



## OPEN Raiding nature's genetic toolbox for UV-C resistance by functional metagenomics

Garrett A. Roberts Kingman<sup>1</sup>, Justin L. Kipness<sup>2</sup> & Lynn J. Rothschild<sup>3</sup>✉

As we assess the habitability of other worlds, we are limited by being able to only study terrestrial life adapted to terrestrial conditions. The environments found on Earth, though tremendously diverse, do not approach the multitude of potentially habitable environments beyond Earth, and so limited terrestrial adaptive capabilities tell us little about the fundamental biochemical boundaries of life. One approach to this problem is to use experimental laboratory evolution to adapt microbes to these novel environmental conditions. This approach can be dramatically improved through functional metagenomics, large-scale introduction of foreign genetic material to screen for phenotypes in a new host organism. This takes advantage of Earth's immense biological diversity with high-throughput screening for genetic tools that can facilitate adaptation. We address a key gap in functional metagenomics work by exploring the impact of the experimental parameters chosen for functional metagenomics libraries. Experimental design dictates both fragment size and copy number, and we show that both can have outsized effects on the resultant phenotypes in non-intuitive ways. These results highlight the potential of functional metagenomics for adapting life rapidly to challenging new environments, with important implications in both astrobiology and bioindustry, while also emphasizing the impacts of decisions in experimental design.

A fundamental unanswered question of biology is “What are the environmental limits of life's biochemistry and what adaptations help it approach these limits?” This question has critical applications ranging from astrobiology, where the answers can help us target our search for life beyond Earth, to bioindustry, where the answers can enable use of biological tools in chemically-driven environments that are biologically challenging. Notably, life exists nearly everywhere on Earth where liquid water is present<sup>1,2</sup>, which is at once a testament to its incredible evolutionary adaptability and a severe impediment to our attempts to understand its limits, as its limits are not readily apparent.

Many studies have examined the environmental limits of extant extremophiles<sup>3</sup> and discovered incredible tolerance to heat<sup>4</sup>, cold<sup>5</sup>, salt<sup>6</sup>, nuclear<sup>7,8</sup> and ultraviolet radiation<sup>9,10</sup>, and numerous additional stressors. However, such work, while exceptionally valuable, is fundamentally limited by only being able to study extant organisms that have evolved under environmental conditions found on the modern Earth (or have retained adaptations to earlier conditions). Therefore, their limits likely reflect the limited conditions of Earth rather than true biological constraints. Candidate locales for potential life beyond Earth include Martian perchlorate brines<sup>11</sup> and the clouds of Venus<sup>12</sup>, both of which differ dramatically from everywhere on modern Earth, leaving us ignorant about their compatibility with life.

Similarly, many bioindustrial processes are also limited by tolerance to their imposed environments, such as chemicals introduced for biomass pretreatment in biofuel production from lignins<sup>13</sup> and by the toxicity of the produced molecules<sup>14,15</sup>. Others seek to increase performance at either low temperatures for maximizing energy efficiency<sup>16</sup> or high temperatures for maximizing reaction rates<sup>17</sup>. Large scale industrial biomining, though efficient, is limited in its deployment by the environmental tolerances of key microbes<sup>18</sup>.

Additionally, many of the currently known limits to life reflect the bounds of incidental tolerance to other stressors, rather than the result of direct evolutionary pressures. For example, the most perchlorate-tolerant organisms currently known can grow in concentrations in excess of 1 M<sup>19,20</sup>, despite the most concentrated perchlorate deposits known on Earth barely reaching parts per million<sup>21–23</sup>. Similarly, *Deinococcus radiodurans* can withstand 5,000 Gy of ionizing radiation with no discernible effect on viability<sup>24</sup> even though the most naturally radioactive locale on Earth yields exposures of only 0.4 Gy per year<sup>25</sup>. Thus, these extreme resistances are likely incidental consequences of tolerance to salinity and desiccation<sup>26</sup>, respectively. Other known limits,

<sup>1</sup>NASA Postdoctoral Program in Astrobiology, Ames Research Center, Moffett Field, CA, USA. <sup>2</sup>Department of Cell Biology & Biochemistry, Brown University, Providence, RI, USA. <sup>3</sup>NASA Ames Research Center, Planetary Systems Branch, Moffett Field, CA, USA. ✉email: lynn.j.rothschild@nasa.gov

including low temperature, pressure, and UV-C radiation (200–280 nm, which is considered fully blocked by the modern Earth's atmosphere<sup>27</sup>) may likewise not represent the limit reached by evolutionary pressures. This suggests that in the presence of direct evolutionary pressures, these limits which reflect incidental tolerance could be further extended. Nonetheless, the genetic tools evolved to address these related environmental challenges provide an invaluable starting point if we can figure out how to identify and make use of them.

When studying these limits, even in a one-dimensional manner, for astrobiology or industrial applications, we face two fundamental limitations. First, we don't know if this is a true limit, even using organisms fundamentally similar to our own, and in many cases like the above examples, we have strong reasons to believe it isn't. Second, we only partially understand how they do it, with severe limitations on our ability to use these genetic tools for our own purposes or to enhance them further.

One approach to addressing these problems is through adaptive laboratory evolution, which applies selective pressures to enhance reproductive fitness under the chosen environmental conditions<sup>28,29</sup>. Chance mutations which increase fitness increase in frequency in the experimental population, which can accumulate additional beneficial mutations over time. This technique has been applied successfully in many contexts to favor the evolution of new capabilities, most famously the ability for *Escherichia coli* to metabolize citrate in the Lenski lab's long term evolution experiment<sup>30</sup>, as well as many strains used for bioindustrial production<sup>31,32</sup>. Fitness increases of 50 to 100% in the first few months are common, followed by decreasing rates of gain<sup>28</sup>. However, despite its many successes, adaptive laboratory evolution is limited by the availability and evolvability of adaptive variants. For example, in the Lenski experiment, *E. coli* required 12 years and 31,000 generations to evolve the ability to metabolize citrate<sup>30</sup>, due to multiple required mutational steps that on their own yielded little to no fitness gains. Other desired traits may have equally complex genetic paths and so never be realized by adaptive laboratory evolution experiments despite being biologically possible. This is particularly likely to be the case when considering extraterrestrial environments unlike any on Earth.

Importantly, unlike axenic laboratory cultures, natural evolutionary processes do not occur in isolation and frequently utilize horizontal gene transfer to access molecular tools previously developed by other species<sup>33–36</sup>. Even distantly related species share sufficient molecular mechanisms to transcribe and translate foreign DNA, often acquired through highly mobile plasmids, viruses, and transposable elements<sup>37,38</sup>, which can lead to new capabilities such as perchlorate reduction<sup>39</sup> and enhanced survival in extreme environments<sup>40,41</sup>. This suggests a similar experimental approach may likewise combine the best aspects of the two methods. Indeed, synthetic biology-based approaches utilizing exogenous genetic constructs selected from the literature have provided many dramatic illustrations of expanding an organism's environmental tolerances, including increased resistance to desiccation<sup>42</sup>, salinity<sup>43</sup>, low temperatures<sup>44</sup>, and radiation<sup>45</sup>, as well as the ubiquitous use of antibiotic resistance as an essential tool for virtually all processes involving recombinant DNA. However, these genetic tools are typically identified only through laborious efforts<sup>45</sup> that do not lend themselves to scalability or understanding more than a tiny fraction of what evolution has exquisitely crafted and honed over billions of years. Thus, while the introduction of foreign DNA provides a powerful and complementary mechanism to traditional genetic mutation to produce rapid and dramatic increases in fitness, its applications in synthetic biology are limited by insufficient knowledge of existing natural genetic resources.

