

1 **Microbial Response to Natural Disturbances: Rare Biosphere often**
2 **plays a role**
3

4 Jianshu Zhao^{1,2}, Genevieve Brandt², Zhao Wang³, Dana E. Hunt^{3,4}, Luis M. Rodriguez-
5 R^{5,6,7}, Janet K. Hatt⁵, Konstantinos T. Konstantinidis^{1,2,5,*}

6 ¹Center for Bioinformatics and Computational Genomics, Georgia Institute of Technology,
7 Atlanta, Georgia, USA

8 ²School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

9 ³Duke University, Marine Laboratory, Beaufort, NC 28516, USA

10 ⁴Biology Department, Duke University, Durham, NC, USA

11 ⁵School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta,
12 Georgia, USA

13 ⁶Department of Microbiology, University of Innsbruck, Innsbruck, Austria

14 ⁷Digital Science Center (DiSC), University of Innsbruck, Innsbruck, Austria

15 **Running title: Response of rare biosphere to natural disturbance**

16 *** Address correspondence to:**

17 **Konstantinos Konstantinidis**, kostas@ce.gatech.edu, School of Civil & Environmental
18 Engineering, Georgia Institute of Technology. 311 Ferst Drive, ES&T Building, Room
19 3321, Atlanta, GA, 30332. Telephone: 404-639-4292.
20

21 **Abstract**

22 Understanding how microbial populations respond to disturbances represents a major
23 goal for microbial ecology. While several theories have been advanced to explain
24 microbial community compositional changes in response to disturbances, appropriate
25 data to test these theories is scarce, especially when considering the challenges to define
26 rare vs. abundant taxa and generalists vs. specialists, a prerequisite for testing the
27 theories. Here, we define these two key concepts by employing the patterns of coverage
28 of a (target) genome by a metagenome to define rare populations, and by borrowing
29 concepts from macroecology, the proportional similarity index (PS index), to define
30 generalists. Using these concepts, we found that coastal microbial communities are
31 resilient to major perturbations such as tropical cyclones and (uncommon) cold or warm
32 weather events snaps -in part- due to the response of rare populations, providing support
33 for the insurance hypothesis (i.e., the rare biosphere has the buffering capacity to mitigate
34 the effects of disturbances). Generalists appear to contribute proportionally more than
35 specialists to community adaptation to perturbations like warming, supporting the
36 disturbance-specialization hypothesis, i.e., disturbance favors generalists. Taken
37 together, our results advance understanding of the mechanisms governing microbial
38 populations dynamics under changing environmental conditions and have potential
39 applications for ecosystem management.

40

41 **Keywords:** Rare biosphere, metagenomics, metagenomics assembled genomes
42 (MAGs), disturbance, resilience, generalists and specialists

43 **Introduction**

44 Natural environmental microbial communities generally harbor a few abundant
45 taxa and numerous low abundance, or rare, taxa [1]. In recent years, the importance of
46 the rare biosphere has been increasingly emphasized. Rare biosphere is defined as the
47 low-abundance active or dormant microbial taxa in a given environment at a specific time
48 point, typically showing <0.1% relative abundance. Low-abundance taxa are important
49 contributors to both α - and β -diversity, and rare taxa have been often shown to have
50 important ecological roles in specific ecosystems [2]. Both the number of taxa
51 (phylogenetic diversity) and the genes these taxa carry (functional diversity) are thought
52 to provide an ‘insurance’ or ‘buffering/caching’ capacity for the ecosystem against
53 environmental change [3,4]. For example, a rare coastal, oil-degrading bacterial
54 population thrived (e.g., made up ~30% of the total community) three weeks after an oil
55 spill and became rare two months later, after the oil was degraded [5]. Members of the
56 rare biosphere have also been shown to become abundant in a bacterioplankton
57 community after disturbance and play important roles in maintaining ecosystem
58 processes [6]. Rare taxa can periodically become abundant, and their abundance
59 dynamics may depend on selection pressure such as those imposed by seasonal
60 fluctuations in environmental parameters and substrate availability, and/or ecological
61 processes including dispersal, predation, reactivation from dormancy, functional
62 redundancy, plasticity, and diversification [7,8].

63 Several ecological theories have been advanced to explain the responses of
64 abundant and rare microbial populations to disturbance. For example, the insurance

65 hypothesis suggests that rare biosphere represents a strategy for responding to
66 temporally variable environments, contributing to community resilience [9,10].
67 Specifically, the hypothesis predicts that rare species, probably beyond the limit of
68 detection and thus, not included in estimates of community diversity, may quickly respond
69 to altered environmental characteristics (pulse disturbance) and become abundant before
70 returning to pre-perturbation (low) abundances [11]. On the other hand, the specialization-
71 disturbance hypothesis suggests that niche breadth, rather than relative abundance, of a
72 species is important for how the species responds to disturbance. That is, disturbance is
73 usually expected to affect specialists negatively, while generalists are thought to benefit
74 from disturbance [12]. Defining generalist versus specialist taxa in environmental samples
75 relies on their prevalence/abundance or niche width: generalists generally show no
76 preference for specific environments contrasting with specialists, which are abundant only
77 in some environments or conditions. Both theoretical and laboratory research has shown
78 that generalist taxa can be crucial in maintaining ecosystem functioning under
79 fluctuating/disturbed environments compared to specialists due to their metabolic
80 flexibility [13,14], although defining generalist vs. specialists is practically challenging due
81 to difficulties in defining the niche breadth that each species may occupy and determining
82 relative abundance when this is low (see also below).

83 Although the rare biosphere often plays an essential role in community function
84 and stability, defining rare and abundant members has largely employed arbitrary
85 divisions. Several different relative abundance thresholds have been proposed (e.g.,
86 <1%, 0.1% or 0.01% of the total community), without attention to ecological theory or

87 habitat differences [15-17]. Moreover, according to random sampling theory, defining rare
88 microbial taxa in comparison to abundant taxa based on 16S rRNA gene amplicon
89 sequence data can be subjected to overestimates of relative abundance due to
90 amplification biases (e.g. abundant populations are favored during amplification and/or
91 sequencing) [18]. More importantly, amplicon sequencing provides little information on
92 population-level functional potential and thus, offers limited information on the roles of
93 abundant and rare taxa in maintaining important ecosystem processes. Metagenomics
94 provides the means to capture the functional diversity of rare biosphere and largely
95 sidestep the biased amplification of abundant taxa based on gene-amplicon data [19,20]
96 but it remains challenging to reconstruct the metagenome-assembled genomes (MAGs)
97 of rare taxa due to insufficient sampling (i.e., sequence coverage), and to define rare vs.
98 abundant taxa at the genome level [21].

99 In this study, we sought to quantify how often rare populations respond to
100 disturbances and what functions rare populations provide relative to abundant taxa. In
101 other words, we aimed to directly test the specialization-disturbance and the insurance
102 hypotheses. For this, we sampled surface water at a coastal observatory (PICO) at
103 Beaufort, North Carolina (USA) weekly, for a three-year period, during which disturbance
104 events occurred (initially identified by microbiome changes, see below), with at least 1 to
105 3 weeks of season-typical weather between these events (Figure S1). Disturbance were
106 defined previously by comparing whether overall community composition and structure
107 (e.g. beta diversity) changed by the event (e.g., cyclone, bloom) relative to the
108 composition that is typical for the same season that the event occurred [22]. Shotgun

109 metagenomes before, during and after these events were used to define abundant and
110 rare population based on MAG coverage (e.g., genome depth and breadth covered by
111 mapped reads) and assess how many rare populations responded to the events. In this
112 process, we also defined the limit of detection of our metagenomic effort based on the
113 species abundance distribution curve as well as generalists vs. specialists based on a
114 proportional similarity index (PS index). Therefore, this study not only quantifies the
115 importance of the rare biosphere during community response to disturbance but also
116 outlines a methodology to define challenging concepts for microbial ecology, such as rare
117 versus abundant population and specialists versus generalists.

118

119 **Materials and Methods**

120 *Sample collection and description of disturbance events*

121 Water samples were collected as part of the Piver's Island Coastal Observatory (PICO)
122 time series adjacent to the Beaufort Inlet, Beaufort, NC, USA (34°71.81'N, 76°67.08'W)
123 on a weekly basis [22]. Here, we sequenced 19 samples from January 2011 to December
124 2013 that represented the winter, summer, and spring seasons. For each season in each
125 year, 3 samples were sequenced representing non-disturbance, disturbance, and post-
126 disturbance samples of disturbance events. Disturbance events were identified as
127 significant changes compared to the seasonal microbial community composition and
128 structure based on 16S rRNA gene amplicon sequencing [22] (**Figure S1**, Table S1). We
129 also looked at the actually environmental parameters that might have driven those
130 changes among the parameters measured. Briefly, for the six disturbance events

131 identified previously by Gronniger *et al.* [22], disturbance 1 was a warm and windy week
132 in winter of 2011 (50% less ammonium in the disturbance sample relative to non-
133 disturbance samples), disturbance 2 was the cyclone in summer of 2011 (Hurricane Irene,
134 65% more ammonium, 80.2% more silicates and 78.2% more chlorophyll A), disturbance
135 3 was a rainy week in winter 2012 (10 times lower than the two low tides, 3 times more
136 ammonium), disturbance 5 was a warm week in spring of 2012 (116% more NO_x, 55.9%
137 less silicates and 31.7% less chlorophyll A), disturbance 8 was a warm week in spring
138 2013 (100.1% more chlorophyll A and 59% less ammonium), disturbance 9 was the
139 tropical cyclone in summer 2013 (220.3% more ammonium, 49.6% more silicates and
140 20.2% more chlorophyll A, 1.5 °C higher). We refer to those disturbance events as
141 win11_warm_1, sum11_cyclone_2, win12_rainy_3, spr12_warm_5, spr13_warm_8,
142 sum13_cyclone_9, respectively, with the last number corresponding to the disturbance
143 events label shown in Figure S1 and elsewhere, which is also consistent with our previous
144 publication [22]. Seawater was collected at 10:30 h local time using a 5-liter Niskin bottle
145 centered at 1 m on a peristaltic pump with the tubing open at 1 m and processed within 1
146 h. Standard laboratory methods for determination of water temperature, pH, salinity,
147 dissolved inorganic nutrient concentrations, and chlorophyll-a concentrations were
148 described previously [23]

149

150 *DNA extraction, amplicon sequencing analysis*

151 Microbial biomass was collected by filtering ~1 liter of seawater through a 0.22-micron
152 Sterivex filter (Millipore, Darmstadt, Germany) and the filters were stored at -80 °C until

153 DNA extraction. Genomic DNA was extracted using the phenol-chloroform lysis
154 supplemented with bead beating (60 seconds) and then subsequently cleaned using the
155 Zymo OneStep PCR inhibitor removal kit. Extracted DNA was quantified using a
156 Nanodrop ND-100 before sequencing. 16S rRNA gene amplicons from each sample were
157 sequenced using the primers targeting the V3-to-V4 region of the bacterial and archaeal
158 16S rRNA genes: for 16S F V3, CCTACGGGNGGCWSCAG; and for 16S R V4,
159 GGACTACNVGGGTWTCTAAT as described previously [24]. PCR reactions contained
160 10 ng of template DNA, 1.25 U Econo Taq (Lucigen) and a final concentration of 1× Taq
161 buffer, 200 μM dNTPs, 2 mM MgCl₂ and 0.5 μM of each primer. PCR reactions were
162 performed with the following protocol: 98°C for 30 seconds followed by 35 cycles at 98°C
163 for 10 seconds, then 55°C for 30 seconds and 72°C for 30 seconds, with a final extension
164 at 72°C for 2 minutes. Triplicate reaction mixtures per sample were pooled and gel
165 purified. Paired-end 250-bp sequencing of barcoded amplicons was performed on an
166 Illumina MiSeq running v2 chemistry at the Duke Center for Genomic and Computational
167 Biology.

168 USEARCH v11.0.667 was used for quality control and merging of paired-end reads. We
169 first trimmed low-quality bases from the sequences using a 10- nucleotide (nt) window
170 with a Q30 running-quality threshold. Paired-end sequences with a ≥ 10 -nt overlap and
171 no mismatches were then merged. We performed a final filtering step to discard low-
172 quality merged sequences with a length of < 400 bp and/or a maximum expected error
173 of > 1 . Amplicon sequence variants (ASVs) were then identified using the UNOISE3
174 algorithm in USEARCH [25], which has been shown to be even more accurate than

175 DADA2 [26]. Taxonomy classification of representative ASV sequences was performed
176 using SINTAX (-sintax_cutoff 0.8, a significance threshold similar to the 50% bootstrap
177 cutoff accuracy of the RDP naïve classifier) against SILVA v138.1 in USEARCH [27,28].
178 MacQIIME v1.9.1 was used for rarefaction, alpha diversity, beta diversity and community
179 composition analysis [29].

180

181 *Metagenomic sequencing, quality control and coverage estimation*

182 DNA was extracted with the MoBio Power Soil kit (MoBio Inc. Carlsbad, CA, USA). 10 ng
183 of DNA was then sheared to 300 bp using the Covaris LE220 and size was selected using
184 SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing,
185 and ligation of Illumina compatible adapters using the KAPA-Illumina library creation kit
186 followed by 5 cycles of PCR to enrich for the final library. These libraries were sequenced
187 with 2 150 nt reads on the Illumina HiSeq 2500 1T platform at either Duke's Genome
188 Sequencing or the Department of Energy's Joint Genome Institute for 300 cycles. Adapter
189 trimming and demultiplexing of sequenced samples was carried out by the instrument.
190 Raw reads were trimmed using Trimmomatic with default parameters [30] and then
191 checked using FastQC (<https://github.com/s-andrews/FastQC>). Sequence subsampling
192 to account for sequencing depth variation among libraries was done using Seqtk and
193 specifically, the subseq command with the same random seed for forward and reverse
194 reads (<https://github.com/lh3/seqtk>). Sequencing coverage estimation and Nonpareil
195 diversity were calculated using Nonpareil v3.0 with -kmer option [31].

196

197 *16S rRNA gene-carrying read extraction from metagenomes and closed reference OTU*
198 *picking*

199 Prokaryotic 16S rRNA gene-carrying metagenomic reads were extracted using metaxa2
200 [32]. USEARCH closed_ref workflow in USEARCH v11.0.667 were performed to pick
201 closed reference OTUs (not ASVs since we cannot cluster 16S short sequences extracted
202 from metagenome) from the extracted 16S rRNA gene-carrying (16S) short reads with
203 default identity 97% [29,33]. Briefly, USEARCH was used to align extracted short 16S
204 rRNA gene reads against the non-redundant SILVA database v138.1. If the semi-global
205 alignment identity of a query sequence to a database sequence was better than 97%, the
206 input sequence was assigned to that OTU. Otherwise, the input sequence was discarded
207 and not assigned to an OTU (an identity threshold of 99% was also used but only small
208 differences were observed in the results in terms of community composition, especially
209 at higher than the family taxonomic levels). Extracted 16S rRNA gene reads were rarefied
210 to 8000 sequences/sample before downstream analysis (Figure S3). Note that closed
211 reference OTU clustering may not find an OTU for a (query) sequence that is not
212 represented with a closely related sequence in the SILVA database, contrasting with de
213 novo OTU/ASV clustering using amplicons. However, such sequences were rather rare
214 in our datasets, representing <5% of the total sequences for each sample; hence, no
215 further action was taken to deal with this issue. Downstream analyses including diversity
216 and compositions analysis were performed using MacQIIME v1.9.1.

217

218 *Metagenome assembly, contig taxonomic classification, coverage calculation and*
219 *functional gene diversity analysis*

220 Quality controlled short reads were assembled with Megahit v2.1.2 (parameters: --meta -
221 -min_contig 1000) [34]. To classify the assembled contigs taxonomically, Centrifuge was
222 used to search against the RefSeq complete genome collection with default parameters
223 [35]. Genes were predicted on contigs using Prodigal v2.6.3 (-p meta) [36]. After mapping
224 short reads onto each contig and the genes of a contig using bwa-mem2 [37], contig/gene
225 coverage was calculated by the CoverM v0.6.1 (<https://github.com/wwood/CoverM>),
226 contig workflow with abundance metric option metabat (--methods metabat --min-read-
227 percent-identity 0.95 --min-read-aligned-percent 0.75). Diamond v0.9.22 was used to
228 perform functional annotation of predicted genes against the Swiss-Prot database
229 (Diamond blastp -k 1 --id 40 --query-cover 70 --max-hsps 4 -e 0.0001) [38]. The matching
230 Swiss-Prot reference sequences were mapped to GO terms and filtered for molecular
231 functions. The Chao-Shen estimate of Shannon entropy H for molecular functions and
232 cellular processes was calculated using the R package Entropy based on the observed
233 read counts (coverage) of genes assigned to the same molecular function GO term. This
234 estimation adjusts for missing species (here GO terms) and sample coverage. The
235 exponential of the estimated Shannon entropy was used to convert the statistic to true
236 diversity (1D) with units of effective GO terms as described recently [39]. The fraction of
237 the total proteome devoted to extracellular proteins for each MAG was predicted using
238 psortb [40].

239

240 *Population genome binning, and bin/MAG refinement*

241 Maxbin 2, Metabat2 and CONCOCT were individually applied to contig binning with
242 default parameters [41-43]. Resulting MAGs were first refined (quality improvement)
243 using DAS Tools with searching engine USEARCH [33,44]. Refined MAGs were further
244 checked by CheckM unique command, to ensure that no contigs were binned to the same
245 MAG more than once and one contig was not binned to different MAGs. Mis-binned
246 contigs were removed from MAGs. Contig coverage depth and tetranucleotide frequency
247 of each MAG were subsequently evaluated to determine whether binned MAGs likely
248 represent chimeric sequences using an in-house R script (available at
249 https://github.com/jianshu93/bin_check). CheckM lineage_wf was then used for MAG
250 quality assessment with default parameters [45]. The quality score was defined as
251 $\text{completeness} - 5 * \text{contamination} + \text{Stain heterogeneity} * 0.5$. Medium to high quality MAGs
252 were defined as quality score larger than 0.5. To further check whether those MAGs
253 binned from each sample represent sequence-discrete populations (species) [46],
254 competitive read mapping on the MAGs followed by recruitment plot were performed via
255 the scripts in the enveomics R package [47] (a new pipeline based on the scripts is also
256 available here: https://github.com/jianshu93/RecruitmentPlot_blast). Briefly, contig
257 sequences of all MAG from the same sample were labeled and then pooled together as
258 one genome database. Subsequently, blastn search was performed to map quality-
259 controlled short reads to the contigs (-task blastn -id 95% -max_target_seqs 500). The
260 blastn tabular output was filtered to remove mapped reads with low alignment ratio (<90%
261 of read length) and to only keep the best match for each query read according to identity.

262 Tied matches were also removed before creating recruitment plot using the
263 BlastTab.recplots2.R. MiGA quality_wf with MyTaxa option was then performed to further
264 validate the taxonomic identity of contigs binned into MAGs [48] and also calculate gene
265 coding density, %G+C content, and other descriptive statistics such as contig length [49].

266

267 *MAG dereplication and classification*

268 Dereplication of MAGs was performed using dRep v2.2.4 with an ANI threshold of 95%
269 by the -fastANI option (minimum completeness 70%, maximum contamination 10% and
270 quality score > 0.5 for filtering MAGs before dereplication) [50]. Quality information from
271 CheckM was passed to dRep to select the best quality MAG as representative of each
272 resulting 95% ANI cluster. MAGs were classified against the Genome Taxonomy
273 Database v214 [51] using GTDB-Tk classify_wf workflow [52], which classifies MAGs
274 based on their placement in a reference tree inferred using pplacer analysis of a set of
275 120 bacterial and 122 archaeal concatenated gene markers, combined with FastANI for
276 the species-level assignments [53,54]. To further confirm the taxonomy classified by
277 GTDB-Tk, especially those that were not classified with confidence by GTDB-Tk (e.g.,
278 distantly related to their best matches found in GTDB), we also used the MiGA workflow
279 miga classify_wf against the type material database [49]. The lowest taxonomic level to
280 which MiGA considered the assignment of the query MAG as significant was kept and
281 compared to the GTDB-Tk classification results.

282

283 *Relative abundance calculation and functional gene annotation of MAGs*

284 The relative abundance of each dereplicated MAG was calculated by competitively
285 mapping reads from each sample to the entire dereplicated MAG collection using bowtie2
286 and then SAMTools to generate sorted BAM files for each MAG [55,56]. Bam files were
287 subsequently filtered using the CoverM workflow (--min-read-percent-identity 0.95 --min-
288 read-aligned-percent 0.75) and only reads with alignment ratio and identity larger than
289 75% and 95%, respectively, were kept. Truncated Mean Depth 80% (TAD80) was then
290 calculated as a proxy for relative DNA abundance, which normalizes for highly conserved
291 or variable regions of the genome [57]. TAD80 estimates were further normalized by
292 genome equivalents based on MicrobeCensus [58] to account for average genome size
293 differences among the sample and provide the final (normalized) relative abundance
294 estimates, using an in-house script (https://github.com/jianshu93/Competitive_mapping)
295 [57]. Functional annotation of the dereplicated MAGs were performed using
296 MicrobeAnnotator and DRAM [59,60]. Briefly, MicrobeAnnotator searches multiple
297 reference protein databases iteratively and combines results from KEGG Orthology (KO),
298 Enzyme Commission (E.C.), Gene Ontology (GO), Pfam and InterPro, and returns the
299 matching annotations together with key metadata. DRAM profiles microbial
300 (meta)genomes for metabolisms known to impact ecosystem function across biomes like
301 carbon degradation, photosynthesis, methanogenesis, etc.

