scientific reports

Check for updates

OPEN

Genomic characterization of rare earth binding by *Shewanella oneidensis*

Sean Medin[1], Alexa M. Schmitz[1], Brooke Pian[1], Kuunemuebari Mini[2], Matthew C. Reid[3], Megan Holycross[4], Esteban Gazel[4], Mingming Wu[1] & Buz Barstow[1]✉

Rare earth elements (REE) are essential ingredients of sustainable energy technologies, but separation of individual REE is one of the hardest problems in chemistry today. Biosorption, where molecules adsorb to the surface of biological materials, offers a sustainable alternative to environmentally harmful solvent extractions currently used for separation of rare earth elements (REE). The REE-biosorption capability of some microorganisms allows for REE separations that, under specialized conditions, are already competitive with solvent extractions, suggesting that genetic engineering could allow it to leapfrog existing technologies. To identify targets for genomic improvement we screened 3,373 mutants from the whole genome knockout collection of the known REE-biosorbing microorganism *Shewanella oneidensis* MR-1. We found 130 genes that increased biosorption of the middle REE europium, and 112 that reduced it. We verified biosorption changes from the screen for a mixed solution of three REE (La, Eu, Yb) using Inductively Coupled Plasma Mass Spectrometry (ICP-MS) in solution conditions with a range of ionic strengths and REE concentrations. We identified 18 gene ontologies and 13 gene operons that make up key systems that affect biosorption. We found, among other things, that disruptions of a key regulatory component of the arc system (*hptA*), which regulates cellular response to anoxic environments and polysaccharide biosynthesis related genes (*wbpQ*, *wbnJ*, *SO_3183*) consistently increase biosorption across all our solution conditions. Our largest total biosorption change comes from our *SO_4685*, a capsular polysaccharide (CPS) synthesis gene, disruption of which results in an up to 79% increase in biosorption; and *nusA*, a transcriptional termination/anti-termination protein, disruption of which results in an up to 35% decrease in biosorption. Knockouts of *glnA*, *pyrD*, and *SO_3183* produce small but significant increases (≈ 1%) in relative biosorption affinity for ytterbium over lanthanum in multiple solution conditions tested, while many other genes we explored have more complex binding affinity changes. Modeling suggests that while these changes to lanthanide biosorption selectivity are small, they could already reduce the length of repeated enrichment process by up to 27%. This broad exploratory study begins to elucidate how genetics affect REE-biosorption by *S. oneidensis*, suggests new areas of investigation for better mechanistic understanding of the membrane chemistry involved in REE binding, and offer potential targets for improving biosorption and separation of REE by genetic engineering.

Rare Earth Elements (REE), typically referring to the lanthanides (lanthanum to lutetium) and sometimes scandium and yttrium, are essential ingredients for sustainable energy technologies including high strength lightweight magnets used in electric vehicles and wind turbines[1,2]; room temperature superconductors[3]; lightweight high-strength alloys[4,5]; high-efficiency lighting[6]; and battery anodes[7]. All of these applications put an increasing demand on the global REE supply chain. As the world demand for sustainable energy grows[8], developing a sustainable supply chain for high-purity REE is critical[9].

[1]Department of Biological and Environmental Engineering, Cornell University, Cornell University, 228 Riley-Robb Hall, Ithaca, NY 14853, USA. [2]Department of Sciences and Technology Studies, Cornell University, Ithaca, NY 14853, USA. [3]School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA. [4]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY 14853, USA. ✉email: bmb35@cornell.edu

Current methods for refining REE often involve harsh chemicals, high temperatures, high pressures, and generate a considerable amount of toxic waste[10–12]. These processes give sustainable energy technologies reliant on REE a high environmental and carbon footprint.

The majority of REE chemical separations utilize commercially available organic solvents and extractants[13]. All lanthanides exist as trivalent cations and the ionic radius difference between the largest rare earth, La^{3+}, and the smallest rare earth, Lu^{3+}, is only 0.17 Å[14]. This means that separations of adjacent or near-adjacent REE pose an enormous challenge for conventional chemical methods, requiring organic solvent extractions in extremely long mixer settler devices[15]. This results in large amounts of toxic waste being generated. As a consequence, due to its high environmental standards, the United States has no capacity to produce purified REE. Furthermore, only two REE purification plants exist outside of China[11,12,16].

Biomining is a promising alternative to conventional mining technologies, and already supplies 5% and 15% of the world's gold and copper[17]. We anticipate that a REE-biomining system will operate in three steps: (1) bioleaching metals from an ore or end-of-life feedstock like a magnet; (2) separating the lanthanides from all other metals present in the leachate (*e.g.*, uranium and thorium from an ore, or iron from a magnet); and (3) separating individual lanthanides. We anticipate that the second and third steps will be carried out separately, removing the engineering challenge of simultaneously discriminating between lanthanide and non-lanthanide and individual lanthanides. Significant progress has been made in developing microorganisms for the first bioleaching step of the biomining process[18,19], and in the second total lanthanide separation step[20–23].

New biological and chemical methods[24–27] have recently been developed to address the challenges of total[20–23,28–30], light and heavy REE[31], and individual[32–34] REE-separations. For example, lanmodulin, a REE-binding protein discovered in methylotrophic bacteria, is selective for REE in the presence of molar amounts of competing metal cations[20,21]. Meanwhile, lanthanide binding tags (LBTs), attached either to the surface of *Caulobacter crescentus*[22] or an engineered curli biofilm[23] can selectively bind REE in the presence of competing metals. Finally, both *Methylorubrum extorquens*[31], and *E. coli* engineered with surface-displayed LBTs[28] are able to preferentially accumulate heavy lanthanides from a mixed solution of lanthanides. Despite numerous advances in total REE separation from competing metals, advancement of separation of individual REE from total REE remains a challenge.

Biosorption and desorption from the surface of a microbial cell offers an environmentally-friendly route for individual REE-separation. Biosorption can provide high metal binding capacity at low cost[35,36]. The cell surface, containing proteins[37], lipids[38], and polysaccharides[39], offers a rich chemical environment for selectively binding and releasing REE. The membranes of both gram-negative and gram-positive bacteria contain sites that bind REE[40]. Bonificio et al. have demonstrated nascent REE-separation capability by biosorption and desorption under decreasing pH by *Shewanella oneidensis* MR-1[2]. Furthermore, Bonificio et al. demonstrated that, under specialized conditions, REE separations are competitive with solvent extraction using *Roseobacter* sp. AzwK-3b[2].

However, knowledge of binding sites for biosorption and their mechanisms of action remains poorly understood[28,41,42]. Prior studies indicate that REE binding on the membrane is primarily driven by phosphate and carboxylate binding sites on both gram-negative and gram-positive bacteria[40,43]. An EXAFS study by Takahashi et al. found that heavy REE were likely enriched by surface complexation with multiple phosphate sites while carboxylate and single phosphate sites generally contributed to a relatively more uniform distribution of REE binding[43]. Moriwaki et al. found that teichoic acid defective strains of *Bacillus subtilis* (a gram-positive microbe) had severely reduced REE binding when freeze dried as a powder[42,44]. Although these are undoubtedly key binding sites, they are unlikely to be the only ones. Teichoic acids are rich in phosphate sites, but spectroscopic data suggests that REE-binding carboxylate sites were also present on the cell wall[43]. Furthermore, the key sources of REE binding on gram-negative bacteria remains uncertain. There has been speculation that Lipid A phosphate groups significantly contribute to REE biosorption but testing of a mutant with reduced free Lipid A phosphate groups in *E. coli* did not find a significant change of REE binding in the conditions tested[28].

New whole genome engineering methods allowing for targeted mutagenesis of genomic loci have enabled substantial improvements in complex phenotypes including lycopene production[45], naringenin and glucaric acid production[46], and spectinomycin resistance[47]. We hypothesize that by mutagenizing genomic loci (particularly regulatory regions, but also small sections of protein coding sequences) involved in REE biosorption, we can engineer REE-biosorption selectivity.

However, despite the promising capability identified by Bonificio et al., and prior advances in mechanistic understanding, there is no obvious roadmap for genetic engineering of biosorption to reduce binding of non-REE metals and improve binding selectivity for specific REE that would enable separation with biosorption to leapfrog existing technologies.

In this work we comprehensively profile the genetics of REE biosorption in *S. oneidensis*. We chose to focus our efforts on *S. oneidensis* because it was found to have nascent REE separation capability[33]; has a fast doubling time; has advanced genetic engineering tools[48–50]; and has an existing whole genome knockout collection that we could use for our genetic screen[51,52]. In our work, we intended our screening to be as expansive as possible. While previous works have found that lower pH was associated with higher selectivity for heavier REE[33,40,43], we specifically chose a relatively high pH, so that we could identify the full breadth of genes involved in REE biosorption. While not every gene we found would affect selectivity for particular REE or REE in general, we felt confident that the set of genes we uncovered would include all, if not most of, the genes that do affect REE selectivity.

In this work, we begin by conducting a high-throughput genetic screen to identify genes involved in biosorption of Eu. As this screen contains numerous possible sources of systematic (variations in growth phase and bacterial density across each 96-well plate) and random noise (variations in reagent concentration, inoculum quantity, and cross-contamination due to splatter) we conduct a global analysis of our genetic screen results to identify which gene categories are statistically enriched with hits from our screen. This increases our confidence in which biological mechanisms truly contributed to biosorption. Finally, we take a closer look at some of our

genes of interest to partially validate our results as well as to identify how total biosorption and REE specificity changes across several different solution conditions.

Results

Genetic screen finds 242 genes that influence europium biosorption

We screened 3373 unique members of the *S. oneidensis* whole genome knockout collection[51,53] for genes that control biosorption of the middle REE europium (Eu) using an Arsenazo III (As-III) colorimetric screen[54] (Figs. 1A and S1, Materials and Methods). We hypothesized that given (1) the tendency of the affinity of many bacterial REE binding sites to have a preference for middle REE or (2) monotonically change with atomic number[20] (*e.g.*, the site will bind heavy REE more strongly than middle REE, and middle more strongly than light REE), and (3) that no lanthanide-binding site is likely to be totally selective for an individual lanthanide, that the middle-REE Eu is most likely to bind to most sites. As a result, screening for mutants with altered Eu-biosorption would reveal the largest number of mutants with both altered total biosorption and altered selective biosorption. For example, knocking out a site with no selective preference for lanthanides will reduce Eu-, La-, and Yb-binding equally. At the same time, because it is unlikely that any site is perfectly selective, the position of Eu in the lanthanide series means that an assay using it will detect more changes due to disruption of selective sites. For example, knocking out a site with high La-affinity will still reduce Eu-binding somewhat, but might not impact Yb-binding as much. Likewise, knocking out a site with high Yb-affinity will also reduce Eu-binding, and is likely to reduce La-binding even less.

In total, we found 130 gene disruption mutants that have significantly higher Eu-biosorption, and 112 that have significantly lower Eu-biosorption (Fig. 1B, Dataset S1). We note that there was a slight (2.9%), but statistically significant ($p < 0.0001$), decrease in the average growth OD of our biosorption outliers compared to the non-outliers. This suggests that mutations that induce biosorption changes are more likely to have a small growth defect compared with mutations that do not. We cannot, however, rule out the possibility that we had a bias towards mis-identifying low-optical density strains as biosorption outliers.

Confirming replicability of genetic screen results

To test the efficacy of our screening procedure, we randomly chose 64 mutants from 7 plates (18 high biosorption outliers, 14 low biosorption outliers, and 32 non-outliers) and analyzed 3 biological replicates of each following similar procedures to our original screen and having each replicate in a new plate position. 62 of our mutants of interest grew enough to be able to analyze the data. We found that 53% (9/17) high biosorption outliers recapitulated the original screen and 31% (4/13) of low biosorption outliers. In total, 43% of the outliers recapitulated (13/30). 97% (31/32) of the non-outliers from the original screen remained non-outliers.

18 Gene ontologies are significantly enriched among genes influencing REE biosorption

Using Fisher's exact test, we analyzed each set of gene disruption mutants—those with higher or lower Eu-biosorption—for enrichment of gene ontologies, as defined by the Gene Ontology Consortium[55,56], to identify trends in the overall set of genes contributing to biosorption. We identified 18 gene ontologies that were significantly enriched ($p < 0.05$) and had more than one representative gene within our genetic screen results (Fig. 1C, Dataset S2, Materials and Methods). Ten were enriched among genes whose disruption increases biosorption while eight were enriched among genes whose disruption decreases biosorption. Ontologies discussed in this work whose gene disruptions increase Eu-biosorption include pyrimidine synthesis, polysaccharide synthesis, and histidine kinase activity ontologies, while one ontology discussed in this work whose gene disruptions decrease Eu-biosorption is pilus assembly.

13 Operons are significantly enriched in genes influencing REE biosorption

Using computationally produced predictions for operons in *S. oneidensis*[57,58] and Fisher's exact test, we identified thirteen operons with statistically significant ($p < 0.05$) enrichments of genes whose disruptions produced differential biosorption in the Arsenazo-III screen (Fig. S2, Dataset S3, Materials and Methods). Out of the thirteen, three operons had highly significant enrichment of hits ($p < 0.001$). Identifying enriched operons allowed us to further refine our analysis of which functions contribute to biosorption. For example, while 'polysaccharide synthesis' is highlighted as relevant from our ontology analysis, our operon analysis points us to two specific polysaccharide and O-antigen synthesis operons ('PS1' and 'PS2') that impact biosorption (Fig. 2A,B). Our operon analysis also specifies the MSHA pilus—previously highlighted more generally in the 'pilus assembly' ontology in our ontology enrichment analysis—as important to REE biosorption as 12 out of the 15 gene disruptions in this operon produced lower Eu-biosorption in our screen. Many, but not all, of the genes identified in these three operons are also highlighted in the corresponding ontologies.