Functional metagenomics is a technique to experimentally discover gene function by cloning random genetic sequences from a diverse range of species into a highly diverse library pool in a tractable laboratory organism such as *E. coli*. Isolates from this library, each containing a single genetic fragment, can be assayed individually or the entire pooled library can be subjected to adaptive laboratory evolution to select for any sequences that increase fitness. This latter approach attempts to replicate and extend the natural process of horizontal gene transfer to efficiently screen diverse taxa for transferable cassettes conferring resistance to extreme environments, taking advantage of our greatest natural resource: Earth's tremendous biodiversity. Unlike targeted synthetic biology design, it is not biased towards already known genes or pathways, and does not even depend on the ability to culture or isolate an organism<sup>46</sup>, although functionality is strongly dependent on the activity of the original promoter in the host cell and typically expression levels vary widely<sup>47</sup>. Previous functional metagenomics work has yielded many successes in identifying new enzymes<sup>47</sup> as well as increasing tolerance to specific acute environmental stresses<sup>48,49</sup>.

To be truly useful, several key variables that play critical roles in functional metagenomic screening must be characterized. For example, for a given DNA source and host organism, the two fundamental variables in experimental design are fragment size and library copy number. These have not been explored systematically and it is not clear which values should be preferred. For instance, although larger inserts may contain intact clusters of related genes performing similar tasks<sup>50</sup>, they may also contain unrelated, deleterious sequences that impose a fitness cost<sup>37,51</sup>. Similarly, while it seems intuitive that more expression of a beneficial genetic tool further improves fitness, this may not be correct. Thus, a higher gene dosage may lead to benefits from higher gene expression, but may also disrupt the availability of some transcription factors to perform other functions or create a counterproductive metabolic load. Here, in an effort to increase the utility of functional metagenomics approaches, we looked at the importance of these foundational variables, insert size and copy number, and found that both can have outsized impacts on the efficacy of the introduced genetic material in non-intuitive ways.

## Results

### Creation and selection of functional metagenomic libraries

To explore these ideas, we created metagenomic libraries with varying insert sizes and studied the relationship between these key experimental parameters and their ability to increase fitness of the transformed *Escherichia coli* (strain Epi300 T1R). We used genomic DNA (gDNA) from 98 species: 96 described in a previous study of UV-resistant extremophiles<sup>9</sup>, the well-characterized polyextremophile *Deinococcus radiodurans*<sup>10,52</sup>, and the robust *E. coli* strain K12. Pooled gDNA was fragmented by sonication for varying amounts of time to yield a

range of sizes, which were then separated by gel electrophoresis and cut from the resulting agarose gel in six size ranges (< 1 kb, 1–3 kb, 3–6 kb, 6–10 kb, 10–20 kb, and > 20 kb), and purified.

The purified gDNA from different size ranges were cloned into the variable copy number pCC1FOS vector backbone and introduced to *E. coli* strain Epi300 T1R through viral transduction following the manufacturer's instructions (Lucigen #CCFOS059, CopyControl™ HTP Fosmid Library Production Kit), yielding thousands of independent colonies on LB agar plates (see [Methods](#)). Colonies from each insert size range were resuspended and pooled in LB to begin the library screening assay (Fig. 1). Each library pool was exposed to 20 s UV-C radiation (2.65 W/m<sup>2</sup> at 254 nm), estimated to kill 99.9% of *E. coli* with baseline tolerance. UV-C radiation was chosen for the focus of this study because it represents a novel environmental stress on present day Earth as the stratospheric ozone layer attenuates solar UV radiation below ~290 nm<sup>27</sup> while also approximating the peak absorbance and thus vulnerability of nucleic acids and polypeptides<sup>53</sup>. This exposure was chosen to provide a large selective advantage to any cells containing an insert that increased UV resistance, while minimizing stochastic loss that could arise through overly harsh conditions. Survivors were allowed to recover for 24 h in fresh LB, and the process was repeated daily for 10 days. On the final day, survivors were plated on LB agar plates to yield the final UV-resistant isolates. Through this pooled and iterative selection, the different library entries compete directly against each, as increased UVR tolerance directly leads to increased representation in the heterogeneous culture. This facilitates identification of those conferring the strongest selective advantage rather than surveying for all constructs conferring advantage. We purified and transformed these selected constructs into new *E. coli* Epi300 T1R before functional testing to eliminate the potential for background genomic mutations that may have accumulated and also influence UV resistance.

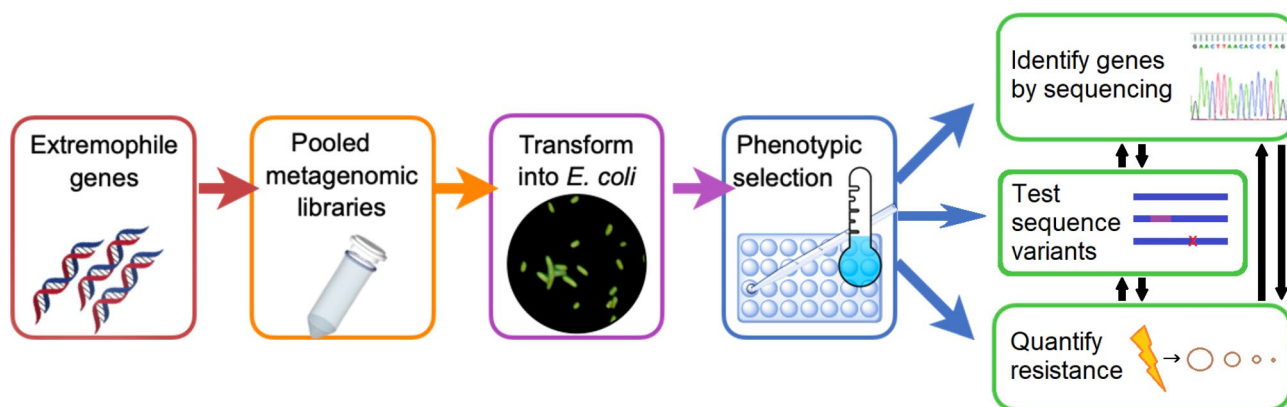
### Functional testing of library isolates

To examine the impact of library insert size, we tested the UV-C radiation resistance of cultures containing constructs from each size range. Confluent cultures were washed in phosphate buffered saline (1x PBS, Fisher Scientific BP2944-100) to remove UV-absorbing organics, exposed to varying amounts of UV radiation, and survival quantified through serial dilution and plating (see [Methods](#)). The smallest insert size range (< 1 kb) did not differ significantly from the negative control, but all larger inserts increased UV tolerance by > 1,000-fold (Fig. 2a). These trends were consistent across a range of UV radiation exposures (Fig. 2b). The smallest successful construct (1–3 kb size range) was significantly less UV-resistant than the larger constructs. This suggests it contained an effective but not complete resistance-conferring sequence. Although the largest insert sizes also underperformed relative to the strongest constructs (3–6 kb inserts), the loss of performance was far more severe for the undersized inserts.

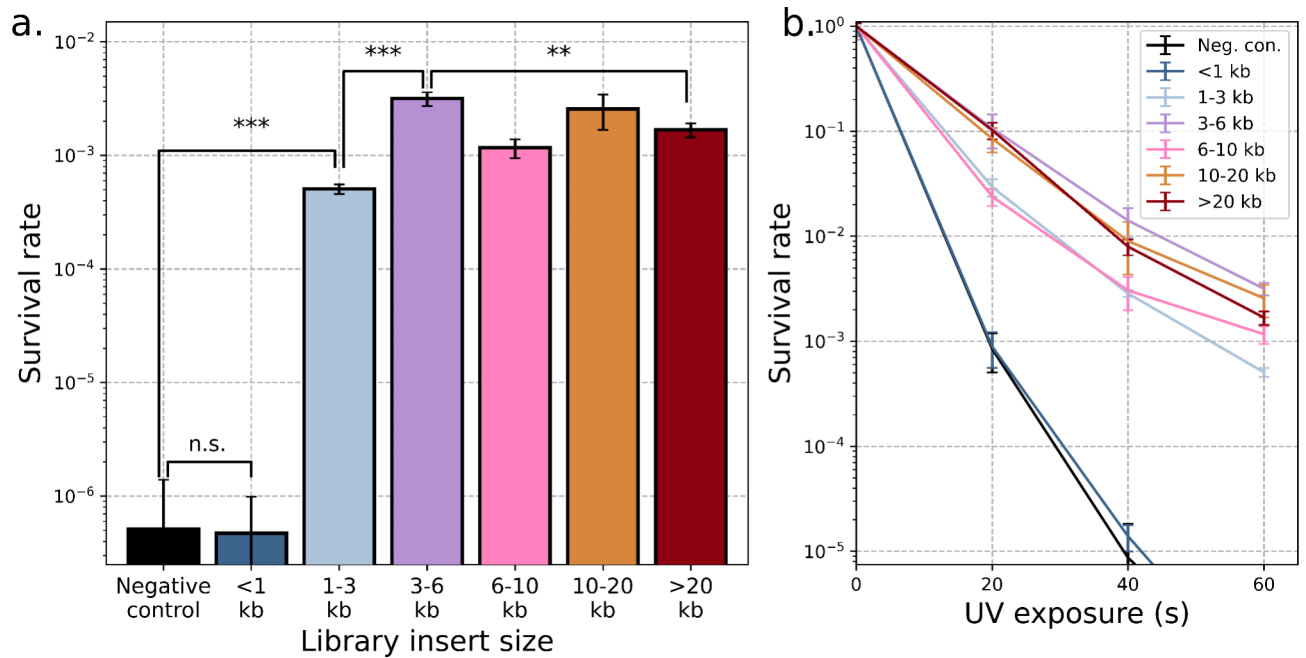
### Sequencing of functional metagenomic library isolates

