NAR Genomics and Bioinformatics, 2024, **6**, lqae110
https://doi.org/10.1093/nargab/lqae110
Advance access publication date: 24 August 2024
Standard Article

Synonymous *rpsH* variants: the common denominator in *Escherichia coli* adapting to ionizing radiation

Katharina Stemwedel [1], Nadin Haase [1], Simon Christ [1], Natalia V. Bogdanova [2] and Sophia Rudorf [1,*]

[1]Leibniz University Hannover, Institute of Cell Biology and Biophysics, Hannover, 30167, Germany
[2]Hannover Medical School, Radiation Oncology, Hannover, 30625, Germany
[*]To whom correspondence should be addressed. Tel: +49 511 76216175; Email: rudorf@cell.uni-hannover.de

Abstract

Ionizing radiation (IR) in high doses is generally lethal to most organisms. Investigating mechanisms of radiation resistance is crucial for gaining insights into the underlying cellular responses and understanding the damaging effects of IR. In this study, we conducted a comprehensive analysis of sequencing data from an evolutionary experiment aimed at understanding the genetic adaptations to ionizing radiation in *Escherichia coli*. By including previously neglected synonymous mutations, we identified the *rpsH* c.294T > G variant, which emerged in all 17 examined isolates across four subpopulations. The identified variant is a synonymous mutation affecting the 30S ribosomal protein S8, and consistently exhibited high detection and low allele frequencies in all subpopulations. This variant, along with two additional *rpsH* variants, potentially influences translational control of the ribosomal *spc* operon. The early emergence and stability of these variants suggest their role in adapting to environmental stress, possibly contributing to radiation resistance. Our findings shed light on the dynamics of ribosomal variants during the evolutionary process and their potential role in stress adaptation, providing valuable implications for understanding clinical radiation sensitivity and improving radiotherapy.

Introduction

Ionizing radiation (IR) poses a hazardous impact on all organisms by inducing damage to proteins and DNA. The predominant contributor to toxicity, whether it stems from protein or DNA damage within cells, remains a subject of ongoing investigation (1–5). A comprehensive understanding of the underlying mechanisms holds the potential to enhance the medical applications of radiation, particularly in therapy and diagnostic scans (6). IR inflicts damage directly to DNA and indirectly through the generation of reactive oxygen species (ROS) via intracellular water radiolysis (3,4). These ROS molecules engage with neighboring molecules, amplifying the detrimental effects beyond the initial target site (7). Direct DNA damage manifests as alterations in bases and strand breakage. If two single-strand breaks are in proximity on opposite strands, a double-strand DNA break (DSB) can occur (3,5). DSBs represent the most lethal form of DNA damage, causing DNA replication stalling, replication fork collapse, and posing significant challenges for repair. A single unrepaired DSB can lead to cell death (1,3). The intricate mechanisms and genes implicated in the response to irradiation continue to be subjects of active investigation.

Bruckbauer *et al.* performed an evolutionary experiment to study the emergence of radiation resistance in *Escherichia coli* (8,9). In a canonical evolutionary experiment, a random fraction of the population is eliminated by a bottleneck, where bottleneck size and growth phases have an impact on the fixation probability for a beneficial mutation. Rather large bottleneck sizes, with a dilution rate of 0.1–0.2, minimize the chance of loosing rare beneficial mutations (10), whereas smaller bottlenecks are capable of delaying hitchhiking mutations (11). In the study of Bruckbauer *et al.*, four biological replicates (IR9, IR10, IR11, IR12) of the parental strain *E. coli* K12 MG1655 were subjected to IR in 100 cycles of selection. The intensity applied in each cycle was aimed at killing 99% of the population. The appropriate irradiation stress was determined by the surviving fraction after the irradiation. Therefore, the bottleneck event did not randomly transfer bacteria to the next regrowth step, but favored those that had the best chance of survival. In every even round and after a regrowth phase, the whole population of each replicate was sequenced. Additionally, after 50 and 100 cycles of selection, five isolates were picked and sequenced (8,9). A graphical overview of the datasets used in this work is given in Figure 1. The analyzed bacteria are named using the substrain, selection cycle and, if applicable, number of isolate: IR9-100-1 is from the substrain IR9 after the 100th selection cycle, isolate 1. This nomenclature was introduced in the original publications (8,9).

Bruckbauer *et al.* showed that the adaption was specific to irradiation, as the response to other DNA damaging agents in general was not improved. The analysis of proteome oxidation showed for IR10-100-1 a great reduction of IR-induced hydroxylation, which was comparable to previously observed IR-induced hydroxylation levels in *Deinococcus radiodurans*,

Received: May 29, 2024. **Revised:** July 18, 2024. **Editorial Decision:** August 6, 2024. **Accepted:** August 8, 2024

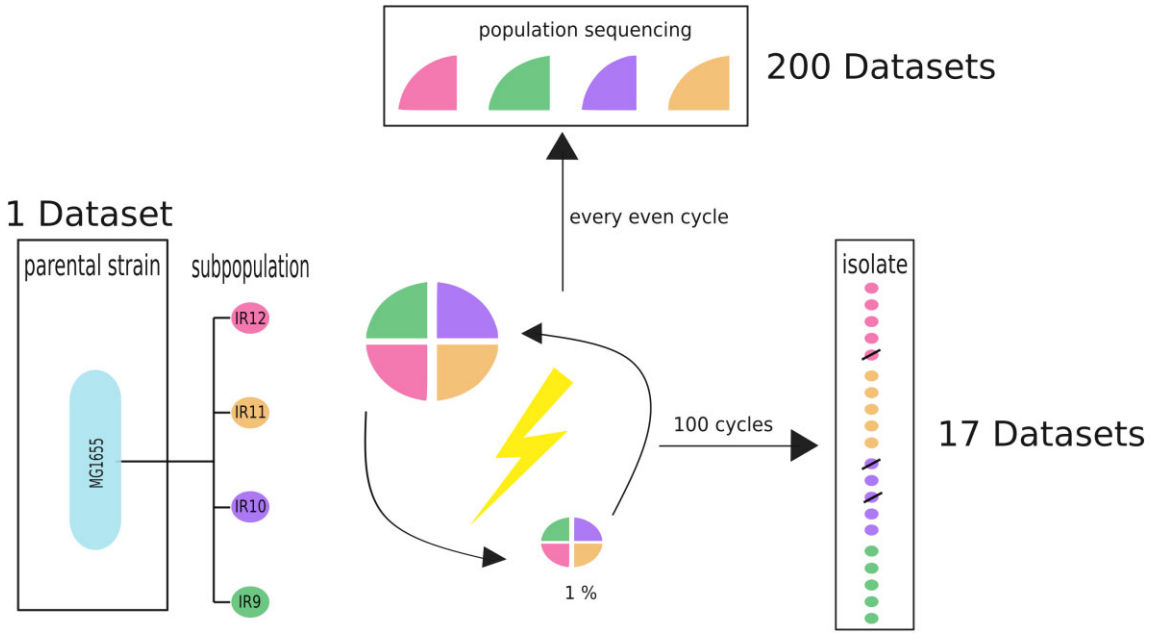

Figure 1. Overview of the experimental design. The parental strain *E. coli* K12 MG1655 was divided in four substrains that evolved independently. In each selection cycle, the irradiation intensity was adjusted so that 1% of the population survived. After every even round of selection, each subpopulation was sequenced. After the 100th selection cycle, for each subpopulation five isolates were picked and sequenced. Three out of the 20 datasets were excluded from further analysis, because they were generated using Illumina MiSeq sequencing instead of Illumina HiSeq (9).

a highly radioresistant microorganism (9). In the genetic analysis conducted by Bruckbauer *et al.* the focus was on variants with non-synonymous changes in proteins, and on the experimental verification of the impact on irradiation survival of promising candidate variants. Variants in the genes *recA*, *recD*, *recN*, *rpoC* and *rpoB* were shown to increase IR resistance (8,12).

Synonymous mutations in protein-coding genes change only the nucleotide composition, but not the corresponding amino acid. Therefore, the primary structure of the protein is not altered, and synonymous mutations are often considered to be 'silent' mutations with no effect (13). However, codon usage in several organisms, including *E. coli*, is correlated to gene expression levels and biased towards tRNAs that are most abundant and with optimal binding strength (14). Overall, synonymous mutations have an impact on fitness and possibly on protein functionality (13,15,16). The effects are currently proposed to be due to mRNA stability and translation efficiency (14–16). Additionally, regulatory processes on the level of translation, cotranslational folding, transcription and posttransciptional processes are affected (15,17).

Therefore, in this study, we adopt a comprehensive approach to characterize variants within the Bruckbauer dataset, encompassing both non-synonymous and synonymous mutations, along with an exploration of the pathways, cellular structures, and mechanisms implicated by these variants. Through this inclusive analysis, we unveil adaptation processes linked to ribosomes. Specifically, our investigation of the detection frequencies of variants in the 17 isolates after 100 cycles of selection highlights a single variant consistently present in the protein-coding sequences across all samples: *rpsH* c.294T > G, a variant that potentially impacts the regulation of the *spc* operon. The ubiquitous presence of this variant in all isolates strongly suggests its role as a beneficial mutation.