Regulatory analysis of the arc system highlights several other genes important to biosorption

A disruption in the histidine kinase *hptA* was found to substantially increase biosorption in our genetic screen and *hptA* contributes to the enrichment of the 'histidine kinase' ontology group among hits. Since HptA is a regulatory protein, we speculated that its impact on biosorption must be caused by indirect activation or repression of downstream genes and thus sought to pinpoint the genes responsible.

HptA is part of the two-component Anoxic Redox Control or Aerobic Respiration Control (Arc) system in *S. oneidensis*. ArcS phosphorylates HptA in the absence of O_2, and HptA in turn phosphorylates the response regulator ArcA[59,60]. While we did not have gene disruptions we could screen for *arcS* or *arcA,* it is reasonable to assume that disruption of *hptA* produces a similar effect to disruption of *arcS*. We found that individual

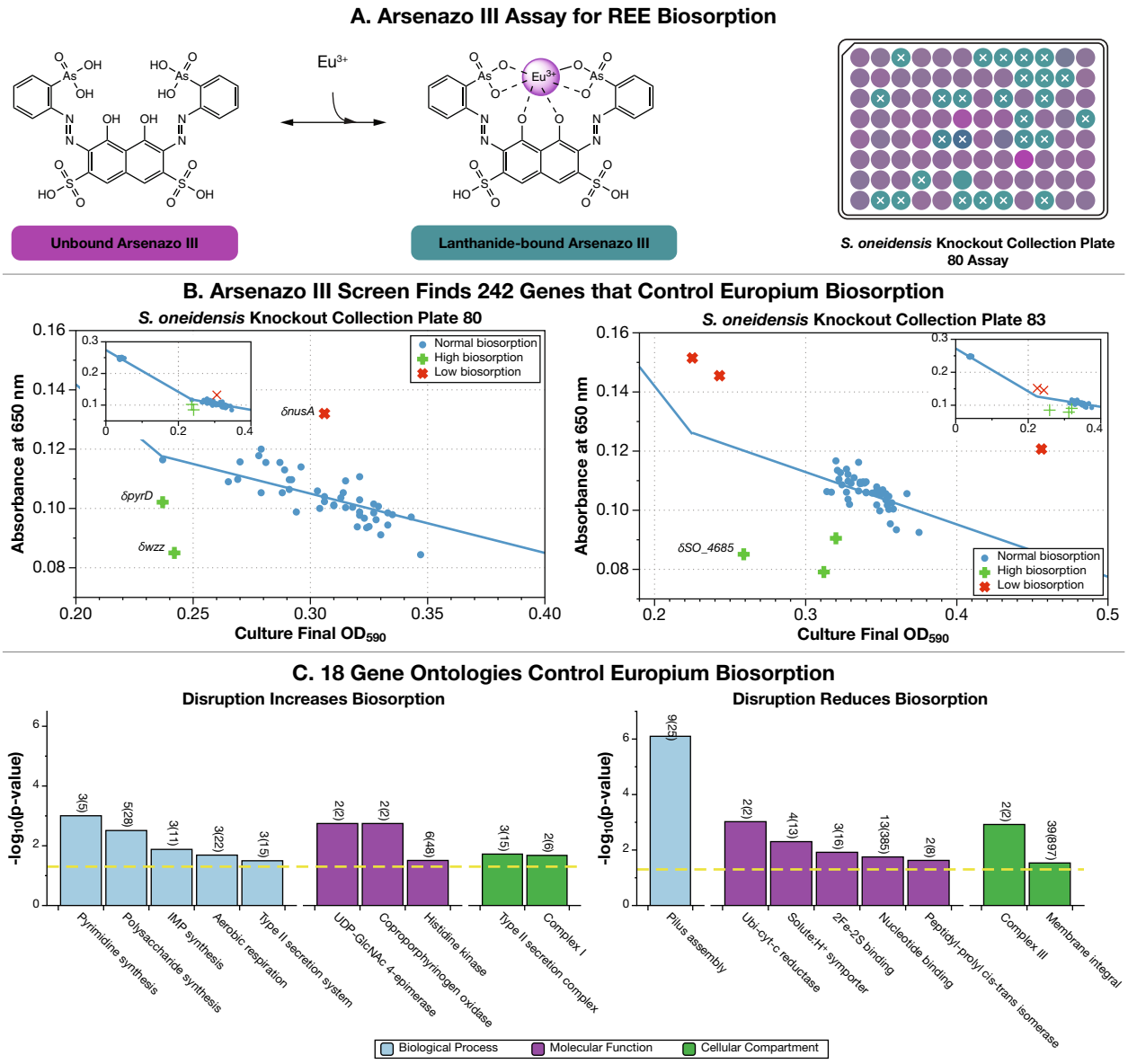

Figure 1. Screening the *Shewanella oneidensis* whole genome knockout collection finds 242 genes representing 18 gene ontologies that control Eu-biosorption. We used the Arsenazo III (As-III) competitive assay for europium- (Eu-) binding to screen 3,373 unique members of the *S. oneidensis* whole genome knockout collection to identify mutants with modified REE-biosorption capability. (**A**) Unbound As-III absorbance peaks at ≈ 530 nm (resulting in a cyan color), while Eu-bound As-III (proposed structure) absorbance peaks at ≈ 650 nm (purple). Right panel shows a computer-generated image of a sample assay plate derived from spectroscopic data. Higher biosorption by *S. oneidensis* results in a lower concentration of Eu-As-III and hence lower 650 nm absorption (the well will be more purple-colored) while lower biosorption results in a higher concentration of Eu-As-III (the well will be more cyan-colored). Additional information on the high-throughput screen is presented in Online Methods and Fig. S1. (**B**) The As-III screen found 242 genes that control Eu-biosorption (Dataset S1). The largest source of Eu-biosorption variability in the screen is due to bacterial density differences between mutants. For most mutants, the optical density of the culture at the start of the biosorption screen will map onto As-III absorption at 650 nm by a linear piecewise function (shown as a blue solid line). Mutants shown as red diagonal crosses had significantly less biosorption than the plate average. Mutants shown as green horizontal crosses had significantly higher biosorption than the plate average (mutants shown as blue dots are not significantly different from the average). (**C**) Gene ontology enrichment analysis found that 18 ontologies were enriched with mutants discovered by the As-III screen. The yellow dotted line indicates a p-value of 0.05. We only show results with p-values below 0.05 and gene ontologies with > 1 representative mutant. Numbers above each bar indicate the number of significant biosorption genes within each ontology in the screen results relative to the number in the *S. oneidensis* genome. Precise definitions of each gene ontology are shown in Dataset S2. IMP: inosine 5′-monophosphate; UDP-GlcNAc 4-epimerase: UDP-N-acetylglucosamine 4-epimerase; Ubi-cyt-c reductase: ubiquinol-cytochrome-c reductase.

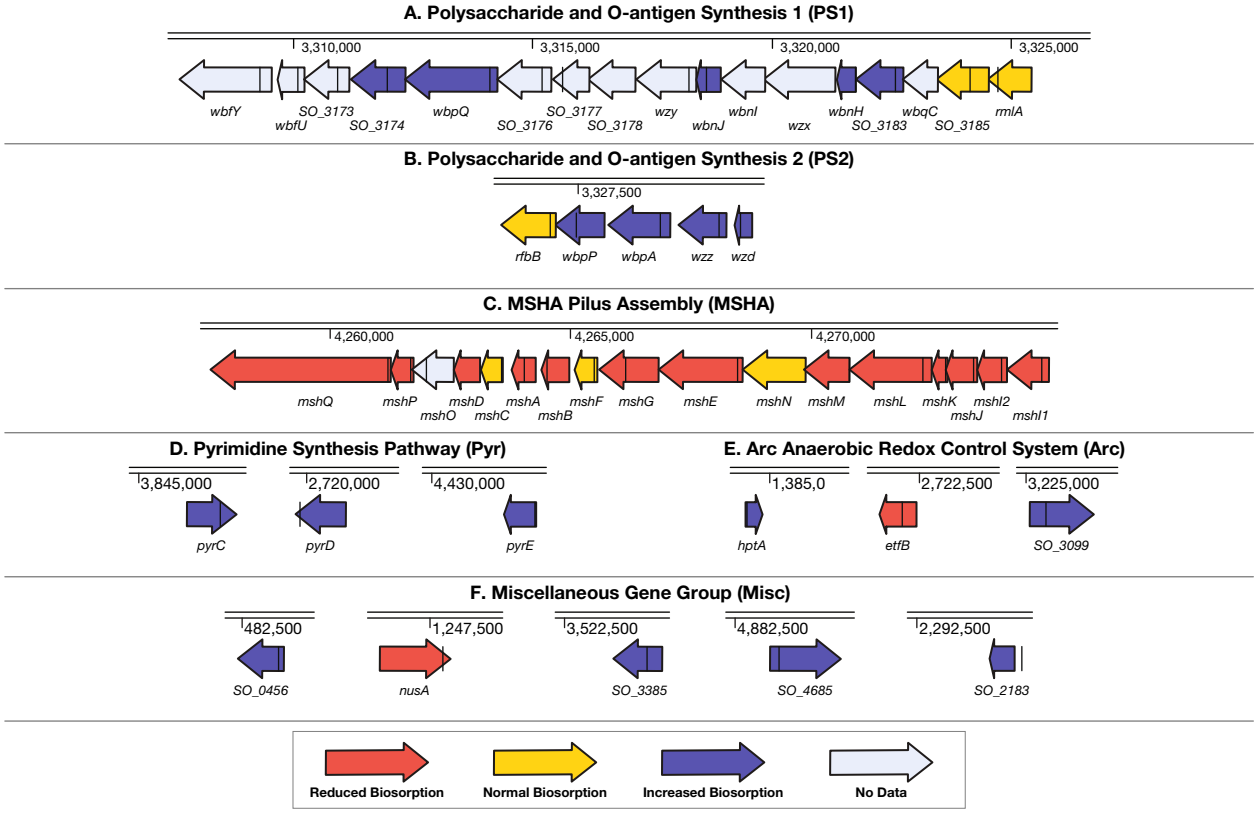

Figure 2. Operon enrichment, ontology enrichment, and regulatory analyses pinpoint 6 groups of genes that influence multiple mechanisms behind Eu-biosorption by *S. oneidensis*. (**A**–**C**) The results of the high-throughput Eu-biosorption screen (Dataset S1) of the *S. oneidensis* knockout collection were analyzed to find operons with statistically-significant enrichments of hits (Dataset S3). The location of the transposon disruption in each gene is marked as a black line. Here we show three operons that are the most statistically-significant results of this analysis. (D) The pyrimidine synthesis pathway was selected by ontology enrichment analysis (Dataset S2). (**E**) One gene involved in the Anaerobic Redox Control (Arc) regulatory system (*hptA*), as well as two genes regulated by Arc whose knockouts produced differential biosorption were also selected for further analysis. (**F**) Finally, five genes whose knockouts produced some of the largest changes to Eu-biosorption were also selected for further analysis.

disruption of 29 of the 604 genes whose activity is affected by an *arcS* deletion (and thus likely a *hptA* deletion as well)[59] produced significant changes in Eu-biosorption (Table S1).

Of the ArcS-regulated genes that may contribute to biosorption, several disruptions decreased biosorption (such as δ*etfB*, a disruption of subunit B of the e^- transfer flavoprotein[61]); several increased biosorption (such as δ*SO_3099*, a disruption of an outer membrane long-chain fatty acid receptor[61]); and several were also found to contribute to enriched ontologies (such as δ*pyrE* which is involved in pyrimidine biosynthesis[61]). ('δ' indicates a gene disruption mutant and 'Δ' indicates a gene deletion mutant. This notation was previously used by Rowe et al.[62] and Schmitz et al.[19]).

Six groups of genes that influence multiple mechanisms of REE biosorption were chosen for detailed analysis

Our As-III biosorption screen was limited because it could analyze only a single REE. We thus sought to expand our analysis by looking at how our gene disruptions impacted biosorption of multiple REE. We selected six groups of genes representing a wide range of cellular functions for detailed analysis with the Inductively Coupled Plasma Mass Spectrometry (ICP-MS)—an instrument for making robust measurements of concentrations of multiple elements—in order to validate the results we found with the As-III assay. We focused on groups of genes rather than individual outliers due to the uncertainty introduced by an $n = 1$ screen. Finding multiple genes with similar functions or with similar locations in the genome that impacted biosorption heightened our confidence that these genes were truly significant. We selected these groups based on gene ontology enrichment, operon enrichment, and regulatory analyses from our Eu-screen biosorption data.

For three of these groups, we selected operons of interest: polysaccharide and O-antigen synthesis operons 1 and 2 (PS1 and PS2; Fig. 2A,B) and the MSHA pilus operon (MSHA; Fig. 2C). Within the MSHA operon, we chose to look at δ*mshQ*, δ*mshD*, δ*mshC*, δ*mshA*, δ*mshB*, δ*mshL*, δ*mshJ* because these are all predicted to be either outer membrane proteins or found on the pilus appendages[63]. All of these, except for δ*mshC*, had significantly lower biosorption in the As-III Eu-biosorption screen.

