RNA atlas of human bacterial pathogens uncovers stress dynamics linked to infection

Kemal Avican1✉, Jehad Aldahdooh2,3, Matteo Togninalli4,5, A. K. M. Firoj Mahmud1, Jing Tang2,3, Karsten M. Borgwardt4,5, Mikael Rhen6 & Maria Fällman1✉

Bacterial processes necessary for adaption to stressful host environments are potential targets for new antimicrobials. Here, we report large-scale transcriptomic analyses of 32 human bacterial pathogens grown under 11 stress conditions mimicking human host environments. The potential relevance of the in vitro stress conditions and responses is supported by comparisons with available in vivo transcriptomes of clinically important pathogens. Calculation of a probability score enables comparative cross-microbial analyses of the stress responses, revealing common and unique regulatory responses to different stresses, as well as overlapping processes participating in different stress responses. We identify conserved and species-specific ‘universal stress responders’, that is, genes showing altered expression in multiple stress conditions. Non-coding RNAs are involved in a substantial proportion of the responses. The data are collected in a freely available, interactive online resource (PATHOgenex).
**Results**

**Cataloging stress response of cross-microbial human pathogens.** The majority of bacteria in the PATHOgenex RNA atlas represent pathogens causing worldwide health problems. Most strains are commonly used in the microbial research community and are diverse in terms of Gram staining, phylogeny, and oxygen requirement (Fig. 1a). We exposed 32 bacterial pathogens to 10 infection-relevant stress conditions. Also species-specific in vitro virulence inducing conditions were included if described in literature. As control for differential expression analyses, we utilized unexposed, exponentially grown bacteria (Fig. 1a, Table 1, Supplementary Data 1, and Supplementary Fig. 1a). We obtained an average of 12.2 million reads for each RNA-depleted barcoded library, which far exceeds the 2–3 million reads considered sufficient to determine differentially expressed genes from bacteria with high significance (Supplementary Data 2). Hierarchical clustering of expression profiles clustered replicates of the same sample together, indicating robust measurements of gene expression in all conditions and strains (Supplementary Fig. 1b).

To assess quality of read mappings, also transcript length coverage was evaluated for each strain using the replicate with the lowest number of total reads. This showed accurate distribution of reads along the transcript length even in the poorest sequencing libraries, indicating good mapping quality (Supplementary Fig. 2).

Differential expression analysis showed that all bacteria dynamically responded to the stresses by regulating 62–90% of their genes in at least one condition (Fig. 1b). The highest fraction of regulated genes was measured under hypoxia, nutritional downshift, and stationary phase (Fig. 1c). Accuracy of the analyses was verified by differential expression of genes expected to be regulated under certain stress conditions (Supplementary 1).
Fig. 1 Stress responses of cross-microbial human pathogens. a Phylogenetic clustering of bacterial species included in the study together with a schematic illustration of the experimental setup for 11 infection-related stress exposures (see also Table 1) and RNA-seq library preparation with RNAtag-Seq allowing multiple samples (36 in this study) per library used for obtaining 1122 transcriptomes deposited in the PATHOgenex RNA atlas. Phylogenetic orders, Gram staining groups, and oxygen dependency are indicated by color. b Proportion of genes differentially regulated in at least one of the conditions for each species. The linked heat maps show proportions of regulated genes for each stress condition. The scales to the right show the span from zero (white) to the highest (green) proportion (%) of regulated genes in the stress condition with the highest percentage of regulated genes. Heat map squares marked with a cross indicate that RNA-seq was not performed for that species. The dot plot showing the percentages of genes differentially regulated for bacterial groups with aerobic and microaerophilic growth, f for three main phylogenetical orders included in this study, n = 32 species were examined for As, Li, Nd, Ns, Oss, and Oxs; n = 26 were examined for Bs; n = 31 were examined for Hyp; n = 31 were examined for Sp; n = 30 were examined for Vic in c, n = 21 Gram-negative and n = 8 Gram-positive; n = 17 aerobic and n = 15 microaerophilic bacterial species were examined in d and e, respectively. n = 4, n = 14, and n = 8 bacterial species from Betaproteobacteria, Gammaproteobacteria, and Bacilli were examined in e. Data are presented as mean values in c and as mean values ± SD in d, e, and f. The significance between the groups in d, e, and f was calculated with two-tailed Multiple t-test using Holm-Sidak method by Prism Graphpad version 8.2.0. ** indicates p-value = 0.0033. Source data are provided as a Source Data file.

Fig. 3. The bid gene encoding bacterioferritin ferroxidase, shown to be upregulated under iron starvation23,24, was upregulated under low iron condition in all strains harboring the gene. Similarly, proW encoding a permease that is part of the osmotically inducible ProU ABC transporter system25, was upregulated in many of the tested strains under osmotic stress. Further, dps encoding DNA protection during starvation protein, shown to be positively regulated in an RpoS-dependent manner during stationary phase26, was upregulated in many strains during stationary phase. The observed downregulation in E. coli...
(UPEC) and some other species such as *Streptococcus pyogones* could be due to OxyR mediated upregulation in exponential phase. As expected, *cstA*, encoding carbon starvation protein, was induced under nutritional downshift in the majority of strains harboring the gene. We also observed upregulation of *cstA* paralogs in strains harboring these. The observed upregulation of *ahpC*, encoding alkyl hydroperoxide reductase subunit during oxidative stress, is in accordance with what is known for many bacteria. This gene was downregulated in *H. pylori* and not regulated for *F. tularensis* and *Haemophilus influenzae*, suggesting possible involvement of other hydrogen peroxidases. The gene encoding the heat shock protein GroL was upregulated in the latter group (Fig. 1d). Exceptions here were *Aggregatibacter actinomycetemcomitans*, *Campylobacter jejuni*, *F. tularensis*, and *Neisseria meningitidis* (Fig. 1b). The reason for this discrepancy in response to bile is not clear, but it is believed that Gram-positive bacteria are more sensitive to bile in general, and that the Gram-negative outer membrane with lipopolysaccharides contributes to bile protection, which might explain the lower response. However, bacterial species adapted to the mammalian intestine are often resistant to bile, and since the majority of intestinal bacteria in the strain collection are found in the Gram-negative group, we cannot exclude that this bias also is reflected in the comparison.

### Clustering genes into gene groups for cross microbial comparisons

Comparing differential gene expression across different bacterial species is important for broad and comprehensive understanding of gene regulation and function in adaptation to new environments. Therefore, we clustered genes from the 32 strains into gene groups using two different orthology and homology approaches: one based on functional orthologs using KEGG’s annotation tool GhostKOALA for KO groups, and one based on isofunctional homologs using the PATRIC database with Patric Global Family (PGFam). KO numbers are assigned genes harboring the same function, independent of sequence homology, while PGFam numbers are assigned genes with same function and sequence homology (isofunctional homologs) as indicated by RAST and CoreSEED. PGFam groups are more specific in terms of function and homology, but with fewer genes per group. The tools assigned 6054 KO numbers to 63,831 genes, representing 60.7% of totally 105,088 genes in the dataset, and 27,999 PGFam numbers to 97,565 genes, representing 92.8% of the total number of genes (Fig. 2a).

