



## OPEN ACCESS

## EDITED BY

Mingtao Huang,  
South China University of Technology, China

## REVIEWED BY

Pengsong Li,  
Beijing Forestry University, China  
Ashish Runthala,  
Koneru Lakshmaiah Education Foundation, India

## \*CORRESPONDENCE

María-Eugenia Guazzaroni  
✉ meguazzaroni@ffclrp.usp.br

RECEIVED 27 July 2023

ACCEPTED 30 August 2023

PUBLISHED 28 September 2023

## CITATION

Huanca-Juarez J, Nascimento-Silva EA,  
Silva NH, Silva-Rocha R and Guazzaroni M-E  
(2023) Identification and functional analysis of  
novel protein-encoding sequences related to  
stress-resistance.  
*Front. Microbiol.* 14:1268315.  
doi: 10.3389/fmicb.2023.1268315

## COPYRIGHT

© 2023 Huanca-Juarez, Nascimento-Silva,  
Silva, Silva-Rocha and Guazzaroni. This is an  
open-access article distributed under the terms  
of the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted which  
does not comply with these terms.

# Identification and functional analysis of novel protein-encoding sequences related to stress-resistance

Joshelin Huanca-Juarez<sup>1,2</sup>, Edson Alexandre Nascimento-Silva<sup>1,2</sup>,  
Ninna Hirata Silva<sup>1</sup>, Rafael Silva-Rocha<sup>3</sup> and  
María-Eugenia Guazzaroni<sup>2\*</sup>

<sup>1</sup>Department of Cell and Molecular Biology, Ribeirão Preto School of Medicine (FMRP) – University of São Paulo (USP), São Paulo, Brazil, <sup>2</sup>Department of Biology, Faculty of Philosophy, Sciences and Letters of Ribeirão Preto (FFCLRP) – University of São Paulo (USP), São Paulo, Brazil, <sup>3</sup>ByMyCell Inova Simples, São Paulo, Brazil

Currently, industrial bioproducts are less competitive than chemically produced goods due to the shortcomings of conventional microbial hosts. Thus, is essential developing robust bacteria for improved cell tolerance to process-specific parameters. In this context, metagenomic approaches from extreme environments can provide useful biological parts to improve bacterial robustness. Here, in order to build genetic constructs that increase bacterial resistance to diverse stress conditions, we recovered novel protein-encoding sequences related to stress-resistance from metagenomic databases using an *in silico* approach based on Hidden-Markov-Model profiles. For this purpose, we used metagenomic shotgun sequencing data from microbial communities of extreme environments to identify genes encoding chaperones and other proteins that confer resistance to stress conditions. We identified and characterized 10 novel protein-encoding sequences related to the DNA-binding protein HU, the ATP-dependent protease ClpP, and the chaperone protein DnaJ. By expressing these genes in *Escherichia coli* under several stress conditions (including high temperature, acidity, oxidative and osmotic stress, and UV radiation), we identified five genes conferring resistance to at least two stress conditions when expressed in *E. coli*. Moreover, one of the identified HU coding-genes which was retrieved from an acidic soil metagenome increased *E. coli* tolerance to four different stress conditions, implying its suitability for the construction of a synthetic circuit directed to expand broad bacterial resistance.

## KEYWORDS

stress resistance, bacterial robustness, industrial biotechnology, synthetic biology, metagenomics

## 1. Introduction

Industrial biotechnology focuses on working with biological systems, such as cell factories or enzymes, rather than using fossil raw materials to produce antibiotics, chemicals, and energy sources (Festel, 2020). As a result, it has attracted much attention for its environmentally friendly and sustainable methods, which could pave the way for a paradigm shift from petroleum to bio-based production (Yu et al., 2019). Therefore, this industry is positioned as a solution to

critical global concerns such as climate change mitigation, food security, ocean pollution reduction, and biodiversity conservation (Rosemann and Molyneux-Hodgson, 2020). Microbial Cell Factories, however, are subjected to harsh extracellular conditions such as high temperature, low pH, and metabolite toxicity, resulting in lower productivity than generated on a laboratory scale and, consequently, higher production costs (Gong et al., 2017; Jiang et al., 2020). Highly robust and tolerable strains towards non-natural or even toxic substrates or products and abiotic conditions are required since they can utilize cheaper carbon and energy sources, such as waste residues or biomass hydrolysates (Blombach et al., 2022).

Abiotic stress conditions—such as those present in industrial bioprocesses—produce considerable cellular damage in non-adapted microorganisms, affecting the structure and stability of various biomolecules, including nucleic acids, membrane lipids, and proteins (Rothschild and Mancinelli, 2001; Merino et al., 2019). To counteract these damage conditions, cells activate different repair mechanisms, such as DNA/RNA repair proteins, proteases to remove unfolded proteins, and multifunctional enzymes, such as RecA, Hup, RecR, RecO, PriA, and RecN. These enzymes are involved in DNA condensation, removal of broken DNA, and DNA recombination repair (Grosjean and Oshima, 2007; Wang et al., 2011; Somayaji et al., 2022). There is also a subset of activated post-translationally protein under specific stress conditions known as molecular chaperones, which act in several processes involving cellular protein folding, unfolding, and homeostasis (Saibil, 2013; Voth and Jakob, 2017). Considering that protein denaturation is a persistent direct or indirect result of stress, chaperones are vital in defensive cellular mechanisms (Jacob et al., 2017).

In addition to the enzymes mentioned above, there are other proteins of particular relevance as protective biological agents against stress damage to essential macromolecules. For instance, the DNA-binding protein HU (histone-like protein) has been shown to bind with high specificity to aberrant DNA formations such as nicks and strand breaks to preserve its integrity (Kamashev and Rouviere-Yaniv, 2000). The Clp protease complex consists of a proteolytic component, ClpP, in association with any of the ATPases, ClpA or ClpX, granting substrate specificity (Wang et al., 1997; Khan et al., 2022). Proteases, in general, play an essential role in protein quality control by removing short-lived regulatory proteins as well as misfolded and damaged proteins, which are toxic to the cell (Yu and Houry, 2007). Lastly, DnaJ is a homolog of eukaryotic HSP40 and acts as a cofactor of DnaK, stimulating its ATPase activity through the J domain, while additional domains give specificity to the system (Cheetham and Caplan, 1998; Ghafoori et al., 2017; Kampinga et al., 2019). In addition to participating in other metabolic tasks, these molecular chaperones have been shown to play an essential role in protein degradation (Huang et al., 2001).

In recent years, various studies have focused on the capacity of bacteria to survive in extreme environments in order to understand the mechanisms of biochemical adaptation to harsh conditions (Morozkina et al., 2010; Guazzaroni et al., 2015; Mirete et al., 2016). Extremophiles live in a wide range of hostile conditions and have evolved a set of adaptations to cope with these situations (Coker, 2019). However, most of them are not cultivable, which limits their exploration in biotechnological applications (Rappé and Giovannoni, 2003; Alves et al., 2018b). Culture-independent approaches help harness the biological activity and build components of this difficult-to-grow

biodiversity (van der Helm et al., 2018). In this context, functional and sequence-based metagenomics have been shown to be effective in the identification of novel genes that confer resistance to extreme conditions, such as enzymes, antibiotics, and other bioactive molecules derived from a variety of environments (Alves et al., 2018a). Functional metagenomic approaches allowed to identify novel genes relevant to biotechnology, involving the creation of expression libraries with thousands of metagenomic clones and activity-based screens in order to find genes that encode a function of interest (Guazzaroni et al., 2013; Mirete et al., 2016). On the other hand, the search for sequences of interest can also be done by sequence identification in (meta)genomic databases; however, low-level expression of heterologous proteins in commonly used hosts and low quality of (meta)genomic annotations are the main disadvantages in both approaches, respectively (Alves et al., 2018b; van der Helm et al., 2018). Furthermore, sequence-based screenings generally employ sequence pair aligners (such as BLAST, BLAT, Minimap2, KMA, and Bowtie) to detect regions of similarity between two biological sequences (Altschul et al., 1990; Kent, 2002; Langmead and Salzberg, 2012; Clausen et al., 2018). These tools constrain the homologous detection process since the percentage of identity is superior to a certain threshold, so more distant evolutionary sequences are lost (Skewes-Cox et al., 2014).