An additional group of disruptions in non-contiguous genes (δ*pyrC*, δ*pyrD*, and δ*pyrE*) was selected for detailed study based on their contribution to the enrichment of the pyrimidine biosynthesis gene ontology (Pyr; Fig. 2D).

In addition to genes contributing to ontology and operon enrichment, we chose to analyze disruptions in genes from the Arc system, including a disruption of the Arc system histidine kinase δ*hptA* as well as disruptions of two genes regulated by the Arc system: δ*etfB* and δ*SO_3099* (Arc; Fig. 2E).

As a miscellaneous group, we selected five gene disruption mutants that were independent of any identified grouping but produced strong changes in biosorption (Misc; Fig. 2F). A mutant in the gene coding for the transcriptional termination/anti-termination protein NusA (δ*nusA*)[61] (also see Fig. 1B, left panel) had lower biosorption. Meanwhile disruption of *SO_4685*, which codes for a protein involved in extracellular capsular polysaccharide synthesis (CPS)[61] produced higher biosorption (also see Fig. 1B, right panel). Likewise, disruption of *SO_3385* which codes for a transcriptional activator of singlet oxygen protection[61] produced higher biosorption. Finally, an insertion 150 bp upstream of *SO_2183*, which codes for a protein involved in biosynthesis of the cell wall component peptidoglycan[61], had higher biosorption.

ICP-MS validates differential biosorption of genes from selected gene groups

ICP-MS measurements of mixed REE-biosorption largely validated the results of the high-throughput biosorption assay (Figs. 3A–D, Table S2). We measured biosorption of lanthanum (La; representing light REE), Eu (for middle REE), and ytterbium (Yb; for heavy REE) under four solution conditions (Materials and Methods) by 25 gene disruption mutants representing the six groups of genes chosen for detailed analysis (Figs. 3A–D) and six clean deletion mutants (Fig. 3E–H).

In industrial settings, REE are processed in a wide array of combinations and concentrations, and with a wide range of competing metal concentrations. We chose to explore two axes of interest to characterize our selected insertion mutants: low and high ionic strength (provided by sodium chloride); and (2) low and high total initial REE concentration. We tested multiple ionic strength concentrations to investigate the possibility of cation competition for REE binding sites and the ability of ionic strength to change protein configurations among other potential impacts. In short, we have four solution conditions: low ionic, low initial REE concentration (LL); low ionic, high REE (LH); high ionic, low REE (HL); and high ionic, high REE (HH). Biosorption solution conditions used are detailed in Table 1.

As a benchmark for comparison, we selected 4 quasi-wild-type (qWT) transposon insertion mutants from the *S. oneidensis* whole genome knockout collection that had the insertion in a location unlikely to impact biosorption (Materials and Methods). We found that the genuine wild-type *S. oneidensis* showed at least 13% higher biosorption than the average qWT in every solution condition (Fig. S4), indicating that the presence of a transposon insertion alone may affect biosorption. As we were interested in how knocking out the gene qualitatively affected biosorption rather than the absolute change from the wild-type, we thus chose to compare our notable transposon mutants with the average of our qWT mutants rather than the true wild-type.

Our multiple-REE biosorption assay recapitulated significant increases or decreases in biosorption from the As-III screen for 54% (13/24) of disruption mutants tested for the LL environment (Fig. 3A), 62% (15/24) for the LH (Fig. 3B) and HL (Fig. 3C) environments, and 42% (10/24) for the HH (Fig. 3D) environment. 79% (19/24 genes) of our insertions are validated in at least one environment (Fig. 3 and Table S2). We note that, as each of these conditions are different from the conditions of the genetic screen in initial REE content, solution composition, bacterial growth phase, and bacterial density, we did not expect every result from the screen to be recapitulated in these follow up experiments. Recapitulation of most of our results in at least some conditions, however, remains an encouraging validation of our initial screen. Furthermore, the percentage of insertions chosen by our operon and ontology enrichment method that were validated in at least one environment is much higher than the percentage of randomly chosen insertions that replicate (79% *vs.* 43%). This suggests that our enrichment analysis approach does improve the chances of finding insertions that produce robust changes to biosorption.

Disruptions to 9 genes produced higher biosorption in all environments: disruption of the uncharacterized protein Wzd (δ*wzd*) (Polysaccharide Synthesis Operon 1), disruption mutants for all 3 of the chosen Polysaccharide Synthesis Operon 1 genes, disruptions to all 3 genes related to the Arc system (including δ*etfB*, which had lower biosorption in the genetic screen), and an insertion upstream of the LD-transpeptidase encoding gene *SO_2183* and disruption to the CPS-synthesis gene *SO_4685* from the Miscellaneous group. In fact, δ*SO_4685* produced our largest observed increases in total biosorption ranging between 31% (in LL) and 79% (in HH) higher than the qWT.

The insertions for *SO_0456* and *SO_3385* (both in the Miscellaneous group) had significantly higher biosorption than the qWT in every solution condition except for HH.

Only one gene disruption produced consistently lower total REE-biosorption. Specifically, disruption of the last 10% of the coding region for the transcriptional termination/anti-termination protein NusA (δ*nusA*; the Miscellaneous group), produced our largest observed reductions in total biosorption ranging between 11% (in LL) and 35% (in HH) lower than the qWT.

Five gene disruption mutants showed a notable discrepancy in total biosorption between low ionic strength (Fig. 3A,B,E,F) and high ionic strength (Fig. 3C,D,G,H) environments. Most notably, the MSHA genes showed the greatest environment dependency in their biosorption results. For example, δ*mshJ* had significantly lower biosorption in the high ionic strength cases (HL and HH; Fig. 3C,D), but no significant change in the low ionic strength cases. Meanwhile, δ*mshA* and δ*mshB* had lower biosorption capabilities in the low ionic strength environments (LL and LH; Fig. 3A,B), and either no significant change or a biosorption increase in the high ionic strength environments.

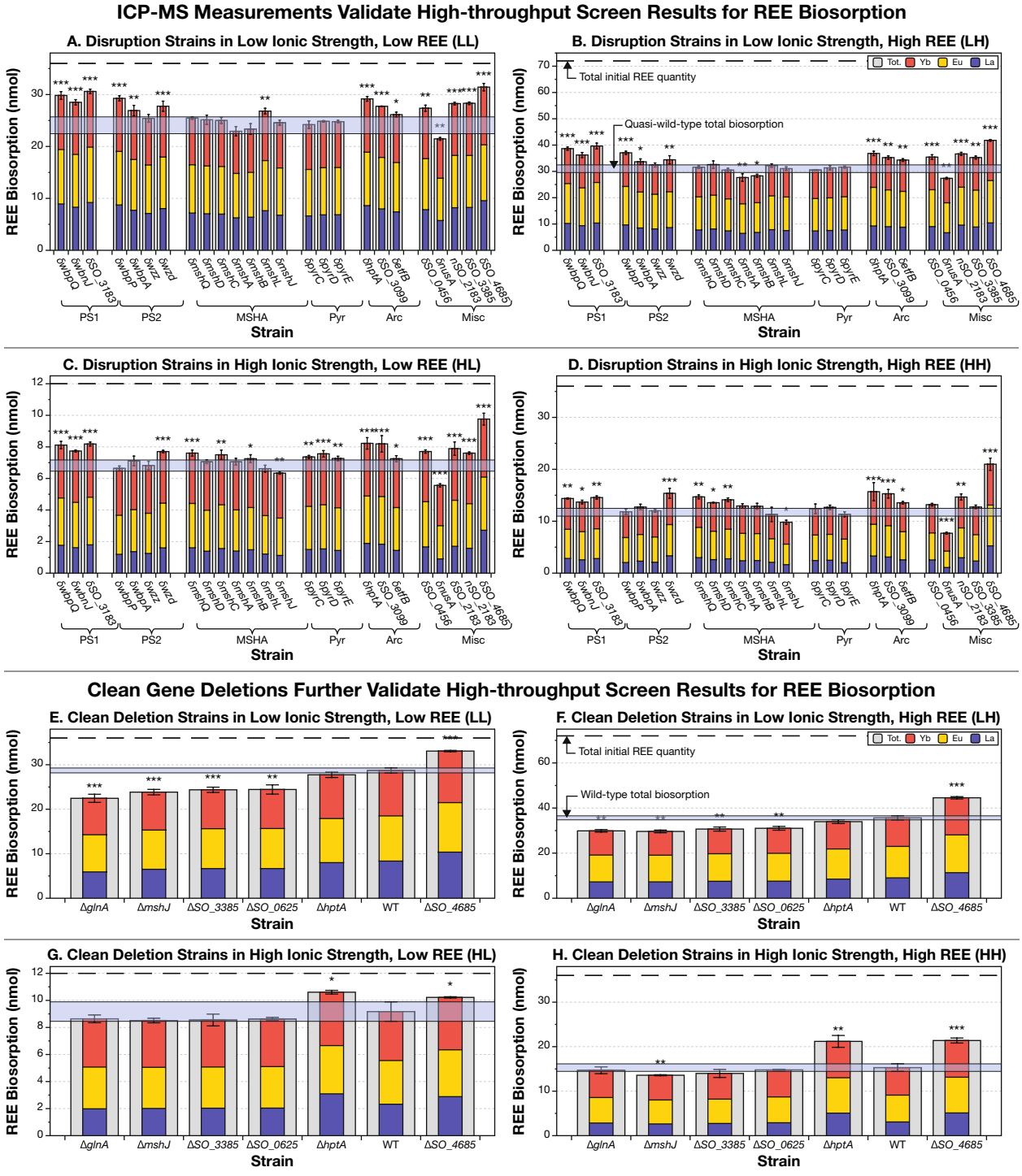

Figure 3. ICP-MS measurements validate the results of high-throughput Eu-biosorption screening in up to 79% of cases. Bar plots show levels of lanthanum (blue), europium (yellow), ytterbium (red), and total REE (grey) biosorption for each strain. The error bar indicates the standard deviation on the total biosorption of three biological replicates. The number of stars above each bar indicates the statistical significance of the measurement difference from quasi-wild-type (**A**–**D**) and wild-type (**E**–**H**): *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. δ indicates a transposon insertion mutant (in panels **A** to **D**), while Δ indicates a clean deletion mutant (in panels **E**–**H**). Note the *nSO_2183* mutant which indicates that the transposon is near to, but not within *SO_2183*. Cross-checks of As-III Eu-biosorption assay and ICP-MS measurements with transposon mutants are shown in Table S2. (A) The low ionic strength, low initial REE concentration environment matches 53% of the As-III screen (Table S2). (**B**) The low ionic strength, high initial REE concentration environment (LH) recapitulates the highest percentage (63%) of results of the As-III screen. (**C**) The high ionic strength, low REE environment (HL) reproduces 63% of significant changes to biosorption. (**D**) The high ionic strength, high initial REE environment reproduces the smallest number (42%) of results from the As-III screen. (**E**–**H**). Clean deletion mutants replicated at least some of the results of transposon mutant measurements in three of four cases.

Environment	Media composition	Initial REE concentration	Initial total REE quantity (nmol)
LL (Low Ionic, Low Initial REE)	10 mM MES + 10 mM NaCl	30 µM each La, Eu, Yb	36
LH (Low Ionic, High Initial REE)	10 mM MES + 10 mM NaCl	60 µM each La, Eu, Yb	72
HL (High Ionic, Low Initial REE)	10 mM MES + 50 mM NaCl	10 µM each La, Eu, Yb	12
HH (High Ionic, High Initial REE)	10 mM MES + 50 mM NaCl	30 µM each La, Eu, Yb	36

Table 1. Solution environments for detailed REE biosorption measurements. All solutions were adjusted to pH 5.5.

All three pyrimidine biosynthesis gene insertions had significantly higher biosorption (between 7 and 11% higher than the qWT) in the HL environment but registered no significant difference in any other environment (Fig. 3D).

Clean gene deletion mutants largely verify biosorption results for gene disruption mutants

We created clean deletion mutants for four genes whose disruption conferred standout biosorption changes: Δ*mshJ*, Δ*hptA*, Δ*SO_3385*, Δ*SO_4685*. Three of four of these clean deletion mutants at least partially reproduced the results of the corresponding disruption mutants. While insertion mutants are effective at knocking out gene function, they do not always successfully mimic a true single gene knockout. While the *S. oneidensis* knockout collection was designed to mitigate polar effects[52,64,65], it is possible that the early parts of the knockout gene can still create a partially functional product[66], which could account for the discrepancies between deletion and disruption strains. We also created two additional mutants for genes identified by the Arsenazo III genome-wide screen (Dataset S1): Δ*SO_0625* (a knockout for periplasmic cytochrome c oxidase regulatory protein) and Δ*glnA* (for glutamine synthetase).

Clean deletion of the pilus biogenesis gene *mshJ* (Δ*mshJ*) had significantly lower biosorption (11 to 17%) in three of the four environments tested (all except HL where the biosorption level was lower, just not statistically significantly). The main difference between the insertion and the clean deletion mutant is that Δ*mshJ* produces lower biosorption in the low ionic strength environments, while δ*mshJ* has no significant change compared to the qWT in those environments.