#### Revealing probability of gene groups to be differentially expressed

To predict probability of genes within KO or PGFam groups to be regulated under certain conditions, we formulated an equation (Eq. 1) that computes a stress condition-specific score indicating probability to be differentially expressed (PTDEX). The equation was employed to gene groups with at least 2 genes (5340 KO groups and 11,353 PGFam groups) (Supplementary Data 3 and 4). The PTDEX score takes into account how conserved the genes in the group are among the 32 pathogens, what proportion of the genes are differentially regulated under a particular stress condition, how well this regulation is preserved among the species in the database, and what number of genes are regulated. We calculated stress condition-specific PTDEX scores for KO and PGFam groups as

\[
\text{PTDEX} = \frac{n_{\text{genes}}(\text{on})}{n_{\text{genes}}(\text{total})} \cdot \frac{n_{\text{species}}(\text{on})}{n_{\text{species}}(\text{total})} \cdot \frac{\log_2(1 + n_{\text{genes}}(\text{on}))}{\log_2(1 + n_{\text{species}}(\text{on}))}
\]

The equation was developed in three parts, evaluating the different criteria mentioned above. In the first part, proportion of regulated genes in a particular gene group under a certain stress condition was calculated. This calculation gives a value regardless of large (well-conserved) or small (less-conserved) gene groups, hence gives no indication regarding how well this regulation is conserved among the species in the database. Therefore, in the second part, the frequency of this regulation among the species in the database was calculated. Here the number of observations at the species level, not strain level was calculated, as there are species with more than one strain in the database. This was done to ensure that a likely similar regulation of same genes in different strains of a species not affects the frequency. As bacterial strains in the database are diverse in terms of phylogenetical order and gene content, they are not expected to harbor genes from all gene groups. Therefore, frequency is calculated as number of species having at least one regulated gene divided by square root of

| Condition                    | Growth Medium  | Growth Temperature | OD_{400} | Stress exposure       | Exposure Time |
|------------------------------|----------------|--------------------|----------|-----------------------|---------------|
| Control (Ctrl)               | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | —                     | 10 min        |
| Acidity stress (As)          | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | pH: 3-5               | 10 min        |
| Bile stress (Bs)             | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | 0.5% Bile Salts       | 10 min        |
| Low iron (Li)                | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | 250 μM 2,2-dipyridyl   | 10 min        |
| Hypoxia (Hyp)                | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | Low oxygen            | 3-4 h         |
| Nutritional downshift (Nd)   | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | 1X M9 salts           | 30 min        |
| Nitrosative stress (Ns)      | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | 250 μM Spermine NONOate| 10 min        |
| Oxidative stress (OxS)       | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | 0.5-10 mM H_{2}O_{2}   | 10 min        |
| Osmotic stress (Oss)         | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | 0.5 M NaCl            | 10 min        |
| Stationary phase (Sp)        | Ambient Med.   | Ambient Temp.      | 0.5-2.5  | Stationary phase      | 3-16 h        |
| Temperature (Tm)             | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | 41°C                  | 20 min        |
| Vir. ind. cond. (Vic)        | Ambient Med.   | Ambient Temp.      | 0.1-0.5  | Varies                | Varies        |
strains harboring the gene times strains in the database. The first two parts of the score provide relative values of proportion of regulated genes and frequency of their regulation, but not number of regulated genes, thus reflecting only conservation of gene groups and their regulation. Therefore, in the third part, the log2 value of total number of genes that are regulated in the gene groups was calculated and value of one was added to avoid obtaining a factor of zero for non-regulated genes. The log2 value was used to minimize the effect of well conserved gene groups against poorly conserved gene groups. This was very critical for PTDEX score equation to give higher scores to more particular stress associated gene groups then well-conserved genes unlike to other statistical analysis such as hypergeometric test. While the lowest PTDEX is zero due to no regulated genes, we observed the highest PTDEX score (2.83) for PGF_00016395 (rpmD) under hypoxia and stationary phase. This group has 29 genes encoded in 28 species of which 23 genes in 23 species are differentially regulated under hypoxia and stationary phase. The PTDEX score is a transformation of differential expression of individual genes from different species under each stress condition to a single value (Fig. 2b) that can be used for comprehensive understanding of gene regulation in bacteria.

As the calculation measures conservation of the genes and their regulation among tested bacteria, the PTDEX scores of gene groups that are highly conserved among the 32 strains but not regulated in many would have similar PTDEX score with gene groups that are less conserved but regulated in a high proportion of strains harboring the gene. For example; leuS (PGF_06812369) is conserved in 31 strains and regulated in only 11 species under nutritional downshift and gcvPA (GF_00008774) is conserved in 6 strains and regulated in 4 species have similar PTDEX score (0.43 and 0.44, respectively) for nutritional downshift. Score values of ≥0.25, a value where at least 50% of the genes in the groups are differentially regulated regardless of the number of genes in the group was considered as ‘high PTDEX score’ (Supplementary Fig. 4a).

To test the power of PTDEX scores, we ranked KO and PGFam groups based on PTDEX scores in low iron and oxidative stress and plotted those representing the highest 20 (Fig. 2c). Many of the genes with high KO and PGFam PTDEX scores for low iron represented genes associated with iron uptake and iron homeostasis. Similarly, the genes with highest scores in oxidative stress represented genes indicative for the different stress conditions shown in Supplementary Fig. 3a have high PTDEX scores (Supplementary Fig. 4b). Interestingly, we observed similar levels of PTDEX scores for KO and PGFam groups. Similar PTDEX score for relatively smaller PGFam groups likely reflects a more similar regulation of isofunctional homologs than of functional orthologs. Therefore, for more accurate conclusions, we used PGFam PTDEX scores for the remainder of our analyses.
Comparisons of stress responses between evolutionary distinct bacterial groups. We next re-clustered genes of Gram-negative and -positive strains into G- and G+ PGFam groups and recomputed the PTDEX scores separately (Supplementary Data 5 and 6). We then generated a similarity matrix with the Pearson correlation coefficient for PTDEX scores from 7105 G- PGFam groups, representing 47,468 genes, and 2903 G+ PGFam groups, representing 14,200 genes, for each stress condition. There were significant differences in responses by Gram-negative versus -positive strains, which supports previous studies35 (Fig. 3a). Transcription factors, especially global regulators, are poorly conserved, and transcriptional regulation has been found to be more flexible than their target genes36. In support of this, an analysis of differential expression of transcription factors in response to the different stresses revealed high diversity between Gram-negative and -positive strains (Supplementary Fig. 5).

