Contribution Evenness: A functional redundancy metric sensitive to trait stability in microbial communities

Taylor M. Royalty (1,2) and Andrew D. Steen (2,1)*

University of Tennessee Departments of (1) Earth and Planetary Sciences and (2) Microbiology

Running title:

Measuring the functional redundancy of microbiomes

*Corresponding: asteen1@utk.edu

Abstract

The concept of functional redundancy has received considerable attention in both the macroecology and microbial ecology literature. As a result, multiple metrics of functional redundancy have been proposed. These vary in how they weight trait levels, species abundance, functional richness, and species richness. Here we present a new functional redundancy metric tailored for community-aggregated traits, traits that quantified at the community level and are quantitatively partitioned among species. We call this metric contribution evenness (CE) because it measures how evenly species contribute to a community-aggregated trait. As CE is an evenness measurement, it ranges from 0 and 1, where 0 corresponds to a single species contributing to a community-aggregated trait and 1 corresponds to all species contributing equally. Using *in silico* extinction simulations, we demonstrated that CE reflects the resilience of an ecosystem function to species extinction, a hypothesized ecological consequence of functional redundancy. As a positive control and to illustrate how CE can be used with sequence data, we analyzed the functional redundancy of eight nitrogen-transforming pathways using 2,631 metagenome-assembled genomes from 47 marine sites. CE for marine nitrogen cycle marker genes was consistent with our conceptual understanding of which nitrogen pathways are more redundant in the ocean. We found that, on average, the NH_4^+ assimilation pathway was the most functionally redundant (0.44 ± 0.08) while dissimilatory nitrate reduction was the least redundant (0.005 ± 0.005). As demonstrated here, CE provides a promising framework for measuring trait redundancy resilience in microbiomes.

Introduction

Recent advances in sequencing technology and sampling coverage have facilitated the discovery that important ecological functions are frequently replicated across co-occurring species in an ecosystem (1). This phenomenon is referred to as "functional redundancy". The ecological implications of functional redundancy are not fully understood, but one hypothesized consequence of functional redundancy is increased function stability in a community. Numerous theoretical studies predict that traits held by multiple species are more resilient to species extinction, a property called "trait resilience" (2–5). In order to test hypotheses about the ecological or biogeochemical implications of functional redundancy, a quantitative and generally applicable metric is required.

Several metrics for functional redundancy exist in the literature. The diversity of metrics reflects the diverse sources of data used to calculate functional redundancy and the diverse purposes for how this concept has been applied. Functional redundancy has been measured as the ratio of community richness to the theoretical maximum number of functional entities (6), median metabolic overlap of genomic traits (7), expected variability in trait space among community members (8), the degree of covariance in functional diversity and biodiversity (e.g., Galand et al., 2018; Micheli & Halpern, 2005; Miki, Yokokawa, & Matsui, 2013), and the simple fraction of populations in a community that bear a given function (12–14). These different metrics treat community structure differently. For instance, some metrics are sensitive to species abundances (8, 9), richness (6, 10), or neither (7). The functional redundancy metrics above (except for the simple fraction of taxa bearing a function) represent ecological function in terms of niche, where a niche is thought of as a point in a hyperdimensional trait space (15).

53 Communities with higher taxa density in trait space are interpreted as being more functionally
54 redundant.

55 However, the large number of potential environmentally-relevant microbial functions and
56 the low phylogenetic conservation of genes for those functions often makes it impossible to
57 define niches and functional guilds in a meaningful and tractable way (1). Thus, it can be
58 challenging to apply existing functional redundancy metrics to microbial ecosystems. An
59 alternative approach is to evaluate functional redundancy with respect to individual functions
60 rather than with respect to niches. Here, we propose a community metric designed to quantify the
61 functional redundancy of individual functions in a microbial community, something that is
62 currently lacking in the literature. We focus on microbial traits represented as nucleic acid
63 sequences (DNA and RNA), i.e., gene and transcript sequences. These sequence traits are
64 regularly measured in microbial ecology studies, used to infer microbial metabolism, easily
65 compared in metacommunity analysis, and are easily integrated into trait-based models (16).

66 One approach to measure community-level processes is with community-aggregated
67 parameters. Community-aggregated parameters describe ecosystem processes observed on the
68 community level. The contribution of each species to an ecosystem-level process is calculated as
69 the contribution of individuals to the process, weighted by their species' abundance. Species
70 contributions are then summed to reflect the community-level processes (17). Here, we propose
71 leveraging community-aggregated parameters to measure community functional redundancy.
72 Although community-aggregated parameters are applicable to a diverse suite of traits (e.g.,
73 respiration, metabolite production, etc.), we opt to narrow the focus of this work to nucleic acid-
74 based traits—genes and transcripts. Such traits are widely used in the microbial community
75 literature as proxies for biogeochemistry and are easily comparable in metacommunity analysis

(18, 19). Using species contributions, we can calculate the relative contribution by species to an observed community-level processes with an evenness measurement (20). Briefly, the "true diversity" of contributors is calculated using species contributions and traditional diversity metrics (21, 22). The observed diversity is normalized to a theoretical maximum diversity (richness) and made "unrelated", in the sense used by Chao and Ricotta (20), to community richness. Our approach can be understood as functional redundancy because it measures how evenly species in a community contribute to an ecosystem process—the more evenly species contribute to an ecosystem process, the more functionally redundant the community is with respect to that process. Accordingly, we term our measure of functional redundancy "contribution evenness" (CE).

To demonstrate the efficacy of CE, we first performed simulations to optimize the value of a constant called diversity order to use in CE calculations (21, 22). Second, we performed an *in silico* experiment to test how well CE captures trait resilience compared to other, existing functional redundancy metrics. Finally, we demonstrate the applicability of CE with real-world data by analyzing functional redundancy for eight nitrogen-transforming pathways found in metagenomic data (23) in *TARA* Oceans draft metagenome-assembled genomes (MAGs) (24). Overall, we found CE appealing because it is easily calculated from conventional shotgun genomic sequence data sets, provides insights on the relative contribution by community members to community-level processes, and is sensitive to the resilience of traits to species extinction.

Methods and Theory

Definition of Functional Redundancy

98 Here, we define a microbial community, C, with S species, where the ith species has a
99 relative abundance, p_i. (For simplicity, we refer to ecological units as species, although it would
100 be possible to perform this analysis at other taxonomic levels – strains, say, or genera – or even
101 to perform the analysis at the level of guilds in order to evaluate resource competition among
102 guilds, e.g., sulfate-reducers competing with methanogens for hydrogen electrons). We therefore
103 represent C as:

104 $$C = \{p | p_i > 0 \; \forall \; p_i; \sum_{i=1}^{S} p_i = 1\} \quad (1$$

105 where p is the set of relative abundances of species, such that any given relative abundance for
106 the ith species is greater than zero and all species relative abundances sum to one. Equation 1
107 represents a traditional abundance distribution (25).