To better understand the cause of the increased tolerances, we purified plasmids from resistant isolates in each size class (ZymoPure™ Plasmid Miniprep Kit, Zymo Research #D4211) and sequenced them in their entirety (Elim Biopharmaceuticals, Hayward, CA USA). Several – but not all – constructs contained the archetypal DNA repair gene *recA*<sup>54</sup>. The gene *recA*, which is non-functional in Epi300 T1R *E. coli*, is not a surprising discovery and has been previously discovered in prior functional metagenomic screens for enhanced resistance to UV radiation<sup>48</sup>. This makes it an ideal positive control and test case to explore the importance of insert size, copy number, and genomic DNA origin in functional metagenomic analysis.

Two distinct genetic sources in the functional metagenomics libraries yielded multiple independent isolates containing *recA* across a range of insert sizes: *E. coli* strain K12 and *Pantoea agglomerans* (formerly known as *Enterobacter agglomerans* or *Erwinia herbicola*), a plant-associated bacterium commonly found on flowers and trees that receive strong sunlight<sup>55,56</sup> (Fig. 3). We observed similar increases in UV tolerance with sequences from both sources. Intriguingly, although the species are widely divergent across their genomes<sup>57</sup>, the *recA* coding sequences are 86% identical at the nucleotide level and 96% similar at the amino acid level, with the majority of differences being concentrated in the C-terminus. In comparison, the adjacent gene in both species, *recX*, does not align at the nucleotide level and is only 64% similar at the amino acid level. This may reflect more stringent



**Fig. 1.** Experimental schematic of metagenomic libraries and screening.



**Fig. 2.** Increased UV tolerance from metagenomic library selection. **(a)** Survival after exposure to  $2.65 \text{ W/m}^2$  of UV exposure (254 nm) for 60 s for the most successful library entries in six different insert size ranges. The < 1 kb insert did not differ significantly from the negative control ( $p = 0.95$ ). The 1–3 kb insert is significantly more UV-tolerant than the negative control ( $p = 2.11 \times 10^{-6}$ ) but significantly less UV-tolerant than the 3–6 kb insert ( $p = 4.25 \times 10^{-5}$ ). The 3–6 kb insert is also more UV-tolerant than the > 20 kb insert ( $p = 0.0021$ ). **(b)** The constructs showed consistent differences in survival rate across a range of UV exposures. Error bars represent standard deviation across  $n = 3$  replicates.



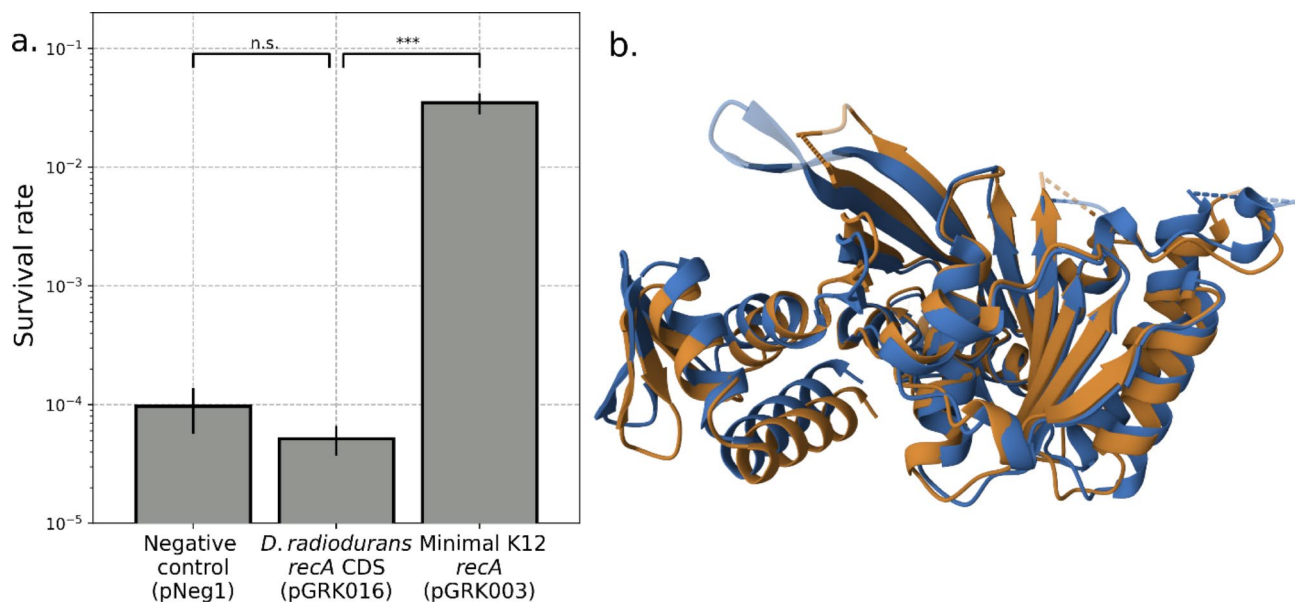
**Fig. 3.** Diversity of *recA*-containing fragments recovered. The extent of fragments originating from strain K12 **(a)** and *P. agglomerans* **(b)** is denoted by horizontal bars under their respective genetic maps. The genetic maps denote the start of each new gene (horizontal grey bars) with a vertical bar and highlight *recA* in red.

sequence constraint at the *RecA* locus, recent horizontal gene transfer between the *E. coli* and *P. agglomerans* lineages, or experimental selection bias if only more similar sequences can function effectively in *E. coli*.

We also cloned and tested the *recA* gene from *Deinococcus radiodurans* and observed no effect on UV tolerance. The eponymous radiation resistance of *D. radiodurans* has been hypothesized to be a consequence of exceptional DNA repair enzymes, including RecA<sup>58</sup>. We did not recover any sequences from *D. radiodurans* in our functional metagenomic assay. We postulate that could be the consequence of differing gene expression mechanisms. Replacing the coding sequence of the K12 *recA* with *D. radiodurans recA*, while maintaining the K12 regulatory sequences, also resulted in UV tolerance indistinguishable from the negative control (Fig. 4a), despite the peptide sequences being 76% similar and having highly similar protein structures (RMSD = 1.83, TM-score = 0.92)<sup>59–61</sup> (Fig. 4b). The large impact of these differences is consistent with prior mechanistic studies of the two homologs<sup>62</sup> and highlights how homologous proteins that nominally perform the same functions may do so in different and incompatible ways. It is also possible that the *recA* protein from *D. radiodurans* did not express as expected, perhaps due to differences in GC content, codon bias, transcript stability, different chaperones, or other mechanisms that may limit genetic compatibility between some organisms.