However, despite substantial differences in PGFam groups, parts of the responses were shared by Gram-negative and -positive bacteria, where conditions showing the highest overlaps were hypoxia, nutritional downshift, and stationary phase (Fig. 3b). These analyses showed usage of PTDEX score as a novel approach providing in-depth observations of common and specific regulations of the same genes in different species, simplifying cross microbial comparisons.

Identification of intersections between stress responses. Reponses to different stressors are expected to partly overlap due to involvement of the same molecular pathways and functional diversity of gene products. Many responses are expected to involve halted growth, enabling translational adjustments and redirection of resources. We next used PTDEX scores of G- (G+) PGFam groups to identify overlaps between different stress responses. There were higher degrees of overlaps in Gram-negative compared to that of Gram-positive bacteria, where especially hypoxia, stationary phase, and nutritional downshift overlapped to high extent. Also responses to low iron, oxidative, and nitrosative stress, as well as responses to acidic and osmotic stress overlapped (Fig. 3c). To reveal overlapping pathways and processes, we implemented a co-expression module identification algorithm37 on G- and G+ PGFam PTDEX scores. This generated six modules of PGFam groups for Gram-negatives showing common patterns of PTDEX scores of PGFam groups in certain combinations of stress conditions, and 3 modules in Gram positives. KEGG pathway mapping of genes in these modules indicated overlapping pathways and processes (Fig. 3d, Supplementary Data 5). The higher number of genes involved in overlapping pathways in N-module-1 and P-module-1 is indicative of more global responses to hypoxia, nutritional downshift and stationary phase. In accordance, genes encoding RpsA and RplA, involved in ribosome biogenesis and ArgH and LysC involved in amino acid biosynthesis were differentially regulated under those conditions in both Gram-negative and -positive strains (Fig. 3d). As seen in N-module-2, genes encoding ribonucleoside-diphosphate reductase subunits (nrdA, nrdB), and AMP nucleosidase (amn) involved in Purine-Pyrimidine metabolism were regulated in response to low iron, nitrosative stress and oxidative stress, conditions known to induce DNA damage38. Accordingly, nrdA and nrdB were previously shown to be induced upon exposure to DNA damaging agents in E. coli39. The dataset did not allow adequate comparisons between phylogenetical orders due to the uneven distribution of strains; only Bacilli representing Gram-positive strains (9 strains) and dominance of Gammaproteobacteria among Gram-negative strains (3 Epsilonproteobacteria, 4 Betaproteobacteria, 14 Gammaproteobacteria).

Revealing universal stress responders including known antimicrobial targets. Participation of gene products in responses to multiple stresses can be key to conservation during evolutionary diversification of species. To retrieve genes encoding these universal stress responders (USRs), we selected PGFam groups with high PTDEX scores (≥0.25) in at least six conditions and containing at least 11 genes from Gram-negative strains (21 in total) and 4 from Gram-positive strains (9 in total) (Supplementary Data 7). We identified 168 USRs (from 6465 individual genes), where functional clustering with PATRIC subsystems40 revealed that USRs are involved in basic biological processes such as metabolism, energy generation, ribosome biogenesis, amino acid biosynthesis, cell division, RNA processing, membrane transport (Supplementary Fig. 6). Interestingly, 9 of the USR genes encode targets for antibiotics, where mutations in their sequences been shown to confer antibiotic resistance (Supplementary Fig. 6). Hence, novel putative antibiotic targets might be found among remaining 159 USRs. In accordance, nrdD and nrdG genes encoding Class III ribonucleotide reductase, display high PTDEX score in all conditions has indeed been suggested as a target for compounds to inhibit cell growth41.

An attractive approach for future antimicrobials is the development of narrow-spectrum or even species-specific drugs to avoid cross-resistance in non-targeted bacteria, and we therefore also retrieved genes representing species-specific USRs (Supplementary Data 8). This group includes genes that are specific to a particular species in the PATHOgenex dataset and differentially regulated under at least 6 stress conditions. We identified 2194 species-specific USRs of which 960 were annotated as genes encoding hypothetical proteins, indicating that a considerable portion of species-specific stress response is still unknown and remain to be explored.

Stress responses identified in PATHOgenex are relevant for infection in vivo. To reveal the in vivo relevance of the data collection of in vitro obtained transcriptomes, we employed expression profiles for clinically relevant pathogens obtained during in vivo infection and compared these to PATHOgenex data. We utilized recently published transcriptomes of the Gram-negative P. aeruginosa in cystic fibrosis lungs42, and the Gram-positive S. aureus (MSSA) in acute murine osteomyelitis43. The comparisons showed that the majority of (75–83%) of the in vivo regulated genes of both pathogens were co-regulated in at least one of the in vitro stress conditions (Fig. 4a). Notable, 3–12% of the regulated genes represented USRs, highlighting the potential importance of those genes for adaptation to real host environments. In vivo regulated genes not co-regulated in any of the in vitro conditions indicate existence of in vivo specific regulation. This analysis showed that mimicking host stress conditions is a relevant approach for studies aiming at understanding of stress dynamics linked different infection scenarios.

To reveal specific stress responses associated with these infections, the number of genes for each in vitro stress condition that also was regulated during infection was determined. Hypoxia, nutritional downshift, and stationary phase responses were excluded as these conditions are complex and involve multiple overlapping stress responses. This showed that responses to nitrosative stress and osmotic stress were pronounced during P. aeruginosa lung infection (Fig. 4b). P. aeruginosa is known to be exposed to low iron, oxidative stress, and osmotic stress in microenvironments of cystic fibrosis lungs42,44,45. However, this is mostly based on functional annotation analysis using Gene Ontology and KEGG pathway mapping. Although such analyses provide information of infection dynamics, they are based on very
low number of functionally annotated genes. Comparing to data in PATHOgenex, we could retrieve more comprehensive information about ongoing stress responses and found that response to nitrosative stress during *P. aeruginosa* lung infections have been underestimated. The nitrosative stress related genes differentially regulated in infected lungs involve for example the *narGHJI* operon and *narK* encoding proteins are involved in repair of damage caused by reactive nitrogen species. Also *kgu* operon consisting of *kguE, kguK, kguT*, and *kguD* genes encoding proteins involved in 2-ketogluconate biosynthesis and catabolism pathways were induced during infection and nitrosative stress only. Response to osmotic stress was also rather pronounced during lung infection.
Fig. 4 Stress responses identified in PATHOgenex are relevant for infection in vivo and can be used for determination of environmental stresses for pathogens at different infection niches. a Differentially expressed genes obtained from differential expression analysis of in vivo transcriptomes of P. aeruginosa in cystic fibrosis lungs50 and S. aureus (MSSA) during acute murine osteomyelitis46 were mapped to in vitro stress responses of the corresponding species. Log2 fold changes in vivo vs. in vitro (control) are shown with genomic localization of genes for each species. The genes that are co-regulated (up/up or down/down) between in vivo and in vitro stress conditions are shown in gray or blue (conserved USRs) bubbles where the size of the bubble indicates number of conditions showing co-regulation with in vivo. Genes with in vivo specific regulation, which has no co-regulation at any of the stress conditions, are shown with red bubbles. The bars in the upper left corner of each bubble plot indicate the proportion of genes that are similarly regulated in vivo and in vitro (grey and blue, where blue indicate USRs) and proportion of genes showing in vivo specific regulation (red). b The number of genes that are co-regulated during infection for each of the PATHOgenex in vitro stress conditions for P. aeruginosa and c, for S. aureus. Hypoxia, nutritional downshift, and stationary phase were not included due to the presence of variety of stress responses under those conditions. Source data are provided as a Source Data file.