Clean deletion of the transcriptional regulator gene *hptA* (Δ*hptA*) produced large increases in biosorption (16 to 39%) in high ionic strength environments (HL and HH; Fig. 3G,H) just like the transposon mutant (δ*hptA*) (Dataset S1, Fig. 3C,D). However, unlike the transposon mutant, Δ*hptA* did not produce significantly different biosorption from the wild-type in low ionic strength environments (LL and LH; Figs. 3E and F).

Complete knockout of the capsular EPS biosynthesis gene SO_4685 (Δ*SO_4685*) showed a significant increase in biosorption (15 to 40%) in every environment just like the insertion mutant (Fig. 3).

Clean deletion mutants for *SO_0625* and *glnA* produced significantly lower biosorption than the wild-type in the low ionic strength cases, and non-statistically significantly lower biosorption in the high ionic strength cases.

Nine insertion mutants have notable modification of individual lanthanide binding preference

We next examined if, in addition to changes to total REE biosorption, our mutants produced changes to the relative biosorption affinity for particular REE over others. Out of the 25 insertion mutants we conducted follow up ICP-MS analysis on, nine of our mutants appeared to produce robust changes to biosorption preferences for particular REE.

Notably, our qWT already has a marked preference for heavier REE. This preference for heavier REE increases with ionic strength. From an initially equimolar mixed REE solution of La, Eu, and Yb, the qWT-biosorbed fraction contains between ≈ 19% and ≈ 28% La; ≈ 37% and 40% Eu; and ≈ 35% and 44% Yb (Fig. S4).

For most transposon insertion mutants, under most of the four solution conditions tested, individual REE biosorption is linearly related to total biosorption (Fig. 4, Materials and Methods) over a finite range (note the finite extent of dashed black lines in Fig. 4A,B).

However, every transposon insertion mutant that we tested had at least one REE biosorption result in at least one environment that deviated from the linear individual-to-total relationships established for most mutants under most conditions (Fig. 4A–C, and S4). It is notable, however, that having used a p-value of only 0.05 while performing 12 significance tests per mutant, our statistical test was not very stringent. We thus narrowed our criteria to select for 'robust' results by focusing on only those mutants with enhanced (or decreased) relative biosorption of a particular REE in the same direction in multiple environments or when a particular REE had enhanced or lowered relative biosorption (again, in the same direction) in more than half the genes in a particular group of interest.

Seven mutants produced robust results according to our criteria. These highlighted mutants are summarized in Fig. 4E.

Disruption of genes in the Polysaccharide Synthesis 1 operon tend to increase Yb-binding and decrease La-binding for high ionic strength environments (Figs. 4E, S6). For example, disruption of *SO_3183* increases relative Yb-binding under HH by 2.2% and reduces relative La-binding by 4.1%.

Disruption of two of the four genes in the Polysaccharide Synthesis 2 operon (*wzz* and *wbpA*) produces a significant increase in Eu and a significant decrease in Yb biosorption in the LH condition (Fig. S6). The disruption of *wbpA* has a significant relative increase in binding of Eu under the LL condition as well.

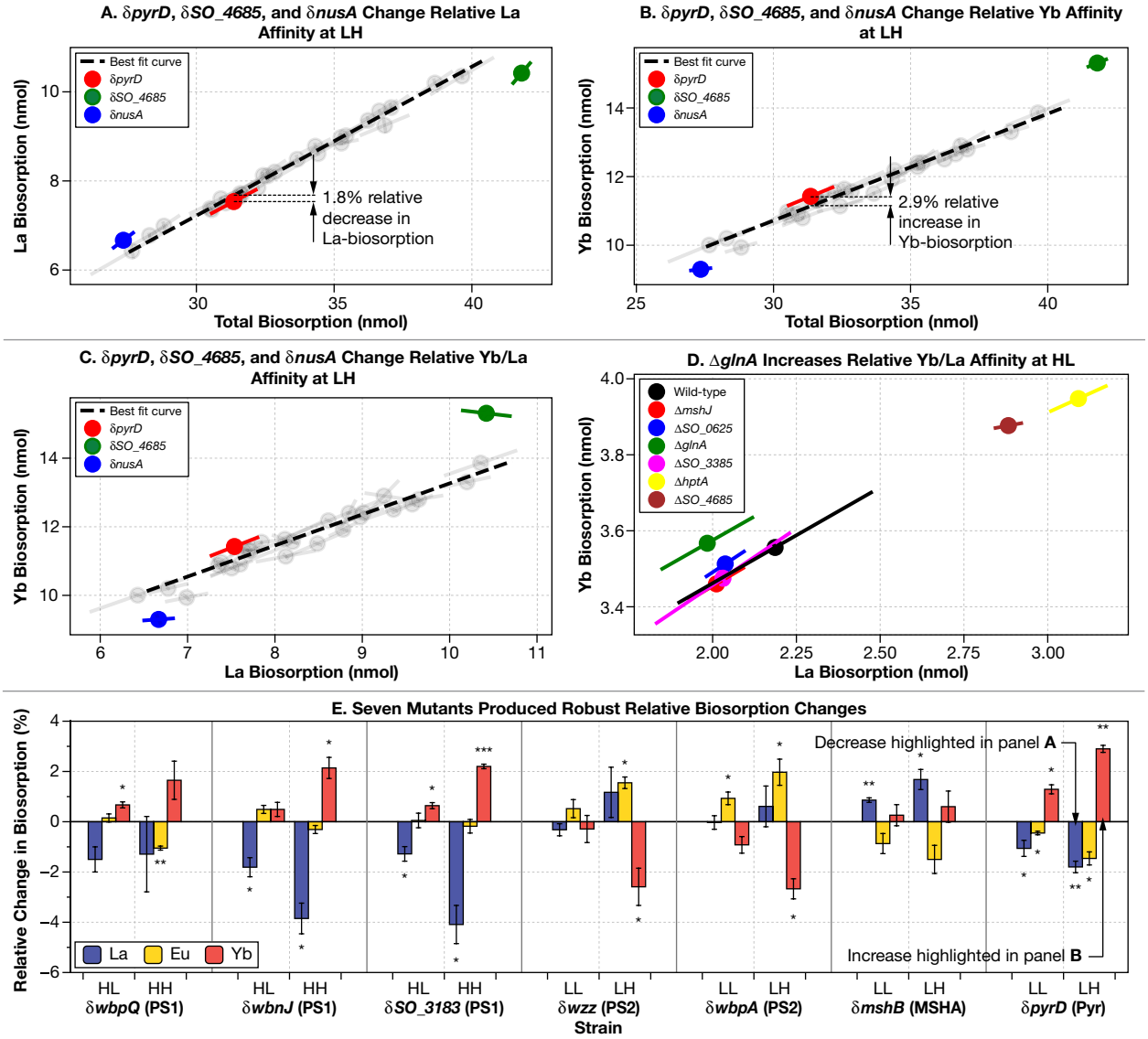

Figure 4. Nine gene disruption mutants make notable changes to REE-biosorption selectivity. (**A** to **C**) For most transposon insertion mutants (including those with modified total REE biosorption), under most of the four solution conditions tested, individual REE biosorption is linearly related to total biosorption or individual biosorption of either of the other 2 REE tested (grey circles, and the black dashed fit lines in panels **A**–**C**) over a finite range of REE biosorption (note the finite extend of dashed black lines in panels **A**–**C**). (**A**–**D**) Individual points indicate the mean values of the mutants and the error bars show the standard deviation along the axis of maximal variation. (**A**–**C**) We highlight changes in La and Yb affinity for δ*pyrD* in the LH environment as well as two mutants (δ*nusA* and δ*SO_4685*) who's total biosorption was too small (δ*nusA*) or large (δ*SO_4685*) to compare to our finite line of best fit, yet nonetheless clearly had different La and Yb affinity relative to the other mutants. In particular, note how both δ*nusA* and δ*SO_4685* have similar La biosorption to other mutants in (**C**) yet had very different Yb biosorption. (**D**) We had insufficient data to perform a line of best fit on our clean gene deletion data, yet it is clear from this plot that the *glnA* deletion has an increase in relative Yb/La affinity. (**E**) Here, we display all our mutants with robust biosorption changes (mutants that enhanced or decreased relative biosorption in multiple environments, or when a particular REE had enhanced or lowered relative biosorption in more than half the genes in a particular group of interest). The number of stars above or below each bar indicates the statistical significance of the measurement difference from quasi-wild-type: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. δ indicates a transposon insertion mutant, Δ indicates a clean deletion mutant. Error bars indicate standard deviation of three biological replicates. *PS1 and 2* Polysaccharide Synthesis 1 and 2; *MSHA* MSHA Pilus Assembly; *Pyr* Pyrimidine Synthesis.

Disruption of *mshB* (δ*mshB*) produces significant reductions in La-binding in low ionic strength conditions, likely at the expense of Eu-binding (although these changes are not significant).

Disruption of *pyrD* produces significant increases in Yb-binding coupled to reductions in La- and Eu-binding in both LL and LH. Under LH, the increase in Yb-binding of 2.9% is one of the largest significant increases in relative REE binding.

While disruption of *pyrC* under HH produces the largest significant change, increasing La-binding by 5.8%, it did not meet our robustness metric.

Two insertion mutants from the Miscellaneous group (δ*nusA* and δ*SO_4685*) have total biosorption levels that are so different from the rest of our insertion mutants that we did not include them in our formal analysis of relative REE changes (they were outside of the finite linear fit region in Fig. 4A–C). However, in the case of the LH environment, we still found a clear way of illuminating relative REE affinity changes (Fig. 4C). δ*nusA* produces very similar La-binding to other transposon insertion mutants that were in-range for our analysis. At the same time, it had a much lower relative level of Yb biosorption. This made it clear that δ*nusA* had relatively higher La-binding and relatively lower Yb-binding. Similarly, δ*SO_4685* had a similar level of La-binding compared to other in-range strains, but bound a much greater amount of Yb, implying a relative increase in preference for heavy versus light REE.

Discussion

Our genetic screen for biosorption reveals layers of the outer surfaces (inner and outer membranes, and periplasmic layer) of *S. oneidensis* that modulate access to REE-binding sites in *S. oneidensis.* These layers include polysaccharides (synthesized by Polysaccharide Synthesis Operons 1 and 2), MSHA pili (synthesized and assembled by the MSHA Pilus Assembly Operon), and a variety of outer membrane proteins (SO_0456, SO_3099, MshQ, MshL, MshJ).

Disruption of polysaccharide synthesis operon 1 raises biosorption

Many of the genes coded by PS1 are responsible for the synthesis of O-antigens, a major component of the lipopolysaccharide (LPS) layer on the outer membrane of *S. oneidensis*[67]. Disruption of all genes selected for further analysis in PS1 increase total biosorption under all solution conditions (Fig. 3) and generally increase relative Yb-binding and decrease relative La-binding in high ionic strength conditions (Fig. 4E). Among the three genes tested in the PS1 group, only δ*SO_3183* is directly implicated in polysaccharide synthesis (it was highlighted in our gene ontology analysis), although the similar biosorption effects of each of our three mutants seem to suggest that they may all be part of a single pathway.

O-antigens themselves have not been directly implicated in REE-biosorption, but it has been theorized that phosphate groups on LPS components below the O-antigens, such as Lipid A, could be responsible for some REE-binding[28]. It is possible that knocking out these genes eliminates certain O-antigens, exposing these phosphate groups on the membrane. Since phosphate groups tend to have a stronger affinity for heavier REE, this could explain why there was a relative increase in biosorption of heavier REE.

Disruption of polysaccharide synthesis operon 2 modifies the cell membrane and REE biosorption

The disruption mutants selected for in-depth analysis belonging to Polysaccharide Synthesis Operon 2 (PS2) generally cause significant increases in REE-biosorption in at least some cases, although the results were not necessarily consistent from gene to gene. For example, while δ*wbpP* and δ*wbpA* significantly raise biosorption only in the low ionic strength cases, δ*wzd* increases biosorption in every case and δ*wzz* fails to significantly alter total biosorption at all. We suspect this is due to each of the genes in this group having a unique impact on biosorption.

We speculate that disruption of WbpP (δ*wbpP*) raises REE-biosorption in low ionic strength conditions due to its significant role in membrane composition. WbpP transforms UDP-N-acetyl-D-glucosamine to the UDP-N-acetyl-D-galactosamine[68]. In *P. aeruginosa*, WbpP plays a role in the synthesis of B-band O-antigens, a component of the lipopolysaccharide layer[69]. In *V. vulnificus*, deletion of *wbpP* causes the failure of CPS (capsular polysaccharide) formation[70]. The deletion also results in increased cell aggregation, hydrophobicity, and adherence to abiotic surfaces, all suggestive of substantial membrane changes[70]. Without more *S. oneidensis* specific data, it is impossible to know the exact mechanism for δ*wbpP* having increased biosorption in only the low ionic strength cases. However, one possibility could be that, like in *P. aeruginosa*, B-band O-antigens are deleted. Consequently, binding sites that are normally covered by those O-antigens could be revealed that bind to REE only in low ionic strength environments.

Likewise, disruption of *wbpA* (δ*wbpA*) also raises biosorption in low ionic strength conditions and plays a role in membrane composition. WbpA, like WbpP, takes UDP-N-acetyl-D-glucosamine as a substrate (but transforms it to uronic acid[68] instead), and is thought to be a key protein in O-antigen biosynthesis in *P. aeruginosa*[71].