Not all constructs with demonstrated radiation resistance contained *recA*. For example, one construct conferred almost the same increase in UVR tolerance while containing a completely unrelated 9.8 kb insert from





**Fig. 4.** *D. radiodurans recA* does not increase UV tolerance. **(a)** *E. coli* transformed with either the *E. coli* strain K12 minimal construct (pGRK003), a variant of that construct with the coding sequence containing the same upstream promoter and downstream terminator sequences but with the *D. radiodurans recA* coding sequence (pGRK016), or a negative control plasmid (pNeg1) were exposed to 30 s of UV-C radiation (2.65 W/m<sup>2</sup> at 254 nm) and survival quantified by spot dilution assay. The indicated p-values, from left, are 0.11 and 0.00014 by two-tailed t-test. Error bars represent standard deviation across  $n = 3$  replicates. **(b)** Overlay of the RecA protein structures from *E. coli* strain K12 (blue) and from *D. radiodurans* (orange) shows high overall similarity as well as subtle differences, such as in  $\alpha$ -helix phasing.

K12. This sequence contains a cluster of genes involved in stress responses, including *yciT*<sup>63,64</sup>, *yciH*<sup>65</sup>, *osmB*<sup>66</sup>, *lapAB*<sup>67,68</sup>, *ribA*<sup>69,70</sup>, and *acnA*<sup>71</sup>. Previous work has highlighted the potential for stress response genes to play an important role in handling several distinct stressors. For example, the overexpression of exogenous *irrE* in *E. coli* increases both radioresistance<sup>72</sup> and salt tolerance<sup>43,73</sup>, emphasizing that even genes primarily associated with osmotic stress, such as *osmB*, can facilitate increased radiation resistance. Future work to molecularly dissect this construct could definitively identify which gene or genes underlie the improved resistance to UV radiation. This highlights how clusters of adjacent genes can collectively yield a phenotype comparable to single key genes, and suggests that larger insert sizes are valuable to allow such discoveries. It also illustrates the richness and potential of genomic resources available even in closely related species or within species.

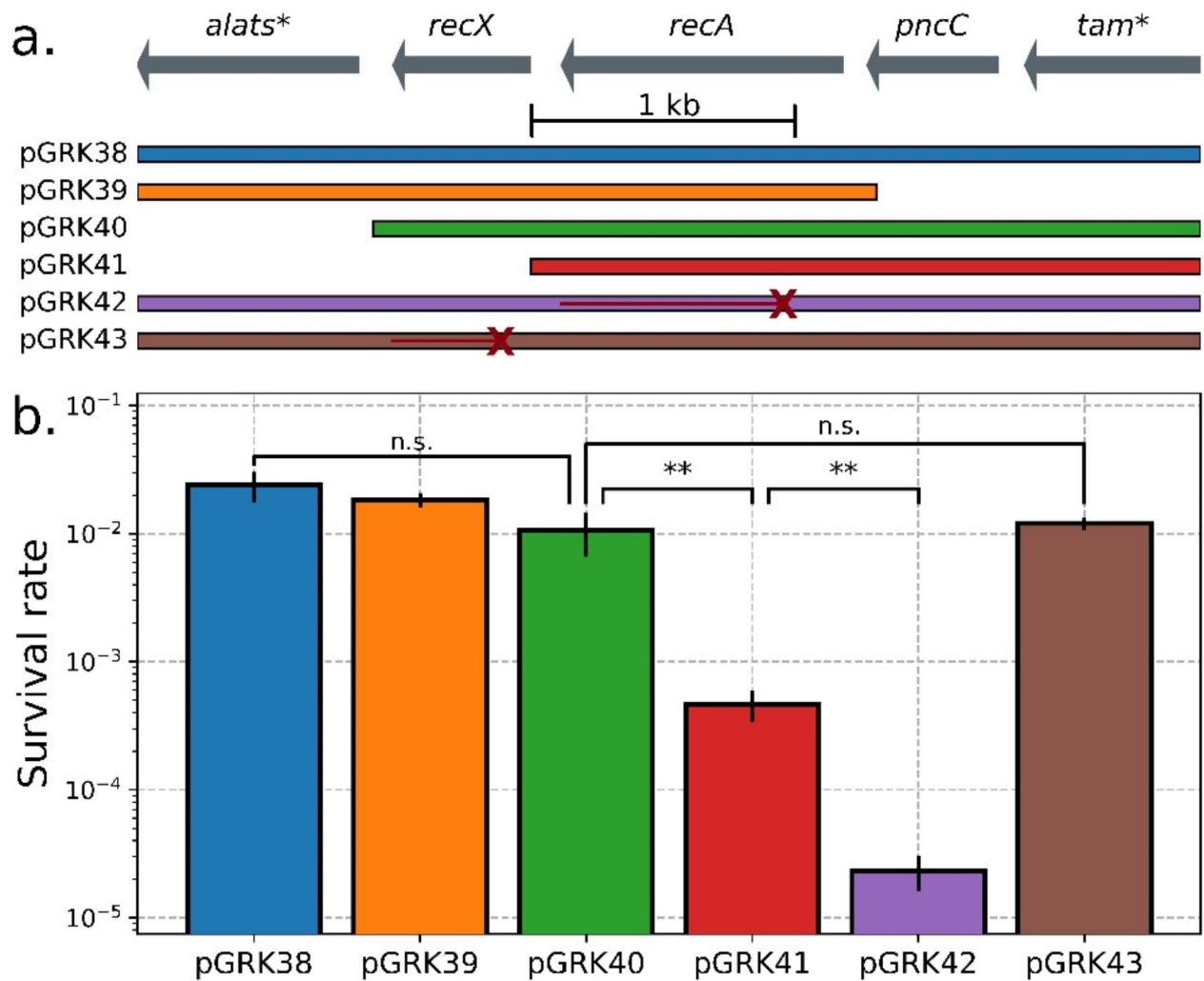
### Molecular dissection of UV resistance

To understand the difference in UV resistance between the smallest and larger *recA*-containing inserts (Figs. 1, 2 and 3 kb vs. 3–6 kb), we created variants of the most resistant construct (pGRK38, 4045 bp insert from *P. agglomerans*) (Fig. 5). Deleting sequences upstream (pGRK39, orange) or downstream (pGRK40, green) of *recX* had no effect on UV resistance, but additionally deleting *recX* (pGRK41, red) significantly reduced UV resistance. Critically, this is not an effect of the *recX* protein, as the introduction of an early *recX* frameshift mutation causing a premature stop codon (pGRK43, brown) without deleting the nucleotide sequence fully rescues UV tolerance. Rather, the downstream *RecX* sequence must instead play a role in ensuring the correct expression of *recA*, such as through distal elements regulating transcription or modulating mRNA stability. Consistent with these ideas, *recX* has previously been shown to be co-transcribed with *recA*<sup>74</sup>. This suggests that it is not enough for a library insert to contain the causative gene and the adjacent intergenic regions, and that neighboring sequences may be important for maximizing functional gains independent of their coding status.

### Impact of expression level

We next explored the impact of plasmid copy number and gene dosage more generally to examine the hypothesis that adding more of a gene that increases UV resistance would lead to further increases in UV resistance. We utilized the variable copy number of the pCC1FOS plasmid backbone to test this, which can be induced to increase from single copy to > 10 copies per cell. However, contrary to our hypothesis, we found that increasing the copy number of the largest *recA*-containing fragment actually resulted in a small but significant decrease in UV resistance (Fig. 6a, middle). Increasing plasmid copy number had no effect on the negative control. Further, moving the core *recA* coding sequence and intergenic regions to the very high copy number pUC19 backbone (the whole *recA* and negative control fragments could not be moved to the high copy number backbone due to their large size) resulted in a dramatic decrease in UV resistance below even the negative control (Fig. 6a, right).