On the other hand, computational methods based on profiles, such as Hidden Markov Model (HMM), allow the detection of remote protein homologs (Kirsip and Abroi, 2019; Radivojac, 2022). In this context, profiling methods gained sensitivity by incorporating position-specific information into the alignment process and quantifying variation between family members at each position (Madera and Gough, 2002; Skewes-Cox et al., 2014). Some software based on HMM profiles were shown to be useful for identifying sequences of interest, such as IDOP or CryProcessor, used for detecting toxin genes (Shikov et al., 2020; Díaz-Valerio et al., 2021). In the same way, deep learning has been shown to be a powerful machine learning approach in predicting DNA sequence affinities and identifying new genes of interest, such as resistance genes or transcription factors (Arango-Argoty et al., 2018; Oliveira Monteiro et al., 2022). Considering the above, we aimed to identify microbial genes related to stress resistance by mining sequences from available metagenomic datasets from extreme environments using HMM profiles. For this achievement, we focused on searching novel protein sequences associated with the DNA-binding protein HU, the ATP-dependent protease ClpP, and the chaperone protein DnaJ. Then, we characterized their functional performance in *Escherichia coli* under five stress conditions (high temperature, acidity, oxidative and osmotic stress, and UV radiation). Through this *in silico* approach, we identified 10 new protein-encoding sequences related to stress-resistance that exhibited the potential to enhance resistance to various stress conditions in bacteria. These findings open up possibilities for the development of synthetic gene circuits aimed at broadening bacterial resistance capabilities.

## 2. Materials and methods

### 2.1. Bacterial strains and culture conditions

The bacterial strain *E. coli* DH10B was employed as a host in the stress resistance experiments. It was routinely grown in M9GlycAA

(1x M9 salts, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 0.1% casamino acids, 1% glycerol) under aerobic conditions at 37°C. Liquid cultures were cultivated in a shaker at 37°C and 200 rpm. For plasmid maintenance, the growth M9GlycAA was supplemented with 150 µg/mL ampicillin (Ap). For stress assays, overnight cultures were diluted to a final OD<sub>600</sub> 0.05 in 5 mL of M9GlycAA with antibiotic and incubated in an orbital shaker at 37°C and 200 rpm. When cultures reached the early exponential growth phase (OD<sub>600</sub> = 0.3–0.4), stress tests were initiated.

## 2.2. Selection of metagenomes from public databases

Raw shotgun metagenomes were curated, chosen, and downloaded from the MG-RAST database, hosted by the University of Chicago and the Argonne National Laboratory.<sup>1</sup> The data were then pre-processed using FastQC to check the sequence's quality, and when necessary, low-quality sequences were excluded using the Trimmomatic tool (Bolger et al., 2014). After that stage, metagenomes were assembled into contigs using the MEGAHIT program (Li et al., 2015), and an initial annotation was performed using Prokka.

## 2.3. Selection of genes of interest and construction of HMM profiles

HMM profiles were built based on selected stress-resistance genes previously documented in the literature. For this, we selected genes of nucleic acid binding protein genes (RBP, HU, DPS); DNA replication protein genes (GyrA, RecA, DnaA), chaperone and protease genes (ClpP, ClpX, ClpA, ClpC, ClpE, ClpL, DnaJ, DnaK); and CRISPR system genes (Cas1, Cas2, Cas9). The nucleotide sequences were obtained from the National Center for Biotechnology Information (NCBI) database, and alignment was performed with MAFFT (Multiple Alignment using Fast Fourier Transform) (Katoh et al., 2002). The profiles were created using the HMMer3 tool (Eddy, 2009), which was also used to detect homologous sequences by comparing the HMM profiles generated and the sequences of previously selected metagenomes. The results were analyzed and correlated with the literature, aiming to obtain 10 candidate genes. The selection process of protein sequences for further functional characterization was based on the following characteristics: (1) abundance of the proteins in the different environments, (2) metagenomic genes related to stress resistance in the literature, and (3) protein function. To continue the analysis and selection of sequences, a BLASTp was performed in order to identify the protein identity percentage.<sup>2</sup>

## 2.4. *In silico* analysis and 3D-structure model generation

Dendrograms for each class of protein were constructed based on the 23–25 amino acid sequences using MEGA X software. The

homologous sequences of selected proteins in different organisms from the NCBI database were retrieved in FASTA format. Alignment was done using the Muscle algorithm (Edgar, 2004) and used to build the dendrogram. The Neighbor-joining statistical method was applied, and the accuracy of the phylogenetic analysis was predicted by using 1,000 bootstrapping replications. The 3D-structure models were generated using the AlphaFold2 collab server (Jumper et al., 2021). All structures were analyzed using the PyMol v. 2.5.2 software<sup>3</sup> where the RMSD value of the alpha carbon atoms was considered. Proteins were aligned with T-Coffee v. 13.45.60.cd84d2a, and the alignment results were visualized in Jalview v. 2.11.2.3.

## 2.5. Cloning of the selected genes

All amino acid sequences previously selected by bioinformatics analysis were converted to nucleotides sequences and optimized for expression in *E. coli* using the EMBOSS Backtranseq tool.<sup>4</sup> Each nucleotide sequence was cloned under control from the constitutive promoter BBa\_J23106 (*Pj106*) and the ribosome binding site (RBS) BBa\_B0034. Promoter and RBS sequences were obtained from the iGEM Foundation Standard Biological Parts Catalog (Registry).<sup>5</sup> The genetic construct was assembled into the pUC19 vector. The synthesis of all plasmids containing the promoter, RBS, and candidate genes cloned into the pUC19 vector was requested from the company Integrated DNA Technologies (IDT). *E. coli* DH10B competent cells were transformed with the resulting plasmids.

## 2.6. Growth rate measurement and fitness cost calculation

Cell growth was monitored using a Victor X3 plate reader (PerkinElmer, Inc.), as described by de Siqueira et al. (2020). Microbial cultivation was evaluated at 30°C for 8 h. During this incubation period, the optical density at 600 nm was measured every 30 min. All experiments were performed with four biological replicates. For fitness cost calculation we followed the methodology described by de Siqueira et al. (2020).

## 2.7. Heat shock experiments

Exponential phase cultures of *E. coli* DH10B (OD<sub>600</sub> = 0.3–0.4) carrying an empty plasmid or the pUC19 vector with *hup*, *clpP*, or *dnaJ* genes were diluted at 10<sup>-1</sup> in PBS buffer (pH 7.2). Then, dilutions were incubated at 55°C, and aliquots were removed at 0 and 30 min. To determine CFU/mL, serial dilutions were made, and three 25 µL droplets (technical replicates) were placed onto M9GlycAA agar with Ap. Plates were incubated at 37°C for 24 h. Survival percentage was expressed as the number of CFU/mL remaining after heat treatment divided by the initial CFU/mL. *E. coli* DH10B carrying empty pUC19 plasmid was used as a negative control. Each experiment was performed at least three times.

1 <https://www.mg-rast.org/>

2 <https://blast.ncbi.nlm.nih.gov>

3 <http://www.pymol.org/>

4 [https://www.ebi.ac.uk/Tools/st/emboss\\_backtranseq/](https://www.ebi.ac.uk/Tools/st/emboss_backtranseq/)

5 <http://parts.igem.org>

## 2.8. Acid resistance assay

Exponential phase cultures of *E. coli* DH10B (OD<sub>600</sub> = 0.3–0.4) carrying an empty plasmid or the pUC19 vector with *hup*, *clpP*, or *dnaJ* genes were diluted at 10<sup>-1</sup> in PBS buffer (pH 7.2) or M9GlycAA acidified with HCl (pH = 3.0) and then incubated at 37°C. Aliquots were taken at 0 and 60 min, respectively. To determine CFU/mL, serial dilutions were plated and enumerated as previously described. Survival percentage was the number of CFU/mL remaining after acidity treatment divided by the CFU/mL at zero time. *E. coli* DH10B carrying empty pUC19 plasmid was used as a negative control. Each experiment was repeated at least three times.

## 2.9. Oxidative stress experiments

Exponential phase cultures of *E. coli* DH10B (OD<sub>600</sub> = 0.3–0.4) carrying an empty plasmid or the pUC19 vector with *hup*, *clpP*, or *dnaJ* genes were diluted at 10<sup>-1</sup> in PBS buffer (pH 7.2) or M9GlycAA + hydrogen peroxide 12 mM and incubated at 37°C. Aliquots were taken at 0 and 60 min, respectively. Survival percent was calculated as above. *E. coli* DH10B carrying empty pUC19 plasmid was used as a negative control. Each experiment was repeated at least three times.

