

UC Berkeley

UC Berkeley Previously Published Works

Title

Genome-Resolved Meta-Omics Ties Microbial Dynamics to Process Performance in Biotechnology for Thiocyanate Degradation

Permalink

<https://escholarship.org/uc/item/50g6126b>

Journal

Environmental Science and Technology, 51(5)

ISSN

0013-936X

Authors

Kantor, Rose S

Huddy, Robert J

Iyer, Ramsunder

et al.

Publication Date

2017-03-07

DOI

10.1021/acs.est.6b04477

Peer reviewed

23 ABSTRACT

24
25 Remediation of industrial wastewater is important for preventing environmental
26 contamination and enabling water reuse. Biological treatment for one industrial
27 contaminant, thiocyanate (SCN^-), relies upon microbial hydrolysis, but this process is
28 sensitive to high loadings. To examine the activity and stability of a microbial community
29 over increasing SCN^- loadings, we established and operated a continuous-flow bioreactor
30 fed increasing loadings of SCN^- . A second reactor was fed ammonium sulfate to mimic
31 breakdown products of SCN^- . Biomass was sampled from both reactors for
32 metagenomics and metaproteomics, yielding a set of genomes for 144 bacteria and one
33 rotifer that constituted the abundant community in both reactors. We analyzed the
34 metabolic potential and temporal dynamics of these organisms across the increasing
35 loadings. In the SCN^- reactor, *Thiobacillus* strains capable of SCN^- degradation were
36 highly abundant, whereas the ammonium sulfate reactor contained nitrifiers and
37 heterotrophs capable of nitrate reduction. Key organisms in the SCN^- reactor expressed
38 proteins involved in SCN^- degradation, sulfur oxidation, carbon fixation, and nitrogen
39 removal. Lower performance at higher loadings was linked to changes in microbial
40 community composition. This work provides an example of how meta-omics can increase
41 our understanding of industrial wastewater treatment and inform iterative process design
42 and development.

43 44 INTRODUCTION

45
46 Microbial communities in biotechnology have historically been treated as black boxes,
47 but as molecular methods have improved, our knowledge of these systems has deepened.
48 Increasingly, ‘meta-omics’ methods are being used to investigate critical processes and
49 potential weak points in biotechnology, such as nitrogen and phosphorus removal or
50 bulking in wastewater treatment.¹⁻³ In particular, specialized treatment of industrial
51 wastewater has benefited from a genome-resolved meta-omics approach⁴⁻⁷ using high-
52 throughput sequencing of community genomic DNA (metagenomics) or RNA
53 (metatranscriptomics), or spectral characterization of proteins (metaproteomics). These
54 data can be used to identify the key species involved in processes of interest. Improved
55 understanding of the activities and abundances of these organisms under varying
56 conditions could inform design and operation of these systems.

57 Thiocyanate (SCN^-) is a widespread industrial contaminant found at especially
58 high concentrations (up to 4000 mg/L) in gold mining effluents. If not remediated, it can
59 affect human health and aquatic organisms.⁸⁻¹⁰ Notably, SCN^- is inhibitory toward iron-
60 and sulfur-oxidizing microorganisms used in bio-oxidation processes at some gold mines
61 (such as BIOX®) and therefore must be removed before wastewater can be recycled
62 within a mining site or discharged into the environment. SCN^- can be biodegraded by
63 chemolithoautotrophic bacteria as a source of energy,¹¹⁻¹⁵ by heterotrophic organisms as a
64 sole source of nitrogen,^{16,17} and by complex microbial consortia.¹⁸ The initial degradation
65 products are ammonium, carbon dioxide, and reduced sulfur compounds. An industrial-
66 scale process known as Activated Sludge Tailings Effluent Remediation (ASTER™,
67 Outotec, South Africa) successfully treats SCN^- -containing wastewater at several gold
68 mines.¹⁹

69 Inoculated with sludge from the ASTER™ process, a long-running laboratory-
70 scale SCN⁻-fed bioreactor (known as the “SCN⁻ stock reactor”) at the University of Cape
71 Town contains a characterized, diverse microbial community.²⁰⁻²² Previous work on this
72 community implicated several abundant *Thiobacillus* spp. in SCN⁻ degradation due to the
73 presence of an SCN⁻ operon in the genomes of these autotrophic bacteria. Results also
74 suggested the potential for nitrogen removal by *Thiobacillus* spp. and other community
75 members, and the presence of heterotrophs.²¹ Questions remained regarding community
76 stability at different SCN⁻ loadings, expression of the observed metabolic potential, and
77 the importance of inter-organism interactions, especially for nitrogen removal. The SCN⁻
78 stock reactor provided the inoculum for the bioreactors established in this study.

79 We used time-series genome-resolved metagenomics, in combination with
80 metaproteomic analyses of the samples from final time point, to track changes in the
81 microbial community of a newly-inoculated SCN⁻ bioreactor operated with increasing
82 loadings. To enrich for organisms that can use and remove the nitrogen produced by
83 SCN⁻ degradation, a second reactor with the same inoculum was fed ammonium sulfate
84 (NH₄(SO₄)_{1/2}) and molasses but no SCN⁻. We describe the microbial community
85 structure, protein expression, and replication rates in both reactors during the experiment.
86 Our analysis linked shifts in community membership to changes in reactor function,
87 highlighted organisms and metabolic pathways active under high-SCN⁻ conditions, and
88 supported the importance of biofilm in this system.

90 MATERIALS AND METHODS

91
92 **Reactor set-up, inoculation, and operation:** Two continuous stirred tank reactors
93 (CSTRs) were inoculated with homogenized biofilm and planktonic samples from the
94 long-running SCN⁻ stock reactor at the University of Cape Town. Reactors were stirred
95 with a pitched-blade impeller at 270 rpm and sparged with filtered air at 900 mL/min.
96 One reactor was fed KSCN while the other was fed NH₄(SO₄)_{1/2} at equivalent nitrogen
97 loadings in order to mimic the end-products of thiocyanate degradation. Both reactors
98 were also fed molasses (150 mg/L) and KH₂PO₄ (0.28 mM) to provide supplemental
99 nutrients. Feed contained increasing amounts of KOH to modulate reactor pH as
100 necessary (**Figure 1**) and small amounts of 5 N KOH were added directly to reactors if
101 observed pH was ≤ 6.5. Bicarbonate (4 g/L) was added to the feed to buffer the system
102 from day 112 to day 136.

103 The reactors were run in batch-fed mode until SCN⁻ degradation was stably
104 observed in the SCN⁻ reactor, at which time both reactors were switched to continuous
105 feeding at a residence time of 42 hours (day 5). The hydraulic retention time (HRT) of
106 both reactors was lowered from 42 hours to 12 hours (days 5-68) and then maintained at
107 12 hours while the feed concentration of SCN⁻ or equivalent NH₄(SO₄)_{1/2} was increased
108 stepwise. The reactor was allowed to stabilize between each step to reach steady state
109 (**Figure S1**).

110
111 **Sampling:** Samples of biomass from each reactor were taken for metagenomic
112 sequencing just before increases in feed concentration (**Figure 1** and **Table S1**).
113 Approximately 0.5 g (wet-weight) of biofilm was scraped from the wall of each reactor
114 with sterile spatula and stored at -60 °C. Paired samples of planktonic biomass were

115 collected by filtering 300 mL of the liquid phase from each reactor onto a sterile 0.22 μm
116 filter. Biomass was gently washed off the filter with sterile water, pelleted, and stored at -
117 60 $^{\circ}\text{C}$ until further analysis. Filtered media was returned to the reactor to maintain
118 chemical continuity.

119

120 **Chemical analysis:** Bulk liquid was sampled daily for chemical analysis, filtered through
121 a 0.22 μm filter, pH analyzed, and frozen at -20 $^{\circ}\text{C}$ until further analysis. SCN^- was
122 measured using High Performance Liquid Chromatography as described previously.²¹ Ion
123 chromatography was performed to quantify nitrate and sulfate (Supporting Information).

124

125 **DNA extraction and sequencing:** DNA was extracted using a NucleoSpin[®] soil
126 genomic DNA extraction kit (Machery-Nagel, Germany) with the inclusion of a repeated
127 extraction step, according to the manufacturer's instructions. Paired-end Illumina TruSeq
128 libraries with either tight insert fragment sizes of 800 bp or regular insert sizes of 500 bp,
129 depending on the sample, were prepared at the Joint Genome Institute (Walnut Creek,
130 CA) (**Table S1**). Libraries were sequenced on an Illumina HiSeq-2500 in rapid run mode
131 to yield 250 bp paired-end reads.

