Redox traits characterize the organization of global microbial communities

Salvador Ramírez-Flandes[a,b,c,1], Bernardo González[d,e], and Osvaldo Ulloa[a,b,1]

[a]Departamento de Oceanografía, Universidad de Concepción, 4070386 Concepción, Chile; [b]Instituto Milenio de Oceanografía, Universidad de Concepción, 4070386 Concepción, Chile; [c]Programa de Doctorado en Ingeniería de Sistemas Complejos, Universidad Adolfo Ibáñez, 7941169 Santiago, Chile; [d]Laboratorio de Bioingeniería, Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, 7941169 Santiago, Chile; and [e]Center of Applied Ecology and Sustainability (CAPES), 8331150 Santiago, Chile

Edited by Paul G. Falkowski, Rutgers, The State University of New Jersey, New Brunswick, NJ, and approved January 9, 2019 (received for review October 12, 2018)

The structure of biological communities is conventionally described as profiles of taxonomic units, whose ecological functions are assumed to be known or, at least, predictable. In environmental microbiology, however, the functions of a majority of microorganisms are unknown and expected to be highly dynamic and collectively redundant, obscuring the link between taxonomic structure and ecosystem functioning. Although genetic trait-based approaches at the community level might overcome this problem, no obvious choice of gene categories can be identified as appropriate descriptive units in a general ecological context. We used 247 microbial metagenomes from 18 biomes to determine which set of genes better characterizes the differences among biomes on the global scale. We show that profiles of oxidoreductase genes support the highest biome differentiation compared with profiles of other categories of enzymes, general protein-coding genes, transporter genes, and taxonomic gene markers. Based on oxidoreductases' description of microbial communities, the role of energetics in differentiation and particular ecosystem function of different biomes become readily apparent. We also show that taxonomic diversity is decoupled from functional diversity, e.g., grasslands and rhizospheres were the most diverse biomes in oxidoreductases but not in taxonomy. Considering that microbes underpin biogeochemical processes and nutrient recycling through oxidoreductases, this functional diversity should be relevant for a better understanding of the stability and conservation of biomes. Consequently, this approach might help to quantify the impact of environmental stressors on microbial ecosystems in the context of the global-scale biome crisis that our planet currently faces.

microbial ecology | functional traits | oxidoreductases | biomes | metagenomics

Biological communities are conventionally described as assemblages of species whose ecological roles are known or predictable from their observable morphological characteristics. In the early twentieth century, Lotka and Volterra pioneered the development of theoretical ecology using species numbers as the master variable in differential equations that describe the interactions and complexity of ecological systems (1). Since then, most theoretical ecologists have used species numbers as the ecological unit for developing an extensive body of theory, which includes elaborate mathematical models to explain the dynamics of populations and communities (1). In practice, this approach requires the categorization of every observed individual into a taxonomic unit —which is not a trivial task in some cases (2), and it is definitively a problem in microbial ecology (3–5). In the latter context, microbial ecologists face three main problems. First, observable morphological attributes do not provide sufficient discriminatory or functional characterization. Second, the isolation of microbial species to assess their physiology and ecological function is rarely possible, a phenomenon that is related to the so-called Great Plate Count Anomaly (6). Third, prokaryotic genomes are highly dynamic, mainly due to pervasive horizontal gene transfers and the effect of mobile DNA elements and phages (7). Microbial ecologists have employed molecular taxonomic markers, primarily the small subunit ribosomal RNA gene, to address the first and second problems, thereby operationally defining species and estimating their abundances and taxonomic diversity (8). This taxonomic approach has been used to explain and predict the microbial dynamics in diverse environments (9, 10). In such a context, the Earth Microbiome Project initiative has recently reported microbial taxonomic diversity per biome on a global scale with the use of standardized protocols to provide an organized and complete catalog of microbes (11). However, several studies have reported inconsistent taxonomical correlations under apparently similar ecological scenarios, finding better consistency only when using multiple protein-coding genes as traits and when the whole community is analyzed as the ecological unit (12–16). This has been performed in an attempt to address the third above-mentioned problem. After all, it is the function, not the taxonomic information, that has the actual ecological relevance (17). Unfortunately, the selection of the ecologically relevant categories of protein-coding genes for use is not evident in the broad context of planetary biomes (6, 18, 19). We analyzed 247 metagenomes from 18 biomes (Fig. 1) to tackle this issue and to determine under which specific nonexclusive set of genes the differences between biomes are the highest. These gene sets included protein-coding genes with associated orthology in the KEGG database (a typical approach in trait-based analyses), enzyme-coding genes, transporter-associated

Significance

Biological communities are conventionally described as assemblages of species, whose ecological roles are known or predictable from their observable morphology. In microbial ecology, such a taxonomic approach is hindered by limited capacity to discriminate among different microbes, which bear highly dynamic genomes and establish complex associations. Approaches based on culture-independent functional genes profiling might overcome these problems, but a set of usable established genes in a general situation is still lacking. We show that genes related to reduction-oxidation (redox) processes separate microbial communities into their corresponding biomes. This redox-based characterization is linked to the microbial energetics of ecosystems and to most biogeochemical cycles and might be useful for assessing the impact of environmental degradation on the ecosystem services, underpinned by microorganisms.

