UNIVERSIDAD ADOLFO IBÁÑEZ
FACULTAD DE INGENIERÍA Y CIENCIAS
PhD IN COMPLEX SYSTEMS ENGINEERING

STUDY ON THE USE OF GENETIC FUNCTIONAL TRAITS TO CHARACTERIZE MICROBIAL COMMUNITIES

Salvador Ramírez Flandes

Thesis Director: Dr. Bernardo González
Thesis Co-director: Dr. Osvaldo Ulloa

Submitted as a requirement for the degree of PhD in Complex Systems Engineering at Universidad Adolfo Ibáñez, April 2019

Contents

ABSTRACT

Biological communities are conventionally described as assemblages of species, whose ecological roles are known or predictable from their observable morphology. In microbial ecology, however, such a taxonomic approach is hindered by our limited knowledge of the functions of most microorganisms, which often alter their genetic material through diverse mechanisms. To tackle these problems, microbial ecologists have used culture-independent genetic approaches to study the whole pool of functional genes at the community level. However, this approach requires dealing with gene categories not necessarily related to the ecology of the organisms, such as functions associated with DNA replication or cellular division. In this work it is demonstrated that genes encoding oxidoreductases characterize the microbial communities better than other categories of genes, including those associated with taxonomy. Additionally, with this approach, the role of microbial communities of the different ecosystems in biogeochemical cycles becomes readily apparent. The importance of this result is, however, limited by the coverage of known genetic functions over the total pool of metagenomic genes, which is currently around the fifty percent in sampled metagenomes. To help increasing this reduced coverage of known functions, a methodology for the recognition of the protein-coding potential of DNA sequences is proposed. Alternative applications of this methodology are discussed. The results of this study pave the way for a better assessment and evaluation of microbial ecosystem services from different environments of our planet. This improved diagnostic of microbial ecosystems can be possible, for example, by focusing directly on the diversity of redox functions encoded in the metagenomes of microbial communities, rather than on their taxonomic structures. Thus, this approach should help in developing better management and conservation policies that effectively include not only iconic species or colonies, such as polar bears or coral reefs, but also microorganisms.

CHAPTER 1. INTRODUCTION

In broad terms, Ecology is the study of the interactions among living organisms and their respective environments. The German zoologist Ernst Haeckel coined the word ecology in 1866, but early studies in this discipline can be traced back to ancient Greece, mainly associated with Aristotle. The organisms are a natural basic level of organization through which living entities in Nature can be described (Fig. 1), and their categorization into kinds or species (e.g., the use of *the* word "dog" to refer to *any* particular dog) might be one of the most ancient human tasks. However, as far as we know today, Aristotle was the first in writing records of an organized and hierarchical classification of organisms. This authoring recurrence is not surprising as ecology depends on a previous classification of organisms, and modern studies make use of a system based on the Linnaean taxonomy on this matter. In this context, the first attempts for establishing a mathematical theory of ecology were the works of Alfred Lotka and Vito Volterra in the decade of 1920, with the predator-prey model. The variables in the differential equations of mathematical models are numbers of species acting in specific ecological roles —predator and prey in this case. Most subsequent works since have used species numbers as the ecological unit to develop a body of theories to understand and predict the dynamics of populations and communities (1). The crucial aspects of this approach are the underlying assumptions that a) all individuals can be categorized into species, and b) the roles or functions of these interacting species are known for a proper setting of the parameters in the corresponding models.

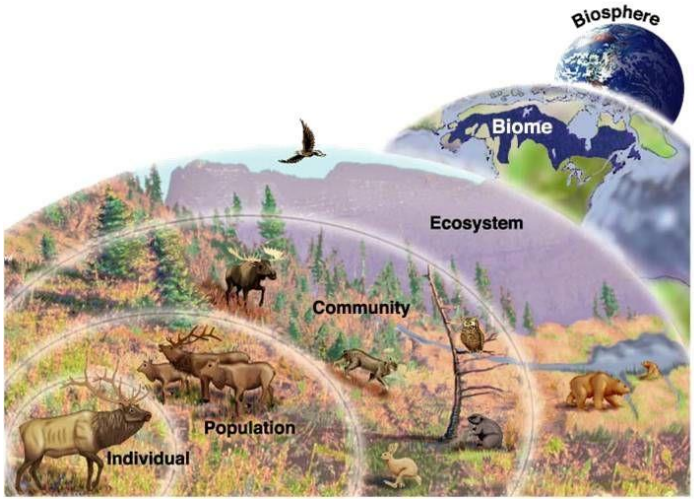


Figure 1. The standard ecological organization. Individuals or organisms are at the basic level of the ecological organization, which aggregated as species compose a population. Agregations of populations in turn define a community, which taken together with environmental conditions make up an ecosystem. Similar ecosystems over the planet are called biomes, and the aggregation of all of the biomes on Earth finally constitutes the biosphere.

Leaving aside the philosophical debate that the concept of species have arisen in the context of evolutionary biology (2), researchers have struggled to find general laws in community ecology utilizing this species-centric approach (3, 4). One reason for this might be the fact that species are complex and dynamical constructs, with multiple potential ecological functions, thus capable of establishing numerous and changing interactions with other species. In this regard, some authors have claimed that community ecology has lost its way by focusing on pairwise species interactions; an approach that has succeeded in explaining few-species systems but that has failed in providing general principles about many species communities (5). These authors also suggest that a return to a trait and environment-focused route, highlighting how functional traits are distributed across gradients in the light of what characterizes the diverse niches should have success where the species-centric approach has not (5).

In microbial ecology, this situation acquires even greater relevance for several reasons. First, the observable attributes of most microorganisms do not provide sufficient discriminatory or functional characterization, making the concept of species been even more problematic in this context (6-8). 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 (9). Accordingly, it has been estimated that cultured bacterial species represent only a tiny fraction of the total bacteria species on Earth (10, 11). Third, prokaryotic genomes are highly dynamic, mainly due to pervasive horizontal gene transfers, genome streamlining (12, 13), and the effect of mobile DNA elements and phages (14, 15). Microbial ecologists have employed molecular taxonomic markers, primarily the small subunit ribosomal RNA gene (SSU rRNA gene), to address the first problem indicated above, thereby operationally defining species (16). Subsequently, culture-independent approaches for studying environmental DNA have been used to estimate the abundances and taxonomic diversity of microorganisms in their natural environments (17, 18). This approach has been termed metagenomics (19), and it is intended to address the second problem described above.

By assuming that the taxonomic structure of microbial communities is a sound predictor of their functioning, this approach has been utilized to explain the microbial dynamics in several specific environments (20, 21). When the taxonomic rank of species has been unable to produce detectable patterns, some authors have found ecological coherence at a coarser taxonomic level, such as phylum or class (22-25). However, several studies have reported plain taxonomical correlations under apparently similar ecological scenarios, finding consistency only when using multiple protein-coding genes as traits and when the whole community is analyzed as the ecological unit (26-31). After all, it is the function, not the taxonomic information that has the actual ecological relevance (32). This approach has been called whole genome metagenomics (18, 33) or shotgun metagenomics (34), and it represents an attempt to address the third abovementioned problem. Unfortunately, a selection of ecologically relevant

categories of protein-coding genes is not evident in the broad context of planetary biomes (9, 35, 36). The use of the whole pool of functional genes at the microbial community level (shotgun metagenomics) requires dealing with profiles comprising thousands of gene categories not necessarily related to the ecology of the microorganisms such as those associated with the DNA replication or cellular division, among many others.

In these conditions, it is natural to ask whether the use of the whole set of functional genes in a metagenome is the best way to characterize its corresponding microbial community, or if it is possible to establish a subset of them that can do better. The first step to answering that question is to precisely define what a characterization of a microbial community is. The second step is to define a procedure to compare such characterizations in order to establish which of them provide the most ecologically relevant view of the microbial communities. A third step would be to evaluate, or at least to explore if that presumed subset might be within the set of genes for which a function is not yet defined. This last possibility should not be neglected as roughly a half of the metagenomic sequences are currently regarded as without a known function (37, 38).

Those are the research topics of this thesis. Under the natural assumption that the abundances of genetic traits in the microbial communities are determined by the conditions of their corresponding environments, a characterization of a microbial community is defined as a set of relative abundances of genetic traits (gene profiles) that distinguishes it from those inhabiting different biomes. In Chapter 2, it is shown that the set of genes related to molecular redox functions (oxidoreductase genes) can characterize the microbial communities from different biomes better than other sets of genes, including those associated with the taxonomy. This analysis considered only genes with known function. In Chapter 3, a methodology for the recognition of the protein-coding potential of DNA sequences is proposed to discover new gene orthologies that might potentially enhance the characterization of the microbial communities by helping assigning functions to presently unknown protein encoding sequences. Alternative applications of this development are discussed in Chapter 4.

CHAPTER 2. REDOX TRAITS CHARACTERIZE THE ORGANIZATION OF GLOBAL MICROBIAL COMMUNITIES

This chapter corresponds to a paper published in the journal Proceedings of the National Academy of Sciences of the United States of America (PNAS).

Ramírez-Flandes, S., González, B., & Ulloa, O. (2019). Redox traits characterize the organization of global microbial communities. Proceedings of the National Academy of Sciences, 116(9), 3630-3635.

https://doi.org/10.1073/pnas.1817554116

MANUSCRIPT TEXT

ABSTRACT

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 characterize the differences among biomes at the global scale. We show that profiles of oxidoreductase genes support the highest biome differentiation when 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.

SIGNIFICANCE STATEMENT

Biological communities are conventionally described as assemblages of species, whose ecological roles are known or predictable from their observable morphology. In microbial

ecology, such 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.

INTRODUCTION

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 (SSU rRNA 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 done in an attempt to address the third abovementioned problem. After all, it is the function, not the taxonomic information, which 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 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).

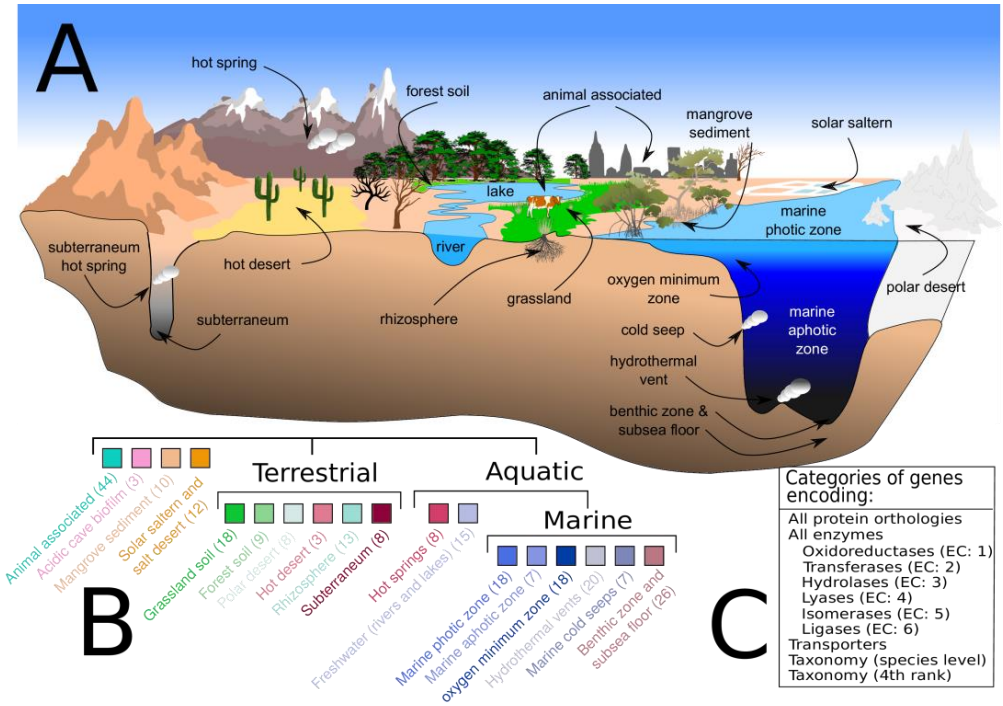


Figure 1. Biomes and categories of genes. A) Sketch of the biomes from which metagenomes (as proxies for microbial communities) were included in this study. The animal-associated biome included metagenomes from terrestrial animals only. A complete list and origin of these metagenomes can be found in the SI Appendix, Table S1. **B)** Organized list of the biomes illustrated in A. The number of metagenomes per biome is shown in parenthesis beside 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 4th rank taxonomy typically corresponds to a phylum in the prokaryotic taxonomy (see SI Appendix for details).

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).

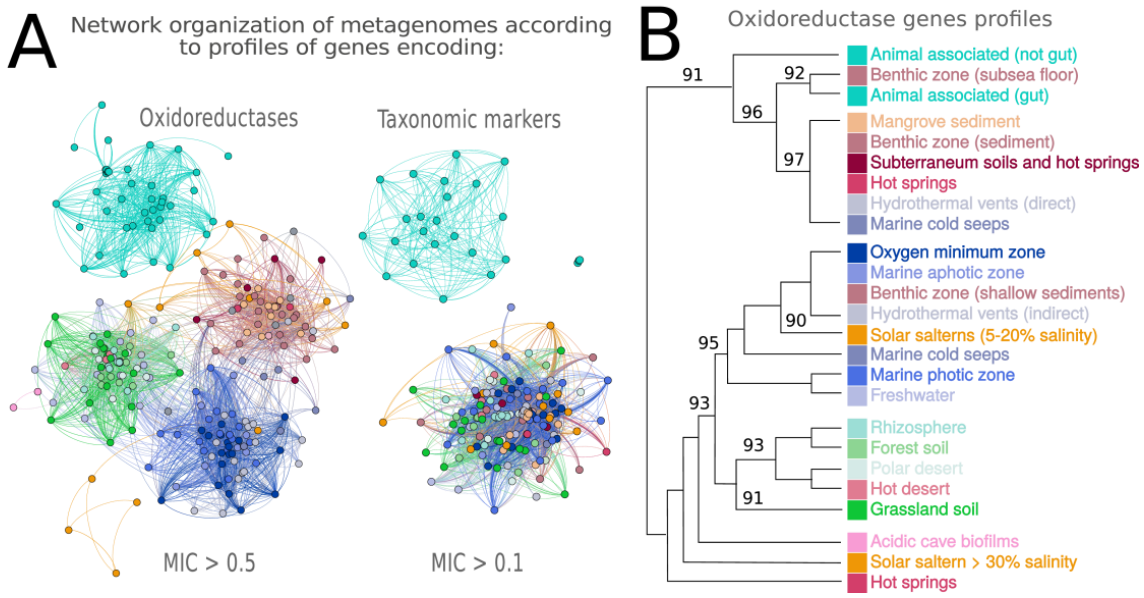


Figure 2. Association of microbial metagenomes and biomes. A) Network representation of the microbial metagenomes by profiles of oxidoreductase and taxonomic gene ranks. Nodes correspond to metagenomes, colored according to their biome of origin (Fig. 1). Edges represent maximal information coefficients (MIC). 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, 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 Figures 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.

Our results showed that redox functions support the highest statistical differentiation amongst biomes when taxonomic and functional sets of genes were compared (Table 1, SI Appendix Table S2, and Fig. S5). 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. 2A, 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. 2B, SI Appendix Figs. S1, 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, sub-seafloor 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 abovementioned 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).

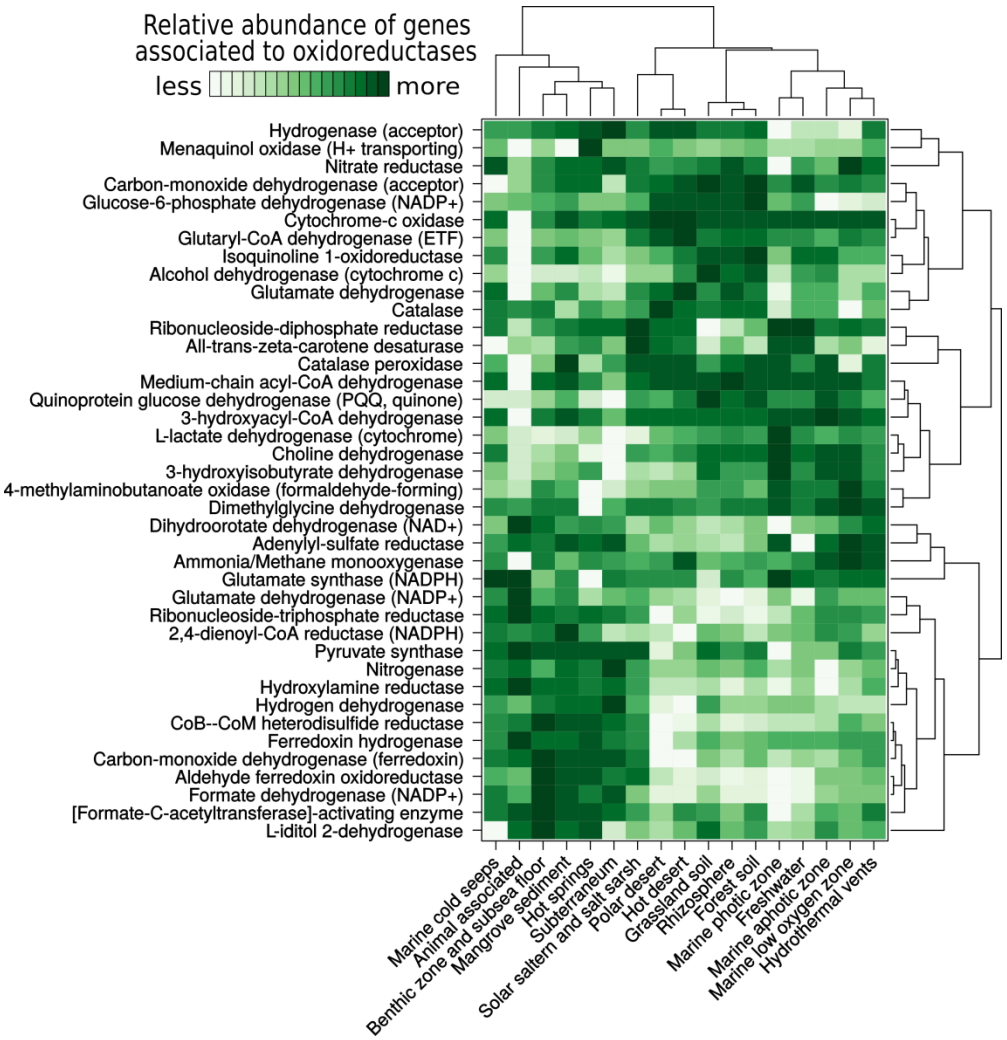


Figure 3. Distinctive oxidoreductase genes associated with biomes. These genes were determined by statistically testing that the average of ranks of each oxidoreductase gene within each biome was significantly different from the average rank in other biomes. Dark and light shades in this figure refer to relative abundances, high and low, respectively. Thus, the ranks for this figure were reversed, as a low rank indicates high relative abundance. These values were scaled for better visualization, which means that the color shades can only be compared horizontally. Some of these distinctive genes encode oxidoreductases 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.