132

133 **Metagenomic assembly, binning, and annotation:** Reads from each sample were
134 trimmed based on quality scores using sickle (<https://github.com/najoshi/sickle>) and then
135 assembled independently with idba_ud²³. Binning of the assembled scaffolds was
136 performed using ggKbase (ggkbase.berkeley.edu) based on scaffold taxonomy, percent
137 GC, and sequencing coverage. Within each assembly, bins were refined and added using
138 differential abundance data visualized in emergent self-organizing maps (ESOMs) as in
139 Sharon *et al.*²⁴ (see Supplemental Information). Each ESOM was trained and visualized
140 using databionic ESOM tools (**Figure S2**) ([http://databionic-](http://databionic-esom.sourceforge.net/index.html)
141 [esom.sourceforge.net/index.html](http://databionic-esom.sourceforge.net/index.html)).²⁵ In two samples (planktonic inoculum and SCN^-
142 reactor T2 biofilm), subassemblies using 1/60th or 1/50th of the reads, respectively, were
143 performed to improve assembly of the most abundant organism as previously described
144 by Hug *et al.*²⁶

145 Many bins were redundant given the recurrence of organisms across the time
146 series experiment. Nucmer²⁷ was used to align sequences and identify sets of bacterial
147 genomes sharing $\geq 98\%$ nucleotide identity across $> 50\%$ of the sequence. The best bin
148 was chosen based on genome completeness and length for inclusion in a de-replicated
149 dataset. Genomes without replicates were also included, except for two known
150 contaminant genomes ('Candidatus Altiarchaeum hamiconexum' and an
151 Epsilonproteobacterium) from another sequencing run on the same lane. Bins were
152 excluded from the final de-replicated dataset if they contained < 36 of 51 single copy
153 genes (SCG) or > 8 multi-copy SCG prior to curation (see **Table S2A**). One recurring
154 eukaryotic genome bin, one chloroplast, several mitochondria, phages, eukaryotic
155 viruses, and plasmid bins were included in the de-replicated dataset. De-replicated
156 bacterial genomes were curated using ra2.py, an automated curation method that uses
157 coverage and paired-end read information to find and reassemble or mask regions with
158 mis-assemblies

159 (https://github.com/christophertbrown/fix_assembly_errors/releases/tag/2.00).²⁸ Curation
160 used the reads of the sample from which the genome originated.

161 Annotation of genome bins used reciprocal ublast²⁹ searches against KEGG³⁰ and
162 UniRef100,³¹ as well as single-direction searches against UniProt.³² Functional genes and
163 marker genes were identified by annotations and using hmmsearch (HMMER 3.1b2;
164 <http://hmmer.org/>) with Hidden Markov Models (HMMs) from TIGRFAM (v15.0),
165 PFAM,³³ and with custom HMMs³⁴ (accessible at <https://github.com/banfieldlab>).
166

167 **Community composition:** Bowtie2 was used with default settings to map reads from
168 each sample to the de-replicated dataset (**Figure S3**). The resulting files were filtered
169 using mapped.py (<https://github.com/christophertbrown/mapped>) to remove reads that
170 mapped with > 3 mismatches. Coverage for each genome in each sample was calculated
171 and values $\leq 1x$ were converted to zero. Genome coverage values were then normalized
172 by dividing by the number of reads for each sample and then multiplying by the number
173 of reads in the largest sample. Normalized coverage was used as a proxy for the relative
174 abundances of organisms across samples (**Figure S4**). A concatenated ribosomal protein
175 tree containing references and sequences from SCN⁻ bioreactors was constructed as
176 previously described³⁵ (**Figure S5**). Datasets for each of 16 ribosomal proteins were
177 aligned independently with MUSCLE, alignments were trimmed and columns with 99%
178 gaps were removed in Geneious, and trimmed alignments were concatenated. RAxML
179 was used to generate a maximum likelihood phylogeny under the LG + gamma model.
180

181 **Variant analysis and Replication rate calculations:** see Supporting Information.
182

183 **Protein extraction and proteomic data analysis:** Proteins were extracted as previously
184 described³⁶ and ~1 mg of protein was subjected to trichloroacetic acid precipitation and
185 subsequent digestion with trypsin. Proteolytic peptides were analyzed via an online nano
186 2D LC-MS/MS system interfaced with hybrid LTQ-Orbitrap-Velos MS (ThermoFisher
187 Scientific). Subsequent processing of the collected spectra was done using Myrimatch³⁷,
188 with the de-replicated set of genomes as the database (Supporting Information). Peptide
189 identifications were quality-filtered to < 1% false discovery rate. Analysis of proteins
190 involved in key metabolic pathways considered spectral counts for unique peptides and
191 total spectral counts for each protein from two technical replicates.
192

193 **Data availability:** Raw read data is accessible at NCBI under accession number
194 SRP056932 (<http://www.ncbi.nlm.nih.gov/sra/SRP056932>) and genome accession
195 numbers may be found in **Table S2A**. Genome bins and sequences for scaffolds, genes,
196 and proteins can be viewed and downloaded at [http://ggkbase.berkeley.edu/scnpilot-](http://ggkbase.berkeley.edu/scnpilot-dereplicated/organisms)
197 [dereplicated/organisms](http://ggkbase.berkeley.edu/scnpilot-dereplicated/organisms). Proteomics data is available at
198 <https://massive.ucsd.edu/ProteoSAFe/datasets.jsp> (MassIVE ID MSV000080104).
199

200 RESULTS AND DISCUSSION

201

202 **Reactor chemistry and efficiency:** In the newly-inoculated SCN-fed and NH₄(SO₄)_{1/2}
203 reactors, the loading rate was increased across 238 days. Samples were taken for
204 metagenomic analysis after HRT reached 12 h. In the SCN⁻ reactor, the SCN⁻ removal
205 rate consistently increased to match the increasing loading rate to a maximum of 1.07
206 mmol.h⁻¹ (**Figures 1A** and **S1A**). On further increase to 1.43 mmol.h⁻¹, the SCN⁻ removal

207 rate decreased and efficiency dropped to near 50%. On average, the stoichiometry
208 between the SCN^- removal rate and sulfate output was 1.05:1, near the 1:1 ratio expected
209 based on known SCN^- degradation mechanisms coupled to complete oxidation of the
210 sulfide released.

211 As loading increased, the thickness of the biofilm that formed on all surfaces
212 within the reactor increased. During one period early in the experiment, nitrate (NO_3^-)
213 output reached up to 30% of nitrogen input as SCN^- (days 86-107) and fluctuated
214 thereafter, reaching a maximum of 64% of nitrogen input. After day 200, nitrate output
215 remained consistently low and sulfate accumulated. Base was added periodically to
216 counter acidification as loadings increased. The SCN^- removal rate decreased during one
217 period of high pH that resulted from over-correction of the feed pH (**Figure 1A**).

218 In the $\text{NH}_4(\text{SO}_4)_{1/2}$ reactor, the rate of sulfate leaving the reactor rose steadily
219 throughout the experiment, matching the sulfate loading rate and indicating that little
220 sulfate was retained in biomass or converted to other forms (**Figure 1B** and **S1B**). The
221 nitrate output rate increased with decreasing HRT and then increased more slowly as
222 biofilm established and thickened. Overall, higher nitrate effluent concentrations were
223 measured in the $\text{NH}_4(\text{SO}_4)_{1/2}$ reactor than in the SCN^- reactor (**Figure S1B**).

224
225 **Genome recovery over the sample series:** Biofilm from the two reactors was sampled
226 at four time points (T1-T4) during the experiment, and concurrent samples of planktonic
227 biomass were collected at T2 and T3 in the SCN^- reactor and T2-T4 in the $\text{NH}_4(\text{SO}_4)_{1/2}$
228 reactor. The inoculum for these reactors (T0), biofilm and planktonic biomass taken
229 from the SCN^- stock reactor, was also sampled (**Figure 1** and **Table S1**). Independent
230 metagenomic assemblies were performed for each sample and 789 bacterial genome bins
231 were reconstructed (**Figure S2**). Two genomes represented highly abundant organisms,
232 and subassemblies substantially improved these genomes (see Methods). Dereplication
233 across the time-series yielded a non-redundant set of 144 draft-quality genomes (**Tables**
234 **1, S2A**). Eukaryotic, mitochondrial, chloroplast, phage and plasmid genomes were also
235 recovered and de-replicated (**Tables 1, S2B**). No Archaea were detected in this system,
236 consistent with previous studies²⁰⁻²². Mapping reads from each assembly demonstrated
237 that this non-redundant genome set accounted for between 72.5 and 93.2% of the data
238 (**Figure S3**). This level of genome recovery approaches that reported for much simpler
239 communities such as those from the infant gut³⁸. The de-replicated metagenomic dataset
240 was used as a database for proteomic searches and accounted for 34, 32, and 15% of
241 high-quality peptides from the SCN^- reactor biofilm, $\text{NH}_4(\text{SO}_4)_{1/2}$ reactor biofilm, and
242 $\text{NH}_4(\text{SO}_4)_{1/2}$ reactor planktonic samples, respectively, at T4. This level of identification is
243 comparable to that seen with the same type of analysis on infant gut metaproteomes
244 paired to metagenomic databases³⁹.