Author contributions: S.R.-F. designed research; S.R.-F. performed research; S.R.-F. analyzed data; S.R.-F., B.G., and O.U. wrote the paper; and B.G. and O.U. supervised research.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: sram@udec.cl or oulloa@udec.cl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1817554116/-/DCSupplemental.

Published online February 11, 2019.

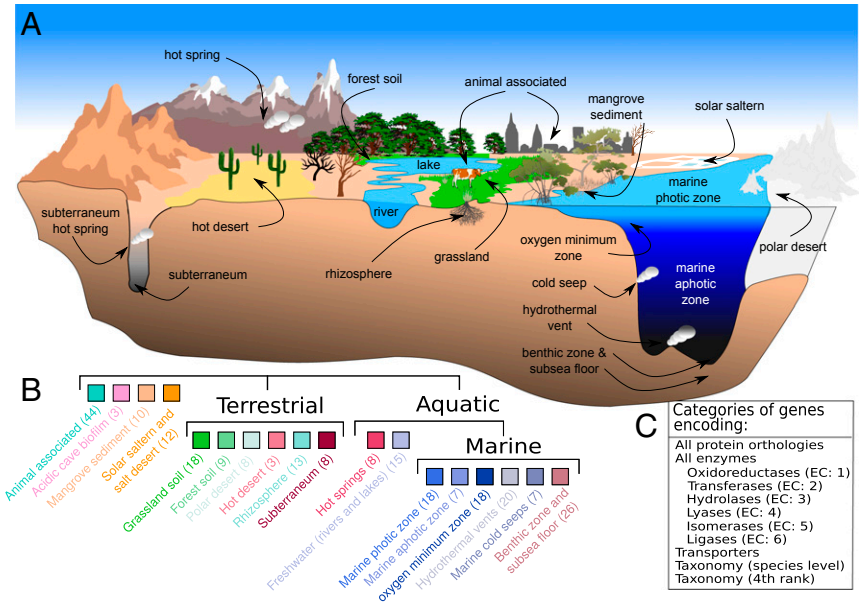

Fig. 1. Biomes and categories of genes. (*A*) Sketch of the biomes from which metagenomes (as proxies for microbial communities) were included in this paper. The animal-associated biome includes metagenomes from terrestrial animals only. A complete list and origin of these metagenomes can be found in the *SI Appendix*, Table S1. (*B*) Organized list of the biomes illustrated in *A*. The number of metagenomes per biome is shown in parentheses besides the biome name, which is displayed in the color code utilized in the rest of the figures. (*C*) Categories of gene profiles considered in the analyses. All protein orthologies refer to the protein orthologies present in the KEGG protein database. The fourth rank taxonomy typically corresponds to a phylum in the prokaryotic taxonomy (see *SI Appendix* for details).

genes, and taxonomic marker genes (Fig. 1). We found that the set of genes that were encoding enzymes better differentiated the biomes than the other gene categories. In particular, the profiles of genes that were encoding oxidoreductases composed the set with the highest cohesion and separation of biome groups, suggesting that they can better describe the association of the microbial communities to their respective biomes. In addition, we found no correspondence in biome maximum diversity between the functional and the taxonomic approaches. An oxidoreductase-based description of microbial communities also serves as a convenient proxy for an energetic description of ecosystems as these proteins are responsible for redox reactions, which are the processes by which every living organism uses energy from and modify the chemical characteristics of the environment (20).

Results and Discussion

From an ecological point of view, the functions of communities represent the most relevant information about an ecosystem. In microbial ecology, when these functions are fine grained to molecular processes through functional genes, it is natural to ask whether all of them have the same ecological relevance to differentiate one biome from another (Fig. 1). Our results show that redox functions supported the highest statistical differentiation among biomes when taxonomic and functional sets of genes were compared (Table 1 and *SI Appendix*, Fig. S5 and Table S2). The discriminatory power of oxidoreductase genes for grouping biomes can be visualized in networks of correlations using different gene categories with metagenomes as nodes (microbial communities, colored according to biome origin) and correlations as edges (Fig. 2*A* and *SI Appendix*, Figs. S3 and S4). Metagenomes from different biomes were more separated in the networks of oxidoreductases than in the network of taxonomic markers, which is the visual expression of the better cohesion and separation results as shown in Table 1. Hierarchical clusterings of these profiles (Fig. 2*B* and *SI Appendix*, Figs. S1 and S2) revealed the following three main groups of biomes: a group of apparent anoxic or suboxic biomes (animal associated, some hot springs, subterranean ecosystems, marine sediments, subseafloor,