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 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. 2B 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 the 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).

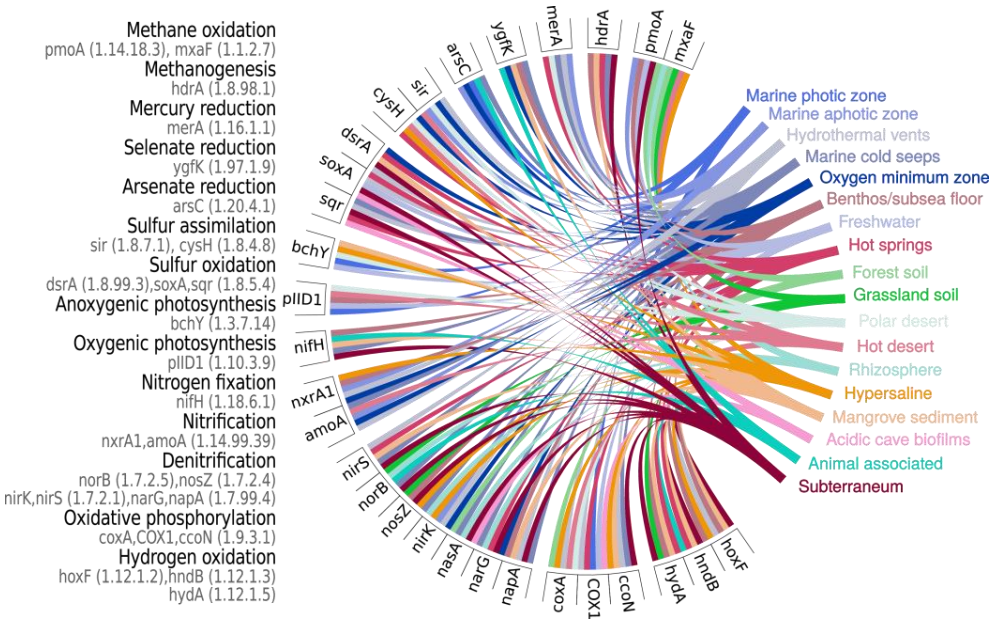


Figure 4. Biogeochemically relevant processes per biome by oxidoreductase genes relative abundances. Oxidoreductase genes associated with biogeochemical processes and their top five biomes where they were ranked the highest. The biomes per genes are in clockwise order, starting from the biome where the 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 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 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. 5A). 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. 5A). 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 under-protection (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. 5B).

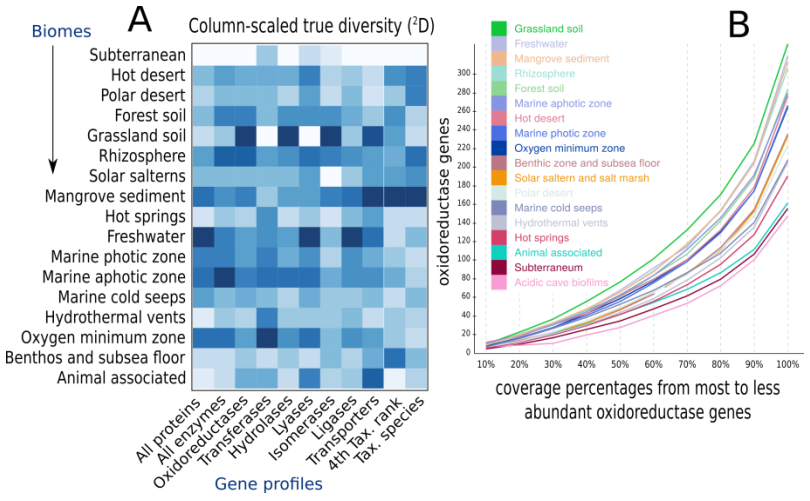


Figure 5. Microbial diversity of biomes. **A)** Heatmap plot constructed with the inverse Simpson diversity index (true-diversity with q=2) of the taxonomic and functional profiles for the metagenomes, averaged per biome. Dark color shades indicate high diversity. These average values were scaled per 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 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 in average covered ca. 45% of the total pool of oxidoreductase genes.

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, 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 macro-communities (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 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, Figure 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 Figure 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 non-parametric correlations [Maximal Information Coefficient (MIC) and Spearman]. **Diversity estimation**: Each profile of categories, for all the metagenomic datasets (Fig. 1), was first resampled by a coverage percentage of 95%. True diversity was calculated on the resampled datasets with 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 2A) were constructed by writing the graph in the Graph Exchange XML format (GEXF) 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 Figure 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 and the Chilean National Commission for Scientific and Technological Research Grant Fondecyt 1161483 (O.U.) and the Center of Applied Ecology and Sustainability [CAPES] (B.G).

REFERENCES

1. Maynard-Smith J (1974) *Models in Ecology*.
2. Hey J (2001) The mind of the species problem. *Trends Ecol Evol* 16(7):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(9):116.
5. Doolittle WF, Zhaxybayeva O (2009) On the origin of prokaryotic species. *Genome Res*

19(5):744–756.

6. Boon E, et al. (2014) Interactions in the microbiome: communities of organisms and communities of genes. *FEMS Microbiol Rev* 38(1):90–118.

7. Darmon E, Leach DRF (2014) Bacterial genome instability. *Microbiol Mol Biol Rev* 78(1):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* 87(12):4576–4579.

9. Fuhrman JA, et al. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci U S A* 103(35):13104–9.

10. Faust K, Raes J (2012) Microbial interactions: from networks to models. *Nat Rev Microbiol* 10(8):538–550.

11. Thompson LR, et al. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 551(7681):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(34):14288–14293.

13. Nemergut DR, et al. (2016) Decreases in average bacterial community rRNA operon copy number during succession. *ISME J* 10(5):1147–1156.

14. Louca S, Parfrey LW, Doebeli M (2016) Decoupling function and taxonomy in the global ocean microbiome. *Science (80-)* 353(6305):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 E, et al. (2008) Functional metagenomic profiling of nine biomes. *Nature* 452(7187):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(1):5–24.

18. Green JL, Bohannan BJM, Whitaker RJ (2008) Microbial biogeography: from taxonomy to traits. *Science* 320(5879):1039–1043.

19. Nemergut DR, et al. (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* 77(3):342–356.

20. Falkowski PG (2015) From light to life. *Orig Life Evol Biosph* 45(3):347–350.

21. Igai K, et al. (2016) Nitrogen fixation and nifH diversity in human gut microbiota. *Sci*

Rep 6(August):1–11.

22. Kiene RP, Hoffman Williams LP (1998) Glycine betaine uptake, retention, and degradation by microorganisms in seawater. *Limnol Oceanogr* 43(7):1592–1603.

23. Eloe EA, et al. (2011) Going deeper: Metagenome of a hadopelagic microbial community. *PLoS One* 6(5):e20388.

24. Roulier MA, Palenik B, Morel FMM (1990) A method for the measurement of choline and hydrogen peroxide in seawater. *Mar Chem* 30(C):409–421.

25. Schertl P, Danne L, Braun H-P (2017) 3-Hydroxyisobutyrate dehydrogenase is involved in valine and isoleucine degradation in A. thaliana. *Plant Physiol* 175(September):pp.00649.2017.

26. Bartual A, et al. (2014) Polyunsaturated aldehydes from large phytoplankton of the Atlantic Ocean surface (42°N to 33°S). *Mar Drugs* 12(2):682–699.

27. Pérez-Pantoja D, González B, Pieper DH (2010) Aerobic degradation of aromatic hydrocarbons. *Handbook of Hydrocarbon and Lipid Microbiology*, pp 799–837.

28. Teufel R, et al. (2010) Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc Natl Acad Sci* 107(32):14390–14395.

29. Moore JC, et al. (2004) Detritus, trophic dynamics and biodiversity. *Ecol Lett* 7(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. doi:10.1155/2016/4904087.

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(11):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(5):1388–1400.

33. Falkowski PG, Fenchel T, Delong EF (2008) The microbial engines that drive Earth’s biogeochemical cycles. *Science (80-)* 320(5879):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(MAY):1–16.

35. Morris RM, et al. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420(December):806–810.

36. Diaz RJ, Rosenberg R (2008) Spreading dead zones and consequences for marine

ecosystems. *Science (80-)* 321(5891):926–929.

37. Wright JJ, Konwar KM, Hallam SJ (2012) Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* 10(6):381–394.

38. Breitburg D, et al. (2018) Declining oxygen in the global ocean and coastal waters. *Science (80-)* 359(6371). doi:10.1126/science.aam7240.

39. Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ (2012) Microbial oceanography of anoxic oxygen minimum zones. *Proc Natl Acad Sci U S A* 109(40):15996–6003.

40. Lozupone C, Knight R (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* 104(27):11436–11440.

41. Torsvik V, Øvreås L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr Opin Microbiol* 5(3):240–245.

42. Wilson JB, Peet RK, Dengler J, Pärtel M (2012) Plant species richness: The world records. *J Veg Sci* 23(4):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(1):23–29.

44. Krause S, et al. (2014) Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front Microbiol* 5(MAY):1–10.

45. McGill BJ, Enquist BJ, Weiher E, Westoby M (2006) Rebuilding community ecology from functional traits. *Trends Ecol Evol* 21(4):178–185.

46. Webster NS, Wagner M, Negri AP (2018) Microbial conservation in the Anthropocene. *Environ Microbiol*. doi:10.1111/1462-2920.14124.

SUPPLEMENTARY INFORMATION APPENDIX (SI APPENDIX)

Data collection. The metagenomic datasets used in this work were collected from selected metagenomic studies of diverse microbial communities in recent years. The selection of metagenomes was guided by literature trying to cover the biomes with at least three metagenomes. Only "Whole Genome Amplified" (WGA) metagenomes sequenced with Roche 454 or Illumina technologies were selected because their output data has been demonstrated to provide comparable views of the sampled communities (*1*). This process resulted in 247 grouped metagenomes in the 18 depicted biomes in Figure 1. The specific sources and references for these metagenomic datasets are listed in Table S1. For the datasets stored in the NCBI Sequence Read Archive (SRA), the fastq-dump program from the NCBI SRA toolkit version 2.8.2 was used (publicly available at https://github.com/ncbi/sra-tools/wiki/Downloads). For the datasets in the MG-RAST system, the application programming Interface at https://api.metagenomics.anl.gov was used with the Linux program cURL (https://curl.haxx.se).

Sequence analysis. A simplified sketch of the main procedures, described here, is illustrated in Figure S6. The collected datasets were renamed according to the third column in Table S1, and their reads were directly compared through translated sequence alignments with the sequences in a subset of the KEGG protein database (*2*), the TCDB Transporters database (*3*), and a custom database (RIBPROTSDB) that was composed of all ribosomal protein sequences from the complete and draft genomes in the NCBI GenBank website[1]. The KEGG protein database used in this study was a subset of the original that included both prokaryotic and eukaryotic sequences but only with an assigned KEGG orthology (hereinafter referred to as the KEGG database for simplicity). All these massive sequence alignments were carried out using the BLASTX algorithm in the DIAMOND software package (*4*) with the options "–sensitive" and "—max-target-seqs 25". The idea behind the latter option was to select only the "best" hit on the target database for each query sequence, and thus we could have used a value of 1 for this parameter. However, the DIAMOND documentation does not explicitly describe what "best" means in this context. Therefore, we let DIAMOND output its best 25 hits and within those we selected the best hit according to the criterion that follows. First, there are several criteria to evaluate alignments. For example, some authors use the alignment length and the sequence identity for assessing homology, whereas others use an E-value or a bit-score cutoff. Alignment filtering based on alignment length and sequence identity ignores the sequence similarity information provided by the sequence alignment analysis using substitution matrices like BLOSUM62 (default option in most BLASTX algorithm implementations, DIAMOND in

[1] ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/assembly_summary.txt and same link changing "bacteria" for "archaea".

particular). On the other hand, a bit-score somehow encodes a measure that combines alignment length and sequence similarity into a single number. The alignment filtering based on an E-value cutoff is only appropriate when a single reference database is used because it depends on the size of the database. Thus, as we use multiple reference databases (Kegg protein sequences with KO, TCDB and a database of ribosomal protein sequences), and since we wanted to use a single criterion comparable across different reference databases, we opted to select the hit with the highest bit-score, considering a minimum cutoff value of 50. After this process, counts of categories in the reference databases were obtained for each metagenome. With the counts in the KEGG database, profiles for KEGG orthologies (KO), enzymes (EC), oxidoreductases (EC1), transferases (EC2), hydrolases (EC3), lyases (EC4), isomerases (EC5) and ligases (EC6) genes were created. The EC numbers were computed using the KO information and the "ko2ec" mapping from the KEGG distribution. With the corresponding counts in the TCDB database, profiles of transporters were constructed using the respective TCDB IDs in that database. The taxonomic profiles at the species and 4th rank level were constructed with information from the GenBank ID (GI), associated with the best BLASTX hit (whenever its corresponding bit-score was higher than or equal to 50) in the RIBPROTSDB and converted to a NCBI taxonomy ID for each sequence (using the database ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi_taxid_prot.zip). The translation from NCBI taxonomic IDs to names was carried out using the NCBI taxonomy database (ftp://ftp.ncbi.nih.gov/pub/taxonomy/). An example for the 4th rank taxonomy, in the taxonomic lineage is as follows:

Bacteria; FCB group; Bacteroidetes/Chlorobi group; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides; Bacteroides vulgatus; Bacteroides vulgatus PC510

where, the 4th rank level taxonomy would be *Bacteroidetes*, while the species rank would be *Bacteroides vulgatus PC510*. Note that the first NCBI rank taxonomy, associated with "cellular organisms" was omitted. Generally, but not always, the 4th rank corresponds to phylum as the prokaryotic taxonomy is not always complete. Profiles for this special "taxonomy rank" were included because higher rank taxonomies in microbial ecology can achieve more coherence than the species level (5). We used ribosomal proteins to construct the taxonomic profiles because the typical taxonomic marker gene, ssu_rRNA (16S or 18S), has normally very low representation in WGA metagenomes, resulting in reduced sample sizes, thereby potentially affecting the power of the statistical tests. The ribosomal proteins were selected because they rarely participate in horizontal gene transfers, just like the ssu_rRNA gene.

Count profiles were then used to create two different sets of normalized profiles: ranking and variance stabilized profiles. The former set was created with a custom script while the latter set was created using the DSeq2 package (*6*) in the R statistical package. The DSeq2 function used was "varianceStabilizingTransformation" with the parameter fitType='local'. Normalization of

the microbiome data is necessary to allow the comparison of datasets with different sizes. Historically, the most popular normalization methods for analyzing microbiome data have been the use of ratios (dividing counts by the total number of counts per dataset) and rarefaction, both of which have known statistical problems (*7*). More recently, variance stabilizing methods that make use of the negative binomial distribution have gained popularity in these analyses. However, these methods convert the natural count numbers into artificial real numbers that are sometimes difficult to interpret in downstream analyses (because, for example, the order of the relative abundances of the elements is lost within a dataset). On the other hand, ranking analysis provides easy-to-interpret data while at the same time is useful in reducing the complexity of heterogeneous data by weakening the order (without losing it) of raw numbers coming from sources with possible noise. For these reasons, ranking profiles were preferably used in most subsequent analyses, unless otherwise stated.

Group variances analyses. Permutational multivariate analysis of variance (PERMANOVA) analyses (Adonis2 implementation from the "vegan" R package, with default number of permutations = 999) were used to test whether the predetermined grouping (biomes) of metagenomes (microbial communities) were determined by the different similarity (correlation) matrices created as described below for the different set of genes (Table 1, Table S2 and Figure S5). For the analysis shown in Table 1, the variance-stabilized counts were used because the ranking gives more weight to the differences among the rarest features, as the rarer a feature is, the higher is its rank number (rankings start from the most abundant with number 1) and thus, its variability. This artificial high variation might affect the MIC (maximal information coefficient) correlation of profiles (more details about correlations is in the section "Correlations, distances, networks, heatmaps and hierarchical clustering" below). To test this effect, we used ranking profiles over different coverage percentages (covering from most to less abundant) of the metagenomes to verify that with lower coverage percentages (i.e., more of the rare elements ignored) the higher was the F-statistic of the PERMANOVA differentiating the biome groups, and the oxidoreductases progressively and consistently dominated this differentiation at 70%, 60% and 50% (Table S2). These results suggested that the rankings of the "long tail" of rare genes (in each category) were affecting the correlations, justifying the use of variance-stabilized counts in this particular analysis by reducing the high variability of the ranks with large numbers (rare elements) (Table 1). However, the results in Table 1 were still not clear as many enzyme categories had rather similar values for the F-statistic. To unambiguously determine which set of enzyme genes produce the highest separation of metagenomes into biome groups, a third PERMANOVA analysis was performed. This time the analysis included bootstrapping (to provide statistical significance of differences among the resulting F-statistic values) and removal of variables or methods from the previous PERMANOVA analyses that might potentially affect the comparison such as different profile sizes (last column in Table 1), different dataset sizes (metagenomes sequenced at different

depths), the use of normalization methods (such as ratios or ranking-based methods) and the use of a reference database (Kegg protein sequences with assigned orthology) that might potentially include bad annotated sequences. To this end, all metagenomes were re-analyzed with mi-faser (*9*) against the GS+ database of sequences from experimentally verified enzymatic functionality (included in the mi-faser software distribution). This database currently has a limited number of sequences to be used as a reference database in general metagenomic studies (the version used here has only 2865 sequences), but that is not a problem for this particular analysis (unlike other analyses in this study, aimed at characterizing the microbial communities qualitatively with reasonable completeness). The output of this process was a set of profiles of counts for the different enzymes [counted with the reads associated with the genes sharing the same Enzyme Commission (EC) number] for each metagenome. With these data, there are two ways in which a bootstrapping analysis can be done. The first is to resample (with replacement) a fixed number of enzymes from these profiles (for each category), and the other is to pre-select a fixed number of enzymes (for each category) and to resample the sequences of each metagenome until a fixed number of counts on these pre-selected enzymes is reached. The second alternative was chosen to achieve homogeneity of the total counts among all resamples, making the particular counts directly comparable. With this method, the use of rankings or variance stabilization methods was avoided, although this method can be seen as a type of normalization by rarefaction on the pre-selected set of enzymes for each category. The size of the sets of the pre-selected enzymes for this analysis was set to 50 (roughly, half of the smaller profile size in Table 1). For each enzyme category, the set of these 50 enzymes were selected from the most abundant ones in the metagenomes across all of the biomes as follows:

1. For each biome, the enzyme ranks of the associated metagenomes were averaged, resulting in profiles of average ranks of enzymes per biome.
2. The top enzyme from its corresponding profile (without replacement) until reaching a set of 50 enzymes was selected, one at a time per biome.