MSHA genes have highly environmentally dependent effects on biosorption

The importance of the interaction of the solution environment with gene disruption on biosorption is most strongly illustrated by disruptions of the MSHA Pilus Assembly Operon (MSHA) genes. While 6/7 of the gene disruptions tested (all except δ*mshC*) had lower biosorption in the original screen, only 3/7 had significantly lower biosorption in any of the solution conditions selected for follow up testes and none of them had significantly lower biosorption for every solution condition.

MshA is responsible for forming the main subunit of the pilus and knocking it out thus has a major impact on the MSHA pilus[72]. δ*mshA* had significantly lower biosorption in the low ionic strength conditions. This seems to suggest that the MSHA pilus plays an important role in binding to REE in low ionic strength cases. At the same

time, δ*mshA* has no significant change in biosorption in the high ionic strength conditions, suggesting that the high NaCl concentration is preventing REE from binding to the pili.

Disruption of pyrimidine synthesis group increases REE-biosorption under high ionic strength, low REE conditions

Disruption mutants of the Pyrimidine Synthesis Group genes *pyrC*, *pyrD*, and *pyrE* all increase biosorption in the high ionic strength, low REE condition. These genes form a section of a pathway in pyrimidine metabolism that produces Orotidine 5′-monophosphate (OMP) from carbamoyl aspartate. The exact mechanism through which these genes effect biosorption remains uncertain and require further investigation.

Binding site changes from single gene knockouts tend to have multiple effects

While a handful of gene disruption mutants we looked at had consistent biosorption changes across every condition tested, for many of our other genes, changes in biosorption levels were inconsistent across different solution conditions. Some of these discrepancies have simple explanations. For example, several gene insertion mutants (such as polysaccharide synthesis protein δ*wbpP*) had higher biosorption for low ionic strength, but similar biosorption to the qWT for high ionic strength. In the case of δ*wbpP*, it is possible that this is because elimination of *wbpP* results in increased accessibility to binding sites that are capable of binding to REE in low ionic strength cases but not high ionic strength cases—possibly due to competition with sodium ions for binding sites. Thus, knocking out *wbpP* increases REE biosorption in the low ionic strength cases, but has no effect in the high ionic strength cases.

Some other features of our data necessitate more complicated explanations. Most prominently, not a single strain had statistically significant and consistent changes in relative affinity for individual REE across all environments.

Intuitively, one might expect that most gene insertion/deletions that affect biosorption would affect a single binding site. That the gene would encode a single outer membrane protein or a protein that alters some single REE binding lipid or polysaccharide found on the membrane. Based on our genetic screen results, this appears to rarely be the case. More often, gene knockouts likely cause a cascade of effects on other genes resulting in changes to multiple binding sites. Gene disruptions can also have confounding effects unrelated to REE binding site composition. It is possible that some gene disruptions impact the shape of the bacteria or perhaps the optical density to bacteria ratio (since the optical density is what we use to normalize the bacterial density from assay to assay). Perhaps some of these gene disruptions affect secretions of *S. oneidensis* that might interfere with biosorption by competing with surface binding sites[40]. Alternatively, a mutation could alter the OD:surface area ratio, making it an artificial outlier in the REE-biosorption screen without truly affecting biosorption. Finally, earlier works have found that growth phase can impact metal biosorption[73]. A gene knockout might not affect membrane composition but instead introduce a growth defect that affects the bacterial growth phase, and hence biosorption, at the sampled optical density. Thus, while our work may answer the question of what genes are important to biosorption, the reasons why they are important almost uniformly requires further investigation.

Changes to lanthanide preference meaningfully improves REE purification process

To meaningfully understand the size of the lanthanide preference changes created by the mutants we have isolated, we created a simple model to get an order of magnitude estimate for how these preference changes affect a lanthanide purification process (Figs. S7 and S8, Table S3, Note S1). In Fig. S7 and Note S1 we present a simplified system for lanthanide enrichment that uses repeated biosorption and elution. A mixed solution of REE is run through a column and allowed to bind to immobilized bacteria. After equilibration the free fraction (the liquid) is removed from the column and moved to a wash collection container. Next, the bound fraction is eluted (for example by a pH swing[33,74]). The eluant is reloaded (or sent to another column) for further purification. In our model, we only follow the biosorbed/eluted REE solution at each stage of the process. We design our system such that 50% of our REE are biosorbed to the bacteria at each stage (Note S1). We also assume that there is only a single binding site type on our bacteria (Note S1). While our results in this article make it clear that this assumption is not entirely realistic—we would need a lot more data to simulate a realistic enrichment process with multiple binding sites,—we feel this assumption will be effective in putting our separation improvements in context.

We compare the enrichment process for one of our strains, δ*wbpA*, and our calculated baseline (Fig. S8 and Table S3). The baseline is what an average transposon insertion strain's REE biosorption distribution would look like if it biosorbed the same total amount of REE as δ*wbpA* (see Fig. 4E for details about this baseline). We note that, since there are apparently multiple binding sites on each strain, we cannot authoritatively say what the separation factor between the REE will be as the REE distribution changes. We chose to look at δ*wbpA* (low ionic strength) as it has similar separation factor improvements in environments with different starting REE concentrations.

We calculated Eu enrichment from an equimolar solution of Eu, Yb and La by δ*wbpA*, using measured separation factors from the low ionic strength solution conditions (LL and LH). Despite the two conditions (LH and LL) having different percent increases in europium biosorption (1% and 2% respectively), we obtained similar results. For the LL condition, we found δ*wbpA* produced a 27% reduction in the number of steps (30 for the baseline, 22 steps for the mutant) compared to the baseline required to produce a 99% purity europium solution. In the LH condition, δ*wbpA* produced a 25% reduction (28 for the baseline, 21 for the mutant) in the number of steps to 99% purity (Table S3).

Conclusions

We have conducted one of the most comprehensive screens of the genetics of REE-biosorption (or any element for that matter) to date. The genes and gene clusters we discovered to be involved in Eu-biosorption are directly involved in membrane composition and gives us and fellow investigators an extensive catalog of entry points for further investigation of REE-biosorption.

At the outset of this work, we anticipated that the genetic screen of REE-biosorption would identify a suite of genes encoding individual REE-binding sites on the surface of *S. oneidensis*—either in the form of outer membrane proteins or proteins that create compounds found on the outer membrane. The reality is much more complex.

As expected, many of the gene disruptions that affect total REE-binding and modify the REE-binding preference of *S. oneidensis* likely have functions that directly affect the outer surface. Both Polysaccharide Synthesis Operon 1 and 2 are involved in making O-antigens on the lipopolysaccharide layer. The MSHA genes synthesize pili on the outer membrane of *S. oneidensis*. Finally, two members of the Miscellaneous Group are related to other outer membrane structures: SO_2183 is involved in synthesis of the peptidoglycan layer and the SO_4685 protein is involved in synthesis of capsular polysaccharides.

Two individual gene disruptions stood out in their impact on total biosorption. In every solution condition, disrupting *SO_4685* resulted in the highest biosorption observed, while disrupting *nusA* resulted in the lowest biosorption observed.

Furthermore, we identified 10 gene disruptions or deletions that affect the biosorption selectivity preference. Despite the apparently small size of changes to biosorption preference (≈ 1 to 4%; Fig. 4) caused by disruptions to single genomic loci, these changes might produce large reductions in the length of a repeated enrichment process for individual lanthanides (Fig. S8). For example, in a simplified model of *S. oneidensis* binding (Fig. S7), the 2% increase in europium binding for δ*wbpA* resulted in up to a 27% reduction in the number of enrichment steps needed to reach 99 and 99.9% purity (Fig. S8, Note S1, Table S3). It is possible that a knockout of *wbpA* plus up-regulation of *pyrD* produced by swapping its endogenous promoter for a stronger version (*e.g.*, J23100 or Tet promoter[75]) could produce a mutant with even better separation characteristics.

The key result of the screen, however, is that there appears to be few, if any, dominant players in REE-biosorption in *S. oneidensis*. Instead of one dominant type of binding site like teichoic acids in *Bacillus subtilis*, many different types of sites, encoding structural features ranging from pili to lipopolysaccharides to outer membrane proteins, contribute to the adsorption of REE to the *S. oneidensis* membrane. This suggests that making more substantial improvements to REE selectivity on the membrane might require many parallel edits to *S. oneidensis*' genome.

Our genetic screen results provide a roadmap for creating strains of *S. oneidensis* with improved binding of individual REE. With this new knowledge of key biosorption genes, we could use multi-site genome engineering tools (*e.g.*, MAGE[45]) to mutate regulatory and coding regions of genes identified in this study to alter the selectivity of *S. oneidensis*' membrane for REE. Iterative applications of this process would not only help with creating *S. oneidensis* strains with increased affinity for specific REE but would also further refine our knowledge of which genes are most important to REE biosorption and refine gene targets for subsequent rounds of mutagenesis. Perhaps a very good mutant might have 10–50 genes that control the membrane composition downregulated or deleted (including possibly *wbpA*) and 10–50 or more genes upregulated (possibly including *pyrD*). We believe that it would be very difficult to both discover the ideal combination and engineer this with traditional genetic engineering strategies, but targeted multiplexed mutagenesis could produce a mutant like this with a few rounds of mutagenesis.

This genetic study advances our knowledge of the mechanisms responsible for REE biosorption in *S. oneidensis*. Our study points to key genes that control membrane composition that are traditionally responsible for cell adhesion and biofilm formation in REE biosorption. This open doors to many areas of study for future investigators seeking to better understand the mechanics of REE biosorption. Furthermore, this knowledge will allow us to engineer bacterial cells with greater affinity for individual REE in order to produce a new, environmentally benign method of extracting and separating REE from other elements and each other.

Materials and methods

Media preparation

Bm20 media

Bm20 is composed of 20 mM MES buffer, 8.6 mM ammonium chloride, 0.5 mM magnesium sulfate, 1.7 mM ammonium sulfate, and 5 mL/L Trace Mineral Supplement (ATCC). The media was adjusted to pH 5.5 with 1 M NaOH.

Genome-wide REE biosorption screen

Introduction

We screened the *S. oneidensis* whole genome knockout collection[51,52] for biosorption of the rare earth element (REE) europium. We hypothesized that using europium, which lies in the middle of the REE size range, would allow us to maximize the number of genes we could discover.

We screened a subset of the knockout collection comprising 67 96-well plates that covers 3,472 unique genes (Dataset S1) (the *S. oneidensis* knockout collection is comprised of ≈ 50% blank wells). Due to the failure of some mutants in the collection to grow, we screened 3,373 unique genes in total. Cross-contamination did not appear to be an extensive problem as, on average, we had only 1.6 contaminated wells (wells that were not supposed to have bacteria in them yet grew anyway) per plate. The maximum number of contaminated wells in a plate was 10. If we assume that the filled-to-filled cross-contamination rate is the same as the filled-to-blank rate, then

we get a total contamination rate of 1.6/48 = 3.3%. To put this into context, the mis-labeling rate on the original gold standard Keio whole genome knockout collection for *E. coli* was 4%[76]. Second, given that most mutants (3373 – 242 = 3131 of them) have a similar biosorption phenotype, cross-contamination is far more likely to mask mutants with an outlier biosorption phenotype than create an artificial hit.

The genetic screen was conducted over the course of several weeks and was divided into batches, each of which took two days to process. Typical batch sizes were between 4 and 8 plates.

Replication of Shewanella oneidensis whole-genome knockout collection

Microplates from the *S. oneidensis* knockout collection were replicated from a master collection (stored at −80 °C) with a pin-tool replicator (EnzyScreen CR1000) into a flat-bottom polypropylene plate (Part no. 655261, Greiner) containing 150 µL of LB media per well with 30 mg L^{-1} Kanamycin. Newly inoculated plates were incubated at 30 °C shaking at 800 rpm in a high-throughput microplate shaker (Infors Multitron Pro) for between 16 and 20 h. The following morning, 3 µL of culture was transferred to a new plate containing fresh LB with 30 mg L^{-1} Kanamycin. The newly diluted cultures were grown for 5–6 h.

Biosorption assay

Upon removing our plates from the incubator, we diluted 40 µL of culture from each plate into a new round-bottom polystyrene 96-well plate (Part no. 650101, Greiner) with 150 µL of bm20 media (see Media Preparation). After this transfer, we took an optical density (OD) measurement of our plate with a plate reader (Biotek Synergy 2) at 590 nm. We called this OD the "growth OD" because it was proportional to the final OD that the bacteria grew to.

After taking the OD, we centrifuged each plate in a swinging bucket centrifuge with micro-well plate adaptors (Eppendorf 5810R) at a speed of 3214 × *g* for 7 min. After that, we positioned a 96-well pipette to remove supernatant from the edges of the wells of the plate. We removed as much supernatant as possible because components of the growth media (data not shown) can bind to rare earth elements and interfere with biosorption. We then rinsed the bacteria by adding 170 µL of bm20 to each well and resuspended by shaking. We then repeated the centrifuging and removing of supernatant steps. After these two rinsing steps, the concentration of growth media components, including the kanamycin, should have been negligible.

We added 200 µL of bm20 with approximately 25 µM of europium to the rinsed *S. oneidensis* cells and resuspended by vortexing. We then shook the plate in the plate reader or in a plate vortexer for 10 min while biosorption occurred. Previous research has shown that bacterial cells tend to reach full biosorption capacity within 10 minutes[29,40]. The shaking also helps to prevent bacterial aggregation. Upon the completion of the shaking step, we took another OD measurement (called the "final OD") to have an estimate of how many bacteria were present in each well for the biosorption assay.