involving expression of pscU, pscF, osmC, PA1323, and PA1324, previously shown to be induced under osmotic stress50 and betaA, betaB genes encoding choline dehydrogenase and glycine betaine aldehyde dehydrogenase involved in osmoadaptation51 (Supplementary Data 9). Similar analyses for S. aureus infection in the murine osteomyelitis model showed a different pattern, with many stress responses engaged, where low iron and osmotic stress responses, were relatively lower than the others (Fig. 4c). The response to acidic stress was most pronounced and involved elevated expression of purC, E and F52. Osteomyelitis is associated with high level of inflammation and bone tissue destruction via osteoclastic resorption of bones53. Osteoclasts secrete hydrogen ions, collagenase, cathepsin K, and hydrolytic enzymes during acute infection where the hydrogen ions lead to

ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-23588-w | www.nature.com/naturecommunications
dissolution of bone minerals and an acidic microenvironment, which was reflected in our comparison with pronounced acidic stress response (Fig. 4c). We also noted increased expression of a gene (RSAU_000352), encoding a super-antigen-like exotoxin during acute infection and in acidic stress only in vitro, suggesting a possible functional role for this toxin under acidic microenvironments during infection (Supplementary Data 9). The second most pronounced response during the acute *S. aureus* infection was measured for virulence inducing condition, which in this case was treatment of bacteria with human serum (Fig. 4c). Taken together, comparison of in vivo bacterial responses to stress responses in PATHOgenex could provide insight to environmental cues at in vivo infection sites and reveal adaptation strategies employed by bacterial pathogens.

ncRNAs in bacterial stress responses. Analysing overall transcription, we observed that a substantial proportion of RNAs were transcribed from non-coding DNA regions, including sRNA, regulatory ncRNAs, 5′ untranslated regions (UTRs), 3′ UTRs, and intergenic regions (noteworthy, this does not include transcription of cis-antisense RNA encoded on CDS’s opposite strand). The proportion is expected to be even higher than calculated, since probably many sRNAs were excluded during sample preparation due to the >100 nt cut-off for library preparation. In some bacteria, expression of non-CDSs was higher than that of CDSs in hypoxia, nutritional downshift, and stationary phase conditions. For example, 72% of reads mapped to *V. cholerae* genome were transcripts encoded from non-coding regions under hypoxia. Corresponding numbers for *E. faecalis* were 62% under hypoxia and 60% at stationary phase, and for *N. meningitidis* was 53% under nutritional downshift (Supplementary Data 2). Noteworthy, for most species, the percent of transcripts from non-CDS was increased upon stress exposure in comparison to control (Fig. 5a). Similar proportions of non-CDS transcripts and also stress-induced upregulations were also shown in a previous study of *Salmonella*18.

The observed stress associated increases of transcripts from non-CDS in many cases correlated with increased expression of tmRNA, also known as SsrA, a conserved and abundant RNA in bacteria responsible to restore translation in detrimental situations54. Induction of tmRNA were particularly high at stationary phase, hypoxia, and upon nutritional downshift, which indeed are conditions associated with halted growth (Fig. 5b and Supplementary Data 10). In line with this, *smpB*, encoding a protein interacting with tmRNA, was identified as a USR with high PTDEX scores in those conditions. (Supplementary Fig. 6). The tmRNA level was however not correlated with expression ratios of non-CDSs for all species (Fig. 5a, b). For example in *Klebsiella pneumoniae*, expression of the carbon storage regulatory ncRNA CsrB, involved in many biological processes55, correlated with expression ratios of non-CDS regions in many stress conditions except stationary phase (Fig. 5a, d). For stationary phase we found high extent read mapping to 5′-UTR and 3′-UTR regions of a CDS (KPN_01149) encoding a hypothetical protein, which has 2 paralogs with diverse UTR stations. Another abundant ncRNA with potential to impact transcription widely spread among different bacteria is 6S RNA58. In *E. coli*, 6S RNA bind the σ70 RNA polymerase and inhibit transcription, promoting adaption to stationary phase and environmental stresses59. Since expression profiles of many other bacteria suggest similar involvement of 6S RNA in adaption to stationary phase, but also potentially different roles in other60, we analyzed of 6S RNA expression during different stresses in the PATHOgenex strains. The proportion of 6S RNA was here found to rather stable under most stress conditions in comparison to control, but increased under hypoxia, stationary phase, and virulence inducing conditions in some species (Supplementary Fig. 8). The proportion of 6S RNA in *C. jejuni* and *M. tuberculosis* stood out among the other species with higher expression levels. 6S RNA in *Mycobacteria*, known as Ms1 RNA has been reported to have altered binding to the RNA polymerase, not requiring the holoenzyme61, which might suggest an alternative role of this variant. However, also here the highest expression levels are seen for stationary phase and hypoxia, which is in line with that of many traditional 6S RNAs.