302

303 *Proportional similarity index for defining generalist and specialist MAGs*

304 Levin's niche breadth index (average relative abundance of species in different
305 environments) has been recently used to define generalists vs. specialists for microbial
306 populations [57,61,62]. However, this metric assumes equal availability of resources,
307 which is hardly the case in natural environments. A metric based on the proportion
308 similarity index, which defines generalists and specialists independently of their absolute
309 abundance and occupancy, was used here to circumvent this limitation according to
310 previous studies [63,64]. Specifically, habitat breadth was defined as the proportional
311 similarity (PS) index [63] relating the proportion of a population (or species) found in each
312 category of samples to the proportion of sampling effort (total samples) for that category
313 as follows: $PS_{index} = 1 - 0.5 \sum_i |p_i - q_i|$, where p_i is the number of cells (we use relative
314 abundance, i.e., TAD/genome equivalent) of the target species in samples of category i ,
315 divided by the total number of cells from that species in all available samples, while q_i is
316 the number of all the cells in samples of category i , divided by all the cells in all samples.
317 Thus, PS index values express variation in the habitat breadth of a species, which is an
318 important aspect of the realized niche of the species (preferences for
319 samples/environments), and range between 0 and 1 for the broadest possible and the
320 narrowest possible niche, respectively (i.e., a population is restricted exclusively to the
321 rarest category of sample types and consequently, is absent in all other types).
322 Genomespecies captured in only one water sample were classified as rare species and
323 excluded from the habitat breadth analysis as no reliable measure of niche breadth could
324 be calculated for such genomespecies. For our time series dataset, samples were first
325 grouped into three main types (or environments) according to seasonality, and then to

326 two or three sub-types within each main type according to the year that the corresponding
327 samples was obtained (Table S3). The PS index was calculated separately for each
328 category of samples (the three main types and the eight sub-types), generating two
329 scores (PS index_{subtype} & PS index_{maintype}). Species that scored for both categories in the
330 upper third percentile, indicating a broad niche, were classified as habitat generalists.
331 Species that scored for both categories in the bottom third percentile were classified as
332 habitat specialists, and were further subdivided into regular habitat specialists (rank-
333 transformed; that is, the rank of the PS values, PS index_{subtype} > PS index_{maintype}) and strict
334 habitat specialists (rank-transformed; PS index_{subtype} < PS index_{maintype}) [64](Table S3).
335 According to this definition, species were classified as generalists or specialists
336 independent of their relative abundance in one category of samples. This definition allows
337 abundant (in some samples but not all) and widespread species (not equally abundant
338 for all samples) to be also classified as habitat specialists when high proportions of
339 individuals are found in a single category of samples (e.g., single season) but not in all
340 category of samples [64]. We also compared this PS index with the Levin's Breadth Index,
341 and also the PS index implementation in MicroNiche package, which provides null model
342 tests with respect to the limit of detection [65]. Specifically, for PS index by the
343 MicroNiche, we used the principal component 1 of all available resource variables (e.g.,
344 DIC, Chlorophyll a, NH₄⁺, NO_x, PO₄³⁻, SiO₄⁴⁻) as the resource parameter in MicroNiche.

345

346 *Diversity analysis and statistics*

347 All diversity and statistical analysis were performed using the R software (version 4.0.5)
348 and its vegan, stats and ggplot2 packages for figures.

349

350 **Results**

351 *Change in overall microbial community composition in response to disturbance events*

352 Shifts in microbial community diversity did not show a consistent pattern between
353 disturbance events. The Chao1 index based on either extracted 16S rRNA gene (or 16S)
354 fragments recovered in metagenomes or 16S amplicon sequences neither increased nor
355 decreased systematically by the winter disturbance events (win11_warmer_1,
356 win12_windier_3) (Figure 1 (a)). Summer disturbance events (sum11_cyclone_2,
357 sum13_cyclone_9), which represented cyclone events in year 2011 and 2013, either
358 increased or decreased diversity despite representing similar weather changes (Figure
359 1). Nonpareil diversity, a reference database-free metric combining richness and
360 evenness based on shotgun metagenomic data, also showed similar patterns to the
361 Chao1 results reported above (Figure 1(b)). A consistent pattern was observed that, no
362 matter whether disturbances increased or decreased diversity, diversity recovered after
363 each disturbance, indicating that the sampled microbial communities were resilient. The
364 resilience was also supported by community structure analysis based on NMDS of Mash
365 distances of whole metagenomes, or amplicon and extracted metagenomic reads
366 carrying 16S fragments (Figure 1(c) and Figure S4). Functional diversity analysis showed
367 that molecular function diversity showed limited change with disturbance despite

368 substantial diversity and composition changes (Figure S6), indicating high functional
369 redundancy among the sampled microbial communities.

370

371 *Identifying abundant vs rare microbial populations*

372 A total of 394 medium or high-quality MAGs (i.e., quality score >0.5) were obtained after
373 quality control and before de-replication at 95% ANI. For each sample, contigs of different
374 MAG fell into separate clusters based on dimension reduction of contig information (i.e.,
375 contig coverage and kmer profile) using the mmgenome2 software (Figure S14),
376 confirming that most MAGs were likely not chimeric and represent species clusters.
377 Recruitment plot showed that each MAG represented a sequence discrete population
378 within the sample that it was recovered from, e.g., reads mapping between 85%-95%
379 nucleotide identity were rare compared with reads showing >95% identity to the MAG
380 (Figure S15, left and right bottom panels). These results were consistent with previous
381 findings suggesting recognizable prokaryotic species may exist for natural communities
382 [46]. There were 198 non-redundant MAGs after dereplication. Among these, three
383 archaea MAGs were found, belonging to the newly proposed family *Candidatus*
384 *Poseidonaceae* (formerly subgroup MGIIa) and one MAG belonging to *Ca.*
385 *Thalassarchaeaceae* (formerly subgroup MGIIb). Among the bacterial populations,
386 although order *Pelagibacterales* was very abundant at the read level (9.6% to 22.5% of
387 the total metagenome/community), only one *Pelagibacterales* MAG was recovered with
388 medium quality (quality score 0.57). All other *Pelagibacterales* MAGs were low in
389 completeness and were removed at the MAG quality control check step. However, at the

390 contig level, 4.4% to 5.2% contigs were classified as *Ca. Pelagibacter*, consistent with
391 high relative abundance of this group at the read level. Inability of current genome binning
392 algorithms to handle high intra-population diversity and recover representative genomes
393 has been recently noted for this microbial group [66]. Most other MAGs were assigned
394 to the classes of *Flavobacteriia* (phylum *Bacteroidetes*; recently renamed to *Bacteroidota*
395 (21), *Alphaproteobacteria* and *Gammaproteobacteria* (phylum
396 *Proteobacteria/Pseudomonadota* (33 and 17) with a few MAGs assigned to classes of
397 other phyla (Table S2). These dereplicated MAGs represented collectively 5.1% to 18.2%
398 of the total metagenomic reads for different samples, consistent with the high overall
399 diversity of our samples that renders assembly and binning processes limiting.

400

401 Next, we pooled the recovered 198 non-redundant MAGs and competitively mapped
402 reads from each of the 19 samples to calculate MAG relative abundance in order to
403 subsequently draw the (genomo-)species abundance distribution (i.e., the abundance of
404 each MAG is plotted against its rank among all MAGs according to their relative
405 abundance; Figure 2). We used the resulting species-abundance curve to examine if
406 there is any natural discontinuity or inflection point that could be used to define abundant
407 vs. rare taxa more reliably than previous arbitrary threshold in abundance (discussed
408 above). The species abundance curve was modeled robustly by a log scale model (Figure
409 2, Figure S8 (a) and (b), $R^2=0.732\sim0.902$, $P < 0.01$), and we observed no area of
410 discontinuity in the data. A sharp decrease in terms of curve slope was observed around
411 0.1% relative abundance, however (Figure S7). We also observed a corresponding sharp

412 decrease in MAG breadth coverage (how much of the genome is covered by reads) for
413 relative abundance around 0.1% (Figure 2 and S8 (a) and (b)). Specifically, breadth
414 coverage was 90% or less at this level of relative abundance and decreased quickly for
415 less abundant genomospecies. Based on these results, we defined MAGs as rare when
416 they showed relative abundance less than 0.1% and a coverage breadth less than 90%
417 in that sample; we adjusted this threshold on a per-sample basis, depending on the
418 sample's species-abundance curve (i.e., sharp decrease in abundance and breadth
419 coverage). In contrast, abundant MAGs were defined as those showing >0.1% relative
420 abundance and coverage breadth more than 90%. Note that robust detection of a MAG
421 and estimation of its relative abundance is achieved with breadth 10% or higher [67].
422 Hence, MAGs are reliably detectable at 0.1% relative abundance or somewhat lower, and
423 the 0.1% abundance threshold is not a mere artifact of sequencing effort applied (e.g.,
424 inability to detect a MAG). Our definition seems to largely agree with the literature on
425 defining rare biosphere as well [68].

426

427 *Testing the insurance hypothesis: rare MAGs often contribute to community response.*

428 We first tested for the predicted signature of the insurance hypothesis in each of the six
429 major disturbance events sampled by our time series metagenomic datasets. Since our
430 metagenomic sampling is part of a much longer (3-year) 16S amplicon-based time series
431 dataset, we selected samples for metagenome sequencing to represent the pre-
432 disturbance according to the 16S diversity and composition relative to additional samples
433 available in the time series for the same season that the disturbance event took place.

434 That is, these samples did not have a significant difference compared to the previous
435 week, and thus served as useful controls for assessing the effects of the disturbance [69].
436 We first checked the overall MAG relative abundance changes across all disturbance
437 events and found that for each event, many MAGs showed clear changes when
438 comparing disturbed samples with pre-disturbed samples (Figure S9). We found that for
439 all 6 disturbance events, there were between 2 to 22 microbial populations represented
440 by MAGs (Figure 3) that switch from being rare to being abundant during the event based
441 on the criteria mentioned above to define rare MAGs. Further, these MAGs represented
442 between 0.5% to 6.9% of the total community based on the number of reads mapping on
443 the MAGs (relative to the total reads of the sample), depending on the event considered
444 (Figure S10). There were also 8 to 29 MAGs that switched from being abundant to rare,
445 representing 0.4% to 1.8% of the total in the corresponding (disturbed) sample. For each
446 event examined, there were always more MAGs assigned to the abundant-to-rare
447 category than to the rare-to-abundant category, except for event 8 (spr13_warmer_8)
448 (Figure 3). Further, the fraction of reads mapping to rare-to-abundant MAGs was higher
449 than that of abundant-to-rare MAGs in the disturbed samples in three events out of total
450 six (Figure S10). Note that about 5.1% to 18.2% of the total reads in a sample were
451 mapped to all available MAGs (not only MAGs in the 3 categories above but also MAGs
452 assigned to rare-to-rare category), which represents a substantial fraction of the microbial
453 communities sampled; much higher fraction than previous, isolate- or lab-based
454 approaches to study similar questions were able to assess.

455 For event #5 representing an extreme warm week during the spring time
456 (spr12_warming_5), there were 21 rare MAGs that became abundant, while 29 abundant
457 MAGs switched to rare. Sixteen abundant MAGs remained abundant during this event,
458 representing 4.5% of the total community. The total relative abundance of MAGs that
459 switched from rare to abundant was around 7% while the abundant MAGs that switched
460 to rare made up about 1% of the community in the disturbed sample (but was 5.3% before
461 the disturbance; Figure S10). Further analysis showed that the rare MAGs that became
462 abundant during this warming event (spr12_warming_5) encode more metabolic
463 pathways related to carbohydrates degradation and fewer pathways related
464 photosynthesis compared to abundant-to-rare MAGs (Figure S16). These results
465 suggested that there were more carbon compounds in the surface seawater during this
466 warming event compared to pre-disturbance sample, and microbial populations that were
467 more efficient in utilizing those carbon gained a competitive advantage over those that
468 lacked these pathways (e.g., abundant to rare MAGs). Amplicon based analysis also
469 showed consistent results: a “spring bloom” indicating overturn in the phytoplankton was
470 obvious based on the amplicon datasets [22]. However, two weeks after the event, 14 of
471 the MAGs that had become abundant from rare during the event became rare again
472 (77.8% of the category) while 11 MAGs that became rare from abundant became
473 abundant again (66.7% of the category), revealing that the sampled microbial
474 communities were resilient. These results reveal that the rare populations provided
475 insurance for ecosystem functioning when undergoing a short-term, strong disturbance
476 event. That said, the importance of abundant populations that remained abundant cannot

477 be underemphasized because the latter MAGs represented a higher fraction of the total
478 community than rare-to-abundant MAGs in all events, including the disturbance event 5
479 (spr12_warming_5) mentioned above (Figure S10).

480 There are only three disturbance events with available post-disturbance samples:
481 the other two being win11_storm_2 and win12_warming_3. For win12_warming_3, we
482 observed a relatively small number of MAGs changing category, i.e., becoming rare-to-
483 abundant or abundant-to-rare (25% and 33.3% of the total MAGs in each category,
484 respectively) compared to the warming event mentioned above. For win11_storm_2,
485 44.4% of MAGs became rare-to-abundant and 25% of MAGs became abundant-to-rare,
486 and all these MAGs recovered to their pre-disturbance level (category) after the event.
487 For the remaining disturbance events, we observed 50% (out of 4 MAGs in total) and
488 42.8% (out of 7 MAGs in total) of the rare-to-abundant and abundant-to-rare categories
489 to recover to pre-disturbance levels for the win11_warming_1 event, respectively, while
490 50% (out of 2 MAGs) and 25% (out of 8 MAGs in total) of the same categories recovered
491 post-disturbance for the sum13_storm_9. Overall, except for event win12_warming_3,
492 about half or more of the taxa that represented rare-to-abundant populations recovered
493 after the event (that is, returned to rarity), further indicating that while rare populations
494 play a role in community resilience by responding to niches created by the disturbance,
495 most of these taxa return to being rare post-disturbance. However, except for
496 spr12_warming_5, less than a half of the abundant-to-rare populations recovered
497 spr12_warming_5 to become abundant again, indicating that rare and abundant
498 populations may respond differently to disturbances.

499

500 *Testing the disturbance-specialization hypothesis: generalists are favored by disturbance*

501 We additionally examined the types of organisms that responded to disturbance in term
502 of generalist vs. specialist lifestyle. For the latter, we used the proportion similarity index
503 that measures population environmental preference by examining abundance changes
504 across categories of available samples (PS index) (Table S3). We categorized the upper
505 and bottom third of MAGs based on ranked PS index (see Materials and Methods for
506 details) as generalists and specialists, respectively. Notably, our ranked based definition
507 of generalists is generally consistent with the Levin's Breadth Index and PS index
508 implementation in MicroNiche (Figure S19), but it is more appropriate for the data
509 available here as explained in the Methods section. For rare populations that became
510 abundant after the disturbance, a larger fraction of them were generalists (60% to 78%)
511 than specialists (16% to 24%) except for the first winter disturbance, which was in fact a
512 resilience event (community recovered to pre-disturbance) (Figure 4). On the other hand,
513 for abundant populations that became rare, most of them were specialists, except for
514 sum11_cyclone_2 and sum13_cyclone_9, which were summer cyclone disturbance
515 events (Figure S11). Functional gene content analysis revealed that generalists had
516 smaller genome sizes and more compacted genomes than specialists (Figure S12), and
517 a slightly higher fraction of extracellular function proteins encoded in their genomes
518 (Figure S13). These results support the disturbance-specialization hypothesis to explain,
519 at least partially, responses to nearly all disturbance events.

520 For winter (warming) disturbances (win11_warming_1 and win12_warming_3), the
521 populations that remain abundant were identified mostly as specialists and only a small
522 fraction of them were generalists. In contrast, for summer disturbances (sum11_storm_2
523 and sum13_storm_9), most MAGs that remained abundant were generalists (6 and 9
524 respectively), suggesting that these generalists could survive strong water body mixing
525 disturbances caused by cyclones. For spr12_warming_5, the minority of abundant MAGs
526 that remained abundant before and during the event were identified as generalists (4 out
527 of 11), contrasting with spr13_warming_8, in which the majority of abundant MAGs that
528 remained abundant before and during the event were generalists (17 out of 21). These
529 patterns suggested that abundant populations of the community, which could be either
530 generalists or specialists, are generally resistant to the effects of the disturbance events
531 studied, consistent with the results reported above at the whole-community and MAG
532 levels.

533

534 **Discussion**

535 Quantifying the importance of rare populations for microbial community response
536 to disturbances represents a cornerstone question in microbial ecology and is key to
537 testing ecological theory as well as modeling the effects of global change on ecosystem
538 function and diversity. We observed that a large fraction of the detected MAGs remained
539 abundant before, during and after disturbance (resistant) for all the disturbance events
540 studied (Figure 3, Figure S10), revealing that these genomospecies are insensitive to
541 short term disturbances and provide stability for the ecosystem. We also found that rare

542 prokaryotic populations became abundant after the disturbance event and made up
543 between 0.3% and 7% of the total community, depending on the event considered.
544 However, more than half of these populations returned to rarity one or two weeks after
545 each disturbance, consistent with the concept of conditionally rare taxa (CRT) defined by
546 Shade and colleagues [7] and also the results of amplicon based analysis of same
547 samples [22]. Such conditionally rare taxa may therefore provide resilience to the system
548 undergoing disturbances and our approach here quantified this contribution in terms of
549 relative abundance of the total community. As community assembly theory also predicts,
550 CRT that are deterministically assembled (that is, selected by disturbance conditions)
551 may contribute to the ecosystem processes/functions as much as the abundant taxa do
552 [3]. In agreement with these interpretations, we also found that rare-to-abundant
553 population genomes harbored more metabolic pathways related to carbohydrate
554 degradation compared to abundant-to-rare populations, the latter were clearly disfavored
555 by the spr12_warming_5 (spring turnover blooms) disturbance events. It should be
556 mentioned, however, that this pattern was clear only in one of the two warming events
557 (spr12_warming_5; the other being spr13_warming_8) and despite the shared
558 characteristics between the two events such as the increase of chlorophyll A
559 concentration, an indicator of primary production.

560

561 The lack of universal patterns among the disturbance events is likely due, at least
562 in part, to the fact that the disturbance events were all largely distinct from one another,
563 even for similar winter “warming” disturbances, and each event favored different MAGs

564 and functional traits (so high inter-event diversity). For example, for the two summer
565 hurricane disturbance events, we did not observe consistent rare-to-abundant population
566 dynamic patterns (e.g., the same genomospecies did not show similar changes), partially
567 because these cyclone events caused the strong mixing of both costal sediment and
568 seawater. We also saw different patterns in the warming winter (win11_warm_1) and
569 rainy winter (win12_rainy_3) disturbances, which, nonetheless, was not surprising
570 because these two disturbance events were characterized by different nutrient changes
571 such as ammonium concentrations (Table S1, pico131 and pico256 for 2 events
572 respectively). Those inconsistent responses, in general, are somewhat expected
573 because capturing natural disturbance events and associated metadata is challenging
574 from the perspective of sampling. Although replicating disturbances is not possible (e.g.,
575 all storms are not identical), (biological) replicate samples are possible, and we will try to
576 obtain such replicate filters in future studies. Also, a lower number and quality of MAGs
577 were recovered for the datasets representing the second hurricane event due to relatively
578 lower sequencing coverage (Figure S2, sample pico551 and pico552 were 1/3 of other
579 samples in terms of size), which somewhat limited our resolution into the population
580 dynamic patterns. Deeper sequencing efforts could further facilitate the study of patterns
581 of rare population dynamics.

