ENVIRONMENTAL MICROBIOLOGY

A Broad Continuum of *E. coli* Traits in Nature Associated with the Trade-off Between Self-preservation and Nutritional Competence

Estela Ynes Valencia[1] · Jackeline Pinheiro Barros[1] · Thomas Ferenci[2] · Beny Spira[1]

Abstract

A trade-off between reproduction and survival is a characteristic of many organisms. In bacteria, growth is constrained when cellular resources are channelled towards environmental stress protection. At the core of this trade-off in *Escherichia coli* is RpoS, a sigma factor that diverts transcriptional resources towards general stress resistance. The constancy of RpoS levels in natural isolates is unknown. A uniform RpoS content in *E. coli* would impart a narrow range of resistance properties to the species, whereas a diverse set of RpoS levels in nature should result in a diverse range of stress susceptibilities. We explore the diversity of trade-off settings and phenotypes by measuring the level of RpoS protein in strains of *E. coli* cohabiting in a natural environment. Strains from a stream polluted with domestic waste were investigated in monthly samples. Analyses included *E. coli* phylogroup classification, RpoS protein level, RpoS-dependent stress phenotypes and the sequencing of *rpoS* mutations. The most striking finding was the continuum of RpoS levels, with a 100-fold range of RpoS amounts consistently found in individuals in the stream. Approximately 1.8% of the sampled strains carried null or non-synonymous mutations in *rpoS*. The natural isolates also exhibited a broad (>100-fold) range of stress resistance responses. Our results are consistent with the view that a multiplicity of survival-multiplication trade-off settings is a feature of the species *E. coli*. The phenotypic diversity resulting from the trade-off permits bet-hedging and the adaptation of *E. coli* strains to a very broad range of environments.

Keywords RpoS · Trade-offs · Stress resistance · Polymorphism · *Escherichia coli* · Natural isolates

Introduction

Evolutionary pathways are often limited by trade-offs, meaning that not all properties of bacteria can be optimised simultaneously. Traits crucial to bacterial competitiveness are frequently subject to trade-offs [1], and one pair of mutually exclusive properties is between resistance to environmental stresses and the ability to grow rapidly, classically called a survival-multiplication trade-off [2]. The molecular basis of this resource allocation trade-off is well understood in *E. coli* [3]. Bacterial general stress resistance is positively regulated by RpoS (σ^S), a sigma factor that controls the transcription of genes associated with bacterial survival under stressful situations and during the stationary phase, conditions that elevate the level of RpoS [4–6]. Vegetative growth, on the other hand, is promoted by the interaction of RNA polymerase with the housekeeping sigma factor RpoD (also known as σ^D or σ^{70}) [7]. At the core of the resistance/growth trade-off is the competition between the sigma factors RpoD and σ^S for transcriptional space and the allocation of resources to growth or resistance. In the exponential growth phase in the absence of stress, RpoD does not compete with other sigma factors for RNA polymerase and can direct transcription towards growth functions. However, upon entering the stationary phase or with any reduction in growth rates, σ^S accumulates and competes with σ^D for

✉ Beny Spira
  benys@usp.br

[1] Departamento de Microbiologia, Instituto de Ciências Biomédicas Universidade de São Paulo, São Paulo, SP, Brazil

[2] Present address: School of Life and Environmental Sciences, University of Sydney, 6/403 Pacific Highway, Sydney, New South Wales 2070, Australia

the association with the core RNA polymerase [3, 4, 8], resulting in lower expression of RpoD-dependent genes. This sigma factor competition in a strain determines whether RpoD (growth) or σ^S-dependent genes (resistance) are transcribed and how the trade-off is set by the level of RpoS [1].

If a population is submitted to conditions conducive to growth in the absence of environmental stresses, the cost of stress protection is relieved by the selection of mutations in *rpoS* or in genes that control *rpoS* expression. Additionally, null or low-RpoS bacteria are selected in the laboratory under several conditions of persistent stationary phase, nutrient limitation and bacterial competition [9–14]. In addition to mutations in *rpoS* itself, other mutations can reduce the amount of RpoS. For instance, mutations in *hfq* and in *spoT* have been co-selected with *rpoS* mutations in populations growing under steady-state glucose [9, 15] or phosphate limitation [13]. RpoS content and function are controlled at multiple levels (transcription, translation, post-translation, protein stability) and governed by several different inputs, including regulatory proteins and small RNAs (RssB, Crl, Hfq, ArcZ, DsrA, RprA) and the SpoT-generated alarmone (p)ppGpp, which are themselves regulated by a multitude of other factors [5, 16–18]. Thus the stress-growth trade-off might be affected by many mutational targets.

An assessment of RpoS variability focusing on *rpoS* gene mutations, without measuring protein levels in isolates, misses an important point, however. The intrinsic levels of RpoS (and stress resistance) in a cell are affected by multiple gene products, not just *rpoS*, as noted above. The variation of RpoS levels in natural isolates has not been assessed, except in a few clinical examples [19]. In other studies, the quantitation of RpoS protein in the different isolates was either not reported [20] or measured in a few isolates [21].

We used a longitudinal approach to assess the status and level of RpoS in a potentially stressful secondary habitat of *E. coli*. Bacterial samples were collected monthly from an urban polluted stream and assayed for RpoS and RpoD levels in all 328 isolates. The presence of *rpoS* mutations in these isolates was confirmed by gene sequencing. Additionally, an independent high-throughput sequencing analysis revealed the presence of *rpoS* mutations in non-cultured samples. We also investigated the association of RpoS variation with the diversity of phenotypes amongst the natural isolates, in terms of stress sensitivity as well as the negative impact on RpoD-dependent effects. These results overall suggest that the trade-off between growth and resistance to environmental stresses is subject to constant selection and leads to a continuum of characteristics amongst natural isolates, ensuring that some strains will have the appropriate set of responses to new challenges. Our results provide strong experimental evidence for the notion that trade-offs are a major source of diversity in biological systems [1, 22].

Methods

Collection of Water Samples, Isolation and Selection of *E. coli* Strains

We collected monthly water samples for a 12-month period from the Pirajuçara stream (23°33′53.31″S; 46°42′49.86″W) in São Paulo, SP, Brazil. This is a heavily polluted stream into which domestic sewage is discharged without any treatment. Each water sample was tested for temperature, pH, dissolved oxygen concentration and turbidity. The first three analyses were performed in loco, while turbidity (OD_{860}), which roughly indicates the concentration of organic matter in the samples [23], was measured in a spectrophotometer in the laboratory. For the measurement of dissolved oxygen, ready-made kits (Chemetrics Inc.) were used.

Bacteria were isolated by first plating 0.1 ml of a 100X dilution of the water samples in 0.9% sterile NaCl on MacConkey medium followed by incubation at 30°C for 24 h. Red colonies were re-isolated on MacConkey plates to guarantee the purity of the colonies and subjected to biochemical tests by inoculation in selective media EPM, Mili and citrate (Probac do Brasil). Colonies that showed the expected pattern for *E. coli* (indol positive, citrate negative, urease negative and hydrogen sulphide negative) were re-streaked on L-agar plates and incubated overnight at 30°C. Isolated colonies from these plates were suspended in 15% glycerol and frozen at −70°C. Samples diluted 100× were spread on MacConkey plates, giving rise to approximately 200 colonies/plate following an overnight incubation at 30°C. From this, we estimated that the average concentration of culturable Gram-negative bacteria in the stream waters was about 2×10^5 bacteria/ml. Normally one hundred lactose-positive colonies from each sample were re-isolated on MacConkey plates and submitted to a series of biochemical tests (described in the Methods section). On average, about 1/3 of the Lac$^+$ colonies were identified as *E. coli*, but a few of those strains were later found out to belong to other Enterobacteriaceae species.

Phylogroup Determination

The *E. coli* isolates were classified according to the phylogenetic group (A, B1, B2, C, D, E, F and Clade I) described by Clermont et al. [24]. The PCR amplifications were performed in 50-μl total volume containing 25-μl Master Mix PCR GoTaq (Promega), 1 ml of each primer (ChuA.1b - 5′-ATGGTACCGGACGAACCAAC; ChuA.2 - 5′-TGCCGCCAGTACCAAAGACA; YjaA.1b - 5′-CAAACGTGAAGTGTCAGGAG; YjaA.2b - 5′-AATGCGTTCCTCAACCTGTG; TspE4C2.1b - 5′-CACTATTCGTAAGGTCATCC; TspE4C2.2b − 5′-AGTTTATCGCTGCGGGTCGC; AceK.f - 5′-AACGCTATTCGCCAGCTTGC; ArpA1.r -5′-TCTCCCCA

TACCGTACGCTA) and 1 ml of a fresh colony suspension in deionised water that was used as a template. To distinguish between groups C and E, the specific primers trpAgpC.1/trpAgpC.2 (5′-AGTTTTATGCCCAGTGCGAG/5′-TCTG CGCCGGTCACGCCC) or ArpAgpE.f/ArpAgpE.r (5′-GATTCCATCTTGTCAAAATATGCC/5′-GAAA AGAAAAAGAATTCCCAAGAG) were used. We conducted the PCR reactions as follows: 4 min 94°C; 30 cycles of 5 s at 94°C, 20 s at 57°C (group E) or 59°C (quadruplex or group C); and a final extension step of 5 min at 72°C. The amplicons were resolved on a 1.8% agarose gel.