The PATHOgenex RNA atlas provides opportunities for exploring gene expression in human pathogens. All transcription data have been collected in an interactive database designed to serve researchers, the PATHOgenex RNA atlas (www. pathogenex.org). The data can be browsed by selecting strain(s) of interest and searched with one or multiple locus tags, protein ID, gene product, and PGFam groups. The database provides information of expression of pathways and operons across different stress conditions. As an example, we show that the three type VI secretion systems (T6SSs) operons in *P. aeruginosa* PAO1 are regulated under different stress conditions (Supplementary Fig. 9a). Such data can help in predicting operon structures, regulation of genes in operons under different stress conditions, or regulation of genes in specific genomic regions such as plasmids and genomic islands. PATHOgenex RNA atlas also allows the retrieval of most regulated gene groups under certain stress condition(s) in a wide range of bacterial pathogens. We show the top 5 highly regulated gene groups under virulence inducing conditions as an example (Supplementary Fig. 9b). The gene with highest PTDEX score under virulence condition was *adhE*, encoding bi-functional acetaldehyde-CoA dehydrogenase and alcohol dehydrogenase, previously shown to be important for *E. coli* and *S. enterica* virulence62,63. We could also see that *adhE* is differentially regulated in Gram-positive and -negative bacteria by retrieving differential expression levels (Supplementary Fig. 9c). Moreover, the database can also provide additional information on gene expression level in individual species under all tested conditions (Supplementary Fig. 9d).

Discussion

Bacterial stress responses, which include coordinated regulation and interactions between different gene products involved in a variety of biological processes are critical for survival of pathogens and therefore attractive potential targets for future antimicrobials. However, genetically diverse pathogens harbor both orthologues and distinct biological processes to adapt to stresses64, which hitherto has hampered investigations aiming at deciphering stress responses at global level. To meet this challenge, we have generated a high-quality resource with transcriptome datasets for comprehensive cross microbial analyses of orthologues and distinct biological processes. Comparisons of in vivo regulated genes of *S. aureus* and *P. aeruginosa* showed that the stress responses selected covered a majority of responses that occur during infection. The high quality of the PATHOgenex atlas rely on the
accuracy, depth, and dimension of the data. All experiments have been performed under similar conditions and resulting data have been handled similarly. This makes PATHOgenex a unique resource with high potential to provide novel information and also constitute a unique source for Big Data approaches. In contrast to challenging properties of existing data collections such as noisy, incomplete, or limited datasets originated from different labs with different experimental setups and bioinformatics analyses, the PATHOgenex RNA atlas provides clean and complete datasets for machine learning algorithms aiming to understand complex biological problems in infection biology.

The high number and diversity of bacterial species in the PATHOgenex datasets allowed generation of a score system, providing response-specific PTDEX scores for iso-functional gene groups. These scores can be used to analyze the big dataset in terms of differential expression and regulation of common and

Fig. 5 Complex stress conditions commonly involve differential expression from non-coding regions. a Ratio of percent of reads mapped to non-CDSs under each stress condition in comparison to control for all tested strains. Different stress conditions are indicated by color (right). The gray line indicates Stress/Control value equal to one. b Proportion of reads mapped to tRNA sequences of total number of mapped reads under each stress condition for all tested strains. c Proportion of reads mapped to tRNA sequences of total number of reads mapped to non-CDSs under each stress condition for all tested strains. d Proportion of reads mapped to the K. pneumoniae CsrB and to the region covering KPN_01149 CDS with 5′-UTR and 3′-UTR sequences under each stress condition. e Number and alignments of reads mapped to KPN_01149 and its UTRs in stationary phase. Green indicates forward reads, red indicates reverse reads, and blue indicates paired reads. The vertical dashed line indicates the position of the TSS of KPN_01149. f Proportion of reads mapped to SRS42 in S. aureus MRSA and MSSA genomes under each stress condition. g Number and alignments of reads mapped to SRS42 in S. aureus MRSA and MSSA strains in stationary phase. Colors represent reads as noted in e. Different stress conditions are indicated with colored dots on the right. The means of three replicates in each condition were used for the calculation in a-d, and f. Source data are provided as a Source Data file.
specific genes associated with certain responses in different bacterial groups. We here used the score to identify general USRs, a group of conserved genes involved in responses to multiple stresses in different bacteria, as well as species-specific USRs, which all can be interesting for exploring novel antimicrobial targets. The PTDEX score, which is indicative of regulation under particular stress condition(s), also provides hints about the function of products of genes encoding hypothetical and uncharacterized proteins. PATHOgenex can therefore be useful for designing experiments to reveal function of these gene products. This in turn can contribute with new information for functional annotations of microbial genes, which currently suffers from being poorly annotated, lacking information for a high proportion of genes. There is a clear need to increase the number of functionally annotated genes to facilitate biological understandings of transcriptomic data and the PATHOgenex RNA atlas can partly complement with expression levels and gene regulation data.

We also show examples where data in PATHOgenex can be used to get information of stress conditions encountered by pathogens during in vivo infections. Compared to the commonly used Gene Ontology and KEGG pathway mapping with a low number of functionally annotated genes, the use of PATHOgenex for comparative analyses increases possibilities to retrieve new information of bacterial microenvironments. We also found that significant proportions of transcriptional responses to stressful environments where from non-coding regions of the genomes. Besides known ncRNAs such as tmRNA and CsrB, we could show examples of ncRNAs that significantly increased in abundance during certain stresses in K. pneumoniae and S. aureus. Although not in the scope of this study, but given the high clinical relevance of these two pathogens, the potential role and function of these ncRNAs should be investigated further. Taken together, we provide comprehensive information of gene expression at a single gene level in individual species, and at a broader level including gene groups, PTDEX scores and expression in multiple species. Results including different cross-microbial comparisons, retrieval of information about bacterial stress responses during infection, as well as the overall screen of transcripts from non-coding regions of bacteria exposed to environmental stresses shows the high potential of the PATHOgenex databases as a data source. The dataset and associated findings have strength to give depth to comparative genomics studies and to more focused analyses of different pathogens. All together, PATHOgenex offers a rich source for researchers to generate novel hypotheses and design experiments accordingly, by that providing new opportunities for novel discoveries.

Methods

Bacterial growth and stress exposures. All bacterial strains shown in Fig. 1a were grown in their optimal growth medium and temperatures in laboratories specialized in each species. The strength of the stress with the different agents and associated exposure time was designed to be as similar and relevant as possible for each species (Supplementary Data 1). Minor changes were necessary for certain conditions; for example, the “low pH level” used for H. pylori strains were lower than for the others. Three bacterial cultures were grown overnight and then subcultured to a new culture vial with 1.3:1–1:10 dilution. The cultures were grown until exponential phase with an OD600 of 0.1–0.6 and exposed to the 11 stresses separately. Un-exposed cultures at exponential growth were used as controls for differential expression analysis. For nutritional downshift, bacterial cultures were spun down at 5,000 × g for 2 min, the supernatant removed, the pellets resuspended in 1× M9 supplemented with 0.1 M MgCl2 and 0.1 M CaCl2, and incubated for 30 minutes at ambient temperature for each strain. For hypoxia, 2 ml screw cap tubes were filled with bacterial culture at exponential growth and the caps were screwed on without leaving space for air. The stress exposures were stopped by adding 0.5% (final concentration) phenol-methanol solution (Supplementary Fig. 1a).