Table 1. Different profiles of genes separating metagenomes into biomes groups

Category	PERMANOVA F statistic	Cohesion	Separation	Cohesion + separation	Profile size
All KEGG protein orthologies	10.66858	0.59033	−0.33468	0.25565	6,789
All enzymes	10.64720	0.69027	−0.43105	0.25922	1,826
Oxidoreductases	12.47645	0.69076	−0.42962	0.26113	484
Transferases	9.64637	0.76038	−0.52017	0.24021	541
Hydrolases	12.16139	0.70443	−0.45038	0.25405	423
Lyases	11.21940	0.75581	−0.51344	0.24237	211
Isomerases	7.80187	0.80306	−0.57657	0.22649	103
Ligases	9.97776	0.86361	−0.69899	0.16463	94
Transporters	7.23900	0.44584	−0.25863	0.18721	1,869
Taxonomy (species)	1.83457	0.13495	−0.06620	0.06875	4,011
Taxonomy (fourth rank)	2.81049	0.36668	−0.24392	0.12275	365

Different sets of profiles of relative gene abundances (with stabilized variances) were evaluated to determine under which of them the separation of metagenomes (microbial communities) into biome groups was most significant. The PERMANOVA statistical test (all P values < 0.001) indicates that gene profiles of oxidoreductases were the set of genes with higher statistically supported differences between the biomes (the higher the F statistic, the more likely to reject the null hypothesis of no differences among groups). The cohesion/separation of biomes groups confirmed that result (third, fourth, and fifth columns). The last column of this table indicates the total number of subcategories that each category displayed in this paper. For example, considering all of the datasets analyzed, we found 484 oxidoreductase genes and 4,011 species. To resolve the low differences among the F statistics of some enzyme categories, additional analyses were required, and their results can be found in *SI Appendix, Group Variances Analyses*, Fig. S3, and Table S2.

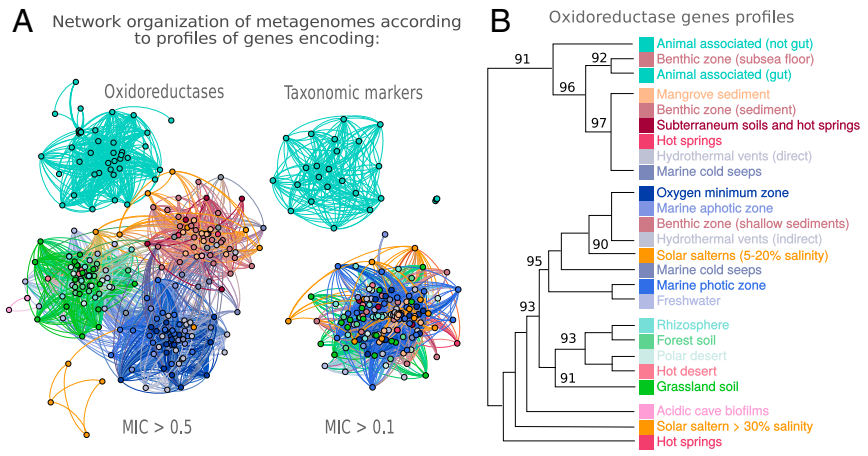

Fig. 2. Association of microbial metagenomes and biomes. (*A*) Network representation of the microbial metagenomes by profiles of oxidoreductase and taxonomic gene ranks. The nodes correspond to metagenomes, colored according to their biome of origin (Fig. 1). The edges represent maximal information coefficients (MICs). In the network associated with oxidoreductases (*Left*), all MIC ≥ 0.5 are shown. The network associated with taxonomic profiles (*Right*) was drawn with all MIC ≥ 0.1 as in this case, and these values were significantly lower. These taxonomic edge weights were increased by 0.4 to give visual balance to the plot. The differential clustering of biomes (colors) is explained by the values of cohesion and separation from Table 1 (see also *SI Appendix*, Figs. S3 and S4). (*B*) A simplified version (topology only and grouped per biomes) of the hierarchical clustering of the metagenomes based on oxidoreductase gene profiles (*SI Appendix*, Fig. S1). Support values higher than 90% are shown in the plot.

and mangrove sediments), a group of aquatic biomes (freshwater and different types of marine ecosystems), and a group of soil-associated biomes (grassland, forest, deserts, and rhizosphere). Note that environments associated with oxygen minimum zones did not cluster with the first above-mentioned group. The oxygen-limited condition shared by these ecosystems is not reflected in this clustering because the microorganisms in the pelagic low-oxygen environments mainly exploit chemolitoautotrophic metabolisms instead of the anaerobic degradation of organic matter that normally occurs in, for example, anoxic sediments. This analysis also showed that metagenomes from extreme ecosystems, such as acidic cave biofilms, some hot spring systems, and hypersaline environments, cluster outside of these three main groups.