Growth Media

We grew bacterial cultures in lysogeny broth (LB) [25], LB without NaCl (for the dehydration stress assay) or semi-rich medium A (for AP assays) [26]. When required, 50-µg/ml kanamycin or 20-µg/ml chloramphenicol was added.

Enzyme-Linked Immunosorbent Assay (ELISA) of RpoS

We devised a modified ELISA protocol based on the method reported by Arif et al. [27]. The main changes to the original protocol we implemented were a sequential freeze-thaw of bacterial cultures for cell disruption and the use of lower concentrations of primary antibodies and longer incubation times. This latter adjustment resulted in a considerable reduction in the use of primary antibodies, an important feature specially when many hundreds of assays are performed. Briefly, we grew 2-ml cultures of each strain overnight (16–18 h) at 37°C in LB medium. The cultures were then centrifuged at 6000 ×g for 5 min at 25°C, and the bacterial pellets were washed with 1 ml of 50-mM Tris-Cl pH 8, centrifuged again and suspended in 1-ml lysis buffer (50-mM Tris-Cl pH 8, 0.5-mg/ml lysozyme, 0.1-mM PMSF and 0.06% DNAse). The suspensions were moved to ice for 30 min and then submitted to 7 cycles of freeze-thawing in liquid nitrogen. The samples were then centrifuged at 18,000 ×g for 20 min at 4°C to precipitate the debris. The supernatant containing the total protein extract was diluted (1:1) with Tris-buffered saline (TBS) pH 7.5 (50-mM Tris, 150-mM NaCl), and 100 µl of the diluted samples were used to coat the wells in a 96-well ELISA plate (COSTAR 3590). We then incubated the plates for 16 h at 4°C, washed twice with TBST buffer (TBS containing 0.05% Tween-20) and blocked with 100-µL Blotto (TBS, 1% skim milk powder) for 1 h at room temperature. The plates were washed twice with TBST, and then we added100 µl of a 1:1500 anti-RpoS or 1:7000 anti-RpoD monoclonal antibodies (Neoclone) diluted in Blotto to each well and incubated for 16 h at 4°C. We washed the wells 3 times with TBST and added 100-µl goat anti-mouse peroxidase-conjugated

antibody (diluted 1:1000) to each well, followed by incubation for 1 h at room temperature. The plates were washed again 7 times with TBST, and 50 µL of a 3,3′,5,5′-tetramethylbenzidine (TMB) substrate solution (Life Technologies) was added to each well. The plates were quickly covered with aluminium foil and incubated at room temperature for 15 min. Twenty-five µl of 1-M phosphoric acid was then added to stop the reaction. The absorbance 450 nm of each well was measured in an Epoch microplate spectrophotometer (Biotek). We calculated the relative amount of RpoS by dividing the A_{450} value of RpoS by that of RpoD. We assayed each isolate at least twice with each primary antibody—anti-RpoS and anti-RpoD. Duplicates of MG1655 and its derivatives (MG1655 Δ*rpoS* and MG1655 Δ*rssB*) were included in every 96-well plate. To normalise the RpoS/RpoD ratios across many different plates, each individual RpoS/RpoD value was divided by the MG1655 RpoS/RpoD value of the corresponding plate.

RpoS Sequencing

The *rpoS* ORF was amplified by PCR using the forward oligonucleotide 429F (5′-GGAACAACAAGAAGTTAAGG) and one of the following reverse primers: *rpoS* reverse ext (5′-AGCCTCGCTTGAGACTGGCC-3′); 10086-R (5′-GTGTTAACGACCATTCTCGG-3′); 1730-R (5′-GTAT GGGCGGTAATTTGAC-3′); E2348-R (5′-AAAG GCCAGCCTCGCTTGA-3′); *rpoS*21 (5′-AGCG TTCCATCAGTTACGACAGC-3′); *rpoS*23 (5′-GACT CAGGGGTTCTGGATTGTGACC-3′). We carried out the reactions using the GoTaq® Green Master (Promega) mix as recommended by the manufacturer. The amplicons were purified with the Wizard SV Gel and PCR Clean-Up System kit (Promega) and sequenced using the oligos described above and the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). The sequencing products were resolved in an ABI 3730 DNA Analyser (Life Technologies—Applied Biosystem). The *rpoS* sequences have been deposited in the GenBank database under the accession numbers MT912641–MT912680.

Glycogen Accumulation and Catalase Activity Assays

For the qualitative assessment of glycogen accumulation, we spotted 5 µl of each LB overnight culture (OD_{600} = 1.0) on L-agar plates and incubated overnight at 37°C, followed by incubation for 24 h at 4°C, at which time we flooded the plate with an iodine solution [28]. The level of glycogen accumulation was qualitatively assessed according to the intensity of the brown colour produced by the staining—0, 1, 2 or 3. Catalase activity was qualitatively assessed by dispensing 10 uL of hydrogen peroxide on top of overnight bacterial patches

grown on L-agar plates and recording the level of bubbling on a scale of 0, 0.5 and 1.0.

Alkaline Phosphatase Assay

Bacteria were seeded in 24-well plates containing medium A supplemented with 1-mM Pi and incubated at 37°C for 16 h. On the next day, we diluted the cultures 1/100 in medium A (without Pi addition) and incubated again at 37°C for 24 h. We transferred 0.2 ml of each culture to a 96-well microplate to measure turbidity at 600 nm, and in parallel 20 µl of each culture was then transferred to a new 96-well plate containing 180 µl of 1-mg/ml p-nitrophenylphosphate dissolved in 1-M Tris pH 8. The plate was immediately moved to the Epoch microplate spectrophotometer, where the absorbance at 410 nm of each well was measured at 50-s intervals for 10 min. The slope of the increase in A_{410} was used to calculate the enzyme specific activity using the formula: $\frac{slope(A_{410}) \times 10}{OD_{600}}$

Stress Assays

For the acid stress assays, we grew bacteria overnight in LB medium followed by dilution to an approximate final concentration of 2×10^3 cells per ml in 0.9% NaCl or in modified EG buffer (0.4% glucose; 73-mM K_2HPO_4; 17-mM $NaNH_4HPO_4$; 0.8-mM $MgSO_4$; 10-mM citrate; 1.5-mM glutamate; pH 2.0). The bacterial suspensions in EG buffer were incubated at 37°C, and 100-µl samples were withdrawn at 5, 15, and 30 min, seeded on L-agar plates and incubated overnight at 37°C. In parallel, 100-µl aliquots from the bacterial suspensions in 0.9% NaCl were plated on L-agar and incubated overnight to determine the precise initial concentration of bacteria. On the next day, we counted the CFU and calculated the per cent of survival by dividing the number of CFU at the various time points by the initial bacterial concentration. The dehydration stress assay was conducted according to the protocol described by Chen and Goulian [29]. Briefly, 10 µl of each bacterial culture grown for 18 h in LB (–NaCl) were inoculated in a 96-well plate well. We placed the plate in a Tupperware containing silica gel blue 4–8 mm (Contemporary Chemical Dynamics) and subsequently incubated at 25°C for 3 days. Following the incubation period, we added 100-µl LB (–NaCl) to each well. The bacterial suspensions were immediately diluted in phosphate-buffered saline (PBS) and plated on L-agar to determine the CFU of the surviving bacteria. Per cent survival was calculated by the following formula:

$$\%survival = \left(\frac{CFU\,after\,dehydration}{Initial\,CFU} \right) \times 100$$

The initial number of bacteria in each well was determined by serially diluting 100 µl of the overnight culture in 0.9% NaCl and plating on L-agar.

DNA Extraction and High-Throughput Analysis of *rpoS* in Water Samples

We collected two independent water samples from the Pirajuçara stream, which were immediately transported to the laboratory. The extraction of genomic DNA from the water samples was adapted from the protocol described by Tabatabaei et al. [30]. Three liters of water were filtered through 40-mm qualitative filter paper (Cienlab, Brazil) followed by another filtration through 11-mm paper (Whatman 1001-932 Grade 1). The samples were centrifuged for 15 min at 8600 ×g, the pellet was suspended in 10-ml of 0.5-M EDTA (pH 8.0) and incubated at room temperature for 10 min. We then added 10 ml of lysis buffer (10-mM Tris, 1-mM EDTA and 2-mg/ml lysozyme, pH 8.0) and incubated the samples at 37°C for 30 min. SDS was added to a final concentration of 0.5 %, and the samples were further incubated at 70°C for 15 min. An equal volume of aqueous phenol/chloroform (1:1) was added followed by centrifugation at 8600 ×g for 10 min. at 4°CThe phenol/chloroform extraction was performed twice. The DNA in the aqueous phase was precipitated by adding 1/10 volume of sodium acetate (3 M, pH 5.2) and an equal volume of cold isopropanol, followed by incubation at −20°C for 15 min. Samples were then centrifuged at 8600 ×g for 10 min at 4°C. The gDNA pellet was washed with 70% ethanol and centrifuged at 8600 ×g for 10 min at 4°C. Finally, the gDNA pellet was incubated at 37°C for 1 h and suspended in 1-ml TE buffer (10-mM Tris-HCl, 1-mM EDTA, pH 8.0). We further purified the gDNA samples using the QIAamp DNA Stool kit (Qiagen) as recommended by the manufacturer.