Total RNA isolation. One milliliter of triplicate bacterial cultures exposed to the stresses and un-exposed control culture were immediately pelleted by centrifugation at 5,000 × g at room temperature for 2 min after adding phenol–ethanol. The supernatants were removed and pellets resuspended in 0.5 ml Trizol for Gram-negative bacteria. For Gram-positive bacteria, the pellets were homogenized in Trizol by pipetting up and down 15 times. For Gram-positive bacteria, culture suspensions in Trizol were transferred to previously cooled bead beater tubes containing 0.1-mm glass beads and treated with Mini-Beadbeater (BioSpec Products Inc, USA) twice at a fixed speed for 45 s, and then cooled on ice for 1 min between the treatments. Culture homogenates were incubated at room temperature for 5 min and then supplemented with 0.2 ml chloroform, thoroughly mixed by shaking 10 times, and incubated for 3 min. After centrifugation at 12,000 × g at 4 °C for 15 min, the aqueous upper phase was carefully transferred to new RNase-free tubes. An equal volume of 1× Tris-HCl ethanol was added to the aqueous phase and mixed. The preparation was continued using the Direct-Zol RNA Miniprep Plus (Zymo Research, USA) RNA purification kit protocol. Total RNAs were eluted in RNase free water in RNase free tubes. The total RNA concentrations were measured using the Qubit BR RNA Assay Kit (ThermoFisher Scientific, USA) and RNA integrity confirmed on a 0.8% agarose gel in TBE buffer.

RNA-seq library preparation with RINtag-Seq. All rRNA-depleted RNA-seq library preparations were performed according to Shishin et al. (2015), with minor modifications. RINtag-Seq allows multiple library preparations in one tube, with initial tagging of total RNA samples with modified (5’P and 3’ Spc3) DNA barcoded adaptors, each harboring a unique 8 nt sequence to demultiplex individual libraries after sequencing. We combined the library preparation of three bacterial species from 30 exposed samples and one control sample, resulting in 36 library preparations in one tube for each bacterial strain. To tag the 36 replicates, 36 unique barcoded adapters were used with every three barcodes used for samples from the same stress conditions in all bacterial species for consistency (Supplementary Data 2). A total of 100 ng total RNA was used for each cDNA replicate. The cDNA was fragmented with Thermosensitive Alkaline Phosphatase buffer for 3 min at 94 °C, DNase treated, and dephosphorylated with a combination of TURBO DNase and Thermosensitive Alkaline Phosphatase in 1× FastAP buffer for 30 min at 37 °C. Fragmented, DNase-treated, and dephosphorylated total RNAs were cleaned with a 2× reaction volume of Agencourt RNAClean XP beads. Cleaned total RNAs were incubated with 100 pmol of the unique DNA barcode adapters at 70 °C for 3 min, and then ligated with T4 RNA ligase 1 for 90 min at 22 °C. The ligation was stopped and the enzyme denatured by the addition of RLT buffer. The 36 denatured ligation mixes were then pooled and cleaned in a Zymo Clean & Concentrator-5 column according to the manufacturer’s 200 nt cut-off protocol. The RNA samples were incubated with 1 μl of RNase-free water. Ribosomal RNA was depleted using the Ribo-ZeroTM Magnetic Gold Kit (Bacteria) according to the manufacturer’s instructions. The first-strand cDNA of each pool was generated using an AffinityScript Multiple Temperature cDNA synthesis kit with 50 pmol of AR2 primer at 55 °C for 55 min. The RNA was degraded by adding a 10% reaction volume of 1 N NaOH at 70 °C for 12 min and the reaction neutralized with an 18% reaction volume of 0.5 M acetic acid. After cleaning the reverse transcription primers with a 2× reaction volume of RNAClean XP beads, the 3Tr3 adapter was ligated with T4 RNA ligase 1 at 22 °C with overnight incubation. The second ligation was cleaned first with a 2×, and secondly a 1× reaction volume of RNAClean XP beads. The cleaned cDNA was used as the template for PCR reaction with FailSafeTM PCR enzyme mix using 12.5 pmol 2P_univPS as forward primers and 12.5 pmol ScriptSeqTM Index (barcode) PCR primers as reverse primers. The PCR cycles were as follows: 95 °C for 3 min, followed by 12 cycles at 95 °C for 30 s, 55 °C for 30 s, and 68 °C for 3 min, and then finishing at 68 °C for 7 min. The PCR product was cleaned first with a 1.5×, and secondly 0.8×, reaction volume of AMPure beads and eluted in RNase-free water. The library concentrations were measured using the QubitTM dsDNA HS Assay Kit and the library insert size determined by the Agilent DNA 1000 Kit in a 2100 Electrophoresis Bioanalyser Instrument (Agilent, USA). The oligonucleotides used in RINtag-Seq sequencing are shown in Supplementary Table 1.

RNA-seq data analysis. The RNA-seq libraries generated by RINTag-Seq were sequenced by either single-end or paired-end Illumina sequencing at SciLifeLab, Stockholm. Each library harboring a pool of 36 RNA samples was demultiplexed according to the unique 8 nt on the ligated barcode adaptors to separate reads from each replicate. The number of reads for the different barcoded samples did not reveal significant variation across conditions and species, excluding bias regarding the depletion of one or more barcodes during demultiplexing. The sequencing reads generated in this study were deposited in GEO with accession number GSE152295. Demultiplexed reads were then mapped to the genome of the species to which the RNAs belong in an annotation-independent manner. The accession numbers of the RefSeq reference genomes used for each strain can be found in Supplementary Data 1. The number of reads mapped to CDSs, rRNAs, and tRNA was calculated according to the annotation files. Read mapping to reference genomes showed efficient rRNA depletion, with only 0.1–5% reads mapping to rRNAs and tRNAs, except for Campylobacter jejuni, two of the H. pylori strains, and Neisseria gonorrhoeae, which had higher proportions mapping to rRNA (Supplementary Data 2). The library sequencing for these species was deep enough to cover 94% of the coding sequences (CDSs), with >10 reads per CDS in the least efficiently depleted library. The number of reads mapped to the non-coding regions...
were calculated by subtracting the reads mapped to the annotated regions from the total number of mapped reads. The sequencing reads from primary transcripts of *K. pneumoniae* with accession number SRR409498 were downloaded from the Sequence Read Archive (SRA) and mapped to the reference genome used in this study. In vivo and in vitro RNA-seq reads of *S. aureus* strain 6850 (accession number ERP005459) and *P. aeruginosa* PA01 (accession number GSE119356) were downloaded from GEO and mapped to RefSeq reference genome CP006706 and NC_002716, respectively. The mappings of in vivo originates reads were performed with at least 95% correct matching to discriminate reference genome-specific reads and avoid mapping of host transcripts. Reads from *P. aeruginosa* lung lower peripheral lobe sample were excluded due to low transcriptome coverage. All of the demultiplexing and read mapping steps were performed in CLC Genomic Workbench (Qiagen, USA).