The group of biomes with apparent anoxic conditions shared distinctive oxidoreductase genes related to methanogenesis, sulfide oxidation, denitrification, hydrogen oxidation, nitrogen fixation, and aromatic aldehydes oxidation (Fig. 3). The animal-associated metagenomes analyzed here were highly diverse, but most of them were related to the digestive systems of animals, making this group slightly biased toward the functional genes that are represented more in these microbial communities. Thus, the functions associated with these diverse biomes should be interpreted with care as it is unlikely that, for example, the human tongue dorsum supports microbial communities exploiting hydrogen oxidation processes. Indeed, hierarchical clusterings separated the microbial communities associated with the parts at the end of the digestive system of animals (cecum, gut, and stool) from other animal-associated metagenomes (human oral mucosa, tongue dorsum, supragingival plaque, anterior nares, and posterior fornix; *SI Appendix*, Fig. S2). Although the latter subgroup of microbial communities can also be associated with potentially anoxic microhabitats, the former subgroup was found to be functionally closer to the communities from the marine sediments and subsea-floor ecosystems, mainly because of the shared redox functionalities for the degradation of organic matter under anoxic conditions. Notably, gut-associated microbiomes displayed nitrogen fixation capabilities too (Fig. 3), which is consistent with the recent observations (21).

Marine microbial communities were best characterized by a group of oxidoreductases that includes dimethylglycine dehydrogenase, sarcosine oxidase, and choline dehydrogenase (Fig. 3). These enzymes are involved in the synthesis and degradation of glycine betaine, which is an effective and widely used compatible solute for coping with saline stress (22). Indeed, most algae and some invertebrates produce and accumulate glycine betaine as an intracellular osmolyte (22). Thus, marine microorganisms might take advantage of the availability of this substrate in seawater and can convert it to formate, which can then be used as an energy source or directed to one-carbon metabolism for biosynthesis (23). A direct precursor of glycine betaine is choline, which is also abundant in seawater, as it can represent up to 0.39% of the dry weight of algae (24). A distinctive oxidoreductase gene present in marine microbial communities was 3-hydroxyisobutyrate dehydrogenase, which has been found to play a role in amino acid catabolism (25), as a source of alternative substrates for respiration under metabolic stress situations. Another representative of oxidoreductase encoded in the metagenomes of these microbial communities is aldehyde dehydrogenase NAD+. Polyunsaturated aldehydes are commonly produced by diatoms as a chemical defense against grazers, and their concentrations in seawater can potentially affect the bacterial community structure and diversity (26).

Microbial communities associated with soil were mainly characterized by oxidoreductase genes related to the degradation of aromatic compounds for the carbon source [alcohol dehydrogenase cytochrome *c*, isoquinoline 1-oxidoreductase, catechol 2,3-dioxygenase, homogentisate 1,2-dioxygenase (27) and phenylacetyl-CoA 1,2 epoxidase (28) (Fig. 3)]. This representation might be explained by the fact that most primary production in soils is returned to the environment as detritus (29), which can be rich in aromatics as they constitute a significant part of lignin in higher plants (27). Genes encoding betaine aldehyde dehydrogenase were also distinctive in soil-associated microbial communities. This enzyme is involved in the biosynthesis of glycine betaine as a compatible solute for alkaline-saline stress (30). In fact, reports indicate that many soil environments are highly alkaline, and transient conditions, such as drought, can significantly increase the alkalinity within cells (31). Additionally, plant root exudates can change the soil chemistry, sometimes creating microhabitats of increased alkalinity (30). Thus, soil microbial communities seem to be genetically prepared to resist saline-alkaline stress by synthesizing their cellular defenses, unlike marine microbial communities that apparently rely more on the environmental availability of glycine betaine or its direct precursors, such as choline or sarcosine. Despite freshwater biome grouping with the marine biomes, its associated microbial communities still share similarities in the abundances of some oxidoreductase genes with the soil biomes, such as in the case of betaine aldehyde dehydrogenase, carbon monoxide dehydrogenase (acceptor), and stearoyl-CoA 9-desaturase (Fig. 3). This observation might be related to the results of a recent study that suggest that freshwater ecosystems might connect the otherwise separated microbial communities (32).

Although most biogeochemical processes are widely distributed across different environments (33), some oxidoreductase genes associated with these processes appear to be unimportant