We centrifuged each plate once more at a speed of 3214 × *g* for 10 min. We then transferred 100 µL of the supernatant to a new flat bottom polystyrene plate (Greiner bio-one ref: 655101). We used the plate reader to take a spectrum of absorbance from 580 to 680 nm in increments of 10 nms. These are our "blank" measurements. We then added 100 µL of a solution of 60 µM concentration of Arsenazo III dye dissolved in pH 3.5 20 mM MES buffer to each well with the 96-well pipette. Arsenazo III has been established to be a reliable indicator for free REE concentration at this pH[54,77]. We shook the plate for 4 min and then we used the plate reader to measure the absorbance in the same wavelength range as the blanks. These absorbance measurements served as a proxy measurement for the concentration of rare earth elements that were not adsorbed by the bacteria.

Analysis of genome-wide REE biosorption screen

Challenges of identification of mutants with differential biosorption

Optical density and As-III absorbance measurements were corrected to account for blemishes on the exterior of assay plates, especially due to centrifugation. We knew that it was impossible for our final optical density after rinsing the bacteria to be higher than the optical density prior to rinsing. We thus took our "final" optical density measurement to be the minimum of the optical density prior to rinsing and the optical density after we have finished preparing our biosorption assay. Likewise, As-III absorbance measurements were background corrected by subtracting a blank measurement taken prior to As-III addition.

Non-uniform growth is the biggest challenge in identification of mutants with genetic modifications that affect REE biosorption. Due to the dynamics of ligand-receptor binding chemistry, we would expect that cultures with higher densities will produce higher overall biosorption but have lower biosorption per cell.

We used deviations from a linear piecewise relationship between optical density at 590 nm (OD_{590}) and As-III absorbance to identify mutants with truly differential biosorption (Fig. 1B). The OD to As-III absorbance function varied from plate to plate, likely due to slightly different growth conditions. As a result, we treated every plate separately when identifying outliers.

The linear piecewise function relating OD_{590} and As-III is made of 2 sections. For $OD_{590} \gtrapprox 0.2$, As-III absorbance was linearly related to OD_{590}. For $OD_{590} \lessapprox 0.2$, we took advantage of the large number of blank wells in each 96-well plate to measure As-III absorbance with no biosorption. Thus, for the earlier part of the piecewise function, we fit a line between a point marking the average OD_{590} and As-III absorbance of blank wells and the point with the smallest OD_{590} in the set used to create the second part of the function. We found that most points in each plate generally fell along our piecewise function.

To find biosorption outliers, we calculated the standard deviation of the distances of each point from the linear piecewise function for each plate. We marked any point as significant if it had a distance of more than 2 standard deviations from our piecewise function (Fig. 1B).

Arsenazo III assay quality control

Quality-control procedures were used to flag low-quality data during analysis. Our procedures specifically looked for anomalous optical density measurements and absorbance spectra patterns.

Plots of final optical density versus growth optical density form a compact line (Fig. S1A). Datapoints where the final OD_{590} was more than three standard deviations above the average value were flagged. Datapoints where the final OD_{590} was below average were not flagged, as these indicated either that the growth OD_{590} measurement was off (which would not affect the final analysis) or that there was a large loss of bacteria during rinsing (which also would not affect the final analysis).

Arsenazo III absorbance measurements were quality-controlled by ratiometric analysis. Errors in the As-III absorbance measurement can arise due to errors in pipetting the As-III stock, or by the presence of dust or scratches on the assay plate surfaces. For the range of As-III absorbance measurements observed in our assay, the relationship between the 650 nm absorbance and the 680:650 absorbance ratio was linear (Fig. S1B,C). Extreme outliers could occasionally skew this analysis, so we eliminated the five data points furthest from the original linear fit and created a second line of best fit. Data points that lay more than three standard deviations away from the line of best fit were flagged.

Manual inspection was used as the final step in quality control resulting in the selection of 240 mutants for further analysis. Each of the 294 data points that had outlier biosorption measurements in the screen were manually examined. We paid special attention to the datapoints flagged with a high final OD_{590} or with an anomalous absorbance spectrum. We eliminated 12 datapoints that were listed as blank in the *S. oneidensis* knockout collection catalog, but that displayed cross-contamination. We eliminated 7 datapoints that lacked transposon location information. Finally, we removed strains where we judged that the OD_{590} to As-III absorbance piecewise function did not provide a reliable estimate for significance due to lack of surrounding data.

Gene ontology enrichment analysis

We followed the gene ontology enrichment analysis procedures laid out in Schmitz et al.[19], except we did not use InterProScan to collect gene ontology data. In brief, we used DIAMOND[78] to assign annotated protein models with the closest BLAST hit using the Uniref90 database (downloaded from https://www.uniprot.org/uniref/), an E-value threshold of 10^{-10}, and a block size of 10. We used the output of this search to assign gene ontologies with BLAST2GO[79].

We performed a gene ontology enrichment analysis with the BioConductor topGO package[80] using the default weight algorithm, the TopGO Fisher test, with a p-value threshold of 0.05.

We performed separate gene ontology enrichment analyses for mutants with significantly higher and lower biosorption (Fig. 1C, Dataset S2). Following the transposon mutant collection screen, we found that Δ*SO_0625* and Δ*glnA* both produced lower Eu-biosorption and added them to the lower biosorption category.

Operon enrichment analysis

Operon enrichment analysis was used as a complement to ontology enrichment analysis to identify groups of genes involved in REE-biosorption (Figs. 2, S2, and Dataset S3).

Operon memberships in the *S. oneidensis* genome were predicted by the union of results from operon predictions by MicrobesOnline[57] and ProOpDB[58]. In most cases, the two sources produced highly similar results.

We used Fisher's exact test to calculate if the set of genes whose disruption conferred differential biosorption was enriched in operons with more than one hit. We compared the total number of gene disruptions that significantly affected biosorption (out of the total number of genes assayed in our genetic screen) to the number of gene disruptions within the operon that significantly affected biosorption (out of the total number of genes we looked at within that operon in our genetic screen). Fisher's exact test was conducted with the fishertest function in MATLAB.

We also applied Fisher's exact test to calculate if the sets of gene disruptions that either increased or decreased biosorption were enriched in operons that had more than one gene disruption that increased or decreased biosorption, respectively.

Spot check of mutants identified in whole genome screen

We sought to test how well our screen did at correctly identifying biosorption outliers in our screen. To this end, we re-screened a subset of both mutants we identified as outliers in our original screen and ones we identified as non-outliers. We chose this subset from plates seven through eleven of the collection. We analyzed all 32 outlier strains that appeared in these plates and 32 random non-outlier strains. We checkerboarded the strains (one blank well between every strain) and had three replicates for each strain. The replicates were spread out onto different areas of the plates.

The assay itself was carried out almost identically to the original screen. The only difference is that after replicating plates seven through eleven, we needed to re-array mutants of interest onto new plates. This was done using a Norgren Colony Picker. Besides this extra growth step, no other changes were made. We note that none of the blank wells had contamination in the final assay.

Analysis of the data was also very similar to the original screen. Instead of fitting a line of best fit to all the mutant data, we only fit the line of best fit to the non-outlier points. We also slightly modified our method of looking for absorbance spectrum outliers. We removed the manual checks of the spectrum because we felt that, for a sanity check such as this, we should use uniformly applied criteria. We chose to instead look for absorbance spectra outliers by comparing the 650 nm wavelength to the ratio of the 660 nm and 650 nm. We did this because, in our manual absorbance analysis for the original screen, many of the ones that looked the most off had a noticeably irregular 660/650 nm ratio suggesting that perhaps we should have used this ratio to begin with.

Because we had three replicates instead of one, we also modified our method of determining which mutants were significant. In the original screen, we required a significance of 2 standard deviations from the mean. In our new screen, we required that at least two replicates had 1 standard deviation from the mean in the same direction. At first, this may seem like a less stringent criteria since we'd expect a normal distribution of data to have non-outlier points meet this criterion 17% of the time. Our points do not, however, follow a normal distribution. We believe that we occasionally have noisy points that contribute significantly to the standard deviation. We usually don't have more than one of these high-noise points per strain. Requiring only two points allows us to deal with the cases where a strain has a single high-noise point. The use of 1 standard deviation instead of 2 is also validated by the very small number of non-outliers in the original screen appearing to be outliers in this spot check. If we truly had 17% of strains off, we would expect five-six points to appear to be outliers instead of the single one we have.

Confirmation of transposon mutant identity

We validated the identity of transposon mutants from the *S. oneidensis* knockout collection that we conducted follow up analyses on using site specific PCR. The verification reactions used a common primer that bound to the Himar transposon, and a mutant specific primer that bound to the genomic region predicted to be outside the transposon. The identities of all but one transposon mutant was correctly predicted in the *S. oneidensis* knockout collection catalog. The single mutant that was mis-identified was originally annotated as δ*arcA* (δ*SO_3988*), but later found to be 147 bp upstream of *SO_2183*.

Note on glnA and SO_0625

The 67 plates that we screened and conducted our analyses on contains mutants whose transposon locations are well characterized by double sequencing verification. This is validated by only a single mutant out of 25 tested having the insertion in an unexpected location. The collection, however, has 10 additional plates with mutants whose transposon locations had some ambiguity. We screened these plates and identified one mutant that we found to be a particularly large outlier in biosorption. The location for the transposon in this mutant was determined be in either *glnA* or *SO_0625*. We decided to make clean deletions for these two genes instead of isolating the transposon mutants.

Construction of gene deletion mutants

Clean deletion mutants were constructed to validate the results of transposon screening for *hptA*, *SO_4685*, *mshJ*, and *SO_3385* as well as for *glnA* and *SO_0625*. Deletions were made by homologous recombination using a suicide vector containing a kanamycin resistance cassette flanked by 1000 bp upstream and downstream sequences surrounding the gene of interest. Mutants that had undergone a second recombination (removing the gene of interest and the kanamycin cassette) or reversion (where the gene of interest was recovered) were selected by a sucrose counter selection. Mutants with a clean deletion were separated from revertants by PCR screening[62]. Primers used for gene deletion are listed in Dataset S4.

To ensure that the gene deletion process did not introduce additional changes to the *S. oneidensis* genome, we checked REE-biosorption by revertants recovered in the process of deleting two of the genes. In both cases, REE-biosorption was statistically indistinguishable from the true wild-type (p-value < 0.05) (Fig. S5).

Analytical measurement of biosorption with ICP-MS

We explored biosorption in four different solution conditions (detailed in Table 1) with three rare earth elements: La (representing light REE), Eu (representing middle REE), and Yb (representing heavy REE). In every condition, bacterial culture density was normalized to the same optical density.

Bacterial strains of interest were retrieved from glycerol stocks frozen at − 80 °C and recovered on LB agar plates (with 50 mg L^{-1} kanamycin for transposon insertion strains). We picked three single colonies for each strain and inoculated them into separate wells containing 200 µL of LB (with 50 mg L^{-1} kanamycin for transposon insertion strains) in 96-well flat-bottom polypropylene plates (Greiner Bio-One ref: 655261) and incubated them at 30 °C overnight shaking at 800 rpm.

The following morning, we back-diluted 30 µL from each well into culture tubes containing 3 mL of LB (with 50 mg L^{-1} kanamycin for transposon mutants). Cultures were incubated at 30 °C until they reached an optical density (OD_{590}) of between 1.3 and 1.45.

Each culture was used for 4 biosorption experiments in each of the different conditions detailed in Table 1. Each culture was split into two 1.7 mL centrifuge tubes, pelleted at 7800 × g, resuspended in 1 mL of buffer, and then pelleted one more time at 7800 × g. The first tube was rinsed with a low ionic strength buffer (20 mM MES, 20 mM NaCl, adjusted to pH 5.5 with 5 M NaOH) and the second tube with a high ionic strength buffer (20 mM MES, 100 mM NaCl, adjusted to pH 5.5 with 5 M NaOH).

We resuspended the rinsed cells in 600 µL of the same respective buffer, took the optical density (OD), then divided each culture into two new tubes with a final OD of 0.85, a volume of 400 µL, and either high or low REE concentrations as follows. For cultures in low ionic strength solution (final concentrations of 10 mM NaCl and 10 mM MES), the low REE solution contained 30 µM each of lanthanum, europium, and ytterbium and the high REE solution contained 60 µM of each. For cultures in high ionic strength solution (final concentrations of 50 mM NaCl and 10 mM MES), the low REE solution contained 10 µM each of lanthanum, europium, and ytterbium and the high REE solution contained 30 µM of each.

The cultures were incubated for 10 min with the REE, prior to a new round of centrifugation (again, 2 min at 7800 × g). We then transferred 300 µL of our supernatant to 0.45 µm Supor 96-well filter plates (Pall Corporation ref:8029). We previously found that these plates adsorb at most a few percent of the REE in solution. Additionally,

we expect the filters to be saturated with REE and thus our estimate for the total amount of biosorbed REE does not depend on how much REE was adsorbed by the filter. From the filtered supernatant, we prepared our ICP-MS samples as described below.