One possible explanation is that the increased copy number of the *recA* promoter could be binding transcription factors with multiple targets, thus reducing their availability to activate other DNA damage

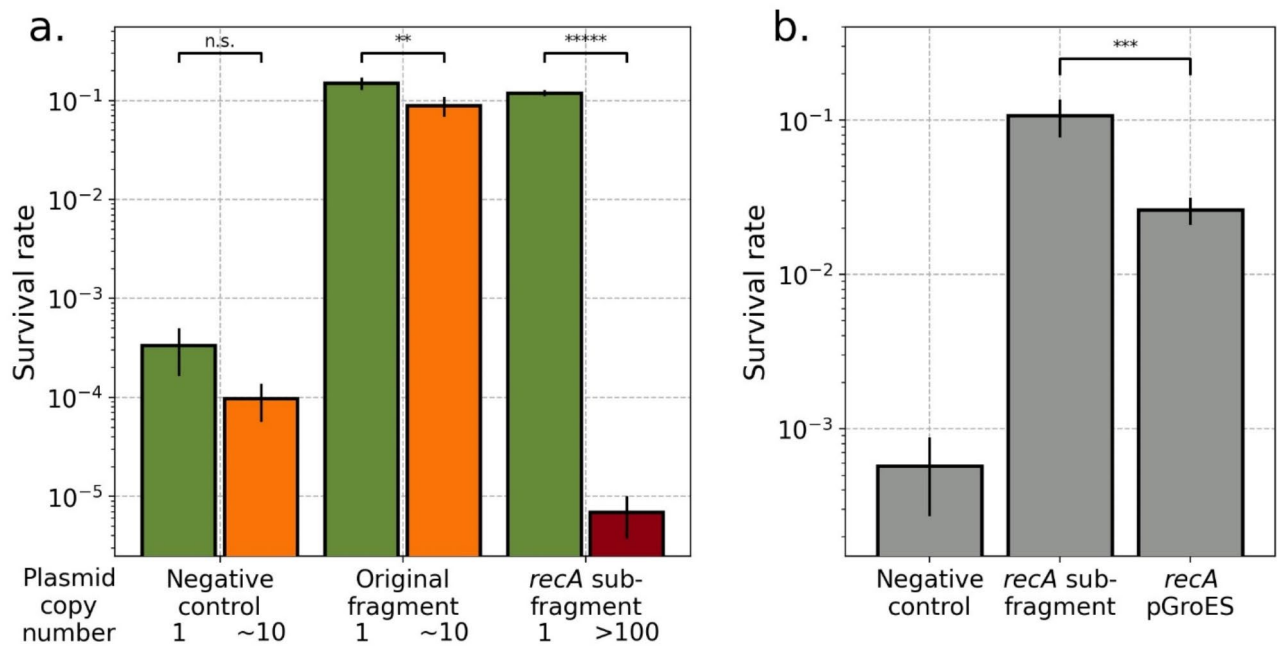


**Fig. 5.** Molecular dissection of UV resistant insert from *Pantoea agglomerans*. **(a)** A genetic map of the metagenomic library sequence derived from *P. agglomerans* (top) and derived variants, either complete (pGRK38), with fragments deleted (pGRK39, pGRK40, pGRK41), or with frameshift mutations causing premature stop codons in *recA* (pGRK42) or *recX* (pGRK43). The genes on the ends, *tam\** and *alats\**, are incomplete fragments. **(b)** *E. coli* containing the constructs from (a) were exposed to 30 s UV radiation (2.65 W/m<sup>2</sup> at 254 nm) and survival quantified by spot dilution assay. Significant differences were not observed between any fragments containing both *recA* and *recX* ( $p=0.067$ ), but the fragment lacking *recX* (pGRK41) had significantly lower survival ( $p=0.023$ ) though still higher than the nonfunctional *recA* (pGRK42,  $p=0.0087$ ). Restoring the *recX* sequence but not its activity (pGRK43) fully rescues this UV sensitivity and does not significantly differ from the intact sequence ( $p=0.66$ ). Error bars represent standard deviation across  $n=3$  replicates.

response genes. A second possible explanation is that the *recA* protein, which functions in conjunction with a large number of other proteins, including *recF*, *recO*, *recR*, *dinI*, *recX*, *rdgC*, *psiB*, *uvrD*, and *ssb*<sup>75</sup>, disrupts the stoichiometric balance of these other proteins when overexpressed, likewise reducing their effectiveness. To distinguish between these possibilities, we overexpressed *RecA* from a strong promoter (pGroES) at low plasmid copy number and observed (Fig. 6b) that UV resistance was diminished relative to the native promoter, but still significantly elevated compared to the negative control. This indicates that both proposed mechanisms play a role, but that the larger contribution is from copy number. This specific example illustrates the more general idea that adding a useful genetic tool may not be beneficial if added at the wrong dosage, and may even be detrimental.

## Discussion

As we explore the frontiers of biology beyond studying extant life into the novel challenges of astrobiology and bioindustrial applications, Theodosius Dobzhansky's aphorism, "Nothing in biology makes sense except in the light of evolution"<sup>76</sup> is more relevant than ever. To explore what is possible, whether optimizing production of



**Fig. 6.** Survival does not scale with copy number or expression level. **(a)** *E. coli* were transformed with either a negative control plasmid (pNeg1, left), large 35 kb fragment containing *recA* and adjacent sequences (pGRK006, center), or smaller fragment containing only *recA* and its immediate promoter and terminator (pGRK003 and pGRK010, right) maintained at differing copy numbers (green low, orange medium, red high) and were exposed to 30 s of UV radiation (2.65 W/m<sup>2</sup> at 254 nm) and survival was quantified by spot dilution assay. Survival was not affected by copy number in the negative control but was adversely affected in the two *recA* fragments (p-values, from left, are 0.057, 0.011, and 3.33e-7, by two-tailed t-test). **(b)** The *recA* promoter (from pGRK003, center) was replaced by the strong pGroES promoter (pGRK015, right) and samples were exposed to 20 s of UV radiation. Survival was decreased relative to the original promoter ( $p=0.0036$ ). For both, error bars represent standard deviation across  $n=3$  replicates.

biomolecules in the presence of challenging chemicals or evaluating the potential habitability of extraterrestrial environments, we need to develop methodology for efficiently expanding the environmental limits of life. This requires emulating as many of evolution's innovations as possible and to reuse existing genetic tools in new applications. As evidenced by the tremendous resilience of some organisms to perchlorate<sup>19,20</sup> and ionizing radiation<sup>21–23</sup> levels exceeding any found on Earth by many orders of magnitude, even seemingly novel environmental challenges may have at least partial genetic solutions amongst Earth's incredible biodiversity, if they can be found and redeployed.

The genetic tools developed by these organisms to address terrestrial environmental challenges are unlikely to be a complete solution for approaching the boundaries of life. For example, although salt export is a key adaptive strategy of many halophiles<sup>77</sup>, there are no known perchlorate export proteins, constituting a key gap in the genetic arsenal of life in the context of Martian perchlorate brines. In the future, *de novo* designed genes may be able to address these missing pieces and push the boundaries of life further still.

We have here built upon previous functional metagenomic studies to systematically explore the foundational variables of insert size and copy number and shown that they can have significant and counterintuitive effects. The specific size range that performed best here (3–6 kb) may not prove to be the ideal balance of opposing size concerns for all applications (for example, modulating membrane lipid composition may involve multiple sequential genes and benefit from a larger insert size), but rather illustrates how even a fragment containing the critical gene and flanking intergenic sequences may nonetheless dramatically underperform (Figs. 2 and 5).

On the other hand, additional increases in insert size do not yield further gains, and actually result in a small decrease in gained resistance. This could result from additional genes in the larger constructs that do not contribute to UV resistance performing activities that increase cellular stress either by their function or simply the metabolic load of increased insert size. This is consistent with prior studies that found selection against extraneous sequences in natural horizontal gene transfer<sup>34,51,78</sup>. However, while the oversized inserts (even those 10x larger) performed worse, it was a marginal difference, in sharp contrast with the more dramatic performance reduction of undersized inserts (Fig. 2). We therefore recommend erring on the side of larger inserts, which can subsequently be reduced in size as needed.