The high-throughput DNA sequencing analysis was performed by BPI-Biotecnologia (Botucatu, São Paulo, Brazil). Briefly, DNA was quantified in a Qubit® 3.0 Fluorometer using the kit Qubit™ dsDNA BR Assay (Thermo Fisher Scientific). PCR amplification was performed using GoTaq® Colourless Master Mix 2x (Promega) and two oligo pairs, one for the 5′ half and the other for the 3′ half of the gene—*rpoS*_F1 TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGATGAGTCAGAATACGCTGAA; *rpoS*_R1 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG CCTTTACGATGTGAATCGGC; *rpoS*_F2 TCGTCGGC AGCGTCAGATGTGTATAAGAGACAG ATCGTAAA GGAGCTGAACGT; *rpoS*_R2 GTCTCGTGGGCTCG GAGATGTGTATAAGAGACAGTTACTCGCGG AACAGCGCTTCG.

The PCR consisted of 3 min at 94°C, followed by 2 cycles of 45 s at 94°C, 60 s at 50°C, 60 s at 72°C and a final extension

of 10 min at 72°C. The amplification product was purified using magnetic beads Agencourt AMPure XP (Beckman Coulter) as specified by the manufacturer. Next, the DNA samples were indexed for the generation of clusters with the help of Nextera XT Index (Illumina) kit as recommended by the manufacturer. The DNA libraries were quantified by real-time PCR in the thermocycler QuantStudio 3 Real Time (Applied Biosystems) and using the Library Quantification Kit KAPA-KK4824 (Illumina/Universal - Roche). An equimolar pool of DNA molecules at 2 nM each was produced. Those were submitted to sequencing in an Illumina MiSeq system using the MiSeq Reagent Kit V2 Nano 500 cycles—reads of 2 × 250 bp.

The bioinformatics analysis was performed using the CLC Genomic Workbench version 7.0 platform. The sequences obtained were filtered for quality and aligned against the *rpoS* coding region of *E. coli* strain MG1655 (NC_000913.3 (2866559..2867551, complement)). The final output was a table containing all SNPs and their characteristics (Table S4).

Statistical Analysis

We calculated the standard error of the mean by using the formula $S.E.M. = \frac{S.D.}{\sqrt{n}}$, where S.D. is the standard deviation and n is the number of samples [31]. Data were evaluated for statistical significance using a two-tailed heteroscedastic Student's t-test or by ANOVA followed by Tukey's post hoc analysis using the JASP software [32].

Results

RpoS Status in *E. Coli* Isolates from a Polluted Stream

We collected water samples monthly over an entire year from the Pirajuçara stream, a heavily polluted waterway that crosses São Paulo's metropolitan area. This polluted secondary environment contains *E. coli* from multiple sources, mostly from human excrement. Physical and chemical characteristics of each water sample (temperature, pH, oxygen content (DOB) and turbidity) are depicted in Table S1. The ranges of pH, temperature, DOB and turbidity were 7.03–8.44, 18–23°C, 0.4–5.0 mg/L and 0.07–0.22, respectively. The June sample (n° 9) was exceptional, in that both the oxygen concentration and turbidity were considerably higher than in other samples. The average concentration of culturable Gram-negative bacteria on MacConkey plates was about 2×10^5 bacteria/ml of stream water. Colonies on MacConkey plates were re-isolated, frozen without further culturing and submitted to a series of biochemical tests to identify *E. coli* bacteria in the sample (described in the Methods section). At the end of 12 months, 328 *E. coli* isolates were identified (about 27 isolates/month).

To establish the phylogenetic diversity of *E. coli* in our environmental collection and also to detect changes in the distribution of the different phylogroups across the monthly samples, the isolates were assigned to their respective subgroups by the method of Clermont et al. [24]. Table S2 shows the character of each isolate, and Table 1 summarises the proportion of isolates from each phylogroup. Most isolates belonged to phylogroups A, B1 and C (66% of the total),

Table 1 Phylogroups of the *E. coli* strains isolated from the Pirajuçara stream

Month	A	B1	B2	C	D	E	F	U*	Number of isolates per month
1	2	5	3	8	1	2	3	1	25
2	7	2	2	10	2	1	5	1	30
3	5	13	3	4	2	0	1	3	31
4	2	2	4	6	0	0	1	1	16
5	8	5	0	8	3	1	1	2	28
6	6	5	1	11	2	0	1	0	26
7	5	9	5	6	7	0	1	0	33
8	7	3	2	7	3	3	2	0	27
9	10	5	4	3	2	1	1	0	26
10	3	11	2	4	3	0	0	3	26
11	8	5	6	5	4	0	0	2	30
12	12	0	4	6	6	0	1	1	30
% of each phylogroup	22.80%	19.80%	10.90%	23.70%	10.60%	2.40%	5.20%	4.60%	328

*Unknown group

typical of human *E. coli* commensal strains [33]. Surprisingly, strains belonging to phylogroup C were the most common (23.7% of the total), in contrast with other studies that showed that the predominant groups in both animals and humans are A, B1, B2 and D [34]. Thirty per cent of the isolates belonged to phylogroups B2, D, E and F, and the remaining 4% could not be classified despite being identified as *E. coli* in the diagnostic tests. Interestingly, the proportion of the various phylogroups were randomly distributed across the monthly samples (Table 1), and there was no obvious correlation between specific phylogroups and the water environmental parameters depicted in Table S1 (temperature, pH, oxygen and organic matter concentrations). Every monthly sample contained a rich cross section of *E. coli* types. This multiplicity of *E. coli* strains containing all known phylotypes of *E. coli* allowed a species-wide analysis of the survival-multiplication trade-off.

Quantitative Analysis of RpoS

A modified high-throughput ELISA protocol that provides a reliable estimation of RpoS relative to RpoD concentration was devised to obtain a quantitative assessment of RpoS content in the 328 isolates. The RpoS/RpoD ratio is indicative of the relative level of RpoS in each strain because RpoD, the main vegetative sigma factor in *E. coli*, displays a near-constant cellular concentration [7]. Bacterial total protein was extracted as described in detail in the "Methods" section.

We quantified the RpoS/RpoD ratio in all 328 isolated strains from the Pirajuçara stream. Each strain was assayed at least twice (biological replicates). To enable the comparative analysis of all isolates tested across many different plates, each ELISA plate also contained protein samples of the control strains MG1655 (*rpoS*+), MG1655 Δ*rpoS*::Kan and MG1655 Δ*rssB*::Kan, whose RpoS/RpoD ratio averages are shown in Fig. S1 and are in agreement with previous studies [35–37]. The mean RpoS/RpoD ratio (±S.E.M.) of each of the 328 isolates is presented in Table S2. While the mean and median RpoS/RpoD ratios of the entire collection were, respectively, 0.92 and 0.86, close to the standardised value (RpoS/RpoD ratio = 1), of the K-12 strain MG1655, the range of RpoS/RpoD ratios was extremely dispersed, ranging from 0 to almost 2.5 with a continuum of ratios between these extremes (Fig. 1a).