Having established these sets of 50 representative enzymes (one set for each enzyme category, Table S3), all metagenomic datasets were resampled 100 times by randomly selecting sequences until these 50 enzymes reached a total count of 1000. With these 100 resamples for each enzyme category, PERMANOVA analyses for group differences were carried out, and the results are shown in Figure S3. ANOVA analysis of the F-values for each resample showed a strong difference of these values among the different categories (p-value $< 10^{-16}$). A post-hoc analysis of these differences was done with the "pairwise t-test," correcting p-values with the Bonferroni method in the R-statistical software, resulting in a significant difference between every pair of groups (all p-values $< 10^{-16}$, except for the Lyases vs. Ligases whose p-value was $< 10^{-14}$).

Cohesion and separation of clusters. To further quantify how the original biome of the sample might be determined by different metagenomic profiles of gene counts, the cohesion and separation of groups (biomes) that resulted from the different sets (profiles) of genes was analyzed. An optimal characterization of any grouping or clustering seeks high cohesion within groups and high separation between groups. An intuitive measure of cohesion is the average of the pairwise correlations of the elements within groups, which in this case correspond to correlations of profiles for each metagenome within particular biomes (see below for details about the type of correlation employed). Maximization of this average was followed to accomplish high cohesion. On the other hand, the separation between groups was estimated with the average of pairwise correlations of profiles for each metagenome to every other metagenome outside the same biome. Minimization of this average was followed to separate the groups (low correlation to outer elements, so separation will be the negative value of this average). Thus, a natural function to optimize both cohesion and separation is simply the sum of both values.

Correlations, distances, networks, heatmaps and hierarchical clustering. For each pair of profiles of rankings from all metagenomes, a measure of correlation between them was calculated using the maximal information coefficient, MIC/MINE (*10*). Given the high-volume of data, the RapidMIC implementation of this algorithm was used (*11*), which is a multithreaded version of MINE written in C++. With these correlations, a distance between every pair of metagenomes (X and Y) was calculated as 1–abs(correlation(X,Y)). The networks of metagenomes were constructed by writing a graph description in the GEXF format with the correlations for each metagenome to other metagenomes (MIC correlation ≥ 0.7 in the case of the network associated with oxidoreductases, and ≥ 0.1 for the network associated with taxonomic profiles). The GEXF files were then graphically rendered by using the Gephi software with default parameters (*12*). The "OpenOrd" layout algorithm (with default parameters) was used for these graphs because it can make use of edge weights but mainly because it is aimed at better distinguish clusters. The hierarchical clustering of metagenomes was computed with the R package pvclust (*13*) with 10000 permutations to give the branches statistical significance. Here, a similar distance, as described above, was used but with the Spearman correlation because the computation of the MIC is a time-consuming process for large data. The use of this distance allowed 10000 bootstrap permutations to be carried out in a reasonable time period. The heatmap of Figure 3 was created with the "levelplot" function from the package "lattice" in the R statistical software. A typical method to determine the distinctive elements in groups of data (distinctive oxidoreductases per biome in this case) is SIMPER (*14*), which determines the elements that contribute more to the dissimilarity of the groups by using the Bray–Curtis measure of similarity. The use of this method (SIMPER) was avoided mainly because it has been recently demonstrated that this method can potentially lead to wrong conclusions (*15*). Thus, the distinctive oxidoreductase genes per biome were obtained as follows: for each biome,

statistical hypothesis tests were conducted for each oxidoreductase gene under the assumption that its average ranking was equal to the corresponding average in all the other biomes, keeping only the oxidoreductase genes (for each biome) that rejected that null hypothesis with a p-value < 0.05. The top three (by the lowest ranking within each biome group, as lower ranking means higher abundance) oxidoreductase genes from the whole set of them were considered. If we have two p-values for two different t-tests (two elements being tested for their "distinctiveness" in a biome), for example, p-value1 = 0.003 (with avg. rank = 100) and p-value2 = 0.005 (with avg. rank = 50), as both p-values were under the cutoff of 0.05, we selected the second element (despite its higher p-value) because it had a lower average rank (again, the lower the rank, the higher its relative abundance). Many oxidoreductase genes selected in this way were shared among many related biomes, but the unique set of them was considered, resulting in the oxidoreductase names displayed in Figure 3. To statistically assess these averages differences the t-student test in R was used for simplicity and coherence with other analyses in this work. Even when these data are not necessarily normal, the robustness of the t-test under non-normal, large data has been previously demonstrated (*16*). Nonetheless, this analysis was also carried out using the Mann–Whiney–Wilcoxon test, and the results were practically the same, with less than 5% of the distinctive elements having differences. Therefore, the results of the analysis with the t-test were used.

Diversity estimation. Each profile of categories for all the metagenomic datasets was resampled, using the Turing-Chao improved coverage estimator, by a coverage of 95% to standardize samples by completeness rather than fixed sizes (*17*, *18*). Subsequently, true diversity (effective numbers of types) was calculated with the parameter q = 2 (*19*). This measure of diversity is also called the inverse Simpson index, and it has the characteristic of giving more weighing to the most abundant types (more than the Shannon index, and much more than a raw richness index). This index was also selected because it is the least affected index by heterogeneous sample sizes and inventory completeness (*20*). The diversity of biomes was calculated as the average of diversities of all metagenomic datasets from each biome (Table S1). The heatmap in Figure 5 was created using the "levelplot" function from the package "lattice" in R with input data from the above described averages. The curves of coverage of oxidoreductase and taxonomic marker genes in Figure 4B and 4C were obtained as follows: first, all datasets (metagenomes) were resampled at 95% of coverage as described above. Afterwards, a simple coverage diversity estimator was defined as the number of categories (oxidoreductase or taxonomy-associated genes) that were present at a given percent coverage for each of these new 100% resamples. With these data, an estimator for percent values of 10%, 20%, 30%… 100% was calculated for each metagenome. An average per biome was calculated using the values that have just been described. With these average values, a Scalable Vector Graphics representation was written using a custom script program. The use of typical rarefaction analyses was avoided because it has been demonstrated that these analyses

are not appropriate when the size of samples significantly varies among datasets (*7*, *17*), which is the case in this study due to the consideration of very heterogeneous datasets of metagenomes obtained with different sequencing techniques, different sequencing depths, different times, etc. Note that the taxonomic diversities assessed here can be underestimated because only ribosomal proteins from sequenced genomes were used for this determination instead of the more common 16S gene (the taxonomic coverage of ribosomal proteins from sequenced genomes is lower than the taxonomic coverage of 16S genes). Thus, this taxonomic diversity estimation should be considered for comparative purposes only, which was the objective of this analysis, but not as absolute taxonomic diversities.

FIGURES

The figures of this supplementary material are presented in the following pages.

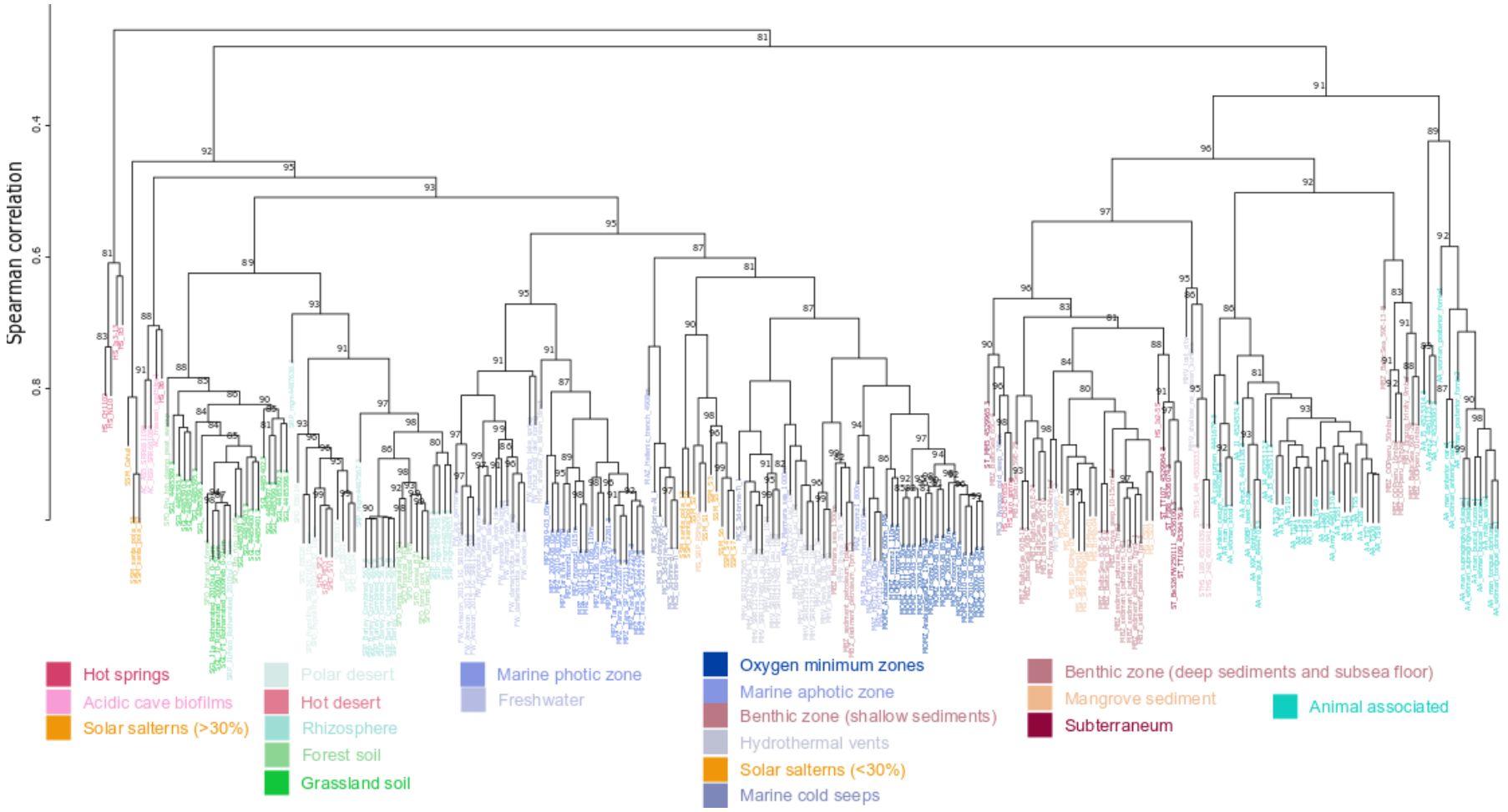


Figure S1

Hierarchical clustering of metagenomes based on profiles of genes encoding oxidoreductases. The distance metric used for this clustering was 1-correlation (Spearman), and support values were obtained with the pvclust R package with 10^4 permutations.

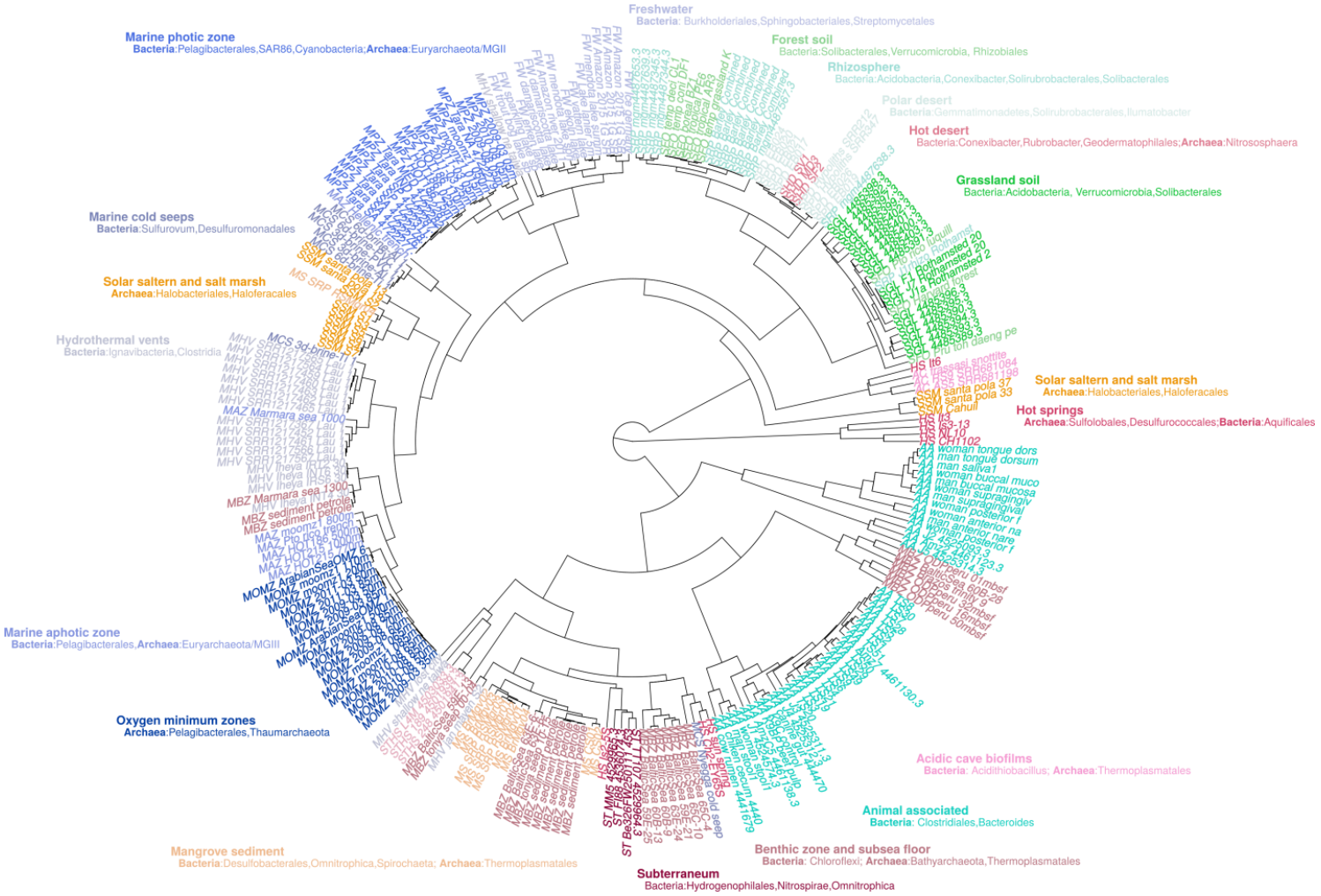


Figure S2

Hierarchical clustering of metagenomes based on profiles of genes encoding oxidoreductases. This plot is the circular version of Figure S1 with an alternative visualization. The colors indicate the biomes from which each metagenome was sampled. The origin of these metagenomes can be found in Table S1. Some representative taxa were included here for each biome (another representation of taxa per biome can be found in Figure S7).

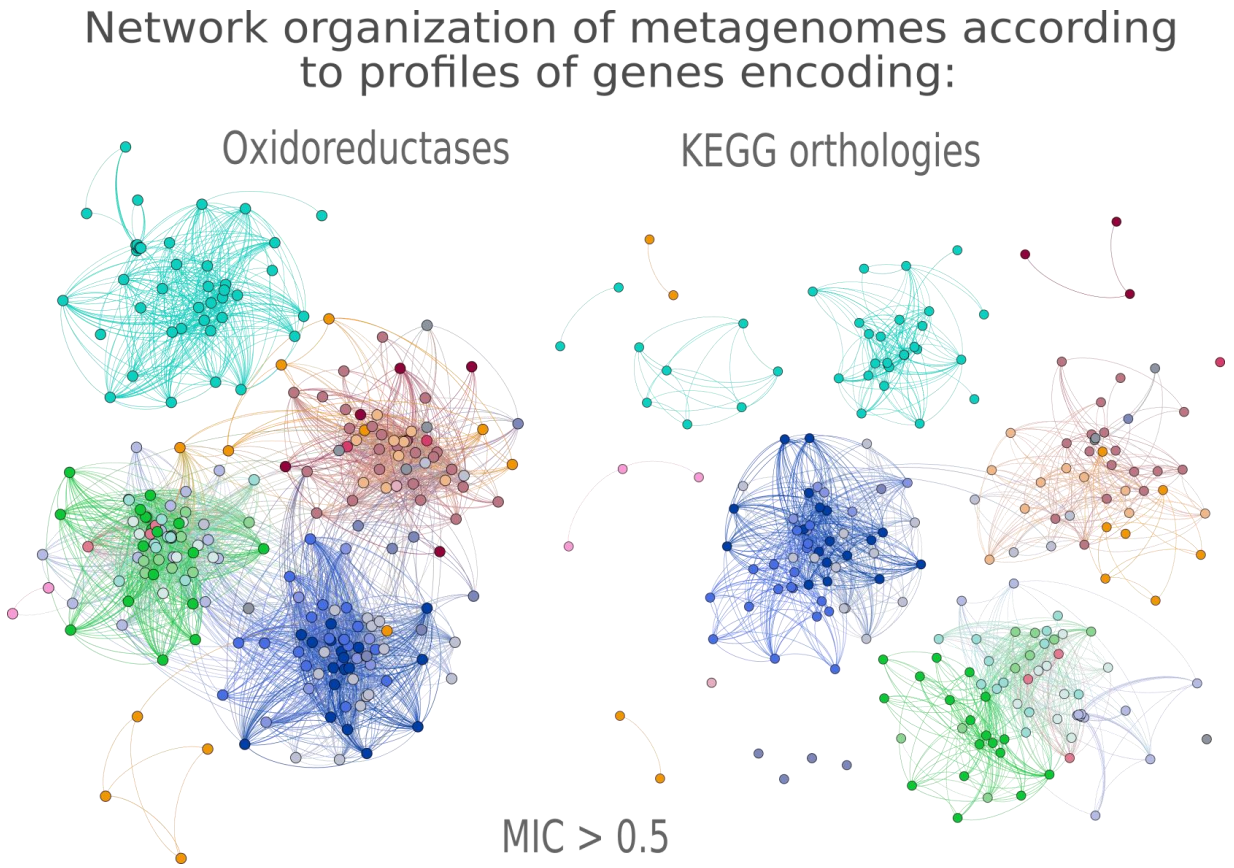


Figure S3

Network representation of the microbial metagenomes used in this study by oxidoreductase and “all KEGG orthologies” profiles of gene ranks. Nodes represent metagenomes colored according to the biome of origin (Fig. 1). Edges represent MIC correlations higher than 0.5 in both cases. Networks were rendered in the Gephi software with the OpenOrd layout algorithm, which it is aimed at better distinguish clusters. Note how the network associated with KEGG orthologies (right) has better separation than the network associated with oxidoreductases (left). However, the latter is superior in the cohesion of the clusters. These effects are the reflection of the numbers in Table 1 (third and fourth columns).