In contrast, higher copy number and higher expression can yield a markedly detrimental effect, sometimes even relative to the original, unmodified strain (Fig. 6). This defies the intuition that more of the correct genetic tool should yield further benefits and highlights the need for caution with these parameters. Importantly, this does not imply that low copy number backbones are superior, as previous functional metagenomics studies have

obtained excellent results with high copy number plasmid backbones<sup>49,79,80</sup>, but rather that different specific insert sequences yield optimal performance under different conditions. Consequently, screening functional metagenomic libraries at a given copy number, whether high or low, may result in overlooking or misevaluating some entries. Therefore, libraries should be screened at multiple copy numbers to maximize thoroughness, though low copy number screening is particularly important for the eventuality of genome integration for enhanced stability. Similarly, cDNA-derived expression libraries, although not examined directly here, are likely to see great benefit from testing transcripts under a variety of promoters rather than a single strong promoter.

In nature, adaptations to environmental conditions are typically the result of multiple independent genetic changes<sup>30,81,82</sup>, and so to evolve life beyond its current capabilities, we should expect to likewise require structured exploration of steps beyond the work performed here and in other functional metagenomic studies of identifying individual genetic tools, regardless of their performance. One critical next step is to systematically test combinations of such tools for additive fitness gains. This may be done through intraculture recombination<sup>83</sup>, iterative plasmid construction, or iterative genomic integration. Likewise, even strongly selected genetic tools are unlikely to be fully optimized for either the intracellular environment of their new host or for the specific environmental challenge of interest. Thus, additional performance may be realized from these tools through subsequent regulatory and/or coding changes, either through traditional adaptive laboratory evolution or facilitated by an orthogonal replication system to more rapidly sample sequence variants<sup>84</sup>. Thus viewed in the light of natural evolutionary processes, organized raiding of nature's genetic toolbox for individual tools is a critical first step of many to realize the common demands of both astrobiology and bioindustry to extend the limits of life.

## Methods

### Template DNA sources and purification

Glycerol stocks of 106 previously isolated UV-resistant microbes were thawed and grown for 72 h on agar plates as previously described<sup>9</sup>. Ninety six of the 106 described cultures grew, and only these were used in subsequent steps. Each culture was seeded into 100  $\mu$ L of its respective liquid medium and grown for 48 h at 30 C. Samples were pooled in batches of 10, and purified by the NEB Monarch<sup>®</sup> Genomic DNA Purification Kit (NEB #T3010S) according to the manufacturer's instructions. DNA preparations from *Deinococcus radiodurans* strain R1 (ATCC 13939) and *E. coli* strain K12 were purified as controls. Each DNA preparation was quantified by Nanodrop and pooled equally by mass to yield a single final sample containing DNA from all 98 source cultures.

### Size selection

The pooled DNA sample was subjected to four different conditions: no treatment, 1 s sonication, 10 s sonication, or 45 s sonication [Sonicator Ultrasonic Processor XL2020, Misonix Inc, with microtip at level 4.5] to generate fragments at a range of sizes. The treated samples were run on a 0.7% low melt-agarose (Promega #V211) gel overnight at 20 V, and the desired insert band size ranges were cut from the gel with a razor, dissolved with GELase (Lucigen #E0032-1D), and purified by ethanol precipitation.

### Library construction

Functional metagenomic libraries were prepared from the size-selected DNA samples with the CopyControl<sup>™</sup> HTP Fosmid Library Production Kit (Lucigen #CCFOS059), following manufacturer's instructions, with adjustment of the insert and backbone mass during ligation to account for the differing insert sizes being tested. Following lentiviral transduction, cultures were plated on LB plates with 30  $\mu$ g/mL chloramphenicol for amplification and quantification. Plates were then suspended in LB to begin repeated growth and UV exposure for selection of the pooled libraries, as described below. This yielded a total of 7,678 independent colonies, with more colonies from the larger insert sizes. Based on the expected insert sizes, these libraries span a total of 160 Mb (approximately 53 genomes averaging 3 Mb each), thus including roughly coverage of up to half of the total input DNA in the resultant metagenomic library. An additional library construction process with the same input DNA but fewer size categories yielded 70,950 colonies spanning an estimated 1259 Mb. Isolates from this second set of libraries yielded similar resistances and underlying sequences to those from the libraries described here.

### UV exposure

Confluent overnight cultures were collected by centrifugation at 5600 x g, resuspended in 1x PBS, then centrifuged and resuspended in PBS an additional time to remove organic matter from the growth media that may absorb UV radiation. Droplets containing 100  $\mu$ L of culture in PBS were then placed in Petri dishes in a fixed location in a NU-425-400 biological safety cabinet and the germicidal UV light (peak emission at 254 nm) activated (for time depending on the specific experiment). Irradiance was measured with a Solar Light PMA2100 sensor fitted with a germicidal UVC sensor as 2.65 W/m<sup>2</sup>. Droplets were then collected and used for either continued UVR selection or quantification by dilution spot assay.

### UV radiation selection

Libraries were screened for entries that conferred increased resistance to UV radiation by daily UV exposure (as described above) and recovery periods for 10 days. After UV exposure, the samples were diluted in 3 mL LB with 30  $\mu$ g/mL chloramphenicol for overnight recovery and growth. This process was then repeated for 10 days. After the final UV exposure, the samples were spread on LB agar plates to allow for picking and characterization of individual colonies.



### Quantifying survival by dilution spot assay

All reported survival rates derive from triplicate assays of each construct, both with and without UV exposure (for a total of six measurements per construct per condition). Cultures were prepared and either exposed or not exposed to UV radiation as described above. All samples were serially diluted by 10-fold and 10  $\mu$ L droplets of each dilution were plated on LB agar plates and incubated at 37 °C overnight. The resulting colonies were counted at the lowest dilution with distinct colonies, and these counts were used to compute surviving cells in the undiluted sample. Survival rates were calculated by dividing the surviving cell count for each of the UV-exposed samples by the average cell count of the unexposed samples.

### Copy number induction

Plasmid pCC1FOS was induced to high copy number by the addition of Lucigen CopyControl Induction Solution (Lucigen #CCIS125) and incubation at 37 °C with vigorous shaking for 24 h, following the manufacturer's instructions.

### Molecular cloning

Variants of the successful library entries were made following standard molecular biology techniques. Briefly, fragments were PCR amplified with Q5 polymerase (NEB #M0492S) with tagged primers (detailed primer and plasmid construction tables below). Amplicons were visualized on an agarose gel (0710-500G, VWR Life Sciences) stained with 0.5  $\mu$ g/mL ethidium bromide by an Azure 200 Gel Imager (Azure Biosystems). Residual template was digested by DpnI (NEB #R0176L) and products column purified (Zymo Research DNA Clean & Concentrator, #D4004). When homology-based cloning methods were used, fragments were then combined through HiFi DNA Assembly (NEB #M5520AA) and transformed into chemically competent Epi300 T1R *E. coli*. When restriction digest-based cloning methods were used, DNA was digested with BamHI-HF (NEB #R3136L) and SacI-HF (NEB #R3156L) in 1x rCutSmart buffer (NEB #B6004S), products gel-purified (Zymo Research Zymoclean Gel DNA Recovery Kit, #D4002), and ligated with T4 DNA Ligase (NEB #M0202L) and transformed into chemically competent Epi300 T1R *E. coli*. The resulting colonies were picked, grown overnight in LB with chloramphenicol (30  $\mu$ g/mL, Sigma #C0378) or carbenicillin (50  $\mu$ g/mL, Fisher Scientific #BP26481), plasmid DNA extracted and purified (Zymo Research Plasmid Miniprep Kit #D4211), and sequenced by Elim Biopharmaceuticals. Further details on the construction of all genetic constructs are given in Supplemental Table S1 and Supplemental Table S2. The Supplemental Materials also contain their full sequences.

### Statistical methods

All error bars represent standard deviation. All statistical tests are performed as unpaired, two-sided Student's t-test. Unless otherwise noted, all reflect results of  $n = 3$  replicates.

### Data availability

Sequences of all genetic constructs have been deposited in GenBank with accession numbers PQ202963-PQ202979 (additional details in Table S3). Physical data in the form of purified plasmids and bacterial strains will be stored in the Rothschild Lab for at least 5 years and available upon request to Dr. Rothschild.