Fig. 1 **a** RpoS levels in 328 *E. coli* isolates. Each isolate was assayed for RpoS and RpoD by ELISA using specific monoclonal antibodies. The RpoS/RpoD ratio of each sample was calculated and divided by that of MG1655 in the corresponding 96-well plate. Each bar represents the mean (± S.E.M.) of at least two independent cultures. **b** Distribution chart of RpoS levels of isolates containing either no mutations or synonymous mutations (detailed in Table 2). Each triangle corresponds to one isolate RpoS level. **c** Distribution chart of AP levels in the collection. Each dot corresponds to the AP activity of one isolate. **d** Distribution chart showing the survival level of isolates challenged with dehydration stress. **e** Distribution chart showing the survival level of isolates challenged with acid stress. Black stars correspond to the isolates carrying non-synonymous mutations. The blue star in panels **c**–**e** corresponds to strain MG1655
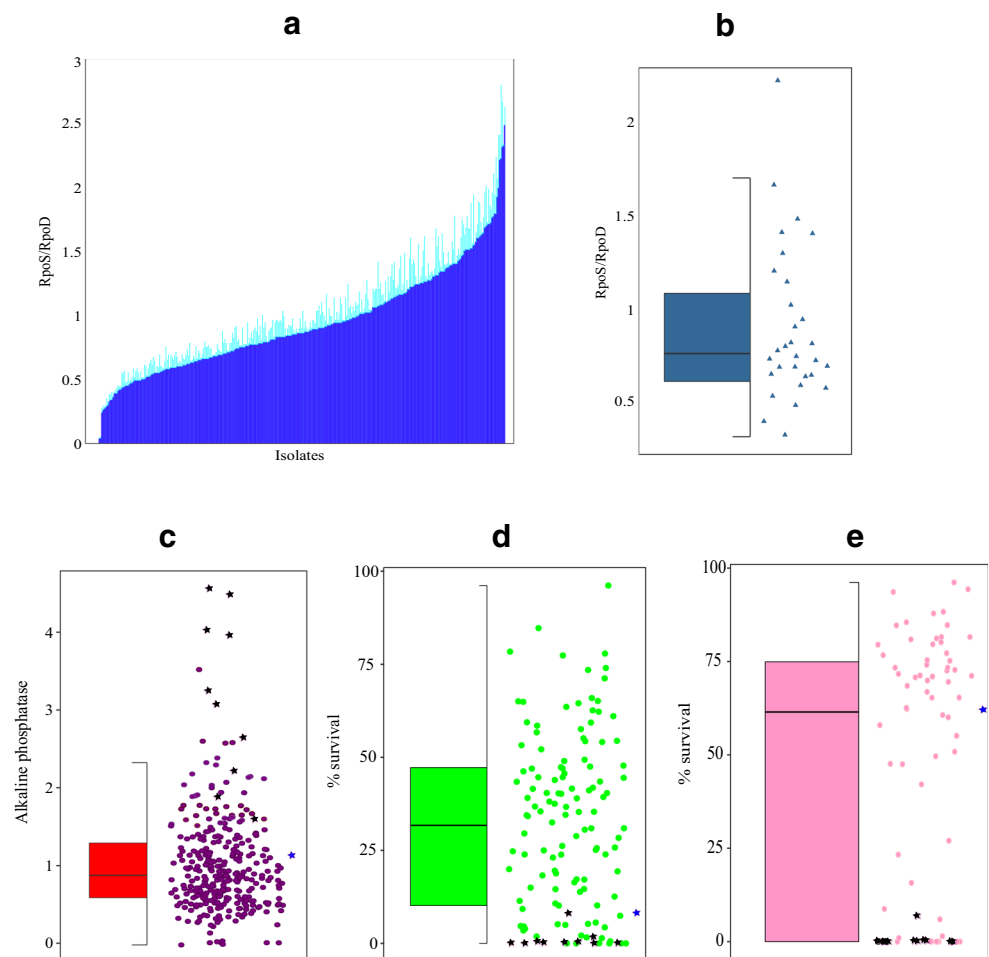
As shown in Fig. 2, the high diversity of RpoS/RpoD ratios in the collection was neither a monthly sample-to-sample variation nor was it a function of the sampled *E. coli* population in terms of the phylotypes present. In fact, a remarkably wide distribution of RpoS levels was observed in every monthly sample (Fig. 2a). The only sample that displayed a somewhat different distribution of RpoS, with slightly higher levels, was month 9 (statistically different from months 1, 3, 5, 7, 10 and 11, with $p<0.025$). Interestingly, month 9 corresponds to June which was the sample with the lowest pH (7.03) and the highest oxygen level (5.0 mg/L) and turbidity ($OD_{600} = 1.15$). Possibly this was the most stressed environment

resulting in the selection of high-RpoS strains, but it still did not purge the diversity of co-cohabiting *E. coli* strains.

The diversity of RpoS/RpoD ratios was not restricted to a particular phylotype of *E. coli*, as shown in Fig. 2b. Unsurprisingly, the range of variation was the least (less than 3-fold) in group E, as this group harbours the lowest number of isolates compared to all other phylotypes. There was a small variation in the mean RpoS/RpoD ratio between groups, with phylogroup F displaying the highest mean RpoS/RpoD ratio (1.11) and the isolates belonging to the Unknown group showing the lowest mean (0.79). A statistical analysis (ANOVA followed by post hoc Tukey) showed significant differences

Fig. 2 **a** Longitudinal distribution of RpoS levels in isolates collected across 1 year. Each plot shows the monthly distribution of RpoS levels; black lines inside the boxes represent the medians. Coloured circles represent the RpoS level of each isolate. Black circles represent strains with non-synonymous mutations in *rpoS* (see Table 2). * denotes statistically significant distribution of RpoS levels compared to month 9 ($p<0.025$). **b** Distribution of RpoS levels across phylogroups. Individual RpoS levels were ordered according to the isolate phylogroup: A, B1, B2, C, D, E, F or U (unknown). Each coloured square represents an isolate and the black line inside the boxes represents the median
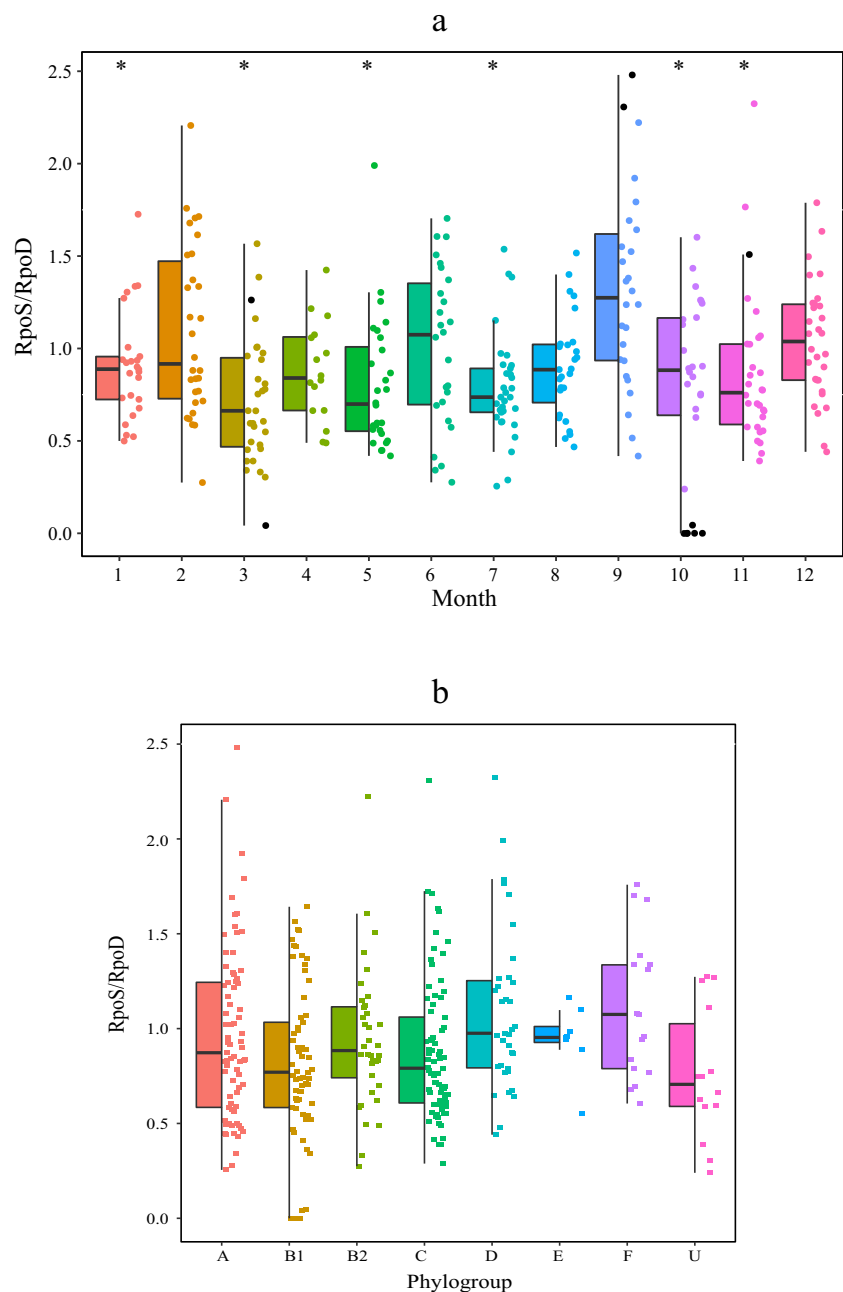
only between phylogroups B1 and D ($p = 0.022$). All 6 RpoS-null strains belonged to phylogroup B1, but 5 out of 6 RpoS-null isolates are likely clonal (see next section).

rpoS Sequence of Selected Isolates

To find out the genetic basis of the various RpoS levels, we sequenced the *rpoS* gene in forty isolates (Table 2). The isolates chosen for sequencing represent a wide range of RpoS/RpoD ratios as listed in Table S2. Also included were the six isolates in which the RpoS protein was not detected, as well as those with the highest RpoS/RpoD ratios. Sequencing of the RpoS-negative isolates J19, J21, J22, J23 and J24 revealed that they carry the same mutation: a 59 bp insertion at position 158–159, resulting in the emergence of a premature stop codon. Interestingly, this exact insertion has already been reported in strains RH90 (GenBank: Z14967.1) and UM122 (GenBank: Z14968.1), both K-12 derivatives. These five *rpoS*⁻ strains were isolated from the same water sample and are likely to share a clonal origin. The bacterial samples were not cultivated before plating so it is possible that the water sample contained faecal contamination biased from a single source. The other RpoS protein-negative strain, C24, carried a one-base deletion at position 191, resulting in a frameshift. In contrast, isolate C20 displayed a higher than average RpoS level (1.26) but revealed an IS4 insertion at position 775, which introduced a premature stop codon and resulted in a truncated protein. A Western blot analysis confirmed the presence of a faster-migrating RpoS band in this strain (Fig. S2). Interestingly, strains I19 (RpoS level = 2.30), I25 (RpoS level = 2.48) and K22 (RpoS level = 1.51) carry the amino acid substitutions G126W, G309D and I125Q, respectively. These non-conservative substitutions might have reduced the proteolytic turnover of the protein [38], thus increasing its level; these amino acid substitutions also affected RpoS function as shown below (Table 4).