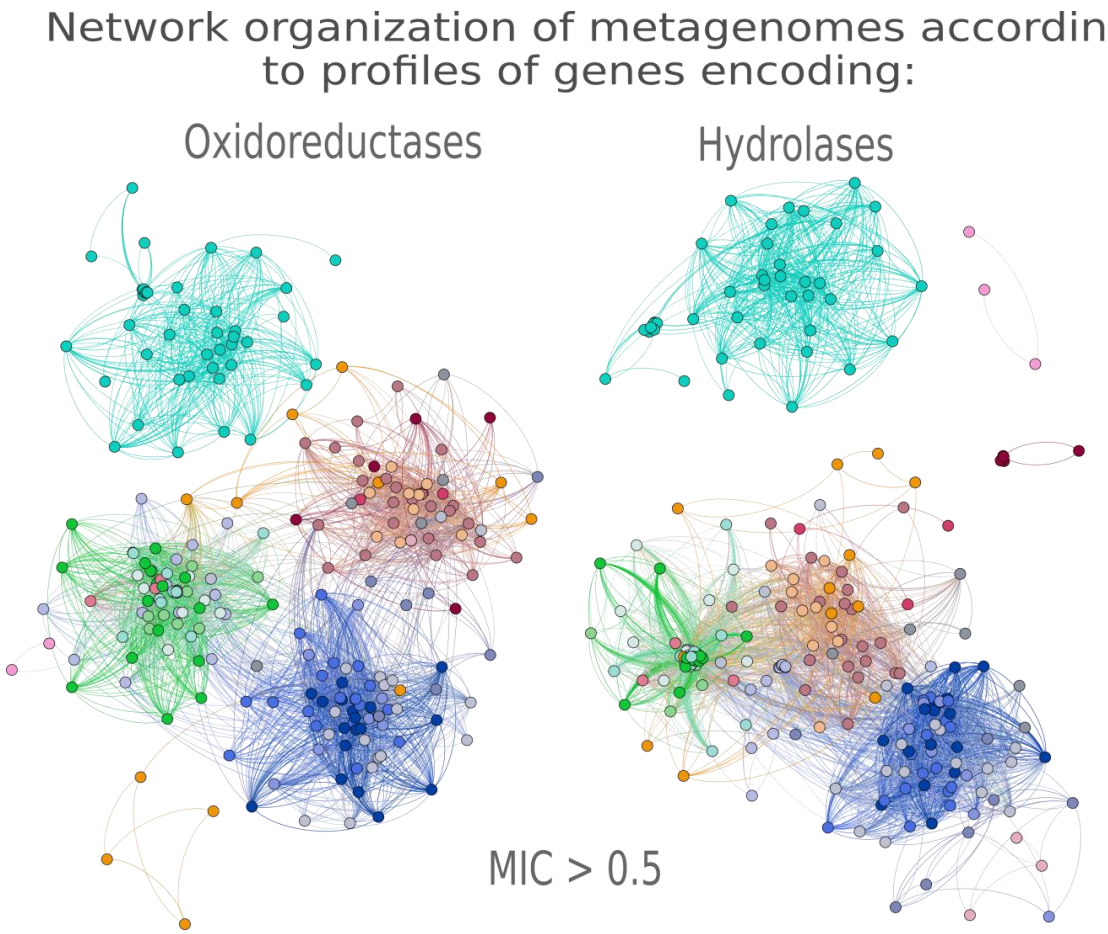


Figure S4

Network representation of the microbial metagenomes used in this study by oxidoreductase and hydrolase profiles of gene ranks. Profiles of hydrolases achieved the second best F-statistic in the PERMANOVA analysis whose results are shown in Table 1. As in Figures 2 and S3, nodes represent metagenomes colored according to the biome of origin (Fig. 1). Edges represent MIC correlations higher than 0.5 in both cases. Networks were rendered in the Gephi software with the OpenOrd layout algorithm, which it is aimed at better distinguish clusters. Note how the network associated with hydrolases (right) has a slightly better cohesion than the network associated with oxidoreductases (left). However, the network of oxidoreductase genes profiles is better separating the different clusters. These effects are the reflection of the numbers in Table 1 (third and fourth columns).

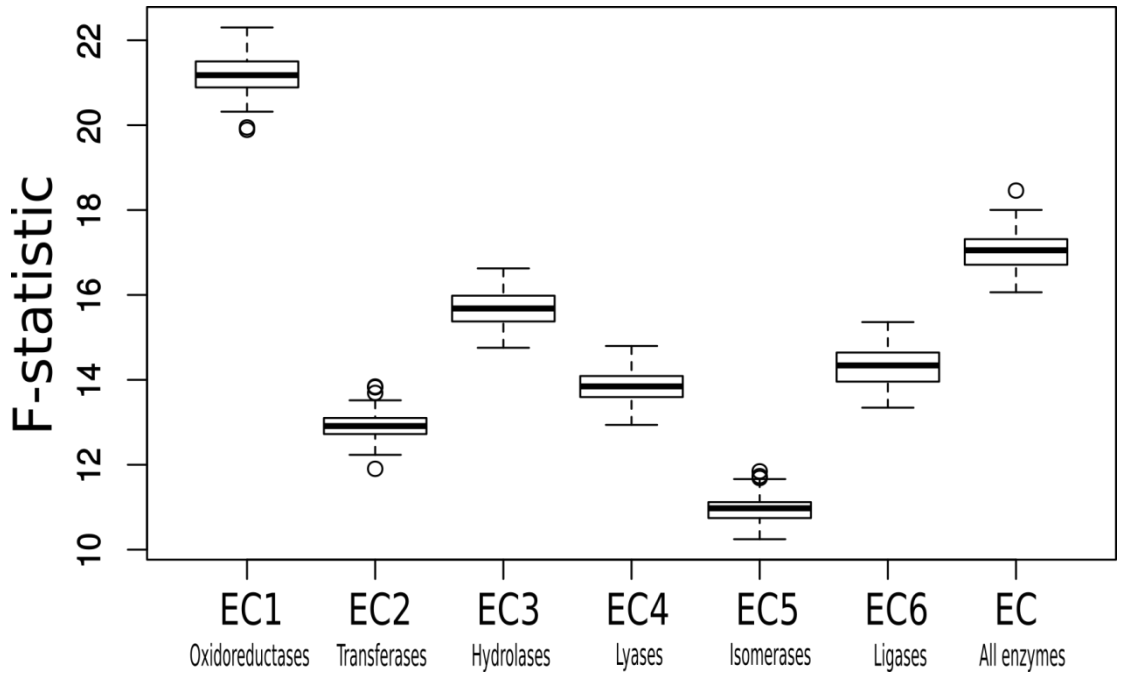


Figure S5

Boxplot with the values of F-statistic from PERMANOVA analyses for biome groups on resamples of metagenomes. To unambiguously determine which set of enzyme genes produce the highest separation of metagenomes into biome groups, PERMANOVA analyses were performed using fixed-size resamples of metagenomes over fixed-size profiles of enzyme categories. Considering that the higher the F-statistic, the more likely is to reject the null hypothesis of no differences between groups, the category of oxidoreductase genes (Enzyme Commission 1, EC1) showed the highest separation of biomes. These results confirm those from Table 1, but the difference is that here we removed variables or methods from the previous analysis that might affect the comparison, such as different profile size (last column in Table 1), different dataset sizes (metagenomes sequenced at different depths), the use of normalization methods (such as ratios, rankings or rarefaction methods) and the use of a reference database (Kegg protein sequences with assigned orthology) that might potentially include bad annotated sequences. More details of this analysis can be found in the text of this supplementary material (section Group variances analysis).

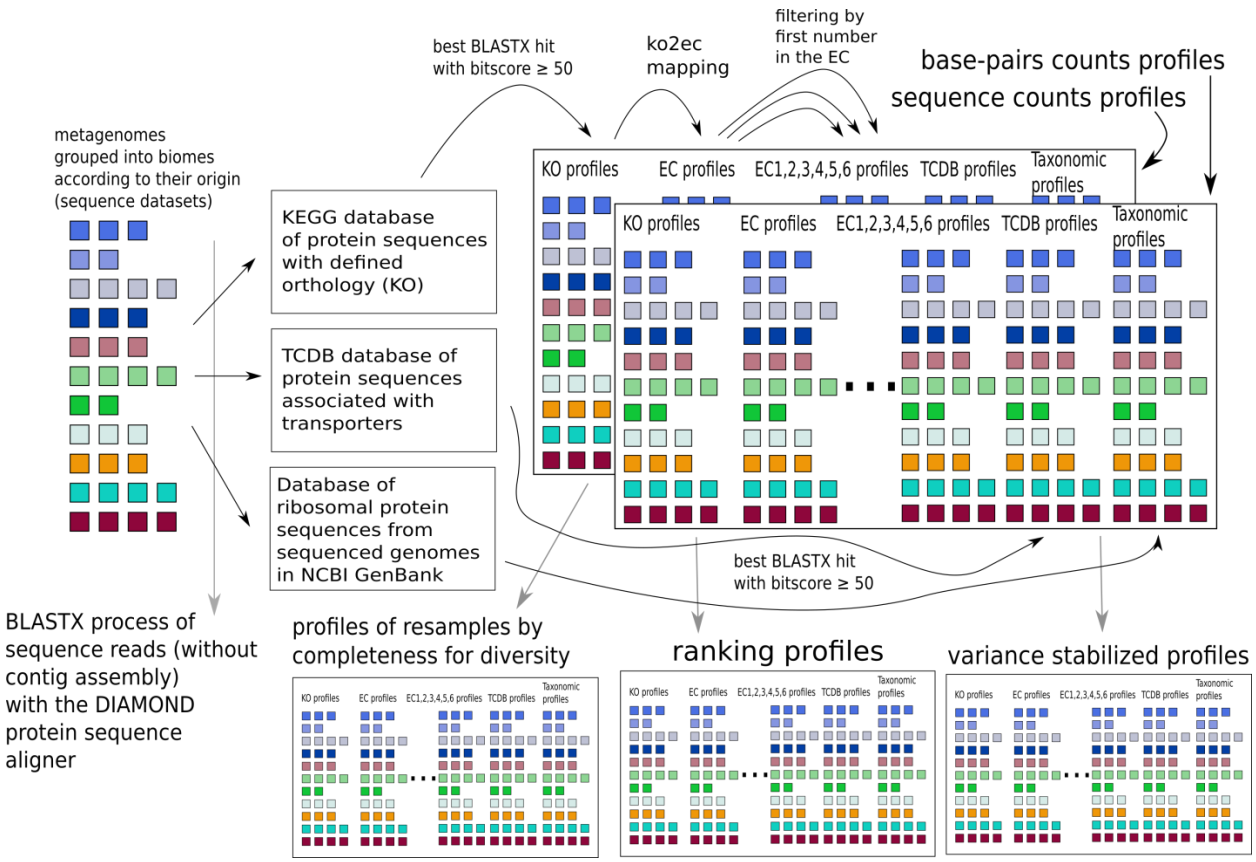


Figure S6

Sketch of the sequence analysis process of the metagenomic datasets for constructing the data profiles. Each metagenomic dataset of sequences (colored squares on the left) was assigned a biome (color) according to its origin (Figure 1 and Table S1). For simplicity, this figure does not include all the biome colors or the exact number of metagenomes per biome (for precise information about this see Figure 1 and Table S1). Every metagenomic dataset was aligned (at sequence reads level) to three reference databases (indicated in the figure), resulting in two sets of profiles, sequence counts profiles and base-pairs counts (alignment length) profiles (upper right) from which three sets of profiles were constructed: profiles of counts of resamples by completeness for diversity analyses; ranking profiles and variance stabilized profiles (bottom). We did not include in this sketch the analysis carried out with mi-faser (Figure S5) as it deviates from the normal pipeline depicted in this figure (e.g., it does not make use of bit-score information). More detailed description of this can be found in the text of this supplementary material.

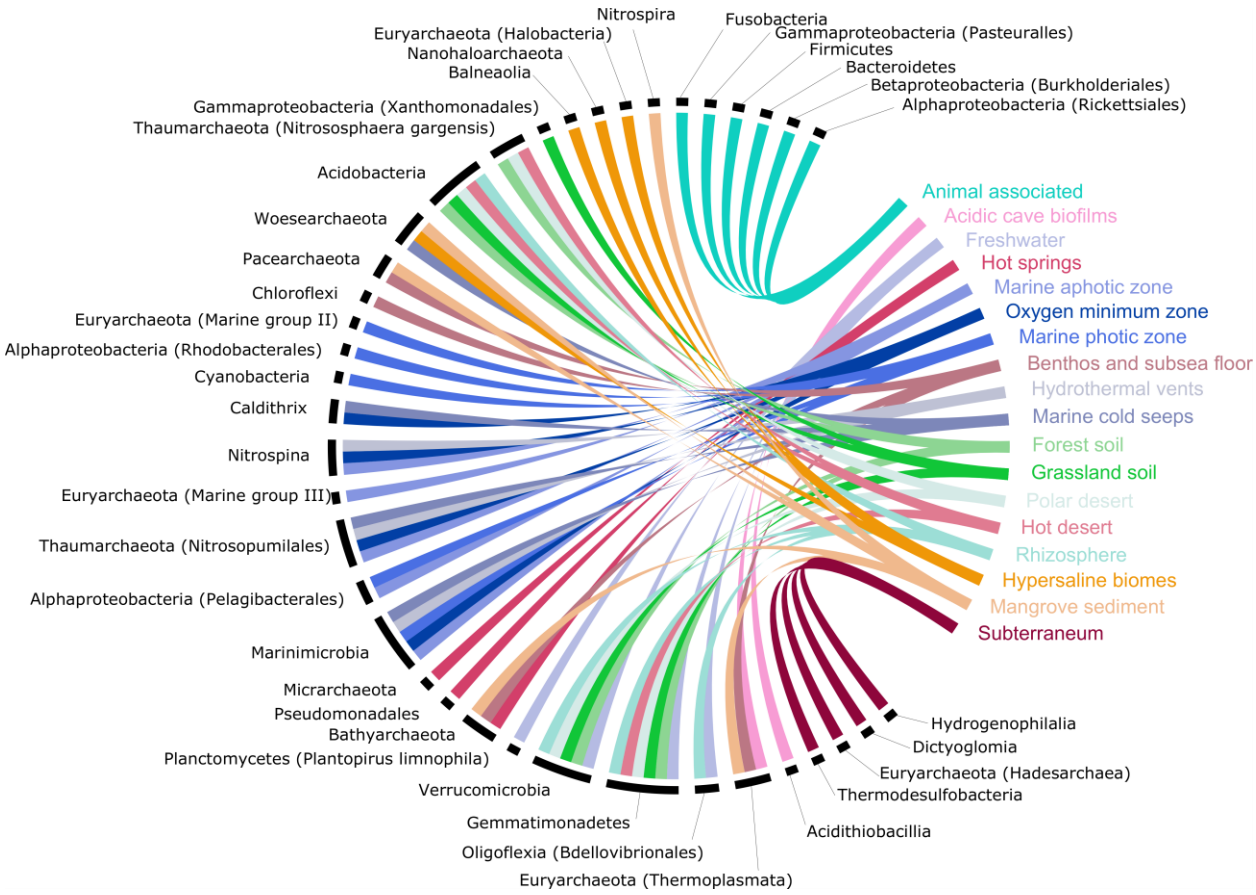


Figure S7

Representative taxa per biome. Taxonomic representatives per biomes were estimated with the relative abundances of genes encoding ribosomal proteins. Only the top represented taxa per biomes are shown in this plot. The biomes per taxon are in clockwise order, starting from the biome where the taxon was best ranked. For example, Verrucomicrobia was found best ranked in the following biomes in this order: freshwater, forest soil, grasslands, polar deserts and rizhosphere.

TABLES

The tables of this text are presented in the following pages.