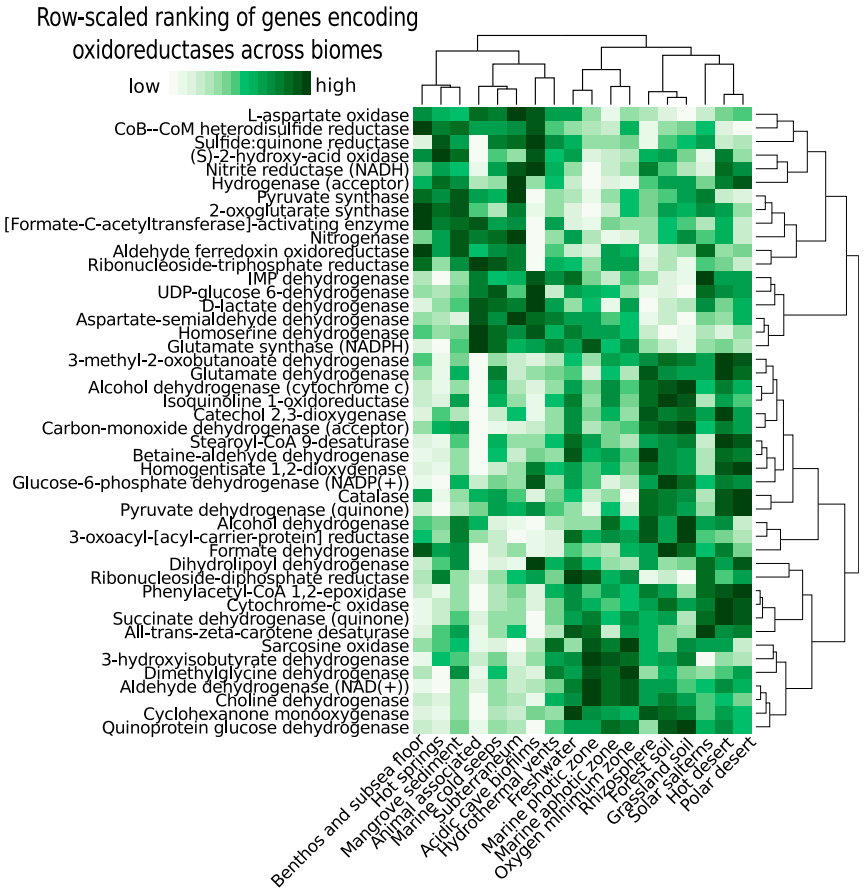

Fig. 3. Distinctive oxidoreductase genes associated with biomes. These genes were determined by statistically testing that the average of rankings of each oxidoreductase gene within each biome was significantly different from the average ranking in other biomes. Dark and light shades in this figure refer to relative abundances, high and low, respectively. Thus, the rankings for this figure were reversed as a low rank indicates high relative abundance. These values were scaled for better visualization, which means that color shades can only be compared horizontally. Some of these distinctive genes encode oxidoreductases considered associated with important biogeochemical and biochemical processes. For example, CoB-CoM heterodisulfide reductase (methanogenesis), sulfide:quinone reductase (sulfide oxidation), nitrite reductase NADH (denitrification), hydrogenase (hydrogen oxidation), nitrogenase (nitrogen fixation), and aldehyde ferredoxin oxidoreductase (aromatic aldehydes oxidation). Hierarchical clusterings using these values were calculated for convenient grouping of both biomes and oxidoreductase genes.

for soil and aquatic biomes. This apparent conflict can be explained by the fact that, frequently, the most abundant microbes in these environments are heterotrophs [e.g., members of *Acidobacteria* in soils (34) and SAR11 clade in the ocean (35)]. Thus, although nitrification, denitrification, sulfur oxidation, and carbon fixation also occur in terrestrial and aquatic ecosystems, their genetic markers are significantly less abundant than the oxidoreductase genes related to heterotrophic metabolisms (*SI Appendix*, Table S4). On the other hand, the biomes from the apparently anoxic group (typically harboring fewer heterotrophs) appeared prominently in many of these processes, such as, for example, methanogenesis, hydrogen oxidation, nitrogen fixation, sulfur oxidation, nitrification, and denitrification (Fig. 4). In addition, the oxidative phosphorylation process under suboxic conditions (associated with Cbb3 oxidase, encoded by the *ccoN* gene, Fig. 4) appeared to be best ranked in these biomes. Despite the pelagic low-oxygen marine biome was not clustered in this group of biomes (Figs. 2*B* and 3 and *SI Appendix*, Figs. S1 and S2), their metagenomes displayed high genetic representation associated with some of these processes, such as nitrification, denitrification, and sulfur oxidation (Fig. 4). This fact has been described as the beginning of a progressive rerouting of the energy flow into the microbial pathways as oxygen declines in marine ecosystems in detriment of higher trophic levels (36–38). Such progression ends in the extreme situation in which all benthic energy is processed as hydrogen sulfide (36) with concomitant accumulation of nitrite in the intermediate case of the anoxic marine zones (39). Low-oxygen areas in the ocean have rapidly expanded in the past decades, and they are expected to further increase as a consequence of global warming (36, 38). This, in turn, can be affected by the greenhouse gases that are emitted in marine low-oxygen zones as a by-product of anaerobic microbial pathways (36, 38, 39).

The extraordinary dispersal potential of microbes is usually expressed through the old tenet "everything is everywhere, but the environment selects," which a recent study extends to "every gene is everywhere, but the environment selects" (32). This fact suggests that measures of diversity for conducting large-scale studies of biomes in microbial ecology should include not only the richness, but also the evenness of the distribution of gene categories. By using the inverse Simpson index, we found that microbial taxonomic diversity does not correlate with microbial functional diversity. In our analysis, microbial communities from mangrove sediments were found to be the most taxonomically diverse (Fig. 5*A*). This result is consistent with findings of the recent studies that show that some sediment environments can be more diverse than soils (40), which, in turn, have been traditionally considered to be the ecosystems with the highest microbial diversity (41). However, regarding oxidoreductase genes, grassland soils and rhizospheres were found to be the most diverse biomes (Fig. 5*A*). This finding correlates with observations in plant diversity that suggest that, in the fine grain, grasslands are the most diverse soil biomes, harboring up to ∼90 different plant species per square meter (42). It is noteworthy that the temperate grasslands are currently among the biomes that face the highest ecological risk due to the extensive habitat loss and underprotection (43). To give a quantitative example of the microbial diversity of oxidoreductase genes in grasslands, consider that, on average, ∼130 of their most abundant categories were needed to cover the 70% of the total abundance of these genes. The same coverage percentage needed only ∼40 of the most abundant categories in the subterranean and acidic cave biofilms biomes (Fig. 5*B*).