108 Our goal was to calculate how evenly different species in a community, C (equation 1),
109 contribute to a community-aggregated parameter, or ecosystem function, T, at a fixed time and
110 space. To do this, we first determined the relative contribution, τ_i, that the ith species makes
111 towards T such that:

112 $$\tau_i = \frac{p_i t_i}{\sum_{i=1}^{n} p_i t_i} \quad 2)$$

113 where t_i is the species-specific contribution level and p_i is the relative abundance the ith species in
114 C. In equation 2, τ_i represents the relative contribution that the ith species contributes to a
115 community-aggregated parameter. We note that the functional analysis and preprocessing of
116 metagenomes and transcriptomes varies in the literature and public databases. For example, the
117 *TARA Oceans* metagenomes have been analyzed as assembled, un-binned metagenomes, where
118 individual genes were assigned taxonomy, annotation, and abundances (26), and as metagenome-

119 assembled genomes (24, 27). Although the fundamental meaning behind τ does not change
120 across these different datasets, the *in silico* approach to calculating τ may slightly differ
121 depending on how sequence data is preprocessed. For un-binned metagenomes, τ_i simply
122 corresponds to the proportion of reads associated with the ith species which map to function, T.
123 Similarly for transcriptomes, τ_i simply corresponds to the proportion of transcripts expressed by
124 the ith species which map to function, T. Assuming the community sequence spaces has been
125 randomly sampled, these scenarios implicitly account for the number of gene copies (t) and the
126 relative abundance (p) of the ith species. In the case of metagenome-assembled genomes, relative
127 contribution level can be calculated with equation 2 such that t_i is gene copies and abundance is
128 measured as the relative proportion of reads mapping to single-copy marker genes for the ith
129 species.

130 As an example, imagine that we have a metagenome that has been binned into MAGs.
131 Our hypothetical community has three species: a Cyanobacterium, Alphaproteobacterium, and a
132 Gammaproteobacterium. Mapping metagenome reads to single-copy marker genes reveals that
133 the Cyanobacteria, the Alphaproteobacterium and Gammaproteobacterium each represent 50%,
134 25%, and 25% of the community, respectively. We wish to calculate the functional redundancy
135 of this community with respect to the production of a secondary metabolite. We use the presence
136 of a gene, *xxxA*, as a proxy for an organisms' potential to produce this secondary metabolite. The
137 MAGs reveal that the Cyanobacterium and Alphaproteobacterium contain 5 copies of *xxxA* each,
138 while the Gammaproteobacterium lacks the *xxxA* gene. Thus, the relative contribution from each
139 species to secondary metabolite production, when using *xxxA* as a proxy, is τ=0.67, τ=0.33, and
140 τ=0, respectively. Using these relative contribution values, we proceed to calculate how evenly
141 species contribute towards the secondary metabolite production.

To do this, we reformulate the abundance distribution for community, C, as an abundance distribution with respect to τ (relative contribution distribution), such that:

$$C_{\tau,all} = \left\{\tau \middle| \tau_i \geq 0 \; \forall \; \tau_i; \sum_{i=1}^{n} \tau_j = 1\right\} \quad (3.1$$

and

$$C_{\tau} = \left\{\tau \in C_{\tau,all} \middle| \tau_i > 0 \; \forall \; \tau_i; \sum_{i=1}^{\sigma} \tau_i = 1\right\} \quad (3.2$$

where again, τ_i is the relative contribution by the ith species to T, such that all τ sum to equal 1. The community represented by equation 3.1 includes species not contributing to the community-aggregated parameter. The community represented by equation 3.2 is a subset of the one represented by equation 3.1, where C_τ contains only species from $C_{\tau,all}$ contributing towards a community-aggregated parameter, which we refer to as "contributors." As such, we adopted the variable, σ, which represents the richness of contributors. The effective richness of contributors towards T was calculated using equation 3.2 and the Tsallis entropy index (21) such that:

$$D_q = \begin{cases} \left(\sum_{i=1}^{\sigma} {\tau_i}^q\right)^{\frac{1}{(1-q)}} & q \neq 1 \\ exp\left(\sum_{i=1}^{\sigma} -\tau_i \log \tau_i\right) & q = 1 \end{cases} \quad (4$$

where τ_i is the relative contribution by i^{th} species in C_τ towards T, and D_q is the effective richness of contributors, and q is the diversity order, which weights the importance of species contributing more or less towards T when calculating D_q. This is analogous to diversity order weighting the importance of rare versus abundant species when measuring diversity of species abundances (22). With diversity order q=0, equation 4 simplifies to the number of contributors in equation 3.2 (i.e., σ). Diversity order q=0 removes any influence in the variability of τ when calculating D_q. Increasing the diversity order subsequently decreases the perceived contribution of

contributors with smaller τ. Diversity orders q=1 and q=2 correspond to Shannon's index. and Simpson's index.

We then normalize D_q by community richness S, which is equal to Ds if every species were to contribute to T equally), and adjust evenness to account for the functional dependence that the lower evenness bound has on community richness (20). As such, functional redundancy spans the full range from 0 to 1, such that:

$$CE = \frac{D_q - 1}{S - 1} \quad (5$$

Here, functional redundancy, *CE*, indicates how evenly species contribute to a trait, *T*. For the hypothetical community described above, the effective richness (D_q, q=0.5—see *Results*) of species possessing *xxxA* was ~1.55, and the functional redundancy, with respect to *xxxA*, was *CE*=0.47 (q=0.5; see *Results*).

Simulations of Functional Redundancy as a Function of Resilience to Species Extinction

In order to test the fidelity with which CE measures resilience of a community-aggregated parameter, we simulated artificial communities with varying values of species richness (*S*), richness of contributors (σ), and contribution levels by contributors (*t)*. First, the richness of contributors was randomly chosen from a uniformly distributed range of 1 to 100. Next, the richness of species not contributing to the community-aggregated parameter was randomly chosen from a uniformly distributed range of 0 to 100, so that some artificial communities were dominated by contributors while others by non-contributors. The relative abundance, *p*, of individual species was modeled log-normally (28, 29) such that:

$$R(r) = R_0 e^{-a^2 r^2} \quad (6$$

183 where R was the rth rank, R_0 was the maximum abundance, and a was the inverse width of the
184 lognormal distribution. Values of a were randomly selected according to a uniform distribution
185 from 0 (even) to 0.4 (more uneven), at 0.01 intervals. After establishing species relative
186 abundances, we randomly chose σ contributing taxa from our lognormally distributed
187 community. The contribution level, t, was then randomly assigned to these species from a
188 uniformly distributed range of 0.01 to 1, at intervals of 0.01. We generated and calculated R for
189 1,000 artificial communities.