Materials and methods

Bioinformatic pipeline for obtaining variants from DNA sequencing files

The DNA sequences were downloaded from NCBI's Sequence Read Archive (SRA) using the SRA toolkit 2.11.2 (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software). Data files were stored as forward and reverse FASTQ files using the command *fastq-dump –split-files*.

The bioinformatic pipeline is adopted from Galaxy Project (18,19) and the generated script is provided in the Supplement ('DNAtoVariants.sh'). Script and all bash commands were executed with GNU bash version 5.0.17 (20). The reference genome for the alignment was *E. coli* K-12 substrain MG1655 (U00096.3), which was downloaded from NCBI and is consistent with the reference genome used in the original analysis by Bruckbauer et al. (8,9). Alignment was conducted with *Burrows–Wheeler Aligner (BWA) 0.7.1* (21). Post-processing steps for the mapped reads were performed with *SAMtools 1.14* and involved exclusion of unmapped reads or reads with an unmapped mate, and deduplication of the remaining reads (22). Variants were called with *freebayes v1.3.5* (23) and stored in a variant calling file (VCF). Found variants were left-aligned and normalized using *BCFtools 1.14* (22). This step ensures that the variants are represented in the same way they are stored in annotation databases. Variant annotation was performed using *SnpEff 5.0e* with the prebuild reference genome folder named *Escherichia_coli_str_k_12_substr_mg1655* (24). Descriptions of the variants were extracted from the *SnpEff* HTML summary file.

Validation of the applied bioinformatic pipeline

Our bioinformatic approach closely follows the pipeline employed by Bruckbauer *et al.* (9). The alignment method (BWA),

Table 1. Number and types of variants after 100 cycles of selection

Isolate	Total	SNP	Deletion	MNP	Insertion	Mixed
IR9-100-1	832	753	62	13	1	3
IR9-100-2	865	771	72	17	2	3
IR9-100-3	804	727	59	13	2	3
IR9-100-4	800	713	64	16	1	6
IR9-100-5	867	779	72	15	0	1
IR10-100-1	540	493	36	7	1	3
IR10-100-2	553	518	26	6	1	2
IR10-100-4	567	517	37	9	2	2
IR11-100-1	502	468	25	6	2	1
IR11-100-2	501	464	27	6	3	1
IR11-100-3	507	464	30	9	1	3
IR11-100-4	505	465	28	7	2	3
IR11-100-5	497	464	24	6	2	1
IR12-100-2	624	577	32	13	1	1
IR12-100-3	673	624	28	18	2	1
IR12-100-4	641	591	33	12	1	4
IR12-100-5	628	576	33	14	1	4

reference genome (MG1655 U00096.3), and annotation tool (*SnpEff*) were kept consistent. However, we utilized the most recent version of *SnpEff* (5.0e in this work, compared to 4.0e in Bruckbauer et al.). Notable modifications in data processing included the elimination of the downsampling step. We also substituted *freebayes* for the *callvariants.sh* tool from *BBMap* and integrated all called variants into the analysis. The exclusion of downsampling aimed to prevent the removal of reads with rare variants in regions of high sequencing depth, ensuring their detectability (25).

To validate our bioinformatic pipeline, we conducted a comparison of the number of SNPs identified in each isolate with the published SNPs by Bruckbauer et al. (Supplementary Table S3 in (9)). The results exhibited overall comparability, with discrepancies ranging from 0.19% (IR10-100-4) to 6.18% (IR11-100-5). In approximately 82% of the cases, our pipeline identified more SNPs. The exact numbers of SNPs, along with the percentage differences, are presented in Supplementary Table S1. Additionally, we compared the number of transitions, transversions, synonymous variants, stops gained and lost, and starts lost (see Supplementary Table S2). In each of these categories, our pipeline detected more (for 6 events) or fewer features (for 15 events). The most significant differences (%) were observed in the identification of rare events such as stop and start losses (up to three times per subpopulation).

We conducted a detailed comparison of allele frequencies, focusing on variants in ribosomal genes in IR9 throughout the entire selection process (see Supplementary Table S3). This comparison encompassed 47 allele frequencies between selection cycles 28 and 100, ranging from 5.2% to 61%. The called allele frequencies were consistent for 27% of the time (±0.5 %). In 36% of the cases, our pipeline called higher allele frequencies, and in 32% lower allele frequencies, with differences of 3.2% at most. Two instances of variants were exclusively detected by the original pipeline, both with allele frequencies just above the limit of detection (5.2% and 5.5%), suggesting that our pipeline's limit of detection might be slightly higher than 5%. Conversely, our pipeline successfully identified variants with allele frequencies starting at 5.4% and demonstrated consistency with the original pipeline. The downsampling step appeared to have no significant impact on the calling of allele frequencies. However, our pipeline detected three variants in the gene *rpsH* that were not identified by the original pipeline. These variants were frequently detected in IR9 with low allele frequencies, and in this context, the downsampling step might have led to the discarding of too many reads with the variant.

Overall, both bioinformatic pipelines yielded similar results. There is no clear evidence indicating that one pipeline is more sensitive than the other. Since it cannot be definitively determined which results better reflect the populations, the variants identified by our pipeline and the original can be considered equivalent.

Table 2. Characteristics of the gene sets $GS_{complete}$: this gene set encompasses all 5533 distinct variants identified across the 17 isolates from the four different subpopulations. These variants have an impact on 4490 genes, which either contain one or more of the variants or are indirectly influenced by a variant in a different gene. For instance, the annotation tool *SnpEff* considers all genes in the same operon to be affected. GS_{coding}: set of genes with a variant in the coding region in at least one of the isolates. $GS_{always(affected)}$: set of genes impacted by a variant in all 17 isolates. These genes either harbored a variant in the coding region or were indirectly influenced by a variant. The procedures for variant analysis leading to gene set creation are outlined in the Materials and methods section. Gene type information is extracted from the genes.txt file generated by *SnpEff*

	$GS_{complete}$	GS_{coding}	$GS_{always(affected)}$
Genes	4490	2452	492
RNA-coding genes	86	25	16
Pseudogenes	178	88	20
Unknown genes	5	1	-
Protein-coding genes	4221	2338	456

Table 3. Variants present in all isolates after 100 selection cycles. *: The *SnpEff* annotation is aligned with the forward strand, while the coding strand of *rpsH* is the complement strand. Therefore, 3466252 A → C induces a change from T to G at the 294th nucleotide, denoted as *rpsH* c.294T > G

Position	Reference → Alternate	Description
3446252	A → C *	rpsH 30S subunit
3781140	A → G	intergenic region between lldD and trmL
3781145	C → G	intergenic region between lldD and trmL
3781150	T → G	intergenic region between lldD and trmL
3781152	A → G	intergenic region between lldD and trmL
4070483	A → G	intergenic region between yihR and yihS

Variant and gene ontology (GO) analysis

During the annotation step with *SnpEff*, gene lists are generated and stored in .genes.txt files. These files contain genes directly affected (variants within the coding region) or indirectly affected (variants located outside the coding region) by one or more variants. The .genes.txt files from all isolates were combined to create a comprehensive list of genes affected by variants in any isolate. The analysis of .genes.txt files was conducted in *R 4.1.3* (26), excluding columns related to upstream/downstream variants and probable variant impact (high, low, moderate, modifier). For genes with variants within their coding region, the sum of the remaining columns was calculated. Genes with a row sum of zero were excluded, and the remaining genes were considered directly affected by at least one variant. This process was performed individually for each isolate, and the resulting gene lists from all 17 isolates were

combined to obtain the final list of genes with variants in the coding region after 100 rounds of selection.

To identify variants present in all samples, VCF files of all isolates containing normalized variants were compressed, indexed (*bgzip -c*), and merged using *bcftools concat*. Subsequently, the merged file was analyzed with the bash command *uniq -c -w 20*, which compares the first 20 characters of each line and lists unique variants along with their quantities. This ensures only the position, reference nucleotide(s), and alternate nucleotide(s) are considered for comparison, preventing multiple listings of identical variants due to different quality scores. The resulting output indicates the number of occurrences for each variant in the 17 isolates, facilitating the identification of variants present in all samples. Variants found in the resequencing data of the founder strain MG1655 were also excluded. The resulting file was then annotated using *SnpEff 5.0e*.

To identify key GO terms, we performed statistical overrepresentation tests using the online version of *PANTHER 17.0* (27) with Fisher's Exact Test and the false discovery rate (FDR) as correction factor.

Allele frequency, detection frequency and durability analysis

The DNA extractions after every even selection cycle were conducted from a mixed population of the evolved subpopulations (9). The allele frequency represents the proportion of a variant within the population and can be used to estimate its abundance. For example, an allele frequency of 0.5 indicates that 50 % of the population carries the variant. To calculate the allele frequency of a variant, the counts of the reference haplotype (RO) and the alternate haplotype (AO) were extracted from the annotated VCF file. The allele frequency was then computed by dividing the count of AO by the sum of AO and RO. Variants with an allele frequency under 5 % are below the detection limit of the method.