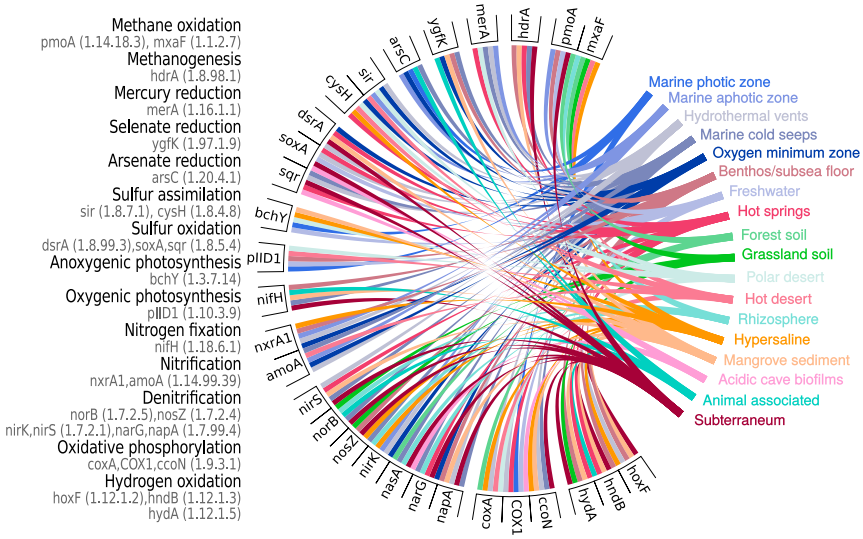

Fig. 4. Biogeochemically relevant processes per biome by oxidoreductase genes relative abundances. Oxidoreductase genes associated with biogeochemical processes and their top five biomes where they were ranked the highest. The biomes per genes are in clockwise order, starting from the biome where the gene was best ranked. For example, the dissimilatory sulfite reductase gene (*dsrA;* involved in sulfur oxidation and reduction) was found best ranked in the following biomes in this order: hydrothermal vents, subterranean habitats, mangrove sediments, hot springs, and oxygen minimum zones.

The choice of relevant variables is a critical step in the analysis of any complex system. In microbial ecology, the taxonomic structure of communities has typically been considered a proxy for the microbial ecosystem's functioning, even though it is often unable to resolve functional genetic traits (44). The need for alternative trait-based approaches has been claimed for years (45), but there has been no agreement on the selection of a relevant set of genes necessary for its practical application (6, 18, 19). In this paper, we evaluated different sets of genes for this purpose, finding that oxidoreductase genes are a convenient choice. The set of transporter genes also has this potential, but its power to differentiate biomes was found to be lower. This is most likely as these genes also suffer from significant redundancy (e.g., there are different transporters for the same substrate, depending on their affinities). Other groups of enzyme genes, such as those associated with hydrolases, also supported a proper separation of biomes (Table 1 and *SI Appendix*, Fig. S4); however, they are slightly related to biogeochemical processes, mainly through the carbon cycle. In contrast, oxidoreductases are directly involved in most biogeochemical processes and nutrient recycling in every environment. Thus, the diversity of these functions should be relevant to better understand the stability and conservation of biomes, affected by the high disparities between ecosystem conversion and conservation across biomes, which has been recognized as comprising an ongoing biome crisis (43). Indeed, conservation efforts have mainly focused on particular species or local macrocommunities (e.g., polar bears and coral reefs, respectively) but not on the microbial ecological functions that sustain trophic levels, biogeochemical cycles, and the ecosystem services that are derived from them. This omission