Received: 4 August 2024; Accepted: 18 December 2024

Published online: 02 January 2025

### References

- Merino, N. et al. Living at the extremes: Extremophiles and the limits of life in a planetary context. *Front. Microbiol.* **10**, 780 (2019).
- Belilla, J. et al. Hyperdiverse archaea near life limits at the polyextreme geothermal dallol area. *Nat. Ecol. Evol.* **3**, 1552–1561 (2019).
- Rothschild, L. J. & Mancinelli, R. L. Life in extreme environments. *Nature* **409**, 1092–1101 (2001).
- Stetter, K. O. Hyperthermophiles in the history of life. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **361**, 1837–1842 (2006). discussion 1842–3.
- Murray, A. E. et al. Microbial life at  $-13$  °C in the brine of an ice-sealed Antarctic lake. *Proc. Natl. Acad. Sci.* **109**, 20626–20631 (2012).
- Oren, A. Thermodynamic limits to microbial life at high salt concentrations. *Environ. Microbiol.* **13**, 1908–1923 (2011).
- Petit, P., Hayoun, K., Alpha-Bazin, B., Armengaud, J. & Rivasseau, C. First isolation and characterization of bacteria from the core's cooling pool of an operating nuclear reactor. *Microorganisms* **11** (2023).
- Thornley, M. J. Radiation resistance among bacteria. *J. Appl. Bacteriol.* **26**, 334–345 (1963).
- Paulino-Lima, I. G. et al. Extremely high UV-C radiation resistant microorganisms from desert environments with different manganese concentrations. *J. Photochem. Photobiol B* **163**, 327–336 (2016).
- Makarova, K. S. et al. Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. *Microbiol. Mol. Biol. Rev.* **65**, 44–79 (2001).
- Lauro, S. E. et al. Multiple subglacial water bodies below the south Pole of Mars unveiled by new MARSIS data. *Nat. Astronomy*. <https://doi.org/10.1038/s41550-020-1200-6> (2020).
- Morowitz, H. & Sagan, C. Life in the clouds of Venus? *Nature* **215**, 1259–1260 (1967).
- Balan, V. Current challenges in commercially producing biofuels from lignocellulosic biomass. *ISRN Biotechnol.* **463074** 2014 (2014).
- Jin, M., Balan, V., Gunawan, C. & Dale, B. E. Quantitatively understanding reduced xylose fermentation performance in AFEX™ treated corn stover hydrolysate using *Saccharomyces cerevisiae* 424A (LNH-ST) and *Escherichia coli* KO11. *Bioresour Technol.* **111**, 294–300 (2012).
- Lam, F. H., Ghaderi, A., Fink, G. R. & Stephanopoulos, G. Biofuels engineering alcohol tolerance in yeast. *Science* **346**, 71–75 (2014).
- Mhetras, N., Mapare, V. & Gokhale, D. Cold active lipases: Biocatalytic tools for greener technology. *Appl. Biochem. Biotechnol.* **193**, 2245–2266 (2021).
- Deng, B. et al. Improving the activity and thermostability of PETase from *Ideonella sakaiensis* through modulating its post-translational glycan modification. *Commun. Biol.* **6**, 1–10 (2023).

18. Chen, J., Liu, Y., Diep, P. & Mahadevan, R. Harnessing synthetic biology for sustainable biomining with Fe/S-oxidizing microbes. *Front. Bioeng. Biotechnol.* **10**, 920639 (2022).
19. Heinz, J. et al. Bacterial growth in chloride and perchlorate brines: Halotolerances and salt stress responses of *Planococcus Halocryophilus*. *Astrobiology* **19**, 1377–1387 (2019).
20. Heinz, J., Krahn, T. & Schulze-Makuch, D. A new record for microbial perchlorate tolerance: Fungal growth in NaClO<sub>4</sub> brines and its implications for putative life on Mars. *Life* **10** (2020).
21. Erickson, G. E. *Geology and Origin of the Chilean Nitrate Deposits*. <https://pubs.usgs.gov/publication/pp1188> (1981). <https://doi.org/10.3133/pp1188>
22. Vega, M., Nerenberg, R. & Vargas, I. T. Perchlorate contamination in Chile: Legacy, challenges, and potential solutions. *Environ. Res.* **164**, 316–326 (2018).
23. Calderón, R. et al. Perchlorate levels in soil and waters from the Atacama Desert. *Arch. Environ. Contam. Toxicol.* **66**, 155–161 (2014).
24. Moseley, B. E. & Mattingly, A. Repair of irradiation transforming deoxyribonucleic acid in wild type and a radiation-sensitive mutant of *Micrococcus radiodurans*. *J. Bacteriol.* **105**, 976–983 (1971).
25. Dartnell, L. R., Desorgher, L., Ward, J. M. & Coates, A. J. Modelling the surface and subsurface martian radiation environment: Implications for astrobiology. *Geophys. Res. Lett.* **34**, L02207 (2007).
26. Mattimore, V. & Battista, J. R. Radioresistance of *Deinococcus radiodurans*: Functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. *J. Bacteriol.* **178**, 633–637 (1996).
27. Döhring, T. SPIE. Critical discussion on the UV absorption properties of Earth's atmosphere. In *Third International Conference on Applications of Optics and Photonics* vol. 10453, 128–137 (2017).
28. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution—Principles and applications for biotechnology. *Microb. Cell. Fact.* **12**, 64 (2013).
29. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O. & Feist, A. M. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metab. Eng.* **56**, 1–16 (2019).
30. Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7899–7906 (2008).
31. Lee, S. & Kim, P. Current status and applications of adaptive laboratory evolution in industrial microorganisms. *J. Microbiol. Biotechnol.* **30**, 793–803 (2020).
32. Wang, G. et al. Recent progress in adaptive laboratory evolution of industrial microorganisms. *J. Ind. Microbiol. Biotechnol.* **50**, kuac023 (2022).
33. Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 (2008).
34. Van Etten, J. & Bhattacharya, D. Horizontal gene transfer in eukaryotes: Not if, but how much? *Trends Genet.* **36**, 915–925 (2020).
35. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: Building the web of life. *Nat. Rev. Genet.* **16**, 472–482 (2015).
36. Arnold, B. J., Huang, I. T. & Hanage, W. P. Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* **20**, 206–218 (2022).
37. Daubin, V. & Szöllösi, G. J. Horizontal gene transfer and the history of life. *Cold Spring Harb Perspect. Biol.* **8**, a018036 (2016).
38. Rossoni, A. W. et al. The genomes of polyextremophilic cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions. *Elife* **8**, e45017 (2019).
39. Melnyk, R. A. et al. Identification of a perchlorate reduction genomic island with novel regulatory and metabolic genes. *Appl. Environ. Microbiol.* **77**, 7401–7404 (2011).
40. Schönknecht, G. et al. Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**, 1207–1210 (2013).
41. Feng, S., Powell, S. M., Wilson, R. & Bowman, J. P. Extensive gene acquisition in the extremely psychrophilic bacterial species *Psychroflexus torquis* and the link to sea-ice ecosystem specialism. *Genome Biol. Evol.* **6**, 133–148 (2014).
42. Billi, D. et al. Engineering desiccation tolerance in *Escherichia coli*. *Appl. Environ. Microbiol.* **66**, 1680–1684 (2000).
43. Pan, J. et al. IrrE, a global regulator of extreme radiation resistance in *Deinococcus radiodurans*, enhances salt tolerance in *Escherichia coli* and *Brassica napus*. *PLoS One.* **4**, e4422 (2009).
44. Ferrer, M., Chernikova, T. N., Yakimov, M. M., Golyshin, P. N. & Timmis, K. N. Chaperonins govern growth of *Escherichia coli* at low temperatures. *Nat. Biotechnol.* **21**, 1266–1267 (2003).
45. Hashimoto, T. et al. Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nat. Commun.* **7**, 12808 (2016).
46. Reitner, J. & Thiel, V. *Encyclopedia of Geobiology* (Springer, 2011).
47. Leis, B., Angelov, A. & Liebl, W. Screening and expression of genes from metagenomes. *Adv. Appl. Microbiol.* **83**, 1–68 (2013).
48. Lamprecht-Grandío, M. et al. Novel genes involved in resistance to both ultraviolet radiation and perchlorate from the metagenomes of hypersaline environments. *Front. Microbiol.* **11**, 453 (2020).
49. Diaz-Rullo, J. et al. Mining for perchlorate resistance genes in microorganisms from sediments of a hypersaline pond in Atacama Desert, Chile. *Front. Microbiol.* **12**, 723874 (2021).
50. Fischbach, M. A., Walsh, C. T. & Clardy, J. The evolution of gene collectives: How natural selection drives chemical innovation. *Proc. Natl. Acad. Sci. U S A.* **105**, 4601–4608 (2008).
51. Hall, R. J., Whelan, F. J., McInerney, J. O., Ou, Y. & Domingo-Sananes, M. R. Horizontal gene transfer as a source of conflict and cooperation in prokaryotes. *Front. Microbiol.* **11**, 1569 (2020).
52. Setlow, J. K. & Duggan, D. E. The resistance of *Micrococcus radiodurans* to ultraviolet radiation. *Biochim. Biophys. Acta* **87**, 664–668 (1964).
53. Glasel, J. A. Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *Biotechniques* **18**, 62–63 (1995).
54. Horii, T., Ogawa, T. & Ogawa, H. Organization of the recA gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. U S A.* **77**, 313–317 (1980).
55. Johnson, K. B., Stockwell, V. O., Sawyer, T. L. & Sugar, D. Assessment of environmental factors influencing growth and spread of *Pantoea agglomerans* on and among blossoms of pear and apple. *Phytopathology* **90**, 1285–1294 (2000).
56. Loncaric, I. et al. Typing of *Pantoea agglomerans* isolated from colonies of honey bees (*Apis mellifera*) and culturability of selected strains from honey. *Apidologie* **40**, 40–54 (2009).
57. Kumar, S. et al. TimeTree 5: An expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, msac174 (2022).
58. Kim, J. I. et al. RecA protein from the extremely radioresistant bacterium *Deinococcus radiodurans*: Expression, purification, and characterization. *J. Bacteriol.* **184**, 1649–1660 (2002).
59. Xing, X. & Bell, C. E. Crystal structures of *Escherichia coli* RecA in a compressed helical filament. *J. Mol. Biol.* **342**, 1471–1485 (2004).
60. Rajan, R. & Bell, C. E. Crystal structure of RecA from *Deinococcus radiodurans*: Insights into the structural basis of extreme radioresistance. *J. Mol. Biol.* **344**, 951–963 (2004).
61. Bittrich, S., Segura, J., Duarte, J. M., Burley, S. K. & Rose, Y. RCSB protein Data Bank: Exploring protein 3D similarities via comprehensive structural alignments. *Bioinformatics* **40** (2024).
62. Kim, J. I. & Cox, M. M. The RecA proteins of *Deinococcus radiodurans* and *Escherichia coli* promote DNA strand exchange via inverse pathways. *Proc. Natl. Acad. Sci.* **99**, 7917–7921 (2002).