We investigated the duration for which ribosomal variants remained detectable throughout the selection process. Sequencing was performed every even round of selection, resulting in a total of 50 sequencing events for each lineage (9). The detection frequency is defined as the ratio of detection events to sequencing events. Since Illumina MiSeq sequencing was excluded from the analysis, the number of sequencing events varies between cell lines: 48 sequencing events for IR12, 49 for IR10, and 50 for both IR9 and IR11. The term 'durability' describes the number of rounds in which a specific variant remains detectable.

Results and discussion

Variants observed after 100 cycles of selection under the evolutionary pressure of irradiation

We re-analyzed a sequencing dataset published by Bruckbauer *et al.* (8,9) using an adapted bioinformatic pipeline. The change with the greatest impact is the omission of a downsampling step, which may lead to a higher probability of detecting rare variants (see Methods section for technical details). In each isolate, we found between 510 and 880 variants, which is very similar to the variants described by Bruckbauer et al. and the differences (see Supplementary Table S1–S3) lie within a range of analytical uncertainty, as both pipelines may call a few false-positive and false-negative variants (28). Variants appeared randomly in the whole chromosome (see Supplementary Figure S1). The different subpopulations showed distinct numbers of variants: IR9 had 800 and more variants in each isolate and was the lineage with the most variants, followed by subpopulations IR12 (650 ± 25), IR10 (553 ± 14), and IR11 (502 ± 5). The predominant type of variants observed were single nucleotide polymorphisms (SNPs), constituting approximately 90 % of all variants in each isolate. Deletions ranked as the second most frequent type, accounting for about 6 % of variants in each isolate. Additionally, other variants such as MNPs (multiple nucleotide polymorphisms), insertions, and mixed variants were identified, listed in descending order of frequency. A comprehensive breakdown of the number and types of variants is provided in Table 1. These findings are consistent with the original analysis of Bruckbauer *et al.* (4).

The effects of a variant can be further characterized based on the region (e.g. downstream, intergenic) and type (e.g. missense, stop gained). It is noteworthy that a single variant may manifest multiple effects. Selected effects and their relative occurrences are summarized in Supplementary Table S4, complemented by a visual representation in Figure 2. Approximately 90% of all variants predominantly manifest as up- or downstream alterations. Missense variants rank as the second most prevalent, constituting around 4%. Synonymous variants make up approximately 2%, while nonsense variants are present at about 0.2%. Although other regions and types of effects (e.g. frameshift, intergenic, stop lost) are observed, their incidence is comparatively lower. The detailed breakdown of variants for each isolate of the four cell lines is complemented by our study.

Additionally, we evaluated the ratio between transitions (AT $\rightarrow$ GC) and transversions (AT $\rightarrow$ CG/TA, GC $\rightarrow$ CG/TA). If the probability for each of the six possible events is uniform, the expected transition-to-transversion ratio would be 0.5 (30). In experimental studies, a transition-to-transversion ratio of 1.27 (56% transitions, 44% transversions) is observed in wild-type *E. coli* (29). In our analyses, the transition-to-transversion ratios extracted from *SnpEff* HTML annotation files are lower for most isolates. Overall, this finding is consistent with the original study. However, we analyzed the 17 isolates after the 100th selection cycle, whereas the previously published results refer to the whole population sequencing after round 100 (4). Transitions make up approximately 48% to 53% of the observed SNPs, while transversions account for a corresponding range of 52% to 47%. The only exceptions are IR11-100-3, which exhibits a transition-to-transversion ratio close to the wild-type (1.30), and IR10-100-4 and IR11-100-4, which both have a transition-to-transversion ratio of 1.39, corresponding to approximately 58% transitions and 42% transversions (see Supplementary Table S4). Environmental and nutritional stress are shown to have an impact on the mutational bias and the transition-to-transversion ratio in *E. coli*, but a particular gene-regulatory network has not yet been identified (31). The heightened prevalence of transversions might suggest a more pronounced impact on protein and gene expression modification. Transversions alter the amino acid composition of proteins with a greater likelihood compared to transitions. Additionally, they can influence the DNA structure, such as the minor groove width (32). The scrutinized sequences exhibit a G/C content of approximately 48%, slightly lower than the 50.8% expected for *E. coli* K12 MG1655 (33). A/T-rich se-

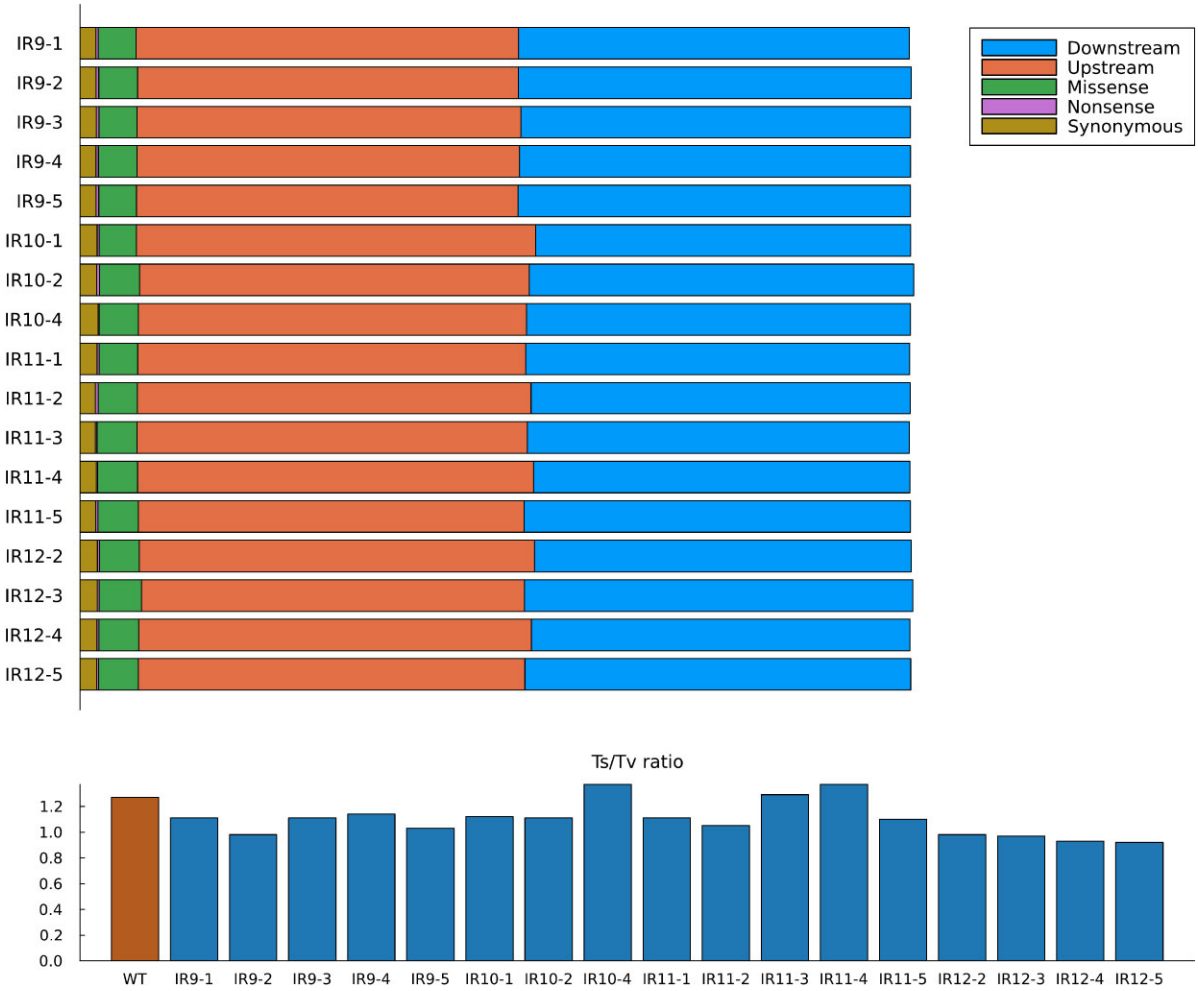

Figure 2. Relative occurrence of selected variant effects of the analyzed isolates after 100 cycles of selection are shown in the stacked bar chart. Most of the variants are located down- or upstream of a gene. Nonsense variants are very rare (approximately 0.2%). The ratio of transitions to transversions (Ts/Tv) are represented in the bar plot. The first (orange) bar is the Ts/Tv ratio of wildtype *E. coli* taken from Lee *et al.* (29). The Ts/Tv ratios are in most of the evolved isolates lower compared to the wildtype, but these variable seems to be very specific for the populations with differences within one subpopulation (e.g. IR10).

quences feature narrower minor grooves, while G/C-rich sequences have wider minor grooves (34). Alterations in DNA shape can hinder the binding of transcription factors and modulate the activity of regulatory elements (32,34). If the evolutionary selection bias toward IR resistance leans toward an increased A/T content, it might be attributed to the narrower minor grooves in the DNA. This could result in a more compact DNA structure and potentially reduce the occurrence of lesions during the irradiation process.