Almost all sequenced isolates carry a Q33E substitution relative to MG1655, as commonly found in *E. coli* natural isolates [39]. Also interesting is that several isolates in Table 2 harboured at least one synonymous *rpoS* change. Strain H19 carries 27 synonymous substitutions, which are also found in 20 other sequenced *E. coli* genomes in the database. Likewise, the C942T substitution, common to many Pirajuçara's isolates, is also found in many sequenced *E. coli* genomes. It is also worth noticing that the low non-synonymous/synonymous ratio indicates that *rpoS* is under strong negative/purifying selection, as expected for an important gene and as already observed by others [20].

It is particularly noteworthy that a broad range of RpoS/RpoD ratios was exhibited by isolates with wild-type *rpoS* gene sequences as well as in strains carrying only synonymous substitutions. As shown in Fig. 1b, a wild-type RpoS sequence can be randomly associated with any RpoS/RpoD ratio over the range found. The important conclusion from this finding is that natural isolates vary in RpoS status even without coding changes in *rpoS*.

Culture-Independent Analysis of *rpoS* in the Pirajuçara Stream

High-throughput gene analysis was also conducted to assess whether *E. coli rpoS* polymorphisms are evident in water samples without growth in the laboratory. Two independent water samples were collected from the Pirajuçara stream after completion of the monthly sampling and immediately processed to extract genomic DNA. The DNA samples were quantified and submitted to PCR amplification with *rpoS* specific primers, resulting in two amplicons, each with 500 bp. The generated libraries were purified, quantified and submitted to high-throughput sequencing analysis. The sequence data were filtered to encompass only *E. coli rpoS*. Table S4 shows the output of this analysis, and Table 3 summarises the results obtained. Using a cut-off of frequency > 0.5% and quality > 25, sample no.1 displayed 79 substitutions, of which 25 were non-synonymous. Nine small deletions and one small insertion that resulted in frameshifts were also observed. The other sample (n° 2) contained 122 single substitutions, 36 of which were non-synonymous. In addition, 9 small deletions and 1 small insertion were counted. As expected, the low ratio of non-synonymous/synonymous substitutions in both samples indicates that *rpoS* is under strong purifying selection. Of the 221 polymorphisms observed in both DNA samples (substitutions+indels), 61 were unique to one of the samples, while 80 were present in both samples.

In conclusion, the existence of *rpoS* polymorphisms and the presence of deleterious mutations in a secondary environment could be confirmed in the absence of lab culture. Nevertheless, the very nature of this experiment cannot provide a precise quantitative assessment of the proportion of *rpoS*-negative variants in the samples.

Heterogeneity in RpoS-Dependent Phenotypes

RpoS directly or indirectly modulates the expression of hundreds of genes [6, 40–44]. Growth and vegetative functions of genes transcribed by RpoD, including growth on alternative carbon sources [13, 21] or alkaline phosphatase synthesis [12], are negatively affected by elevated RpoS. On the stress response side, RpoS is the master regulator of the general stress response [45, 46] and is needed to elicit, amongst others, resistance to acid and dehydration stresses [29, 47]. RpoS also positively controls the accumulation of glycogen [28] and the synthesis of catalase [21, 48]. The variation in RpoS content demonstrated above was likely to be associated with a

Table 2 *rpoS* sequence of isolates

Strain	RpoS ratio	Synonymous mutations	Non-synonymous mutations	AA change
A06	0.7	G381A, C384T, G819A	C97G	Q33E
A09	0.5	C942T	C97G	Q33E
B05	0.6	T163C, T357G, T472C, T573C, C732T	C97G	Q33E
B06	2,2	C942T	C97G	Q33E
B08	1,5	C942T	C97G	Q33E
B10	1,3	C942T	C97G	Q33E
B27	0.8	C372T, G971A	C97G	Q33E
C08	0.6	C372T, T573C, C732T	C97G	Q33E
C18	0.3	—	C97G	Q33E
C20	1,3	—	C97G, IS4 ins 775-776	Q33E, frameshift
C24	0	—	C97G, 191A del	Q33E, frameshift
C25	0.4	C942T	C97G	Q33E
D07	1,4	—	C97G	Q33E
D09	0.7	C264T, T270T, T357G, T462C, T549C, G564A, T573C, G819A	C97G	Q33E
D15	0.5	T163C, T357G, T472C, T573C, C732T	C97G	Q33E
E02	0.6	C942T	C97G	Q33E
E07	0.9	—	C97G	Q33E
F26	0.7	G952A	C97G	Q33E
G27	1,4	T163C, T357G, T472C, T573C, G699A, C732T, G990A	C97G	Q33E
H19	1	T42C, A81G, G84A, A93T, C123T, A132G, G144A, A147G, A204G, T237C, C267T, T270A, C273T, C285T, G288A, T295C, G306C, A309G, A312G, T346C, C349T, T357G, T450C, T462C, T549C, G564A, T573C	C97G	Q33E
I09	1,7	—	C97G	Q33E
I19	2,3	C942T	C97G, G376T	Q33E, G126W
I25	2,4	C942T	G376T, G926A	G309D
I26	1,2	C942T	C97G	Q33E
J14	0.8	—	C97G	Q33E
J16	0.8	—	C97G	Q33E
J19	0	—	59 bp ins 158-159	Frameshift
J21	0	—	59 bp ins 158-159	Frameshift
J22	0	—	59 bp ins 158-159	Frameshift
J23	0	—	59 bp ins 158-159	Frameshift
J24	0	—	59 bp ins 158-159	Frameshift
J26	0.7	G951A	C97G	Q33E
K04	0.7	C942T	C97G	Q33E
K08	1,1	—	C97G	Q33E
K14	0.9	—	C97G	Q33E
K17	0.6	T163C, T462C, T573C, C732T	C97G	Q33E
K22	1,5	T163C, T357G, T462C, T573C, G699A, C732T, G990A	C97G, T374A	Q33E, L125Q
K26	0.6	—	C97G	Q33E
L06	0.7	C942T	C97G	Q33E
L09	0.8	T163C, T357G, T462C	C97G	Q33E

phenotypic diversity associated with the survival-growth balance across the isolates. We thus studied the characteristics of the Pirajuçara stream isolates to test the heterogeneity of stress responses controlled by RpoS.

The quantified characteristics of isolates in terms of alkaline phosphatase activity (negatively regulated by RpoS), as well as dehydration survival and acid survival (positively regulated by RpoS), are presented in Table S2 and Table S3,

Table 3 Metagenome analysis of *rpoS* sequences in the Pirajuçara stream

Mutations	Sample 1	Sample 2
Non-synonymous	25	36
Insertions	1	1
Deletions	9	9
Synonymous	54	86
Total polymorphisms	89	132

respectively. As is evident in Fig. 1c, d and e, the isolates displayed a very broad distribution of all RpoS-regulated properties (varying over 100-fold for each activity). The distinct position of the K-12 MG1655 reference strain in each assay is also shown as blue dots in the panels in Fig. 1c–e. The strong conclusion from this analysis is that *E. coli* natural isolates exhibit a remarkably wide range of stress resistance capabilities and impacts on RpoD-transcribed genes.

Further evidence for the heterogeneity of RpoS-related phenotypes is shown in Table S2. A range of responses was also found in assays of catalase activity and iodine staining (a measure of glycogen content) in the isolates tested. These assays were more qualitative (just a 3–4-step scale) so less able to detect a continuum of responses. Nevertheless, both catalase and glycogen contents were non-uniform across the 328 members of the collection.

So how do the strain variations in Fig. 1 and Table S2 correlate with RpoS/RpoD ratios? In *E. coli* K-12, all the assayed properties used here are trustworthy indicators of RpoS status in isogenic strains [12, 13, 48]. Several of the

phenotypes (iodine staining, catalase activity and growth on alternative carbon substrates) have also been used to differentiate between *rpoS*[+]- and *rpoS*[-]-deficient strains in natural populations [20, 21]. The correlation between various RpoS-dependent phenotypes and RpoS level has been analysed in a few clinical isolates from a patient [19] but not in multiple natural isolates of *E. coli* from a secondary habitat.

The expression of *phoA*, an RpoD-dependent gene that encodes alkaline phosphatase and is negatively regulated by RpoS [49, 50], was highest in the six RpoS-negative strains, which displayed an average AP activity of 3.96 E.U. (solid black symbols in Fig. 1c), considerably above the AP average of the collection as a whole (1.02 E.U.). Fourteen RpoS-positive isolates displayed high AP levels (>2.04 E.U., i.e. more than twice the collection AP mean). On average, these isolates presented relatively low RpoS/RpoD ratios (0.7 versus 0.92, which was the collection mean). A correlational analysis of RpoS and AP levels showed that these two traits were only moderately correlated (Pearson's correlation coefficient = −0.22), although the correlation is better when the strains carrying non-synonymous mutations are excluded (Pearson's correlation coefficient = −0.30). Presumably, polymorphisms other than in *rpoS* contribute to the diversity of AP expression levels in these natural isolates.