582

583 How to define rare populations or species has been a challenging task, and
584 arbitrary thresholds based on relative abundance of 16S rRNA gene sequences (e.g.,
585 0.1%) have been commonly used [15]. Here we defined rare populations or species

586 based on the species-abundance curve that was derived from the sequencing depth and
587 breadth coverage of recovered MAGs (normalized by total genome equivalents), while
588 considering the limit of detection of our metagenomic sequencing effort. Specifically, the
589 TAD80 metric used for calculating abundance was larger than the genome breadth
590 coverage that corresponded to the limit of detection, which was defined as -
591 $\ln(0.9)/\text{genome equivalents}$ [70]. Therefore, the commonly derived 0.1% threshold for rare
592 taxa by our analysis, and our definition of rare MAGs, were robust. Although our derived
593 threshold often matched the previous (arbitrary) thresholds, we anticipate that the
594 threshold may differ for other samples and/or environments. Hence, the approach
595 outlined here based on species abundance curve should be useful for future studies.
596 Further, the TAD80 (coverage depth), a metric used to define population abundance, can
597 remove highly conserved regions and regions recently subjected to horizontal gene
598 transfer. If those regions are not removed, the sequencing depth, and thus abundance
599 can be overestimated. TAD80 can also remove regions with high gene-content micro-
600 diversity, which tend to underestimate depth. Thus, TAD80 provides reliable estimations
601 of relative abundance and thus, species abundance curve [57,71]. It should be
602 mentioned, however, that for the species abundance curves we did not observe a clear
603 inflection point that could be used to define rare species in a more natural and objective
604 manner compared to the use of a predetermined threshold (e.g., 0.1% abundance).
605 Instead, the curve often appeared to be a monotonic, log-normal decrease with no
606 obvious inflection points but with a sharp change in slope. Hence, the threshold to define
607 rare taxa may appear somewhat arbitrary even with a species abundance curve available,

608 although we do recommend the TAD-80-based methodology outlined above as a more
609 robust and well-defined approach that takes into account (normalizes for) different
610 sequencing efforts between samples (see also below).

611

612 Rare taxa have been defined based on several different approaches. For instance,
613 Debroas et al. defined rare biosphere as unassembled reads with low sequence depth
614 coverage, without obtaining population genomes, and bioinformatically annotated those
615 unassembled reads by mapping them to pre-annotated databases in order to infer
616 functions carried by the rare taxa [72]. However, this approach cannot link these functions
617 with specific taxa and derive species abundance dynamics since these unassembled
618 short reads are typically not linked to each other and/or a phylogenetic marker. Our
619 approach to define rare population based on MAG abundances can link taxa with their
620 functions, providing an important advantage over previous literature, albeit the number of
621 taxa (or MAGs) that we were able to study was relatively small (e.g., the MAG has to be
622 abundant enough in at least one sample of the series to be recoverable by sequencing).
623 Further, binning rare population MAGs is subjected to artifacts (more so than for abundant
624 MAGs) because such populations show low sequencing coverage, and thus are typically
625 represented by short contigs that are not easy to bin (Figure S14, white and orange
626 datapoints) [73,74]. Long read metagenome sequencing could improve assembly and
627 binning even of low coverage MAGs, especially in combination with new binning
628 algorithms, such as those employing deep variational auto-encoder [75] and graph-based
629 embedding method [76] or both [77].

630

631 To further understand the key functional and/or ecological differences between
632 taxa that responded to disturbances by changes in their abundance relative to those that
633 did not, we tested the disturbance-specialization hypothesis that generalists are favored
634 while specialists are disfavored by disturbance. Our results provided some support for
635 this hypothesis based on the PS index, which is a direct way to measure taxon relative
636 abundance changes across the entire dataset to define generalists vs. specialists. Our
637 conclusion that more generalists taxa than specialists are favored by most disturbance
638 events studied here is consistent with those of recent studies based on laboratory
639 mesocosms [78]. Specifically, Chen and colleagues concluded that generalist taxa are
640 more metabolically flexible. Consistently, we found that generalists have -on average- a
641 smaller and more compacted genomes, which probably provides an (more) efficient
642 metabolic strategy and a selective advantage during changing conditions. The smaller
643 genome size is also consistent with the Black Queen Hypothesis that has been used to
644 explain the ecological success of small genomes [79]. Further, we found that generalists
645 have a slightly higher fraction of their genome devoted to extracellular protein functions
646 compared to specialists, which could be another reason why generalists were favored by
647 the disturbance events. That is, extracellular proteins represent mostly enzymes that
648 degrade or transport public goods that presumably remain abundant throughout the
649 disturbance.

650

651 It should be noted, however, that there is no standard or universally accepted
652 definition for generalist/specialist except probably for Levin's niche breadth index [62,80]
653 and Proportional Similarity Index [63]. We preferably employed the PS index because it
654 allows abundant and widespread species to be also classified as habitat specialists when
655 they show relatively high abundance in a single category of samples (so prevalence in
656 more samples matters in addition to their abundance or absolute abundance) [57,64]. We
657 obtained largely consistent results to those reported above with the PS index when we
658 used the PS implementation in MicroNiche package, as well as with Levin's niche breadth
659 definition (Figure S19)[57]. Further, the PS index, based on absolute abundance data, is
660 commonly used in macroecology compared to Levin's definition [64]. We used relative
661 abundance of microbial taxa as compared to the absolute abundance typically used in
662 macroecology since absolute abundances were not available for our datasets. Performing
663 this type of analysis with absolute abundances and a more detailed measurement of
664 nutrient availability in each sample in future studies, could further corroborate the
665 conclusions presented here.

666

667 It should also be noted, however, that the definition of rare populations used could
668 affect our findings and conclusions. For example, the sequencing effort applied
669 presumably affects the limit of detection established, and thus our definition of rare taxa.
670 Under-sampling (in term of sequencing coverage) of microbial communities, especially in
671 highly complex environmental settings, leads to poor assembly and MAG recovery.
672 Further, this limitation could also affect the species abundance curve and thus, the

673 threshold for rare taxa derived from the curve. For example, for non-subsampled reads,
674 we obtained a total of 412 dereplicated MAGs compared to 198 when we subsampled the
675 datasets to provide similar sequencing effort across datasets, further indicating that more
676 MAGs can be obtained with a higher sequencing effort. More available MAGs could
677 somewhat affect the species abundance distribution curve, and thus our empirical
678 definition of rare population changes (e.g., threshold to be mostly around 0.05% instead
679 of the 0.1% used) (Figure S17). Sequencing coverage could also affect the PS index
680 because with more sequencing effort some undetected genomes in under-sampled
681 samples could become detectable. Nonetheless, we expect that this sequence coverage
682 limitation applies evenly to all samples and rare taxa, and thus should not significantly
683 affect our PS index values for most taxa and our derived conclusions. It would be valuable
684 to see if our conclusions would change with higher sequencing coverage and absolute
685 abundance measurements of each species in future studies of our sampling and other
686 sites.

687

688 Our work also showed that natural microbial community are remarkably resilient at
689 the whole community level. Both diversity (Figure 1c and Figure S4 a and b, MAG
690 recovery results section) and community composition (Figure S5) recovered after
691 disturbance, consistent with previous findings in a lake ecosystem and elsewhere [81].
692 Metagenomic-based functional diversity analysis showed that functional redundancy
693 could play a key role in this resilience because we observed a decoupling between
694 taxonomic diversity changes and functional diversity changes (Figure S6), which

695 presumably reflects functional redundancy among the taxa that changed in abundance
696 [82]. The recovery of the rare-to-abundant populations also supported the strong
697 resilience of the sampled communities (that is, going back to being rare after disturbance)
698 even though these populations typically accounted for a smaller fraction of the community
699 typically, i.e., 1-7% of total. The fact that we did not observe high similarity in the microbial
700 community responses to similar disturbance events in the same season underscores the
701 great diversity of microbial communities as well as our inability to measure the key
702 parameters changes by each event e.g., diversity of organic compounds released or
703 exact physicochemical properties that may slightly differ among similar disturbance
704 events. Year to year variation, e.g., stochastic birth/death of microbial populations, could
705 also be another reason we did not see consistent responses among similar events [83].
706 Multiple samples that represent the prevailing, not-disturbance-associated, microbial
707 communities in each season would be needed to quantify the effect of stochastic
708 processes on the results reported here. For instance, it is possible that some of the MAGs
709 that are reported above to change from the abundant to the rare categories during the
710 event (or vice versa) might represent such stochastic processes -as opposed to
711 deterministic processes caused by the event. We believe that the effect of such stochastic
712 processes on our results is limited because in a couple samples that the post-disturbance
713 sample resembled closely the pre-disturbance sample in terms of 16S-based microbial
714 community composition, and thus allow us to assess stochastic processes, such as for
715 win12_rainy_3 (see Fig. S4), the MAG abundances were much more similar between the
716 pre- and post- disturbance samples relative to the disturbance sample (Fig. S18).

717 Therefore, most of the MAGs reported to change between the abundant and rare
718 categories are presumably due to the effects of the events rather than stochastic
719 processes.

720

721 In conclusion, we found microbial communities in the coastal ocean of Southeast
722 USA to be both resistant and resilience (depending on whether the corresponding
723 populations were abundant and stayed abundant or became rare during the event and
724 bounced back to be abundant after the event, respectively) against natural disturbances,
725 and provided evidence in support of the disturbance-specialization and the insurance
726 hypotheses. Further, we provided a new approach based on the species abundance
727 curve to define rare vs abundant populations as well as generalists versus specialists
728 based on a time-series metagenomic sampling. These definitions and approaches could
729 be helpful for future ecological studies aiming at answering questions related to the role
730 and importance of the rare biosphere. Collectively, the findings presented here advance
731 our understanding of natural microbial community response to disturbance, and thus,
732 could be useful for ecosystem management, e.g., microbiome rescue [8] in a changing
733 world.

734

735 **Data Availability**

736 Raw metagenomic reads, amplicon reads can be found at NCBI, under accession number
737 PRJNA803723.

738

739 **Author Contribution and Funding**

740 J.Z, L.M.R and K.T.K designed the work. Z.W and D.H did the sampling and
741 environmental analysis. J.Z and G.B did the bioinformatic analysis. J.Z and K.T.K wrote
742 the manuscript. This work was supported, in part, by the US National Science Foundation
743 to D.E.H. (OCE 1416665) and to K.T.K. (OCE 1416673 and DEB 1831582).

744

745 **Acknowledgment**

746 We want to thank PACE at Georgia Tech for providing computational resources.

747

748

749

750

751

752 **References cited**

- 753 1. Pedrós-Alió, C. (2012) The Rare Bacterial Biosphere. *Annual Review of Marine Science*
754 4, 449-466. [10.1146/annurev-marine-120710-100948](https://doi.org/10.1146/annurev-marine-120710-100948)
- 755 2. Reid, A. and Buckley, M. (2011) The rare biosphere.
- 756 3. Jia, X. *et al.* (2018) Community Assembly Processes of the Microbial Rare Biosphere.
757 *Trends in Microbiology* 26, 738-747. <https://doi.org/10.1016/j.tim.2018.02.011>
- 758 4. Jousset, A. *et al.* (2017) Where less may be more: how the rare biosphere pulls
759 ecosystems strings. *The ISME Journal* 11, 853-862. [10.1038/ismej.2016.174](https://doi.org/10.1038/ismej.2016.174)
- 760 5. Karthikeyan, S. *et al.* (2019) “Candidatus Macondimonas diazotrophica”, a novel
761 gammaproteobacterial genus dominating crude-oil-contaminated coastal sediments. *The*
762 *ISME Journal* 13, 2129-2134. [10.1038/s41396-019-0400-5](https://doi.org/10.1038/s41396-019-0400-5)
- 763 6. Sjöstedt, J. *et al.* (2012) Recruitment of Members from the Rare Biosphere of Marine
764 Bacterioplankton Communities after an Environmental Disturbance. *Applied and*
765 *Environmental Microbiology* 78, 1361-1369. doi:10.1128/AEM.05542-11
- 766 7. Shade, A. *et al.* (2014) Conditionally Rare Taxa Disproportionately Contribute to Temporal
767 Changes in Microbial Diversity. *mBio* 5, e01371-01314. doi:10.1128/mBio.01371-14
- 768 8. Shade, A. (2023) Microbiome rescue: directing resilience of environmental microbial
769 communities. *Current Opinion in Microbiology* 72, 102263.
770 <https://doi.org/10.1016/j.mib.2022.102263>

- 771 9. Kurm, V. *et al.* (2019) A low proportion of rare bacterial taxa responds to abiotic changes
772 compared with dominant taxa. *Environmental Microbiology* 21, 750-758.
773 <https://doi.org/10.1111/1462-2920.14492>
- 774 10. Yachi, S. and Loreau, M. (1999) Biodiversity and ecosystem productivity in a fluctuating
775 environment: The insurance hypothesis. *Proceedings of the National Academy of*
776 *Sciences* 96, 1463-1468
- 777 11. Shade, A. *et al.* (2012) Fundamentals of Microbial Community Resistance and Resilience.
778 *Frontiers in Microbiology* 3. 10.3389/fmicb.2012.00417
- 779 12. Vázquez, D.P. and Simberloff, D. (2002) Ecological specialization and susceptibility to
780 disturbance: conjectures and refutations. *The American Naturalist* 159, 606-623
- 781 13. Chen, Y.-J. *et al.* (2021) Metabolic flexibility allows bacterial habitat generalists to become
782 dominant in a frequently disturbed ecosystem. *The ISME Journal*. 10.1038/s41396-021-
783 00988-w
- 784 14. Richmond, C.E. *et al.* (2005) The role of environmental generalist species in ecosystem
785 function. *Ecological modelling* 188, 279-295
- 786 15. Alonso-Sáez, L. *et al.* (2015) The hidden seasonality of the rare biosphere in coastal
787 marine bacterioplankton. *Environmental Microbiology* 17, 3766-3780.
788 <https://doi.org/10.1111/1462-2920.12801>
- 789 16. Galand, P.E. *et al.* (2009) Ecology of the rare microbial biosphere of the Arctic Ocean.
790 *Proceedings of the National Academy of Sciences* 106, 22427-22432
- 791 17. Ji, M. *et al.* (2020) Distinct assembly mechanisms underlie similar biogeographical
792 patterns of rare and abundant bacteria in Tibetan Plateau grassland soils. *Environmental*
793 *Microbiology* 22, 2261-2272. <https://doi.org/10.1111/1462-2920.14993>
- 794 18. Venette, R.C. *et al.* (2002) Strategies and Statistics of Sampling for Rare Individuals.
795 *Annual Review of Entomology* 47, 143-174. 10.1146/annurev.ento.47.091201.145147
- 796 19. Wang, Y. *et al.* (2017) Quantifying the Importance of the Rare Biosphere for Microbial
797 Community Response to Organic Pollutants in a Freshwater Ecosystem. *Applied and*
798 *Environmental Microbiology* 83,
- 799 20. Hua, Z.-S. *et al.* (2015) Ecological roles of dominant and rare prokaryotes in acid mine
800 drainage revealed by metagenomics and metatranscriptomics. *The ISME Journal* 9, 1280-
801 1294. 10.1038/ismej.2014.212
- 802 21. Karthikeyan, S. *et al.* (2020) Genome repository of oil systems: An interactive and
803 searchable database that expands the catalogued diversity of crude oil-associated
804 microbes. *Environmental Microbiology* 22, 2094-2106. [https://doi.org/10.1111/1462-
805 2920.14966](https://doi.org/10.1111/1462-2920.14966)
- 806 22. Gröniger, J.L. *et al.* (2022) Rapid changes in coastal ocean microbiomes uncoupled with
807 shifts in environmental variables. *Environmental Microbiology*,
- 808 23. Ward, C.S. *et al.* (2017) Annual community patterns are driven by seasonal switching
809 between closely related marine bacteria. *The ISME Journal* 11, 1412-1422.
810 10.1038/ismej.2017.4
- 811 24. Yung, C.-M. *et al.* (2016) Insensitivity of diverse and temporally variable particle-
812 associated microbial communities to bulk seawater environmental parameters. *Applied*
813 *and environmental microbiology* 82, 3431-3437
- 814 25. Edgar, R.C. (2016) UNOISE2: improved error-correction for Illumina 16S and ITS
815 amplicon sequencing. *BioRxiv*, 081257
- 816 26. Peng, X. and Dorman, K.S. (2020) AmpliCl: a high-resolution model-based approach for
817 denoising Illumina amplicon data. *Bioinformatics* 36, 5151-5158
- 818 27. Edgar, R.C. (2016) SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS
819 sequences. *bioRxiv*, 074161. 10.1101/074161

- 820 28. Pruesse, E. *et al.* (2007) SILVA: a comprehensive online resource for quality checked and
821 aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids research* 35,
822 7188-7196
- 823 29. Caporaso, J.G. *et al.* (2010) QIIME allows analysis of high-throughput community
824 sequencing data. *Nature methods* 7, 335-336
- 825 30. Bolger, A.M. *et al.* (2014) Trimmomatic: a flexible trimmer for Illumina sequence data.
826 *Bioinformatics* 30, 2114-2120. 10.1093/bioinformatics/btu170
- 827 31. Rodriguez-R, L.M. *et al.* (2018) Nonpareil 3: Fast Estimation of Metagenomic Coverage
828 and Sequence Diversity. *mSystems* 3, e00039-00018. doi:10.1128/mSystems.00039-18
- 829 32. Bengtsson-Palme, J. *et al.* (2015) METAXA2: improved identification and taxonomic
830 classification of small and large subunit rRNA in metagenomic data. *Molecular ecology*
831 *resources* 15, 1403-1414
- 832 33. Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST.
833 *Bioinformatics* 26, 2460-2461
- 834 34. Li, D. *et al.* (2015) MEGAHIT: an ultra-fast single-node solution for large and complex
835 metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674-1676.
836 10.1093/bioinformatics/btv033
- 837 35. Kim, D. *et al.* (2016) Centrifuge: rapid and sensitive classification of metagenomic
838 sequences. *Genome research* 26, 1721-1729
- 839 36. Hyatt, D. *et al.* (2012) Gene and translation initiation site prediction in metagenomic
840 sequences. *Bioinformatics* 28, 2223-2230. 10.1093/bioinformatics/bts429
- 841 37. Vasimuddin, M. *et al.* (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for
842 Multicore Systems. 2019 IEEE International Parallel and Distributed Processing
843 Symposium (IPDPS).
- 844 38. Buchfink, B. *et al.* (2015) Fast and sensitive protein alignment using DIAMOND. *Nature*
845 *methods* 12, 59-60
- 846 39. Heritier-Robbins, P. *et al.* (2021) Beach sand oil spills select for generalist microbial
847 populations. *The ISME Journal*. 10.1038/s41396-021-01017-6
- 848 40. Yu, N.Y. *et al.* (2010) PSORTb 3.0: improved protein subcellular localization prediction
849 with refined localization subcategories and predictive capabilities for all prokaryotes.
850 *Bioinformatics* 26, 1608-1615. 10.1093/bioinformatics/btq249
- 851 41. Alneberg, J. *et al.* (2014) Binning metagenomic contigs by coverage and composition.
852 *Nature Methods* 11, 1144-1146. 10.1038/nmeth.3103
- 853 42. Kang, D.D. *et al.* (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient
854 genome reconstruction from metagenome assemblies. *PeerJ* 7, e7359
- 855 43. Wu, Y.-W. *et al.* (2015) MaxBin 2.0: an automated binning algorithm to recover genomes
856 from multiple metagenomic datasets. *Bioinformatics* 32, 605-607.
857 10.1093/bioinformatics/btv638
- 858 44. Sieber, C.M. *et al.* (2018) Recovery of genomes from metagenomes via a dereplication,
859 aggregation and scoring strategy. *Nature microbiology* 3, 836-843
- 860 45. Parks, D.H. *et al.* (2015) CheckM: assessing the quality of microbial genomes recovered
861 from isolates, single cells, and metagenomes. *Genome research* 25, 1043-1055
- 862 46. Rodriguez-R, L.M. *et al.* (2021) Reply to: "Re-evaluating the evidence for a universal
863 genetic boundary among microbial species". *Nature Communications* 12, 4060.
864 10.1038/s41467-021-24129-1
- 865 47. Rodriguez-R, L.M. and Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for
866 specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* 4,
867 e1900v1901. 10.7287/peerj.preprints.1900v1