245
246 **Community structure and metabolic potential:** Hierarchical clustering of samples
247 based on their community compositions grouped the samples first by the reactor, then by
248 type of biomass and time point from which the samples were taken (**Figure S4**).
249 Clustering organisms by abundance delineated several distinct groups: a small subset of
250 organisms was present in both reactors, while other subsets were found at high-
251 abundance in SCN^- community or the $\text{NH}_4(\text{SO}_4)_{1/2}$ community. Still other organisms
252 were abundant primarily in the inoculum (**Figure S4**). The taxonomic identity of

253 genomes was determined via phylogenetic reconstruction using ribosomal proteins
254 (Figure S5). In order to identify key organisms in the bioreactor communities, we
255 characterized the metabolic potential encoded and expressed by each genome with
256 respect to the key processes of SCN⁻ degradation, sulfur, ammonium, and nitrite
257 oxidation, denitrification, and carbon fixation (Figure S4, Table S2A).

258
259 **SCN⁻ removal and sulfur cycling:** Four genomes, *Thiobacillus_1*, *Thiobacillus_3*,
260 *Thiobacillus_4* and *Afipia_1*, contain one of two known types of SCN⁻ hydrolases^{11,40}.
261 The corresponding organisms were abundant only in the SCN⁻ reactor (Figure 2), and
262 proteomics data support activity of these organisms in SCN⁻ degradation, sulfur oxidation
263 and carbon fixation (Figure 3). Peptides detected by proteomics matched to predicted
264 proteins in the recently described SCN⁻ operon from *Thiobacillus* spp.²¹ Specifically, the
265 genes in the operon were detected by proteomics with spectral hits in at least one of the
266 three *Thiobacillus* genomes that contained this operon. Interestingly, an SQR-like protein
267 in this operon had the highest count of unique spectral hits for all three genomes (Table
268 S3). The *cbiM*-like protein hypothesized to be involved in cobalt metabolism was the
269 only protein from the operon not detected in proteomics.

270 Consistent with the genome of *Thiobacillus denitrificans*,⁴¹ all six recovered
271 *Thiobacillus* genomes encode the potential for autotrophic growth on sulfur compounds
272 via numerous sulfur-related genes from multiple pathways (Table S2A). In a
273 transcriptomics-based study of *Thiobacillus denitrificans*, some of these genes were
274 constitutively expressed (*sox*, *rDsr*, *apr*, *atps*) whereas others are upregulated under
275 denitrifying conditions (e.g., sulfide quinone reductase).⁴² We identified all sulfur
276 oxidation genes in proteomics for several of the *Thiobacilli* described here (Figure 3).

277 In addition to the four genomes encoding thiocyanate-degradation, twenty-two
278 other genomes possess the Sox pathway (including at least 4 of *soxX*, *Y*, *Z*, *A*, and *B*, with
279 or without *soxCD*; Table S2A). Of these, the genomes with the highest normalized
280 coverages in the SCN⁻ reactor were *Burkholderiales_6*, *Thiobacillus_2* and *Rhizobiales_3*
281 (Figure 2). Sulfur oxidation may proceed from sulfide to elemental sulfur or sulfate,
282 likely determined by the availability of electron acceptors, as discussed below. Sulfur
283 globules may be produced by SCN-degrading *Thiobacillus* spp., which use the reverse
284 dissimilatory sulfite reductase (*rDsr*) pathway instead of *soxCD*⁴³. In turn, other sulfur
285 oxidizers may use this elemental sulfur and any excess sulfide produced by SCN⁻
286 degradation. Since chemical data showed that SCN⁻ was completely converted to sulfate,
287 and proteomic data showed expression of Sox proteins from genomes lacking known
288 thiocyanate hydrolases (Figure 3), we suspect that sulfur species were passed from SCN-
289 degraders to the rest of community as “metabolic handoffs”. Additionally, some of these
290 organisms may use as-yet-unidentified pathways for SCN⁻ degradation.

291
292 **Community dynamics and SCN⁻ removal across increased loadings:** As SCN⁻
293 loadings and SCN⁻ degradation rate increased (Figure 1), the relative abundance of
294 *Thiobacillus* spp. whose genomes encode SCN-degradation also increased (Figure 2),
295 with *Thiobacillus_1* alone accounting for 38% of all reads at T2. During operation at the
296 two final loading rates (T3 and T4), the *Thiobacillus_1* population decreased in relative
297 abundance, concordant with a decrease in SCN⁻ degradation rate and reactor efficiency.

298 Given this observation, we looked for other changes associated with loss of reactor
299 efficiency.

300 First, a read mapping-based sequence variance analysis of the *Thiobacillus_1*
301 population in each sample showed that it was largely clonal throughout the time series
302 but two distinct strains were present at the last time point, where the relative strain
303 proportions were ~60 and 40% (**Figures S6A and S6B**). The genome for the second
304 strain was not recovered, but a scaffold corresponding to the SCN^- operon was identified
305 among the un-binned metagenomic data from this time point. We noted a few differences
306 in the protein sequences of genes contained in this operon, which could in principle affect
307 the efficiency of SCN^- degradation relative to the dominant strain (**Figure S7**).

308 A second change in the community was the increase in relative abundance of
309 *Burkholderiales_6*, which became dominant in T3 and T4 (**Figure S8**). No known genes
310 for SCN^- degradation were found in the *Burkholderiales_6* genome, but previous studies
311 have isolated and characterized a strain of *Burkholderia phytofirmans* capable of
312 thiocyanate degradation with acetate as a carbon source.⁴⁴ The genome of
313 *Burkholderiales_6* contains genes encoding the *sox* pathway, and the corresponding
314 proteins were detected in proteomics (**Figure 3**). Hence we infer that the
315 *Burkholderiales_6* organism likely used reduced sulfur species for mixotrophic growth,
316 and may have increased in relative abundance concordant with the accumulation of
317 organic matter in the reactor (in the form of biomass), perhaps outcompeting
318 *Thiobacillus_1* for SCN^- as a source of nitrogen and energy. Overall, the results suggest a
319 transition from autotrophic to mixotrophic / heterotrophic thiocyanate degradation at high
320 thiocyanate loadings and after long periods of reactor operation.

321 A third change in the reactor community that could have affected SCN^-
322 degradation was substantial algal growth at the last time point. This was visually
323 observed in the planktonic portion of the SCN^- reactor and could have affected microbial
324 population dynamics and reactor efficiency. Lastly, the increase in residual SCN^-
325 concentrations in the reactor may have led to toxicity effects including lower bacterial
326 replication rates (see below). This in turn could have reduced the SCN^- degradation rate,
327 creating a negative feedback effect on reactor performance.

328 Overall, we speculate that the decline in reactor efficiency at high loading rates
329 occurred when the capacity for SCN^- degradation was exceeded. The abundance of
330 *Thiobacillus* cells may have been insufficient to meet the demand for SCN^- degradation
331 owing to a maximum specific SCN^- degradation rate. Alternatively, degradation may
332 have been inhibited or the per-cell rate of degradation may have decreased. Since
333 metagenomic data provide relative abundance information, the apparent decrease in
334 *Thiobacillus_1* relative abundance may have been due to increases in abundance of other
335 organisms, such as *Burkholderiales_6*. Further studies are needed that apply
336 measurements of absolute, species-specific biomass and metabolic rates.

337

338 **Nitrogen removal and dynamics over time:** Since SCN^- degradation releases nitrogen
339 in the form of ammonium, we looked for possible mechanisms of nitrogen cycling and
340 removal to N_2 . No anamox genes were detected in any genome and, based on identified
341 genes and genomes, the conversion of ammonium to nitrate occurred aerobically. A
342 single genome in the dataset, *Nitrospira_1*, encoded the potential for ammonium
343 oxidation (**Figure S4**). The *Nitrospira_1* organism became enriched at early time points

344 in the $\text{NH}_4(\text{SO}_4)_{1/2}$ reactor and later in the SCN^- reactor (**Figure 2**) and was active at the
345 final time point, based on proteomic data (**Figure 3**). Two predicted nitrite oxidizers,
346 Nitrobacter_1 and Nitrobacter_2, were present at low abundances in the $\text{NH}_4(\text{SO}_4)_{1/2}$
347 reactor (**Figure 2**) and were so low in abundance in the SCN^- reactor that their genomes
348 did not assemble. However, proteins for nitrite oxidation corresponding to one of these
349 genomes were detected in samples from both reactors (**Figure 3**). Low abundance but
350 high activity has been observed previously for other nitrite oxidizing bacteria, and some
351 have hypothesized that high nitrite oxidation activity may be a requirement for growth,
352 given the low energy yield of this metabolism.⁴⁵ Despite the low relative abundance of
353 both predicted ammonium and nitrite oxidizers, nitrate was detected in the effluent of
354 both reactors during the initial ramping phases (**Figure 1**).