Table S1

Row	Number of datasets	Biome	Dataset name prefix used in this study	Source
1	2	Acidic cave biofilms	AC_AS5, AC_RS9	Metagenomic evidence for sulfide oxidation in extremely acidic cave biofilms (*21*)
2	1	Acidic cave biofilms	AC_frassasi	Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm (*22*)
3	1	Animal associated (Canine gut)	AA_canine_gut	MG-RAST id 4444703.3
4	1	Animal associated (Chicken cecum)	AA_chicken_cecum	MG-RAST id 4440283.3
5	1	Animal associated (Cow rumen)	AA_cow_rumen	MG-RAST id 4441679.3
6	18	Animal associated (Human gut)	AA_TS	A core gut microbiome in obese and lean twins (*23*)
7	5	Animal associated (Human gut)	AA_J	Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes (*24*). (MG-RAST ids: 4524574.3, 4525093.3, 4525311.3, 4525312.3, 4525314.3)
8	3	Animal associated (Human gut)	AA_Amz	Human gut microbiome viewed across age and geography (*25*). (three representative datasets were used with MG-RAST ids: 4461123.3, 4461130.3, 4461138.3)
9	12	Animal associated (Human anterior nares, supragingival plaque, stool, buccal mucosa, tongue dorsum, posterior fornix)	AA_man, AA_woman	Strains, functions and dynamics in the expanded Human Microbiome Project (*26*). (selected samples from NCBI-SRA: SRR1804441, SRR1804072, SRR1804057, SRR1804053,SRR1804209, SRR1804011, SRR1804840, SRR1804073, SRR1803288, SRR1804442, SRR1804835, SRR1804059)
10	2	Animal associated (Canine gut)	AA_K9	Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice (*27*)

11	3	Freshwater (River)	FW_Amazon_2015	Metagenomic and metatranscriptomic inventories of the lower Amazon River (*28*). (three representative metagenomes were used with NCBI sra ids: SRR1796118, SRR1796234 and SRR1796236)
12	1	Freshwater (River)	FW_Amazon_2011	Metagenomics of the water column in the pristine upper course of the amazon river (*29*)
13	1	Freshwater (Lake)	FW_Lake_lanier	Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem (*30*)
14	1	Freshwater (Ice)	FW_Ice_germany	Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome (*31*)
15	8	Hot spring	HS	Comparative metagenomics of eight geographically remote terrestrial hot springs (*32*)
16	16	Oxygen minimum zone	MOMZ_2009,2010,2011	Initiative for the study of microbial communities in low oxygen zones. CAMERA project CAM_P_000692
17	2	Oxygen minimum zone	MOMZ_Arabian	Metagenomic analysis of nitrogen and methane cycling in the Arabian sea oxygen minimum zone (OMZ) (*33*)
18	8	Oxygen minimum zone, Marine aphotic zone, Marine photic zone	MOMZ_moomz1, MAZ_moomz1, MPZ_moomz1	Microbial metatranscriptomics in a permanent marine oxygen minimum zone (*34*)
19	4	Marine benthic zone and subseafloor	MBZ_ODPperu	Metagenomic signatures of the Peru margin subseafloor biosphere show a genetically distinct environment (*35*)
20	1	Marine benthic zone and subseafloor	MBZ_Brazos_trinity	Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes (*36*)
21	2	Marine benthic zone and subseafloor	MBZ_Marmara_sea	Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara (*37*)
22	7	Marine benthic zone and subseafloor	MBZ_sediment_petroleum	Metagenomic and geochemical characterization of pockmarked sediments overlaying the Troll petroleum reservoir in the North Sea (*38*)
23	2	Marine benthic zone and subseafloor	MBZ_tonya_seep	A metagenomic study of methanotrophic microorganisms in coal oil point seep sediments (*39*)
24	11	Marine benthic zone and subseafloor	MBZ_BalticSea	Metagenomes from deep Baltic Sea sediments reveal how past and present environmental conditions determine microbial community composition (*40*)

25	6	Marine cold seep	MCS_3d-brine, MCS_6d_brine	Synchronized dynamics of bacterial niche-specific functions during biofilm development in a cold seep brine pool (*41*)
26	1	Marine cold seep	MCS_nyegga	Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments (*42*)
27	1	Marine aphotic zone (trench)	MAZ_Hellenic_trench	Metagenomic analysis of hadopelagic microbial assemblages thriving at the deepest part of Mediterranean sea, Matapan-Vavilov Deep (*43*)
28	1	Marine aphotic zone (trench)	MAZ_Pto_rico_trench	Going deeper: metagenome of a hadopelagic microbial community (*44*)
29	6	Marine photic zone, Marine aphotic zone	MAZ_HOT, MPZ_HOT	Comparative metagenomic analysis of a microbial community residing at a depth of 4000 meters at station ALOHA in the North Pacific Subtropical Gyre (*45*)
30	9	Marine photic zone	MPZ_Tara	Structure and function of the global ocean microbiome (*46*)
31	12	Hydrothermal vent	MHV_SRR*_Lau	Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center (*47*)
32	2	Hydrothermal vent	MHV_shallow_ne_taiwan	Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system (*48*)
33	1	Hydrothermal vent	MHV_jan	Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics (*49*).
34	4	Mangrove sediment	MS_BrMgv	The microbiome of Brazilian mangrove sediments as revealed by metagenomics (*50*)
35	2	Mangrove sediment	MS_CS	Rhizosphere microbiome metagenomics of gray mangroves (*Avicennia marina*) in the Red Sea (*51*)
36	4	Mangrove rhizosphere	MS_SRP_RSMgr	Rhizosphere microbiome metagenomics of gray mangroves (*Avicennia marina*) in the Red Sea (*51*)
37	6	Polar desert	SPD_EB	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (*52*)
38	2	Polar desert	SPD_Hypoliths	Comparative metagenomic analysis reveals mechanisms for stress response in hypoliths from extreme hyperarid deserts (*53*)
39	3	Hot desert	SHD	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (*52*)

40	2	Soil forest (tropical forest)	SFO_tropical	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (*52*)
41	1	Soil forest (boreal forest)	SFO_boreal	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (*52*)
42	1	Soil forest (temperate deciduous forest)	SFO_temp_deci	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (*52*)
43	1	Soil forest (temperate coniferous forest)	SFO_coni	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (*52*)
44	1	Soil forest (temperate grassland)	SFO_temp_grassland	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (*52*)
45	1	Soil forest (temperate forest)	SFO_hardvard_forest	Community transcriptomics reveals universal patterns of protein sequence conservation in natural microbial communities (*54*) (only metagenomes were considered from this study)
46	1	Soil forest (tropical forest)	SFO_Pru_toh_daeng	Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis (*55*)
47	1	Tropical forest	SFO_Pto_rico_luquillo	Luquillo experimental forest study, Puerto Rico. NCBI SRA id SRP001743
48	3	Grassland	SGL_*_Rothamsted	Structure, fluctuation and magnitude of a natural grassland soil metagenome (*56*)
49	15	Grassland	SGL_4485	Impact of long-term N,P,K and NPK fertilization on the compositional and potential functions of the bacterial community in grassland soil (*57*)
50	1	Rhizosphere	SRP_J1rhizo_Rothamsted	Structure, fluctuation and magnitude of a natural grassland soil metagenome (*56*)
51	6	Rhizosphere	SRP_Barley	Structure and function of the bacterial root microbiota in wild and domesticated barley (*58*)
52	6	Rhizosphere	SRP_mgm4487	Functional congruence of rhizosphere microbial communities associated to leguminous tree from Brazilian semiarid region (*59*)
53	1	Solar saltern	SSM_Cahuil	Metagenome sequencing of the microbial community of a solar saltern crystallizer pond at Cahuil lagoon, Chile (*60*)
54	7	Salt desert	SSM_S[1-7]	A snapshot of microbial communities from the Kutch, one of the largest salt deserts in the World (*61*)
55	4	Solar saltern	SSM_santa_pola	New abundant microbial groups in aquatic hypersaline environments (*62*)

56	3	Subterraneum (subterranean hot springs)	STHS_	Diverse sulfur metabolisms from two subterranean sulfidic spring systems (*63*)
57	5	Subterraneum	ST_	A metagenomic window into carbon metabolism at 3km depth in Precambrian continental crust (*64*)

Table S2

Different profiles of ranking of genes separating metagenomes into biomes groups at different coverage percentages of sequence information

Category	PERMANOVA F statistic (90% coverage)	PERMANOVA F statistic (80% coverage)	PERMANOVA F statistic (70% coverage)	PERMANOVA F statistic (60% coverage)	PERMANOVA F statistic (50% coverage)	Profile size
All KEGG protein orthologies	15.96617	18.74589	20.63887	21.94649	23.12362	6,789
All enzymes	22.16813	23.06131	23.51528	23.5733	24.62584	1,826
Oxidoreductases	21.46566	24.47962	**26.29865**	**28.984**	**29.13182**	484
Transferases	23.61112	23.60262	22.85	22.35255	22.01938	541
Hydrolases	22.88798	23.70751	23.82132	24.0186	25.78586	423
Lyases	25.94873	25.4859	24.55671	24.48155	26.4212	211
Isomerases	**26.83865**	**27.80602**	25.64842	25.71113	18.75029	103
Ligases	26.51808	25.78329	22.90783	22.54884	20.60366	94
Transporters	17.50794	19.65666	21.3858	23.37097	25.19628	1,869
Taxonomy (species)	2.400858	2.582031	2.850984	3.148067	3.713205	4,011
Taxonomy (4th rank)	4.289553	4.673616	5.246327	5.441026	6.74631	365

Different sets of profiles of rankings of gene abundances were evaluated at different coverage percentages (starting from the most abundant) to determine under which of them the separation of metagenomes (microbial communities) into biome groups is most significant. The PERMANOVA statistical test (all p-values < 0.001) indicates that when we leave out at least the 30% (or 70% of coverage percentage) of the rarest genes in each category, the profiles of oxidoreductases start to be the category with higher statistically supported differences between the biomes (highest value in bold typeface, the higher the F-statistic, the more likely is to reject the null hypothesis of no differences between groups). This is most likely due to the fact that the rarest genes have the higher rank values (rankings start from the most abundant with number 1), thus producing a higher variation in the metagenomes associated with the same biomes, affecting the MIC correlation. For this reason we opted for presenting this statistics in Table 1 with the 100% of the sequence information (coverage percentage) from each metagenome but with stabilized variances of the sequences (base pairs) counts. Hence, homogeneous weights are given to abundant and rare genes. More details of this analysis can be found in the text of this supplementary material (section Group variances analysis).

Table S3

Selected enzymes (50 for each enzyme category) from the gold-standard (GS+) database in the mi-faser package. Oxidoreductases: rows 1-50, transferases: rows 51-100, hydrolases: rows 101-150, lyases: rows 151-200, isomerases: rows 201-250, ligases: rows 251-300 and all enzymes combined: rows 301-350. The names of the enzymes in this table were retrieved directly from the GS+ database without reformatting.

Row	EC number	Oxidoreductase
1	1.4.1.4	NADP-specific glutamate dehydrogenase (NADP-GDH) (EC 1.4.1.4)
2	1.11.1.21	Catalase-peroxidase (CP) (EC 1.11.1.21) (Hydroperoxidase I) (HPI) (Peroxidase/catalase)
3	1.4.7.1	Ferredoxin-dependent glutamate synthase 2 (EC 1.4.7.1) (FD-GOGAT)
4	1.2.7.3	Gapor Gor Glyceraldehyde-3-phosphate:ferredoxin oxidoreductase
5	1.6.1.2	NAD(P) transhydrogenase subunit alpha (EC 1.6.1.2) (Nicotinamide nucleotide transhydrogenase subunit alpha) (Pyridine nucleotide transhydrogenase subunit alpha)
6	1.8.1.19	SudB Sulfide dehydrogenase subunit beta
7	1.7.99.4	Periplasmic nitrate reductase (EC 1.7.99.4)
8	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase FabG (EC 1.1.1.100) (3-ketoacyl-acyl carrier protein reductase) (Beta-Ketoacyl-acyl carrier protein reductase) (Beta-ketoacyl-ACP reductase)
9	1.11.1.6	Catalase (EC 1.11.1.6)
10	1.18.6.1	nifH Fe protein of nitrogenase
11	1.8.5.4	Sulfide-quinone reductase (SQR) (EC 1.8.5.4) (Sulfide:quinone oxidoreductase)
12	1.2.4.1	Pyruvate dehydrogenase E1 component (PDH E1 component) (EC 1.2.4.1)
13	1.1.1.22	UDP-glucose 6-dehydrogenase YwqF (UDP-Glc dehydrogenase) (UDP-GlcDH) (UDPGDH) (EC 1.1.1.22)
14	1.17.1.10	Formate dehydrogenase alpha subunit FdhA
15	1.1.1.205	Inosine-5'-monophosphate dehydrogenase (IMP dehydrogenase) (IMPD) (IMPDH) (EC 1.1.1.205)

16	1.17.4.1	Ribonucleoside-diphosphate reductase subunit beta (EC 1.17.4.1) (Ribonucleotide reductase small subunit)
17	1.10.3.10	Cytochrome bo(3) ubiquinol oxidase subunit 1 (EC 1.10.3.10) (Cytochrome b562-o complex subunit I) (Cytochrome o ubiquinol oxidase subunit 1) (Cytochrome o subunit 1) (Oxidase bo(3) subunit 1) (Ubiquinol oxidase chain A) (Ubiquinol oxidase polypeptide I) (Ubiquinol oxidase subunit 1)
18	1.10.3.9	Photosystem II protein D1 1 (PSII D1 protein 1) (EC 1.10.3.9) (Photosystem II Q(B) protein 1)
19	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase subunit beta (EC 1.2.4.4) (Branched-chain alpha-ketoacid dehydrogenase E1 component subunit beta) (BCKADH E1-beta)
20	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (EC 1.2.1.12) (NAD-dependent glyceraldehyde-3-phosphate dehydrogenase)
21	1.1.1.42	Isocitrate dehydrogenase [NADP] (IDH) (EC 1.1.1.42) (IDP) (NADP(+)-specific ICDH) (Oxalosuccinate decarboxylase)
22	1.2.7.10	Oxalate oxidoreductase subunit delta (OOR delta subunit) (EC 1.2.7.10)
23	1.2.7.4	cdhA acetyl-CoA decarbonylase/synthase complex subunit alpha
24	1.8.1.9	Thioredoxin reductase (TRXR) (EC 1.8.1.9)
25	1.17.2.1	Nicotinate dehydrogenase subunit B (EC 1.17.2.1) (Nicotinate degradation protein B) (Nicotinate dehydrogenase large subunit)
26	1.2.1.11	Aspartate-semialdehyde dehydrogenase (ASA dehydrogenase) (ASADH) (EC 1.2.1.11) (Aspartate-beta-semialdehyde dehydrogenase)
27	1.97.1.12	Photosystem I P700 chlorophyll a apoprotein A1 (EC 1.97.1.12) (PsaA)
28	1.1.1.37	Malate dehydrogenase (EC 1.1.1.37)
29	1.8.1.4	Dihydrolipoyl dehydrogenase (EC 1.8.1.4) (Dihydrolipoamide dehydrogenase) (E3 component of pyruvate and 2-oxoglutarate dehydrogenases complexes) (Glycine cleavage system L protein)
30	1.1.1.85	3-isopropylmalate dehydrogenase (EC 1.1.1.85) (3-IPM-DH) (Beta-IPM dehydrogenase) (IMDH)
31	1.12.99.6	HyaB Hydrogenase-1 large chain

32	1.6.5.2	Glutathione-regulated potassium-efflux system ancillary protein KefF (Quinone oxidoreductase KefF) (EC 1.6.5.2)
33	1.15.1.2	Desulfoferrodoxin (Dfx) (EC 1.15.1.2) (Superoxide reductase) (SOR)
34	1.4.1.2	NAD-specific glutamate dehydrogenase (NAD-GDH) (EC 1.4.1.2) (NAD(+)-dependent glutamate dehydrogenase)
35	1.1.1.23	Histidinol dehydrogenase (HDH) (EC 1.1.1.23)
36	1.11.1.1	NADH peroxidase (NPXase) (Npx) (EC 1.11.1.1)
37	1.9.3.1	Cytochrome c oxidase polypeptide 2A (EC 1.9.3.1) (Cytochrome c ba(3) subunit IIA) (Cytochrome c oxidase polypeptide IIA) (Cytochrome cba3 subunit 2A)
38	1.4.1.1	Alanine dehydrogenase (EC 1.4.1.1)
39	1.1.1.271	GDP-L-fucose synthase (EC 1.1.1.271) (GDP-4-keto-6-deoxy-D-mannose-3,5-epimerase-4-reductase)
40	1.4.1.16	Meso-diaminopimelate D-dehydrogenase (DAPDH) (Meso-DAP dehydrogenase) (EC 1.4.1.16)
41	1.3.5.1	Succinate dehydrogenase flavoprotein subunit (EC 1.3.5.1)
42	1.1.1.49	Glucose-6-phosphate 1-dehydrogenase (G6PD) (EC 1.1.1.49)
43	1.3.1.9	Enoyl-[acyl-carrier-protein] reductase [NADH] FabI (ENR) (EC 1.3.1.9) (NADH-dependent enoyl-ACP reductase)
44	1.8.99.2	AprA Adenylylsulfate reductase, subunit A
45	1.2.99.2	Carbon monoxide dehydrogenase large chain (CO dehydrogenase subunit L) (CO-DH L) (EC 1.2.99.2)
46	1.17.7.3	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (flavodoxin) (EC 1.17.7.3) (1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase)
47	1.2.1.41	Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase) (GSA dehydrogenase)
48	1.3.1.14	Dihydroorotate dehydrogenase B (NAD(+)), catalytic subunit (DHOD B) (DHODase B) (DHOdehase B) (EC 1.3.1.14) (Dihydroorotate oxidase B) (Orotate reductase (NADH))
49	1.1.1.2	Aldehyde reductase Ahr (EC 1.1.1.2) (Zinc-dependent alcohol dehydrogenase Ahr)

50	1.1.1.308	Sulfopropanediol 3-dehydrogenase (EC 1.1.1.308) (2,3-dihydroxypropane-1-sulfonate 3-dehydrogenase (sulfolactate forming)) (DHPS 3-dehydrogenase (sulfolactate forming))
Row	**EC number**	**Transferase name**
51	2.7.7.6	DNA-directed RNA polymerase subunit beta' (RNAP subunit beta') (EC 2.7.7.6) (RNA polymerase subunit beta') (Transcriptase subunit beta')
52	2.3.1.54	Formate acetyltransferase 1 (EC 2.3.1.54) (Pyruvate formate-lyase 1)
53	2.7.7.8	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide phosphorylase) (PNPase)
54	2.1.1.13	Methionine synthase (EC 2.1.1.13) (5-methyltetrahydrofolate--homocysteine methyltransferase) (Methionine synthase, vitamin-B12-dependent) (MS)
55	2.7.9.2	Phosphoenolpyruvate synthase (PEP synthase) (EC 2.7.9.2) (Pyruvate, water dikinase)
56	2.7.11.1	Serine/threonine-protein kinase PknA (EC 2.7.11.1)
57	2.7.13.3	Signal-transduction histidine kinase senX3 (EC 2.7.13.3)
58	2.5.1.6	S-adenosylmethionine synthase (AdoMet synthase) (EC 2.5.1.6) (MAT) (Methionine adenosyltransferase)
59	2.1.2.1	Serine hydroxymethyltransferase (SHMT) (Serine methylase) (EC 2.1.2.1)
60	2.7.2.3	Phosphoglycerate kinase (EC 2.7.2.3)
61	2.6.1.16	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) (D-fructose-6-phosphate amidotransferase) (GFAT) (Glucosamine-6-phosphate synthase) (Hexosephosphate aminotransferase) (L-glutamine--D-fructose-6-phosphate amidotransferase)
62	2.5.1.47	Cysteine synthase A (CSase A) (EC 2.5.1.47) (O-acetylserine (thiol)-lyase A) (OAS-TL A) (O-acetylserine sulfhydrylase A)
63	2.7.7.7	pol DNA polymerase, archaea type
64	2.3.1.41	Phenolphthiocerol synthesis polyketide synthase type I Pks15/1 (Beta-ketoacyl-acyl-carrier-protein synthase I) (EC 2.3.1.41)
65	2.8.1.7	Cysteine desulfurase SufS (EC 2.8.1.7)