Of the isolates tested for both acid and dehydration stresses, 18 showed low survival rates under both stresses (Table S3). This group includes the six RpoS-negative strains, C24, J19, J21, J22, J23 and J24, and the four RpoS[+] strains, C20, I19, I25 and K22, that carry an IS4 insertion (isolate C20) or non-conservative amino acid substitutions (Table 4). Isolates C17, C18, C25 and J13 harbour

Table 4 Summary of RpoS-deficient or attenuated isolates in the collection

Isolate	RpoS[1]	Iodine[2]	Catalase[3]	AP[4]	Dehydration[5]	Acid[6]	Sequence[7]
C20	1.26	0	0.5	3.25	0.63	5.97	ins IS4 775-776
C24	0.04	0	0.5	1.94	0.04	0.00	ins A 91-92
I19	2.31	1	1	3.09	0.00	0.00	G126W
I25	2.48	0	0	1.60	2.11	0.00	G309D
J19	0.04	0	0	4.03	0.00	0.00	59 bp ins 158-159
J21	0.00	0	0	2.60	0.00	0.00	59 bp ins 158-159
J22	0.00	0	0	4.49	0.01	0.00	59 bp ins 158-159
J23	0.00	0	0	3.97	0.00	0.00	59 bp ins 158-159
J24	0.00	0	0	4.56	0.00	0.00	59 bp ins 158-159
K22	1.51	1	1	2.21	7.42	0.00	L125Q

[1] RpoS/RpoD ratio assessed by ELISA

[2] Iodine staining (0–3)

[3] Catalase level (0–1)

[4] AP specific activity

[5] % cell survival under dehydration stress

[6] % cell survival under acid stress

[7] WT, wild-type *rpoS* sequence

wild-type *rpoS* sequences but contain low RpoS levels (0.33, 0.30, 0.39 and 0.24, respectively) which might explain their stress sensitivity. The following isolates displayed normal RpoS levels (>0.57), and wild-type sequences, but showed strong sensitivity to both acid and dehydration stresses: B05, C08, D09, K08 and K17. These are true outliers, and, as for AP activity, polymorphisms other than in RpoS might explain the complexity of stress resistance properties in natural isolates. The Pearson's correlation coefficient between RpoS level and bacterial survival under dehydration and acid stress was 0.35 and 0.47, respectively, when the isolates carrying the non-synonymous mutation were excluded. The correlation between bacterial survival to acid stress and survival to dehydration stress was 0.55, consistent with the idea that both stress responses involve a shared mechanism.

Catalase activity in Table S2 was found to show the closest correlation with RpoS status in the 328 isolates. In addition to the six RpoS-negative strains, 12 isolates that responded positively in the ELISA (RpoS level > 0.2) displayed null (catalase score = 0) or low catalase activity (catalase score = 0.5), i.e. 316 isolates (96% of all strains) displayed the phenotype expected from the measured RpoS level. In assays of iodine staining for glycogen, all six RpoS-negative strains and twenty-seven *rpoS*⁺ isolates displayed low colour intensity (iodine = 0 in Table S2). These 27 strains correspond to 8% of the total 328 strong population, denoting that the iodine staining assay showed the expected phenotype for 92% of the population. Two of these "outliers" were shown to carry non-synonymous mutations in RpoS and low survival under stress, which underlines the value of the catalase and iodine staining assays in testing RpoS status.

In summary, all the phenotypes known to be regulated by RpoS show wide distributions, so there is no "typical" level of stress resistance in any phylotype amongst natural isolates of *E. coli*. These findings extend to environmental isolates previous data observed in a small number of human patient isolates that exhibited distinct RpoS levels and resistance properties [19]. Multiple *E. coli* strains with different settings of the resistance/growth trade-off can coexist, as predicted by mathematical models involving different trade-off settings [22]. It can also be concluded that all five phenotypes we tested are reliable but not perfect indicators of RpoS status in diverse natural *E. coli* isolates. In each test, variables other than RpoS level can influence outcomes. The one strain out of 328 which was a true outlier was B05, which had only synonymous mutations in *rpoS* but exhibited all RpoS-negative traits despite being positive for RpoS protein production (RpoS level = 0.59). This strain might carry mutations in other regulatory genes, for example, in *crl*, *iraP*, *cspC* [51] or *spoT* [13, 35, 50].

Discussion

The existence of biodiversity in nature is at least partly explained by trade-offs that organisms face in dealing with the constraints of their environment [52]. Several types of trade-off can lead to diversity [1], including the trade-off between reproduction and survival [2, 22]. Here we reveal the remarkable continuum of traits in a single species of bacteria that are subject to a trade-off with a well-understood molecular mechanism. The resource allocation trade-off based on competition between sigma factors [3, 4, 8, 53] can affect fundamental properties of *E. coli*. Stress resistance and the ability to express vegetative genes exhibited a 100-fold range of variation of such properties in distinct members of the species cohabiting in a complex secondary habitat.

Previous studies have noted variation in RpoS-dependent phenotypes but were limited to a small number of isolates from humans [19] or, more seriously, in using bacterial collections that were compromised by storage and transfer in and between laboratories [35, 53]. In the present study, we focussed on the heterogeneity of RpoS protein levels and RpoS-controlled functions in fresh isolates over many months collected from a secondary habitat of *E. coli*. The isolates analysed in this study were from a river, but likely to be derived from multiple primary sources. The sampled Pirajuçara stream runs across a heavily populated urban area through the western metropolitan area of São Paulo, where many households are not connected to the municipal sewage system. Ours is the first longitudinal study of this kind on *E. coli* strains likely to be derived from a rich diversity of primary hosts. Our sampling covered the breadth of the species, since we found all the known phylotypes representing *E. coli* sub-species diversity. The primary experimental finding in this study is that the content of an important sigma factor, RpoS, has no characteristic value in the species *E. coli* and exhibits a continuum of RpoS/RpoD ratios across 328 members of the species. The observed RpoS/RpoD ratio varied from zero to well above that found in *E. coli* K-12. As discussed below, the different levels of RpoS were not entirely a consequence of polymorphisms in the *rpoS* gene but also occurred in isolates that had wild-type *rpoS* sequences or synonymous mutations.

We found that the only observed statistically significant environmental effect on RpoS/RpoD ratios was in a single month that had different physicochemical parameters to the other 11 river samples; this turbid sample contained isolates with a higher than average RpoS content. It is tempting to speculate that the challenging environmental conditions in this sample selected for high-RpoS strains. Nevertheless, there was no obvious sweep eliminating all low-RpoS strains in this turbid sample, and a diverse range of RpoS levels was still observed, as it was in all other

months. Furthermore, the variation in RpoS content was not limited to any particular phylotype of *E. coli*. So the main conclusion of this longitudinal study is that a continuum of RpoS/RpoD ratios is consistently present across the species.

As would be expected from the role of RpoS in general stress resistance, phenotypes positively dependent on RpoS regulation were also highly variable across the collection. Acid and dehydration stress were as variable between strains as was the level of RpoS protein. RpoS variation was not the sole explanation for this variation. The correlation between RpoS level and stress resistance was strong, but not perfect, such that some strains with significant RpoS/RpoD ratios were relatively sensitive to one or both stresses. In addition to RpoS, strain variation in other regulatory factors is required to explain the differential response to environmental stresses across the 328 isolates.

Negative regulation by RpoS was also a variable across the strain collection. The expression of *phoA*, an RpoD-transcribed gene, was negatively regulated by RpoS, but the extent of the negative effect was as variable as the RpoS level in the species. As with stress resistance, the correlation between RpoS/RpoD ratios and AP was there but not perfect; again, the heterogeneity of other inputs like PhoB-PhoR, the two-component system that regulates *phoA* transcription, and perhaps other specific inputs that might affect AP expression needs to be postulated.

Previously, mutational changes in the survival-growth trade-off were demonstrated in the laboratory so a crucial question is whether RpoS levels and *rpoS* gene integrity are subject to frequent mutational shifts in natural environments. A variability in *rpoS* sequence and functionality both in *E. coli* K-12 laboratory strains [39, 50, 52] and in natural populations [20, 21, 35, 49, 54–58] has been reported. The detected proportion of RpoS-deficient strains in the natural isolate collections showed considerable variation, however, ranging from 0.3 to more than 30%. Table 5 summarises previous findings. In some of these studies, the strain collections have been stored under conditions that favour the selection and fixation of *rpoS*-deficient alleles. These studies usually reported high proportions of RpoS-deficient strains in the collections, ranging from 12 to 36%. Overall, the recorded frequencies of RpoS-deficient strains in *E. coli* strains range from 0.3 to more than 30%. It is quite clear now that the high proportions of RpoS mutants were mostly observed in collections in which the bacteria were stored in stabs for long periods as in the ECOR collection, which has been stored and transferred repeatedly [35, 53]. On the other hand, fresh natural isolates that had not been subject to manipulation displayed low frequencies of *rpoS* mutations [20]. A conclusion from this study was that *rpoS* mutations follow a source-sink dynamics, in which the wild-type allele is favoured in the source environment (natural environments) and the RpoS-inactivating alleles are favoured in the sink environments (laboratory conditions with selection for growth) [20].