355 Searching for mechanisms of nitrate removal, we identified 92 genomes that
356 contained at least one gene involved in denitrification (including *nar*, *nap*, *nirS*, *nirK*,
357 *norB/norZ*, and *nosZ*; **Table S2A**). Fourteen of these genomes encoded the capacity for
358 complete nitrate reduction to N_2 , whereas 49 had only one gene in the pathway. Genomes
359 of complete denitrifiers corresponded to three of the predicted autotrophs implicated in
360 SCN^- removal: Thiobacillus_1, Thiobacillus_3, and Afipia_1 (**Figure 2**). For the
361 Thiobacilli, several denitrification-related complexes were detected with proteomics
362 (**Figure 3**). SCN^- hydrolysis and concomitant sulfide oxidation coupled to denitrification
363 may be possible in these organisms, as was described for *Thioalkalivibrio*.⁴⁶ Based on
364 proteomic evidence, the Burkholderiales_6 organism, which became abundant in T3 and
365 T4 in the SCN^- reactor, also likely contributed to denitrification (**Figure 2** and **3**). All
366 denitrification genes except *norB* were identified in proteomics from the SCN^- reactor
367 biofilm (**Figure 3**). The limited detection of NorB may be an extraction bias artifact due
368 to numerous transmembrane domains in these proteins.^{47,48}

369 In the $\text{NH}_4(\text{SO}_4)_{1/2}$ reactor, three of most abundant bacteria, Rhodanobacter_1,
370 Xanthomonadales_1, and Novosphingobium_1 may have roles in denitrification (**Figure**
371 **2** and **Table S2A**). The potential for dissimilatory nitrate reduction to ammonia (via *nrfA*)
372 was detected in genomes from several members of the Bacteroidetes and one *Aeromonas*
373 sp. but most of these were abundant only in the inoculum (**Table S2A**) and NrfA was not
374 detected in the proteomic data.

375
376 **Changes in bacterial replication rates across the time series:** We used a recently
377 established approach to investigate bacterial replication rates from metagenomics⁴⁹ with a
378 new implementation that reports rates as index of replication (iRep) values.⁵⁰ In the SCN^-
379 reactor biofilm, iRep values for most genomes increased between T0 and T1, suggesting
380 replication proceeded more quickly in newly-forming biofilm than in inoculum biofilm
381 taken from the long-running SCN^- stock reactor (**Figure S7**). Over the remainder of the
382 experiment (T2-T4), iRep values decreased for most genomes, especially toward the end
383 of the experiment. This may have been due to toxicity of residual SCN^- in the reactor
384 media or to spatial limitations for growth within the thickening biofilm.

385 In contrast, iRep values in the $\text{NH}_4(\text{SO}_4)_{1/2}$ reactor biofilm were initially low, but
386 increased from T1 to T3. This is suggestive of a period of adaptation, as organisms
387 adjusted to the new conditions relative to the SCN^- stock reactor (**Figure S9**).

388

389 **Biofilm and planktonic communities:** The planktonic and biofilm portions of the
390 reactor were sampled separately in order to understand whether these represented distinct
391 communities. At T1, the planktonic fraction of each reactor was very dilute, yielding
392 inadequate amounts of DNA for sequencing, and at T4, the planktonic portion of the
393 SCN^- reactor was overgrown with algae. At T2, metagenomes for planktonic samples
394 from both reactors were highly enriched in a rotifer genome, which accounted for 45 and
395 25 % of the sequence data in the $\text{NH}_4(\text{SO}_4)_{1/2}$ and SCN^- reactors, respectively. With
396 microscopy, rotifers were observed grazing on biofilm (**Figure S10**). Other eukaryotes
397 were observed in both reactors, and many of these organisms were at higher relative
398 abundance in the planktonic samples compared to biofilm (**Table S2**). Recurring
399 mitochondrial sequences recovered in the metagenomes corresponded to relatives of
400 *Vermamoeba vermiformis*, *Acanthamoeba* spp., *Naegleria fowleri*, and *Chlorella* sp.,
401 identified based on the phylogenetic profile of their proteins and searches against NCBI-
402 nr (**Table S2B**).

403 Eleven bacterial genomes in the dataset derived from predicted symbionts, as
404 indicated by their phylogenetic affiliations and/or reduced genome sizes (**Figure S4** and
405 **Table S2A**). These included two complete genomes for organisms belonging to the
406 phylum Saccharibacteria^{51,52} (formerly TM7). One organism, TM7_2, was the only
407 putative symbiont found at high abundances in planktonic fractions of both reactors
408 (**Figure 2**).

409 Overall, we estimate that biofilm constituted the majority of the biomass in the
410 reactors, and sloughing may have contributed to the planktonic community. While SCN^-
411 degradation itself does not rely on biofilm,^{53,54} the formation of biofilm effectively
412 uncoupled the HRT from bacterial growth rates, preventing wash-out as the HRT was
413 decreased. This may have allowed *Thiobacillus* spp., and nitrifiers to reach higher
414 population sizes than would otherwise have been possible, thereby converting higher
415 loadings of SCN^- to nitrate. The biofilm likely also provided anoxic conditions promoting
416 denitrification, as suggested by proteomic data (see above).

417
418 **Long-term community stability and phage susceptibility:** We compared the 114
419 bacterial genomes in this study to 86 genomes reconstructed in a prior study of the SCN^-
420 stock reactor and a daughter reactor fed cyanide and SCN^- (CN- SCN^- reactor) conducted
421 two years earlier.²¹ Thirty-one genomes were matched, overlapping by at least a total of 1
422 Mbp at 98% nucleotide identity (**Table S2A**). These included close relatives of the three
423 SCN^- -degrading *Thiobacillus* spp. and Burkholderiales_6 enriched in the SCN^- -treated
424 reactor studied here. Given the importance of these three populations, we looked for
425 evidence of recent phage infections, based on changes to CRISPR loci over time.
426 *Thiobacillus_1* has no CRISPR locus, perhaps making it more susceptible to acquisition
427 of mobile elements and phage attack. The CRISPR locus in *Thiobacillus_3* was identical
428 in all versions of this genome recovered from the current study, but was not recovered in
429 the previous study. The recovered *Thiobacillus_4* genomes belonged to two distinct
430 CRISPR sub-types that differed from one another in 12 spacers at the 3' end of the array:
431 sequences from biofilm and planktonic inoculum samples comprised one version, and
432 sequences from later in the time series (and those recovered previously in the SCN^- stock
433 reactor and the CN- SCN^- reactor) comprised the second version. Importantly, no spacers
434 from *Thiobacillus_3* and *Thiobacillus_4* targeted any sequence in the metagenomes (or

435 previous metagenomes from the SCN^- stock or CN-SCN reactors), suggesting little recent
436 phage interaction.

437

438 **System overview:** Relatives of known SCN^- -degrading chemolithoautotrophs (several
439 *Thiobacillus* spp. and one *Afipia* sp.) are predicted to oxidize the SCN^- -sulfur as their sole
440 energy source under both aerobic and anaerobic conditions in the SCN^- reactor (**Figure**
441 **4A**). Sulfide oxidation may stop at elemental sulfur when parts of the reactor become
442 anaerobic⁵⁵ (**Figure 4B**), and proteomic data suggest that several *Thiobacillus* spp., as
443 well as Burkholderiales_6, coupled sulfur oxidation to denitrification (**Figure 3**). In fact,
444 based on their abundance, and proteomic evidence, these organisms may be the key
445 denitrifiers in the system. Other sulfur oxidizing autotrophs and mixotrophs may use
446 reduced sulfur compounds produced during SCN^- degradation, including elemental
447 sulfur. The combination of sulfide oxidation by SCN^- -degraders and non-degraders may
448 explain the complete conversion of sulfur from SCN^- to sulfate (**Figure 4B**). The
449 breakdown of SCN^- produces ammonium that can be converted to nitrate by autotrophic
450 ammonium and nitrite oxidizers. Overall, some heterotrophs in the system likely
451 contributed to sulfur oxidation and denitrification, and likely also to biofilm formation
452 and biofilm integrity (perhaps via filamentous morphology, see **Figure S10**).
453 Heterotrophs may also break down SCN^- as a source of nitrogen (via an unknown
454 pathway).⁴⁴ However, this SCN^- degradation may be inhibited if there is abundant
455 nitrogen available as ammonium, as observed in some alkaliphiles.¹⁵ Lastly, eukaryotes
456 such as rotifers and amoeba are predators and thus contribute to carbon turnover in the
457 system.