ICP-MS measurements

Our ICP-MS samples were prepared by diluting our biosorption samples 1/25 in 2% trace metal grade nitric acid (JT9368, J.T. Baker, Radnor, PA). Our samples were analyzed using an Agilent 7800 ICP-MS (m/z: La, 139; Eu, 151; Yb, 172) using a rare earth element mix standard which included all the other rare earth elements in addition to the three we analyzed (67,349, Sigma-Aldrich, St. Louis, MO) and a rhodium in-line internal standard (SKU04736, Sigma-Aldrich, St. Louis, MO, m/z = 103). ICP-MS data were analyzed using the program MassHunter, version 4.5. Quality control was conducted by doing periodic measurements (every ten samples) of our standards (the 10, 25, 50, and 100 ppb) and 2% nitric acid blanks. We used the Rh internal standard to account for effects of drift. Repeat standards were analyzed periodically (or every 10 samples) and were quantified with an accuracy of + /− 2.5%.

Comparing transposon containing and wild-type S. oneidensis strains

We found that the average biosorption of transposon insertion strains did not resemble our wild-type bacteria. We theorize that even if the gene a transposon was inserted into did not alter biosorption, it is possible that the transposon itself—or the fact that the insertion mutant strains were grown up with Kanamycin—impacts biosorption. To test this hypothesis, we took four transposon mutants (which we refer to as quasi-wild-type or qWT) that had the transposon in a presumably neutral location and whose biosorption in the As-III screen did not significantly differ from average mutant in the containing plate.

Choice of quasi-WT strains

We expected that a transposon appearing at the end of a gene would have no effect on that gene. We thus picked transposon mutants where the transposon appeared at the very end of the gene. We also ensured that the transposon was at least 300 base pairs away from the start of any other gene to minimize disruption of promoter regions. Finally, we confirmed that the selected disruptions did not have any significant changes in biosorption within our assay. Four transposon mutants were selected at random from the mutants that met these requirements. The end of the genes where the transposon appeared were *SO_4279* for qWT_1, *SO_4707* for qWT_2, *SO_0214* for qWT_3, and *SO_2225* for qWT_4.

Biosorption of each qWT was compared to that of the natural WT and the other qWT strains using a two-sided t-test. We found that the wild-type showed at least 13% higher biosorption compared to the average qWT in every solution condition (Fig. S4). Additionally, the qWT had solution condition-dependent differences in total biosorption compared to each other. qWT_4 had higher ($p < 0.05$) biosorption than the average qWT for the HH solution condition. When we performed pairwise t-tests between our four qWT mutants, we found that in three out of four of our solution conditions, we had some mutant or mutants that had different biosorption than the others. In LL, qWT_2 had significantly higher biosorption than each of the other qWTs. In LH, qWT_2 had higher biosorption than qWT_3 and qWT_4. In HH, qWT_4 had significantly higher biosorption than every other qWT and qWT_3 had significantly lower biosorption than qWT_1.

Methodology for comparing relative REE biosorption

We compared the amount of biosorption for each individual REE to the total REE biosorption for each solution condition. We found that, over a finite range of total REE biosorption, there was generally a linear relationship between individual and total REE biosorption. We thus used our data for our transposon insertion mutants to plot lines of best fit in each solution condition comparing each individual REE biosorption to total REE biosorption. We speculated that, even though our data set consisted almost totally of strains that had outlier biosorption compared to the wild-type, the relative biosorption changes would not be in any particular direction. We thus expect the baseline calculated from these mutants to reasonably resemble the true average of *S. oneidensis* transposon insertion strains.

We excluded our two biggest disruption mutant total biosorption outliers—δ*nusA* and δ*SO_4685*—from our analysis. We left these strains out of our analysis because their total REE biosorption fell outside of the finite range of total biosorption that we felt confident was linear with individual REE biosorption.

Once we had our line of best fit, $f(REE_T)$, for each data point, we calculated the percent change of biosorption of the individual REE of interest (REE_i) on the y-axis compared to the expected value based on the total REE biosorption (REE_T) on the *x*-axis: $\frac{f(REE_T)-REE_i}{f(REE_T)}$. For each strain we calculated the mean and standard error of this percent change. Significance was calculated by doing a two-sided t-test looking to see if our percent change of biosorption was significantly different than 0.

Effects of extra incubation time on biosorption

Previous research has found that, in a sufficiently high pH environment, the longer bacteria were exposed to REE, the less biosorption occurred[40]. The authors theorized that this decrease in biosorption over time was caused by bacterial secretions. Other changes, such as changes to bacterial viability within a non-optimal media, could also be responsible for these time-dependent changes. Since they found that this occurred at pH 5.8 and we conducted our assays at pH 5.5, we conducted experiments to test how much different incubation times could have affected our results. Since the amount of time we let our bacteria mix with REE was fixed inside our experiment, we chose instead to conduct our experiments by changing the amount of time we let our bacteria sit after rinsing was completed, but prior to adding the REE for our assays.

We tested the effects of secretions in two of our biosorption environments, LL and HH. We did not find a statistically significant impact on absolute biosorption or on the separation factor for LL. While our results were not statistically significant, it did appear like there was a clear downward trajectory to the level of biosorption as well as an increase in the Yb/La and Eu/La separation factors. For HH, on the other hand, there was a substantial decrease in the overall biosorption level as well as a substantial increase in the Yb/La and Eu/La separation factors. However, the overall biosorption decrease between the 74 min and 138 min waiting periods lacked statistical significance and was far less than the decrease between the 42 and 74 min measurements. This suggests that the effects of the secretions (or whatever other mechanism is responsible for the change in REE binding) decrease over time—a result that would give us confidence in our experiments given that our bacteria typically sat in their final assay solutions for between 90 and 120 min prior to the completion of our biosorption assay.

Statistical information

Statistics relating to the genetic screen (identifying outliers, gene ontology enrichment analysis, operon enrichment analysis) are all described in their respective sections. All other statistics were performed using two-tailed t-tests with three biological replicates each.

Materials and correspondence

Correspondence and material requests should be addressed to B.B.. Individual strains (up to ≈ 10 at a time) are available at no charge for academic researchers. We are happy to supply a duplicate of the entire *S. oneidensis* knockout collection to academic researchers, but will require reimbursement for materials, supplies and labor costs. Commercial researchers should contact Cornell Technology Licensing for licensing details.

Data availability

Data for all figures is available at https://github.com/barstowlab/ree-selectivity and is archived on Zenodo[81]. Sequences for verification of mutants are available on GenBank under accession codes OR490624, OR490625, OR490626, OR490627, OR490628, OR490629, OR490630, OR490631, OR490632, OR490633, OR490634, and OR490635.

Code availability

The REE-separations model software is available at https://github.com/barstowlab/ree-selectivity and is archived on Zenodo[81].

Received: 14 June 2023; Accepted: 14 September 2023
Published online: 25 September 2023

Bibliography

1. Dent, P. C. Rare earth elements and permanent magnets. *J. Appl. Phys.* **111**, 07A721. https://doi.org/10.1063/1.3676616 (2012).
2. Lucas, J., Lucas, P., Le Mercier, T., Rollat, A. & Davenport, W. *Rare Earths: Science, Technology, Production and Use* (Elsevier Inc., 2014).
3. Schubert, E. F. & Kim, J. K. Solid-state light sources getting smart. *Science* **308**, 1274–1278. https://doi.org/10.1126/science.1108712 (2005).
4. Norman, A. F., Prangnell, P. B. & McEwen, R. S. The solidification behaviour of dilute aluminium–scandium alloys. *Acta Mater.* **46**, 5715–5732. https://doi.org/10.1016/s1359-6454(98)00257-2 (1998).
5. Adesina, O., Anzai, I. A., Avalos, J. L. & Barstow, B. Embracing biological solutions to the sustainable energy challenge. *Chemistry* **2**, 20–51. https://doi.org/10.1016/j.chempr.2016.12.009 (2017).
6. Nazarov, M. & Noh, D. Y. *New Generation of Europium and Terbium Activated Phosphors: From Syntheses to Applications* (Pan Stanford Publishing, 2011).
7. Müller, T. & Friedrich, B. Development of a recycling process for nickel-metal hydride batteries. *J. Power Sources* **158**, 1498–1509. https://doi.org/10.1016/j.jpowsour.2005.10.046 (2006).
8. International Energy Outlook 2021 with Projections to 2050. (2021).
9. Annual Energy Outlook 2020, with projections to 2050. (U.S. Energy Information Administration, Office of Energy Analysis, U.S. Department of Energy, Washington, DC, 2020).
10. Voncken, J. H. L. *The Rare Earth Elements, An Introduction*. (2016).
11. Scheyder, E. *China Set to Control Rare Earth Supply for Years Due to Processing Dominance*, <https://www.reuters.com/article/us-china-usa-rareearth-refining/china-set-to-control-rare-earth-supply-for-years-due-to-processing-dominance-idUSKCN1T004J> (2019).
12. Mining the Future: How China is set to dominate the next Industrial Revolution. (Foreign Policy, 2019).
13. Peiravi, M. *et al.* Chemical extraction of rare earth elements from coal ash. *Miner. Metall. Process.* **34**, 170–177. https://doi.org/10.19150/mmp.7856 (2017).
14. in *CRC Handbook of Chemistry and Physics, 102nd Edition* (Taylor and Francis, 2021).
15. Peiravi, M. *et al.* A review of rare-earth elements extraction with emphasis on non-conventional sources: Coal and coal byproducts, iron ore tailings, apatite, and phosphate byproducts. *Min. Metall. Explor.* **38**, 1–26. https://doi.org/10.1007/s42461-020-00307-5 (2021).
16. Kuan, S. H., Saw, L. H. & Ghorbani, Y. in *International Annual Symposium on Sustainability Science and Management* 105.
17. Johnson, D. B. Biomining—Biotechnologies for extracting and recovering metals from ores and waste materials. *Curr. Opin. Biotech.* **30**, 24–31. https://doi.org/10.1016/j.copbio.2014.04.008 (2014).
18. Reed, D. W., Fujita, Y., Daubaras, D. L., Jiao, Y. & Thompson, V. S. Bioleaching of rare earth elements from waste phosphors and cracking catalysts. *Hydrometallurgy* **166**, 34–40. https://doi.org/10.1016/j.hydromet.2016.08.006 (2016).
19. Schmitz, A. M. *et al.* Generation of a Gluconobacter oxydans knockout collection for improved extraction of rare earth elements. *Nat. Commun.* **12**, 6693. https://doi.org/10.1038/s41467-021-27047-4 (2021).
20. Cotruvo, J. A., Featherston, E. R., Mattocks, J. A., Ho, J. V. & Laremore, T. N. Lanmodulin: A highly selective lanthanide-binding protein from a lanthanide-utilizing bacterium. *J. Am. Chem. Soc.* **140**, 15056–15061. https://doi.org/10.1021/jacs.8b09842 (2018).
21. Deblonde, G. J. *et al.* Selective and efficient biomacromolecular extraction of rare-earth elements using lanmodulin. *Inorg. Chem.* **59**, 11855–11867. https://doi.org/10.1021/acs.inorgchem.0c01303 (2020).