For differential gene expression analysis, the trimmed mean of M values (TMM) normalization method was used to normalize the sequencing depth of the individual libraries. For each bacterial species, the comparisons were performed between multiple groups and the control sample so that the full dataset could be used for fitting the generalized linear model. Using the complete datasets with multiple group comparisons allowed us to determine whether a gene has unstable expression for which the variation is random or is differentially expressed in response to the stress. Read values for genes with a maximum group mean expression (the maximum average TPM value in the statistical comparison group) ≤20 were removed and a threshold of 1.5-fold change with FDR-adjusted p-value <0.05 was employed to determine differential expression.

**Comparisons of in vivo differentially regulated genes of *S. aureus* and *P. aeruginosa* to in vitro stress responses.** For comparison between the *S. aureus* strain 6850 used strain for the mouse infection study and in vitro induced stress responses of *S. aureus* MSSA A476 in PATHOgenex, homologs genes were identified with PATRIC’s Prokaryote Comparison Service. For *P. aeruginosa* PA01 the same strain was used as reference genome. Co-regulated genes in vivo and each in vitro stress conditions were determined with the same differential expression profiles such as down-regulated in both conditions or up-regulated in both conditions.

**Generation of a phylogenetic tree of 32 strains.** A phylogenetic tree of the 32 strains based on NCBi taxonomy was generated with PhyloT in Newick format. The visualization of the tree was achieved using iTol.

**Clustering 32 strains based on expression profiles.** Hierarchical clustering, with Euclidian distance, of 32 strains based on the percent of genes expressed (TPM ≥10) in all 12 conditions, percent of genes expressed in at least one condition, and percent of genes not expressed in any of the conditions was performed with ClustVis.

**Clustering 105,088 genes with orthology/homology.** The KEGG orthology groups were assigned for clustering genes based on functional orthology. The amino acid sequences of all CDSS for each strain were uploaded to GhostKOALA to assign the best KO group in the genus_prokaryotes KEGG GENES database. Clustering based on isofunctional homologs was performed with PATRIC’s Prokaryote Comparison Service. The amino acid sequences of all CDSS for each strain were compared using its own annotated genome, if available, or the closest strain to assign the best PGfam group for each CDS.

**PTDDEX score calculation.** The PTDDEX score was calculated for the orthology/homology groups using Python scripts and Jupyter notebooks, relying on the open-source libraries NumPy and Pandas. The PTDDEX score formula relies on the definition of differentially expressed genes given in “RNA-Seq Analysis” above and the equation shown in Eq. 1, where \( n_{\text{match}}(\text{on}) \) is the number of differentially expressed genes, maximum group mean \( >20 \), fold change \( >1.75 \) with FDR-adjusted p-value <0.05, in the orthology/homology gene group. \( n_{\text{match}}(\text{total}) \) is the total number of genes in the group, \( n_{\text{match}}(\text{on}) \) is the number of species for which at least one of the orthology group genes is differentially regulated, \( n_{\text{match}}(\text{total}) \) is the total number of strains for which at least one of the orthology group genes is present, and \( n_{\text{match}}(\text{database}) \) is the total number of strains in the database (\( n_{\text{match}}(\text{database}) = 32 \) for general PTDDEX scores, 21 for Gram-negative specific, 9 for Gram-positive specific PTDDEX scores). The additional logarithmic component is a weight factor to give more relative importance to orthology/homology groups with more genes.

**Generation of co-expression modules and KEGG pathway enrichments.** The co-expression Modules Identification tool (CEMMI tool) was implemented on PGfam groups of Gram-negative and -positive strains to find PGfam groups with similar PTDDEX score patterns over the 11 stress conditions. PGfam groups associated with each module can be found in Supplementary Data 5 and 6. Each PGfam group associated with the particular module, if possible, was converted in KO groups and KO groups were used in the KEGG orthology database to map pathways.

**PATHOgenex website construction.** PATHOgenex was built using PHP Laravel Framework 6.18.2 and PHP 7.4.4 for server-side data processing, Javascript, ECMAScript 2015 for the front end, and D3.js 5.16.0 and Plotly 1.40.0 libraries for the generation of the interactive visualizations. Linux distribution CentOS-8 with the 64-bit kernel 4.18.0 running on four processor cores and 64 Gb of RAM is used to host the web service on the in-house computational cluster.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability** The datasets with sequencing reads and processed data generated/analyzed during this study are available at GEO under accession number GSE152295. Previously published *P. aeruginosa* RNA-seq data from mouse osteomyelitis model and pure culture are available at GEO under accession number GSE119356 and RNA under the accession number PRRJEB003, respectively. The PATHOgenex RNA atlas with global expression profiles of the 32 bacterial pathogens under 11 stress conditions and an un-exposed control condition as well as PTDDEX scores of PGfam gene groups are publicly available at www.pathogenex.org. Source data are provided with this paper.

**Code availability** The custom scripts employed to calculate PTDDEX scores from differential expression analysis data are available at Zenodo (https://zenodo.org/record/4708491#.Y1D1oOutV5E).