458

459 **Engineering of SCN^- degradation by a microbial community:** The consortium
460 described here can completely hydrolyze SCN^- and oxidize sulfide under a range of SCN^-
461 loadings but reduced performance occurred at higher loadings. Smaller increases in
462 concentration to reach higher loadings may lead to sustained reactor performance by
463 allowing microbial cell numbers and associated volumetric degradation rates to keep pace
464 with input SCN^- . For treating such high SCN^- concentrations at low retention times
465 (>12.9 mM or 750 ppm in 12 hours), an attached growth design may have better SCN^-
466 and nitrogen removal.

467 Studies on such biofilm-based systems for SCN^- treatment have demonstrated the
468 effectiveness of rotating biological contact (RBC)^{56,57} and fixed bed and fluidized-carrier
469 type reactors.^{58,59} Potential weaknesses of these systems include biofilm overgrowth
470 leading to poor mixing and aeration. The ASTERTM process, modeled here with
471 laboratory-scale CSTRs, was designed to be easy to establish at remote locations and to
472 accommodate suspended solid tailings when necessary.^{53,54} The results of our work on a
473 laboratory-scale CSTR suggest that biofilm can play an important role when this reactor
474 design is operated at high loadings.

475 Our work highlights the applicability of bioinformatics tools to gain a mechanistic
476 understanding of contaminant degradation by a microbial community, to assess
477 community stability, and ultimately, to inform engineering design choices. Others have
478 called for broader use of metagenomics to advance biotechnology, including in
479 wastewater treatment,^{60,61} and this study represents a step toward the use of such
480 techniques in the field. The level of resolution achieved using metagenomics combined

481 with metaproteomics allowed us to access not only phylogenetic classifications and
482 diversity of community members, but also which members express key proteins involved
483 in the degradation process. The dataset and analysis provide valuable information that
484 can be used to generate primers or probes for on-site measurements.

485

486 **ACKNOWLEDGEMENTS**

487 We gratefully acknowledge assistance from Alexander Probst, Andrea Singh, Spencer
488 Diamond, Paula Matheus-Carnavali, A. Wynand Van Zyl, Robert Van Hille, Tanya
489 Hodgson, and Tijana Glavina del Rio. Funding was provided by an NSF Sustainable
490 Chemistry grant (1349278), NSF-GRFP and NSF-GROW with USAID fellowships to
491 RSK, the Department of Science and Technology (DST) and NRF of South Africa
492 through the SARChI Chair in Bioprocess Engineering (UID 64778), and an NRF
493 Research Career Advancement Fellowship to RJH. Oak Ridge National Lab received
494 support from the U.S. Department of Energy, Genome Science Program. The Joint
495 Genome Institute Emerging Technologies Opportunity Program provided sequencing,
496 and was in turn supported by the Office of Science of the U.S. Department of Energy
497 under Contract No. DE-AC02-05CH11231.

498

499 **SUPPORTING INFORMATION AVAILABLE**

500 This information is available free of charge at <http://pubs.acs.org>.

501

502

503 **AUTHOR CONTRIBUTIONS**

504 RSK, RJH, STLH, and JFB designed the study; RSK and JFB wrote the paper; all co-
505 authors read and contributed to revisions of the paper; RJH operated and sampled the
506 bioreactors and conducted analysis for metadata, with contributions from STLH; SGT
507 provided DNA sequencing; RI and RLH collected and processed proteomic data; RSK
508 analyzed metagenomic and proteomic data with contributions from BCT, CTB, and KA.

509 **FIGURES AND TABLES**

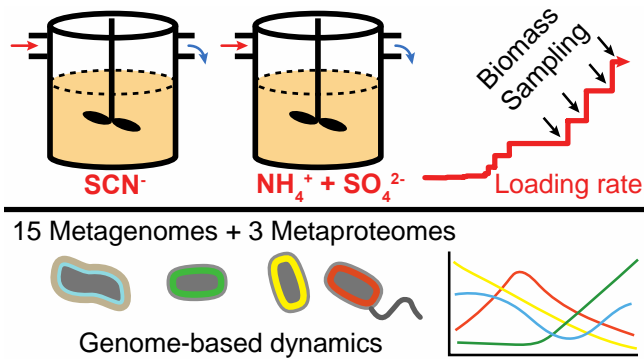
510

511 **Table 1.** Counts and completeness estimated by single copy gene (SCG) inventories for
 512 de-replicated bins resulting from 15 metagenome assemblies.

Bacterial genomes (144)	Count
Genomes \geq 96% complete (49/51 SCG)	111
Genomes in \leq 30 scaffolds and 49/51 SCG	37
Genomes with \geq 90% of sequence in scaffolds $>$ 10 kb	117
Non-bacterial bins (90)	Count
Plasmids and mobile elements	45
Phage	25
Eukaryotes	1
Mitochondria	15
Chloroplasts	1
Viruses	3