Our results showed that a 1.8% prevalence of *rpoS* mutants is compatible with the ones observed by Bleibtreu et al. [20] (1.2%) and Snyder et al. [56]. We believe that this is the most accurate assessment of *rpoS* status in natural isolates for the following reasons: [1] we took measures to avoid the selection of *rpoS* mutants by growing the bacteria as isolated colonies in solid medium plates, where the appearance of *rpoS* mutants is very rare [11] and freezed the colonies in 15% glycerol directly without further growth in liquid media; [2] samples were collected monthly throughout an entire year which might have reduced any bias caused by changes in temperature or other physico-chemical conditions; [3] all 328 isolates were assayed for RpoS levels and normalised against the house-keeping sigma factor RpoD; [4] all isolates were tested for three phenotypic features, iodine staining, catalase and AP activity, that in combination had the power to distinguish a RpoS-negative strain with a 99.7% accuracy (1 outlier—B05—out of 328 strains) [5]; dehydration and acid stress were used to confirm the RpoS status of the isolates, especially of those carrying non-synonymous mutations that resulted in total or partial loss of RpoS function. As a possible explanation of our data in which RpoS content varied without *rpoS* mutations, a recent comparative genomics study has identified several genes associated with *rpoS* regulation that are frequently inactivated in *Salmonella* genomes. Three commonly inactivated genes (*iraP*, *crl* and *cspC*) were genes that modify RpoS expression or function [49]. These could well be the kind of mutations outside *rpoS* that are present in *rpoS*+ strains but add to the extensive heterogeneity in RpoS levels, stress resistance and other RpoS-dependent functions.

Finally, it is worth pointing out that a 1.8% frequency of *rpoS* mutants is many orders of magnitude above the rate of spontaneous null mutations in *E. coli* genes which is about 10^{-7}–10^{-8}/gene [59–61], suggesting an active selective pressure favouring *rpoS* mutations under natural environmental conditions. This relatively high frequency is even more impressive given the conserved nature of RpoS as deduced from the high dN/dS ratio of substitutions observed in the sequencing and high-throughput gene analyses and elsewhere [20]. In addition, several isolates expressed low RpoS/RpoD ratios (between 0.2 and 0.3), which resulted in poor ability to respond to environmental challenges. Altogether, these data indicate that the selective pressures against *rpoS* that were frequently observed in laboratory conditions [10, 11, 13, 14] are likely to play a role in natural settings as well.

In conclusion, a broad distribution of RpoS protein content and activity is characteristic of *E. coli* isolates. The continuum of strain-specific protein level variation implies a broad range of adaptive strategies in the species. This variation can be explained by the frequency of mutations in *rpoS* and also by polymorphisms in genes (other than *rpoS* itself) affecting

Table 5 Summary of studies that analysed RpoS status in natural populations

Source	No. of strains	Storage	%RpoS-negative strains based on			Assessment of RpoS levels	References
			Phenotypic assays	Immunoblots	*rpoS* sequencing		
STEC[1]	82	Unknown[7]	24%[8]	ND	ND	No	Bhagwat et al. (2005) [54]
SLTEC[2]	58	Unknown[7]	17.2%[8]	ND	ND	No	Waterman and Small (1996) [57]
Salmonella typhi	41	Unknown[7]	36.6%[9]	36.6%	ND	No	Robbe-Saule et al. (2003) [55]
Salmonella typhimurium	75	Unknown[7]	0%[9]	0%	ND	No	Robbe-Saule et al. (2003) [55]
E. coli from human hosts	115	−70°C	55.7%[10]	ND	ND	No	White et al. (2011) [58]
E. coli from animal hosts	169	−70°C	10.7%[10]	ND	ND	No	White et al. (2011) [58]
E. coli from urban beaches	2040	Unspecified	0.3%[11]	ND	0.3%[12]	No	Chiang et al. (2011) [21]
ECOR[3]	31	Stabs & −70°C	ND	29%	29%	Yes[15]	Ferenci et al. (2011) [35]
E. coli (different sources)	96	Stabs & −80°C	ND	2.1%	2.1%[13]	No	Snyder et al. (2012) [56]
IAI[4]	82	Stabs & −80°C	36%[10]	ND	36%	No	Bleibtreu et al. (2014) [20]
NILS[5]	82	−80°C	1.2%[10]	ND	1.2%	No	Bleibtreu et al. (2014) [20]
Salmonella enterica	100,000[6]	Various	ND	ND	0.27%[14]	No	Cherry (2020) [49]

[1] Shiga toxin-producing *E. coli*

[2] Shiga-like toxin-producing *E. coli*

[3] ECOR collection of *E. coli* strains

[4] Collection of natural isolates pathogenic and commensal

[5] Collection of natural isolates with low subcultures

[6] In silico analysis

[7] Given the old age and source heterogeneity of the strains, probably stored in stabs

[8] Acid sensitivity assays

[9] Catalase activity assays

[10] Catalase activity and iodine staining assays

[11] Growth on succinate and catalase activity assays

[12] Sequencing of 45 isolates confirmed the presence of *rpoS* null mutations in 6 of them

[13] Only the two RpoS-null strains (based on immunoblots) were sequenced

[14] Only premature stop codons were reported

[15] Western blots gave semi-quantitative estimates of RpoS levels

RpoS protein levels. All these may well be under survival-growth selection pressures and also impact the stress resistance properties of *E. coli*, adding to the inherent heterogeneity of *E. coli*. Similarly, a continuum of physiological properties was also observed in a structural constraint trade-off in *E. coli*, characterised by a wide continuum of endogenous antibiotic susceptibilities [62, 63]. Based on our results, it can be predicted that most of the many trade-offs identified in the bacteria [1] will contribute to phenotypic diversity in many important bacterial traits. The availability of a range of stress resistances and growth fitness levels in all samples reflects the broad diversity within *E. coli* strains, each one adapted to a specific range of hosts and environments.

Author's Contribution EYV carried out most of the lab work and data analysis. JPB isolated the strains and carried out the initial characterizations. TF critically revised and drafted the manuscript, and BS conceived

of the study, designed the study, coordinated the study and drafted the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Availability of Data and Material All data is contained in the manuscript and supplement.

Code Availability Not applicable

Declarations

Competing Interests The authors declare no competing interests.