- 868 48. Luo, C. *et al.* (2014) MyTaxa: an advanced taxonomic classifier for genomic and
869 metagenomic sequences. *Nucleic Acids Research* 42, e73-e73. 10.1093/nar/gku169
- 870 49. Rodriguez-R, L.M. *et al.* (2018) The Microbial Genomes Atlas (MiGA) webserver:
871 taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level.
872 *Nucleic Acids Research* 46, W282-W288. 10.1093/nar/gky467
- 873 50. Olm, M.R. *et al.* (2017) dRep: a tool for fast and accurate genomic comparisons that
874 enables improved genome recovery from metagenomes through de-replication. *The ISME*
875 *journal* 11, 2864-2868
- 876 51. Parks, D.H. *et al.* (2018) A standardized bacterial taxonomy based on genome phylogeny
877 substantially revises the tree of life. *Nature Biotechnology* 36, 996-1004. 10.1038/nbt.4229
- 878 52. Chaumeil, P.-A. *et al.* (2019) GTDB-Tk: a toolkit to classify genomes with the Genome
879 Taxonomy Database. *Bioinformatics* 36, 1925-1927. 10.1093/bioinformatics/btz848
- 880 53. Jain, C. *et al.* (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals
881 clear species boundaries. *Nature Communications* 9, 5114. 10.1038/s41467-018-07641-
882 9
- 883 54. Matsen, F.A. *et al.* (2010) pplacer: linear time maximum-likelihood and Bayesian
884 phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11,
885 538. 10.1186/1471-2105-11-538
- 886 55. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2.
887 *Nature Methods* 9, 357-359. 10.1038/nmeth.1923
- 888 56. Li, H. *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25,
889 2078-2079. 10.1093/bioinformatics/btp352
- 890 57. Rodriguez-R, L.M. *et al.* (2020) Iterative subtractive binning of freshwater chronoserics
891 metagenomes identifies over 400 novel species and their ecologic preferences.
892 *Environmental Microbiology* 22, 3394-3412. <https://doi.org/10.1111/1462-2920.15112>
- 893 58. Nayfach, S. and Pollard, K.S. (2015) Average genome size estimation improves
894 comparative metagenomics and sheds light on the functional ecology of the human
895 microbiome. *Genome biology* 16, 1-18
- 896 59. Ruiz-Perez, C.A. *et al.* (2021) MicrobeAnnotator: a user-friendly, comprehensive
897 functional annotation pipeline for microbial genomes. *BMC Bioinformatics* 22, 11.
898 10.1186/s12859-020-03940-5
- 899 60. Shaffer, M. *et al.* (2020) DRAM for distilling microbial metabolism to automate the curation
900 of microbiome function. *Nucleic Acids Research* 48, 8883-8900. 10.1093/nar/gkaa621
- 901 61. Levins, R. (1968) *Evolution in changing environments* Princeton University Press
- 902 62. Luo, Z. *et al.* (2019) Biogeographic Patterns and Assembly Mechanisms of Bacterial
903 Communities Differ Between Habitat Generalists and Specialists Across Elevational
904 Gradients. *Frontiers in Microbiology* 10. 10.3389/fmicb.2019.00169
- 905 63. Feinsinger, P. *et al.* (1981) A Simple Measure of Niche Breadth. *Ecology* 62, 27-32.
906 <https://doi.org/10.2307/1936664>
- 907 64. Verberk, W.C.E.P. *et al.* (2010) Explaining abundance–occupancy relationships in
908 specialists and generalists: a case study on aquatic macroinvertebrates in standing
909 waters. *Journal of Animal Ecology* 79, 589-601. [https://doi.org/10.1111/j.1365-
910 2656.2010.01660.x](https://doi.org/10.1111/j.1365-2656.2010.01660.x)
- 911 65. Finn, D. *et al.* (2020) MicroNiche: an R package for assessing microbial niche breadth and
912 overlap from amplicon sequencing data. *FEMS microbiology ecology* 96, fiae131
- 913 66. Ruiz-Perez, C.A. *et al.* (2021) Description of *Candidatus Mesopelagibacter*
914 *carboxydoxydans* and *Candidatus Anoxipelagibacter denitrificans*: Nitrate-reducing
915 SAR11 genera that dominate mesopelagic and anoxic marine zones. *Systematic and*
916 *Applied Microbiology* 44, 126185

- 917 67. Castro, J.C. *et al.* (2018) imGLAD: accurate detection and quantification of target
918 organisms in metagenomes. *PeerJ* 6, e5882
- 919 68. Sharon, I. *et al.* (2015) Accurate, multi-kb reads resolve complex populations and detect
920 rare microorganisms. *Genome research* 25, 534-543
- 921 69. Wang, Z. *et al.* (2019) Microbial communities across nearshore to offshore coastal
922 transects are primarily shaped by distance and temperature. *Environmental Microbiology*
923 21, 3862-3872. <https://doi.org/10.1111/1462-2920.14734>
- 924 70. Lindner, B.G. *et al.* (2022) Toward shotgun metagenomic approaches for microbial source
925 tracking sewage spills based on laboratory mesocosms. *Water Research* 210, 117993.
926 <https://doi.org/10.1016/j.watres.2021.117993>
- 927 71. Lindner, B.G. *et al.* (2023) A user's guide to the bioinformatic analysis of shotgun
928 metagenomic sequence data for bacterial pathogen detection. *International Journal of*
929 *Food Microbiology*, 110488
- 930 72. Debroas, D. *et al.* (2022) Seasonal microbial dynamics in the ocean inferred from
931 assembled and unassembled data: a view on the unknown biosphere. *ISME*
932 *Communications* 2, 87. 10.1038/s43705-022-00167-8
- 933 73. Peng, Y. *et al.* (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic
934 sequencing data with highly uneven depth. *Bioinformatics* 28, 1420-1428.
935 10.1093/bioinformatics/bts174
- 936 74. Karst, S.M. *et al.* (2016) mmgenome: a toolbox for reproducible genome extraction from
937 metagenomes. *bioRxiv*, 059121. 10.1101/059121
- 938 75. Nissen, J.N. *et al.* (2021) Improved metagenome binning and assembly using deep
939 variational autoencoders. *Nature biotechnology* 39, 555-560
- 940 76. Lamurias, A. *et al.* (2022) Metagenomic binning with assembly graph embeddings.
941 *Bioinformatics* 38, 4481-4487. 10.1093/bioinformatics/btac557
- 942 77. Lamurias, A. *et al.* (2023) Metagenomic Binning using Connectivity-constrained
943 Variational Autoencoders.
- 944 78. Chen, Y.-J. *et al.* (2021) Metabolic flexibility allows bacterial habitat generalists to become
945 dominant in a frequently disturbed ecosystem. *The ISME Journal* 15, 2986-3004.
946 10.1038/s41396-021-00988-w
- 947 79. Morris, J.J. *et al.* (2012) The Black Queen Hypothesis: Evolution of Dependencies through
948 Adaptive Gene Loss. *mBio* 3, e00036-00012. doi:10.1128/mBio.00036-12
- 949 80. Levins, R. (1968) *Evolution in changing environments* (Princeton, New Jersey,
- 950 81. Shade, A. *et al.* (2012) Lake microbial communities are resilient after a whole-ecosystem
951 disturbance. *The ISME Journal* 6, 2153-2167. 10.1038/ismej.2012.56
- 952 82. Louca, S. *et al.* (2018) Function and functional redundancy in microbial systems. *Nature*
953 *Ecology & Evolution* 2, 936-943. 10.1038/s41559-018-0519-1
- 954 83. Sloan, W.T. *et al.* (2021) Drift dynamics in microbial communities and the effective
955 community size. *Environmental Microbiology* 23, 2473-2483.
956 <https://doi.org/10.1111/1462-2920.15453>
- 957

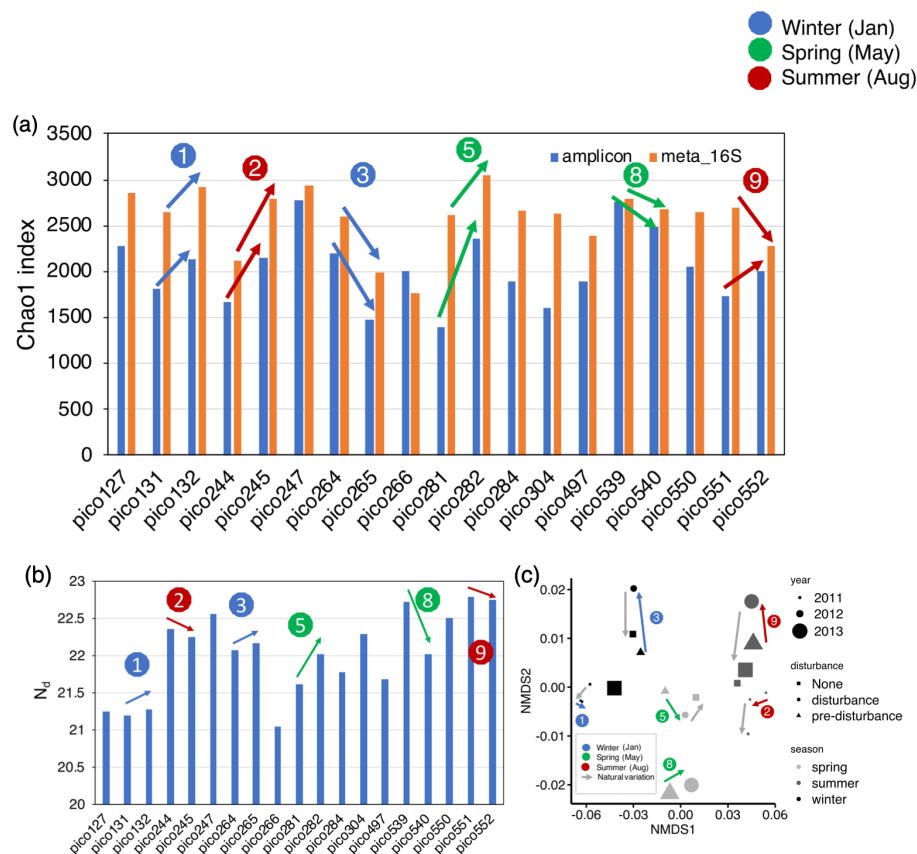


Figure 1. (a) Diversity shifts during disturbance events. Chao1 diversity index (a) for both amplicon 16S rRNA genes (blue) and extracted 16S rRNA gene reads from metagenomes (orange) are shown. Nonpareil diversity (N_d) (b) and Mash distance based NMDS (c) of subsampled metagenomes are also included. Disturbance events are labelled with number (**see Figure S1 for details**) and the color indicates seasonality. Colored arrows in (a), (b) and (c) showed how N_d diversity and metagenomic composition was changed by each disturbance event. Grey arrows in (c) shows the natural variation of metagenomic composition (recovery or natural variation).

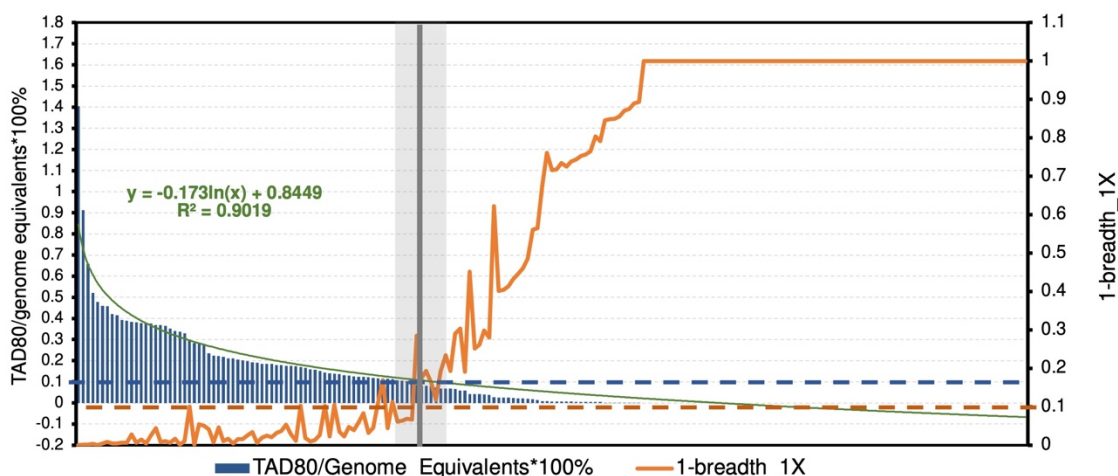


Figure 2. Our approach to define abundance vs. rare MAGs. MAG coverage depth (left y axis, blue bar) and coverage breadth (right y axis, orange line, shown as 1-coverage breadth) distribution for one metagenomic sample (pico127). Thus, X axis is MAG rank by abundance, estimated as coverage depth (i.e., TAD80 values normalized by genome equivalents). Dashed blue and orange line represent normalized coverage depth 0.1% and coverage breadth 0.1, respectively. Grey area and vertical line (center of area in terms of x axis) indicate regions where both coverage depth and breadth drop sharply as abundance rank increases. The vertical line was therefore used to define abundant (MAGs to the left of the line) vs. rare (MAGs to the right of the line) MAGs. Green line is a log fitting of coverage depth vs rank with fitting function shown above the line. For detailed model fitting of coverage depth distribution, see Supplementary figure S7.

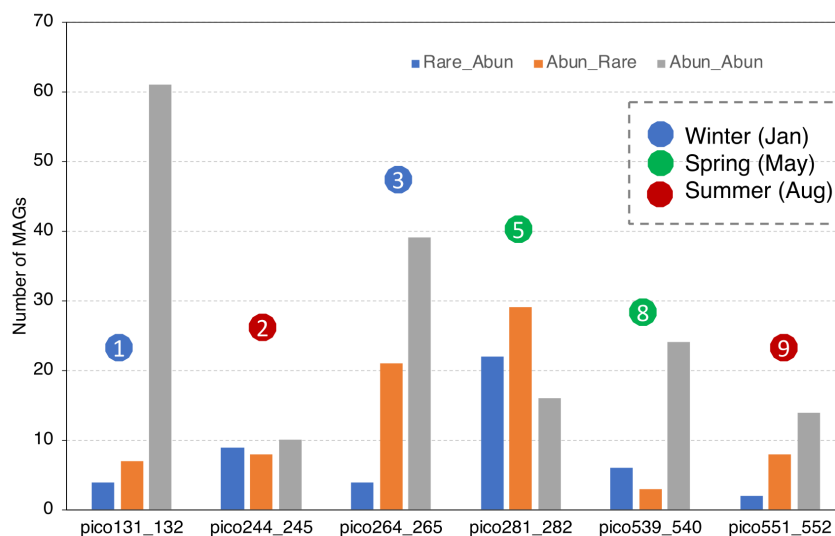


Figure 3. Responses of abundant vs. rare MAGs to the disturbance events. The figure shows the number of MAGs for each of the three categories assessed: abundant MAGs that remained abundant after the event, MAGs becoming abundant from rare, and MAGs becoming rare from abundant for each of the disturbance event. For one given event, if MAG's relative abundance and coverage breath fall below the threshold of being abundant in the pre-disturbance sample and fall above the threshold of abundant in the disturbed sample, this MAG will be in the category of Rare_Abun. Similar rules applied for other two categories. Disturbance events are labelled with a number as in Figure 1 (see Figure S1 for details). See Figure S10 for total relative abundance of MAGs for each category.

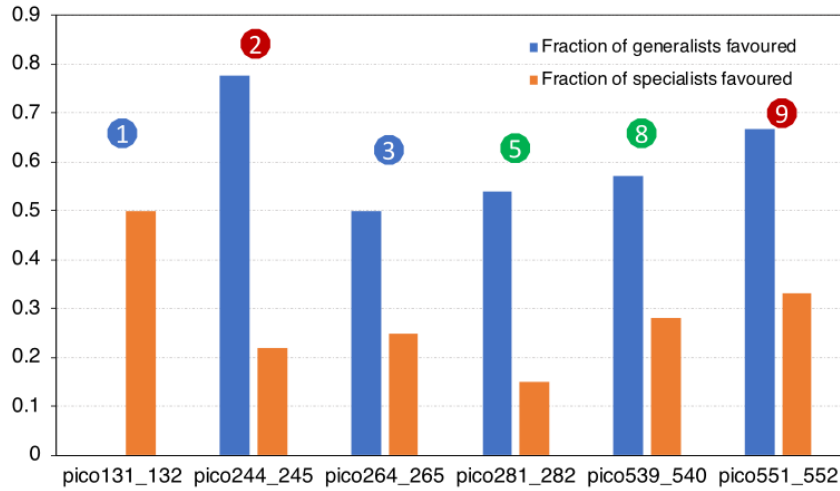


Figure 4. Fraction of specialists vs. generalists selected by each disturbance event in terms of number of MAGs. Selected MAGs are those that became abundant from rare in each disturbance event shown in Figure 3. Disturbance events are labelled with a number (see Figure S1 for details).

Supplementary Figures

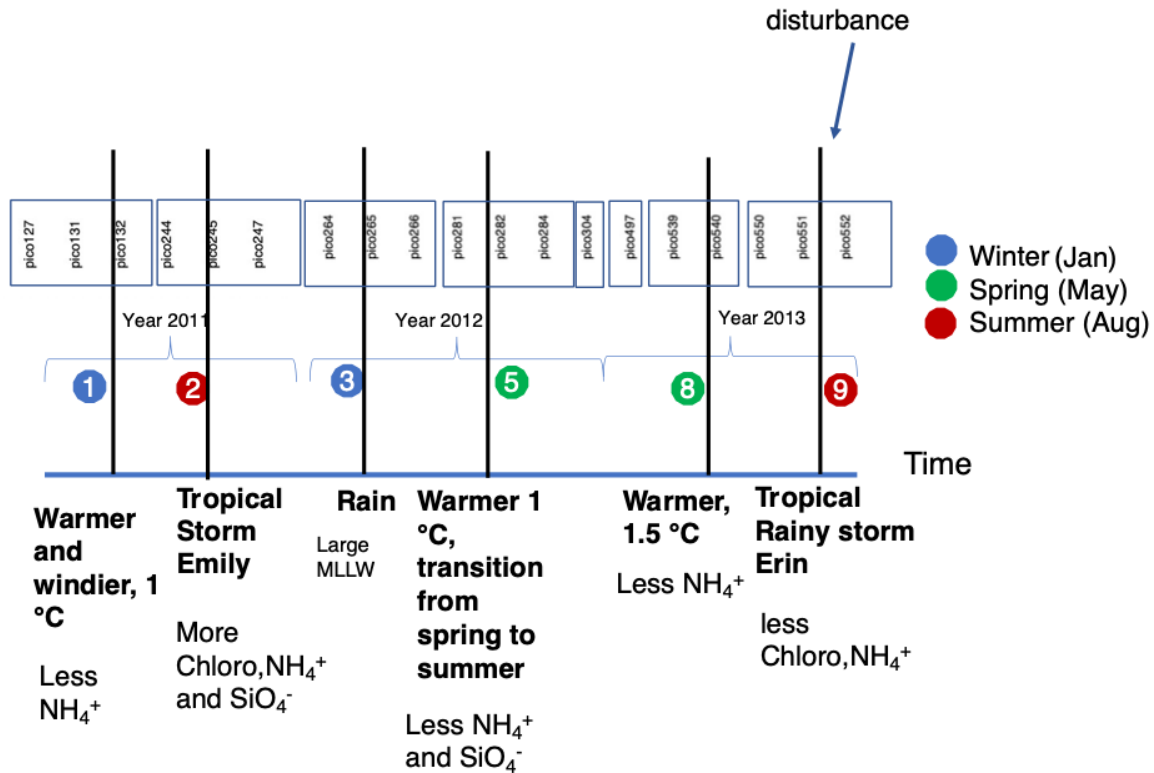


Figure S1. Details of the Pivers Island Coastal Observatory (PICO) time series samples. Sample names are labeled chronologically within boxes, the color of which corresponds to seasonality along the three-year sampling period. Continuous numbers in the name of samples indicate samples taken one week apart (e.g., pico281 and pioc282 are sampled from 2 adjacent weeks, while pico284 is sampled 2 weeks after pico282). Each vertical line indicates a disturbance event and is labelled with a number for convenience. Description below each vertical line shows the details of each disturbance and the most pronounced differences in environmental parameters measured. MLLW stands for the average lowest of the two low tides of a week.

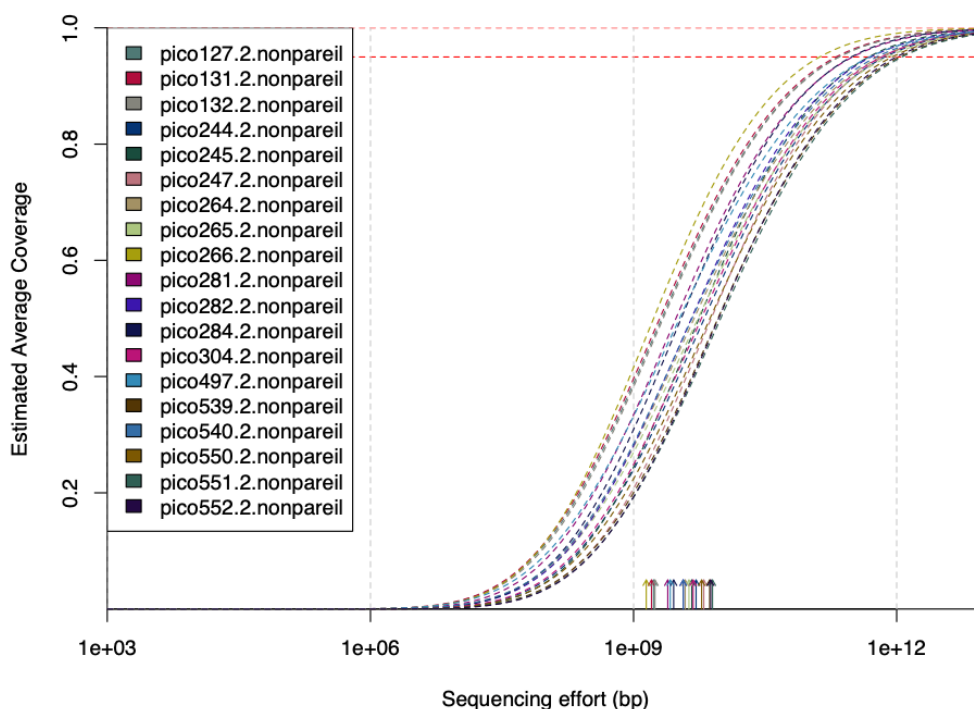


Figure S2. Nonpareil curves showing the coverage of subsampled metagenomes. Solid lines show the estimated average coverage (y-axis) as sequencing efforts increases (x-axis); dashed lines represent the projections for 95% and 99% coverage (horizontal dashed lines on the top). Only reverse reads were used for coverage estimation according to Nonpareil author recommendation for sequences used to not be linked/associated to each other (independent observations); forward reads showed similar curves for each sample (not shown).