513

514



515

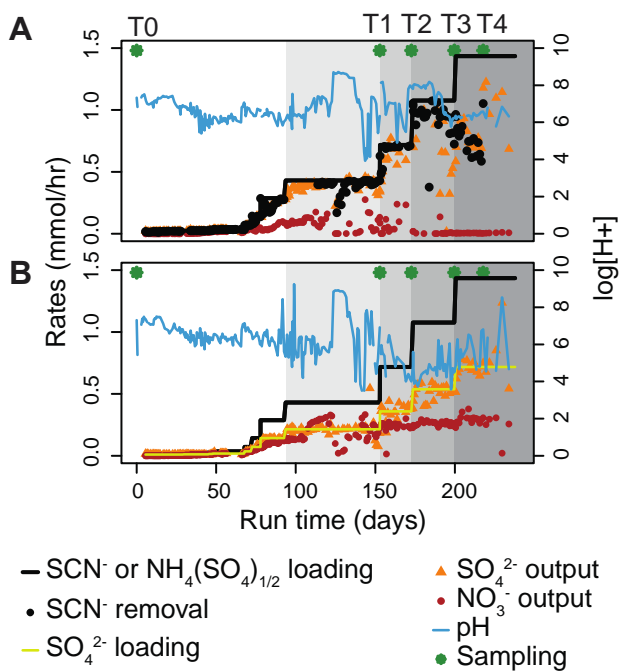
516

517

518

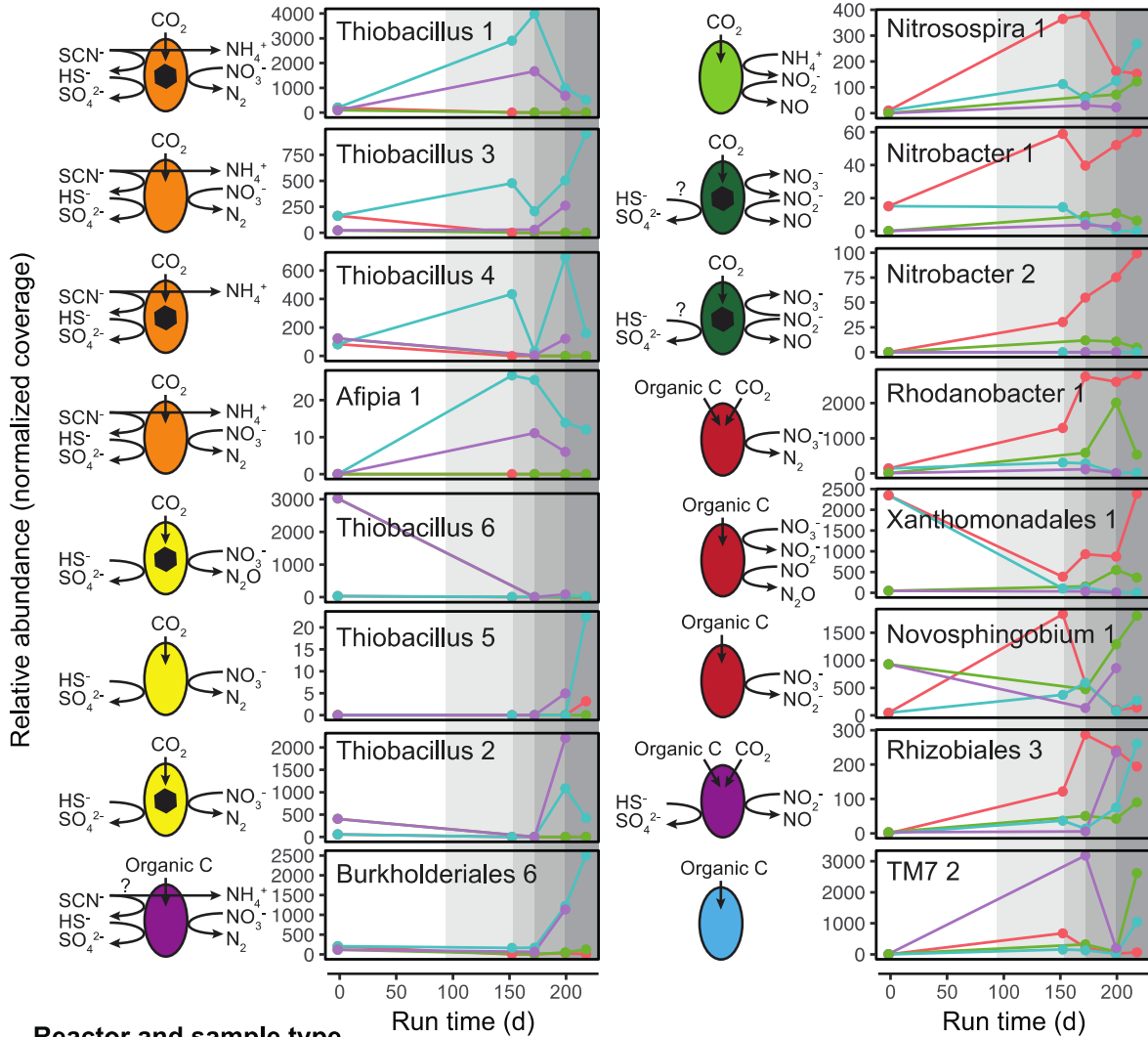
TOC/Abstract art

519



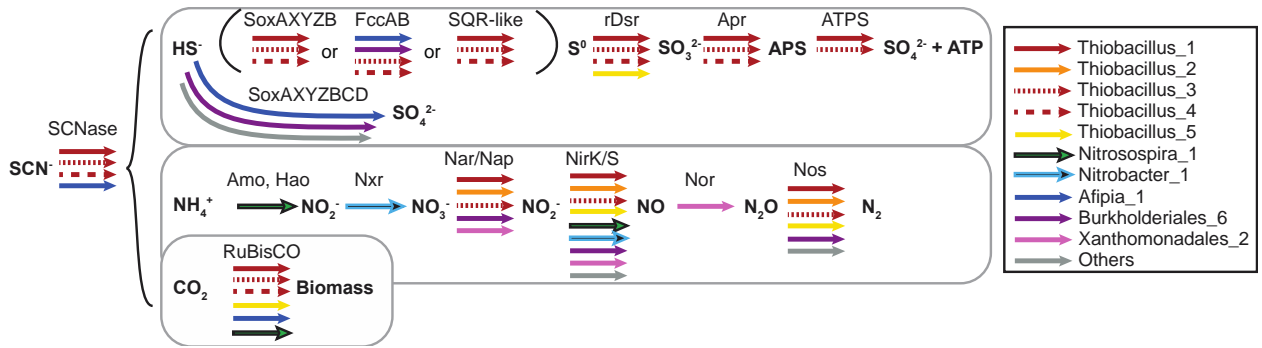
520
521

522 **Figure 1.** Chemical parameters of operation for the SCN⁻ reactor (A), and
523 NH₄(SO₄)_{1/2} reactor (B). Sampling time points are indicated above plots. Gray
524 shading indicates the different loading regimes.



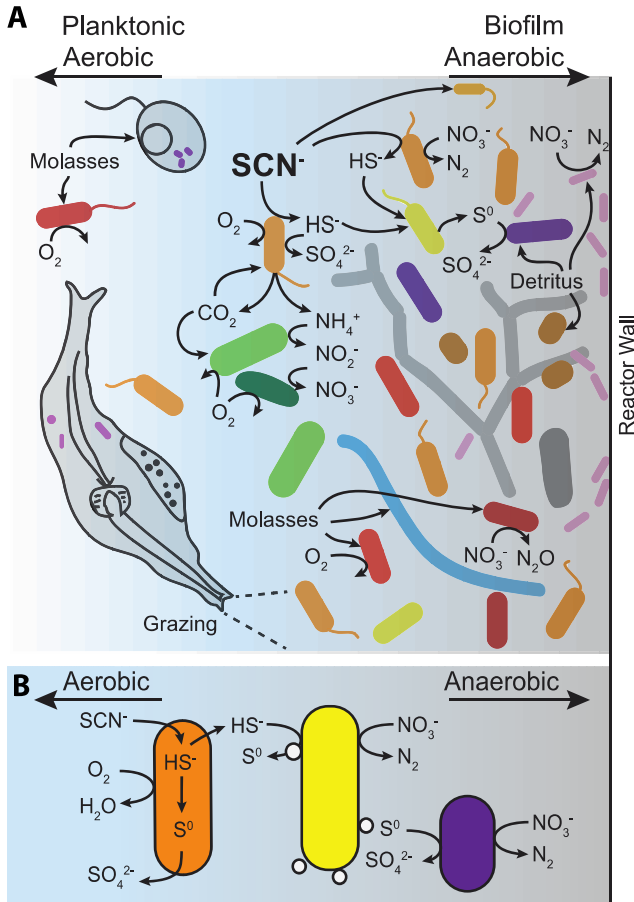
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541

Figure 2. Metabolic potential and relative abundances of key organisms of interest over time in biofilm and planktonic samples from both reactors. Gray shading corresponds to increasing loading rates of SCN⁻ or NH₄(SO₄)_{1/2} as in Figure 1. SCN⁻ degraders (orange) and several key sulfur oxidizers (yellow and purple) were found at higher relative abundance in the SCN⁻ reactor while ammonium and nitrite oxidizers (green) were at higher relative abundances in the NH₄(SO₄)_{1/2} reactor. Several highly abundant heterotrophs (red and blue) and one possible sulfur oxidizing mixotroph (purple) were present in both reactors. Note different y-axis scales. Hexagons indicate carboxysomes where annotations support this prediction.



542
543
544
545
546
547
548
549

Figure 3. Metaproteomics at end point (T4) in SCN⁻ reactor shows expression of genes involved in SCN⁻ degradation and byproduct breakdown in key organisms. Each arrow indicates that the average of unique spectral counts across two technical replicates was ≥ 2 for at least one subunit or component of the enzyme complex involved in the chemical transformation.



550
551
552
553
554
555

Figure 4. SCN^- removal and sulfur, nitrogen, and carbon cycling in the reactor system depicted based on metagenomic analysis.

556 **LITERATURE CITED**

557

- 558 (1) Speth, D. R.; In 't Zandt, M. H.; Guerrero-Cruz, S.; Dutilh, B. E.; Jetten, M. S. M.
 559 Genome-based microbial ecology of anammox granules in a full-scale
 560 wastewater treatment system. *Nature Communications* **2016**, *7*, 11172.
- 561 (2) Sekiguchi, Y.; Ohashi, A.; Parks, D. H.; Yamauchi, T.; Tyson, G. W.;
 562 Hugenholtz, P. First genomic insights into members of a candidate bacterial
 563 phylum responsible for wastewater bulking. *PeerJ* **2015**, *3*, e740.
- 564 (3) Albertsen, M.; Saunders, A. M.; Nielsen, K. L.; Nielsen, P. H. Metagenomes
 565 obtained by “deep sequencing” – what do they tell about the enhanced biological
 566 phosphorus removal communities? *Water Sci. Technol.* **2013**, *68* (9), 1959.
- 567 (4) Hug, L. A.; Beiko, R. G.; Rowe, A. R.; Richardson, R. E.; Edwards, E. A.
 568 Comparative metagenomics of three Dehalococcoides-containing enrichment
 569 cultures: the role of the non-dechlorinating community. *BMC Genomics* **2012**, *13*
 570 (1), 327.
- 571 (5) Lykidis, A.; Chen, C.-L.; Tringe, S. G.; McHardy, A. C.; Copeland, A.; Kyrpides,
 572 N. C.; Hugenholtz, P.; Macarie, H.; Olmos, A.; Monroy, O.; et al. Multiple
 573 syntrophic interactions in a terephthalate-degrading methanogenic consortium.
 574 *The ISME Journal* **2010**, *5* (1), 122–130.
- 575 (6) Nobu, M. K.; Narihiro, T.; Rinke, C.; Kamagata, Y.; Tringe, S. G.; Woyke, T.;
 576 Liu, W.-T. Microbial dark matter ecogenomics reveals complex synergistic
 577 networks in a methanogenic bioreactor. *The ISME Journal* **2015**, *9* (8), 1710–
 578 1722.
- 579 (7) Taubert, M.; Vogt, C.; Wubet, T.; Kleinstuber, S.; Tarkka, M. T.; Harms, H.;
 580 Buscot, F.; Richnow, H.-H.; Bergen, von, M.; Seifert, J. Protein-SIP enables
 581 time-resolved analysis of the carbon flux in a sulfate-reducing, benzene-
 582 degrading microbial consortium. *The ISME Journal* **2012**, *6* (12), 2291–2301.
- 583 (8) Erdogan, M. F. Thiocyanate overload and thyroid disease. *Biofactors* **2003**, *19*
 584 (3-4), 107–111.
- 585 (9) Speyer, M. R.; Raymond, P. The acute toxicity of thiocyanate and cyanate to
 586 rainbow trout as modified by water temperature and pH. *Environmental*
 587 *Toxicology and Chemistry* **1988**, *7* (7), 565–571.
- 588 (10) Watson, S. J.; Maly, E. J. Thiocyanate toxicity to *Daphnia magna*: modified by
 589 pH and temperature. *Aquatic Toxicology* **1987**, *10* (1), 1–8.
- 590 (11) Hussain, A.; Ogawa, T.; Saito, M.; Sekine, T.; Nameki, M.; Matsushita, Y.;
 591 Hayashi, T.; Katayama, Y. Cloning and expression of a gene encoding a novel
 592 thermostable thiocyanate-degrading enzyme from a mesophilic
 593 alphaproteobacteria strain THI201. *Microbiology* **2013**, *159* (Pt 11), 2294–2302.
- 594 (12) Katayama, Y.; Kuraishi, H. Characteristics of *Thiobacillus thioparus* and its
 595 thiocyanate assimilation. *Can. J. Microbiol.* **1978**, *24* (7), 804–810.
- 596 (13) Katayama, Y.; Narahara, Y.; Inoue, Y.; Amano, F.; Kanagawa, T.; Kuraishi, H.
 597 A thiocyanate hydrolase of *Thiobacillus thioparus*. A novel enzyme catalyzing
 598 the formation of carbonyl sulfide from thiocyanate. *J. Biol. Chem.* **1992**, *267* (13),
 599 9170–9175.
- 600 (14) Katayama, Y.; Matsushita, Y.; Kaneko, M.; Kondo, M.; Mizuno, T.; Nyunoya, H.
 601 Cloning of genes coding for the three subunits of thiocyanate hydrolase of