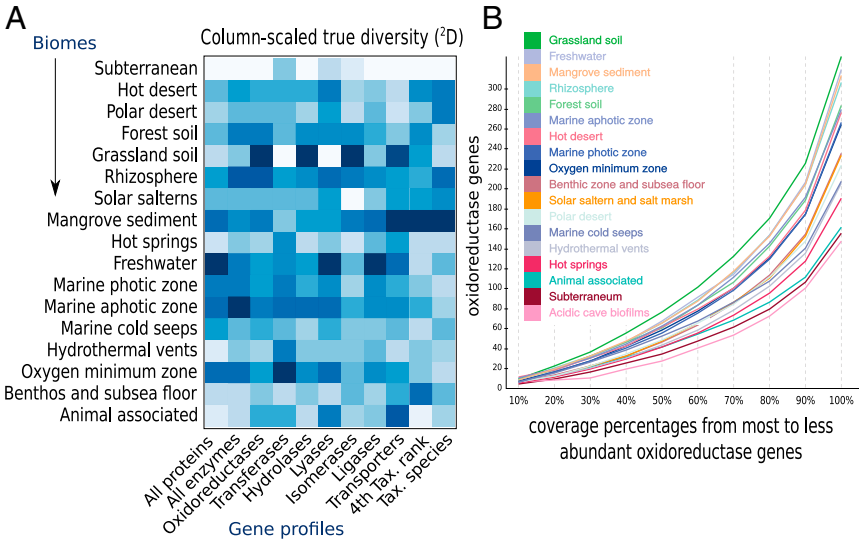

Fig. 5. Microbial diversity of biomes. (*A*) Heat map plot constructed with the inverse Simpson diversity index (true diversity with q = 2) of the taxonomic and functional profiles for the metagenomes, averaged per biome. The dark color shades indicate high diversity. These average values were scaled per profile category for homogenous contrast. Thus, the colors can only be compared along columns, i.e., by biome. For example, regarding oxidoreductase genes, the grassland biome is the most diverse, and the rhizosphere is the second one. On the other hand, the subterranean biome is shown as the less diverse biome in almost every gene category. Note that "All proteins" refer to all proteins with defined orthology in the KEGG database (see *SI Appendix* for details). (*B*) Average number, per biome, of oxidoreductase genes (vertical axis) necessary to cover different percentages of total oxidoreductase genes, counted from the most to less abundant. For example, the 60 most abundant oxidoreductase genes in grassland-associated datasets on average covered ca. 45% of the total pool of oxidoreductase genes.

is likely due to the difficulty of predicting microbial ecosystem dysfunction from environmental stressors using microbial taxonomy information (46). We expect that an oxidoreductase-based description of microbial communities should facilitate this task and help to quantify in future developments the impact of environmental changes on microbial ecosystem functions in the context of the global-scale biome crisis that our planet currently faces.

Materials and Methods

Data Collection and Sequence Analysis. The metagenomic datasets were collected from metagenomic studies of diverse microbial communities in recent years. The selection of metagenomes was guided by literature search, trying to cover the biomes with at least three "whole genome amplified" metagenomes sequenced with 454 or Illumina technologies. This process resulted in 247 metagenomes, grouped in 18 biomes (Fig. 1). The sources of these datasets are listed in the *SI Appendix*, Table S1. The sequences of these datasets were aligned against different protein sequence databases (*SI Appendix*, Fig. S6) using the BLASTX algorithm of the DIAMOND software with a bit-score cutoff of 50. With these alignment results, the different profiles listed in Fig. 1 and Table 1 were constructed.

Group Variances Analyses. The PERMANOVA statistical test was used to assess and compare the degree of separation of metagenomes (microbial communities) into biome groups by using the data profiles (Table 1) with dissimilarity matrices constructed with distances calculated based on nonparametric correlations [MIC and Spearman].

Diversity Estimation. Each profile of categories, for all of the metagenomic datasets (Fig. 1), was first resampled by a coverage percentage of 95%. True diversity was calculated on the resampled datasets by using the inverse Simpson index. The diversity per biome was calculated as the average of the diversities of all metagenomic datasets from each biome (*SI Appendix*, Table S1).

Networks and Clustering. For each pair of profiles described above, a distance between them was calculated as 1 correlation (correlation as the pairwise maximal information coefficient between the profiles). The networks of metagenomes (Fig. 2*A*) were constructed by writing the graph in the graph exchange XML format and rendered using the Gephi software with the OpenOrd network layout. The hierarchical clustering of biomes was computed with the R package Pvclust with 10^4 permutations and with a distance based on the Spearman correlation. The genes in Fig. 3 were selected as the top three oxidoreductase genes from each biome whose average ranking was lower than the total average. More details about all these procedures can be found in the *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by the Millennium Science Initiative Grant IC120019, the Chilean National Commission for Scientific and Technological Research Fondecyt Grant 1161483 (to O.U.), and the Center of Applied Ecology and Sustainability (CAPES) (to B.G.).