RNA-coding genes less affected by variants than protein-coding genes

In a new approach to explore the specific genes modified during the evolutionary process towards IR resistance, the variants were grouped into three different gene sets depending on the context of their appearance, see Table 2. These gene sets reveal that over 50% of the 4703 genes in *E. coli* are directly altered by a variant and about 95% of genes are affected by a variant. Despite the fact that 95% of the genome was affected by a variant, only about 40% of RNA-coding genes were affected (relative to 229 RNA-coding genes in the whole genome (35)). One explanation for this observation could be a tightly regulated system that is sensitive towards alterations. This is also supported by the fact that approximately 11% of the RNA-coding genes are altered, whereas about 50% of protein-coding genes and pseudogenes are changed by a variant. Further investigation of the RNA-coding genes, which were directly altered by a variant, may reveal important signalling pathways and interactions. It also underlines the importance of RNA-coding genes in the adaption process to a changing or challenging environment.

Gene Ontology (GO) analysis points towards adaptations in genes associated with translation and ribosomes

The generated gene sets underwent GO term overrepresentation tests to explore susceptibility to variants under irradiation as a selective pressure. For GS_{coding}, 2410 out of 2452 genes were analyzed, as 42 genes could not be mapped due to PANTHER limitations in handling small RNAs and some pseudogenes. Significant enrichments were observed across all three GO categories, ranging from 1.07 (cellular anatomical entity) to 1.29 (transmembrane transporter activity) fold enrichment, as illustrated in Figure 3. The overrepresented GO

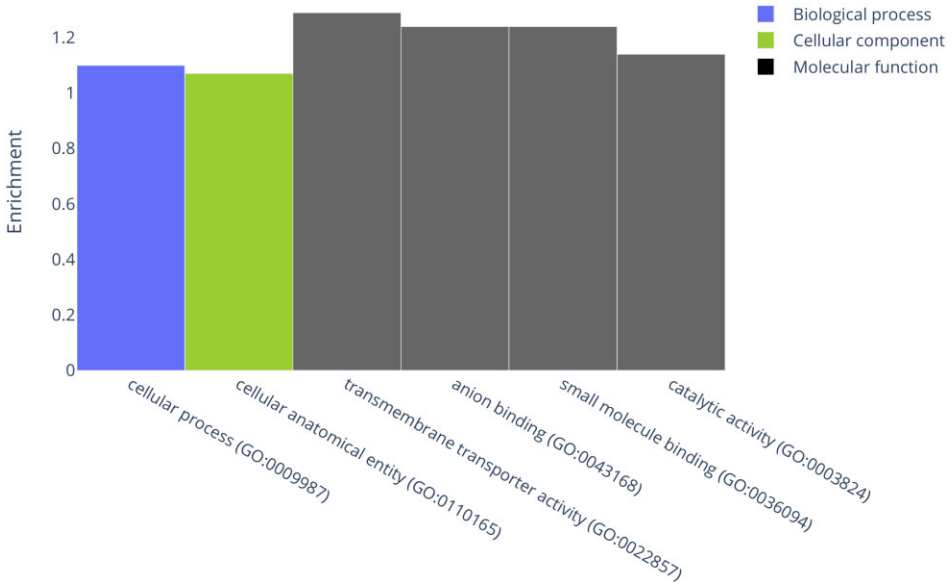

Figure 3. PANTHER overrepresentation analysis of GS_{coding}. Out of the 2452 genes assigned to this gene set, PANTHER uniquely mapped 2410 genes. The presented results include significant findings with *P*-values <0.05 and the most specialized child GO term for each root term.

terms 'cellular process', 'metabolic process', 'cellular anatomical entity' and 'catalytic activity' are direct children of their root terms. Ancestor charts for the overrepresented GO terms are presented in the Supplementary Figure S2. GS_{coding} encompasses all genes carrying a variant in a coding region in any of the isolates. While it has the potential to detect parallel adaptation pathways towards IR resistance, it did not yield insights into general adaptations due to the rather general and modestly enriched (1.07–1.29) GO terms.

The overrepresentation analysis for the gene set $GS_{always(affected)}$ examined 458 out of the 492 genes, excluding tRNAs, sRNAs, rRNAs, and pseudogenes. Significant enrichments were observed in all three GO categories, ranging from 3.2 ('ribosome assembly') to 4.52 ('cytosolic large ribosomal subunit') fold enrichment, as depicted in Figure 4. $GS_{always(affected)}$ is not restricted to genes directly affected by a variant, acknowledging the potential for gene expression modification by up- or downstream variants. However, it exclusively considers genes affected in every isolate, providing insights into critical adaptation processes regulated by specific genes. The overrepresented GO terms in $GS_{always(affected)}$ converge towards adaptations in translational and ribosomal genes. Ancestor charts for these terms are detailed in the Supplementary Figure S3, emphasizing their specificity compared to GO terms overrepresented in GS_{coding}.

G-quadruplex motif is affected by four intergenic variants

We investigated the presence of shared variants among the isolates, and the distribution of how frequently a single variant appeared among the 17 isolates is depicted in the Supplementary Figure S4. A total of 3464 variants occurred only once, with approximately 400 variants found in 2–5 samples. Given the selection of five isolates for each subpopulation, a decline in frequency after five occurrences was anticipated. The number of variants shared by six or more isolates ranged from zero to ten, with six variants being shared by all 17 isolates. In Table 3, the six variants present in all isolates are displayed, all of which, to our knowledge, have not been described before.

Five of these variants are in a predicted G-quadruplex motif region and four of the variants (genomic positions 3781140, 3781145, 3781150, 3781152) in the 5' UTR of the *trmL* gene (36). The motif stretches from the genomic position 3781135 to 3781155, ending 60 nucleotides away from the start codon. The variants at positions 3781140, 3781140 and 3781145 appear in all four subpopulations after the second round of selection with an allele frequency of about 8%. The variant at position 3781152 is detected for the first time after the second round of selection in IR10, after the fourth selection round in IR11 and after the eighth round in IR9 and IR12 (see Supplementary Figure S7). After the initial detection, the variants do not occur continuously. Remarkably, the detection frequencies are ranging from 73% (3781152, 3781150 in IR10) to 100% (3781140 in IR9 and IR10). Variants with similar frequencies are at position 3781120 (average 94%) and 3781133 (average 82%) in the conserved cytosine rich region 5' of the motif. As these variants are detected as SNPs and none are fixed in a population, the variants may have a similar biological effect and may not necessarily occur in the same individuum simultaneously. However, very rarely we observed the double variants n.3781150TGA > GGG (IR9 round 86 and 92, IR10 round 72, 76 and 82, IR11 round 78, IR12 round 86). Strikingly, all variants within the motif are changing a nucleotide to a guanine and in the conserved 5' region to a cytosine (see Supplementary Figure S7). We speculate that the G-quadruplex structure is disrupted by these variants as at least one of the three necessary loops is altered. This is supported by Kaplan *et al.*, who also suggest that mutations in the G-tract prevent the formation of the G-quadruplex (36). Additionally, structural analyses conducted by Kaplan *et al.* showed this particular G-quadruplex to be able to switch between G-quadruplex and hairpin-like topology (36). This switching could be impaired by the variants observed by us, and the region could therefore be structurally fixed.

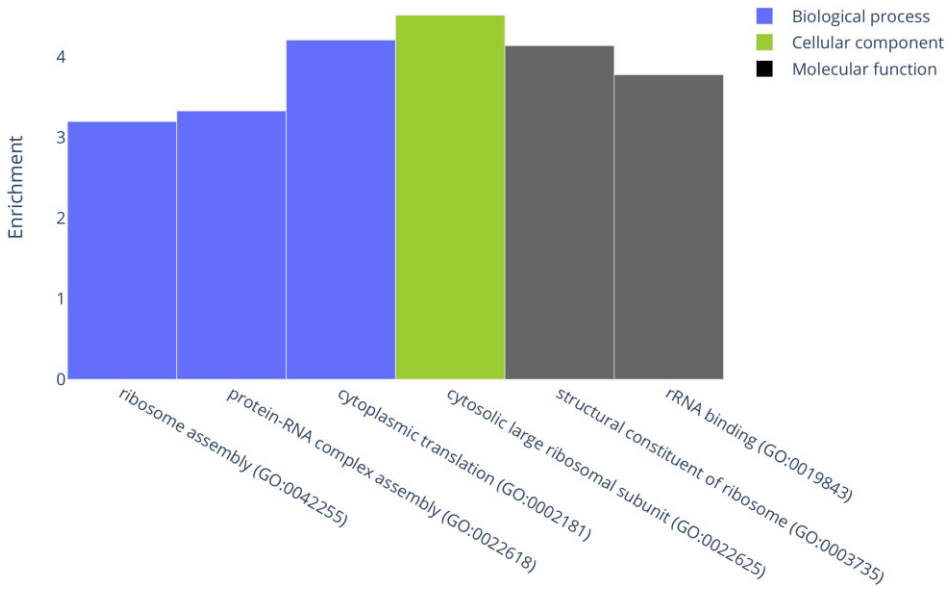

Figure 4. PANTHER overrepresentation analysis of the gene set $GS_{always(affected)}$. This gene set comprised 492 genes, of which PANTHER uniquely mapped 458 genes. The results presented include significant findings with *P*-values <0.05 and the most specialized child GO term for each root term.