190 The resilience of a community-aggregated parameter to extinction was tested by a
191 simulation in which resilience to extinction was quantified as the average fraction of species
192 richness (S) that must go extinct in order to decrease community-aggregated parameter by 99%.
193 Species were randomly removed from the community until summed τ was reduced to 1% of its
194 original level. This choice of threshold allowed for the evaluation of communities where all
195 species were contributors. In such communities, all species would have to go extinct to remove
196 100% of the ecosystem function. We think that it is also justified ecologically, because dormant
197 microbes in the rare biosphere can serve as a deep reservoir of traits, and it is not clear that fully
198 removing a trait from an ecosystem realistically happens (30). The extinction simulation was
199 performed 100 times for each of the individual 1,000 artificial communities. The resilience to
200 extinction generated from the 100 simulations were then averaged for each of the individual
201 1,000 artificial communities. We made no *a priori* assumptions about the optimal diversity order
202 (i.e., q) to use in equations 4 and 5. We performed the simulation described above varying
203 diversity orders from 0 to 2 at 0.1 intervals. The best diversity order was determined by fitting
204 generalized additive model (GAM) regressions between R and resilience to extinction, and then
205 choosing the diversity order which resulted in a GAM fit with the lowest mean absolute error

among residuals. All GAM fits were performed using the gam function from the R package mgcv (31). Last, we performed the same simulations and calculated *R* using other methods described in the literature. A summary of these measurements is provided in Table 1. An R package which calculates functional redundancy (equations 1-5) and performs these simulations is available at https://github.com/taylorroyalty/funfunfun. *A Case Study of Functional Redundancy in the Global Ocean Nitrogen Cycle*

Draft metagenome-assembled genomes (MAGs) from Tully, Graham et al. (2018) were downloaded from NCBI (32). Data on metagenome sequencing effort (number of short reads), MAG relative abundance estimates (pre-corrected for genome size), ocean depths, and ocean regions, defined in (24), were downloaded from Figshare (https://doi.org/10.6084/m9.figshare.5188273). MAG relative abundances from each site were multiplied by the respective site's sequencing effort to convert relative abundances into abundance count data. Abundance counts for all sites were normalized by the lowest abundance (greater than 0) among all sites. MAG abundances were then rarefied using the rrarefy function from the R package vegan (33). We chose a sampling effort of 14,265, as this sampling effort corresponded to the site with lowest sampling effort (i.e., sequencing effort). We tabulated nitrogen-cycle marker genes in individual MAGs by blasting MAG proteins against nitrogen-cycle marker gene databases using the BLASTP algorithm (*E*-value=10^{-10}) (34). Nitrogen cycle marker genes were categorized into eight categories: N_2 fixation (*nifH*), nitrification (*amoA*, *hao*), NO_3 reduction (*napA*, *narG*), dissimilatory NO_3 reduction (*nrfA*), assimilatory NO_3/NO_2 reduction (*narB*, *nasA*, *nasB*, and *nirB*), NH_4^+ assimilation (*gdhA*, *glnA*), N remineralization (*gltB*), and denitrification (*nirK*, *nirS*, *norB*, and *nosZ*) (23). Individual BLAST databases were constructed for each nitrogen cycle category. Gene names associated with individual categories

229 were searched for in the Swiss-Prot database (35). All proteins matching gene names were
230 subsequently downloaded and used in BLAST database construction. Functional redundancy was
231 then calculated for individual stations, at individual depths, for each nitrogen cycle category
232 using equations 1-5. We used a diversity order, q=0.5, based on the GAM sensitivity analysis.
233 Table 2 contains protein names, gene names, corresponding EC, and entry name for all genes
234 used in the above analysis.

235 **Results and Discussion**

236 *Representing microbial communities as species abundance, trait abundance, and relative*
237 *contribution distributions*

238 CE represents functional redundancy using two distributions: the distribution of species
239 abundances within a community, and the distribution of contribution levels among species. Here,
240 we evaluate genomic traits because these are regularly adopted as proxies for microbially-
241 mediated biogeochemistry (18, 26, 36–40). Because trait levels are measured as non-negative,
242 real numbers (17), the relative contribution of individual species to a community-aggregated
243 parameter can be represented as a relative abundance distribution (Figure 1). We determined the
244 relative contribution of individual species to the community-aggregated parameter from the
245 product of species relative abundances and their respective trait abundance distributions
246 (equation 2). We refer to the distribution of the relative contribution of each species to total as
247 the relative contribution distribution (equation 3). Thus, an abundant species that expresses a
248 small amount of trait per cell, and a rarer species that expresses a large amount of trait per cell,
249 can contribute equally with respect to a community-aggregated parameter (Figure 1).

250 *Defining, Quantifying Functional Redundancy, and Choosing a Diversity Order q*

We measured functional redundancy as how evenly species in a community contribute to a community-aggregated parameter and termed this metric as contribution evenness (CE). This definition is consistent with descriptions of functional redundancy as multiple species sharing similar ecological roles or functions (2, 3). The relative contribution by species towards a community aggregated-parameter (equation 2) provides a framework to calculate the effective number, or richness, of species that contribute to that community-aggregated parameter as an ecological function. This effective richness is calculated with traditional diversity indices (equations 3-4) (21, 22). To determine the effective proportion of a community contributing to the community-aggregated parameter, the richness of contributors is normalized by species richness (22). Last, to make meaningful comparisons of evenness between communities with different levels of richness, we made our evenness metric unrelated to community richness (equation 5) (20).

Numerous diversity indices have been proposed in the literature. Three common measures of diversity include species richness, Shannon's diversity index, and Simpson's diversity index, are manifestations of equation 4 when the q term—commonly referred to as the diversity order—is q=0, q=1, and q=2, respectively. The diversity order changes the relative weight placed on rare versus common species when determining the effective richness. We found that GAM fits between functional redundancy and resilience to extinction followed positive, monotonic functions for all diversity orders spanning q=0 to q=2. However, at $q \approx 0.5$, CE was most tightly related to resilience to extinction, our benchmark for functional redundancy efficacy (Figure 2; also compare Fig 3a to 3b). Therefore, we suggest that q=0.5 makes *CE* the best indicator of how resilient traits are to species extinction. In that case, equation 4 simplifies to:

$$D_q = \left(\sum_{i=1}^{\sigma} \sqrt{\tau_i}\right)^2 \quad (7$$

and CE becomes:

$$CE = \frac{\left(\sum_{i=1}^{\sigma} \sqrt{\tau_i}\right)^2 - 1}{S-1} \quad (8$$

This is the formulation of CE that we recommend for general use.

Functional Redundancy and Trait Resilience to Species Extinction

A tangible ecological consequence of functional redundancy is resilience of an ecosystem function to the extinction of species performing that function. When functions are more broadly distributed across species, a function is less likely to disappear due to the extinction of any subset of species. This interpretation is often highlighted in macro-ecological studies (2–4, 8, 41–43). The conceptual relationship between functional redundancy and resilience of an ecosystem function provides an opportunity to demonstrate the fidelity of CE. To do this, CE was calculated for randomly generated communities. Individual species were randomly assigned relative contribution levels towards a community-aggregated parameter. We simulated random species extinction events in the artificial communities until >99% of the original ecosystem function was removed from the community (see *Methods*).

It is important to emphasize that these simulations were not designed to mimic how extinction occurs in nature. Instead, these simulations empirically demonstrate the degree to which an ecosystem function is concentrated in species. The more widely an ecosystem function is distributed across species, the more species must go extinct before the ecosystem function is lost. We observed a positive, monotonic relationship between the resilience to random species extinction events and CE (Figure 3A). When CE was ~0, resilience to species extinction was

295 ~0.5, meaning ~50% of the community, on average, went extinct before an ecosystem function
296 was removed from the community. This is equivalent to only one species performing the
297 ecosystem function. When CE was ~1, resilience to species extinct was ~1, meaning that 100%
298 of the community, on average, had to go extinct to completely remove a function from an
299 ecosystem. In general, the relationship between CE and trait resilience to extinction resembled a
300 saturation curve—although we opted to use GAMs when calculating mean absolute error
301 (previous section) as we made no assumption about the true relationship between CE and
302 function resilience, other than it should be a positive, monotonic function.