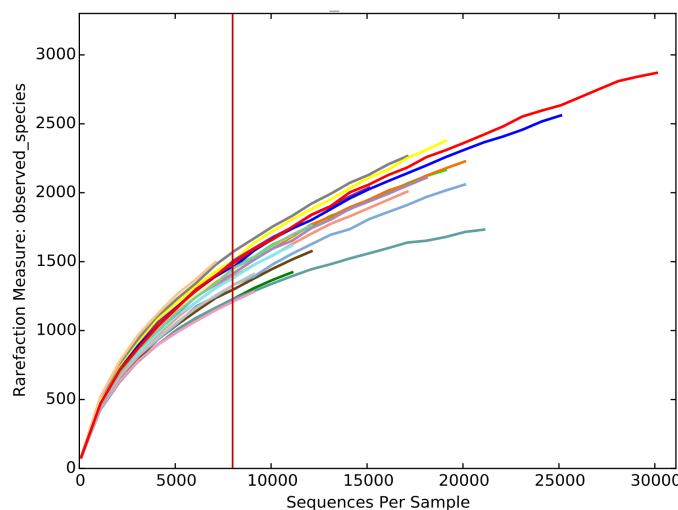


Figure S3. Rarefaction curve for extracted 16S-carrying reads from metagenomes. Each curve represents a sample. The vertical red line shows the number of reads subsampled (8000) for downstream analysis.

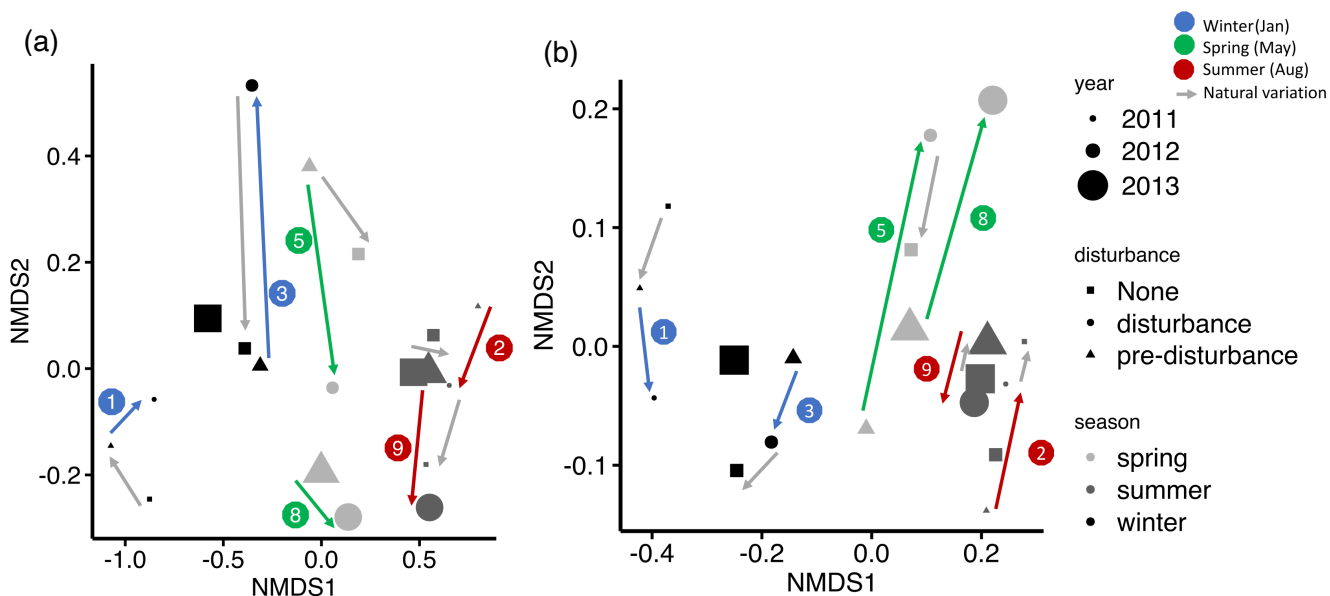


Figure S4. NMDS plot of amplicon 16S rRNA gene sequences (a) and extracted 16S rRNA gene-carrying reads from metagenomes (b). Arrows show the direction that the microbial community composition changed by each disturbance event. Grey arrows show the natural variation of metagenomic composition.

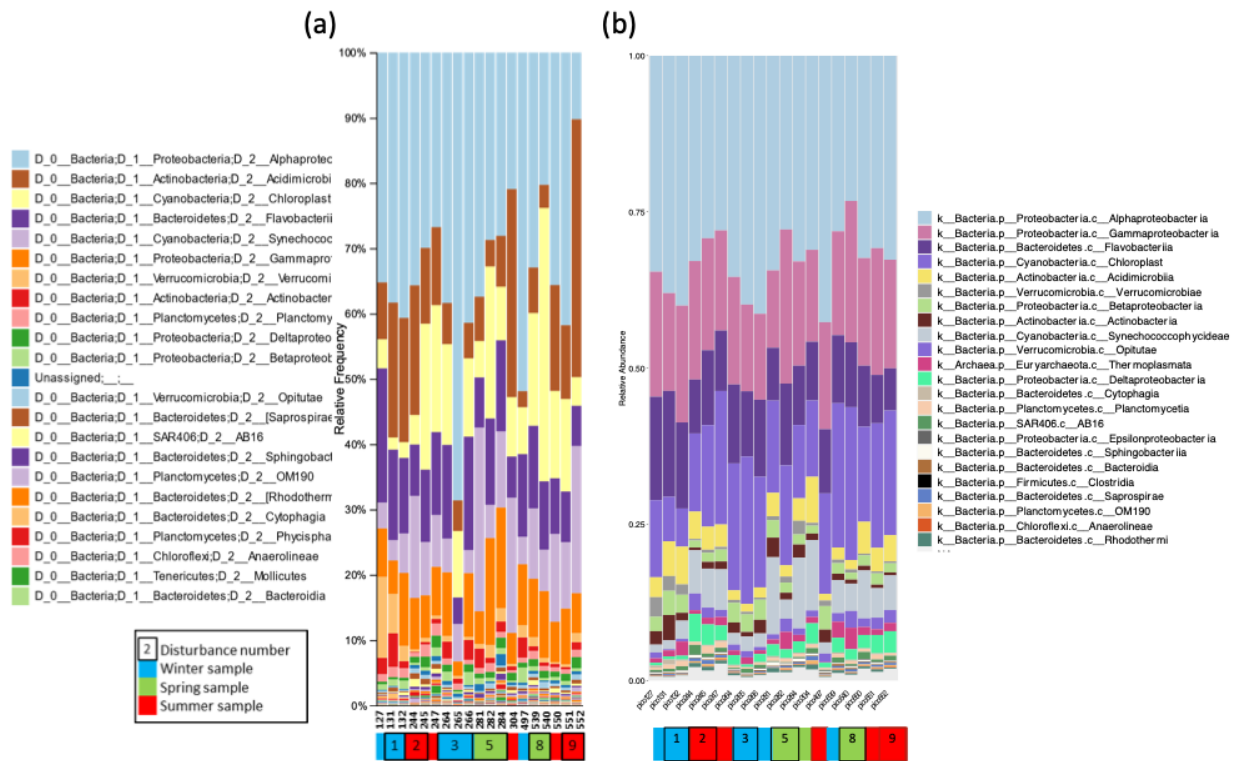


Figure S5. Class level microbial community composition (relative abundance) for amplicon 16S (a) and 16S-carrying reads extracted from metagenome (b). Each column represents a sample, with sample details provided by the color box. Disturbance events are labelled by the same number as in Figure S1. See Figure S1 for detailed explanation for each disturbance event.

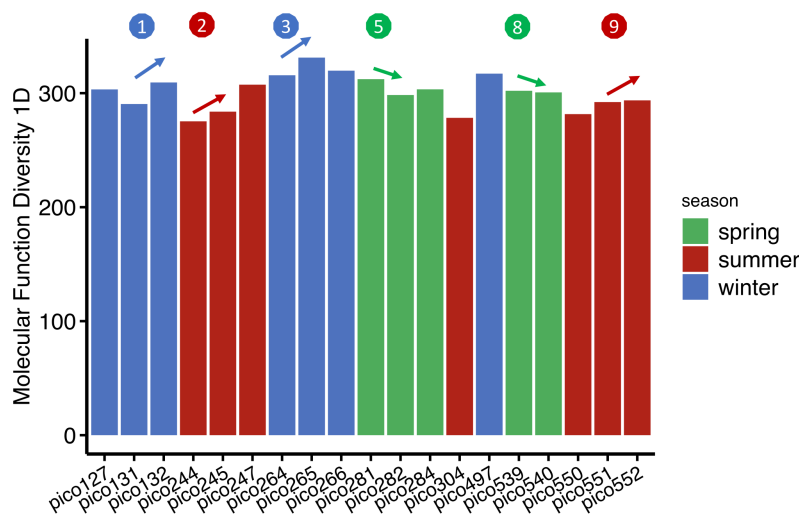


Figure S6. Changes in molecular functional diversity of metagenomes by each disturbance event as revealed by mapping reads to annotated functional pathways (See Methods & Materials).

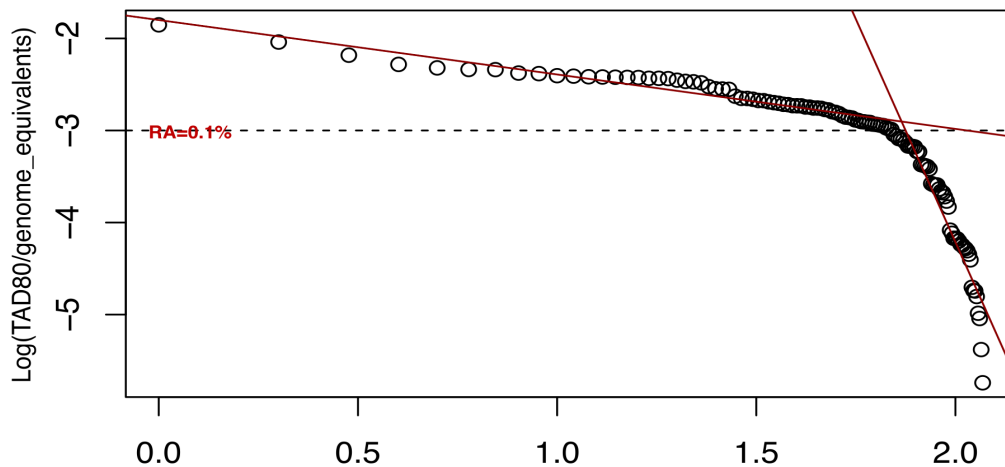


Figure S7. Log-log fitting of normalized sequence coverage depth vs. abundance rank for all MAGs used in the study. Two linear fittings were performed using MAGs with normalized coverage depth larger than 0.1% and smaller than 0.1% respectively ($R^2 > 0.7$). A clear difference in slope of two fit lines indicate a sharp decrease for normalized coverage depth around 0.1%.

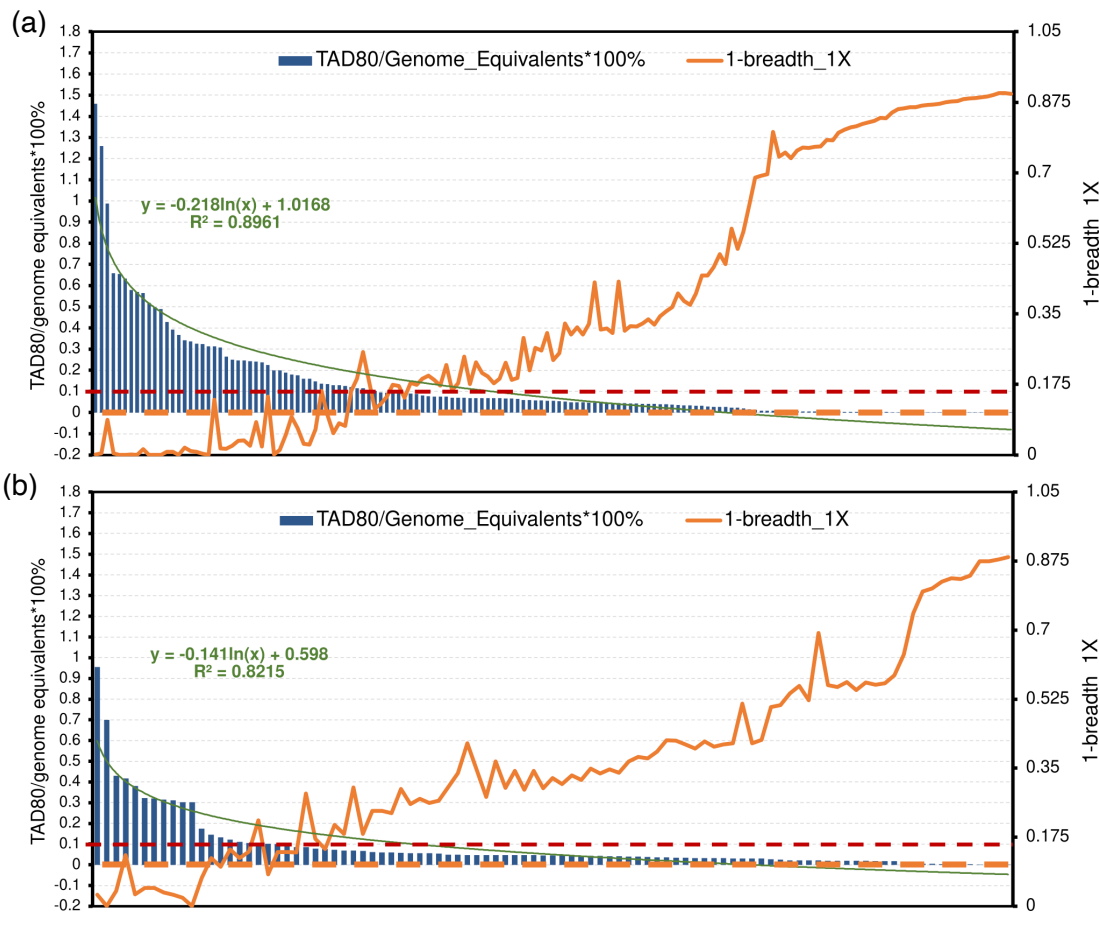


Figure S8. MAG coverage depth (left y axis, blue bar) and coverage breadth (right y axis, orange line, shown as 1- coverage breadth) distribution for two metagenomic samples, pico284 (a) and pico247 (b). This figure shows two additional examples and consistent patterns to those observed in Figure 3. X axis is MAG abundance rank based on coverage depth (TAD80 normalized by genome equivalents). Dashed red and orange lines represent normalized coverage depth 0.1% and coverage breadth 0.1, respectively. Green line is a log fitting of coverage depth vs rank with the corresponding function shown above it. Before subsampling, pico284 has similar sequencing depth with pico127 while pico247 is the shallowest sequenced sample of the three, thus only a few MAGs show high coverage depth (i.e., less reads could be mapped to the dereplicated 198 MAGs).

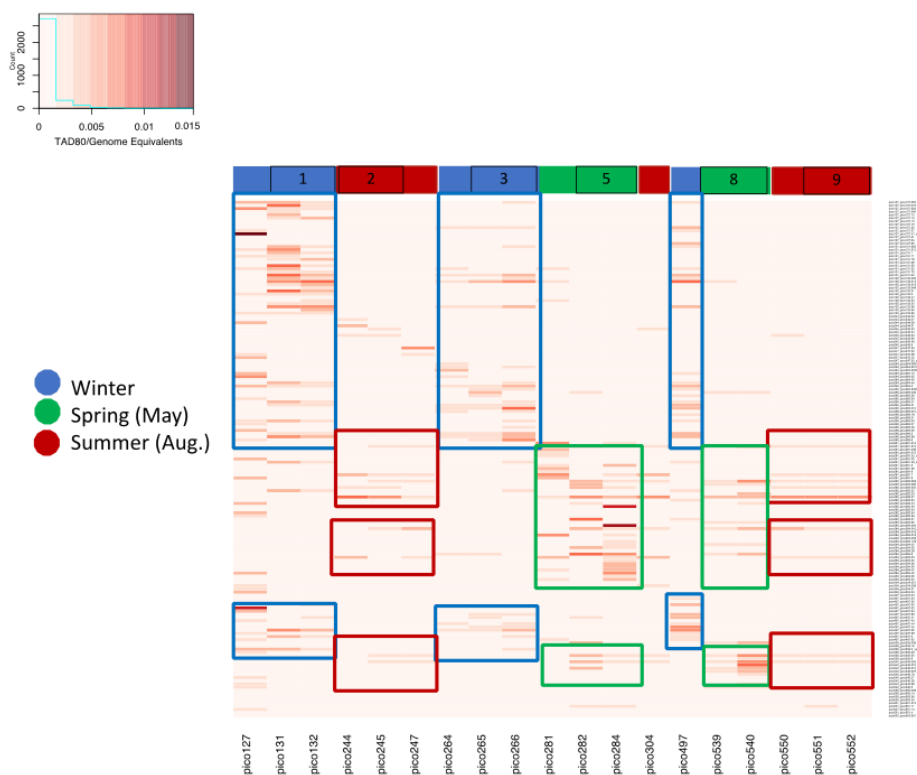


Figure S9. Heatmap of MAG relative abundance. Each row represents a MAG while each column represents a sample (see key for sample designation by color). Disturbance events are labelled by a number as in Figure 1. See Figure S1 for detailed explanation for each disturbance event.

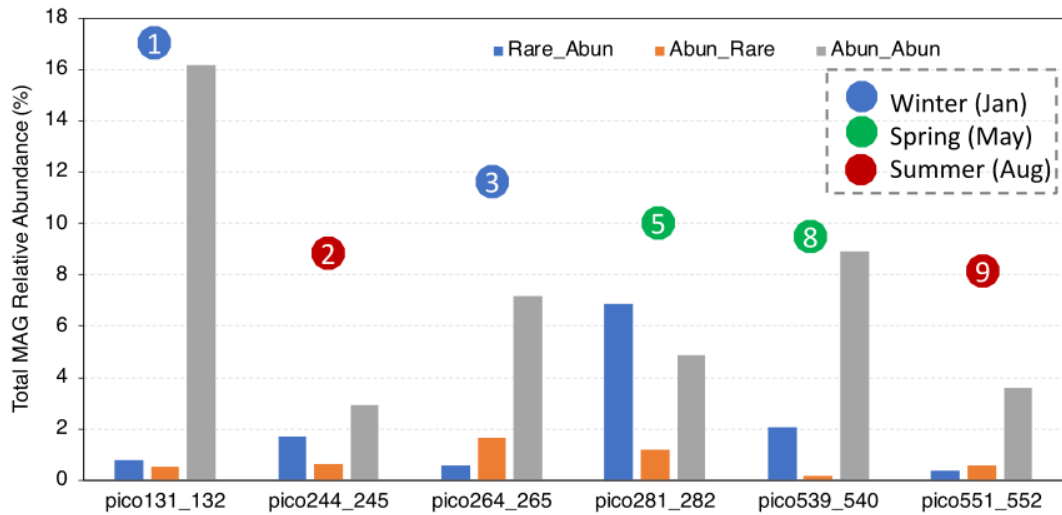


Figure S10. Total relative abundance of MAGs assigned to each category of abundance change. Categories were: remain abundant, become abundant from rare and become rare from abundant for each of the disturbance event (comparing samples before and after each disturbance event). Disturbance events are labelled with number (see Figure S1 for details).

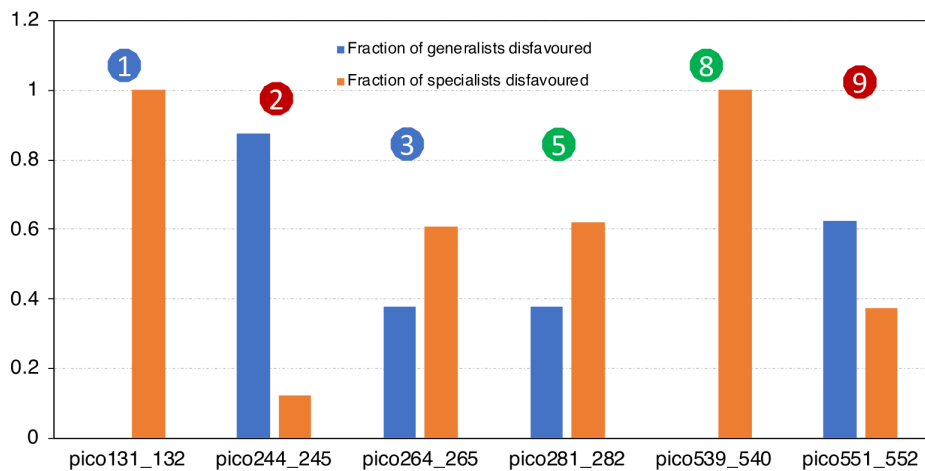


Figure S11. Fraction of specialists and generalists disfavored by each disturbance event. Disfavored MAGs are those that become rare from abundant by each disturbance event (similar to Figure 4 but the opposite pattern). Disturbance events are labelled with number (see Figure S1 for details).

