Burgas,M. (2017) Centrality in the host-pathogen interactome is associated with pathogen fitness during infection. *Nat. Commun.*, **8**, 14092.
4. Van Opijnen,T. and Camilli,A. (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat. Rev. Microbiol.*, **11**, 435–442.
5. Chao,M.C., Abel,S., Davis,B.M. and Waldor,M.K. (2016) The design and analysis of transposon insertion sequencing experiments. *Nat. Rev. Microbiol.*, **14**, 119–128.
6. Chaudhuri,R.R., Morgan,E., Peters,S.E., Pleasance,S.J., Hudson,D.L., Davies,H.M., Wang,J., van Diemen,P.M., Buckley,A.M., Bowen,A.J. *et al.* (2013) Comprehensive assignment of roles for Salmonella Typhimurium genes in intestinal colonization of food-producing animals. *PLoS Genet.*, **9**, e1003456.
7. Gawronski,J.D., Wong,S.M.S., Giannoukos,G., Ward,D. V. and Akerley,B.J. (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 16422–16427.
8. Le Breton,Y., Belew,A.T., Freiberg,J.A., Sundar,G.S., Islam,E., Lieberman,J., Shirtliff,M.E., Tettelin,H., El-Sayed,N.M. and McIver,K.S. (2017) Genome-wide discovery of novel M1T1 group A streptococcal determinants important for fitness and virulence during soft-tissue infection. *PLoS Pathog.*, **13**, e1006584.
9. Miller,D.P., Hutcherson,J.A., Wang,Y., Nowakowska,Z.M., Potempa,J., Yoder-Himes,D.R., Scott,D.A., Whiteley,M. and Lamont,R.J. (2017) Genes contributing to porphyromonas gingivalis fitness in abscess and epithelial cell colonization environments. *Front. Cell. Infect. Microbiol.*, **7**, 378.
10. Subashchandrabose,S., Smith,S.N., Spurbeck,R.R., Kole,M.M. and Mobley,H.L.T. (2013) Genome-Wide detection of fitness genes in uropathogenic escherichia coli during systemic infection. *PLoS Pathog.*, **9**, 1–15.
11. Wang,J., Pritchard,J.R., Kreitmann,L., Montpetit,A. and Behr,M.A. (2014) Disruption of Mycobacterium avium subsp. paratuberculosis-specific genes impairs in vivo fitness. *BMC Genomics*, **15**, 415.
12. Olson,M.A., Siebach,T.W., Griffitts,J.S., Wilson,E. and Erickson,D.L. (2018) Genome-wide identification of fitness factors in mastitis-associated Escherichia coli. *Appl. Environ. Microbiol.*, **84**, e02190-17.
13. Eckert,S.E., Dziva,F., Chaudhuri,R.R., Langridge,G.C., Turner,D.J., Pickard,D.J., Maskell,D.J., Thomson,N.R. and Stevens,M.P. (2011) Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn5Km2 mutants of Escherichia coli O157:H7 screened in cattle. *J. Bacteriol.*, **193**, 1771–1776.
14. Fu,Y., Waldor,M.K. and Mekalanos,J.J. (2013) Tn-seq analysis of vibrio cholerae intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. *Cell Host Microbe*, **14**, 652–663.
15. Gao,B., Vorwerk,H., Huber,C., Lara-Tejero,M., Mohr,J., Goodman,A.L., Eisenreich,W., Galán,J.E. and Hofreuter,D. (2017) Metabolic and fitness determinants for in vitro growth and intestinal colonization of the bacterial pathogen Campylobacter jejuni. *PLoS Biol.*, **15**, e2001390.
16. Bachman,M.A., Breen,P., Deornellas,V., Mu,Q., Zhao,L., Wu,W., Cavalcoli,J.D. and Mobley,H.L.T. (2015) Genome-wide identification of Klebsiella pneumoniae fitness genes during lung infection. *Mbio*, **6**, e00775-15.
17. Wang,N., Ozer,E.A., Mandel,M.J. and Hauser,A.R. (2014) Genome-wide identification of Acinetobacter baumannii genes necessary for persistence in the lung. *Mbio*, **5**, e01163-14.
18. Grant,A.J., Oshota,O., Chaudhuri,R.R., Mayho,M., Peters,S.E., Clare,S., Maskell,D.J. and Mastroeni,P. (2016) Genes required for the

fitness of Salmonella enterica serovar Typhimurium during infection of immunodeficient gp91-/- phox mice. *Infect. Immun.*, **84**, 989–997.

19. Anderson,M.T., Mitchell,L.A., Zhao,L., Mobley,H.L.T. and Welch,R. (2017) Capsule production and glucose metabolism dictate fitness during serratia marcescens bacteremia. *Mbio*. **8**, e00740-17.

20. Hubbard,T.P., Chao,M.C., Abel,S., Blondel,C.J., Abel zur Wiesch,P., Zhou,X., Davis,B.M. and Waldor,M.K. (2016) Genetic analysis of Vibrio parahaemolyticus intestinal colonization. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 6283–6288.

21. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

22. Jehl,P., Manguy,J., Shields,D.C., Higgins,D.G. and Davey,N.E. (2016) ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences. *Nucleic Acids Res.*, **44**, W11–W15.

23. El-Gebali,S., Mistry,J., Bateman,A., Eddy,S.R., Luciani,A., Potter,S.C., Qureshi,M., Richardson,L.J., Salazar,G.A., Smart,A. *et al.* (2019) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427–D432.

24. Riley,M., Abe,T., Arnaud,M.B., Berlyn,M.K.B., Blattner,F.R., Chaudhuri,R.R., Glasner,J.D., Horiuchi,T., Keseler,I.M., Kosuge,T. *et al.* (2006) Escherichia coli K-12: A cooperatively developed annotation snapshot - 2005. *Nucleic Acids Res.*, **34**, 1–9.

25. UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.

26. Cassini,A., Högberg,L.D., Plachouras,D., Quattrocchi,A., Hoxha,A., Simonsen,G.S., Colomb-Cotinat,M., Kretzschmar,M.E., Devleesschauwer,B., Cecchini,M. *et al.* (2019) Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infect. Dis.*, **19**, 56–66.

27. Collignon,P., Beggs,J.J., Walsh,T.R., Gandra,S. and Laxminarayan,R. (2018) Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis. *Lancet Planet. Heal.*, **2**, e398–e405.