303 Here, we recommend the form of CE described in equation 8, as it is optimized for our
304 intended use case, which is measuring relative resilience of traits to species extinctions within
305 communities. However, we can imagine other applications of the CE concept, for which other
306 values of q may be appropriate. Even for species diversity (i.e., equation 4), there is still
307 discussion over the merits of using Simpson's Index (q=2), Shannon's Index (q=1), and richness
308 (q=0) in different situations (44–46). As such, it is important to be explicit about the implications
309 of behind choosing a trait and the diversity order.

310 *CE Measures Resilience Differently than Other Functional Redundancy Metrics*

311 Several methods for evaluating community-level functional redundancy exist in the
312 literature, as detailed in Table 1. While Table 1 certainly does not include all published metrics,
313 the highlighted metrics reflect the diversity in how traits, community properties (i.e., richness
314 and evenness), as well as the desired metric utility (by design) vary among different functional
315 redundancy metrics. Two of the metrics evaluate niche overlap, where niches are defined in
316 multivariate trait space (7, 8). In the case of Ricotta et al. (2016), functional redundancy weights
317 species abundance, leading to a measure of the expected niche dissimilarity from the mean

community niche. In contrast, Hester et al. (2019) evaluates the median metabolic overlap which considers neither species abundance nor trait levels. One metric evaluates niche occupancy, where niches are defined in multivariate trait space (6). In this approach, functional redundancy is taken to be as the ratio of species richness to plausible number niches (unique combinations of trait states), called functional entities. The last metric is informal but has been utilized by various authors. In this approach, functional redundancy is simply interpreted as the proportion of species in a community with a trait (e.g., 12–14).

To highlight the difference between our metric and the others described in Table 1, we repeated the extinction simulations and compared all estimates of functional redundancy with trait resilience to species extinction (Figure 3B-E). In these simulations, only one trait was considered when calculating functional redundancy with all metrics. For the method presented in Mouillot et al. (2014), we assumed there were two functional entities (trait absence and presence). Generally, all functional redundancy metrics trended differently to resilience to species extinction. This observation was not surprising, as most of these methods evaluate niche overlap in some fashion. Among the metrics in Table 1, the fraction of the community possessing a trait behaved most like CE (Figure 3B). We note that the fraction of the community possessing a trait is a special case of CE, with diversity order q=0. Lastly, one inherent limitation of CE is that it only evaluates single community-aggregated parameters at a time. To address functional redundancy with respect to multiple traits perform dimensional reduction (e.g., PCA) on multiple traits and use the resultant factors as trait levels in equation 2. Factors in this case would be interpreted as different niches.

Case Study: Redundancy of Nitrogen Cycle Marker Genes in Marine Microbiomes

340 To demonstrate a real application and provide a positive control for CE results, we
341 measured CE in the marine nitrogen cycle using genomic data from the *TARA Oceans* project
342 (47). Recently, (24) published 2,361 draft metagenome-assembled genomes (MAGs) which
343 represent genomic content from seven oceanographic provinces, three different depths, and four
344 filter size fractions. Relative abundance data used here were reported in (24) and were
345 determined by the proportion of sequence reads mapping to assembled MAGs. For simplicity,
346 we limited our analysis to only MAGs identified in the bacterial size fraction (<3 μm), which
347 corresponds to free-living microbes. The nitrogen cycle was simplified to eight pathways (23)
348 and included N_2 fixation, nitrification, NO_3^- reduction, dissimilatory NO_3^- reduction, assimilatory
349 NO_3^-/NO_2^- reduction, NH_4^+ assimilation, N remineralization, and denitrification. A general
350 summary of the marker genes (downloaded from Swiss-Prot; see Methods) used for quantifying
351 nitrogen cycle pathways in MAGs is summarized in Table 2.

352 Overall, the functional redundancy for individual genomic traits was consistent with our
353 conceptual understanding of primary nitrogen-transforming pathways in the ocean (Figure 4). A
354 clear example was the low functional redundancy in the nitrification pathway. We expected this
355 observation as nitrification metabolisms in oceans are primarily restricted to *Crenarchaeota* (48).
356 Similarly, the N_2 fixation pathway, which is predominantly facilitated by a small group of
357 *Cyanobacteria*, exhibits low functional redundancy. Conversely, the uptake of inorganic N via
358 NH_4^+ and NO_3^-/NO_2^- assimilation are known to be broadly expressed traits (23), and accordingly
359 exhibit high CE. Last, the low functional redundancy in dissimilatory NO_3^- reduction and NO_3^-
360 reduction were not surprising considering these pathways occur in anaerobic and low-oxygen
361 environments, respectively (23).

362 Some variance existed across oceanographic provinces and depth for intra-pathway
363 functional redundancy. However, inter-pathway variance in functional redundancy was generally
364 larger than intra-pathway variance (Figure 4). This suggests that the marine nitrogen cycle
365 marker genes considered here are consistently distributed among communities for MAGs
366 retrieved in Earth's ocean at middle and lower latitudes. This was consistent with the lack of
367 relationship between the metagenome companion metadata (e.g., O_2, inorganic N, PO_4^{3-}, latitude,
368 longitude, etc.) and functional redundancy. The low intra-pathway variance in functional
369 redundancy was not particularly surprising, as saltwater biomes exhibit the lowest alpha-
370 diversity and small beta-diversity among free-living microbiomes (47, 49). Minor shifts in
371 community composition among lower taxonomic ranks (e.g., genus or species) will not
372 drastically alter redundancy in genomic traits as changes in genomic composition is buffered by
373 traits conserved at higher taxonomic ranks (e.g., phylum level) (50). This observation was in
374 agreement with an earlier analysis evaluating cluster of orthologous groups (COGs) composition
375 across *TARA Oceans* metagenomes (47).

376 The disproportionate number of lineages that lack cultured representatives highlights the
377 continued importance of nucleotide sequencing in microbial ecology and the interference of
378 microbial function (51). Here, we demonstrated that CE provides a framework to evaluate the
379 functional redundancy for these traits in a novel way. This approach represents an easy-to-apply,
380 quantitative metric for functional redundancy that faithfully represents the stability of traits to
381 extinctions of species within a community.

382 **Data Availability**

In order to facilitate the use of CE functional redundancy measurements, we created an R package to calculate CE functional redundancy and run extinction simulations. The package is available at https://github.com/taylorroyalty/funfunfun.

Acknowledgments

We gratefully acknowledge the comments of three anonymous reviewers, whose constructive comments substantially improved the manuscript. Funding for this project was provided by a C-DEBI graduate fellowship to TMR and by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0020369 for TMR and ADS. This is C-DEBI contribution number [to be assigned upon manuscript acceptance].

Author Contribution

TMR designed the research, performed research, analyzed data, developed the R package, and wrote the paper. ADS designed the research, developed the R package, and wrote the paper.

397

398 **References**

399 1. Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, Ackermann M,
400 Hahn AS, Srivastava DS, Crowe SA, Doebeli M, Parfrey LW. 2018. Function and
401 functional redundancy in microbial systems. Nat Ecol Evol 2:936–943.

402 2. Naeem S. 1998. Species Redundancy and Ecosystem Reliability 12:39–45.

403 3. Rosenfeld JS, Mall M. 2002. Functional redundancy in ecology and conservation. OIKOS
404 1.