## 2.10. UV radiation experiments

Overnight cultures of *E. coli* DH10B carrying an empty plasmid or a pUC19 vector with *hup*, *clpP*, or *dnaJ* genes were diluted in PBS buffer to 1.0 OD<sub>600</sub>. For UV radiation exposure, 10 μL of each culture were linearly spread onto M9GlycAA plates and irradiated with a germicidal lamp for 0, 5, 10, 15, 20, and 25 s at room temperature. Bacterial growth was evaluated after 24 h of incubation at 37°C. Cell growth up to a certain level of exposure was determined by counting at least 10 colonies in the spot. *E. coli* DH10B carrying empty pUC19 plasmid was used as a negative control. Each experiment was repeated at least three times.

## 2.11. Osmotic stress assay

For continuous NaCl exposure experiments, cells were cultivated in M9GlycAA medium +3.5% NaCl. Cultures were placed in a 96-well plate and incubated at 30°C for 8 h. During this incubation period, the optical density at 600 nm was measured every 30 min using a Victor X3 plate reader (PerkinElmer, Inc.). Growth rate was calculated as indicated above. Untransformed *E. coli* DH10B was used as a negative control. All experiments were performed with four biological replicates.

## 2.12. Data analysis and visualization

Data analysis and generation of growth curves were performed using the R statistical platform (version 4.2.0), using packages *ggpubr* and *ggplot2*. The *mipreadr* package version 0.1.0 was used to import and process the raw data produced by the plate reader. A Student's *t*-test with a 5% significance level was employed to compare the values

obtained. For clustering construction, we provide discrete arbitrary values to the different stress responses presented by clones for comparative purposes. Thus, data were classified as high, medium, and low-stress responses, and values of 2, 1, and 0 were given, respectively. Then, hierarchical clustering and normalization of data were performed using the R statistical platform (version 4.2.0).

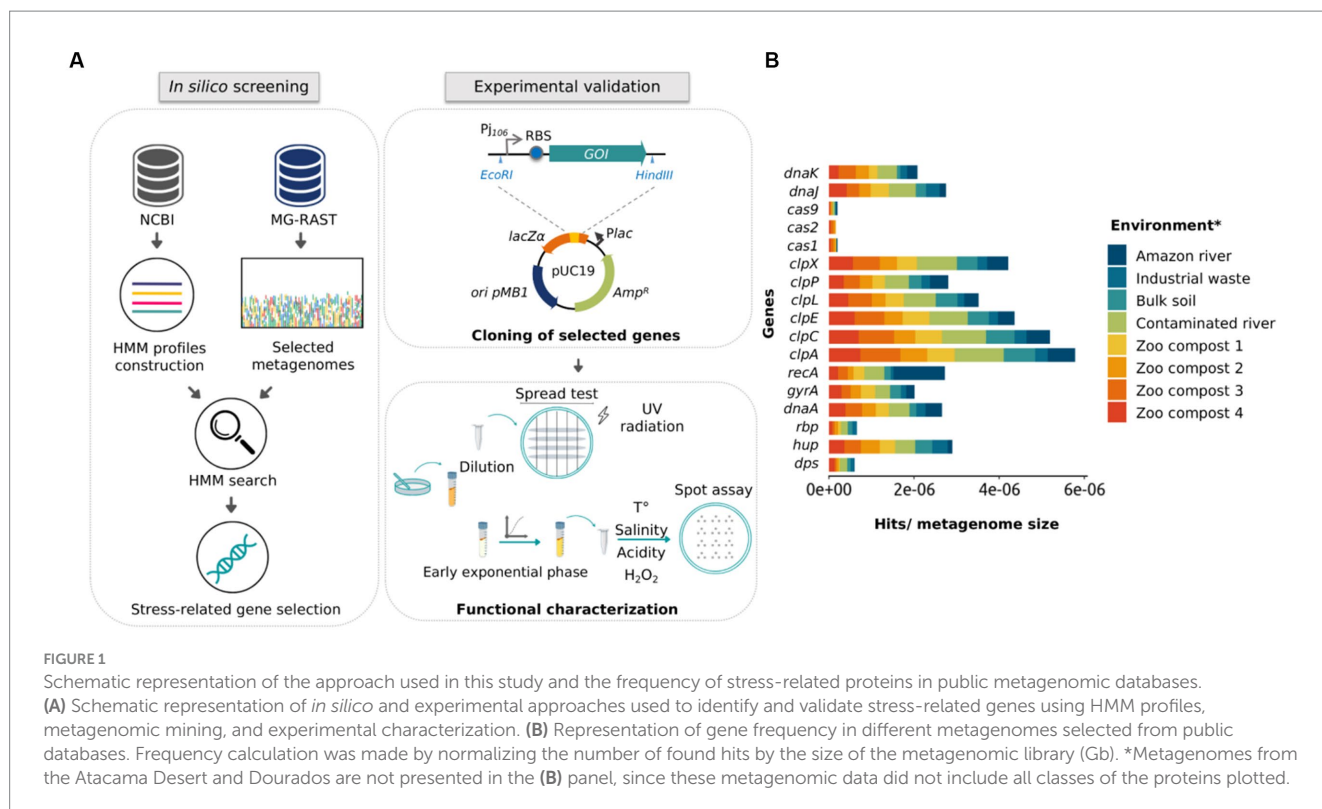
## 3. Results and discussion

### 3.1. *In silico* screening of stress-related genes

The search for new stress-resistance genes is a useful way for expanding robustness in microorganisms of industrial interest (Chen and Jiang, 2018; Wehrs et al., 2019). Although metagenomics offers an excellent alternative for discovering novel functional genes, activity-driven screenings are time-consuming and present low recovery rates (Guazzaroni et al., 2015; van der Helm et al., 2018). On the other hand, sequence-driven screening depends on the quality of the sequences—which present high annotation errors—and does not allow the discovery of new sequences if they have not been previously related to a specific function (Alves et al., 2018b). In turn, profile search methods are more sensitive than pairwise alignment in detecting low percent identity homologs by incorporating position-specific information into the alignment process and quantifying variation across family members at each position (Madera and Gough, 2002; Skewes-Cox et al., 2014). Considering that, we addressed the identification of new genes in metagenomic databases using an *in silico* approach based on HMM profiles. The general computational pipeline began searching metagenomic shotgun sequencing data from 10 microbial communities of extreme environments recruited from the MG-RAST database. Then, we aimed to identify genes encoding chaperones and other proteins—such as proteases and nucleic acid-binding proteins—that confer resistance to stress conditions (Figure 1A).

As the first step for selecting the metagenomes, we selected those who met the following requirements: (i) that were sequenced through the Illumina platform; (ii) that the sequences did not belong to DNA from the microbiota of humans/animals; (iii) that the size of the studies was around 1 gigabyte of DNA data, and; (iv) that metagenomes were originated from environments presenting harsh conditions (or in which there was some marked abiotic stress). Selected metagenomes are listed in Supplementary Table S1, it should be noted that about 85% are from the Brazilian territory. A summary of the metagenome assembly metrics returned, and the annotation results are shown in Supplementary Tables S2, S3, respectively. Next, to create an HMM profile, a multiple sequence alignment of 17 different stress-related proteins deposited in the NCBI database was performed using the HMMer3 server (Supplementary Figure S1). Then, metagenome data were searched against those HMM probabilistic profiles. The resulting profiles were compared with annotated protein regions, generating a total of 6,840 hits (Supplementary Table S4).

For the selection of protein sequences, it was considered that a protein found in a broader range of environments would be linked to the response to more than one stress factor. Figure 1B shows the abundance of the 17 protein families in the assembled metagenomes,



considering the number of hits normalized by the size of the corresponding metagenome. The graph shows that the *rbp* and *dps* genes had the lowest abundance, being disregarded in the subsequent analysis. Then, we carried out an extensive literature search and found that 50%–70% of the recovered published works connected *hup*, *clpP*, *clpX*, *dnaK*, and *dnaJ* genes to stress-defense mechanisms in microbial cells. As mentioned before, ClpP and ClpX protease subunits were reported to work together to produce a complex that degrades misfolded proteins and protects cells against its toxic effect (Baker and Sauer, 2012). However, different studies demonstrated that the ClpP protease subunit can function independently of the ATPase subunit ClpX (Guazzaroni et al., 2013; de Siqueira et al., 2020). On the other hand, the overexpression of the DnaK protein in *E. coli* presents a bactericidal effect (Blum et al., 1992; Jung and Ahn, 2022). Considering that, three families of proteins were selected for functional characterization, HU, ClpP, and DnaJ. To select the most promising protein sequences from the total sequences retrieved by the HMM approach, we evaluated if the chosen sequences were non-truncated when compared to homologous protein sequences using BLASTp. As a result, 24 protein sequences were selected (15 for HU, 6 for ClpP, and 3 for DnaJ), with protein identity percentage of 50%–85% for HU, 83%–97% for ClpP, and 55%–93% for DnaJ (Table 1).