Variant *rpsH* c.294T > G is the only variant found within a protein-coding region and shared across all isolates

Among the variants shared by all isolates, only one variant is located in a protein-coding region: 3466252 A → C, within the gene *rspH*. Since this gene is on the complement strand, it induces a change from T to G at the 294th nucleotide, denoted as *rpsH* c.294T > G. This alteration represents a synonymous mutation of the glycine codon (GGU → GGG) at position 98 of the amino acid chain of the 30S ribosomal protein S8. The S8 protein, an essential component of the small ribosomal subunit, plays a crucial role in subunit assembly and translational regulation of the *spc* operon (37,38). Consequently, a variant in this gene can impact the cell during subunit assembly and/or gene expression regulation.

Variant *rpsH* c.294T > G is uniquely and consistently displaying high detection and low allele frequencies across all subpopulations

Due to the anticipated beneficial effect of the *rpsH* variant (refer to Supplementary Information for the estimation of its random appearance in all datasets, including Figure S8), our interest turned to identifying the selection round when the variant first emerged in the subpopulations. To achieve this, we conducted an additional analysis of the whole-population sequencing data from earlier selection cycles. In three out of the four observed subpopulations (IR9, IR10, IR12), the *rpsH* c.294T > G variant emerged after two cycles of selection. In IR9, the variant remained consistently detectable after cycle 36, and in IR10 after cycle 48. For IR11 and IR12, *rpsH* c.294T > G became continuously detectable after cycle 78 and 74, respectively (refer to Supplementary Figure S5).

To gain insight into the frequency and dominance of variants in ribosomal proteins within the population, we compared the allele and detection frequencies of all variants in the protein-coding region of ribosomal proteins. In Figure 5, the average allele frequency is plotted against the detection frequency for all ribosomal proteins with at least one variant in each subpopulation. Most variants were identified at both low detection frequency and low allele frequency (bottom left quadrants in Figure 5). The variant *rpsH* c.294T > G, observed in all isolates after 100 cycles of selection, was also the only variant with consistently high detection and low allele frequencies in all subpopulations (Figure 5, top left quadrants). Interestingly, a second variant in *rpsH*, *rpsH* c.327T > G, consistently appeared across all subpopulations but with low allele and detection frequencies. This variant in *rpsH* introduces the same glycine codon change from GGU to GGG, akin to *rpsH* c.294T > G, at position 109 of the S8 amino acid chain. Furthermore, a third variant in *rpsH* that also introduces the same codon change, *rpsH* c.300T > G, was identified in all subpopulations with low allele frequency but relatively high detection frequency. These three *rpsH* variants were the exclusive variants in the protein-coding region of ribosomal proteins present in all four subpopulations.

In light of fluctuating allele frequencies, particularly when a variant failed to reach fixation (allele frequency = 1) in the population, we investigated the stability of *rpsH* allele frequencies in comparison to other ribosomal variants. In Figure 6, the allele frequencies of ribosomal variants, sorted by the number of detections, are presented. Notably, the variant of interest, *rpsH* c.294T > G, not only exhibits the highest number of detections in each subpopulation but also maintains a remarkably stable allele frequency. Moreover, *rpsH* c.300T > G stands out as the variant with the second-highest number of detections in IR9, IR11, and IR12. In all subpopulations, this variant was detected at least 20 times, consistently maintaining a highly stable allele frequency. In contrast, the allele frequencies of other ribosomal variants exhibited a broader range of variability across different sequencing events (see Figure 6).

Conclusion

Ionizing radiation poses a highly toxic stress to all organisms, and the mechanisms leading to IR-induced cell death

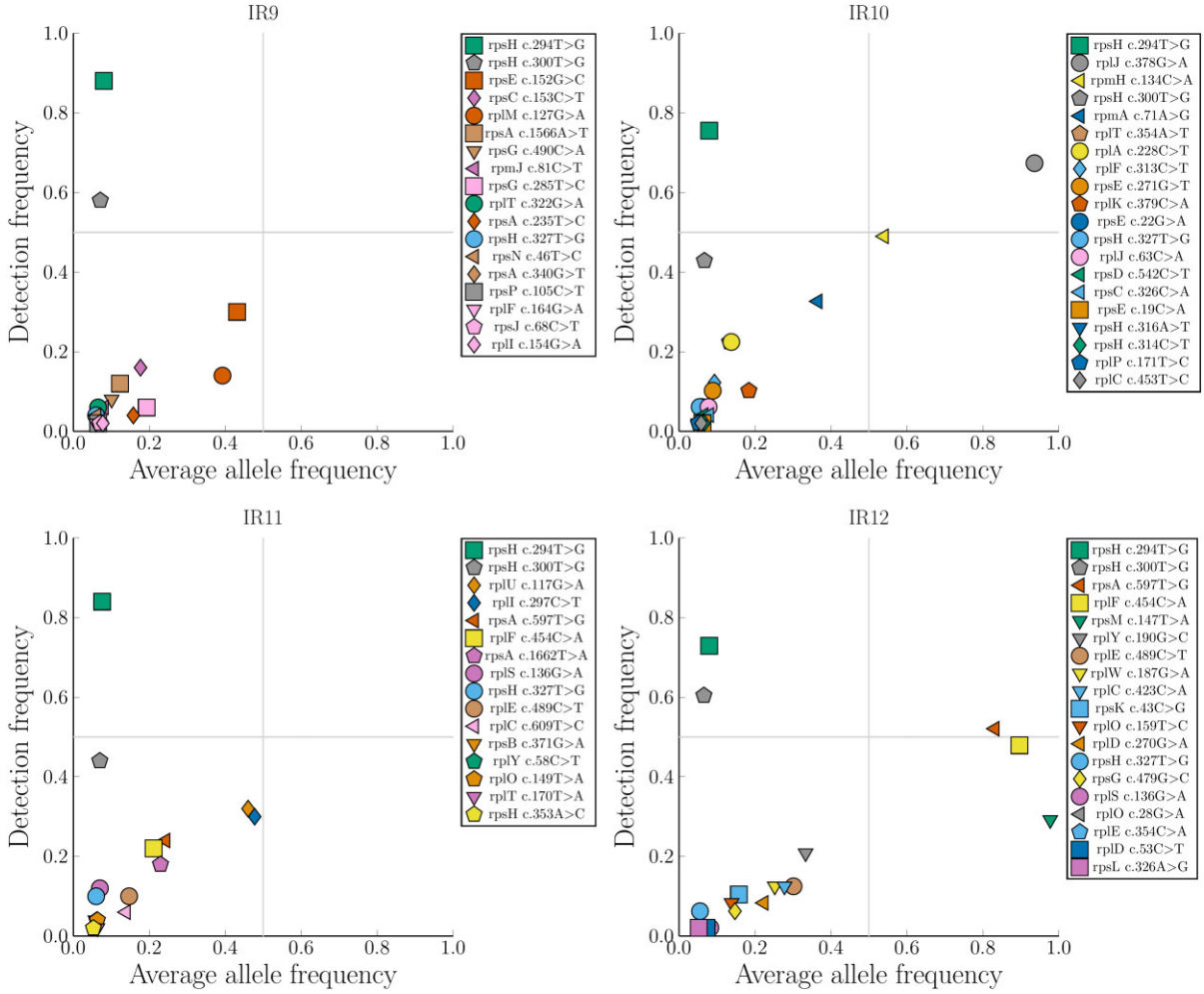

Figure 5. Detection and allele frequencies of ribosomal protein variants. Analysis of sequencing data involved every even round of 100 selection cycles, totaling 50 sequencing events in each lineage (9). Illumina HiSeq sequencing data were exclusively utilized to ensure high comparability, resulting in 50 events for IR9 and IR11, 49 for IR10, and 48 for IR12. The average allele frequency was computed as the mean of all observations, and the detection frequency was determined as the ratio of detection to sequencing events. Both IR10 and IR12 harbored variants with low detection frequencies but high allele frequencies (bottom right quadrants). In IR12, two variants (*rpsM* c.147T > A, green downwards pointing triangle, and *rplF* c.454C > A, yellow square) were initially detected between cycles 50 and 70, reaching an allele frequency of 1 (see Supplementary Figure S5). In IR10, a variant with low detection and high allele frequency (*rpmH* c.134C > A, yellow left pointing triangle) was first detected post-cycle 50 but exhibited more fluctuation, reaching a peak allele frequency of around 90% between cycles 66 and 76 (see Supplementary Figure S5). After 100 cycles of irradiation, all isolates exhibited synonymous mutations in *rpsH* c.294T > G (green square), *rpsH* c.300T > G (grey pentagon), and *rpsH* c.327T > G (blue circle). The *rpsH* c.294T > G variant consistently displayed the highest detection frequency across all subpopulations, with an average allele frequency of approximately 8%.

are still a subject of investigation. Given the medical applications of ionizing radiation, a comprehensive understanding of cellular reactions becomes crucial. This study analyzes DNA sequencing data from an evolution experiment conducted by Bruckbauer et al. (8,9). Their previous work focused on non-synonymous mutations, and promising variants associated with enhanced radiation resistance were identified and subjected to functional analysis. Experimental confirmation highlighted the contribution of variants in genes related to DNA repair (*recA*, *recD*, *recN*, *rpoC*, *rpoB*), ROS-related stress response (*sufD*), ATP synthesis (*atpA*), and nucleotide (*recJ*) and amino acid metabolism (*cadA*) to the IR-resistant phenotype (9).