63. Sun, Z. et al. Density of  $\sigma 70$  promoter-like sites in the intergenic regions dictates the redistribution of RNA polymerase during osmotic stress in *Escherichia coli*. *Nucleic Acids Res.* **47**, 3970–3985 (2019).
64. Gao, Y. et al. Unraveling the functions of uncharacterized transcription factors in *Escherichia coli* using ChIP-exo. *Nucleic Acids Res.* **49**, 9696–9710 (2021).
65. Osterman, I. A. et al. A bacterial homolog *YciH* of eukaryotic translation initiation factor eIF1 regulates stress-related gene expression and is unlikely to be involved in translation initiation fidelity. *RNA Biol.* **12**, 966–971 (2015).
66. Boulanger, A. et al. Multistress regulation in *Escherichia coli*: Expression of *osmB* involves two independent promoters responding either to sigmaS or to the RcsCDB his-asp phosphorelay. *J. Bacteriol.* **187**, 3282–3286 (2005).
67. Klein, G., Kobylak, N., Lindner, B., Stupak, A. & Raina, S. Assembly of lipopolysaccharide in *Escherichia coli* requires the essential LapB heat shock protein. *J. Biol. Chem.* **289**, 14829–14853 (2014).
68. Klein, G. et al. Multiple transcriptional factors regulate transcription of the *rpoE* gene in *Escherichia coli* under different growth conditions and when the lipopolysaccharide biosynthesis is defective. *J. Biol. Chem.* **291**, 22999–23019 (2016).
69. Koh, Y. S., Choih, J., Lee, J. H. & Roe, J. H. Regulation of the *ribA* gene encoding GTP cyclohydrolase II by the *soxRS* locus in *Escherichia coli*. *Mol. Gen. Genet.* **251**, 591–598 (1996).
70. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res.* **30**, 3141–3151 (2002).
71. Tang, Y., Quail, M. A., Artymiuk, P. J., Guest, J. R. & Green, J. *Escherichia coli* aconitases and oxidative stress: post-transcriptional regulation of *sodA* expression. *Microbiology* **148**, 1027–1037 (2002).
72. Gao, G. et al. Expression of *Deinococcus radiodurans* *Ppr1* enhances the radioresistance of *Escherichia coli*. *DNA Repair.* **2**, 1419–1427 (2003).
73. Ma, R. et al. Improved osmotic tolerance and ethanol production of ethanologenic *Escherichia coli* by *IrrE*, a global regulator of radiation-resistance of *Deinococcus radiodurans*. *Curr. Microbiol.* **62**, 659–664 (2011).
74. Pagès, V., Koffel-Schwartz, N. & P Fuchs, R. *recX*, a new SOS gene that is co-transcribed with the *recA* gene in *Escherichia coli*. *DNA Repair.* **2**, 273–284 (2003).
75. Cox, M. M. Regulation of bacterial *RecA* protein function. *Crit. Rev. Biochem. Mol. Biol.* **42**, 41–63 (2007).
76. Dobzhansky, T. Nothing in biology makes sense except in the light of evolution. *Am. Biology Teacher.* **75**, 87–91 (1973).
77. DasSarma, S. & DasSarma, P. Halophiles and their enzymes: Negativity put to good use. *Curr. Opin. Microbiol.* **25**, 120–126 (2015).
78. Arber, W. Horizontal gene transfer among bacteria and its role in biological evolution. *Life* **4**, 217–224 (2014).
79. Mirete, S., Morgante, V. & González-Pastor, J. E. Functional metagenomics of extreme environments. *Curr. Opin. Biotechnol.* **38**, 143–149 (2016).
80. Mirete, S. et al. Salt resistance genes revealed by functional metagenomics from brines and moderate-salinity rhizosphere within a hypersaline environment. *Front. Microbiol.* **6**, 1121 (2015).
81. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481–487 (2005).
82. Roberts Kingman, G. A. et al. Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Sci. Adv.* **7**, eabg5285 (2021).
83. Chu, H. Y., Sprouffske, K. & Wagner, A. Assessing the benefits of horizontal gene transfer by laboratory evolution and genome sequencing. *BMC Evol. Biol.* **18**, 54 (2018).
84. Tian, R. et al. Establishing a synthetic orthogonal replication system enables accelerated evolution in *E. Coli*. *Science* **383**, 421–426 (2024).

## Acknowledgements

We thank Ivan Paulino-Lima for the donation of his collection of UV-resistant microbes and all other members of the Rothschild Lab for their valuable comments and feedback throughout the course of this work. G.A.R.K.'s research was supported by an appointment to the NASA Postdoctoral Program at Ames Research Center, administered by Oak Ridge Associated Universities under contract with NASA.

## Author contributions

G.A.R.K. and L.J.R. conceived of the project. G.A.R.K. designed, performed, and analyzed the experiments, with assistance from J.K. for the experiments pertaining to Fig. 5. G.A.R.K. wrote the paper and prepared all figures with input from L.J.R. All authors reviewed the manuscript.

## Declarations

### Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-83952-w>.

**Correspondence** and requests for materials should be addressed to L.J.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

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

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2024