Received: 6 October 2020; Accepted: 5 May 2021; Published online: 02 June 2021

**References**

1. Malachowa, N., Kobayashi, S. D., Sturdevant, D. E., Scott, D. P. & DeLeo, F. R. Insights into the Staphylococcus aureus-host interface: global changes in host and pathogen gene expression in a rabbit skin infection model. *PLoS ONE* 10, e0117713 (2015).
2. Mandlik, A. et al. RNA-Seq-based monitoring of infection-linked changes in Vibrio cholerae gene expression. *Cell Host Microbe* 10, 165–174 (2011).
3. Avican, K. et al. Reprogramming of Yersinia from virulent to persistent mode revealed by complex in vivo RNA-seq analysis. *PLoS Pathog*. 11, e1004600 (2015).
4. Nuss, A. M. et al. Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host-pathogen transcriptomes. *Proc. Natl Acad. Sci. USA* 114, E791–E800 (2017).
5. Klinkert, B. & Narberhaus, F. Microbial thermosensors. *Cell Mol. Life Sci.* 66, 2661–2676 (2009).
6. Lund, P., Tramonti, A. & De Biase, D. Coping with low pH: molecular strategies in neutrophilic bacteria. *FEBS Microbiol. Rev.* 38, 1091–1125 (2014).
7. Miller, E. F. & Maier, R. J. Ammonium metabolism enzymes aid Helicobacter pylori acid resistance. *J. Bacteriol.* 196, 3074–3081 (2014).
8. Hofmann, A. F. & Hagey, L. R. Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics. *Cell Mol. Life Sci.* 65, 2461–2483 (2008).
9. Gonzalez-Escobedo, G., Marshall, J. M. & Gunn, J. S. Chronic and acute infection of the gall bladder by Salmonella Typhi: understanding the carrier state. *Nat. Rev. Microbiol.* 9, 9–14 (2011).
10. Prouty, A. M., Schwesinger, W. H. & Gunn, J. S. Biofilm formation and interaction with the surfaces of gallstones by Salmonella spp: understanding the carrier state. *Nat. Rev. Microbiol.* 9, 760, 2640–2649 (2002).
11. Loh, J. T., Torres, V. J. & Cover, T. L. Regulation of Helicobacter pylori cagA expression in response to salt. *Cancer Res.* 67, 4709–4715 (2007).
12. Ishikawa, T. et al. Pathoadaptive conditional regulation of the type VI secretion system in Vibrio cholerae O1 strains. *Infect. Immun.* 80, 575–584 (2012).
13. Fang, F. C., Frawley, E. R., Tapscott, T. & Vazquez-Torres, A. Bacterial stress responses during host infection. *Cell Host Microbe* 20, 133–143 (2016).
14. Garbe, J. & Collin, M. Bacterial hydrolysis of host glycoproteins - powerful protein modification and efficient nutrient acquisition. *J. Innate Immun.* 4, 121–131 (2012).
15. Cassat, J. E. & Skaar, E. P. Iron in infection and immunity. *Cell Host Microbe* 13, 509–519 (2013).
16. Vazquez-Torres, A. Redox active thiol sensors of oxidative and nitrosative stress. *Antioxid. Redox Signal* 17, 1201–1214 (2012).
17. Schaffer, K. & Taylor, C. T. The impact of hypoxia on bacterial infection. *FEBS J.* 282, 2260–2266 (2015).
34. Davis, J. J. et al. PATTyFams: protein families for the microbial genomes in the NATURE COMMUNICATIONS | (2021) 12:3282 | https://doi.org/10.1038/s41467-021-23588-w | www.nature.com/naturecommunications

32. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: a Bioinformatics Database and Analysis Resource Center.

40. Wattam, A. R. et al. Improvements to PATRIC, the all-bacterial reference database. J. Bacteriol. 190, 2004–2013 (2008).

18. Kroger, C. et al. An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. Cell Host Microbe 14, 683–695 (2013).

19. Smith, C., Stringer, A. M., Mao, C., Palumbo, M. J. & Wade, J. T. Mapping the Regulatory Network for Salmonella enterica Serovar Typhimurium Invasin. mBio 7(2016).

24. McHugh, J. P. et al. Global iron-dependent gene regulation in bacteria. J. Bacteriol. 194, 5495–5503 (2012).

26. Almiron, M., Link, A. J., Furlong, D. & Kolter, R. A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli. Genes Dev. 6, 2646–2654 (1992).

29. Dubbs, J. M. & Mongkolsuk, S. Peroxide-sensing transcriptional regulators in bacteria. J. Mol. Microbiol. 10, 718–722 (1985).

50. Aspedon, A., Palmer, K. & Whiteley, M. Microarray analysis of the osmotic stress response in Pseudomonas aeruginosa. J. Bacteriol. 188, 2721–2725 (2006).

37. Russo, P. S. T. et al. CEMiTool: a Bioconductor package for performing multivariate data using Principal Component Analysis and heatmap. Nat. Commun. 10, 3397 (2019).

46. Rabin, R. S. & Stewart, V. Dual response regulators (NarL and NarP) interact with dual sensors (NarX and NarQ) to control nitrate- and nitrite-regulated gene expression in Escherichia coli K-12. J. Bacteriol. 175, 3259–3268 (1993).

47. Jayaraman, P. S., Cole, J. A. & Busby, S. J. Mutational analysis of the nucleotide sequence at the FNR-dependent nirB promoter in Escherichia coli. Nucleic Acids Res 17, 135–145 (1989).

48. Zhou, C. Y. & Fey, P. D. The acid response network of Staphylococcus aureus. J. Mol. Biol. 329, 2629 (2018).

54. Keiler, K. C. Physiology of tmRNA: what gets tagged and why? Curr. Opin. Microbiol. 10, 169–175 (2007).

55. Liu, M. Y. et al. The RNA molecule CorA binds to the global regulatory protein CorA and antagonizes its activity in Escherichia coli. J. Bacteriol. 172, 1750–17510 (1999).

56. Berner, A., Nolte, L., Mattino, R. & Kolter, R. Dispersal of 6S RNA among bacteria. RNA Biol. 11, 1746–1748 (2014).

58. Wehner, S., Damm, K., Hartmann, R. K. & Marz, M. Dissemination of 6S RNA bacteria. RNA Biol. 11, 1467–1478 (2014).

59. Wassarman, K. M. 6S RNA: a small RNA regulator of transcription. Curr. Opin. Microbiol. 10, 164–168 (2007).

61. Hnilicova, J. et al. Ms1, a novel sRNA interacting with the RNA polymerase, regulates gene expression and infectivity. J. Anim. Sci. Biotechnol. 9, 569 (2018).

62. Beckham, K. S. et al. The metabolic enzyme AdhE controls the virulence of Pseudomonas fluorescens. J. Bacteriol. 189, 6265–6272 (2007).

68. Zitnik, M. et al. Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities. Inf. Fusion 50, 71–95 (2019).

71. Avican, K., et al. RNA atlas of human bacterial pathogens uncovers stress response genes regulated by the regulatory network in human infection. Proc. Natl Acad. Sci. USA 115, ES125–ES134 (2018).

We are grateful to all the expert labs for opening their lab to us and help with bacterial culturing and stress exposure experiments. We thank Drs. Peter Lind, Teresa Frisan, and Saskia Erttmann for critical reading of the manuscript. The work was supported by Knut and Alice Wallenberg Foundation (No. 2016.0063), Swedish Research Council (No. 2018-02855), and Insamlingsstiftelsen, Medical Faculty at Umeå University to M
Author contributions
K.A. and M.F. conceived and supervised the project, and wrote the manuscript with input from J.A., M.T., F.M., J.T., K.B., and M.R. K.A. performed the experiments and analyzed the data. M.T., K.B., K.A., and M.F. designed and M.T. calculated PTDEX scores. J.A., J.T., K.A., and M.F. designed and J.A. constructed the PATHOgenex RNA atlas and webpage.

Funding
Open access funding provided by Umeå University.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-23588-w.

Correspondence and requests for materials should be addressed to K.A. or M.F.

Peer review information Nature Communications thanks Anais Le Rhun, Jonathan Livny, and the other, anonymous, reviewer for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021