### 3.2. Sequence analysis of selected protein candidates

The 24 selected amino acid sequences were used to construct dendrograms for each protein family. This analysis allowed us to categorize the sequences according to protein sequence identity and

select them based on their position resulting in the dendrogram (Supplementary Figures S2–S4). For this, we took into account as selection criteria (i) sequences presenting a lower percentage of amino acid identity; (ii) that belong to different environmental sources, and; (iii) that were located in different branches of the dendrogram. Those observations enabled us to choose 10 sequences for further functional characterization (Table 1). To verify the presence of conserved residues, sequence alignments (Supplementary Figure S5) and 3D structure models were constructed (Supplementary Figure S6). In the HU sequence alignment, the presence of proline 63 in HU sequences was verified (Supplementary Figure S5A), a highly conserved residue at the ends of the beta strands that surrounds the DNA and which, as part of an Arg-Asn-Pro motif (RNP), stabilizes the twists of the DNA (Grove, 2011). Interestingly, HU.d1 and HU.d2 have valine instead of arginine at position 61. The presence of Arg does not contribute to the stabilization of DNA twists, however, in *Thermotoga maritima*, this motif contains Val instead of Arg, therefore its removal appears to be required to maintain any distinction between perfect duplexes and damaged DNA (Grove, 2011). Ligation to perfect duplexes DNA would allow HU protein to protect them before exposure to stress, thus reducing damage. Additionally, the conserved glycine at position 15 was also observed, a residue that mediates loop flexibility and promotes the thermal stability of HU (Kawamura et al., 1996; Georgoulis et al., 2020) (Supplementary Figure S5A). Likewise, the presence of the Ser-His-Asp catalytic triad was observed in ClpP sequences, a motif fully conserved in members of the ClpP family and which plays an essential role in the cleavage capacity of proteases (Khan et al., 2022) (Supplementary Figure S5B). Similarly, the conserved His-Pro-Asp motif in DnaJ sequences was verified (Supplementary Figure S5C). This conserved motif recognizes and stimulates the ATPase activity of DnaK (Ahmad et al., 2011).

TABLE 1 Summary of the main characteristics of the selected proteins.

Protein name	Source	Length (aa)	Identity percentage	Closest related microorganism <sup>a</sup>
HU.d1	Dourados	91	76	<i>Thermoleophilia bacterium</i>
HU.d2	Dourados	88	74	<i>Solirubrobacterales bacterium</i>
HU.jz	Juazeiro	90	60.23	<i>Sphingomonas koreensis</i>
HU.go	mgpGO	90	70.79	Unclassified bacterium
ClpP.da	Atacama Desert	198	97.92	<i>Rubrobacteraceae bacterium</i>
ClpP.jz	Juazeiro	193	83.87	<i>Candidatus Saccharibacteria bacterium</i>
ClpP.z4	Zoo 4	203	93.88	<i>Actinobacteria bacterium</i>
DnaJ.tt	Tietê	375	55.65	<i>Rhizobium</i> sp. ARZ01
DnaJ.z4	Zoo 4	389	88.11	Unclassified bacterium
DnaJ.jz	Juazeiro	379	93.90	<i>Thermomonas aquatica</i>

<sup>a</sup>Considering the closest similar protein according to BLASTp hits.

We observed that the percentage of identity ranged from 60 to more than 90% in the 10 selected sequences when compared to proteins from BLASTp hits (Table 1), something that in principle would indicate that there is no novelty in the sequences. However, it should be noticed that none of these sequences have been experimentally characterized in previous studies.

Next, we carried out the cloning of the 10 selected sequences in individual genetic constructs (Supplementary Table S5). For this, each nucleotide sequence was cloned under control of the constitutive medium expression promoter P<sub>J106</sub> and the RBS BBa\_B0034, which was reported to present medium translation strength. Constructions were assembled into the pUC19 vector and resulting plasmids were transformed into *E. coli* DH10B (Figure 1A).

### 3.3. Metabolic burden evaluation of the genetic constructs

Considering that metabolic load, defined as the proportion of a host cell's resources—energy molecules or carbon building blocks—that are used to express heterologous proteins, has an important impact in normal cellular processes (Glick, 1995; Wu et al., 2016), we proceeded to evaluate the metabolic burden of the 10 genetic constructs by measuring the growth of the clones in a culture medium without any stress factor. For this, *E. coli* DH10B clones bearing empty plasmid or genetic constructs were grown in M9GlycAA medium (supplemented with Ap when strains carried a plasmid), and the fitness cost was calculated considering the growth rates (Figure 2A; Supplementary Table S6). Unexpectedly, the clone carrying the empty pUC19 plasmid showed the highest fitness cost (42.1%). In accordance with this observation, there are several reports found in the literature or in research discussion forums that describe difficulties related to the growth or transformation efficiency of *E. coli* cells transformed with empty plasmid pUC19 (Toh, 2013; Amirie et al., 2014). This is probably due to a higher number of copies of this plasmid when it is not bearing extra genes, which imposes a high metabolic cost on the cell. To circumvent this limitation, we performed the experiments with clones (bearing empty plasmid or genetic constructs) situated in the same metabolic state, that is, stress tests were just initiated when cultures reached the early exponential growth phase

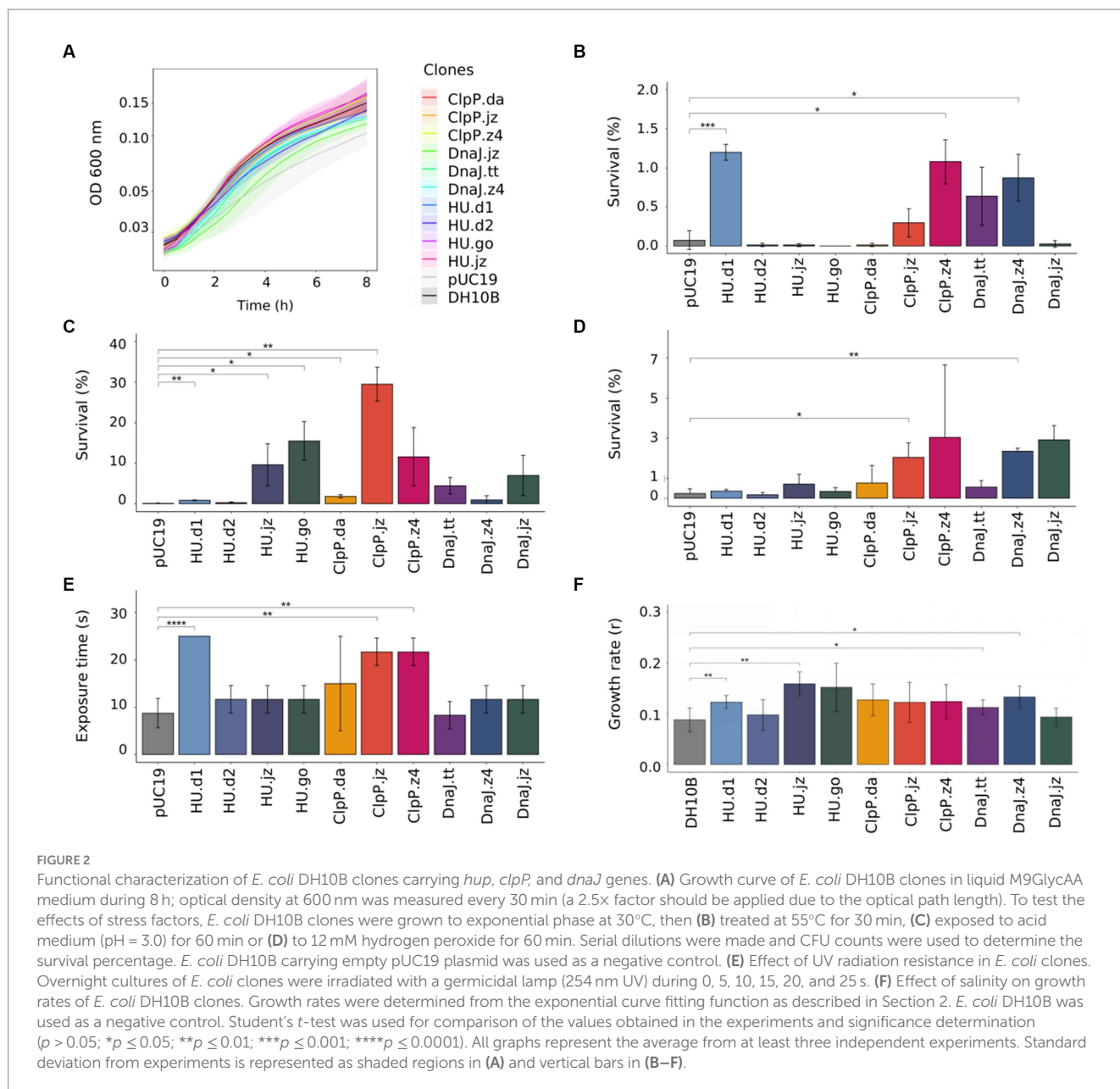
(OD<sub>600</sub> = 0.3–0.4). Also, for comparison purposes, untransformed *E. coli* cells were used as an additional negative control.