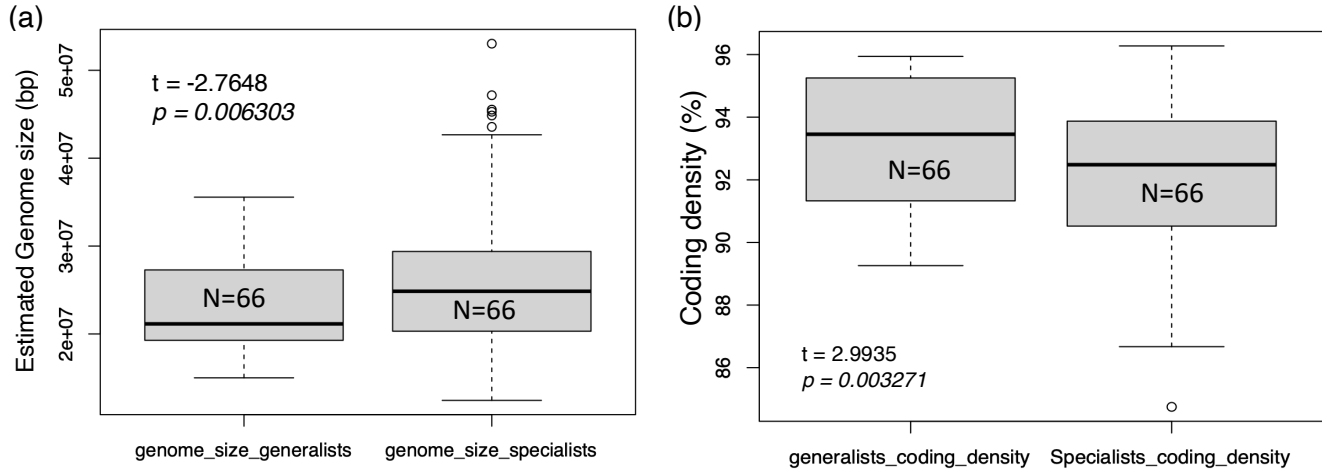


Figure S12. Estimated genome size (a) and coding density (b) differences between generalists and specialists. Both T-test and Mann–Whitney test were significant (note the p-values shown on the graphs).

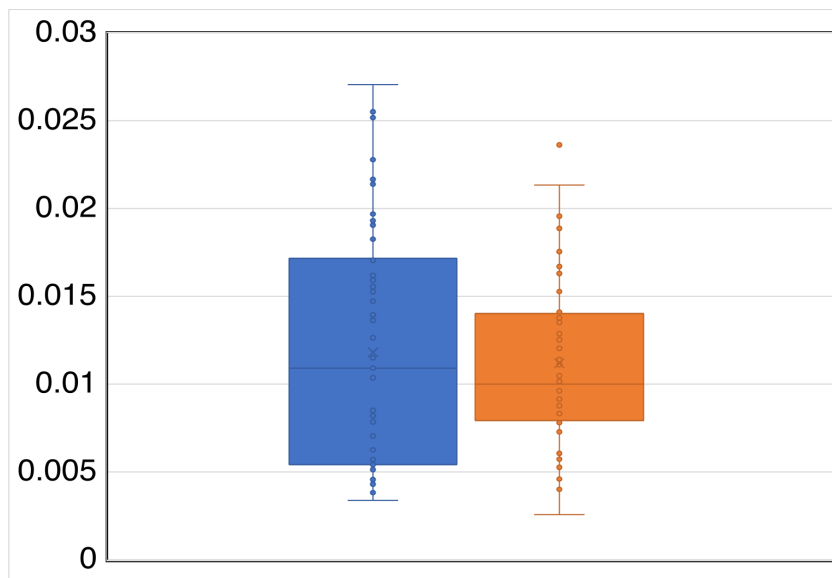


Figure S13. Proteins involved in extracellular activities, as a fraction of the total genes in the genome, for generalists (Blue) and specialists (Orange) taxa identified by this study. The difference is significant at $p < 0.05$ (Mann-Whitney test).

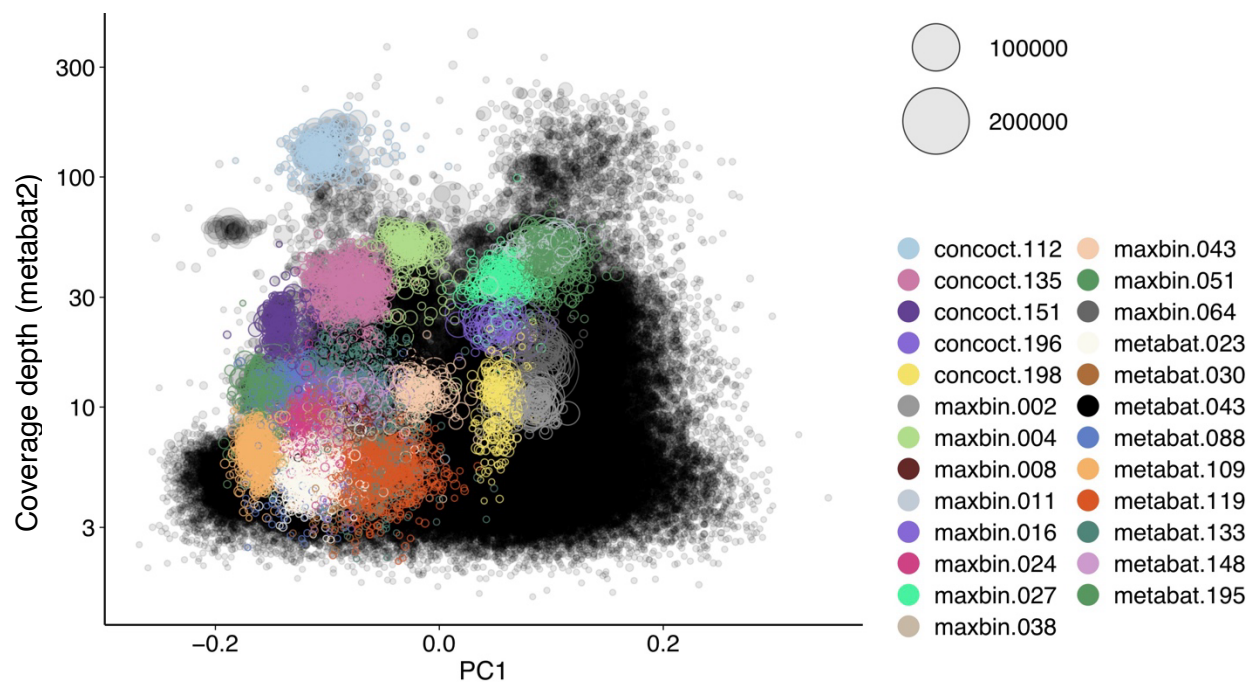


Figure S14. Contig coverage depth vs. principal component 1 of tetranucleotide frequencies for sample pico127 showed that each MAG we binned represents species-like, sequence clusters. Binned contigs by 3 different pieces of binning software (MaxBin2, MetaBAT2 and CONCOCT) are labelled by different color. The size of each circle represents the contig length.

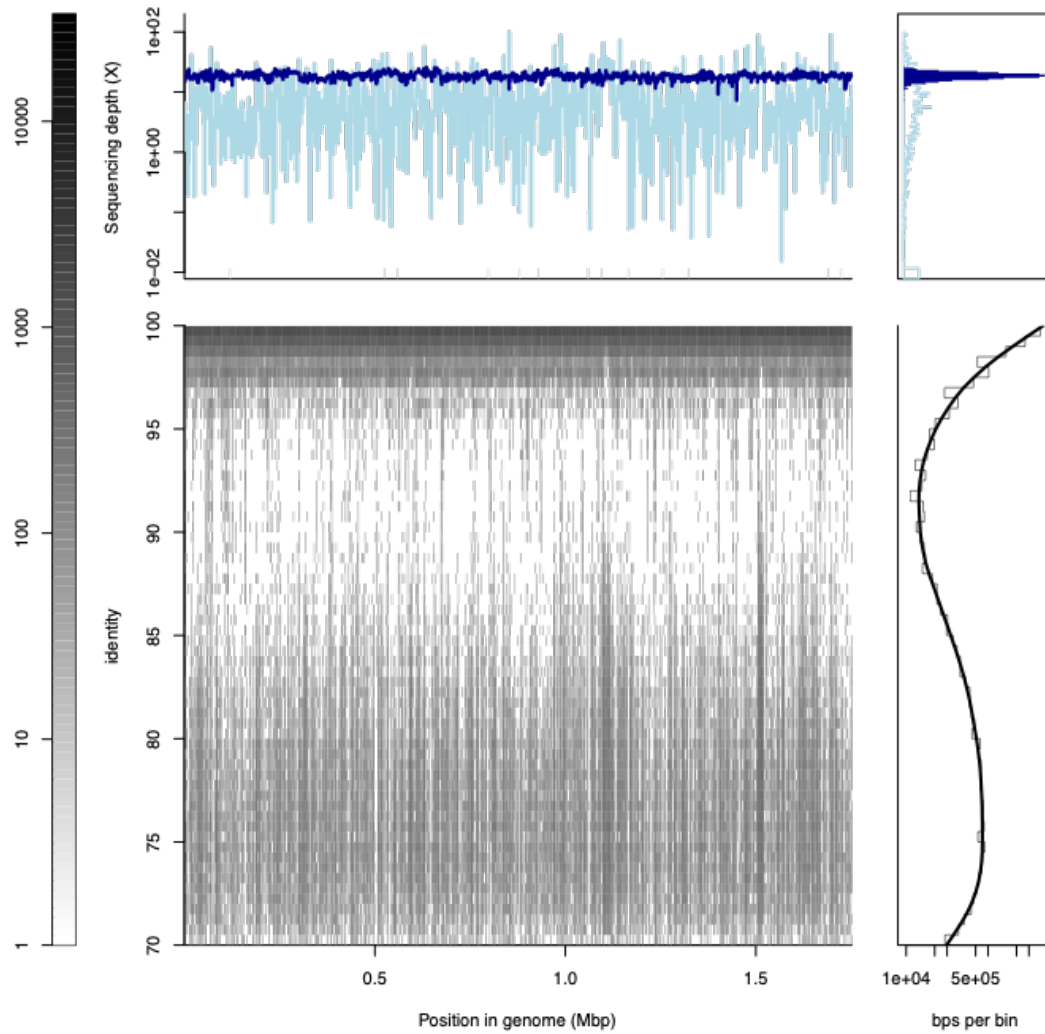


Figure S15. Recruitment plot of one MAG from sample pico497 (pico497_pico497.23.fasta) as an example that the binned MAGs represent sequence discrete populations. The reference MAG represents a sequence-discrete population in the pico 497 metagenome because there are many reads mapping on the MAG with >95% nucleotide, contrasting with reads showing 85-95% identity that are sparse. Average sequence coverage depth of this MAG is ~23X. An interactive version (up left and bottom left panel) of this plot is available: https://github.com/jianshu93/RecruitmentPlot_blast/blob/main/example_out/pico497.23.html.zip (download and then open it in a browser). All Recruitment plots for all MAGs of each sample are available here:

https://github.com/jianshu93/RecruitmentPlot_blast/tree/main/example_out/pico_rec_plot_2
 and here:

https://github.com/jianshu93/RecruitmentPlot_blast/tree/main/example_out/pico_rec_plot.

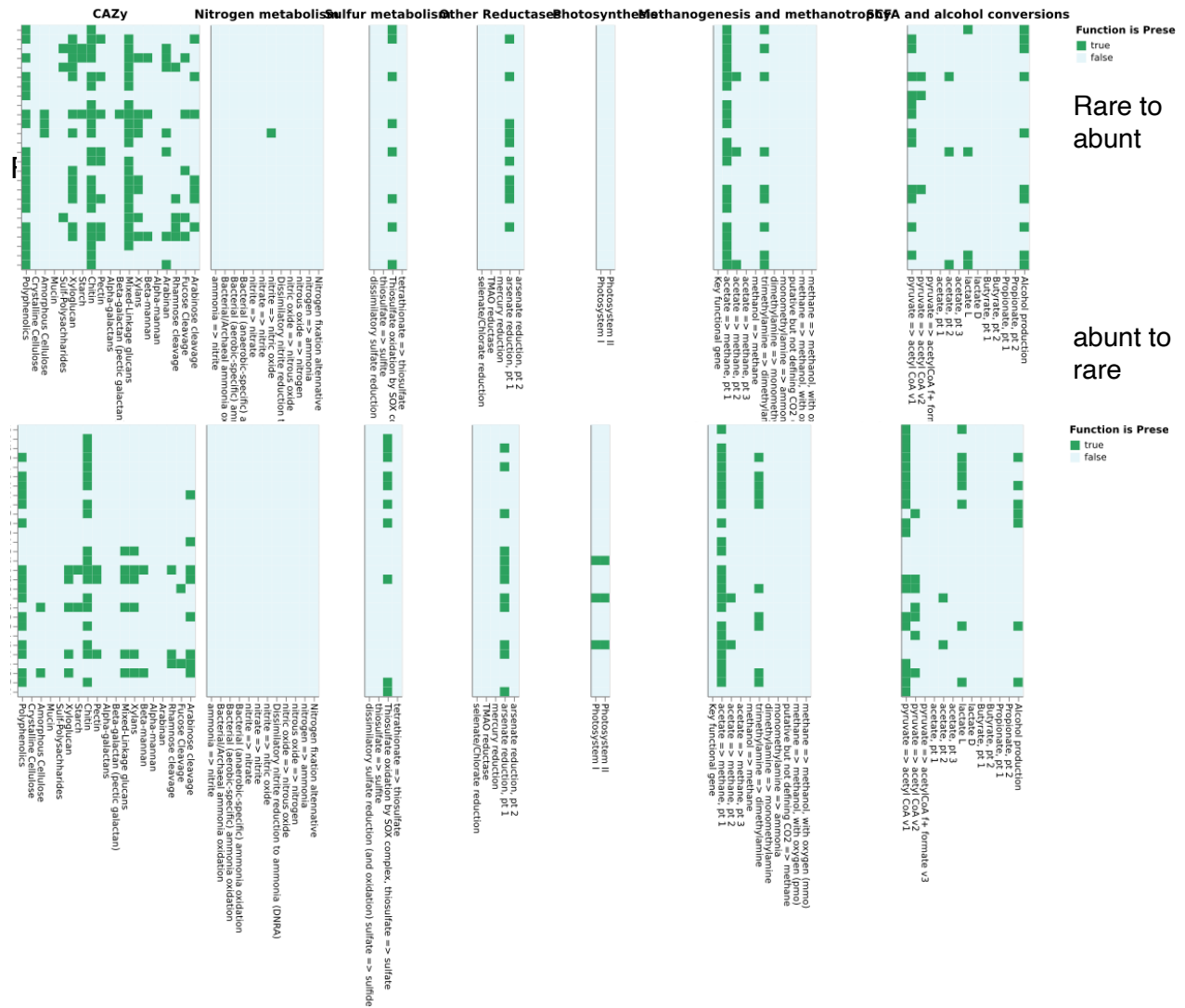


Figure S16. Metabolic pathways encoded in the genome of abundant-to-rare and rare-to-abundant MAGs for disturbance event 5 based on the DRAM software.

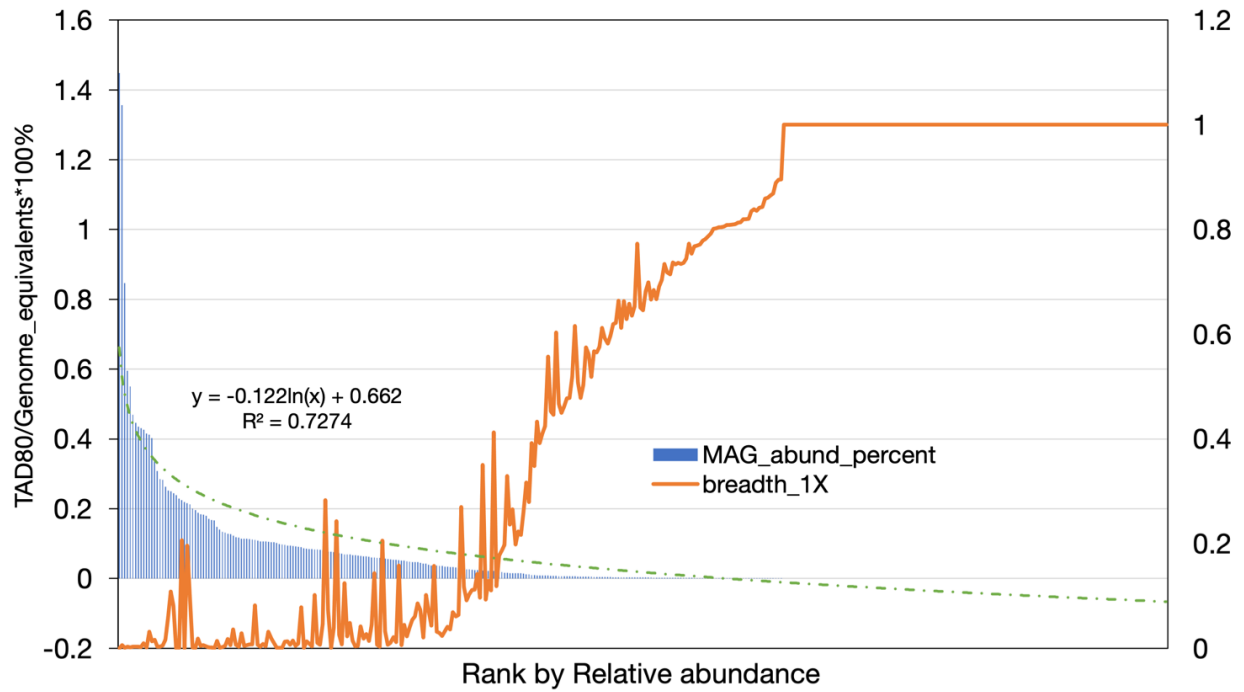


Figure 17. MAG sequence depth (left y axis, blue bar) and breadth (right y axis, orange line, shown as 1- coverage breadth) coverage distribution for pico127 before subsampling shows similar fitted line as the subsampled dataset and a slightly shifted abundance threshold for defining rare taxa (around 0.05).

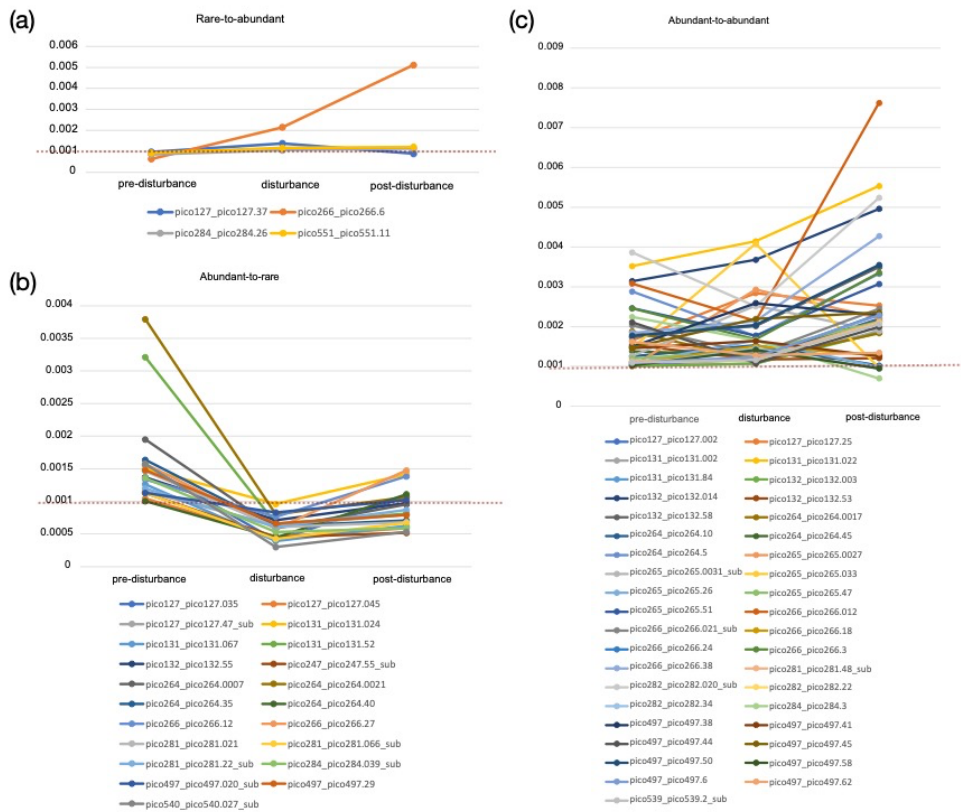


Figure S18. Relative abundance of MAGs assigned to rare to abundant (a), abundant to rare (b) and abundant to abundant (c) categories for disturbance event #3 (winter12_rainy_3 in Figure S1). Relative abundance (y-axis) was estimated as TAD80 sequence depth divided by genome equivalents to normalize for any average genome size differences between the samples as described in the Materials and Methods section. The red dashed line represents the threshold used to define rare taxa. Note the higher similarity in abundances between the pre- and post-disturbance samples relative to the disturbance sample, which indicates that stochastic processes have limited effect in identifying MAGs that change abundance categories due to the disturbance event.