405 4. Fonseca CR, Ganade G. 2001. Species Functional Redundancy, Random Extinctions and
406 the Stability of Ecosystems. J Ecol 89:118–125.

407 5. Cottingham KL, Brown BL, Lennon JT. 2001. Biodiversity may regulate the temporal
408 variability of ecological systems. Ecol Lett 4:72–85.

409 6. Mouillot D, Villéger S, Parravicini V, Kulbicki M, Arias-González JE, Bender M,
410 Chabanet P, Floeter SR, Friedlander A, Vigliola L, Bellwood DR. 2014. Functional over-
411 redundancy and high functional vulnerability in global fish faunas on tropical reefs. Proc
412 Natl Acad Sci U S A 111:13757–13762.

413 7. Hester ER, Jetten MSM, Welte CU, Lücker S. 2019. Metabolic overlap in
414 environmentally diverse microbial communities. Front Genet 10:1–10.

415 8. Ricotta C, de Bello F, Moretti M, Caccianiga M, Cerabolini BEL, Pavoine S. 2016.
416 Measuring the functional redundancy of biological communities: a quantitative guide.
417 Methods Ecol Evol 7:1386–1395.

418 9. Galand PE, Pereira O, Hochart C, Auguet JC, Debroas D. 2018. A strong link between

419 marine microbial community composition and function challenges the idea of functional
420 redundancy. ISME J 12:2470–2478.

421 10. Micheli F, Halpern BS. 2005. Low functional redundancy in coastal marine assemblages.
422 Ecol Lett 8:391–400.

423 11. Miki T, Yokokawa T, Matsui K. 2013. Biodiversity and multifunctionality in a microbial
424 community: A novel theoretical approach to quantify functional redundancy. Proc R Soc
425 B Biol Sci 281.

426 12. Tully BJ, Wheat CG, Glazer BT, Huber JA. 2018. A dynamic microbial community with
427 high functional redundancy inhabits the cold, oxic subseafloor aquifer. ISME J 12:1–16.

428 13. Wohl DL, Arora S, Gladstone JR. 2004. Functional Redundancy Supports Biodiversity
429 and Ecosystem Function in a Closed and Constant Environment. Ecology 85:1534–1540.

430 14. Trivedi CB, Stamps BW, Lau GE, Grasby SE, Templeton AS, Spear JR. 2020. Microbial
431 Metabolic Redundancy Is a Key Mechanism in a Sulfur-Rich Glacial Ecosystem.
432 mSystems 5:1–16.

433 15. Chase JM, Leibold MA. 2003. Ecological Niches, 1st ed. The University of Chicago
434 Press, Chicago.

435 16. Green JL, Bohannan BJM, Whitaker RJ. 2008. Microbial biogeography: From taxonomy
436 to traits. Science (80-) 320:1039–1043.

437 17. Violle C, Navas M, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E, Oikos S, May
438 N, Violle C, Navas M, Vile D, Kazakou E, Fortunel C. 2007. Let the concept of trait be
439 functional! Oikos 116:882–892.

440 18. Krause S, Le Roux X, Niklaus PA, van Bodegom PM, Lennon T. JT, Bertilsson S,
441 Grossart HP, Philippot L, Bodelier PLE. 2014. Trait-based approaches for understanding
442 microbial biodiversity and ecosystem functioning. Front Microbiol 5:1–10.

443 19. Fierer N, Barberán A, Laughlin DC. 2014. Seeing the forest for the genes: Using
444 metagenomics to infer the aggregated traits of microbial communities. Front Microbiol
445 5:1–6.

446 20. Chao A, Ricotta C. 2019. Quantifying evenness and linking it to diversity , beta diversity ,
447 and similarity 100:1–15.

448 21. Jost L. 2006. Entropy and diversity. Oikos 2.

449 22. Hill MO. 1973. Diversity and Evenness : A Unifying Notation and Its Consequences.
450 Ecology 54:427–432.

451 23. Pajares S, Ramos R. 2019. Processes and Microorganisms Involved in the Marine
452 Nitrogen Cycle: Knowledge and Gaps. Front Mar Sci 6.

453 24. Tully BJ, Graham ED, Heidelberg JF. 2018. The reconstruction of 2,631 draft
454 metagenome-assembled genomes from the global oceans. Sci Data 5:1–8.

455 25. McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, Dornelas M,
456 Enquist BJ, Green JL, He F, Hurlbert AH, Magurran AE, Marquet PA, Maurer BA,
457 Ostling A, Soykan CU, Ugland KI, White EP. 2007. Species abundance distributions:
458 Moving beyond single prediction theories to integration within an ecological framework.
459 Ecol Lett 10:995–1015.

460 26. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh HJ, Cuenca M, Field CM,

461 Coelho LP, Cruaud C, Engelen S, Gregory AC, Labadie K, Marec C, Pelletier E, Royo-
462 Llonch M, Roux S, Sánchez P, Uehara H, Zayed AA, Zeller G, Carmichael M, Dimier C,
463 Ferland J, Kandels S, Picheral M, Pisarev S, Poulain J, Acinas SG, Babin M, Bork P, Boss
464 E, Bowler C, Cochrane G, de Vargas C, Follows M, Gorsky G, Grimsley N, Guidi L,
465 Hingamp P, Iudicone D, Jaillon O, Kandels-Lewis S, Karp-Boss L, Karsenti E, Not F,
466 Ogata H, Pesant S, Poulton N, Raes J, Sardet C, Speich S, Stemmann L, Sullivan MB,
467 Sunagawa S, Wincker P. 2019. Gene Expression Changes and Community Turnover
468 Differentially Shape the Global Ocean Metatranscriptome. Cell 179:1068-1083.e21.

469 27. Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, MacLellan SL, Lücker
470 S, Eren AM. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are
471 abundant in surface ocean metagenomes. Nat Microbiol 3:804–813.

472 28. Magurran AE. 1988. Ecological Diversity and Its Measurement, 1st ed. Croom Helm Ltd.

473 29. Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. 2009. Ecology of the rare
474 microbial biosphere of the Arctic Ocean. Proc Natl Acad Sci 106:22427–22432.

475 30. Lennon JT, Jones SE. 2011. Microbial seed banks: The ecological and evolutionary
476 implications of dormancy. Nat Rev Microbiol 9:119–130.

477 31. Wood S. 2017. mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML
478 Smoothness Estimation.

479 32. Federhen S. 2012. The NCBI Taxonomy database. Nucleic Acids Res 40:136–143.

480 33. Philip D. 2003. Computer program review VEGAN , a package of R functions for
481 community ecology. J Veg Sci 14:927–930.

482 34. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2019. GenBank. Nucleic
483 Acids Res 47:D94–D99.

484 35. Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its
485 supplement TrEMBL in 2000. Nucleic Acids Res 28:45–48.

486 36. Falkowski PG, Fenchel T, Delong EF. 2008. The Microbial Engines That Drive Earth's
487 Biogeochemical Cycles. Science (80-) 320:1034–1039.

488 37. Malik AA, Martiny JBH, Brodie EL, Martiny AC, Treseder KK, Allison SD. 2019.
489 Defining trait-based microbial strategies with consequences for soil carbon cycling under
490 climate change. ISME J 445866.