In the presented study, we conducted a re-analysis of the sequencing datasets, focusing on all mutations occurring during the evolutionary experiment, including synonymous variants. Within the set of genes impacted by a variant in all 17 isolates across the four subpopulations, over-represented GO terms were exclusively related to ribosomes and translation. Exploring variants common to all isolates, we identified six variants emerging in all 17 samples, despite distinct evolutionary processes between the four subpopulations. Four of these variants were located in a predicted G-quadruplex motif with the ability to switch between G-quadruplex and hairpin structures and thus possibly act as a regulator (36). The biological effect could target the *trmL* gene, which is located 60 nucleotides downstream of the motif and translates to the tRNA (cytidine/uridine-2'-O)-ribose methyltransferase. To our knowledge, the characterization of the different structures and their functions has not yet been discussed in the literature. G-quadruplexes are also found in the *Deinococcus radiodurans* and human genomes, possibly protecting the genome from radiation damage (39).

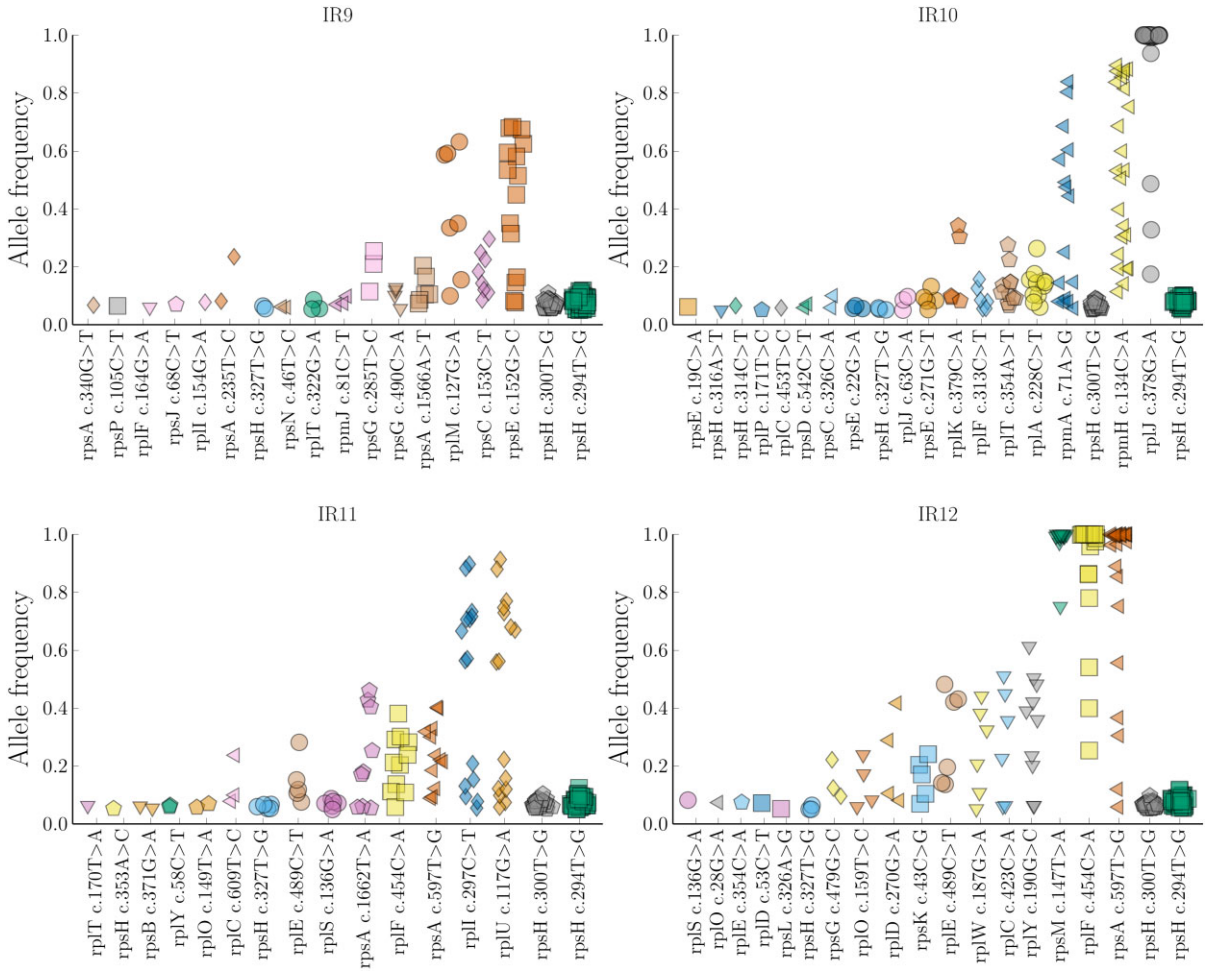

Figure 6. Variants detected in protein-coding region of ribosomal proteins. Similar to Figure 5, 50 sequencing events were scrutinized for IR9 and IR11, 49 for IR10, and 48 for IR12. For each ribosomal protein-coding variant and for each detection (i.e., the variant is identified in a particular sequencing event), the allele frequency is reported. Variants are sorted based on the number of detections. Consistently across all subpopulations, the *rpsH* c.294T > G variant exhibited the highest number of detections. The allele frequency, approximately 8%, remained remarkably stable in all subpopulations. The limit of detection was set at 5%.

Among the variants found in all isolates, only one variant was located in a protein-coding region: the synonymous variant *rpsH* c.294T > G, leading to a GGU to GGG codon change in the mRNA coding for ribosomal protein S8. This variant appeared early in the evolutionary process, with subpopulations IR9, IR10, and IR12 showing its presence after two cycles, and IR11 after four selection cycles. Throughout the evolutionary process, this variant exhibited consistent behavior in all four subpopulations, characterized by a stable low allele frequency and detectability in at least 35 instances. Notably, its behavior differed significantly from variants in other genes encoding ribosomal proteins. Two additional synonymous variants, *rpsH* c.300T > G (S8 Gly100) and *rpsH* c.327T > G (S8 Gly109), both inducing a GGU to GGG glycine codon change, emerged with lower detection and allele frequencies in all subpopulations.

The variant *rpsH* c.294T > G emerged early in the adaptation process and remained conserved in all isolates after 100 cycles of selection. Despite a slight allele frequency variation from 7.6% to 8.5%, the consistent presence suggests a potential beneficial impact on radiation stress. Variants found across all replicates are typically either situated within highly variable genomic regions or confer a beneficial impact. In the case of ribosomal genes, highly variable regions are unlikely. Analyses of the variation in core genes across different *E. coli* genomes demonstrate the conservation of the majority of genes associated with translation, ribosomal structures, and biogenesis. Moreover, gene clusters related to cell cycle control, cell division, and metabolism are also conserved (40). Consequently, the question arises as to whether and how the variant in *rpsH* might exert a beneficial effect.

The variant *rpsH* c.294T > G represents a synonymous variant, regarding the amino acid at position 98 in the 30S ribosomal protein S8 (S8 Gly98). In *E. coli*, *rpsH* is part of the *spc* operon, which includes genes for 11 ribosomal proteins and secY, a secretion protein (41). The small subunit of *E. coli* ribosomes comprises one rRNA and 21 proteins (41), playing a role in mRNA and initiation factor binding, as well as tRNA selection (37,42). S8, an essential gene in *E. coli* (43), is crucial for the small subunit's assembly and translational regulation of the *spc* operon (37,38). S8 Gly98 resides in a highly conserved region of the protein and is part of the translational regulator of *spc* operon expression. The binding site of S8 to the 16S rRNA in the ribosome is biochemically similar to the *spc* mRNA binding site (44). Excess S8 proteins bind to

spc operon mRNA, repressing the translation of all ribosomal proteins in that mRNA (44).

We cannot speculate on a mechanistic explanation for the impact of the variants on the regulation of the *spc* operon. Firstly, the mechanism of translational repression through the formation of the S8/*spc* mRNA complex remains poorly understood (44). Secondly, the positions 98 and 100 of S8 do not engage in interactions with mRNA, as indicated by the structure presented in Supplementary Figure S6 (44,45). The mRNA interaction site is situated in the third cistron, which encodes the ribosomal protein L5, of the *spc* operon. Since the S8 protein is encoded by the fifth cistron (44), none of the *rpsH* variants alter the mRNA site involved in translational repression.

Furthermore, due to the diverse genetic background in the different isolates, it is unlikely that the variant is a hitchhiking mutation. For a closely linked allele to be considered a hitchhiking mutation, it would need to be present in all isolates and evolve before or at a similar time (46). The *rpsH* variants are the earliest to arise in all isolates and occur without other variants in proximity. The small bottlenecks during the evolutionary experiment, which proceeds with the best-adapted 1% of the subpopulation (9), are also well-suited to minimize the chance for hitchhiking mutations (11).