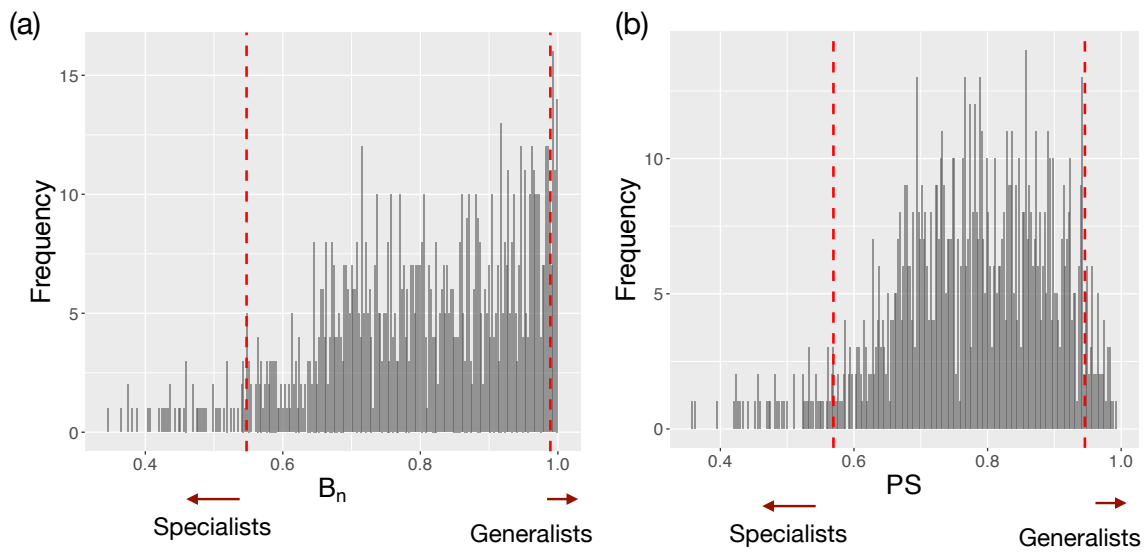


Figure S19. Levin's Breadth index (a) and PS index (b) in MicroNiche package to identify generalists and specialists.

Supplementary Tables

Table S1. Measured environmental variables for the time-series samples used in this study. Boldface numbers denote samples that associated with disturbance events. See Figure S1 for summary of each event.

	DayLength	Insolation	Temp	MLLW ^a	Barometric Pressure	Salinity	Oxygen	Oxygen Saturation	pH	DIC	Chloro	NH4	NOx ^b	PO4	SiO4
pico127	10.18	231.74	7.4	0.62	1005.3	32	9.74	99.8	7.92	1971	3.61	390.63	0.02	0.05	1.79
pico131	10.83	278.76	8.9	0.03	1031.8	32	9.89	104.9	7.92	1967.62	4.1	167.46	0	0.04	1.92
pico132	11.07	296.61	9.9	0.89	1026.7	35	9.37	103.5	7.93	2073.19	2.45	174.05	0	0.04	0.92
pico244	13.28	431.24	29	1.13	1020.9	34	6.55	103	8.08	2106.39	3.26	193.37	0	0.07	2.42
pico245	13.05	418.88	27.8	0.31	1019	35	5.85	90	8.02	2156.02	5.81	319.94	0	0.13	4.36
pico247	12.61	392.56	28	0.25	1012.9	31	6.39	97.1	7.94	1986.3	7.96	218.94	0.09	0.03	13.95
pico264	9.74	199.91	10	0.08	1018.7	31	8.8	94.3	7.92	2096.6	3.11	73.76	0.12	0	2.83
pico265	9.83	206.67	11.4	0.8	1019.6	33.5	8.73	98.1	7.92	2145.7	2.4	303.85	0.19	0	2.05
pico266	9.96	215.95	9.6	0.2	1024.7	33	9.16	99.2	7.89	2157.37	2.38	229.76	0.1	0.01	3.58
pico281	13.45	448.19	21.1	-0.05	1023.5	30.5	7.74	103.1	7.87	2060.23	5.65	153	0.06	0.06	8.25
pico282	13.65	457.68	21.8	1.12	1017.2	34	7.84	106.8	7.95	2157.2	3.86	130	0.13	0	3.64
pico284	13.99	472.05	22.3	0.94	1013.2	33	7.39	102.8	7.92	2119.8	5.9	123.5	0.03	0.02	7.52
pico304	14.08	471.45	29.1	0.36	1016	33	6.74	107.2	8.01	2135.5	6.1	102	0	0	5.52
pico497	10.52	256.02	10.1	0	1021.3	32.75	7.99	86.9	7.9	2115.77	1.82	24.5	0	0.02	1.94
pico539	14.13	477.05	23	0.83	1026.1	34.78	7.03	97.7	7.94	2111.17	3.02	22	0.04	0.05	1.55
pico540	14.23	480.63	24.7	0.18	1020.3	34.57	6.65	95.1	7.95	2107.43	6.8	9	0.04	0.02	1.98
pico550	13.57	446.17	26.9	0.84	1020	34.99	6.25	94.7	7.95	2099.73	5.52	29.5	0	0.03	4.52
pico551	13.35	435.45	28.3	0.48	1013.8	35.41	5.73	89	7.97	2115.97	6.71	95	0.03	0.05	6.76
pico552	13.13	423.42	26.9	0.89	1024	34.97	6.26	94.7	7.95	2106.8	5.46	53	0	0.03	5.29

^amean of the two low tides.

Table S2. Quality and classification of the dereplicated, high quality MAGs used in the study. Quality was assessed based on CheckM results while classification is from GTDB-tk v1.1 against GTDB v207.

Genome name	Completeness	Contamination	classification
pico127_pico127.002	92.02	5.44	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata
pico127_pico127.018	75.91	7.94	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaaceae;g__UBA985;s__
pico127_pico127.023	92.47	6.38	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Porricoccaceae;g__HTCC2207;s__HTCC2207 sp002685195
pico127_pico127.035	90.35	8.53	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Pontimonas;s__
pico127_pico127.045	93.35	4.69	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA4582;s__UBA4582 sp002389265
pico127_pico127.062	82.36	5.96	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Nanopelagiales;f__S36-B12;g__UBA6154;s__
pico127_pico127.12	80.09	4.08	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Algibacter_B;s__
pico127_pico127.14	89.57	4.8	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Ilumatobacteraceae;g__Ilumatobacter_A;s__Ilumatobacter_A sp002711735
pico127_pico127.15	89.93	3.47	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Winogradskyella;s__
pico127_pico127.16	88.44	0.82	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MAG-120531;s__
pico127_pico127.19	77.92	0	d__Archaea;p__Thermoplasmata;c__Poseidonii;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-L1;s__MGIIa-L1 sp002506275
pico127_pico127.25	90.15	1.26	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__LFE01;s__LFE01 sp001642945
pico127_pico127.37	78.24	1.7	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Amylibacter;s__Amylibacter sp900197625
pico127_pico127.47_sub	77.6	1.88	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003045825
pico127_pico127.51	86.6	0.93	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__UBA9145;s__UBA9145 sp003483155
pico127_pico127.64	86.83	2.96	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp002387615
pico127_pico127.69	81.75	1.24	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__CAU-1491;s__
pico131_pico131.002	96.06	6.69	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata
pico131_pico131.013_sub	92.13	9.12	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Nanopelagiales;f__S36-B12;g__UBA6154;s__
pico131_pico131.022	88.06	6.49	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__Puniceispirillum;s__
pico131_pico131.024	91.4	6.18	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003045825
pico131_pico131.047	86.21	3.71	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Porricoccaceae;g__HTCC2207;s__
pico131_pico131.048	82.69	4.69	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Pontimonas;s__
pico131_pico131.067	75.61	10.7	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__OM182;s__OM182 sp003482475
pico131_pico131.1	98.11	4	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Crocinitomicaceae;g__Crocinitomix;s__
pico131_pico131.11	97.43	0.74	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MS024-2A;s__
pico131_pico131.18	88.79	5.73	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Ilumatobacteraceae;g__Ilumatobacter_A;s__Ilumatobacter_A sp002711735
pico131_pico131.28	92.57	0.83	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MAG-120531;s__
pico131_pico131.30	87.02	0.33	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__s__
pico131_pico131.50	87.93	9.84	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__GCA-2707915;s__
pico131_pico131.52	77.93	5.87	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__BACL14;s__
pico131_pico131.78	81.42	1.57	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Crocinitomicaceae;g__UBA4466;s__
pico131_pico131.84	85.91	2.32	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__BACL14;s__
pico132_pico132.003	91.08	5.93	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata

pico132_pico132.014	86.1	8.79	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Punicei spirillaceae;g__Puniceispirillum;s__
pico132_pico132.019	91.97	7.36	d__Bacteria;p__Actinobacteriota;c__Acidimicrobia;o__Acidimicrobiales;f__Ilumatobact eraceae;g__Ilumatobacter_A;s__Ilumatobacter_A sp002711735
pico132_pico132.023	87.93	7.97	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__GC A-2707915;s__
pico132_pico132.046	84	1.72	d__Archaea;p__Thermoplasmatota;c__Poseidoniiia;o__Poseidoniales;f__Poseidoniac eae;g__MGIIa-L1;s__MGIIa-L1 sp002506275
pico132_pico132.049	85.91	1.79	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Ps eudohongiellaceae;g__UBA9145;s__UBA9145 sp003483155
pico132_pico132.21	85.42	1.16	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae ;g__s__
pico132_pico132.3	85.63	1.08	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__TMED109;f__TMED109;g__ GCA-2684605;s__GCA-2684605 sp002684605
pico132_pico132.47	77.3	0.33	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae ;g__MAG-120531;s__
pico132_pico132.53	83.64	4.45	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__SG8- 40;g__UBA3031;s__UBA3031 sp003483505
pico132_pico132.55	85.11	1.03	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__ _UBA10364;s__UBA10364 sp003045825
pico132_pico132.58	84.38	1.97	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Methy lophilaceae;g__BACL14;s__
pico132_pico132.76	93.1	8.82	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae ;g__MS024-2A;s__
pico132_pico132.80	94.9	0.8	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Sphi ngomonadaceae;g__Sphingobium;s__Sphingobium yanoikuyae
pico132_pico132.86	80.73	2.15	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__ _TMED14;s__
pico132_pico132.91	82.4	4.72	d__Archaea;p__Thermoplasmatota;c__Poseidoniiia;o__Poseidoniales;f__Thalassarcha eaceae;g__MGIIb-O2;s__
pico244_pico244.021	86.37	4.42	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodo bacteraceae;g__HIMB11;s__HIMB11 sp003486095
pico244_pico244.22	79.84	2.69	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__ _UBA10364;s__UBA10364 sp003023665
pico244_pico244.27	76.61	7.04	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Salibacteraceae;g__ SHAN690;s__
pico244_pico244.28	87.78	1.98	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Ps eudohongiellaceae;g__UBA9145;s__
pico244_pico244.31	93.49	2.69	d__Bacteria;p__Planctomycetota;c__UBA8108;o__UBA1146;f__UBA1146;g__UBA121 91;s__
pico245_pico245.104	77.54	2.61	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__ _;s__
pico245_pico245.20	92.75	2.67	d__Bacteria;p__Cyanobacteria;c__Cyanobacteriia;o__PCC- 6307;f__Cyanobiaceae;g__Synechococcus_C;s__
pico245_pico245.24	77.3	1.06	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae ;g__Winogradskyella;s__Winogradskyella sp003335675
pico245_pico245.34	87.77	0.57	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae ;g__BACL21;s__
pico245_pico245.36	84.05	2.47	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__ _UBA10364;s__UBA10364 sp003023665
pico245_pico245.46	84.02	5.9	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Lit oricolaceae;g__Litoricola;s__Litoricola sp002691485
pico245_pico245.8	82.36	1.02	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Ps eudohongiellaceae;g__UBA9145;s__
pico247_pico247.26	88.88	2.9	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae ;g__Polaribacter;s__
pico247_pico247.32	83.12	1.08	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__ _UBA10364;s__UBA10364 sp003023665
pico247_pico247.38	89.76	2.07	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Cycl oclasticaceae;g__Cycloclasticus;s__Cycloclasticus sp002700385
pico247_pico247.42	83.72	0.93	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Ps eudohongiellaceae;g__UBA9145;s__
pico247_pico247.55_sub	81.13	1.56	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Punicei spirillaceae;g__Puniceispirillum;s__
pico264_pico264.0007	84.38	5.83	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Ps eudohongiellaceae;g__OM182;s__OM182 sp003482475
pico264_pico264.0017	76.49	7.15	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Ps eudohongiellaceae;g__OM182;s__
pico264_pico264.0021	92.46	2.05	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Lit oricolaceae;g__Litoricola;s__Litoricola sp002691485
pico264_pico264.10	81.85	2.05	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodo bacteraceae;g__Amylibacter;s__Amylibacter sp900197625
pico264_pico264.35	94.24	4.37	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Punicei spirillaceae;g__Puniceispirillum;s__
pico264_pico264.40	92.57	0.7	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae ;g__MAG-120531;s__
pico264_pico264.45	88.73	0.96	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodo bacteraceae;g__Planktomarina;s__Planktomarina temperata
pico264_pico264.5	86.39	1.11	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Po rticoccaceae;g__Porticoccus;s__Porticoccus sp002390525

pico265_pico265.0027	75.02	7.34	d__Archaea;p__Thermoplasmata;c__Poseidonii;o__Poseidoniales;f__Thalassarchaeaceae;g__MGIIb-O2;s__MGIIb-O2 sp002498985
pico265_pico265.0031_sub	82.41	7.5	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__L_FER01;s__L_FER01 sp001642945
pico265_pico265.033	88.62	7.81	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__UA16;g__UBA8752;s__
pico265_pico265.26	82.47	1.82	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Amylibacter;s__Amylibacter sp900197625
pico265_pico265.33	90.21	1.23	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico265_pico265.47	96.2	0.9	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata
pico265_pico265.51	84.95	2.22	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Porticoccaceae;g__Porticoccus;s__Porticoccus sp002390525
pico266_pico266.012	93.2	6.24	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__
pico266_pico266.021_sub	87.71	4.82	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003045825
pico266_pico266.106	77.45	3.64	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__Opitutaceae;g__UBA5691;s__
pico266_pico266.12	92.85	7.13	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__Puniceispirillum;s__
pico266_pico266.18	79.44	2.38	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Amylibacter;s__Amylibacter sp900197625
pico266_pico266.21	92.39	0.22	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MAG-120531;s__
pico266_pico266.24	93.89	0.35	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata
pico266_pico266.27	75.88	2.22	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__OM182;s__OM182 sp003482475
pico266_pico266.28	94.49	0.76	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MS024-2A;s__
pico266_pico266.29	91.47	0.56	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico266_pico266.3	88.18	5.53	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Porticoccaceae;g__Porticoccus;s__Porticoccus sp002390525
pico266_pico266.38	85.84	3.41	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__BACL14;s__
pico266_pico266.6	83.79	0.95	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA7446;s__
pico281_pico281.012	94.59	3.85	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Acidimicrobiales;f__Ilumatobacteraceae;g__Casp-actino5;s__
pico281_pico281.019	94.07	3.61	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__HIMB11;s__HIMB11 sp003486095
pico281_pico281.021	96.24	8.03	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__Puniceispirillum;s__
pico281_pico281.039	93.77	2.75	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Sphingomonadales;f__Kordiimonadaceae;g__s__
pico281_pico281.049	98.9	2.09	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA3478;s__UBA3478 sp003045935
pico281_pico281.066_sub	94.36	8.85	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__OM182;s__OM182 sp003482475
pico281_pico281.069_sub	75.9	3.61	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Actinomycetales;f__Microbacteriaceae;g__Pontimonas;s__
pico281_pico281.072	83.86	7.01	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Acidimicrobiales;f__Ilumatobacteraceae;g__Casp-actino5;s__
pico281_pico281.22_sub	87.42	0.43	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003045825
pico281_pico281.42	89	0.93	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico281_pico281.43	91.08	20.2	d__Bacteria;p__Cyanobacteria;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__Cyanobium_A;s__
pico281_pico281.48_sub	89.74	2.13	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata
pico281_pico281.5	97.61	0.84	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Kordia;s__
pico281_pico281.58	84.6	2.07	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__s__
pico281_pico281.6	77.82	3.06	d__Bacteria;p__Actinobacteriota;c__Actinomycetia;o__Nanopelagicales;f__S36-B12;g__Mxb001;s__
pico281_pico281.7	84.48	5.96	d__Bacteria;p__Cyanobacteria;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__Synechococcus_E;s__
pico281_pico281.8	97.94	2.15	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__UBA10329;g__UBA10329;s__
pico282_pico282.004	93.38	3.08	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__HIMB11;s__HIMB11 sp003486095
pico282_pico282.006	89.5	8.31	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Halieaceae;g__Luminiphilus;s__Luminiphilus sp002691565
pico282_pico282.020_sub	80	7.51	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__BACL14;s__

pico282_pico282.025_sub	76.36	7.79	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Halieaceae;g__Luminiphilus;s__
pico282_pico282.20	87.77	2.32	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA3478;s__
pico282_pico282.22	89.71	1.37	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata
pico282_pico282.23	87.23	3.72	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__Puniceicocaceae;g__s__
pico282_pico282.27	83.35	2.47	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__MED-G52;s__MED-G52 sp002457055
pico282_pico282.34	76.82	5.12	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Amylibacter;s__Amylibacter sp900197625
pico282_pico282.37	90.13	5.07	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Halieaceae;g__Luminiphilus;s__
pico282_pico282.44	98.53	0.83	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA3478;s__UBA3478 sp003045935
pico282_pico282.45	81.33	0.13	d__Archaea;p__Thermoplasmatota;c__Poseidoniiia;o__Poseidoniales;f__Poseidoniaceae;g__MGIIa-K1;s__
pico282_pico282.49	83.19	0.93	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Poricoccaceae;g__HTCC2207;s__
pico282_pico282.53	83.12	6.68	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Winogradskyella;s__Winogradskyella sp003335675
pico282_pico282.54	83.35	0.49	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico282_pico282.56	91.3	0.54	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003023665
pico282_pico282.57	89.14	1.67	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__UBA7434;g__UBA7434;s__UBA7434 sp002480045
pico282_pico282.69	86.83	2.58	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__Puniceispirillum;s__
pico284_pico284.003	92.84	7.42	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Poricoccaceae;g__HTCC2207;s__
pico284_pico284.010_sub	89.04	6.48	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__HIMB11;s__HIMB11 sp003486095
pico284_pico284.016_sub	91.02	1.3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico284_pico284.018_sub	82.19	0.85	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Acidimicrobiales;f__Illumatobacteraceae;g__Casp-actino5;s__
pico284_pico284.039_sub	95.99	2.53	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__Puniceispirillum;s__
pico284_pico284.124	76.07	2.94	d__Bacteria;p__Planctomycetota;c__UBA8742;o__UBA2392;f__g__s__
pico284_pico284.22	88.43	3.02	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__Puniceispirillum;s__
pico284_pico284.26	82.94	1.4	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__LFFER01;s__
pico284_pico284.28	84.73	1.63	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003023665
pico284_pico284.3	94.42	2.42	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Parvibaculales;f__RS24;g__UBA8337;s__UBA8337 sp900197605
pico284_pico284.33	79.31	4.35	d__Bacteria;p__Cyanobacteria;c__Cyanobacteriia;o__PCC-6307;f__Cyanobiaceae;g__Synechococcus_E;s__
pico284_pico284.35	81.59	0.92	d__Bacteria;p__Planctomycetota;c__Phycisphaerae;o__Phycisphaerales;f__Phycisphaeraeae;g__s__
pico284_pico284.38	92.71	6.96	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Parvibaculales;f__RS24;g__UBA8337;s__
pico284_pico284.39	93.71	1.44	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaceae;g__SW10;s__
pico284_pico284.42	91.67	0.68	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Kordia;s__
pico284_pico284.45	75.81	3.76	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__BACL11;g__UBA8444;s__
pico284_pico284.66	93.16	1.71	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Acidimicrobiales;f__UBA11606;g__UBA11606;s__
pico284_pico284.84	82.91	0.95	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Acidimicrobiales;f__Illumatobacteraceae;g__Casp-actino8;s__
pico304_pico304.021	94.06	4.23	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico304_pico304.038	87.9	5.66	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__s__
pico304_pico304.21	75.39	1.24	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Winogradskyella;s__
pico304_pico304.24	82.79	2.21	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__
pico304_pico304.46	86.13	1.6	d__Bacteria;p__Actinobacteriota;c__Acidimicrobiia;o__Acidimicrobiales;f__Illumatobacteraceae;g__Casp-actino5;s__
pico497_pico497.020_sub	89.77	8.4	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__CAU-1491;s__
pico497_pico497.11	84.48	7.34	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Halieaceae;g__Luminiphilus;s__Luminiphilus sp002390485