22. Park, D. M. *et al.* Bioadsorption of rare earth elements through cell surface display of lanthanide binding tags. *Environ. Sci. Technol.* **50**, 2735–2742. https://doi.org/10.1021/acs.est.5b06129 (2016).
23. Tay, P. K. R., Manjula-Basavanna, A. & Joshi, N. S. Repurposing bacterial extracellular matrix for selective and differential abstraction of rare earth elements. *Green Chem.* **20**, 3512–3520. https://doi.org/10.1039/c8gc01355a (2018).
24. Bogart, J. A., Lippincott, C. A., Carroll, P. J. & Schelter, E. J. An Operationally simple method for separating the rare-earth elements neodymium and dysprosium. *Angew. Chem Int. Ed.* **54**, 8222–8225. https://doi.org/10.1002/anie.201501659 (2015).
25. Higgins, R. F. *et al.* Magnetic field directed rare-earth separations. *Angew. Chem. Int. Ed.* **59**, 1851–1856. https://doi.org/10.1002/anie.201911606 (2020).
26. Lin, W. *et al.* Promising priority separation of europium from lanthanide by novel DGA-functionalized metal organic frameworks. *Miner. Eng.* **164**, 106831. https://doi.org/10.1016/j.mineng.2021.106831 (2021).
27. Yang, H. *et al.* Selective crystallization of rare-earth ions into cationic metal-organic frameworks for rare-earth separation. *Angew. Chem. Int. Ed.* **60**, 11148–11152. https://doi.org/10.1002/anie.202017042 (2021).
28. Park, D. M., Brewer, A., Reed, D. W., Lammers, L. N. & Jiao, Y. Recovery of rare earth elements from low-grade feedstock leachates using engineered bacteria. *Environ. Sci. Technol.* **51**, 13471–13480. https://doi.org/10.1021/acs.est.7b02414 (2017).
29. Brewer, A. *et al.* Recovery of rare earth elements from geothermal fluids through bacterial cell surface adsorption. *Environ. Sci. Technol.* **53**, 7714–7723. https://doi.org/10.1021/acs.est.9b00301 (2019).
30. Brewer, A. *et al.* Microbe encapsulation for selective rare-earth recovery from electronic waste leachates. *Environ. Sci. Technol.* **53**, 13888–13897. https://doi.org/10.1021/acs.est.9b04608 (2019).
31. Good, N. M. *et al.* Hyperaccumulation of gadolinium by methylorubrum extorquens AM1 reveals impacts of lanthanides on cellular processes beyond methylotrophy. *Front. Microbiol.* **13**, 820327. https://doi.org/10.3389/fmicb.2022.820327 (2022).
32. Thiele, N. A., Fiszbein, D. J., Woods, J. J. & Wilson, J. J. Tuning the separation of light lanthanides using a reverse-size selective aqueous complexant. *Inorg. Chem.* **59**, 16522–16530. https://doi.org/10.1021/acs.inorgchem.0c02413 (2020).
33. Bonificio, W. D. & Clarke, D. R. Rare-earth separation using bacteria. *Environ. Sci. Technol. Lett.* **3**, 180–184. https://doi.org/10.1021/acs.estlett.6b00064 (2016).
34. Hu, A., MacMillan, S. N. & Wilson, J. J. Macrocyclic ligands with an unprecedented size-selectivity pattern for the lanthanide ions. *J. Am. Chem. Soc.* **142**, 13500–13506. https://doi.org/10.1021/jacs.0c05217 (2020).
35. Abbas, S. H., Ismail, I. M., Mostafa, T. M. & Sulaymon, A. H. Biosorption of heavy metals: A review. *J. Chem. Sci. Technol.* **3**, 74–102 (2014).
36. *Microbial Biosorption of Metals*. (Springer Dordrecht, 2011).
37. Tang, X. *et al.* Profiling the membrane proteome of shewanella oneidensis MR-1 with new affinity labeling probes. *J. Proteome Res.* **6**, 724–734. https://doi.org/10.1021/pr060480ePMID-17269728 (2007).
38. Sohlenkamp, C. & Geiger, O. Bacterial membrane lipids: Diversity in structures and pathways. *FEMS Microbiol. Rev.* **40**, 133–159. https://doi.org/10.1093/femsre/fuv008 (2016).
39. Perez-Burgos, M. & Sogaard-Andersen, L. Biosynthesis and function of cell-surface polysaccharides in the social bacterium *Myxococcus xanthus*. *Biol. Chem.* **401**, 1375–1387. https://doi.org/10.1515/hsz-2020-0217 (2020).
40. Takahashi, Y., Châtellier, X., Hattori, K. H., Kato, K. & Fortin, D. Adsorption of rare earth elements onto bacterial cell walls and its implication for REE sorption onto natural microbial mats. *Chem. Geol.* **219**, 53–67. https://doi.org/10.1016/j.chemgeo.2005.02.009 (2005).
41. Fomina, M. & Gadd, G. M. Biosorption: Current perspectives on concept, definition and application. *Bioresour. Technol.* **160**, 3–14. https://doi.org/10.1016/j.biortech.2013.12.102 (2014).
42. Moriwaki, H. *et al.* Application of freeze-dried powders of genetically engineered microbial strains as adsorbents for rare earth metal ions. *ACS Appl. Mater. Interfaces* **8**, 26524–26531. https://doi.org/10.1021/acsami.6b08369 (2016).
43. Takahashi, Y., Yamamoto, M., Yamamoto, Y. & Tanaka, K. EXAFS study on the cause of enrichment of heavy REEs on bacterial cell surfaces. *Geochim. Cosmochim. Acta* **74**, 5443–5462. https://doi.org/10.1016/j.gca.2010.07.001 (2010).
44. Moriwaki, H., Koide, R., Yoshikawa, R., Warabino, Y. & Yamamoto, H. Adsorption of rare earth ions onto the cell walls of wild-type and lipoteichoic acid-defective strains of Bacillus subtilis. *Appl. Microbiol. Biotechnol.* **97**, 3721–3728. https://doi.org/10.1007/s00253-012-4200-3 (2013).
45. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898. https://doi.org/10.1038/nature08187 (2009).
46. Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci.* **111**, 17803–17808. https://doi.org/10.1073/pnas.1409523111 (2014).
47. Halperin, S. O. *et al.* CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560**, 248–252. https://doi.org/10.1038/s41586-018-0384-8 (2018).
48. Corts, A. D. *Efficient and Precise Genome Editing in Shewanella with Recombineering and CRISPR/Cas9-mediated Counter-selection* PhD thesis, University of Minnesota, (2019).
49. Corts, A. D., Thomason, L. C., Gill, R. T. & Gralnick, J. A. Efficient and precise genome editing in Shewanella with recombineering and CRISPR/Cas9-mediated counter-selection. *ACS Synth. Biol.* **8**, 1877–1889. https://doi.org/10.1021/acssynbio.9b00188 (2019).
50. Corts, A. D., Thomason, L. C., Gill, R. T. & Gralnick, J. A. A new recombineering system for precise genome-editing in *Shewanella oneidensis* strain MR-1 using single-stranded oligonucleotides. *Sci. Rep.* **9**, 39. https://doi.org/10.1038/s41598-018-37025-4 (2019).
51. Anzai, I. A., Shaket, L., Adesina, O., Baym, M. & Barstow, B. Rapid curation of gene disruption collections using Knockout Sudoku. *Nat. Protocols* **12**, 2110–2137. https://doi.org/10.1038/nprot.2017.073 (2017).
52. Baym, M., Shaket, L., Anzai, I. A., Adesina, O. & Barstow, B. Rapid construction of a whole-genome transposon insertion collection for *Shewanella oneidensis* by Knockout Sudoku. *Nat. Commun.* **7**, 13270. https://doi.org/10.1038/ncomms13270 (2016).
53. Baym, M., Shaket, L., Anzai, I. A., Adesina, O. & Barstow, B. Rapid construction of a whole-genome transposon insertion collection for *Shewanella oneidensis*by Knockout Sudoku. *Nat. Commun.* **7**, 13270. https://doi.org/10.1038/ncomms13270 (2016).
54. Hogendoorn, C. *et al.* Facile Arsenazo III-based assay for monitoring rare earth element depletion from cultivation media for methanotrophic and methylotrophic bacteria. *Appl. Environ. Microb.* **84**, e02887-02817. https://doi.org/10.1128/aem.02887-17 (2018).
55. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
56. Consortium GO. The gene ontology resource: Enriching a gold mine. *Nucleic Acids Res.* **49**, D325–D334. https://doi.org/10.1093/nar/gkaa1113 (2021).
57. Dehal, P. S. *et al.* MicrobesOnline: An integrated portal for comparative and functional genomics. *Nucleic Acids Res.* **38**, D396-400. https://doi.org/10.1093/nar/gkp919 (2010).
58. Taboada, B., Ciria, R., Martinez-Guerrero, C. E. & Merino, E. ProOpDB: prokaryotic operon DataBase. *Nucleic Acids Res.* **40**, D627-631. https://doi.org/10.1093/nar/gkr1020 (2012).
59. Lassak, J., Bubendorfer, S. & Thormann, K. M. Domain analysis of ArcS, the hybrid sensor kinase of the *Shewanella oneidensis* MR-1 Arc two-component system, reveals functional differentiation of its two receiver domains. *J. Bacteriol.* **195**, 482–492. https://doi.org/10.1128/JB.01715-12 (2013).
60. Iuchi, S. & Lin, E. arcA (dye), A global regulatory gene in Escherichia coli mediating repression of enzymes in aerobic pathways. *Proc. Natl. Acad. Sci.* **85**, 1888–1892 (1988).

61. Heidelberg, J. F. *et al.* Genome sequence of the dissimilatory metal ion–reducing bacterium Shewanella oneidensis. *Nat. Biotechnol.* **20**, nbt749. https://doi.org/10.1038/nbt749 (2002).
62. Rowe, A. R. *et al.* Identification of a pathway for electron uptake in *Shewanella oneidensis*. *Commun. Biol.* **4**, 957. https://doi.org/10.1038/s42003-021-02454-x (2021).
63. Fitzgerald, L. A. *et al. Shewanella oneidensis* MR-1 Msh pilin proteins are involved in extracellular electron transfer in microbial fuel cells. *Process Biochem.* **47**, 170–174. https://doi.org/10.1016/j.procbio.2011.10.029 (2012).
64. Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A. & Gautheret, D. ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol.* **8**, 11–13. https://doi.org/10.4161/rna.8.1.13346 (2011).
65. Jacobs, M. A. *et al.* Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* **100**, 14339–14344. https://doi.org/10.1073/pnas.2036282100 (2003).
66. Opijnen, T. V., Bodi, K. L. & Camilli, A. Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* **6**, 767–772. https://doi.org/10.1038/nmeth.1377 (2009).
67. Bertani, B. & Ruiz, N. Function and biogenesis of lipopolysaccharides. *EcoSal. Plus* **8**, 10–1128. https://doi.org/10.1128/ecosalplus.ESP-0001-2018 (2018).
68. Edel, M. *et al.* Extracellular riboflavin induces anaerobic biofilm formation in *Shewanella oneidensis*. *Biotechnol. Biofuels* **14**, 130. https://doi.org/10.1186/s13068-021-01981-3 (2021).
69. Bélanger, M., Burrows, L. L. & Lam, J. S. Functional analysis of genes responsible for the synthesis of the B-band O antigen *of Pseudomonas aeruginosa* serotype O6 lipopolysaccharide. *Microbiology* **145**, 3505–3521. https://doi.org/10.1099/00221287-145-12-3505 (1999).
70. Lee, K. J., Kim, J. A., Hwang, W., Park, S. J. & Lee, K. H. Role of capsular polysaccharide (CPS) in biofilm formation and regulation of CPS production by quorum-sensing in Vibrio vulnificus. *Mol. Microbiol.* **90**, 841–857. https://doi.org/10.1111/mmi.12401 (2013).
71. Miller, W. L. *et al.* Biochemical characterization of WbpA, a UDP-N-acetyl-D-glucosamine 6-dehydrogenase involved in O-antigen biosynthesis in Pseudomonas aeruginosa PAO1. *J. Biol. Chem.* **279**, 37551–37558. https://doi.org/10.1074/jbc.M404749200 (2004).
72. Thormann, K. M., Saville, R. M., Shukla, S., Pelletier, D. A. & Spormann, A. M. Initial Phases of biofilm formation in *Shewanella oneidensis* MR-1. *J. Bacteriol.* **186**, 8096–8104. https://doi.org/10.1128/JB.186.23.8096-8104.2004 (2004).
73. Daughney, C. J., Fowle, D. A. & Fortin, D. The effect of growth phase on proton and metal adsorption by Bacillus subtilis. *Geochim. Cosmochim. Acta* **65**, 1025–1035. https://doi.org/10.1016/s0016-7037(00)00587-1 (2001).
74. Park, D. *et al.* A biosorption-based approach for selective extraction of rare earth elements from coal byproducts. *Sep. Purif. Technol.* **241**, 116726. https://doi.org/10.1016/j.seppur.2020.116726 (2020).
75. Yi, Y.-C. & Ng, I. S. Establishment of toolkit and T7RNA polymerase/promoter system in Shewanella oneidensis MR-1. *J. Taiwan Inst. Chem. Eng.* **109**, 8–14. https://doi.org/10.1016/j.jtice.2020.02.003 (2020).
76. Yamamoto, N. *et al.* Update on the Keio collection of Escherichia coli single-gene deletion mutants. *Mol. Syst. Biol.* **5**, 335. https://doi.org/10.1038/msb.2009.92 (2009).
77. Rohwer, H., Collier, N. & Hosten, E. Spectrophotometric study of arsenazo III and its interactions with lanthanides. *Anal. Chim. Acta* **314**, 219–223. https://doi.org/10.1016/0003-2670(95)00279-9 (1995).
78. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368. https://doi.org/10.1038/s41592-021-01101-x (2021).
79. Götz, S. *et al.* High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435. https://doi.org/10.1093/nar/gkn176 (2008).
80. Adrian, A. & Jörg, R. *Gene Set Enrichment Analysis with topGO*, <http://compdiag.molgen.mpg.de/ngfn/docs/2007/sep/topGO_Exercises.pdf> (2007).
81. Medin, S. Data and Code for Genomic Characterization of Rare Earth Binding by Shewanella oneidensis. (2023). https://doi.org/10.5281/zenodo.7960563

Acknowledgements

S.M. was supported by a Cornell Presidential Life Sciences Graduate Fellowship. A.M.S. was supported by a Cornell Energy Systems Institute Postdoctoral Fellowship, and a Small Grant from the Cornell Atkinson Center for Sustainability. This work was supported by Cornell University startup funds, an Academic Venture Fund award from the Atkinson Center for Sustainability at Cornell University, a Career Award at the Scientific Interface from the Burroughs Welcome Fund to B.B., ARPA-E award DE-AR0001341 to B.B, M.H., E.G., and M.W., and a gift from Mary Fernando Conrad and Tony Conrad to B.B.

Author contributions

Conceptualization, S.M. and B.B.; methodology, S.M. and B.B.; investigation, S.M., A.M.S., B.P. K.M., M.R. and B.B; writing—original draft, S.M.; writing—review and editing, S.M., B.B., M.R., M.H., M.W., E.G., and B.B.; funding acquisition, A.M.S., M.H., E.G., M.W., and B.B.; resources, M.R., M.H, E.G., M.W., and B.B.; supervision, M.R., M.W. and B.B.; data curation, S.M. and B.B.; visualization, S.M. and B.B.; formal analysis, S.M.

Competing interests

The authors are pursuing patent protection for engineered organisms using knowledge gathered in this work (US provisional patent Application no. 63/405353). A.M.S. and S.M. are co-founders of, and B.B. is a contributor, and uncompensated advisor to, REEgen, Inc., which is developing genetically engineered microbes for REE bio-mining. The remaining authors have competing interests to declare.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-42742-6.

Correspondence and requests for materials should be addressed to B.B.

Reprints and permissions information is available at www.nature.com/reprints.

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