491 38. Baker BJ, Lazar CS, Teske AP, Dick GJ. 2015. Genomic resolution of linkages in carbon,
492 nitrogen, and sulfur cycling among widespread estuary sediment bacteria. Microbiome
493 3:14.

494 39. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC,
495 Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF.
496 2016. Thousands of microbial genomes shed light on interconnected biogeochemical
497 processes in an aquifer system. Nat Commun 7:13219.

498 40. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M,
499 Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J,
500 Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A,
501 Wiltshire KH, Glöckner FO, Schweder T, Amann R. 2012. Substrate-Controlled
502 Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom.
503 Science (80-) 336:608–611.

504 41. McPherson JM, Yeager LA, Baum JK. 2018. A simulation tool to scrutinise the behaviour
505 of functional diversity metrics. Methods Ecol Evol 9:200–206.

506 42. Tilman D. 2001. Functional Diversity. Encycl Biodivers Second Ed 3:587–596.

507 43. Tilman D, Reich PB, Knops JMH. 2006. Biodiversity and ecosystem stability in a decade-
508 long grassland experiment. Nature 441:629–632.

509 44. Patil GP, Taillie C. 1982. Diversity as a concept and its measurement. J Am Stat Assoc
510 77:548–561.

511 45. Jost L. 2007. Partitioning Diversity into Independent Alpha and Beta Components.
512 Ecology 88:2427–2439.

513 46. Morris EK, Caruso T, Buscot F, Fischer M, Hancock C, Maier TS, Meiners T, Müller C,
514 Obermaier E, Prati D, Socher SA, Sonnemann I, Wäschke N, Wubet T, Wurst S, Rillig
515 MC. 2014. Choosing and using diversity indices: insights for ecological applications from
516 the German Biodiversity Exploratories. Ecol Evol 4:3514–24.

517 47. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B,
518 Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, D'Ovidio F,
519 Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-
520 Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmento H, Vieira-Silva S, Dimier C,
521 Picheral M, Searson S, Kandels-Lewis S, Boss E, Follows M, Karp-Boss L, Krzic U,
522 Reynaud EG, Sardet C, Sieracki M, Velayoudon D, Bowler C, De Vargas C, Gorsky G,
523 Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S,
524 Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG,
525 Bork P. 2015. Structure and function of the global ocean microbiome. Science (80-)

526 348:1–10.

527 48. Wuchter C, Abbas B, Coolen MJL, Herfort L, Van Bleijswijk J, Timmers P, Strous M,
528 Teira E, Herndl GJ, Middelburg JJ, Schouten S, Damsté JSS. 2006. Archaeal nitrification
529 in the ocean. Proc Natl Acad Sci U S A 103:12317–12322.

530 49. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi
531 A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-
532 Baeza Y, González A, Morton JT, Mirarab S, Xu ZZ, Jiang L, Haroon MF, Kanbar J, Zhu
533 Q, Song SJ, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM,
534 Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A,
535 Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R,
536 Agosto Rivera JL, Al-Moosawi L, Alverdy J, Amato KR, Andras J, Angenent LT,
537 Antonopoulos DA, Apprill A, Armitage D, Ballantine K, Bárta J, Baum JK, Berry A,
538 Bhatnagar A, Bhatnagar M, Biddle JF, Bittner L, Boldgiv B, Bottos E, Boyer DM, Braun
539 J, Brazelton W, Brearley FQ, Campbell AH, Caporaso JG, Cardona C, Carroll JL, Cary
540 SC, Casper BB, Charles TC, Chu H, Claar DC, Clark RG, Clayton JB, Clemente JC,
541 Cochran A, Coleman ML, Collins G, Colwell RR, Contreras M, Crary BB, Creer S,
542 Cristol DA, Crump BC, Cui D, Daly SE, Davalos L, Dawson RD, Defazio J, Delsuc F,
543 Dionisi HM, Dominguez-Bello MG, Dowell R, Dubinsky EA, Dunn PO, Ercolini D,
544 Espinoza RE, Ezenwa V, Fenner N, Findlay HS, Fleming ID, Fogliano V, Forsman A,
545 Freeman C, Friedman ES, Galindo G, Garcia L, Garcia-Amado MA, Garshelis D, Gasser
546 RB, Gerdts G, Gibson MK, Gifford I, Gill RT, Giray T, Gittel A, Golyshin P, Gong D,
547 Grossart HP, Guyton K, Haig SJ, Hale V, Hall RS, Hallam SJ, Handley KM, Hasan NA,
548 Haydon SR, Hickman JE, Hidalgo G, Hofmockel KS, Hooker J, Hulth S, Hultman J, Hyde

549 E, Ibáñez-Álamo JD, Jastrow JD, Jex AR, Johnson LS, Johnston ER, Joseph S, Jurburg
550 SD, Jurelevicius D, Karlsson A, Karlsson R, Kauppinen S, Kellogg CTE, Kennedy SJ,
551 Kerkhof LJ, King GM, Kling GW, Koehler A V., Krezalek M, Kueneman J, Lamendella
552 R, Landon EM, Lanede Graaf K, LaRoche J, Larsen P, Laverock B, Lax S, Lentino M,
553 Levin II, Liancourt P, Liang W, Linz AM, Lipson DA, Liu Y, Lladser ME, Lozada M,
554 Spirito CM, MacCormack WP, MacRae-Crerar A, Magris M, Martín-Platero AM, Martín-
555 Vivaldi M, Martínez LM, Martínez-Bueno M, Marzinelli EM, Mason OU, Mayer GD,
556 McDevitt-Irwin JM, McDonald JE, McGuire KL, McMahon KD, McMinds R, Medina M,
557 Mendelson JR, Metcalf JL, Meyer F, Michelangeli F, Miller K, Mills DA, Minich J,
558 Mocali S, Moitinho-Silva L, Moore A, Morgan-Kiss RM, Munroe P, Myrold D, Neufeld
559 JD, Ni Y, Nicol GW, Nielsen S, Nissimov JI, Niu K, Nolan MJ, Noyce K, O'Brien SL,
560 Okamoto N, Orlando L, Castellano YO, Osuolale O, Oswald W, Parnell J, Peralta-
561 Sánchez JM, Petraitis P, Pfister C, Pilon-Smits E, Piombino P, Pointing SB, Pollock FJ,
562 Potter C, Prithiviraj B, Quince C, Rani A, Ranjan R, Rao S, Rees AP, Richardson M,
563 Riebesell U, Robinson C, Rockne KJ, Rodriguezl SM, Rohwer F, Roundstone W, Safran
564 RJ, Sangwan N, Sanz V, Schrenk M, Schrenzel MD, Scott NM, Seger RL, Seguinorlando
565 A, Seldin L, Seyler LM, Shakhsheer B, Sheets GM, Shen C, Shi Y, Shin H, Shogan BD,
566 Shutler D, Siegel J, Simmons S, Sjöling S, Smith DP, Soler JJ, Sperling M, Steinberg PD,
567 Stephens B, Stevens MA, Taghavi S, Tai V, Tait K, Tan CL, Taş N, Taylor DL, Thomas
568 T, Timling I, Turner BL, Urich T, Ursell LK, Van Der Lelie D, Van Treuren W, Van
569 Zwieten L, Vargas-Robles D, Thurber RV, Vitaglione P, Walker DA, Walters WA, Wang
570 S, Wang T, Weaver T, Webster NS, Wehrle B, Weisenhorn P, Weiss S, Werner JJ, West
571 K, Whitehead A, Whitehead SR, Whittingham LA, Willerslev E, Williams AE, Wood SA,

572 Woodhams DC, Yang Y, Zaneveld J, Zarraonaindia I, Zhang Q, Zhao H. 2017. A
573 communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457–463.