66	2.7.2.4	Aspartate kinase Ask_Ect (EC 2.7.2.4) (Aspartokinase)
67	2.3.3.13	2-isopropylmalate synthase (EC 2.3.3.13) (Alpha-IPM synthase) (Alpha-isopropylmalate synthase)
68	2.6.1.83	LL-diaminopimelate aminotransferase (DAP-AT) (DAP-aminotransferase) (LL-DAP-aminotransferase) (EC 2.6.1.83)
69	2.2.1.1	Transketolase 1 (TK 1) (EC 2.2.1.1)
70	2.1.2.11	3-methyl-2-oxobutanoate hydroxymethyltransferase (EC 2.1.2.11) (Ketopantoate hydroxymethyltransferase) (KPHMT)
71	2.7.1.30	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK)
72	2.5.1.19	3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase) (EPSPS)
73	2.2.1.6	Acetolactate synthase large subunit IlvG (ALS) (EC 2.2.1.6) (Acetohydroxy-acid synthase large subunit) (AHAS)
74	2.7.1.11	ATP-dependent 6-phosphofructokinase (ATP-PFK) (Phosphofructokinase) (EC 2.7.1.11) (Phosphohexokinase)
75	2.7.7.9	UTP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) (Alpha-D-glucosyl-1-phosphate uridylyltransferase) (General stress protein 33) (GSP33) (UDP-glucose pyrophosphorylase) (UDPGP) (Uridine diphosphoglucose pyrophosphorylase)
76	2.3.3.9	Malate synthase G (EC 2.3.3.9)
77	2.2.1.7	1-deoxy-D-xylulose-5-phosphate synthase (EC 2.2.1.7) (1-deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS)
78	2.3.1.182	(R)-citramalate synthase (EC 2.3.1.182) (Citramalate synthase)
79	2.3.1.179	3-oxoacyl-[acyl-carrier-protein] synthase 2 (EC 2.3.1.179) (3-oxoacyl-[acyl-carrier-protein] synthase II) (Beta-ketoacyl-ACP synthase II) (KAS II)
80	2.4.1.18	1,4-alpha-glucan branching enzyme GlgB (EC 2.4.1.18) (1,4-alpha-D-glucan:1,4-alpha-D-glucan 6-glucosyl-transferase) (Alpha-(1->4)-glucan branching enzyme) (Glycogen-branching enzyme) (BE)
81	2.7.7.24	Glucose-1-phosphate thymidylyltransferase (EC 2.7.7.24) (dTDP-glucose pyrophosphorylase) (dTDP-glucose synthase)

82	2.7.7.87	Threonylcarbamoyl-AMP synthase (TC-AMP synthase) (EC 2.7.7.87) (L-threonylcarbamoyladenylate synthase) (t(6)A37 threonylcarbamoyladenosine biosynthesis protein YwlC) (tRNA threonylcarbamoyladenosine biosynthesis protein YwlC)
83	2.7.4.6	Nucleoside diphosphate kinase (NDK) (NDP kinase) (EC 2.7.4.6) (Nucleoside-2-P kinase)
84	2.5.1.7	UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7) (Enoylpyruvate transferase) (UDP-N-acetylglucosamine enolpyruvyl transferase) (EPT)
85	2.4.2.14	Amidophosphoribosyltransferase (ATase) (EC 2.4.2.14) (Glutamine phosphoribosylpyrophosphate amidotransferase) (GPATase)
86	2.3.1.9	Acetyl-CoA acetyltransferase (EC 2.3.1.9) (Acetoacetyl-CoA thiolase) (Beta-ketothiolase PhbA)
87	2.4.1.1	Maltodextrin phosphorylase (EC 2.4.1.1)
88	2.4.99.16	Alpha-1,4-glucan:maltose-1-phosphate maltosyltransferase (GMPMT) (EC 2.4.99.16) ((1->4)-alpha-D-glucan:maltose-1-phosphate alpha-D-maltosyltransferase) ((1->4)-alpha-D-glucan:phosphate alpha-D-maltosyltransferase)
89	2.3.1.117	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (EC 2.3.1.117) (Tetrahydrodipicolinate N-succinyltransferase) (Tetrahydropicolinate succinylase)
90	2.8.1.1	Thiosulfate sulfurtransferase YnjE (EC 2.8.1.1)
91	2.7.6.5	GTP pyrophosphokinase YjbM (EC 2.7.6.5) ((p)ppGpp synthase YjbM) (Small alarmone synthase 1) (SAS 1)
92	2.7.6.1	Ribose-phosphate pyrophosphokinase (RPPK) (EC 2.7.6.1) (5-phospho-D-ribosyl alpha-1-diphosphate) (Phosphoribosyl diphosphate synthase) (Phosphoribosyl pyrophosphate synthase) (P-Rib-PP synthase) (PRPP synthase) (PRPPase)
93	2.5.1.55	2-dehydro-3-deoxyphosphooctonate aldolase (EC 2.5.1.55) (3-deoxy-D-manno-octulosonic acid 8-phosphate synthase) (KDO-8-phosphate synthase) (KDO 8-P synthase) (KDOPS) (Phospho-2-dehydro-3-deoxyoctonate aldolase)

94	2.7.4.22	Uridylate kinase (UK) (EC 2.7.4.22) (Uridine monophosphate kinase) (UMP kinase) (UMPK)
95	2.4.2.29	Queuine tRNA-ribosyltransferase (EC 2.4.2.29) (Guanine insertion enzyme) (tRNA-guanine transglycosylase)
96	2.3.3.16	Citrate synthase (EC 2.3.3.16)
97	2.1.3.3	Ornithine carbamoyltransferase (OTCase) (EC 2.1.3.3)
98	2.8.4.4	Ribosomal protein S12 methylthiotransferase RimO (S12 MTTase) (S12 methylthiotransferase) (EC 2.8.4.4) (Ribosomal protein S12 (aspartate(89)-C(3))-methylthiotransferase) (Ribosome maturation factor RimO)
99	2.8.1.8	Lipoyl synthase (EC 2.8.1.8) (Lip-syn) (LS) (Lipoate synthase) (Lipoic acid synthase) (Sulfur insertion protein LipA)
100	2.6.1.2	Glutamate-pyruvate aminotransferase AlaA (EC 2.6.1.2)
Row	**EC number**	**Hydrolase name**
101	3.6.4.12	ATP-dependent helicase/deoxyribonuclease subunit B (EC 3.1.-.-) (EC 3.6.4.12) (ATP-dependent helicase/nuclease RexB)
102	3.6.3.14	ATP synthase subunit alpha (EC 3.6.3.14) (ATP synthase F1 sector subunit alpha) (F-ATPase subunit alpha)
103	3.6.1.1	Putative K(+)-stimulated pyrophosphate-energized sodium pump (EC 3.6.1.1) (Membrane-bound sodium-translocating pyrophosphatase) (Pyrophosphate-energized inorganic pyrophosphatase) (Na(+)-PPase)
104	3.6.3.54	Copper-exporting P-type ATPase A (Protein CopA) (EC 3.6.3.54) (Cu(+)-exporting ATPase)
105	3.6.5.n1	Elongation factor 4 (EF-4) (EC 3.6.5.n1) (Ribosomal back-translocase LepA)
106	3.6.3.12	Potassium-transporting ATPase ATP-binding subunit (EC 3.6.3.12) (ATP phosphohydrolase [potassium-transporting] B chain) (Potassium-binding and translocating subunit B) (Potassium-translocating ATPase B chain)
107	3.4.21.53	Lon protease (EC 3.4.21.53) (ATP-dependent protease La)
108	3.6.4.13	Probable ATP-dependent RNA helicase YfmL (EC 3.6.4.13)

109	3.3.1.1	Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase)
110	3.2.1.23	Beta-galactosidase BgaB (Beta-gal) (EC 3.2.1.23) (Beta-Gal II)
111	3.6.3.8	Calcium-transporting ATPase (EC 3.6.3.8) (Calcium pump)
112	3.4.21.92	ATP-dependent Clp protease proteolytic subunit (EC 3.4.21.92) (Caseinolytic protease) (Endopeptidase Clp) (Heat shock protein F21.5) (Protease Ti)
113	3.2.1.3	Glucan 1,4-alpha-glucosidase SusB (EC 3.2.1.3) (Alpha-glucosidase SusB) (Glucoamylase SusB) (Starch-utilization system protein B)
114	3.1.26.12	Ribonuclease E (RNase E) (EC 3.1.26.12)
115	3.6.3.32	Carnitine transport ATP-binding protein OpuCA (EC 3.6.3.32)
116	3.1.21.2	Endonuclease 4 (EC 3.1.21.2) (Endodeoxyribonuclease IV) (Endonuclease IV)
117	3.2.1.86	6-phospho-beta-glucosidase GmuD (EC 3.2.1.86) (Aryl-phospho-beta-D-glucosidase BglD) (Glucomannan utilization protein D)
118	3.5.1.5	Urease subunit alpha (EC 3.5.1.5) (Urea amidohydrolase subunit alpha)
119	3.4.21.107	Serine endoprotease DegS (EC 3.4.21.107) (Site-1 protease DegS) (S1P protease DegS) (Site-1-type intramembrane protease)
120	3.4.11.2	Aminopeptidase N (EC 3.4.11.2) (Alpha-aminoacylpeptide hydrolase)
121	3.4.17.19	Carboxypeptidase 1 (EC 3.4.17.19) (BsuCP)
122	3.6.3.31	Spermidine/putrescine import ATP-binding protein PotA (EC 3.6.3.31)
123	3.2.1.28	Trehalase (EC 3.2.1.28) (Alpha,alpha-trehalase) (Alpha,alpha-trehalose glucohydrolase)
124	3.5.1.10	Formyltetrahydrofolate deformylase (EC 3.5.1.10) (Formyl-FH(4) hydrolase)
125	3.1.11.5	RecBCD enzyme subunit RecD (EC 3.1.11.5) (Exodeoxyribonuclease V 67 kDa polypeptide) (Exodeoxyribonuclease V alpha chain) (Exonuclease V subunit RecD) (ExoV subunit RecD) (Helicase/nuclease RecBCD subunit RecD)
126	3.4.11.18	Methionine aminopeptidase 2 (MAP 2) (MetAP 2) (EC 3.4.11.18)
127	3.1.3.11	Fructose-1,6-bisphosphatase class 1 (FBPase class 1) (EC 3.1.3.11) (D-fructose-1,6-bisphosphate 1-phosphohydrolase class 1)

128	3.5.4.13	Deoxycytidine triphosphate deaminase (dCTP deaminase) (EC 3.5.4.13)
129	3.1.13.1	Ribonuclease R (RNase R) (EC 3.1.13.1) (Protein VacB)
130	3.5.1.18	Succinyl-diaminopimelate desuccinylase (SDAP desuccinylase) (EC 3.5.1.18) (N-succinyl-LL-2,6-diaminoheptanedioate amidohydrolase)
131	3.4.19.5	Isoaspartyl peptidase (EC 3.4.19.5) (Beta-aspartyl-peptidase) (EcAIII) (Isoaspartyl dipeptidase) [Cleaved into: Isoaspartyl peptidase subunit alpha; Isoaspartyl peptidase subunit beta]
132	3.1.4.52	Putative cyclic-di-GMP phosphodiesterase YjcC (EC 3.1.4.52)
133	3.2.2.27	Uracil-DNA glycosylase (UDG) (EC 3.2.2.27)
134	3.6.3.39	Protein glycosylation K (EC 3.6.3.39)
135	3.2.1.70	Glucan 1,6-alpha-glucosidase (EC 3.2.1.70) (Dextran glucosidase) (Exo-1,6-alpha-glucosidase) (Glucodextranase)
136	3.6.3.20	sn-glycerol-3-phosphate import ATP-binding protein UgpC (EC 3.6.3.20)
137	3.4.25.2	ATP-dependent protease subunit HslV (EC 3.4.25.2) (Heat shock protein HslV)
138	3.2.1.55	Intracellular exo-alpha-(1->5)-L-arabinofuranosidase (ABF) (EC 3.2.1.55) (Intracellular arabinan exo-alpha-(1->5)-L-arabinosidase) (Arabinosidase)
139	3.6.3.27	Phosphate-import ATP-binding protein PhnC (EC 3.6.3.27)
140	3.5.4.25	GTP cyclohydrolase-2 (EC 3.5.4.25) (GTP cyclohydrolase II)
141	3.1.1.61	Chemotaxis response regulator protein-glutamate methylesterase (EC 3.1.1.61)
142	3.5.2.6	Beta-lactamase (EC 3.5.2.6) (Ambler class A beta-lactamase)
143	3.1.26.3	Ribonuclease 3 (EC 3.1.26.3) (Ribonuclease III) (RNase III)
144	3.4.11.4	Peptidase T (EC 3.4.11.4) (Aminotripeptidase) (Tripeptidase) (Tripeptide aminopeptidase)
145	3.4.11.9	Xaa-Pro aminopeptidase (EC 3.4.11.9) (Aminoacylproline aminopeptidase) (Aminopeptidase P II) (APP-II) (X-Pro aminopeptidase)
146	3.6.3.19	Maltose/maltodextrin import ATP-binding protein MalK (EC 3.6.3.19)
147	3.4.25.1	Proteasome subunit alpha (EC 3.4.25.1) (20S proteasome alpha subunit) (Proteasome core protein PrcA)
148	3.5.4.32	8-oxoguanine deaminase (EC 3.5.4.32)

149	3.4.21.88	LexA repressor (EC 3.4.21.88)
150	3.6.1.23	Deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase) (EC 3.6.1.23) (dUTP pyrophosphatase)
Row	**EC number**	**Lyase name**
151	4.2.1.11	Enolase (EC 4.2.1.11) (2-phospho-D-glycerate hydro-lyase) (2-phosphoglycerate dehydratase)
152	4.2.1.46	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)
153	4.2.1.3	Aconitate hydratase A (ACN) (Aconitase) (EC 4.2.1.3) (Iron-responsive protein-like) (IRP-like) (RNA-binding protein) (Stationary phase enzyme)
154	4.1.1.49	Phosphoenolpyruvate carboxykinase [ATP] (PCK) (PEP carboxykinase) (PEPCK) (EC 4.1.1.49)
155	4.1.1.39	Ribulose bisphosphate carboxylase large chain (RuBisCO large subunit) (EC 4.1.1.39)
156	4.3.2.2	Adenylosuccinate lyase (ASL) (EC 4.3.2.2) (Adenylosuccinase) (ASase)
157	4.2.1.20	Tryptophan synthase alpha chain (EC 4.2.1.20)
158	4.1.99.17	Phosphomethylpyrimidine synthase (EC 4.1.99.17) (Hydroxymethylpyrimidine phosphate synthase) (HMP-P synthase) (HMP-phosphate synthase) (HMPP synthase) (Thiamine biosynthesis protein ThiC)
159	4.2.1.47	GDP-mannose 4,6-dehydratase (EC 4.2.1.47) (GDP-D-mannose dehydratase)
160	4.3.3.6	Pyridoxal 5'-phosphate synthase subunit PdxS (PLP synthase subunit PdxS) (EC 4.3.3.6) (Pdx1) (Superoxide-inducible protein 7) (SOI7)
161	4.2.1.2	Fumarate hydratase class II (Fumarase C) (EC 4.2.1.2) (Iron-independent fumarase)
162	4.2.1.36	Homoaconitase small subunit (HACN) (EC 4.2.1.36) (Homoaconitate hydratase)
163	4.2.1.53	Oleate hydratase (EC 4.2.1.53) (Fatty acid double bond hydratase) (Fatty acid hydratase) (Linoleate hydratase) (Myosin cross-reactive antigen) (MCRA)

164	4.2.3.1	Threonine synthase (TS) (EC 4.2.3.1)
165	4.2.1.24	Delta-aminolevulinic acid dehydratase (ALAD) (ALADH) (EC 4.2.1.24) (Porphobilinogen synthase)
166	4.2.3.5	Chorismate synthase (CS) (EC 4.2.3.5) (5-enolpyruvylshikimate-3-phosphate phospholyase) (EPSP phospholyase)
167	4.1.3.36	1,4-dihydroxy-2-naphthoyl-CoA synthase (DHNA-CoA synthase) (EC 4.1.3.36)
168	4.1.2.13	Fructose-bisphosphate aldolase class 2 (FBP aldolase) (FBPA) (EC 4.1.2.13) (Fructose-1,6-bisphosphate aldolase) (Fructose-bisphosphate aldolase class II)
169	4.1.1.20	Diaminopimelate decarboxylase (DAP decarboxylase) (DAPDC) (EC 4.1.1.20)
170	4.1.3.1	Isocitrate lyase (ICL) (EC 4.1.3.1) (Isocitrase) (Isocitratase)
171	4.1.1.32	Phosphoenolpyruvate carboxykinase [GTP] (PEP carboxykinase) (PEPCK) (EC 4.1.1.32) (GTP-dependent phosphoenolpyruvate carboxykinase) (GTP-PEPCK)
172	4.3.2.1	Argininosuccinate lyase (ASAL) (EC 4.3.2.1) (Arginosuccinase)
173	4.1.3.27	Anthranilate synthase component 1 (AS) (ASI) (EC 4.1.3.27)
174	4.1.99.3	Deoxyribodipyrimidine photo-lyase (EC 4.1.99.3) (Cyclobutane pyrimidine dimer photolyase) (CPD photolyase) (DNA photolyase PhrA) (Photoreactivating enzyme PhrA)
175	4.1.2.4	Deoxyribose-phosphate aldolase (DERA) (EC 4.1.2.4) (2-deoxy-D-ribose 5-phosphate aldolase) (Phosphodeoxyriboaldolase) (Deoxyriboaldolase)
176	4.4.1.24	(2R)-sulfolactate sulfo-lyase subunit beta (EC 4.4.1.24) (Sulfolactate sulfo-lyase B)
177	4.6.1.12	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECDP-synthase) (MECPP-synthase) (MECPS) (EC 4.6.1.12)
178	4.1.1.37	Uroporphyrinogen decarboxylase (UPD) (URO-D) (EC 4.1.1.37)
179	4.7.1.1	Alpha-D-ribose 1-methylphosphonate 5-phosphate C-P lyase (PRPn C-P lyase) (EC 4.7.1.1)
180	4.2.99.18	Endonuclease III (EC 4.2.99.18) (DNA-(apurinic or apyrimidinic site) lyase)