### 3.4. Response to stress conditions conferred by the stress-related genes

Microorganisms resistant to several stresses play an essential role in industrial bioprocesses because they permit increased yields and lower production costs (Thorwall et al., 2020; Wang et al., 2022). Production processes generally involve osmotic (e.g., processes involving high salt concentration culture medium), high temperature (such as processes that enable non-sterilized fermentation), acidity (e.g., organic acid production), and oxidative stresses (e.g., accumulation of radical oxygen species) (Patel et al., 2006; Qin et al., 2009; Li et al., 2011; Auesukaree, 2017; Şanlıer et al., 2019; Ye and Chen, 2021). Similarly, bioremediation of polluted environments is also affected by different abiotic parameters since wastewater is produced by various industries such as chemical, petrochemical, and textile (Le Borgne et al., 2008; Wernick et al., 2016; Koshlaf and Ball, 2017; Aliko et al., 2022). Another critical application of stress-resistant microorganisms is in enzyme production processes, in which bacterial hosts need to be active and stable at high salt contents or high temperatures (Margesin and Schinner, 2001; Verma et al., 2021).

In order to study if genetic constructs—bearing independent new sequences coding for the DNA-binding protein HU, the ATP-dependent Clp protease proteolytic subunit, and the molecular chaperone DnaJ—increase *E. coli* resistance to diverse stress conditions, we performed stress shocks assays with *E. coli* clones in five stress conditions: high temperature, acidity, oxidative and osmotic stress, and UV radiation (Figure 2; Supplementary Table S7).

Considering the set of assays in the five stress conditions, we observed that expression of most of the encoding sequences granted stress resistance to the cells in differing degrees (Figure 2). For example, cells expressing *hu.d1* showed a positive response to four stress factors over the ones expressing either of the other hu homologs, with an average survival percentage of over 1.2% in heat shock tests (Figure 2B) and 0.8% in the acidic challenge (Figure 2C). These values represented a fold change of 17 and 9, respectively, in comparison to *E. coli* cells harboring the empty plasmid (0.07% in heat shock assays



and 0.1% in acid shock experiments). These results agree with the previous work of Ogata et al. (1997), which showed that the HU protein maintains the negative supercoiling of DNA during thermal stress and contributes to cellular thermotolerance in *E. coli*. Similarly, the HU role in acid resistance was also corroborated by several studies (Bi et al., 2009; Guazzaroni et al., 2013; de Siqueira et al., 2020; Alves et al., 2023). In the UV radiation tests, cells expressing *hu.d1* also showed a higher percentage of survival (colony growth until 25 s of UV radiation exposure) compared with control cells (growth until 10 s) (Figure 2E; Supplementary Figure S7). This is probably because HU acts against UV radiation stress by preventing the formation of photoproducts and in the RecA-dependent repair process (Miyabe et al., 2000). At the same time, in the salinity assays, cells expressing *hu.d1* reached a growth rate of 0.12. This rate was 33% higher than that achieved by untransformed *E. coli* cells (Figure 2F; Supplementary Figure S8). The HU protein response to osmotic stress

could be involve in the modulation of expression of sigma factor *rpoS* ( $\sigma_{38}$ ), a global regulator whose translation is stimulated in response to high osmolarity (Muffler et al., 1996; Battesti et al., 2011).

In a similar way, cells expressing *clpP.jz* or *dnaJ.z4* showed a positive response when exposed to three different stress factors (acidity, oxidative stress, and UV radiation for cells expressing *clpP.jz*, and heat shock, oxidative, and osmotic stress for cells expressing *dnaJ.z4*). For example, cells expressing *clpP.jz* showed an average survival percentage of around 29% in the acidity challenge (Figure 2C) and 2% in oxidative stress (Figure 2D). These values represent a fold change of 290 and 9 times, respectively, in comparison to negative controls. Moreover, UV radiation tests also showed a significantly higher survival of cells expressing *clpP.jz* compared with cells bearing empty plasmid (20 s versus 10 s) (Figure 2E; Supplementary Figure S7). In accordance with the obtained results, there are several works in the literature describing a connection between this protease and cellular

damage protection in stress conditions (Guazzaroni et al., 2013; Chen et al., 2015; de Siqueira et al., 2020). For instance, a functional metagenomic study from an extremely acidic environment identified a *clpP* gene able to confer resistance to acidity and UV radiation in *E. coli* (Guazzaroni et al., 2013). Also, when this gene was used for the construction of a combinatorial acid stress-tolerance module improved the growth robustness of *E. coli* at low pH (de Siqueira et al., 2020). Additional examples in other microorganisms, such as *Bifidobacterium longum* and *Lactococcus lactis*, also show that ClpP is involved in the acid stress response (Frees and Ingmer, 1999; Wei et al., 2015). It is important to highlight that during oxidative stress, hydrogen peroxide diffuses easily into cells, increasing intramolecular concentrations (Sen et al., 2020). In the cytoplasm, the reactions produce hydroxyl radicals that can damage DNA and other biomolecules,—similar to the effects of protons during acid stress—, endangering the survival of the cell (Arcari et al., 2020; Sen et al., 2020; Figure 3). Hydrogen peroxide triggers the production of reactive oxygen species (ROS), which, in turn, oxidize proteins, resulting in the formation of misfolded proteins (Daly et al., 2007; Wang et al., 2009). These misfolded proteins must undergo degradation. Similarly, after UV radiation exposure, excited sensitizers can directly react with DNA directly by transferring electrons or energy to molecular oxygen, thus generating ROS, which in turn can damage the DNA molecule (Schuch and Menck, 2010).

On the other hand, clones expressing *dnaJ.z4* had an average survival of 0.87% in the heat shock test (Figure 2B) and 2.35% in oxidative stress (Figure 2D), representing more than 12 and 10 times of survival in respective negative control. Also, when the cells expressing *dnaJ.z4* were subjected to saline stress, they reached a 44% higher growth rate compared to untransformed *E. coli* cells (Figure 2F; Supplementary Figure S8). It should be noticed that DnaJ protein is part of the molecular response

machinery to heat shock (Ghafoori et al., 2017), confirming the results found in our study. A recent work in which the *dnaJ* gene from *Bacillus halodurans* was overexpressed in *E. coli* showed improved cell thermotolerance, indicating that DnaJ overexpression was sufficient to increase resistance to high temperatures (Ghafoori et al., 2017). On the other hand, our study is the first one—to the best of our knowledge—to report DnaJ protein protection against oxidative cellular damage.