- 602 *Thiobacillus thioparus* THI 115 and their evolutionary relationships to nitrile
603 hydratase. *Journal of Bacteriology* **1998**, *180* (10), 2583–2589.
- 604 (15) Sorokin, D. Y.; Tourova, T. P.; Lysenko, A. M.; Kuenen, J. G. Microbial
605 thiocyanate utilization under highly alkaline conditions. *Applied and*
606 *Environmental Microbiology* **2001**, *67* (2), 528–538.
- 607 (16) Stratford, J.; Dias, A. E.; Knowles, C. J. The utilization of thiocyanate as a
608 nitrogen source by a heterotrophic bacterium: the degradative pathway involves
609 formation of ammonia and tetrathionate. *Microbiology (Reading, Engl.)* **1994**,
610 *140*, 2657–2662.
- 611 (17) Wood, A. P.; Kelly, D. P.; McDonald, I. R.; Jordan, S. L.; Morgan, T. D.; Khan,
612 S.; Murrell, J. C.; Borodina, E. A novel pink-pigmented facultative methylotroph,
613 *Methylobacterium thiocyanatum* sp. nov., capable of growth on thiocyanate or
614 cyanate as sole nitrogen sources. *Arch Microbiol* **1998**, *169* (2), 148–158.
- 615 (18) Boucabeille, C.; Bories, A.; Ollivier, P.; Michel, G. Microbial degradation of
616 metal complexed cyanides and thiocyanate from mining wastewaters. *Environ.*
617 *Pollut.* **1994**, *84* (1), 59–67.
- 618 (19) van Buuren, C.; Makhotla, N.; Olivier, J. W. The ASTER process: technology
619 development through to piloting, demonstration, and commercialization.
620 *Proceedings of the ALTA 2011 Nickel-Cobalt-Copper, Uranium and Gold*
621 *Conference* **2011**.
- 622 (20) Huddy, R. J.; van zyl, A. W.; van Hille, R. P.; Harrison, S. T. L. Characterisation
623 of the complex microbial community associated with the ASTER™ thiocyanate
624 biodegradation system. *Minerals Engineering* **2015**, *76*, 65–71.
- 625 (21) Kantor, R. S.; van zyl, A. W.; van Hille, R. P.; Thomas, B. C.; Harrison, S. T. L.;
626 Banfield, J. F. Bioreactor microbial ecosystems for thiocyanate and cyanide
627 degradation unravelled with genome-resolved metagenomics. *Environ Microbiol*
628 **2015**, *17* (12), 4929–4941.
- 629 (22) van Zyl, A. W.; Harrison, S. T. L.; van Hille, R. P. Biodegradation of thiocyanate
630 by a mixed microbial population. *Proceedings of the International Mine Water*
631 *Association Conference, 2011* **2011**, 119–124.
- 632 (23) Peng, Y.; Leung, H. C. M.; Yiu, S. M.; Chin, F. Y. L. IDBA-UD: a de novo
633 assembler for single-cell and metagenomic sequencing data with highly uneven
634 depth. *Bioinformatics* **2012**, *28* (11), 1420–1428.
- 635 (24) Sharon, I.; Morowitz, M. J.; Thomas, B. C.; Costello, E. K.; Relman, D. A.;
636 Banfield, J. F. Time series community genomics analysis reveals rapid shifts in
637 bacterial species, strains, and phage during infant gut colonization. *Genome*
638 *Research* **2013**, *23* (1), 111–120.
- 639 (25) Ultsch, A.; Mörchen, F. ESOM-Maps: tools for clustering, visualization, and
640 classification with Emergent SOM. *Data Bionics Research Group, University of*
641 *Marburg, Germany* **2005**.
- 642 (26) Hug, L. A.; Thomas, B. C.; Sharon, I.; Brown, C. T.; Sharma, R.; Hettich, R. L.;
643 Wilkins, M. J.; Williams, K. H.; Singh, A.; Banfield, J. F. Critical
644 biogeochemical functions in the subsurface are associated with bacteria from
645 new phyla and little studied lineages. *Environ Microbiol* **2016**, *18* (1), 159–173.
- 646 (27) Kurtz, S.; Phillippy, A.; Delcher, A. L.; Smoot, M.; Shumway, M.; Antonescu,
647 C.; Salzberg, S. L. Versatile and open software for comparing large genomes.

- 648 *Genome Biol* **2004**, 5 (2), R12.
- 649 (28) Brown, C. T.; Hug, L. A.; Thomas, B. C.; Sharon, I.; Castelle, C. J.; Singh, A.;
650 Wilkins, M. J.; Wrighton, K. C.; Williams, K. H.; Banfield, J. F. Unusual biology
651 across a group comprising more than 15% of domain Bacteria. *Nature* **2015**, 523
652 (7559), 208–211.
- 653 (29) Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.
654 *Bioinformatics* **2010**, 26 (19), 2460–2461.
- 655 (30) Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a
656 reference resource for gene and protein annotation. *Nucleic Acids Research* **2016**,
657 44 (D1), D457–D462.
- 658 (31) Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H.; The UniProt
659 Consortium. UniRef clusters: a comprehensive and scalable alternative for
660 improving sequence similarity searches. *Bioinformatics* **2015**, 31 (6), 926–932.
- 661 (32) The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids*
662 *Research* **2015**, 43 (D1), D204–D212.
- 663 (33) Finn, R. D.; Coghill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.;
664 Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A.; et al. The Pfam
665 protein families database: towards a more sustainable future. *Nucleic Acids*
666 *Research* **2016**, 44 (D1), D279–D285.
- 667 (34) Anantharaman, K.; Brown, C. T.; Hug, L. A.; Sharon, I.; Castelle, C. J.; Probst,
668 A. J.; Thomas, B. C.; Singh, A.; Wilkins, M. J.; Karaoz, U.; et al. Thousands of
669 microbial genomes shed light on interconnected biogeochemical processes in an
670 aquifer system. *Nature Communications* **2016**, 7, 13219.
- 671 (35) Hug, L. A.; Castelle, C. J.; Wrighton, K. C.; Thomas, B. C.; Sharon, I.;
672 Frischkorn, K. R.; Williams, K. H.; Tringe, S. G.; Banfield, J. F. Community
673 genomic analyses constrain the distribution of metabolic traits across the
674 Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome*
675 **2013**, 1 (1), 22.
- 676 (36) Chourey, K.; Jansson, J.; VerBerkmoes, N.; Shah, M.; Chavarria, K. L.; Tom, L.
677 M.; Brodie, E. L.; Hettich, R. L. Direct Cellular Lysis/Protein Extraction
678 Protocol for Soil Metaproteomics. *J. Proteome Res.* **2010**, 9 (12), 6615–6622.
- 679 (37) Tabb, D. L.; Fernando, C. G.; Chambers, M. C. MyriMatch: highly accurate
680 tandem mass spectral peptide identification by multivariate hypergeometric
681 analysis. *J. Proteome Res.* **2007**, 6 (2), 654–661.
- 682 (38) Raveh-Sadka, T.; Thomas, B. C.; Singh, A.; Firek, B.; Brooks, B.; Castelle, C. J.;
683 Sharon, I.; Baker, R.; Good, M.; Morowitz, M. J.; et al. Gut bacteria are rarely
684 shared by co-hospitalized premature infants, regardless of necrotizing
685 enterocolitis development. *eLife* **2015**, 4.
- 686 (39) Xiong, W.; Giannone, R. J.; Morowitz, M. J.; Banfield, J. F.; Hettich, R. L.
687 Development of an enhanced metaproteomic approach for deepening the
688 microbiome characterization of the human infant gut. *J. Proteome Res.* **2015**, 14
689 (1), 133–141.
- 690 (40) Arakawa, T.; Kawano, Y.; Kataoka, S.; Katayama, Y.; Kamiya, N.; Yohda, M.;
691 Odaka, M. Structure of Thiocyanate Hydrolase: A New Nitrile Hydratase Family
692 Protein with a Novel Five-coordinate Cobalt(III) Center. *Journal of Molecular*
693 *Biology* **2007**, 366 (5), 1497–1509.