181	4.4.1.21	S-ribosylhomocysteine lyase (EC 4.4.1.21) (AI-2 synthesis protein) (Autoinducer-2 production protein LuxS)
182	4.1.99.13	(6-4) photolyase (EC 4.1.99.13) ((6-4)DNA photolyase) (DNA photolyase PhrB) (Photoreactivating enzyme PhrB)
183	4.2.1.7	Altronate dehydratase (EC 4.2.1.7) (D-altronate hydro-lyase)
184	4.1.1.23	Orotidine 5'-phosphate decarboxylase (EC 4.1.1.23) (OMP decarboxylase) (OMPDCase) (OMPdecase)
185	4.2.1.115	UDP-N-acetylglucosamine 4,6-dehydratase (inverting) (EC 4.2.1.115) (Pseudaminic acid biosynthesis protein B) (UDP-GlcNAc-inverting 4,6-dehydratase)
186	4.2.3.4	3-dehydroquinate synthase (EC 4.2.3.4)
187	4.2.1.8	D-galactonate dehydratase family member Ent638_1932 (EC 4.2.1.-) (D-mannonate dehydratase) (EC 4.2.1.8)
188	4.3.1.19	L-threonine dehydratase biosynthetic IlvA (EC 4.3.1.19) (Threonine deaminase)
189	4.2.1.10	3-dehydroquinate dehydratase (3-dehydroquinase) (EC 4.2.1.10) (Type I DHQase) (Type I dehydroquinase) (DHQ1)
190	4.1.1.31	Phosphoenolpyruvate carboxylase (PEPC) (PEPCase) (EC 4.1.1.31)
191	4.2.1.135	UDP-N-acetyl-alpha-D-glucosamine C6 dehydratase (UDP-GlcNAc C6 dehydratase) (EC 4.2.1.135) (Protein glycosylation pathway protein F) (UDP-N-acetylglucosamine 4,6-dehydratase (configuration-retaining))
192	4.1.1.48	Indole-3-glycerol phosphate synthase (IGPS) (EC 4.1.1.48)
193	4.1.1.61	Phenolic acid decarboxylase subunit D (PAD) (EC 4.1.1.-) (4-hydroxybenzoate decarboxylase subunit D) (4-hydroxybenzoate DC) (EC 4.1.1.61) (Phenylacrylic acid decarboxylase subunit D) (Vanillate decarboxylase subunit D)
194	4.2.1.1	Carbonic anhydrase 1 (EC 4.2.1.1) (Carbonate dehydratase 1)
195	4.2.2.21	Chondroitin sulfate ABC exolyase (EC 4.2.2.21) (Chondroitin ABC exoeliminase) (Chondroitin ABC lyase II) (Chondroitin sulfate ABC lyase II) (ChS ABC lyase II) (Chondroitinase ABC II) (cABC II) (Exochondroitinase ABC)

196	4.4.1.5	Lactoylglutathione lyase (EC 4.4.1.5) (Aldoketomutase) (Glyoxalase I) (Glx I) (Ketone-aldehyde mutase) (Methylglyoxalase) (S-D-lactoylglutathione methylglyoxal lyase)
197	4.2.1.150	Short-chain-enoyl-CoA hydratase (EC 4.2.1.150) (3-hydroxybutyryl-CoA dehydratase) (Crotonase)
198	4.2.1.17	2,3-dehydroadipyl-CoA hydratase (EC 4.2.1.17) (Enoyl-CoA hydratase)
199	4.1.99.12	3,4-dihydroxy-2-butanone 4-phosphate synthase (DHBP synthase) (EC 4.1.99.12)
200	4.3.1.1	Aspartate ammonia-lyase (Aspartase) (EC 4.3.1.1)
Row	**EC number**	**Isomerase name**
201	5.99.1.3	DNA gyrase subunit A (EC 5.99.1.3)
202	5.99.1.2	DNA topoisomerase 1 (EC 5.99.1.2) (DNA topoisomerase I) (Omega-protein) (Relaxing enzyme) (Swivelase) (Untwisting enzyme)
203	5.3.1.9	Glucose-6-phosphate isomerase (GPI) (EC 5.3.1.9) (Phosphoglucose isomerase) (PGI) (Phosphohexose isomerase) (PHI)
204	5.3.1.23	Methylthioribose-1-phosphate isomerase (M1Pi) (MTR-1-P isomerase) (EC 5.3.1.23) (S-methyl-5-thioribose-1-phosphate isomerase)
205	5.4.99.2	Methylmalonyl-CoA mutase (MCM) (EC 5.4.99.2)
206	5.4.2.10	Phosphoglucosamine mutase (EC 5.4.2.10)
207	5.3.1.13	Arabinose 5-phosphate isomerase GutQ (API) (G-API) (EC 5.3.1.13) (Phosphosugar aldol-ketol isomerase)
208	5.4.2.12	2,3-bisphosphoglycerate-independent phosphoglycerate mutase (BPG-independent PGAM) (Phosphoglyceromutase) (iPGM) (EC 5.4.2.12)
209	5.4.3.8	Glutamate-1-semialdehyde 2,1-aminomutase (GSA) (EC 5.4.3.8) (Glutamate-1-semialdehyde aminotransferase) (GSA-AT)
210	5.2.1.8	FKBP-type peptidyl-prolyl cis-trans isomerase SlyD (PPIase) (EC 5.2.1.8) (Metallochaperone SlyD)
211	5.1.3.23	UDP-2,3-diacetamido-2,3-dideoxy-D-glucuronate 2-epimerase (UDP-alpha-D-GlcNAc3NAcA 2-epimerase) (EC 5.1.3.23) (UDP-2,3-diacetamido-2,3-dideoxy-alpha-D-glucuronic acid 2-epimerase)

212	5.4.2.2	Phosphoglucomutase (PGM) (EC 5.4.2.2) (Alpha-phosphoglucomutase) (Glucose phosphomutase)
213	5.1.3.2	UDP-glucose 4-epimerase (EC 5.1.3.2) (UDP-galactose 4-epimerase) (Uridine diphosphate galactose 4-epimerase)
214	5.4.2.11	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (BPG-dependent PGAM) (PGAM) (Phosphoglyceromutase) (dPGM) (EC 5.4.2.11)
215	5.3.1.1	Triosephosphate isomerase (TIM) (EC 5.3.1.1) (Triose-phosphate isomerase)
216	5.1.3.1	Ribulose-phosphate 3-epimerase (EC 5.1.3.1) (Pentose-5-phosphate 3-epimerase) (PPE) (R5P3E)
217	5.4.99.9	UDP-galactopyranose mutase (UGM) (EC 5.4.99.9) (UDP-GALP mutase) (Uridine 5-diphosphate galactopyranose mutase)
218	5.1.3.13	dTDP-4-dehydrorhamnose 3,5-epimerase (EC 5.1.3.13) (Thymidine diphospho-4-keto-rhamnose 3,5-epimerase) (dTDP-4-keto-6-deoxyglucose 3,5-epimerase) (dTDP-6-deoxy-D-xylo-4-hexulose 3,5-epimerase) (dTDP-L-rhamnose synthase)
219	5.1.3.20	ADP-L-glycero-D-manno-heptose-6-epimerase (EC 5.1.3.20) (ADP-L-glycero-beta-D-manno-heptose-6-epimerase) (ADP-glyceromanno-heptose 6-epimerase) (ADP-hep 6-epimerase) (AGME)
220	5.4.99.18	N5-carboxyaminoimidazole ribonucleotide mutase (N5-CAIR mutase) (EC 5.4.99.18) (5-(carboxyamino)imidazole ribonucleotide mutase)
221	5.1.3.4	L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4) (Phosphoribulose isomerase)
222	5.3.1.6	Ribose-5-phosphate isomerase A (EC 5.3.1.6) (Phosphoriboisomerase A) (PRI)
223	5.1.3.14	UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14) (Bacteriophage N4 adsorption protein C) (UDP-GlcNAc-2-epimerase)
224	5.1.1.3	Glutamate racemase (EC 5.1.1.3)
225	5.4.99.22	Ribosomal large subunit pseudouridine synthase B (EC 5.4.99.22) (23S rRNA pseudouridine(2605) synthase) (rRNA pseudouridylate synthase B) (rRNA-uridine isomerase B)

226	5.4.99.25	tRNA pseudouridine synthase B (EC 5.4.99.25) (Protein p35) (tRNA pseudouridine(55) synthase) (Psi55 synthase) (tRNA pseudouridylate synthase) (tRNA-uridine isomerase)
227	5.1.1.7	Diaminopimelate epimerase (DAP epimerase) (EC 5.1.1.7)
228	5.3.1.5	Xylose isomerase (EC 5.3.1.5)
229	5.4.99.23	Ribosomal large subunit pseudouridine synthase D (EC 5.4.99.23) (23S rRNA pseudouridine(1911/1915/1917) synthase) (rRNA pseudouridylate synthase D) (rRNA-uridine isomerase D)
230	5.1.1.8	4-hydroxyproline 2-epimerase (4Hyp 2-epimerase) (4HypE) (EC 5.1.1.8)
231	5.1.1.1	Alanine racemase, catabolic (EC 5.1.1.1)
232	5.4.4.3	3-hydroxylaminophenol mutase (3HAP mutase) (EC 5.4.4.3) (3-(hydroxyamino)phenol mutase)
233	5.3.1.14	L-rhamnose isomerase (EC 5.3.1.14)
234	5.3.1.28	Phosphoheptose isomerase (EC 5.3.1.28) (Sedoheptulose 7-phosphate isomerase)
235	5.4.99.21	Ribosomal large subunit pseudouridine synthase F (EC 5.4.99.21) (23S rRNA pseudouridine(2604) synthase) (rRNA pseudouridylate synthase F) (rRNA-uridine isomerase F)
236	5.4.1.3	2-methylfumaryl-CoA isomerase (EC 5.4.1.3)
237	5.3.1.24	N-(5'-phosphoribosyl)anthranilate isomerase (PRAI) (EC 5.3.1.24)
238	5.4.99.12	tRNA pseudouridine synthase A (EC 5.4.99.12) (tRNA pseudouridine(38-40) synthase) (tRNA pseudouridylate synthase I) (PSU-I) (tRNA-uridine isomerase I)
239	5.4.4.2	Isochorismate synthase EntC (EC 5.4.4.2) (Isochorismate mutase)
240	5.3.3.2	Isopentenyl-diphosphate Delta-isomerase (IPP isomerase) (EC 5.3.3.2) (IPP:DMAPP isomerase) (Isopentenyl pyrophosphate isomerase)
241	5.4.3.2	L-lysine 2,3-aminomutase (LAM) (EC 5.4.3.2) (KAM)
242	5.1.1.20	L-Ala-D/L-Glu epimerase (AE epimerase) (AEE) (EC 5.1.1.20)
243	5.4.99.20	Ribosomal large subunit pseudouridine synthase E (EC 5.4.99.20) (rRNA pseudouridylate synthase E) (rRNA-uridine isomerase E)
244	5.3.1.17	4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase (EC 5.3.1.17) (5-keto-4-deoxyuronate isomerase) (DKI isomerase)

245	5.4.99.19	Ribosomal small subunit pseudouridine synthase A (EC 5.4.99.19) (16S pseudouridylate 516 synthase) (16S rRNA pseudouridine(516) synthase) (rRNA pseudouridylate synthase A) (rRNA-uridine isomerase A)
246	5.3.2.5	2,3-diketo-5-methylthiopentyl-1-phosphate enolase (DK-MTP-1-P enolase) (EC 5.3.2.5) (RuBisCO-like protein) (RLP)
247	5.3.3.18	1,2-epoxyphenylacetyl-CoA isomerase (EC 5.3.3.18)
248	5.4.99.26	tRNA pseudouridine synthase C (EC 5.4.99.26) (tRNA pseudouridine(65) synthase) (tRNA pseudouridylate synthase C) (tRNA-uridine isomerase C)
249	5.3.1.8	Probable mannose-6-phosphate isomerase GmuF (EC 5.3.1.8) (Glucomannan utilization protein F) (Phosphohexomutase) (Phosphomannose isomerase) (PMI)
250	5.1.3.25	dTDP-L-rhamnose 4-epimerase (EC 5.1.3.25)
Row	**EC number**	**Ligase name**
251	6.3.5.2	GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (GMP synthetase) (GMPS) (Glutamine amidotransferase)
252	6.3.5.3	Phosphoribosylformylglycinamidine synthase subunit PurL (FGAM synthase) (EC 6.3.5.3) (Formylglycinamide ribonucleotide amidotransferase subunit II) (FGAR amidotransferase II) (FGAR-AT II) (Glutamine amidotransferase PurL) (Phosphoribosylformylglycinamidine synthase subunit II)
253	6.2.1.1	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)
254	6.4.1.1	PycA pyruvate carboxylase subunit A
255	6.2.1.5	Succinyl-CoA ligase [ADP-forming] subunit beta (EC 6.2.1.5) (Succinyl-CoA synthetase subunit beta) (SCS-beta)
256	6.4.1.2	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (ACCase subunit beta) (Acetyl-CoA carboxylase carboxyltransferase subunit beta) (EC 6.4.1.2)
257	6.1.1.6	Lysine--tRNA ligase (EC 6.1.1.6) (Lysyl-tRNA synthetase) (LysRS)
258	6.1.1.7	Alanine--tRNA ligase (EC 6.1.1.7) (Alanyl-tRNA synthetase) (AlaRS)

259	6.3.5.7	Glutamyl-tRNA(Gln) amidotransferase subunit A (Glu-ADT subunit A) (EC 6.3.5.7)
260	6.4.1.3	Probable propionyl-CoA carboxylase beta chain 5 (PCCase) (EC 6.4.1.3) (Propanoyl-CoA:carbon dioxide ligase)
261	6.1.1.10	Methionine--tRNA ligase (EC 6.1.1.10) (Methionyl-tRNA synthetase) (MetRS)
262	6.1.1.3	Threonine--tRNA ligase (EC 6.1.1.3) (Threonyl-tRNA synthetase) (ThrRS)
263	6.5.1.2	DNA ligase A (EC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD(+)])
264	6.1.1.11	Serine--tRNA ligase (EC 6.1.1.11) (Seryl-tRNA synthetase) (SerRS) (Seryl-tRNA(Ser/Sec) synthetase)
265	6.1.1.18	Glutamine--tRNA ligase (EC 6.1.1.18) (Glutaminyl-tRNA synthetase) (GlnRS)
266	6.3.4.4	Adenylosuccinate synthetase (AMPSase) (AdSS) (EC 6.3.4.4) (IMP--aspartate ligase)
267	6.1.1.17	Glutamate--tRNA ligase (EC 6.1.1.17) (Glutamyl-tRNA synthetase) (GluRS)
268	6.1.1.20	Phenylalanine--tRNA ligase alpha subunit (EC 6.1.1.20) (Phenylalanyl-tRNA synthetase alpha subunit) (PheRS)
269	6.3.4.5	Argininosuccinate synthase (EC 6.3.4.5) (Citrulline--aspartate ligase)
270	6.5.1.1	DNA ligase C1 (EC 6.5.1.1) (Polydeoxyribonucleotide synthase [ATP])
271	6.1.1.2	Tryptophan--tRNA ligase (EC 6.1.1.2) (Tryptophanyl-tRNA synthetase) (TrpRS)
272	6.1.1.23	Aspartate--tRNA(Asp/Asn) ligase (EC 6.1.1.23) (Aspartyl-tRNA synthetase) (AspRS) (Non-discriminating aspartyl-tRNA synthetase) (ND-AspRS)
273	6.1.1.1	Tyrosine--tRNA ligase (EC 6.1.1.1) (Tyrosyl-tRNA synthetase) (TyrRS)
274	6.3.1.2	Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase)
275	6.1.1.5	Isoleucine--tRNA ligase (EC 6.1.1.5) (Isoleucyl-tRNA synthetase) (IleRS)
276	6.3.5.5	Carbamoyl-phosphate synthase small chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase glutamine chain)
277	6.3.2.1	Pantothenate synthetase (PS) (EC 6.3.2.1) (Pantoate--beta-alanine ligase) (Pantoate-activating enzyme)

278	6.3.2.n2	Pup--protein ligase (EC 6.3.2.n2) (Proteasome accessory factor A) (Pup-conjugating enzyme)
279	6.2.1.3	Long-chain-fatty-acid--CoA ligase FadD15 (FACL) (EC 6.2.1.3) (Acyl-CoA synthetase)
280	6.3.2.45	UDP-N-acetylmuramate--L-alanyl-gamma-D-glutamyl-meso-2,6-diaminoheptandioate ligase (EC 6.3.2.45) (Murein peptide ligase) (UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase)
281	6.1.1.15	Proline--tRNA ligase (EC 6.1.1.15) (Prolyl-tRNA synthetase) (ProRS)
282	6.3.2.8	UDP-N-acetylmuramate--L-alanine ligase (EC 6.3.2.8) (UDP-N-acetylmuramoyl-L-alanine synthetase)
283	6.2.1.30	Phenylacetate-coenzyme A ligase (EC 6.2.1.30) (Phenylacetyl-CoA ligase) (PA-CoA ligase)
284	6.3.2.4	D-alanine--D-alanine ligase B (EC 6.3.2.4) (D-Ala-D-Ala ligase B) (D-alanylalanine synthetase B)
285	6.3.5.4	Asparagine synthetase B [glutamine-hydrolyzing] (AS-B) (EC 6.3.5.4)
286	6.3.4.18	N5-carboxyaminoimidazole ribonucleotide synthase (N5-CAIR synthase) (EC 6.3.4.18) (5-(carboxyamino)imidazole ribonucleotide synthetase)
287	6.1.1.19	Arginine--tRNA ligase (EC 6.1.1.19) (Arginyl-tRNA synthetase) (ArgRS)
288	6.3.1.1	Aspartate--ammonia ligase (EC 6.3.1.1) (Asparagine synthetase A)
289	6.3.2.13	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase (EC 6.3.2.13) (Meso-A2pm-adding enzyme) (Meso-diaminopimelate-adding enzyme) (UDP-MurNAc-L-Ala-D-Glu:meso-diaminopimelate ligase) (UDP-MurNAc-tripeptide synthetase) (UDP-N-acetylmuramyl-tripeptide synthetase)
290	6.4.1.6	Acetone carboxylase gamma subunit (EC 6.4.1.6)
291	6.6.1.1	Magnesium-chelatase 38 kDa subunit (EC 6.6.1.1) (Mg-protoporphyrin IX chelatase)
292	6.3.2.43	Alpha-aminoadipate--LysW ligase LysX (AAA--LysW ligase LysX) (EC 6.3.2.43)
293	6.1.1.12	Aspartate--tRNA ligase (EC 6.1.1.12) (Aspartyl-tRNA synthetase) (AspRS)