574 50. Royalty TM, Steen AD. 2019. Quantitatively Partitioning Microbial Genomic Traits
575 among Taxonomic Ranks across the Microbial Tree of Life. mSphere 4:e00446-19.

576 51. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. 2018. Phylogenetically Novel Uncultured
577 Microbial Cells Dominate Earth Microbiomes. mSystems 3:1–12.

578

Data Accessibility

We provided a function for calculating functional redundancy with CE and an example script at https://github.com/taylorroyalty/functional_redundancy_script.

Author Contribution

TMR designed the research, performed research, analyzed data, and wrote the paper. ADS designed the research, performed research, and wrote the paper.

585 **Tables**

586 **Table 1. A summary of metrics of community-level functional redundancy.**

Authors	Interpretation	Equation	Notes
Here	Effective proportion of species with a trait	$R = \frac{D_q - 1}{S - 1}$	• Accounts for species abundances • Continuous trait states • Individual traits • R spans 0 to 1
(8)	Expected trait dissimilarity among species in a community	$R = 1 - \frac{\sum_i p_i \sum_j p_j \delta_{ji}}{D_2}$; $\delta_{ji} = trait\ distance$; $D_2 =$ Taxonomic diversity order ($q = 2$)	• Accounts for species abundances • Continuous trait states • Individual or multivariate traits • Normalized by a maximum theoretical dissimilarity • R spans 0 to 1
(6)	Ratio of species to number of functional entities (states)	$R = \frac{S}{FE}$; FE = # of possible combinations of trait states for all traits under consideration	• Binary trait states • Individual or multivariate traits • R spans ~0 ($\lim_{FE \to \infty} \frac{S}{FE} = 0$) to S
Various (see text for examples)	Fraction of species with a trait	$R = \frac{v}{S}$	• Binary trait states • Individual traits • R spans 0 to 1
(7)	Median overlap of traits between species	$R = med(O)$; $O = \{o \mid o_{i,j} = \lvert g_i \cap g_j \rvert; i \neq j\}$; $g = taxa$; O = set containing # of overlaps among species	• Binary trait states • $\lvert g_i \cap g_j \rvert$represents the number of overlapping traits g between species i and species j • Individual and multivariate traits • R spans 0 to ∞

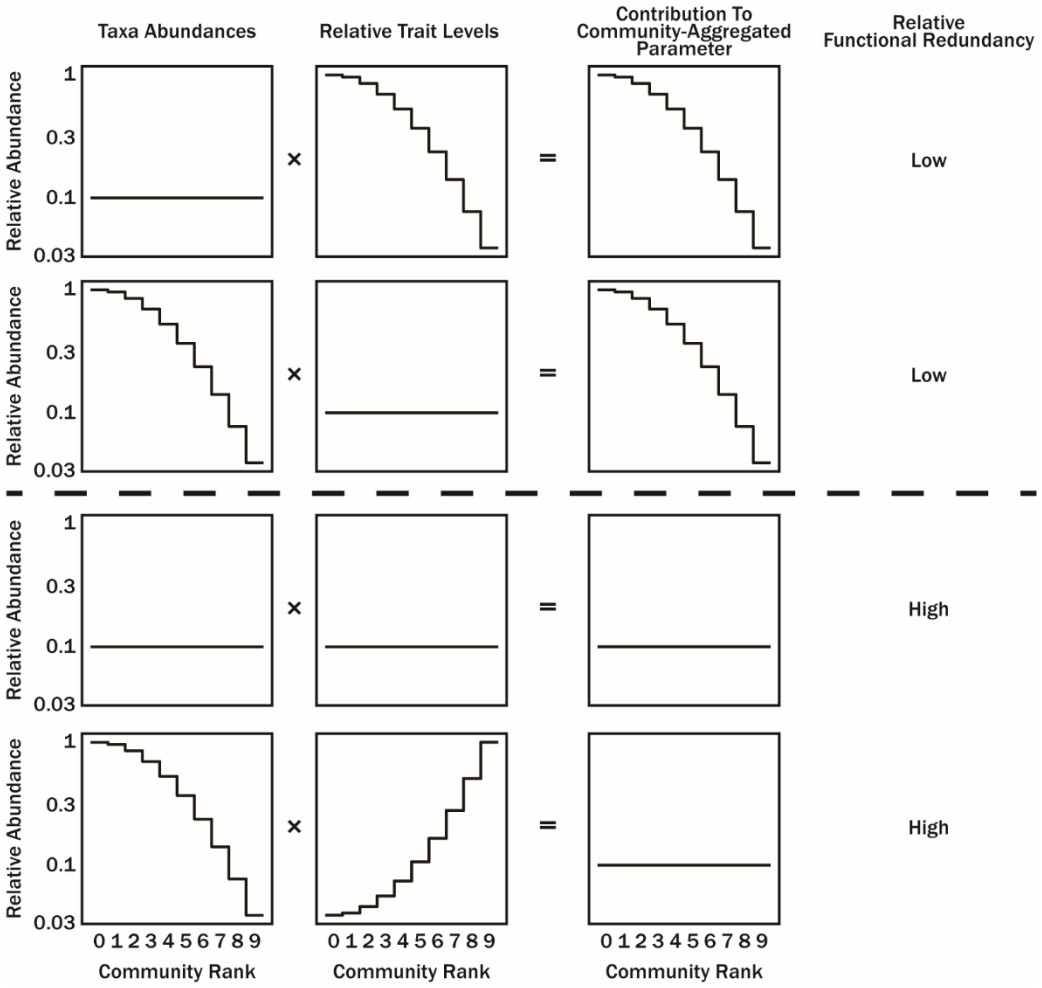
587

588 **Table 2. A summary of marker gene sequences used in the BLASTP analysis of *TARA***
589 ***Oceans* MAGs. All marker genes were downloaded from Swiss-Prot.**

Pathway	Gene Names	EC Number	# Marker Genes
N Remineralization	*gltB*	1.4.1.13	14
N Fixation	*nifH*	1.18.6.1	127
Nitrification	*amoA*	1.14.99.39	2
	hoa	1.7.2.6	1
NO_3^- Reduction	*napA*	1.9.6.1	136
	narG	1.7.5.1	13
Dissimilatory NO_3^- Reduction	*nrfA*	1.7.2.2	77
Assimilatory NO_3^-/NO_2^- Reduction	*narB*	1.7.5.1	3
	nasA	1.7.7.2	1
	nasB	1.7.1.4	3
	nirB	1.7.1.15	5
NH_4^+ Assimilation	*gdhA*	1.4.1.3	14
	glnA	6.3.1.2	173
Denitrification	*nirK*	1.7.2.1	7
	nirS	1.7.2.1; 1.7.99.1	4
	norB	1.7.2.5	2
	nosZ	1.7.2.4	11

590

591 **Figures**



592

593 **Figure 1.**

594 Community taxon abundance and relative contribution level are scalable and provide a measure
595 for the contribution that different species have towards a community-aggregated parameter. The
596 evenness of species-level contribution to the community-aggregated parameter is taken as the
597 functional redundancy.