## 4. Conclusion

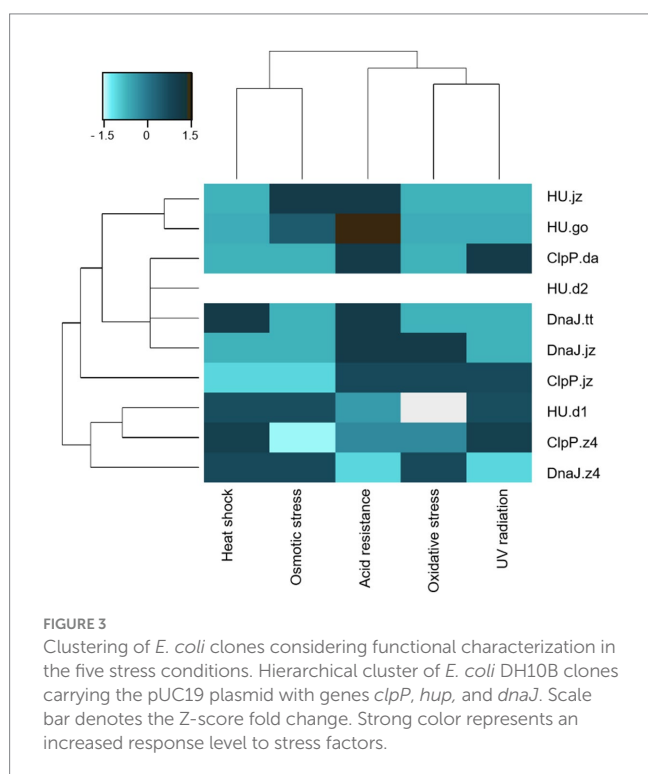
The approach presented here allowed us to identify five novel coding sequences conferring resistance to at least two stress conditions when expressed in *E. coli*. It should be noticed that all expressed proteins, except for HU.d2, demonstrated resistance to at least one of the tested stress conditions (Figures 2, 3). Notably, among all genetic constructs, cells expressing HU.d2 exhibited a significantly higher metabolic cost compared to the other clones (Figure 2A; Supplementary Table S6), which should explain its limited response to stress conditions. Also, it is worth mentioning that the response level varied among clones despite proteins belonging to the same family. This inconsistency may be attributed to the origin of metagenomic DNA from diverse microorganisms, where certain genes tend to be preferentially expressed in hosts possessing transcription/translation machinery more closely resembling the source organism (Gabor et al., 2004).

Clustering of *E. coli* clones stress-responses illustrated that probably the expression of each protein corresponds to specific molecular mechanisms of survival (Figure 3). Notably, UV radiation, acidity, and oxidative stress exhibit a similar response pattern, probably due to the fact that UV radiation and acidity indirectly induce oxidative stress by generating ROS (Schuch and Menck, 2010).

Despite considerable research on stress response available in the literature, the approach presented here has revealed novel aspects of this well-established response, including genes that potentially enhance resistance. On the other hand, from a biotechnological point of view, one of the identified novel coding sequences (*hu.d1*) retrieved from an acidic soil metagenome notably increased microbial tolerance to four different stress conditions, indicating its suitability for the construction of a synthetic circuit directed to expand broad bacterial resistance. Moreover, the assembly of genetic constructs with a combination of the best-identified gene candidates from our study (e.g., *hu.d1*, *clpP.jz*, and *dnaJ.z4*) should lead to improved bacterial endurance. Taken together, this work provides novel biological parts for the engineering of microbial hosts which could be used for extreme industrial applications. Still, further research is necessary to comprehend the mechanisms of action when these proteins are heterologously expressed and to understand the functional behavior of the identified genes in other hosts than *E. coli*.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.





## Author contributions

JH-J: Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. EN-S: Formal analysis, Investigation, Methodology, Writing – review & editing. NS: Formal analysis, Methodology, Software, Writing – review & editing. RS-R: Conceptualization, Formal analysis, Supervision, Writing – review & editing. M-EG: Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the São Paulo State Foundation (FAPESP, award # 2021/01748-5). JH-J was the beneficiary of a CNPq scholarship. M-EG was supported by CNPq Research Productivity Scholarship (award # 302750/2020-7).

## Acknowledgments

The authors would like to thank the lab technician Thalita Riul Prado for her invaluable assistance in the course of this work.

## References

- Ahmad, A., Bhattacharya, A., McDonald, R. A., Cordes, M., Ellington, B., Bertelsen, E. B., et al. (2011). Heat shock protein 70 kDa chaperone/DnaJ co-chaperone complex employs an unusual dynamic interface. *Proc. Natl. Acad. Sci.* 108, 18966–18971. doi: 10.1073/pnas.1111220108
- Aliko, V., Multisanti, C. R., Turani, B., and Faggio, C. (2022). Get rid of marine pollution: bioremediation an innovative, attractive, and successful cleaning strategy. *Sustainability* 14:11784. doi: 10.3390/su141811784
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Alves, L. d. F., Bortolucci, J., Reginatto, V., Guazzaroni, M.-E., and Mussatto, S. I. (2023). Improving *Saccharomyces cerevisiae* acid and oxidative stress resistance using a prokaryotic gene identified by functional metagenomics. *Heliyon* 9:e14838. doi: 10.1016/j.heliyon.2023.e14838
- Alves, L. d. F., Meleiro, L. P., Silva, R. N., Westmann, C. A., and Guazzaroni, M.-E. (2018a). Novel ethanol- and 5-hydroxymethyl furfural-stimulated  $\beta$ -glucosidase retrieved from a Brazilian secondary Atlantic Forest soil metagenome. *Front. Microbiol.* 9:2556. doi: 10.3389/fmicb.2018.02556
- Alves, L. d. F., Westmann, C. A., Lovate, G. L., de Siqueira, G. M. V., Borelli, T. C., and Guazzaroni, M.-E. (2018b). Metagenomic approaches for understanding new concepts in microbial science. *Int. J. Genomics* 2018b:e2312987, 1–15. doi: 10.1155/2018/2312987
- Amirie, M., Cheng, I., Cho, J., and Vu, V. (2014). Attempts to construct an enlarged pUC19 via insertion of HindIII-digested Coliphage  $\lambda$  DNA. *J. Microbiol. Immunol. Infect.* 18, 96–101.
- Arango-Argoty, G., Garner, E., Pruden, A., Heath, L. S., Vikesland, P., and Zhang, L. (2018). DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* 6:23. doi: 10.1186/s40168-018-0401-z
- Arcari, T., Feger, M.-L., Guerreiro, D. N., Wu, J., and O'Byrne, C. P. (2020). Comparative review of the responses of *Listeria monocytogenes* and *Escherichia coli* to low pH stress. *Genes* 11:1330. doi: 10.3390/genes11111330
- Auesukaree, C. (2017). Molecular mechanisms of the yeast adaptive response and tolerance to stresses encountered during ethanol fermentation. *J. Biosci. Bioeng.* 124, 133–142. doi: 10.1016/j.jbiosc.2017.03.009
- Baker, T. A., and Sauer, R. T. (2012). ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim. Biophys. Acta* 1823, 15–28. doi: 10.1016/j.bbamcr.2011.06.007
- Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-mediated general stress response in *Escherichia coli*. *Annu. Rev. Microbiol.* 65, 189–213. doi: 10.1146/annurev-micro-090110-102946

## Conflict of interest

RS-R is employed by ByMyCell Inova Simples. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1268315/full#supplementary-material>

- Bi, H., Sun, L., Fukamachi, T., Saito, H., and Kobayashi, H. (2009). HU participates in expression of a specific set of genes required for growth and survival at acidic pH in *Escherichia coli*. *Curr. Microbiol.* 58, 443–448. doi: 10.1007/s00284-008-9340-4
- Blombach, B., Grünberger, A., Centler, F., Wierckx, N., and Schmid, J. (2022). Exploiting unconventional prokaryotic hosts for industrial biotechnology. *Trends Biotechnol.* 40, 385–397. doi: 10.1016/j.tibtech.2021.08.003
- Blum, P., Ory, J., Bauernfeind, J., and Krška, J. (1992). Physiological consequences of DnaK and DnaJ overproduction in *Escherichia coli*. *J. Bacteriol.* 174, 7436–7444. doi: 10.1128/jb.174.22.7436-7444.1992
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Cheetham, M. E., and Caplan, A. J. (1998). Structure, function and evolution of DnaJ: conservation and adaptation of chaperone function. *Cell Stress Chaperones* 3, 28–36. doi: 10.1379/1466-1268(1998)003<0028:SFAEOD>2.3.CO;2
- Chen, L., Hu, M., Huang, L., Hua, Z., Kuang, J., Li, S., et al. (2015). Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. *ISME J.* 9, 1579–1592. doi: 10.1038/ismej.2014.245
- Chen, G.-Q., and Jiang, X.-R. (2018). Next generation industrial biotechnology based on extremophilic bacteria. *Curr. Opin. Biotechnol.* 50, 94–100. doi: 10.1016/j.copbio.2017.11.016
- Clausen, P. T. L. C., Aarestrup, F. M., and Lund, O. (2018). Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 19:307. doi: 10.1186/s12859-018-2336-6
- Coker, J. A. (2019). Recent advances in understanding extremophiles. *F1000Research* 8:1917. doi: 10.12688/f1000research.20765.1
- Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., Vasilenko, A., Zhai, M., Leapman, R. D., et al. (2007). Protein oxidation implicated as the primary determinant of bacterial radioresistance. *PLoS Biol.* 5:e92. doi: 10.1371/journal.pbio.0050092
- de Siqueira, G. M. V., Silva-Rocha, R., and Guazzaroni, M.-E. (2020). Turning the screw: engineering extreme pH resistance in *Escherichia coli* through combinatorial synthetic operons. *ACS Synth. Biol.* 9, 1254–1262. doi: 10.1021/acssynbio.0c00089
- Díaz-Valerio, S., Lev Hachoen, A., Schöppe, R., and Liesegang, H. (2021). IDOPS, a profile HMM-based tool to detect pesticidal sequences and compare their genetic context. *Front. Microbiol.* 12:664476. doi: 10.3389/fmicb.2021.664476
- Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform.* 23, 205–211.

- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113. doi: 10.1186/1471-2105-5-113
- Festel, G. (2020). Economic aspects of industrial biotechnology. In M. Fröhling and M. Hiete (Orgs.), *Sustainability and life cycle assessment in industrial biotechnology* (p. 53–74). Berlin: Springer International Publishing
- Frees, D., and Ingmer, H. (1999). ClpP participates in the degradation of misfolded protein in *Lactococcus lactis*. *Mol. Microbiol.* 31, 79–87. doi: 10.1046/j.1365-2958.1999.01149.x
- Gabor, E. M., Alkema, W. B. L., and Janssen, D. B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* 6, 879–886. doi: 10.1111/j.1462-2920.2004.00640.x
- Georgoulis, A., Louka, M., Mylonas, S., Stavros, P., Nounesis, G., and Vorgias, C. E. (2020). Consensus protein engineering on the thermostable histone-like bacterial protein HUs significantly improves stability and DNA binding affinity. *Extremophiles* 24, 293–306. doi: 10.1007/s00792-020-01154-4
- Ghafoori, H., Askari, M., and Sarikhan, S. (2017). Molecular cloning, expression and functional characterization of the 40-kDa heat shock protein, DnaJ, from *Bacillus halodurans*. *Process Biochem.* 54, 33–40. doi: 10.1016/j.procbio.2016.12.017
- Glick, B. R. (1995). Metabolic load and heterologous gene expression. *Biotechnol. Adv.* 13, 247–261. doi: 10.1016/0734-9750(95)00004-A
- Gong, Z., Nielsen, J., and Zhou, Y. J. (2017). Engineering robustness of microbial cell factories. *Biotechnol. J.* 12:1700014. doi: 10.1002/biot.201700014
- Grosjean, H., and Oshima, T. (2007). “How nucleic acids cope with high temperature” in *Physiology and biochemistry of extremophiles*. eds. C. Gerday and N. Glansdorff (Chichester: John Wiley & Sons, Ltd), 39–56.
- Grove, A. (2011). Functional evolution of bacterial histone-like HU proteins. *Curr. Issues Mol. Biol.* 13:Artigo 1. doi: 10.21775/cimb.013001
- Guazzaroni, M.-E., Morgante, V., Mirete, S., and González-Pastor, J. E. (2013). Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. *Environ. Microbiol.* 15, 1088–1102. doi: 10.1111/1462-2920.12021
- Guazzaroni, M.-E., Silva-Rocha, R., and Ward, R. J. (2015). Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microb. Biotechnol.* 8, 52–64. doi: 10.1111/1751-7915.12146
- Huang, H.-C., Sherman, M. Y., Kandror, O., and Goldberg, A. L. (2001). The molecular chaperone DnaJ is required for the degradation of a soluble abnormal protein in *Escherichia coli* \*. *J. Biol. Chem.*, 276, 3920–3928. doi: 10.1074/jbc.M002937200
- Jacob, P., Hirt, H., and Bendahmane, A. (2017). The heat-shock protein/chaperone network and multiple stress resistance. *Plant Biotechnol. J.* 15, 405–414. doi: 10.1111/pbi.12659
- Jiang, T., Li, C., Teng, Y., Zhang, R., and Yan, Y. (2020). Recent advances in improving metabolic robustness of microbial cell factories. *Curr. Opin. Biotechnol.* 66, 69–77. doi: 10.1016/j.copbio.2020.06.006
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Jung, M., and Ahn, Y.-J. (2022). Growth-enhancing effect of bacterial and plant heat shock proteins in *Escherichia coli*. *Biocatal. Agric. Biotechnol.* 46:102545. doi: 10.1016/j.cbab.2022.102545
- Kamashev, D., and Rouviere-Yaniv, J. (2000). The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *EMBO J.* 19, 6527–6535. doi: 10.1093/emboj/19.23.6527
- Kampinga, H. H., Andreasson, C., Barducci, A., Cheetham, M. E., Cyr, D., Emanuelsson, C., et al. (2019). Function, evolution, and structure of J-domain proteins. *Cell Stress Chaperones* 24, 7–15. doi: 10.1007/s12192-018-0948-4
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Kawamura, S., Kakuta, Y., Tanaka, I., Hikichi, K., Kuhara, S., Yamasaki, N., et al. (1996). Glycine-15 in the bend between two  $\alpha$ -helices can explain the thermostability of DNA binding protein HU from *Bacillus stearothermophilus*. *Biochemistry* 35, 1195–1200. doi: 10.1021/bi951581l
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202
- Khan, Y. A., White, K. I., and Brunger, A. T. (2022). The AAA+ superfamily: a review of the structural and mechanistic principles of these molecular machines. *Crit. Rev. Biochem. Mol. Biol.* 57, 156–187. doi: 10.1080/10409238.2021.1979460
- Kirsip, H., and Abroi, A. (2019). Protein structure-guided hidden Markov models (HMMs) as a powerful method in the detection of ancestral endogenous viral elements. *Viruses* 11:Artigo 4. doi: 10.3390/v11040320
- Koshlaf, E., and Ball, A. S. (2017). Soil bioremediation approaches for petroleum hydrocarbon polluted environments. *AIMS Microbiol.* 3, 25–49. doi: 10.3934/microbiol.2017.1.25
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Le Borgne, S., Paniagua, D., and Vazquez-Duhalt, R. (2008). Biodegradation of organic pollutants by halophilic bacteria and archaea. *J. Mol. Microbiol. Biotechnol.* 15, 74–92. doi: 10.1159/000121323
- Li, Q., Bai, Z., O'Donnell, A., Harvey, L. M., Hoskisson, P. A., and McNeil, B. (2011). Oxidative stress in fungal fermentation processes: the roles of alternative respiration. *Biotechnol. Lett.* 33, 457–467. doi: 10.1007/s10529-010-0471-x
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Madera, M., and Gough, J. (2002). A comparison of profile hidden Markov model procedures for remote homology detection. *Nucleic Acids Res.* 30, 4321–4328. doi: 10.1093/nar/gkf544
- Margesin, R., and Schinner, F. (2001). Potential of halotolerant and halophilic microorganisms for biotechnology. *Extremophiles* 5, 73–83. doi: 10.1007/s007920100184
- Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.00780
- Mirete, S., Morgante, V., and González-Pastor, J. E. (2016). Functional metagenomics of extreme environments. *Curr. Opin. Biotechnol.* 38, 143–149. doi: 10.1016/j.copbio.2016.01.017
- Miyabe, Q.-M. Z., Zhang, Q. M., Kano, Y., and Yonei, S. (2000). Histone-like protein HU is required for recA gene-dependent DNA repair and SOS induction pathways in UV-irradiated *Escherichia coli*. *Int. J. Radiat. Biol.* 76, 43–49. doi: 10.1080/09553000138998
- Morozkina, E. V., Slutskaia, E. S., Fedorova, T. V., Tugay, T. I., Golubeva, L. I., and Koroleva, O. V. (2010). Extremophilic microorganisms: biochemical adaptation and biotechnological application (review). *Appl. Biochem. Microbiol.* 46, 1–14. doi: 10.1134/S0003683810010011
- Muffler, A., Traulsen, D. D., Lange, R., and Hengge-Aronis, R. (1996). Posttranscriptional osmotic regulation of the sigma(s) subunit of RNA polymerase in *Escherichia coli*. *J. Bacteriol.* 178, 1607–1613. doi: 10.1128/jb.178.6.1607-1613.1996
- Ogata, Y., Inoue, R., Mizushima, T., Kano, Y., Miki, T., and Sekimizu, K. (1997). Heat shock-induced excessive relaxation of DNA in *Escherichia coli* mutants lacking the histone-like protein HU. *Biochim. Biophys. Acta* 1353, 298–306. doi: 10.1016/S0167-4781(97)00105-X
- Oliveira Monteiro, L. M., Saraiva, J. P., Brizola Toscan, R., Stadler, P. F., Silva-Rocha, R., and Nunes da Rocha, U. (2022). PredicTF: prediction of bacterial transcription factors in complex microbial communities using deep learning. *Environ. Microbiome* 17:7. doi: 10.1186/s40793-021-00394-x
- Patel, M. A., Ou, M. S., Harbrucker, R., Aldrich, H. C., Buszko, M. L., Ingram, L. O., et al. (2006). Isolation and characterization of acid-tolerant, thermophilic bacteria for effective fermentation of biomass-derived sugars to lactic acid. *Appl. Environ. Microbiol.* 72, 3228–3235. doi: 10.1128/AEM.72.5.3228-3235.2006
- Qin, J., Zhao, B., Wang, X., Wang, L., Yu, B., Ma, Y., et al. (2009). Non-sterilized fermentative production of polymer-grade L-lactic acid by a newly isolated thermophilic strain *Bacillus* sp. 2–6. *PLoS One* 4:e4359. doi: 10.1371/journal.pone.0004359
- Radivojac, P. (2022). Advancing remote homology detection: a step toward understanding and accurately predicting protein function. *Cell Syst.* 13, 435–437. doi: 10.1016/j.cels.2022.05.006
- Rappé, M. S., and Giovannoni, S. J. (2003). The uncultured microbial majority. *Annu. Rev. Microbiol.* 57, 369–394. doi: 10.1146/annurev.micro.57.030502.090759
- Rosemann, A., and Molyneux-Hodgson, S. (2020). Industrial biotechnology: to what extent is responsible innovation on the agenda? *Trends Biotechnol.* 38, 5–7. doi: 10.1016/j.tibtech.2019.07.006
- Rothschild, L. J., and Mancinelli, R. L. (2001). Life in extreme environments. *Nature* 409, 1092–1101. doi: 10.1038/35059215
- Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* 14, 630–642. doi: 10.1038/nrm3658
- Şanlıer, N., Gökçen, B. B., and Sezgin, A. C. (2019). Health benefits of fermented foods. *Crit. Rev. Food Sci. Nutr.* 59, 506–527. doi: 10.1080/10408398.2017.1383355
- Schuch, A. P., and Menck, C. F. M. (2010). The genotoxic effects of DNA lesions induced by artificial UV-radiation and sunlight. *J. Photochem. Photobiol. B Biol.* 99, 111–116. doi: 10.1016/j.jphotobiol.2010.03.004
- Sen, A., Zhou, Y., and Imlay, J. A. (2020). During oxidative stress the Clp proteins of *Escherichia coli* ensure that iron pools remain sufficient to reactivate oxidized metalloenzymes. *J. Bacteriol.* 202, e00235–e00220. doi: 10.1128/JB.00235-20
- Shikov, A. E., Malovichko, Y. V., Skitchenko, R. K., Nizhnikov, A. A., and Antonets, K. S. (2020). No more tears: mining sequencing data for novel Bt cry toxins with CryProcessor. *Toxins* 12:204. doi: 10.3390/toxins12030204
- Skewes-Cox, P., Sharpton, T. J., Pollard, K. S., and DeRisi, J. L. (2014). Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *PLoS One* 9:e105067. doi: 10.1371/journal.pone.0105067