- 694 (41) Beller, H. R.; Chain, P. S. G.; Letain, T. E.; Chakicherla, A.; Larimer, F. W.;
695 Richardson, P. M.; Coleman, M. A.; Wood, A. P.; Kelly, D. P. The genome
696 sequence of the obligately chemolithoautotrophic, facultatively anaerobic
697 bacterium *Thiobacillus denitrificans*. *Journal of Bacteriology* **2006**, *188* (4),
698 1473–1488.
- 699 (42) Beller, H. R.; Letain, T. E.; Chakicherla, A.; Kane, S. R.; Legler, T. C.; Coleman,
700 M. A. Whole-genome transcriptional analysis of chemolithoautotrophic
701 thiosulfate oxidation by *Thiobacillus denitrificans* under aerobic versus
702 denitrifying conditions. *Journal of Bacteriology* **2006**, *188* (19), 7005–7015.
- 703 (43) Loy, A.; Duller, S.; Baranyi, C.; Mussmann, M.; Ott, J.; Sharon, I.; Béjà, O.; Le
704 Paslier, D.; Dahl, C.; Wagner, M. Reverse dissimilatory sulfite reductase as
705 phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environ*
706 *Microbiol* **2009**, *11* (2), 289–299.
- 707 (44) Vu, H. P.; Mu, A.; Moreau, J. W. Biodegradation of thiocyanate by a novel strain
708 of *Burkholderia phytofirmans* from soil contaminated by gold mine tailings. *Lett.*
709 *Appl. Microbiol.* **2013**, *57* (4), 368–372.
- 710 (45) Baker, B. J.; Sheik, C. S.; Taylor, C. A.; Jain, S.; Bhasi, A.; Cavalcoli, J. D.;
711 Dick, G. J. Community transcriptomic assembly reveals microbes that contribute
712 to deep-sea carbon and nitrogen cycling. *The ISME Journal* **2013**, *7* (10), 1962–
713 1973.
- 714 (46) Sorokin, D. Y.; Tourova, T. P.; Antipov, A. N.; Muyzer, G.; Kuenen, J. G.
715 Anaerobic growth of the haloalkaliphilic denitrifying sulfur-oxidizing bacterium
716 *Thialkalicoccus thiocyanodenitrificans* sp. nov. with thiocyanate. *Microbiology*
717 *(Reading, Engl.)* **2004**, *150* (Pt 7), 2435–2442.
- 718 (47) Hino, T.; Matsumoto, Y.; Nagano, S.; Sugimoto, H.; Fukumori, Y.; Murata, T.;
719 Iwata, S.; Shiro, Y. Structural basis of biological N₂O generation by bacterial
720 nitric oxide reductase. *Science* **2010**, *330* (6011), 1666–1670.
- 721 (48) Matsumoto, Y.; Tosha, T.; Pislakov, A. V.; Hino, T.; Sugimoto, H.; Nagano, S.;
722 Sugita, Y.; Shiro, Y. Crystal structure of quinol-dependent nitric oxide reductase
723 from *Geobacillus stearothermophilus*. *Nat Struct Mol Biol* **2012**, *19* (2), 238–245.
- 724 (49) Korem, T.; Zeevi, D.; Suez, J.; Weinberger, A.; Avnit-Sagi, T.; Pompan-Lotan,
725 M.; Matot, E.; Jona, G.; Harmelin, A.; Cohen, N.; et al. Growth dynamics of gut
726 microbiota in health and disease inferred from single metagenomic samples.
727 *Science* **2015**, *349* (6252), 1101–1106.
- 728 (50) Brown, C. T.; Olm, M. R.; Thomas, B. C.; Banfield, J. F. Measurement of
729 bacterial replication rates in microbial communities. *Nature biotechnology* **2016**.
- 730 (51) Bor, B.; Poweleit, N.; Bois, J. S.; Cen, L.; Bedree, J. K.; Zhou, Z. H.; Gunsalus,
731 R. P.; Lux, R.; McLean, J. S.; He, X.; et al. Phenotypic and physiological
732 characterization of the epibiotic interaction between TM7x and its basibiont
733 *Actinomyces*. *Microb Ecol* **2015**, *71* (1), 243–255.
- 734 (52) He, X.; McLean, J. S.; Edlund, A.; Yooseph, S.; Hall, A. P.; Liu, S.-Y.;
735 Dorrestein, P. C.; Esquenazi, E.; Hunter, R. C.; Cheng, G.; et al. Cultivation of a
736 human-associated TM7 phylotype reveals a reduced genome and epibiotic
737 parasitic lifestyle. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (1), 244–249.
- 738 (53) van Zyl, A. W.; Huddy, R.; Harrison, S. T. L.; van Hille, R. P. Evaluation of the
739 ASTERTM process in the presence of suspended solids. *Minerals Engineering*

- 740 **2015**, 76, 72–80.
- 741 (54) Rahman, S. F.; Kantor, R. S.; Huddy, R. J.; Thomas, B. C.; van zyl, A. W.;
- 742 Harrison, S. T. L.; Banfield, J. F. During mine water bioremediation, the
- 743 presence of mineral particles decreases diversity and the abundance of
- 744 *Thiobacilli* responsible for thiocyanate degradation. *MicrobiologyOpen*.
- 745 (55) Moraes, B. S.; Souza, T.; Foresti, E. Effect of sulfide concentration on
- 746 autotrophic denitrification from nitrate and nitrite in vertical fixed-bed reactors.
- 747 *Process Biochemistry* **2012**, No. 47, 1395–1401.
- 748 (56) Whitlock, J. L. Biological detoxification of precious metal processing
- 749 wastewaters. *Geomicrobiology Journal* **1990**, 8 (3-4), 241–249.
- 750 (57) Stott, M. B.; Franzmann, P. D.; Zappia, L. R.; Watling, H. R.; Quan, L. P.; Clark,
- 751 B. J.; Houchin, M. R.; Miller, P. C.; Williams, T. L. Thiocyanate removal from
- 752 saline CIP process water by a rotating biological contactor, with reuse of the
- 753 water for bioleaching. *Hydrometallurgy* **2001**, 62 (2), 93–105.
- 754 (58) Dictor, M. C.; Battaglia-Brunet, F.; Morin, D.; Bories, A.; Clarens, M. Biological
- 755 treatment of gold ore cyanidation wastewater in fixed bed reactors. *Environ.*
- 756 *Pollut.* **1997**, 97 (3), 287–294.
- 757 (59) Jeong, Y.-S.; Chung, J. S. Biodegradation of thiocyanate in biofilm reactor using
- 758 fluidized-carriers. *Process Biochemistry* **2006**, 41 (3), 701–707.
- 759 (60) Ju, F.; Zhang, T. Experimental design and bioinformatics analysis for the
- 760 application of metagenomics in environmental sciences and biotechnology.
- 761 *Environ. Sci. Technol.* **2015**, 49 (21), 12628–12640.
- 762 (61) Roume, H.; Heintz-Buschart, A.; Muller, E. E. L.; May, P.; Satagopam, V. P.;
- 763 Laczny, C. C.; Narayanasamy, S.; Lebrun, L. A.; Hoopmann, M. R.; Schupp, J.
- 764 M.; et al. Comparative integrated omics: identification of key functionalities in
- 765 microbial community-wide metabolic networks. *npj Biofilms and Microbiomes*
- 766 **2015**, 1, 15007.
- 767