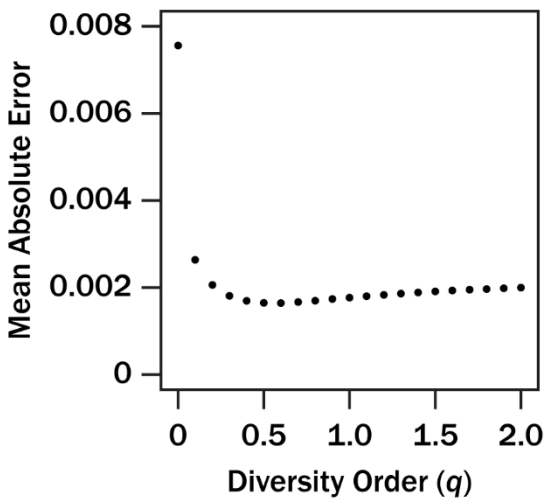


Figure 2.

The sensitivity of mean absolute error between functional redundancy and trait resilience to species extinction in response changing diversity order q in equation 4.

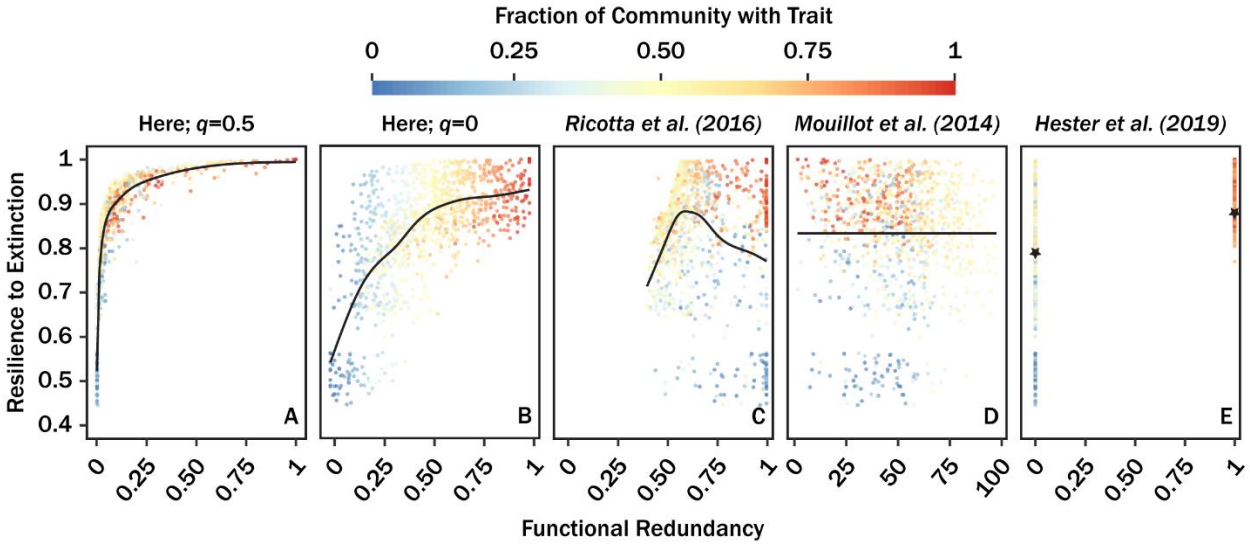


Figure 3.

A comparison of the relationship between selected different functional redundancy metrics and trait resilience (see methods). Functional redundancy calculated with equation 5 (q=0.5) (**A**), equation 5 (q=0) (**B**), methodology from Ricotta et al. (2016) (**C**), methodology from Mouillot et al. (2014) (**D**), and methodology from Hester et al. (2019). Black lines in panels **A-D** correspond to best fit GAMs. The stars in panel **E** correspond to the average trait resilience at R=0 and R=1. The color scale corresponds to the fraction of a community with a trait versus.

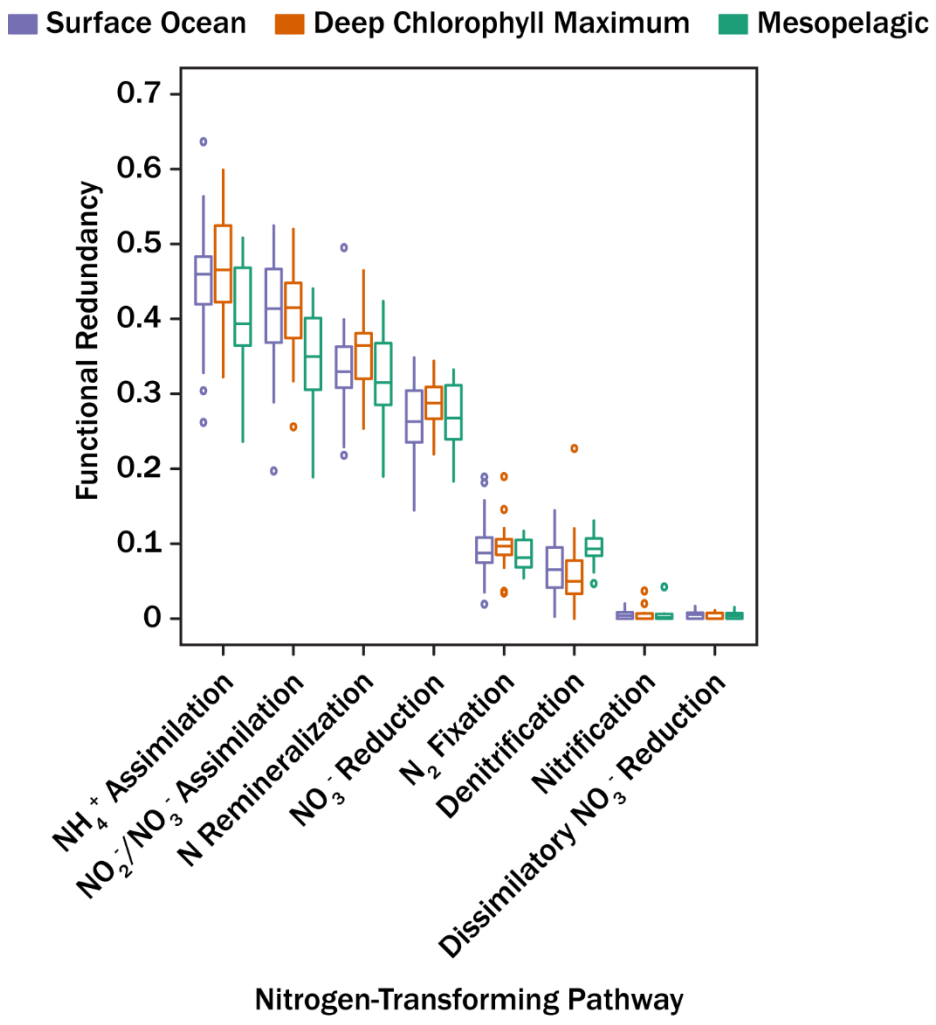


Figure 4.

Boxplots showing functional redundancy calculated for eight nitrogen-transforming pathways in *TARA Oceans* draft MAGs. Different colors correspond to sample depth. Whiskers correspond to 150% of the inter-quantile region.