Codon selection can impact mRNA translation efficiency, and the GGG codon is known to be translated more slowly than the GGU codon (47,48). While a single codon's effect on the translation dynamics of S8 mRNA might be marginal, it could influence translation fidelity. The higher relevance of translation accuracy compared to translation speed is further supported by the fact that the rarest and slowest glycine codon GGA (47,48) is not used as a synonymous substitution. An increased translational accuracy by a single substitution is already shown in archaea and eukaryotes (49). Mordret *et al.* demonstrated that the GGU glycine codon is prone to misreading errors, leading to amino acid substitutions, particularly aspartic acid, serine, and asparagine. A serine substitution at position 98 of the S8 protein results in a fast-growing phenotype due to defective *spc* operon regulation (38). Although the impact of substituting glycine with aspartic acid or asparagine is not experimentally validated, the larger molecular and structural differences suggest potential impairment in *spc* operon regulation. In contrast, the GGG codon of the variant is translated to glycine with higher fidelity; even compared to the GGA codon, which allows mispairing with glutamic acid (50). Two other *rpsH* variants (c.300T > G and c.327T > G) also involve synonymous glycine codon changes (GGU → GGG) at positions S8 Gly100 and S8 Gly109, both in conserved regions of the protein. Substitution of S8 Gly109 by serine leads to a regulation-defective phenotype (38), while the effect of S8 Gly100 substitution remains unexplored. These findings suggest that *E. coli* with wild-type S8 might have defective *spc* operon regulation after irradiation due to increased mispairing of amino acids under stress (51), leading to faster cell growth. With the GGG-glycine codon in the S8 variants, *spc* operon regulation would be maintained, resulting in reduced cell growth compared to the wild type. Reducing the translation of *rpsH* is already associated with depleted cell growth (52). Additionally, ribosome silencing and inactivation are known, but not fully understood, stress responses (51). The observed irradiation protection could be attributed to two aspects: slower-growing cells may be less susceptible to DNA damage (53), and a slower cell cycle could potentially improve DNA repair and post-irradiation recovery. This findings can also contribute to an understanding of slow-cycling subpopulations, which are known to exist in bacteria, yeast and human cells and are associated with resistance mechanisms (54).

The observed low allele frequency of the variants can be attributed to their assumed reduced growth rate. Prior to sequencing, surviving bacteria underwent two rounds of overnight cultures in a nutrient-rich Luria-Bertani broth (9). In this non-irradiated environment, S8 variants with slower growth would be outcompeted by bacteria containing wild-type S8. To obtain an accurate genomic profile of the surviving bacteria, immediate post-irradiation sequencing would be necessary (55). However, this would require cell sorting to distinguish between surviving and dead cells. Such an approach would enable sequencing of both fractions, facilitating a detailed comparison.

These considerations suggest a potential contribution to radiation resistance through translational control of ribosomal operons. The early emergence of these adaptations implies that this could be an initial mechanism for surviving environmental stress. However, mutations conferring an equal contribution to the irradiation-resistant phenotype would be favored due to fitness advantages. The described variant could be viewed as a potential mutation that 'rescues' the population, enabling the subsequent emergence of true resistance mechanisms (56). Conversely, an accelerated cell cycle might contribute to an irradiation-sensitive phenotype. This insight could be valuable when assessing genomic variants in irradiation-sensitive populations. Our results are consistent with an overrepresentation of ribosomal pathways in human breast cancer cells, as well as reduced expression of ribosomal proteins in breast cancer cells with induced radioresistance (57). In addition, G-quadruplex structures are associated with modulating radiation protection in humans (39), and interact with ROS induced by radiation (58). If these findings prove applicable to other organisms, they may contribute to understanding clinical radiation sensitivity and, ultimately, improving radiotherapy.

Data availability

The SRR numbers for all datasets used in this analysis are listed in the Supplemental file. The SRA numbers are stated in the Supplemental Data Sheet 4 of the original publication from Bruckbauer *et al.* (9).

Supplementary data

Supplementary Data are available at NARGAB Online.

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 437345987. Katharina Stemwedel was supported by the Exposé-Scholarship from the Leibniz University Hannover and is supported by the 'Dissertation Plus' scholarship from the Claussen-Simon-Stiftung. Nadin Haase is supported by the Add-on Fellowship of the Joachim Herz Foundation.

Plots were produced using the julia programming language (59) and `Plots.jl` (60). We would like to thank the anonymous reviewers for their valuable suggestions and constructive feedback.

Funding

Deutsche Forschungsgemeinschaft [437345987]; Claussen-Simon-Stiftung; Gottfried Wilhelm Leibniz Universität Hannover; Joachim Herz Stiftung.

Conflict of interest statement

None declared.

References

1. Daly,M.J. (2009) A new perspective on radiation resistance based on Deinococcus radiodurans. *Nat. Rev. Microbiol.*, **7**, 237–245.
2. Krisko,A. and Radman,M. (2010) Protein damage and death by radiation in *Escherichia coli* and *Deinococcus radiodurans*. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 14373–14377.
3. Byrne,R.T., Chen,S.H., Wood,E.A., Cabot,E.L. and Cox,M.M. (2014) *Escherichia coli* Genes and Pathways Involved in Surviving Extreme Exposure to Ionizing Radiation. *J. Bacteriol.*, **196**, 3534–3545.
4. Bruckbauer,S.T., Minkoff,B.B., Yu,D., Cryns,V.L., Cox,M.M. and Sussman,M.R. (2020) Ionizing radiation-induced proteomic oxidation in *Escherichia coli*. *Mol. Cell. Proteom.*, **19**, 1375–1395.
5. Cox,M.M. and Battista,J.R. (2005) *Deinococcus radiodurans*—the consummate survivor. *Nat. Rev. Microbiol.*, **3**, 882–892.
6. Bogdanova,N.V., Jguburia,N., Ramachandran,D., Nischik,N., Stemwedel,K., Stamm,G., Werncke,T., Wacker,F., Dörk,T. and Christiansen,H. (2021) Persistent DNA double-strand breaks after repeated diagnostic CT scans in breast epithelial cells and lymphocytes. *Front. Oncol.*, **11**, 634389.
7. Le Caër,S. (2011) Water radiolysis: influence of oxide surfaces on H2 production under ionizing radiation. *Water*, **3**, 235–253.
8. Bruckbauer,S.T., Trimarco,J.D., Martin,J., Bushnell,B., Senn,K.A., Schackwitz,W., Lipzen,A., Blow,M., Wood,E.A., Culberson,W.S., *et al.* (2019) Experimental evolution of extreme resistance to ionizing radiation in *EscherichiaColi* after 50 cycles of selection. *J. Bacteriol.*, **201**, e00784-18.
9. Bruckbauer,S.T., Martin,J., Minkoff,B.B., Veling,M.T., Lancaster,I., Liu,J., Trimarco,J.D., Bushnell,B., Lipzen,A., Wood,E.A., *et al.* (2020) Physiology of highly radioresistant *Escherichia Coli* after experimental evolution for 100 cycles of selection. *Front. Microbiol.*, **11**, 582590.
10. Wahl,L.M., Gerrish,P.J. and Saika-Voivod,I. (2002) Evaluating the impact of population bottlenecks in experimental evolution. *Genetics*, **162**, 961–971.
11. Raynes,Y., Halstead,A.L. and Sniegowski,P.D. (2014) The effect of population bottlenecks on mutation rate evolution in asexual populations. *J. Evol. Biol.*, **27**, 161–169.
12. Harris,D.R., Pollock,S.V., Wood,E.A., Goiffon,R.J., Klingele,A.J., Cabot,E.L., Schackwitz,W., Martin,J., Eggington,J., Durfee,T.J., *et al.* (2009) Directed evolution of ionizing radiation resistance in *Escherichia coli*. *J. Bacteriol.*, **191**, 5240–5252.
13. Shen,X., Song,S., Li,C. and Zhang,J. (2022) Synonymous mutations in representative yeast genes are mostly strongly non-neutral. *Nature*, **606**, 725–731.
14. Stoletzki,N. and Eyre-Walker,A. (2007) Synonymous codon usage in *Escherichia coli*: selection for translational accuracy. *Mol. Biol. Evol.*, **24**, 374–381.
15. Liu,Y., Yang,Q. and Zhao,F. (2021) Synonymous but not silent: the codon usage code for gene expression and protein folding. *Annu. Rev. Biochem.*, **90**, 375–401.
16. Kristofich,J.C., Morgenthaler,A.B., Kinney,W.R., Ebmeier,C.C., Snyder,D.J., Old,W.M., Cooper,V.S. and Copley,S.D. (2018) Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLoS Genet.*, **14**, 1–25.
17. Bryant,J., Chewapreecha,C. and Bentley,S.D. (2012) Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. *Future Microbiol.*, **7**, 1283–1296.
18. Maier,W., Batut,B., Houwaart,T., Erxleben,A. and Grüning,B. (2021) Exome sequencing data analysis for diagnosing a genetic disease (Galaxy Training Materials). https://training.galaxyproject.org/training-material/topics/variant-analysis/tutorials/exome-seq/tutorial.html (29 September 2022, date last accessed).
19. Batut,B., Hiltemann,S., Bagnacani,A., Baker,D., Bhardwaj,V., Blank,C., Bretaudeau,A., Brillet-Guéguen,L., Čech,M., Chilton,J., *et al.* (2018) Community-Driven Data Analysis Training for Biology. *Cell Systems*, **6**, 752–758.
20. GNU,P. (2007) Free Software Foundation. Bash (5.0.17) [Unix shell program].
21. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, **25**, 1754–1760.
22. Danecek,P., Bonfield,J.K., Liddle,J., Marshall,J., Ohan,V., Pollard,M.O., Whitwham,A., Keane,T., McCarthy,S.A., Davies,R.M., *et al.* (2021) Twelve years of SAMtools and BCFtools. *GigaScience*, **10**, giab008.
23. Garrison,E. and Marth,G. (2012) Haplotype-based variant detection from short-read sequencing. arXiv doi: https://arxiv.org/abs/1207.3907, 20 July 2012, preprint: not peer reviewed.
24. Cingolani,P., Platts,A., Wang,L.L., Coon,M., Nguyen,T., Wang,L., Land,S.J., Lu,X. and Ruden,D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain W1118; Iso-2; Iso-3. *Fly*, **6**, 80–92.
25. Github (2019) Picard toolkit. https://broadinstitute.github.io/picard/ (18 July 2024, date last accessed).
26. R Core Team. (2022) In: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.
27. Mi,H., Muruganujan,A., Huang,X., Ebert,D., Mills,C., Guo,X. and Thomas,P.D. (2019) Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.*, **14**, 703–721.
28. Bush,S.J. (2021) Generalizable characteristics of false-positive bacterial variant calls. *Microbial. Genom.*, **7**, 000615.
29. Lee,H., Popodi,E., Tang,H. and Foster,P.L. (2012) Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E2774–E2783.
30. Wielgoss,S., Barrick,J.E., Tenaillon,O., Cruveiller,S., Chane-Woon-Ming,B., Médigue,C., Lenski,R.E. and Schneider,D. (2011) Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3 Genes Genomes Genetics*, **1**, 183–186.
31. Maharjan,R.P. and Ferenci,T. (2017) A shifting mutational landscape in 6 nutritional states: stress-induced mutagenesis as a series of distinct stress input–mutation output relationships. *PLoS Biol.*, **15**, e2001477.
32. Guo,C., McDowell,I.C., Nodzenski,M., Scholtens,D.M., Allen,A.S., Lowe,W.L. and Reddy,T.E. (2017) Transversions have larger regulatory effects than transitions. *Bmc Genomics*, **18**, 394.
33. Milo,R., Jorgensen,P., Moran,U., Weber,G.M. and Springer,M. (2010) BioNumbers - the database of key numbers in molecular and cell biology. *Nucleic Acids Res.*, **38**, D750–D753.
34. Hancock,S.P., Ghane,T., Cascio,D., Rohs,R., Di Felice,R. and Johnson,R.C. (2013) Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. *Nucleic Acids Res.*, **41**, 6750–6760.
35. Keseler,I.M., Gama-Castro,S., Mackie,A., Billington,R., Bonavides-Martínez,C., Caspi,R., Kothari,A., Krummenacker,M.,