1. Maynard-Smith J (1974) *Models in Ecology* (Cambridge Univ Press, Cambridge, UK).
2. Hey J (2001) The mind of the species problem. *Trends Ecol Evol* 16:326–329.
3. Cohan FM (2002) What are bacterial species? *Annu Rev Microbiol* 56:457–487.
4. Doolittle WF, Papke RT (2006) Genomics and the bacterial species problem. *Genome Biol* 7:116.
5. Doolittle WF, Zhaxybayeva O (2009) On the origin of prokaryotic species. *Genome Res* 19:744–756.
6. Boon E, et al. (2014) Interactions in the microbiome: Communities of organisms and communities of genes. *FEMS Microbiol Rev* 38:90–118.
7. Darmon E, Leach DRF (2014) Bacterial genome instability. *Microbiol Mol Biol Rev* 78: 1–39.
8. Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proc Natl Acad Sci USA* 87: 4576–4579.
9. Fuhrman JA, et al. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci USA* 103:13104–13109.
10. Faust K, Raes J (2012) Microbial interactions: From networks to models. *Nat Rev Microbiol* 10:538–550.
11. Thompson LR, et al.; Earth Microbiome Project Consortium (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551:457–463.
12. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T (2011) Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci USA* 108: 14288–14293.
13. Nemergut DR, et al. (2016) Decreases in average bacterial community rRNA operon copy number during succession. *ISME J* 10:1147–1156.
14. Louca S, Parfrey LW, Doebeli M (2016) Decoupling function and taxonomy in the global ocean microbiome. *Science* 353:1272–1277.
15. Bletz MC, et al. (2016) Amphibian gut microbiota shifts differentially in community structure but converges on habitat-specific predicted functions. *Nat Commun* 7: 13699.
16. Dinsdale EA, et al. (2008) Functional metagenomic profiling of nine biomes. *Nature* 452:629–632.
17. Doolittle WF, Booth A (2017) It's the song, not the singer: An exploration of holobiosis and evolutionary theory. *Biol Philos* 32:5–24.
18. Green JL, Bohannan BJM, Whitaker RJ (2008) Microbial biogeography: From taxonomy to traits. *Science* 320:1039–1043.
19. Nemergut DR, et al. (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* 77:342–356.
20. Falkowski PG (2015) From light to life. *Orig Life Evol Biosph* 45:347–350.
21. Igai K, et al. (2016) Nitrogen fixation and nifH diversity in human gut microbiota. *Sci Rep* 6:31942.
22. Kiene RP, Williams LPH (1998) Glycine betaine uptake, retention, and degradation by microorganisms in seawater. *Limnol Oceanogr* 43:1592–1603.
23. Eloe EA, et al. (2011) Going deeper: Metagenome of a hadopelagic microbial community. *PLoS One* 6:e20388.
24. Roulier MA, Palenik B, Morel FMM (1990) A method for the measurement of choline and hydrogen peroxide in seawater. *Mar Chem* 30:409–421.
25. Schertl P, Danne L, Braun H-P (2017) 3-Hydroxyisobutyrate dehydrogenase is involved in both, valine and isoleucine degradation in *Arabidopsis thaliana*. *Plant Physiol* 175: 51–61.
26. Bartual A, et al. (2014) Polyunsaturated aldehydes from large phytoplankton of the Atlantic Ocean surface (42°n to 33°s). *Mar Drugs* 12:682–699.
27. Pérez-Pantoja D, González B, Pieper DH (2010) Aerobic degradation of aromatic hydrocarbons. *Handbook of Hydrocarbon and Lipid Microbiology*, ed Timmis KN (Springer, Berlin), pp 799–837.
28. Teufel R, et al. (2010) Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc Natl Acad Sci USA* 107:14390–14395.
29. Moore JC, et al. (2004) Detritus, trophic dynamics and biodiversity. *Ecol Lett* 7: 584–600.
30. Nie Y, Wang DQ, Zhao G, Yu S, Wang HY (2016) Effects of betaine aldehyde dehydrogenase-transgenic soybean on phosphatase activities and rhizospheric bacterial community of the saline-alkali soil. *BioMed Res Int* 2016:4904087.
31. Vriezen JAC, de Bruijn FJ, Nüsslein K (2007) Responses of rhizobia to desiccation in relation to osmotic stress, oxygen, and temperature. *Appl Environ Microbiol* 73: 3451–3459.
32. Fondi M, et al. (2016) "Every gene is everywhere but the environment selects": Global geolocalization of gene sharing in environmental samples through network analysis. *Genome Biol Evol* 8:1388–1400.
33. Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* 320:1034–1039.
34. Kielak AM, Barreto CC, Kowalchuk GA, van Veen JA, Kuramae EE (2016) The ecology of Acidobacteria: Moving beyond genes and genomes. *Front Microbiol* 7:744.
35. Morris RM, et al. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420:806–810.
36. Diaz RJ, Rosenberg R (2008) Spreading dead zones and consequences for marine ecosystems. *Science* 321:926–929.
37. Wright JJ, Konwar KM, Hallam SJ (2012) Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* 10:381–394.
38. Breitburg D, et al. (2018) Declining oxygen in the global ocean and coastal waters. *Science* 359:eaam7240.
39. Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ (2012) Microbial oceanography of anoxic oxygen minimum zones. *Proc Natl Acad Sci USA* 109:15996–16003.
40. Lozupone CA, Knight R (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* 104:11436–11440.
41. Torsvik V, Øvreås L (2002) Microbial diversity and function in soil: From genes to ecosystems. *Curr Opin Microbiol* 5:240–245.
42. Wilson JB, Peet RK, Dengler J, Pärtel M (2012) Plant species richness: The world records. *J Veg Sci* 23:796–802.
43. Hoekstra JM, Boucher TM, Ricketts TH, Roberts C (2005) Confronting a biome crisis: Global disparities of habitat loss and protection. *Ecol Lett* 8:23–29.
44. Krause S, et al. (2014) Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front Microbiol* 5:251.
45. McGill BJ, Enquist BJ, Weiher E, Westoby M (2006) Rebuilding community ecology from functional traits. *Trends Ecol Evol* 21:178–185.
46. Webster NS, Wagner M, Negri AP (April 6, 2018) Microbial conservation in the Anthropocene. *Environ Microbiol* 20:1925–1928.