pico497_pico497.113	79.72	14.4	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Akkermansiaceae;g__UBA985;s__
pico497_pico497.23	84.76	0.14	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MAG-120531;s__
pico497_pico497.24	90.86	0.43	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Poricoccaceae;g__HTCC2207;s__HTCC2207 sp002457245
pico497_pico497.29	95.47	0.87	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__Puniceispirillum;s__
pico497_pico497.30	79.3	3.08	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Caulobacterales;f__Maricaulaceae;g__Hellea;s__
pico497_pico497.32	95.22	1.1	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__MS024-2A;s__
pico497_pico497.34	81.81	5.33	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA4582;s__UBA4582 sp002389265
pico497_pico497.38	88.26	2.11	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__L_FER01;s__L_FER01 sp001642945
pico497_pico497.41	85.45	1.04	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA3537;s__UBA3537 sp001735715
pico497_pico497.43	90.17	0.42	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico497_pico497.44	90.65	3.33	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA9926;s__
pico497_pico497.45	91.11	1.53	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA4421;s__
pico497_pico497.50	78.48	1.75	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__Methylophilaceae;g__BACL14;s__
pico497_pico497.58	80.54	1.29	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Amylibacter;s__Amylibacter sp900197625
pico497_pico497.6	94.47	0.48	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata
pico497_pico497.62	83.76	1.83	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Burkholderiales;f__SG8-40;g__UBA3031;s__UBA3031 sp003485335
pico539_pico539.006	96.37	8.93	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico539_pico539.009	90.51	3.66	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Halieaceae;g__Luminiphilus;s__Luminiphilus sp002691565
pico539_pico539.16	84.29	1.59	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Cycloclasticaceae;g__Cycloclasticus;s__Cycloclasticus pugetii
pico539_pico539.2_sub	81.13	0.54	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Planktomarina;s__Planktomarina temperata
pico539_pico539.23	89.97	1.75	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003023665
pico539_pico539.35	91.26	1.92	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__HIMB11;s__HIMB11 sp003486095
pico539_pico539.8	89.68	1.47	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA3478;s__UBA3478 sp003045935
pico540_pico540.003	91.46	3.09	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__HIMB11;s__HIMB11 sp003486095
pico540_pico540.010	96.02	6.44	d__Bacteria;p__Verrucomicrobiota;c__Verrucomicrobiae;o__Opitutales;f__Puniceicoccaceae;g__s__
pico540_pico540.012	87.71	4.7	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Halieaceae;g__Luminiphilus;s__Luminiphilus sp002691565
pico540_pico540.027_sub	87.68	6.04	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__
pico540_pico540.18	84.84	2.43	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA3478;s__
pico540_pico540.27	93.26	1.6	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico540_pico540.28	83.87	0.54	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003023665
pico540_pico540.38	75.54	0.77	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA3478;s__UBA3478 sp003045935
pico540_pico540.5	87.76	1.62	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Puniceispirillales;f__Puniceispirillaceae;g__s__
pico550_pico550.028	95.79	3.04	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico550_pico550.037	87.41	9.97	d__Bacteria;p__Cyanobacteria;c__Cyanobacteria;o__PCC-6307;f__Cyanobiaceae;g__Vulcanococcus;s__Vulcanococcus sp000179255
pico550_pico550.14	84.38	3.39	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Winogradskyella;s__Winogradskyella sp003335675
pico550_pico550.28	82.65	1.48	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__UBA9145;s__
pico550_pico550.45	82.71	1.66	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Schleiferiaceae;g__UBA10364;s__UBA10364 sp003023665
pico550_pico550.49	88.94	2.41	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__UBA3478;s__
pico551_pico551.013	89.74	4.2	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485
pico551_pico551.11	84.34	1.46	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__L_FER01;s__

pico551_pico551.14	85.16	1.79	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__
pico551_pico551.4	82.35	4.85	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Winogradskyella;s__Winogradskyella sp003335675
pico552_pico552.021	90	2.59	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudohongiellaceae;g__UBA9145;s__
pico552_pico552.026	91.23	9.3	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Litoricolaceae;g__Litoricola;s__Litoricola sp002691485

Table S3. Proportional Similarity index for all dereplicated, high quality MAGs. Main type and subtype of PS index are calculated according to the equation in the Material and Methods section.

	PS_maintenance	Rank_transformed_PS_maintenance	PS_subtype	Rank_transformed_PS_subtype	Rank_product	g_s
pico127_pico127.002	0.423458459	56	0.42345846	69	3864	specialists
pico127_pico127.018	0.388136546	36	0.26391317	26	936	specialists
pico127_pico127.023	0.368573156	19	0.21413539	22	418	specialists
pico127_pico127.035	0.63576366	126	0.52752312	115	14490	potential_specialists
pico127_pico127.045	0.368421053	10	0.36842105	47	470	specialists
pico127_pico127.062	0.431132576	64	0.3110947	31	1984	specialists
pico127_pico127.12	0.368421053	11	0.15789474	2	22	specialists
pico127_pico127.14	0.374401623	30	0.19480945	15	450	specialists
pico127_pico127.15	0.396249494	41	0.18572318	13	533	specialists
pico127_pico127.16	0.452641622	74	0.45264162	86	6364	potential_specialists
pico127_pico127.19	0.369623126	21	0.2138458	21	441	specialists
pico127_pico127.25	0.562866455	106	0.51621983	109	11554	potential_specialists
pico127_pico127.37	0.425467381	58	0.42546738	71	4118	potential_specialists
pico127_pico127.47_sub	0.422113545	55	0.36605702	46	2530	specialists
pico127_pico127.51	0.370762565	26	0.35848922	45	1170	specialists
pico127_pico127.64	0.379665689	32	0.22604431	25	800	specialists
pico127_pico127.69	0.473121423	83	0.33280089	39	3237	specialists
pico131_pico131.002	0.436110027	65	0.43611003	77	5005	potential_specialists
pico131_pico131.013_sub	0.420230952	53	0.31395254	32	1696	specialists
pico131_pico131.022	0.561159425	104	0.56115943	123	12792	potential_specialists
pico131_pico131.024	0.507118015	94	0.50711802	108	10152	potential_specialists
pico131_pico131.047	0.379613634	31	0.27880432	28	868	specialists
pico131_pico131.048	0.569082781	108	0.43072985	76	8208	potential_specialists
pico131_pico131.067	0.53950108	99	0.53950108	117	11583	potential_specialists

pico131_pico131.1	0.368421053	12	0.15789474	3	36	specialists
pico131_pico131.11	0.399710489	44	0.39971049	60	2640	specialists
pico131_pico131.18	0.374148902	29	0.19081107	14	406	specialists
pico131_pico131.28	0.477135272	87	0.47713527	103	8961	potential_specialists
pico131_pico131.30	0.369948435	23	0.2088105	19	437	specialists
pico131_pico131.50	0.369439545	20	0.17008161	9	180	specialists
pico131_pico131.52	0.650916648	135	0.58219334	128	17280	potential_specialists
pico131_pico131.78	0.501790492	91	0.42672819	73	6643	potential_specialists
pico131_pico131.84	0.400183247	45	0.40018325	61	2745	specialists
pico132_pico132.003	0.430605081	62	0.43060508	74	4588	potential_specialists
pico132_pico132.014	0.557862425	102	0.55786243	121	12342	potential_specialists
pico132_pico132.019	0.370491173	24	0.17545465	12	288	specialists
pico132_pico132.023	0.37049483	25	0.17157076	10	250	specialists
pico132_pico132.046	0.373272922	28	0.21114827	20	560	specialists
pico132_pico132.049	0.372330492	27	0.35660097	44	1188	specialists
pico132_pico132.21	0.369896297	22	0.19913859	16	352	specialists
pico132_pico132.47	0.475200469	85	0.47520047	100	8500	potential_specialists
pico132_pico132.53	0.631578947	120	0.57501467	126	15120	potential_specialists
pico132_pico132.55	0.531840462	97	0.53184046	116	11252	potential_specialists
pico132_pico132.58	0.400614931	46	0.40061493	62	2852	specialists
pico132_pico132.76	0.410929366	50	0.41092937	65	3250	specialists
pico132_pico132.80	0.368421053	13	0.15789474	4	52	specialists
pico132_pico132.86	0.368421053	14	0.15789474	5	70	specialists
pico132_pico132.91	0.368421053	15	0.15789474	6	90	specialists
pico244_pico244.021	0.583159757	111	0.58315976	130	14430	potential_specialists
pico244_pico244.22	0.877163298	190	0.81938119	192	36480	generalists
pico244_pico244.27	0.822616536	180	0.54485225	119	21420	generalists
pico244_pico244.28	0.736842105	151	0.52308377	112	16912	potential_specialists
pico244_pico244.31	0.368421053	16	0.31578947	33	528	specialists
pico245_pico245.104	0.729576935	150	0.72957694	169	25350	generalists
pico245_pico245.20	0.399304254	43	0.39930425	59	2537	specialists
pico245_pico245.24	0.848327247	184	0.80004032	184	33856	generalists
pico245_pico245.34	0.411401648	51	0.41140165	66	3366	specialists

pico245_pico245.36	0.785388171	164	0.73923476	171	28044	generalists
pico245_pico245.46	0.873322775	189	0.87332278	196	37044	generalists
pico245_pico245.8	0.368421053	17	0.36842105	48	816	specialists
pico247_pico247.26	0.430238065	61	0.32984125	38	2318	specialists
pico247_pico247.32	0.786770663	165	0.74447828	172	28380	generalists
pico247_pico247.38	0.708718616	148	0.60702685	138	20424	generalists
pico247_pico247.42	0.736842105	152	0.5169276	111	16872	potential_specialists
pico247_pico247.55_sub	0.859885256	186	0.64950016	148	27528	generalists
pico264_pico264.0007	0.583828953	112	0.58382895	132	14784	potential_specialists
pico264_pico264.0017	0.631578947	116	0.54933814	120	13920	potential_specialists
pico264_pico264.0021	0.742954057	156	0.69973034	163	25428	generalists
pico264_pico264.10	0.424617986	57	0.42461799	70	3990	specialists
pico264_pico264.35	0.690546661	147	0.65778919	151	22197	generalists
pico264_pico264.40	0.453082912	75	0.45308291	87	6525	potential_specialists
pico264_pico264.45	0.444382305	70	0.44438231	81	5670	potential_specialists
pico264_pico264.5	0.390515182	38	0.39051518	54	2052	specialists
pico265_pico265.0027	0.38236677	33	0.33500781	40	1320	specialists
pico265_pico265.0031_sub	0.552413487	101	0.45518174	90	9090	potential_specialists
pico265_pico265.033	0.85593742	185	0.71532754	166	30710	generalists
pico265_pico265.26	0.446349242	71	0.44634924	82	5822	potential_specialists
pico265_pico265.33	0.80560775	175	0.80560775	186	32550	generalists
pico265_pico265.47	0.443322504	69	0.4433225	80	5520	potential_specialists
pico265_pico265.51	0.39330795	39	0.39330795	55	2145	specialists
pico266_pico266.012	0.497613626	90	0.39715618	58	5220	potential_specialists
pico266_pico266.021_sub	0.51642883	95	0.51642883	110	10450	potential_specialists
pico266_pico266.106	0.477251097	88	0.46489989	95	8360	potential_specialists
pico266_pico266.12	0.744500514	157	0.58240823	129	20253	generalists
pico266_pico266.18	0.446894707	72	0.44689471	83	5976	potential_specialists
pico266_pico266.21	0.467877787	81	0.46787779	96	7776	potential_specialists
pico266_pico266.24	0.43068744	63	0.43068744	75	4725	potential_specialists
pico266_pico266.27	0.560495195	103	0.5604952	122	12566	potential_specialists
pico266_pico266.28	0.416246309	52	0.41624631	67	3484	specialists
pico266_pico266.29	0.790423224	168	0.79042322	180	30240	generalists

pico266_pico266.3	0.395619079	40	0.39561908	57	2280	specialists
pico266_pico266.38	0.403594818	49	0.40359482	64	3136	specialists
pico266_pico266.6	0.421489444	54	0.3944986	56	3024	specialists
pico281_pico281.012	0.562446661	105	0.45012795	84	8820	potential_specialists
pico281_pico281.019	0.681197855	146	0.68119786	160	23360	generalists
pico281_pico281.021	0.784992584	163	0.58052541	127	20701	generalists
pico281_pico281.039	0.263157895	1	0.15789474	7	7	specialists
pico281_pico281.049	0.649566718	134	0.63816212	145	19430	generalists
pico281_pico281.066_sub	0.637376654	127	0.58376149	131	16637	potential_specialists
pico281_pico281.069_sub	0.661892478	139	0.62372096	142	19738	generalists
pico281_pico281.072	0.803858908	173	0.67049961	155	26815	generalists
pico281_pico281.22_sub	0.631578947	117	0.56390397	124	14508	potential_specialists
pico281_pico281.42	0.796773633	171	0.79677363	183	31293	generalists
pico281_pico281.43	0.451883531	73	0.35607904	43	3139	specialists
pico281_pico281.48_sub	0.453392708	76	0.45339271	88	6688	potential_specialists
pico281_pico281.5	0.263157895	2	0.15789474	8	16	specialists
pico281_pico281.58	0.456705756	78	0.45670576	92	7176	potential_specialists
pico281_pico281.6	0.551484855	100	0.37016243	50	5000	potential_specialists
pico281_pico281.7	0.640098372	130	0.58441921	134	17420	potential_specialists
pico281_pico281.8	0.631578947	118	0.31578947	34	4012	potential_specialists
pico282_pico282.004	0.6403223	131	0.6403223	147	19257	generalists
pico282_pico282.006	0.671593237	142	0.67159324	156	22152	generalists
pico282_pico282.020_sub	0.828916846	182	0.8168228	189	34398	generalists
pico282_pico282.025_sub	0.663334117	140	0.6325504	143	20020	generalists
pico282_pico282.20	0.664713248	141	0.66471325	153	21573	generalists
pico282_pico282.22	0.45686551	79	0.45686551	93	7347	potential_specialists
pico282_pico282.23	0.291370205	6	0.29137021	29	174	specialists
pico282_pico282.27	0.677997507	145	0.67799751	159	23055	generalists
pico282_pico282.34	0.493896425	89	0.49389643	106	9434	potential_specialists
pico282_pico282.37	0.635662941	125	0.45667272	91	11375	potential_specialists
pico282_pico282.44	0.814183258	177	0.69188627	162	28674	generalists
pico282_pico282.45	0.632700831	122	0.5434132	118	14396	potential_specialists
pico282_pico282.49	0.278869412	4	0.22598072	24	96	specialists

pico282_pico282.53	0.871388376	188	0.81458879	188	35344	generalists
pico282_pico282.54	0.777679987	161	0.77461966	176	28336	generalists
pico282_pico282.56	0.909869328	196	0.84599718	194	38024	generalists
pico282_pico282.57	0.573632863	110	0.37437544	51	5610	potential_specialists
pico282_pico282.69	0.585079555	113	0.58507956	135	15255	potential_specialists
pico284_pico284.003	0.271301532	3	0.22113624	23	69	specialists
pico284_pico284.010_sub	0.652653101	137	0.6526531	150	20550	generalists
pico284_pico284.016_sub	0.809456719	176	0.73818279	170	29920	generalists
pico284_pico284.018_sub	0.564506002	107	0.45206288	85	9095	potential_specialists
pico284_pico284.039_sub	0.794761335	170	0.61182428	139	23630	generalists
pico284_pico284.124	0.537085871	98	0.31578947	35	3430	specialists
pico284_pico284.22	0.59071919	114	0.59071919	137	15618	potential_specialists
pico284_pico284.26	0.76364092	160	0.70299222	164	26240	generalists
pico284_pico284.28	0.786840161	166	0.74531885	173	28718	generalists
pico284_pico284.3	0.503421902	93	0.49997918	107	9951	potential_specialists
pico284_pico284.33	0.641764445	132	0.58603094	136	17952	generalists
pico284_pico284.35	0.279677509	5	0.17441435	11	55	specialists
pico284_pico284.38	0.633017633	123	0.48839066	105	12915	potential_specialists
pico284_pico284.39	0.296261077	8	0.27554128	27	216	specialists
pico284_pico284.42	0.403237172	48	0.29797401	30	1440	specialists
pico284_pico284.45	0.76130285	159	0.61406967	140	22260	generalists
pico284_pico284.66	0.29618981	7	0.20305605	17	119	specialists
pico284_pico284.84	0.45740534	80	0.35214218	42	3360	specialists
pico304_pico304.021	0.882871728	191	0.78105113	178	33998	generalists
pico304_pico304.038	0.82235419	179	0.81857435	191	34189	generalists
pico304_pico304.21	0.903779827	194	0.88643037	197	38218	generalists
pico304_pico304.24	0.736842105	154	0.52631579	114	17556	generalists
pico304_pico304.46	0.634059936	124	0.62078295	141	17484	generalists
pico497_pico497.020_sub	0.468014967	82	0.46801497	97	7954	potential_specialists
pico497_pico497.11	0.398377457	42	0.20406253	18	756	specialists
pico497_pico497.113	0.790461853	169	0.75942498	174	29406	generalists
pico497_pico497.23	0.475211368	86	0.47521137	101	8686	potential_specialists
pico497_pico497.24	0.440497975	67	0.14141051	1	67	specialists

pico497_pico497.29	0.777987443	162	0.68858344	161	26082	generalists
pico497_pico497.30	0.425662299	59	0.32651485	37	2183	specialists
pico497_pico497.32	0.390184283	37	0.34259029	41	1517	specialists
pico497_pico497.34	0.368421053	18	0.36842105	49	882	specialists
pico497_pico497.38	0.571407805	109	0.47611046	102	11118	potential_specialists
pico497_pico497.41	0.627042251	115	0.58438548	133	15295	potential_specialists
pico497_pico497.43	0.825211701	181	0.8252117	193	34933	generalists
pico497_pico497.44	0.639132273	128	0.47278698	98	12544	potential_specialists
pico497_pico497.45	0.517694348	96	0.46370467	94	9024	potential_specialists
pico497_pico497.50	0.401854744	47	0.40185474	63	2961	specialists
pico497_pico497.58	0.425943775	60	0.42594378	72	4320	potential_specialists
pico497_pico497.6	0.436242542	66	0.43624254	78	5148	potential_specialists
pico497_pico497.62	0.631710406	121	0.57427787	125	15125	potential_specialists
pico539_pico539.006	0.454219914	77	0.45421991	89	6853	potential_specialists
pico539_pico539.009	0.675638785	144	0.67563879	158	22752	generalists
pico539_pico539.16	0.631578947	119	0.48242034	104	12376	potential_specialists
pico539_pico539.2_sub	0.440732288	68	0.44073229	79	5372	potential_specialists
pico539_pico539.23	0.904921777	195	0.85020243	195	38025	generalists
pico539_pico539.35	0.658872813	138	0.65887281	152	20976	generalists
pico539_pico539.8	0.64862534	133	0.63402808	144	19152	generalists
pico540_pico540.003	0.639564838	129	0.63956484	146	18834	generalists
pico540_pico540.010	0.316447668	9	0.31644767	36	324	specialists
pico540_pico540.012	0.675085314	143	0.67508531	157	22451	generalists
pico540_pico540.027_sub	0.723747451	149	0.72374745	167	24883	generalists
pico540_pico540.18	0.651685597	136	0.6516856	149	20264	generalists
pico540_pico540.27	0.817286762	178	0.81728676	190	33820	generalists
pico540_pico540.28	0.865326377	187	0.80893206	187	34969	generalists
pico540_pico540.38	0.739865506	155	0.72652938	168	26040	generalists
pico540_pico540.5	0.501966447	92	0.42111139	68	6256	potential_specialists
pico550_pico550.028	0.895886411	193	0.79539277	182	35126	generalists
pico550_pico550.037	0.757818404	158	0.66745176	154	24332	generalists
pico550_pico550.14	0.843786929	183	0.80327018	185	33855	generalists
pico550_pico550.28	0.473403806	84	0.47340381	99	8316	potential_specialists

pico550_pico550. 45	0.7883071 11	167	0.77279984	175	29225	generalists
pico550_pico550. 49	0.3860925 89	35	0.38609259	53	1855	specialists
pico551_pico551. 013	0.8846800 99	192	0.78746679	179	34368	generalists
pico551_pico551. 11	0.8046699 56	174	0.70581874	165	28710	generalists
pico551_pico551. 14	0.7368421 05	153	0.52631579	113	17289	potential_sp ecialists
pico551_pico551. 4	0.8013391 11	172	0.77648755	177	30444	generalists
pico552_pico552. 021	0.3844007 25	34	0.38440073	52	1768	specialists
pico552_pico552. 026	0.9191910 39	197	0.7931623	181	35657	generalists