Midford,P.E., Muñiz-Rascado,L., *et al.* (2021) The EcoCyc Database in 2021. *Front. Microbiol.*, **12**, 711077.
36. Kaplan,O.I., Berber,B., Hekim,N. and Doluca,O. (2016) G-quadruplex prediction in E. coli genome reveals a conserved putative G-quadruplex-Hairpin-Duplex switch. *Nucleic Acids Res.*, **44**, 9083–9095.
37. Culver,G.M. (2003) Assembly of the 30S ribosomal subunit. *Biopolymers*, **68**, 234–249.
38. Wower,I., Kowaleski,M.P., Sears,L.E. and Zimmermann,R.A. (1992) Mutagenesis of ribosomal protein S8 from *Escherichia coli*: defects in regulation of the Spc Operon. *J. Bacteriol.*, **174**, 1213–1221.
39. Kumari,N. and Raghavan,S.C. (2021) G-quadruplex DNA structures and their relevance in radioprotection. *Biochim. Biophys. Acta Gen. Subj.*, **1865**, 129857.
40. Kaas,R.S., Friis,C., Ussery,D.W. and Aarestrup,F.M. (2012) Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. *BMC genomics*, **13**, 577.
41. Zengel,J.M. and Lindahl,L. (1994) Diverse mechanisms for regulating ribosomal protein synthesis in *Escherichia coli*. In: *Progress in Nucleic Acid Research and Molecular Biology*. Vol. **47**, Academic Press, pp. 331–370.
42. Lafontaine,D.L. and Tollervey,D. (2001) The function and synthesis of ribosomes. *Nat. Rev. Mol. Cell Biol.*, **2**, 514–520.
43. Goodall,E.C.A., Robinson,A., Johnston,I.G., Jabbari,S., Turner,K.A., Cunningham,A.F., Lund,P.A., Cole,J.A. and Henderson,I.R. (2018) The essential genome of *Escherichia coli* K-12. *mBio*, **9**, e02096-17.
44. Merianos,H.J., Wang,J. and Moore,P.B. (2004) The structure of a ribosomal protein S8/spc operon mRNA complex. *RNA*, **10**, 954–964.
45. Merianos,H.J., Wang,J. and Moore,P.B. (2004) The structure of a ribosomal protein S8/spc operon mRNA complex. *RNA*, **10**, 954–964.
46. Stephan,W. (2021) The classical hitchhiking model with continuous mutational pressure and purifying selection. *Ecol. Evol.*, **11**, 15896–15904.
47. Rudorf,S., Thommen,M., Rodnina,M.V. and Lipowsky,R. (2014) Deducing the kinetics of protein synthesis in vivo from the transition rates measured In vitro. *PLoS Comput. Biol.*, **10**, e1003909.
48. Rudorf,S. and Lipowsky,R. (2015) Protein synthesis in E. coli: dependence of codon-specific elongation on tRNA concentration and codon usage. *PLoS One*, **10**, e0134994.
49. Martinez-Miguel,V.E., Lujan,C., Espie-Caullet,T., Martinez-Martinez,D., Moore,S., Backes,C., Gonzalez,S., Galimov,E.R., Brown,A.E.X., Halic,M., *et al.* (2021) Increased fidelity of protein synthesis extends lifespan. *Cell Metab.*, **33**, 2288–2300.
50. Mordret,E., Dahan,O., Asraf,O., Rak,R., Yehonadav,A., Barnabas,G.D., Cox,J., Geiger,T., Lindner,A.B. and Pilpel,Y. (2019) Systematic detection of amino acid substitutions in proteomes reveals mechanistic basis of ribosome errors and selection for translation fidelity. *Mol. Cell*, **75**, 427–441.
51. Njenga,R., Boele,J., Öztürk,Y. and Koch,H.-G. (2023) Coping with stress: how bacteria fine-tune protein synthesis and protein transport. *J. Biol. Chem.*, **299**, 105163.
52. Popella,L., Jung,J., Do,P.T., Hayward,R.J., Barquist,L. and Vogel,J. (2022) Comprehensive analysis of PNA-based antisense antibiotics targeting various essential genes in uropathogenic Escherichia coli. *Nucleic Acids Res.*, **50**, 6435–6452.
53. Alper,T. and Gillies,N.E. (1958) 'Restoration' of Escherichia coli strain B after irradiation: its dependence on suboptimal growth conditions. *Microbiology*, **18**, 461–472.
54. Min,M. and Spencer,S.L. (2019) Spontaneously slow-cycling subpopulations of human cells originate from activation of stress-response pathways. *PLoS Biol.*, **17**, e3000178.
55. Bruckbauer,S. and Cox,M. (2021) Experimental evolution of extremophile resistance to ionizing radiation. *Trends Genet.*, **37**, 830–845.
56. Kayser,J., Schreck,C.F., Gralka,M., Fusco,D. and Hallatschek,O. (2019) Collective motion conceals fitness differences in crowded cellular populations. *Nat. Ecol. Evol.*, **3**, 125.
57. Ying,Y., Bian,L., Meng,Y., Zhang,M., Yao,Y., Bo,F. and Li,D. (2023) Comparative proteomic analysis of irradiation-induced radioresistant breast cancer cells using label-free quantitation. *FBL*, **28**, 244.
58. Wu,S., Jiang,L., Lei,L., Fu,C., Huang,J., Hu,Y., Dong,Y., Chen,J. and Zeng,Q. (2023) Crosstalk between G-quadruplex and ROS. *Cell Death Dis.*, **14**, 37.
59. Bezanson,J., Edelman,A., Karpinski,S. and Shah,V.B. (2017) Julia: a fresh approach to numerical computing. *SIAM Rev.*, **59**, 65–98.
60. Christ,S., Schwabeneder,D., Rackauckas,C., Borregaard,M.K. and Breloff,T. (2023) Plots.jl – a user extendable plotting API for the Julia programming language. *J. Open Res. Softw.*, **11**, 5.

Received: May 29, 2024. **Revised:** July 18, 2024. **Editorial Decision:** August 6, 2024. **Accepted:** August 8, 2024