References

1. Ferenci T (2016) Trade-off mechanisms shaping the diversity of bacteria. Trends Microbiol 24:209–223
2. Stearns S (1992) The evolution of life histories. Oxford University Press, Oxford, New York
3. Ferenci T (2005) Maintaining a healthy SPANC balance through regulatory and mutational adaptation. Mol Microbiol 57:1–8
4. Hengge R (2011) Stationary-phase gene regulation in *Escherichia coli* §. Ecosal Plus 4(2). https://doi.org/10.1128/ecosalplus.5.6.3
5. Landini P, Egli T, Wolf J, Lacour S (2014) sigmaS, a major player in the response to environmental stresses in *Escherichia coli*: role, regulation and mechanisms of promoter recognition. Environ Microbiol Rep 6:1–13
6. Schellhorn HE (2014) Elucidating the function of the RpoS regulon. Future Microbiol 9:497–507
7. Ishihama A (2018) Building a complete image of genome regulation in the model organism *Escherichia coli*. J Gen Appl Microbiol 63:311–324
8. Nyström T (2004) Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? Mol Microbiol 54:855–862
9. Maharjan R, Seeto S, Notley-McRobb L, Ferenci T (2006) Clonal adaptive radiation in a constant environment. Sci 313:514–517
10. Notley-McRobb L, King T, Ferenci T (2002) *rpoS* mutations and loss of general stress resistance in *Escherichia coli* populations as a consequence of conflict between competing stress responses. J Bacteriol 184:806–811
11. Spira B, de Almeida TR, Maharjan RP, Ferenci T (2011) The uncertain consequences of transferring bacterial strains between laboratories - *rpoS* instability as an example. BMC Microbiol 11:248
12. Spira B, Ferenci T (2008) Alkaline phosphatase as a reporter of sigma(S) levels and *rpoS* polymorphisms in different *E. coli* strains. Arch Microbiol 189:43–47
13. Wang L, Spira B, Zhou Z, Feng L, Maharjan RP, Li X, Li F, McKenzie C, Reeves PR, Ferenci T (2010) Divergence involving global regulatory gene mutations in an *Escherichia coli* population evolving under phosphate limitation. Genome Biol Evol 2:478–487
14. Zambrano MM, Siegele DA, Almirón M, Tormo A, Kolter R (1993) Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. Sci 259:1757–1760
15. Maharjan RP, Ferenci T, Reeves PR, Li Y, Liu B, Wang L (2012) The multiplicity of divergence mechanisms in a single evolving population. Genome Biol 13:R41
16. Cavaliere P, Norel F (2016) Recent advances in the characterization of Crl, the unconventional activator of the stress sigma factor sigmaS/RpoS. Biomol Concepts 7:197–204
17. Lalaouna D, Massé E (2016) The spectrum of activity of the small RNA DsrA: not so narrow after all. Curr Genet 62:261–264
18. Mika F, Hengge R (2014) Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. RNA Biol 11: 494–507
19. Levert M, Zamfir O, Clermont O, Bouvet O, Lespinats S, Hipeaux MC, Branger C, Picard B, Saint-Ruf C, Norel F, Balliau T, Zivy M, Nagard HL, Cruvellier S, Chane-Woon-Ming B, Nilsson S, Gudelj I, Phan K, Ferenci T, Tenaillon O, Denamur E (2010) Molecular and evolutionary bases of within-patient genotypic and phenotypic diversity in *Escherichia coli* extraintestinal infections. PLoS Pathog 6:e1001125
20. Bleibtreu A, Clermont O, Darlu P, Glodt J, Branger C, Picard B, Denamur E (2014) The *rpoS* gene is predominantly inactivated during laboratory storage and undergoes source-sink evolution in *Escherichia coli* species. J Bacteriol 196:4276–4284
21. Chiang SM, Dong T, Edge TA, Schellhorn HE (2011) Phenotypic diversity caused by differential RpoS activity among environmental *Escherichia coli* isolates. Appl Environ Microbiol 77:7915–7923
22. Gudelj I, Weitz JS, Ferenci T, Claire Horner-Devine M, Marx CJ, Meyer JR, Forde SE (2010) An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. Ecol Lett 13:1073–1084
23. Hongve D, Akesson G (1998) Comparison of nephelometric turbidity measurements using wavelengths 400–600 and 860nm. Water Res 32:3143–3145
24. Clermont O, Christenson JK, Denamur E, Gordon DM (2013) The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep 5:58–65
25. Miller JH (1992) A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. da Costa Vasconcelos FN, Padilla G, Spira B (2016) *Chromobacterium violaceum* adaptation to low-phosphate conditions. Arch Microbiol 198:269–277
27. Arif A, Dhaliwal M, Panasar B, Yang S (2014) Heat induced RpoS expression may inhibit RpoD-dependent expression of β-galactosidase in *Escherichia coli* B23. J Exp Microbiol Immunol JEMI Vol 18:18–23
28. Hengge-Aronis R, Fischer D (1992) Identification and molecular analysis of *glgS*, a novel growth-phase-regulated and *rpoS*-dependent gene involved in glycogen synthesis in *Escherichia coli*. Mol Microbiol 6:1877–1886
29. Chen AI, Goulian M (2018) A network of regulators promotes dehydration tolerance in *Escherichia coli*. Environ Microbiol 20: 1283–1295
30. Tabatabaei M, Zakaria MR, Rahim RA, Abdullah N, Wright ADG, Shirai Y, Shamsara M, Sakai K, Hassan MA (2010) Comparative study of methods for extraction and purification of environmental DNA from high-strength wastewater sludge. Afr J Biotechnol 9: 4926–4937
31. Cumming G, Fidler F, Vaux DL (2007) Error bars in experimental biology. J Cell Biol 177:7–11
32. JASP Team. 2020. JASP (Version 0.13.1)[Computer software].
33. Escobar-Páramo P, Grenet K, Le Menac'h A, Rode L, Salgado E, Amorin C, Gouriou S, Picard B, Rahimy MC, Andremont A, Denamur E, Ruimy R (2004) Large-scale population structure of human commensal *Escherichia coli* isolates. Appl Environ Microbiol 70:5698–5700

34. Tenaillon O, Skurnik D, Picard B, Denamur E (2010) The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol 8:207–217

35. Ferenci T, Galbiati HF, Betteridge T, Phan K, Spira B (2011) The constancy of global regulation across a species: the concentrations of ppGpp and RpoS are strain-specific in *Escherichia coli*. BMC Microbiol 11:62

36. Muffler A, Fischer D, Altuvia S, Storz G, Hengge-Aronis R (1996) The response regulator RssB controls stability of the sigma (S) subunit of RNA polymerase in *Escherichia coli*. EMBO J 15:1333–1339

37. Peterson CN, Ruiz N, Silhavy TJ (2004) RpoS proteolysis is regulated by a mechanism that does not require the SprE (RssB) response regulator phosphorylation site. J Bacteriol 186:7403–7410

38. Becker G, Klauck E, Hengge-Aronis R (1999) Regulation of RpoS proteolysis in *Escherichia coli*: the response regulator RssB is a recognition factor that interacts with the turnover element in RpoS. Proc Natl Acad Sci U S A 96:6439–6444

39. Subbarayan P, Sarkar M (2004) A comparative study of variation in codon 33 of the *rpoS* gene in *Escherichia coli* K12 stocks: implications for the synthesis of σ\textsuperscriptS. Mol Gen Genomics 270:533–538

40. Dong T, Schellhorn HE (2009) Global effect of RpoS on gene expression in pathogenic *Escherichia coli* O157:H7 strain EDL933. BMC Genomics 10:349

41. Lacour S, Landini P (2004) SigmaS-dependent gene expression at the onset of stationary phase in *Escherichia coli*: function of sigmaS-dependent genes and identification of their promoter sequences. J Bacteriol 186:7186–7195

42. Patten CL, Kirchhof MG, Schertzberg MR, Morton RA, Schellhorn HE (2004) Microarray analysis of RpoS-mediated gene expression in *Escherichia coli* K-12. Mol Genet Genomics MGG 272:580–591

43. Vijayakumar SR, Kirchhof MG, Patten CL, Schellhorn HE (2004) RpoS-regulated genes of *Escherichia coli* identified by random *lacZ* fusion mutagenesis. J Bacteriol 186:8499–8507

44. Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R (2005) Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol 187:1591–1603

45. Battesti A, Majdalani N, Gottesman S (2011) The RpoS-mediated general stress response in *Escherichia coli*. Annu Rev Microbiol 65:189–213

46. Gottesman S (2019) Trouble is coming: Signaling pathways that regulate general stress responses in bacteria. J Biol Chem 294:11685–11700

47. Xu J, Li T, Gao Y, Deng J, Gu J (2019) MgrB affects the acid stress response of *Escherichia coli* by modulating the expression of iraM. FEMS Microbiol Lett 366:123

48. Visick JE, Clarke S (1997) RpoS- and OxyR-independent induction of HPI catalase at stationary phase in *Escherichia coli* and identification of *rpoS* mutations in common laboratory strains. J Bacteriol 179:4158–4163

49. Cherry JL (2020) Selection-driven gene inactivation in *Salmonella*. Genome Biol Evol 12:18–34

50. Spira B, Hu X, Ferenci T (2008) Strain variation in ppGpp concentration and RpoS levels in laboratory strains of *Escherichia coli* K-12. Microbiol 154:2887–2895

51. Tilman D (2000) Causes, consequences and ethics of biodiversity. Nature 405:208–211

52. Atlung T, Nielsen HV, Hansen FG (2002) Characterisation of the allelic variation in the *rpoS* gene in thirteen K12 and six other non-pathogenic *Escherichia coli* strains. Mol Gen Genomics 266:873–881

53. Jishage M, Ishihama A (1997) Variation in RNA polymerase sigma subunit composition within different stocks of *Escherichia coli* W3110. J Bacteriol 179:959–963

54. Bhagwat AA, Chan L, Han R, Tan J, Kothary M, Jean-Gilles J, Tall BD (2005) Characterization of enterohemorrhagic *Escherichia coli* strains based on acid resistance phenotypes. Infect Immun 73:4993–5003

55. Robbe-Saule V, Algorta G, Rouilhac I, Norel F (2003) Characterization of the RpoS status of clinical isolates of *Salmonella enterica*. Appl Environ Microbiol 69:4352–4358

56. Snyder E, Gordon DM, Stoebel DM (2012) *Escherichia coli* lacking RpoS are rare in natural populations of non-pathogens. G3 Bethesda 2:1341–1344

57. Waterman SR, Small PL (1996) Characterization of the acid resistance phenotype and *rpoS* alleles of shiga-like toxin-producing *Escherichia coli*. Infect Immun 64:2808–2811

58. White A, Sibley K, Sibley C, Wasmuth J, Schaefer R, Surette M, Edge T, Neumann N (2011) Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity. Appl Environ Microbiol 77:7620–7632

59. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148:1667–1686

60. Fehér T, Cseh B, Umenhoffer K, Karcagi I, Pósfai G (2006) Characterization of *cycA* mutants of *Escherichia coli*. An assay for measuring in vivo mutation rates. Mutat Res 595:184–190

61. Maharjan R, Ferenci T (2015) Mutational signatures indicative of environmental stress in bacteria. Mol Biol Evol 32:380–391

62. Phan K, Ferenci T (2013) A design-constraint trade-off underpins the diversity in ecologically important traits in species *Escherichia coli*. ISME J 7:2034–2043

63. Phan K, Ferenci T (2017) The fitness costs and trade-off shapes associated with the exclusion of nine antibiotics by OmpF porin channels. ISME J 11:1472–1482