- Somayaji, A., Dhanjal, C. R., Lingamsetty, R., Vinayagam, R., Selvaraj, R., Varadavenkatesan, T., et al. (2022). An insight into the mechanisms of homeostasis in extremophiles. *Microbiol. Res.* 263:127115. doi: 10.1016/j.micres.2022.127115
- Thorwall, S., Schwartz, C., Chartron, J. W., and Wheeldon, I. (2020). Stress-tolerant non-conventional microbes enable next-generation chemical biosynthesis. *Nat. Chem. Biol.* 16:Artigo 2. doi: 10.1038/s41589-019-0452-x
- Toh, S. Y. (2013). The study of exclusion effect of pBR322 using its rop-inactivated mutant, during co-transformation with pBR322 and pUC19; plasmid copy number does not relate to the exclusion of pBR322. *J. Exp. Microbiol. Immunol.* 17, 109–114.
- van der Helm, E., Genee, H. J., and Sommer, M. O. A. (2018). The evolving interface between synthetic biology and functional metagenomics. *Nat. Chem. Biol.* 14, 752–759. doi: 10.1038/s41589-018-0100-x
- Verma, S., Meghwanshi, G. K., and Kumar, R. (2021). Current perspectives for microbial lipases from extremophiles and metagenomics. *Biochimie* 182, 23–36. doi: 10.1016/j.biochi.2020.12.027
- Voth, W., and Jakob, U. (2017). Stress-activated chaperones: a first line of defense. *Trends Biochem. Sci.* 42, 899–913. doi: 10.1016/j.tibs.2017.08.006
- Wang, S., Deng, K., Zaremba, S., Deng, X., Lin, C., Wang, Q., et al. (2009). Transcriptomic response of *Escherichia coli* O157:H7 to oxidative stress. *Appl. Environ. Microbiol.* 75, 6110–6123. doi: 10.1128/AEM.00914-09
- Wang, J., Hartling, J. A., and Flanagan, J. M. (1997). The structure of ClpP at 2.3 Å resolution suggests a model for ATP-dependent proteolysis. *Cells* 91, 447–456. doi: 10.1016/S0092-8674(00)80431-6
- Wang, L., Li, B., Su, R.-R., Wang, S.-P., Xia, Z.-Y., Xie, C.-Y., et al. (2022). Screening novel genes by a comprehensive strategy to construct multiple stress-tolerant industrial *Saccharomyces cerevisiae* with prominent bioethanol production. *Biotechnol. Biofuels Bioproducts* 15:11. doi: 10.1186/s13068-022-02109-x
- Wang, G., Lo, L. F., and Maier, R. J. (2011). The RecRO pathway of DNA recombinational repair in *Helicobacter pylori* and its role in bacterial survival in the host. *DNA Repair* 10, 373–379. doi: 10.1016/j.dnarep.2011.01.004
- Wehrs, M., Tanjore, D., Eng, T., Lievense, J., Pray, T. R., and Mukhopadhyay, A. (2019). Engineering robust production microbes for large-scale cultivation. *Trends Microbiol.* 27, 524–537. doi: 10.1016/j.tim.2019.01.006
- Wei, Y.-X., Ye, L., Liu, D.-B., Zhang, Z.-Y., Liu, C., and Guo, X.-K. (2015). Activation of the chromosomally encoded mazEFBif locus of *Bifidobacterium longum* under acid stress. *Int. J. Food Microbiol.* 207, 16–22. doi: 10.1016/j.ijfoodmicro.2015.04.028
- Wernick, D. G., Pontrelli, S. P., Pollock, A. W., and Liao, J. C. (2016). Sustainable biorefining in wastewater by engineered extreme alkaliphile *Bacillus marmarensis*. *Sci. Rep.* 6:Artigo 1. doi: 10.1038/srep20224
- Wu, G., Yan, Q., Jones, J. A., Tang, Y. J., Fong, S. S., and Koffas, M. A. G. (2016). Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications. *Trends Biotechnol.* 34, 652–664. doi: 10.1016/j.tibtech.2016.02.010
- Ye, J.-W., and Chen, G.-Q. (2021). *Halomonas* as a chassis. *Essays Biochem.* 65, 393–403. doi: 10.1042/EBC20200159
- Yu, A. Y. H., and Houry, W. A. (2007). ClpP: a distinctive family of cylindrical energy-dependent serine proteases. *FEBS Lett.* 581, 3749–3757. doi: 10.1016/j.febslet.2007.04.076
- Yu, L.-P., Wu, F.-Q., and Chen, G.-Q. (2019). Next-generation industrial biotechnology-transforming the current industrial biotechnology into competitive processes. *Biotechnol. J.* 14:1800437. doi: 10.1002/biot.201800437