294	6.3.2.9	UDP-N-acetylmuramoylalanine--D-glutamate ligase (EC 6.3.2.9) (D-glutamic acid-adding enzyme) (UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase)
295	6.2.1.41	3-[(3aS,4S,7aS)-7a-methyl-1,5-dioxo-octahydro-1H-inden-4-yl]propanoyl:CoA ligase (HIP:CoA ligase) (EC 6.2.1.41)
296	6.3.2.3	Glutathione synthetase (EC 6.3.2.3) (GSH synthetase) (GSH-S) (GSHase) (Glutathione synthase)
297	6.2.1.44	3-methylmercaptopropionyl-CoA ligase (MMPA-CoA ligase) (EC 6.2.1.44) (Acyl-CoA ligase)
298	6.3.4.21	Nicotinate phosphoribosyltransferase (NAPRTase) (EC 6.3.4.21)
299	6.3.2.10	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase (EC 6.3.2.10) (D-alanyl-D-alanine-adding enzyme) (UDP-MurNAc-pentapeptide synthetase)
300	6.3.4.20	7-cyano-7-deazaguanine synthase (EC 6.3.4.20) (7-cyano-7-carbaguanine synthase) (PreQ(0) synthase) (Queuosine biosynthesis protein QueC)
Row	**EC number**	**Enzyme name**
301	2.7.7.6	DNA-directed RNA polymerase subunit beta' (RNAP subunit beta') (EC 2.7.7.6) (RNA polymerase subunit beta') (Transcriptase subunit beta')
302	5.99.1.3	DNA gyrase subunit A (EC 5.99.1.3)
303	6.2.1.1	Acetyl-coenzyme A synthetase (AcCoA synthetase) (Acs) (EC 6.2.1.1) (Acetate--CoA ligase) (Acyl-activating enzyme)
304	2.7.11.1	Serine/threonine-protein kinase PknA (EC 2.7.11.1)
305	3.6.3.14	ATP synthase subunit alpha (EC 3.6.3.14) (ATP synthase F1 sector subunit alpha) (F-ATPase subunit alpha)
306	3.6.4.12	ATP-dependent helicase/deoxyribonuclease subunit B (EC 3.1.-.-) (EC 3.6.4.12) (ATP-dependent helicase/nuclease RexB)
307	3.6.1.1	Putative K(+)-stimulated pyrophosphate-energized sodium pump (EC 3.6.1.1) (Membrane-bound sodium-translocating pyrophosphatase) (Pyrophosphate-energized inorganic pyrophosphatase) (Na(+)-PPase)
308	3.6.3.54	Copper-exporting P-type ATPase A (Protein CopA) (EC 3.6.3.54) (Cu(+)-exporting ATPase)

309	1.18.6.1	nifH Fe protein of nitrogenase
310	3.6.5.n1	Elongation factor 4 (EF-4) (EC 3.6.5.n1) (Ribosomal back-translocase LepA)
311	1.11.1.21	Catalase-peroxidase (CP) (EC 1.11.1.21) (Hydroperoxidase I) (HPI) (Peroxidase/catalase)
312	1.8.1.19	SudB Sulfide dehydrogenase subunit beta
313	3.6.4.13	Probable ATP-dependent RNA helicase YfmL (EC 3.6.4.13)
314	3.4.21.53	Lon protease (EC 3.4.21.53) (ATP-dependent protease La)
315	2.7.13.3	Signal-transduction histidine kinase senX3 (EC 2.7.13.3)
316	4.2.1.3	Aconitate hydratase A (ACN) (Aconitase) (EC 4.2.1.3) (Iron-responsive protein-like) (IRP-like) (RNA-binding protein) (Stationary phase enzyme)
317	4.2.1.11	Enolase (EC 4.2.1.11) (2-phospho-D-glycerate hydro-lyase) (2-phosphoglycerate dehydratase)
318	1.4.7.1	Ferredoxin-dependent glutamate synthase 2 (EC 1.4.7.1) (FD-GOGAT)
319	4.2.1.46	dTDP-glucose 4,6-dehydratase (EC 4.2.1.46)
320	6.4.1.1	PycA pyruvate carboxylase subunit A
321	2.1.1.13	Methionine synthase (EC 2.1.1.13) (5-methyltetrahydrofolate--homocysteine methyltransferase) (Methionine synthase, vitamin-B12-dependent) (MS)
322	6.2.1.5	Succinyl-CoA ligase [ADP-forming] subunit beta (EC 6.2.1.5) (Succinyl-CoA synthetase subunit beta) (SCS-beta)
323	2.7.7.8	Polyribonucleotide nucleotidyltransferase (EC 2.7.7.8) (Polynucleotide phosphorylase) (PNPase)
324	2.3.1.54	Formate acetyltransferase 1 (EC 2.3.1.54) (Pyruvate formate-lyase 1)
325	4.1.1.39	Ribulose bisphosphate carboxylase large chain (RuBisCO large subunit) (EC 4.1.1.39)
326	1.2.7.3	Gapor Gor Glyceraldehyde-3-phosphate:ferredoxin oxidoreductase
327	2.5.1.6	S-adenosylmethionine synthase (AdoMet synthase) (EC 2.5.1.6) (MAT) (Methionine adenosyltransferase)

328	1.6.1.2	NAD(P) transhydrogenase subunit alpha (EC 1.6.1.2) (Nicotinamide nucleotide transhydrogenase subunit alpha) (Pyridine nucleotide transhydrogenase subunit alpha)
329	3.6.3.12	Potassium-transporting ATPase ATP-binding subunit (EC 3.6.3.12) (ATP phosphohydrolase [potassium-transporting] B chain) (Potassium-binding and translocating subunit B) (Potassium-translocating ATPase B chain)
330	3.3.1.1	Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase)
331	2.3.1.41	Phenolphthiocerol synthesis polyketide synthase type I Pks15/1 (Beta-ketoacyl-acyl-carrier-protein synthase I) (EC 2.3.1.41)
332	4.1.99.17	Phosphomethylpyrimidine synthase (EC 4.1.99.17) (Hydroxymethylpyrimidine phosphate synthase) (HMP-P synthase) (HMP-phosphate synthase) (HMPP synthase) (Thiamine biosynthesis protein ThiC)
333	6.3.5.2	GMP synthase [glutamine-hydrolyzing] (EC 6.3.5.2) (GMP synthetase) (GMPS) (Glutamine amidotransferase)
334	1.2.4.1	Pyruvate dehydrogenase E1 component (PDH E1 component) (EC 1.2.4.1)
335	6.3.5.3	Phosphoribosylformylglycinamidine synthase subunit PurL (FGAM synthase) (EC 6.3.5.3) (Formylglycinamide ribonucleotide amidotransferase subunit II) (FGAR amidotransferase II) (FGAR-AT II) (Glutamine amidotransferase PurL) (Phosphoribosylformylglycinamidine synthase subunit II)
336	1.8.5.4	Sulfide-quinone reductase (SQR) (EC 1.8.5.4) (Sulfide:quinone oxidoreductase)
337	1.1.1.22	UDP-glucose 6-dehydrogenase YwqF (UDP-Glc dehydrogenase) (UDP-GlcDH) (UDPGDH) (EC 1.1.1.22)
338	2.7.7.7	pol DNA polymerase, archaea type

339	1.10.3.10	Cytochrome bo(3) ubiquinol oxidase subunit 1 (EC 1.10.3.10) (Cytochrome b562-o complex subunit I) (Cytochrome o ubiquinol oxidase subunit 1) (Cytochrome o subunit 1) (Oxidase bo(3) subunit 1) (Ubiquinol oxidase chain A) (Ubiquinol oxidase polypeptide I) (Ubiquinol oxidase subunit 1)
340	5.99.1.2	DNA topoisomerase 1 (EC 5.99.1.2) (DNA topoisomerase I) (Omega-protein) (Relaxing enzyme) (Swivelase) (Untwisting enzyme)
341	2.7.9.2	Phosphoenolpyruvate synthase (PEP synthase) (EC 2.7.9.2) (Pyruvate, water dikinase)
342	5.4.99.2	Methylmalonyl-CoA mutase (MCM) (EC 5.4.99.2)
343	4.2.1.47	GDP-mannose 4,6-dehydratase (EC 4.2.1.47) (GDP-D-mannose dehydratase)
344	2.6.1.16	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] (EC 2.6.1.16) (D-fructose-6-phosphate amidotransferase) (GFAT) (Glucosamine-6-phosphate synthase) (Hexosephosphate aminotransferase) (L-glutamine--D-fructose-6-phosphate amidotransferase)
345	2.6.1.83	LL-diaminopimelate aminotransferase (DAP-AT) (DAP-aminotransferase) (LL-DAP-aminotransferase) (EC 2.6.1.83)
346	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase FabG (EC 1.1.1.100) (3-ketoacyl-acyl carrier protein reductase) (Beta-Ketoacyl-acyl carrier protein reductase) (Beta-ketoacyl-ACP reductase)
347	1.11.1.6	Catalase (EC 1.11.1.6)
348	2.1.2.1	Serine hydroxymethyltransferase (SHMT) (Serine methylase) (EC 2.1.2.1)
349	2.7.2.3	Phosphoglycerate kinase (EC 2.7.2.3)
350	6.1.1.6	Lysine--tRNA ligase (EC 6.1.1.6) (Lysyl-tRNA synthetase) (LysRS)

Table S4

Top 100 oxidoreductase genes by biomes, averaged across corresponding metagenomes.

Animal associated (44 metagenomes)			
row	**Avg.rank**	**EC number**	**Oxidoreductase**
1	7.00 ± 13.38	1.17.4.2	Ribonucleoside-triphosphate reductase
2	7.02 ± 16.04	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
3	10.25 ± 9.24	1.17.4.1	Ribonucleoside-diphosphate reductase
4	14.07 ± 17.36	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
5	14.89 ± 11.42	1.1.1.205	IMP dehydrogenase
6	14.98 ± 13.45	1.3.99.22	Coproporphyrinogen dehydrogenase
7	15.70 ± 11.32	1.1.1.1	Alcohol dehydrogenase
8	16.20 ± 8.72	1.1.1.3	Homoserine dehydrogenase
9	16.30 ± 32.02	1.2.7.1	Pyruvate synthase
10	16.59 ± 20.72	1.4.1.4	Glutamate dehydrogenase (NADP(+))
11	16.98 ± 29.59	1.4.1.13	Glutamate synthase (NADPH)
12	18.41 ± 11.54	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
13	19.36 ± 18.19	1.3.5.1	Succinate dehydrogenase (quinone)
14	22.77 ± 16.48	1.8.1.9	Thioredoxin-disulfide reductase
15	28.50 ± 14.82	1.1.1.95	Phosphoglycerate dehydrogenase
16	28.52 ± 15.79	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
17	30.50 ± 26.60	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
18	32.14 ± 12.93	1.2.1.11	Aspartate-semialdehyde dehydrogenase
19	32.98 ± 42.66	1.2.7.3	2-oxoglutarate synthase
20	34.11 ± 26.78	1.1.1.22	UDP-glucose 6-dehydrogenase
21	36.42 ± 18.46	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
22	38.32 ± 19.31	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
23	39.52 ± 17.45	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
24	39.93 ± 24.04	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
25	40.16 ± 28.37	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
26	40.75 ± 24.36	1.8.1.4	Dihydrolipoyl dehydrogenase
27	40.84 ± 17.60	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
28	41.05 ± 23.03	1.1.5.3	Glycerol-3-phosphate dehydrogenase
29	42.30 ± 38.72	1.7.99.1	Hydroxylamine reductase
30	42.45 ± 17.98	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
31	42.64 ± 31.41	1.1.1.133	dTDP-4-dehydrorhamnose reductase
32	44.05 ± 16.03	1.1.1.85	3-isopropylmalate dehydrogenase
33	44.14 ± 19.39	1.1.1.25	Shikimate dehydrogenase
34	47.59 ± 26.82	1.1.1.23	Histidinol dehydrogenase

35	47.70 ± 20.83	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
36	47.77 ± 36.83	1.4.3.16	L-aspartate oxidase
37	48.84 ± 52.25	1.4.7.1	Glutamate synthase (ferredoxin)
38	49.07 ± 20.53	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
39	51.39 ± 24.41	1.11.1.15	Peroxiredoxin
40	52.16 ± 23.85	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
41	52.84 ± 31.20	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
42	53.48 ± 35.88	1.3.5.4	Fumarate reductase (quinol)
43	54.00 ± 26.07	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
44	56.49 ± 14.80	1.3.1.12	Prephenate dehydrogenase
45	56.55 ± 26.54	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
46	57.86 ± 28.51	1.3.1.14	Dihydroorotate dehydrogenase (NAD(+))
47	58.40 ± 33.51	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
48	58.67 ± 35.90	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
49	59.79 ± 35.22	1.8.1.8	Protein-disulfide reductase
50	61.36 ± 71.38	1.12.7.2	Ferredoxin hydrogenase
51	63.68 ± 21.75	1.5.1.2	Pyrroline-5-carboxylate reductase
52	63.70 ± 46.08	1.5.1.7	Saccharopine dehydrogenase (NAD(+), L-lysine-forming)
53	64.11 ± 29.02	1.2.1.2	Formate dehydrogenase
54	64.57 ± 31.32	1.1.1.27	L-lactate dehydrogenase
55	66.26 ± 43.51	1.1.1.77	Lactaldehyde reductase
56	67.16 ± 36.16	1.3.1.9	Enoyl-[acyl-carrier-protein] reductase (NADH)
57	69.00 ± 34.11	1.1.1.37	Malate dehydrogenase
58	70.95 ± 32.47	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
59	72.00 ± 66.57	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
60	72.41 ± 32.93	1.1.1.29	Glycerate dehydrogenase
61	72.77 ± 30.43	1.18.1.2	Ferredoxin--NADP(+) reductase
62	72.77 ± 41.72	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
63	73.77 ± 36.14	1.1.1.169	2-dehydropantoate 2-reductase
64	76.66 ± 27.12	1.1.1.28	D-lactate dehydrogenase
65	77.14 ± 46.68	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
66	77.53 ± 48.81	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
67	77.77 ± 24.37	1.1.1.346	2,5-didehydrogluconate reductase (2-dehydro-L-gulonate-forming)
68	78.43 ± 27.96	1.4.1.16	Diaminopimelate dehydrogenase
69	79.50 ± 20.08	1.5.1.3	Dihydrofolate reductase
70	81.37 ± 51.60	1.3.8.1	Short-chain acyl-CoA dehydrogenase
71	81.49 ± 30.43	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
72	84.80 ± 51.68	1.17.1.4	Xanthine dehydrogenase
73	85.40 ± 54.90	1.1.1.14	L-iditol 2-dehydrogenase

74	86.41 ± 24.76	1.4.1.1	Alanine dehydrogenase
75	89.59 ± 34.30	1.97.1.9	Selenate reductase
76	89.60 ± 31.66	1.15.1.1	Superoxide dismutase
77	90.62 ± 57.86	1.12.1.3	Hydrogen dehydrogenase (NADP(+))
78	90.68 ± 50.50	1.1.1.57	Fructuronate reductase
79	91.02 ± 65.74	1.1.1.58	Tagaturonate reductase
80	91.28 ± 55.58	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
81	92.00 ± 50.34	1.1.1.271	GDP-L-fucose synthase
82	92.17 ± 43.91	1.1.1.69	Gluconate 5-dehydrogenase
83	94.02 ± 32.47	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
84	95.11 ± 41.34	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
85	95.51 ± 51.94	1.5.3.1	Sarcosine oxidase
86	95.86 ± 31.23	1.2.1.70	Glutamyl-tRNA reductase
87	95.95 ± 31.67	1.20.4.1	Arsenate reductase (glutaredoxin)
88	97.35 ± 25.11	1.7.1.13	PreQ(1) synthase
89	98.72 ± 28.90	1.1.1.18	Inositol 2-dehydrogenase
90	98.89 ± 63.96	1.7.99.4	Nitrate reductase
91	99.76 ± 31.99	1.1.1.6	Glycerol dehydrogenase
92	100.30 ± 46.59	1.6.99.3	NADH dehydrogenase
93	101.98 ± 36.09	1.16.3.2	Bacterial non-heme ferritin
94	102.05 ± 30.37	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
95	102.12 ± 42.64	1.3.98.1	Dihydroorotate oxidase (fumarate)
96	103.14 ± 60.57	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
97	103.98 ± 33.77	1.13.12.16	Nitronate monooxygenase
98	105.43 ± 24.03	1.8.4.8	Phosphoadenylyl-sulfate reductase (thioredoxin)
99	106.10 ± 68.19	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
100	106.39 ± 95.96	1.7.2.5.	Nitric-oxide reductase (cytochrome c)

Acidic cave biofilms (3 metagenomes)			
row	**Avg.rank**	**EC number**	**Oxidoreductase**
101	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
102	3.00 ± 0.82	1.8.98.1	CoB--CoM heterodisulfide reductase
103	3.33 ± 1.25	1.8.5.4	Sulfide:quinone reductase
104	3.67 ± 0.47	1.17.4.1	Ribonucleoside-diphosphate reductase
105	4.00 ± 1.41	1.4.1.13	Glutamate synthase (NADPH)
106	6.00 ± 0.00	1.8.1.4	Dihydrolipoyl dehydrogenase
107	7.00 ± 0.00	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
108	11.33 ± 3.40	1.11.1.21	Catalase peroxidase
109	12.00 ± 3.56	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
110	12.33 ± 1.70	1.7.1.15	Nitrite reductase (NADH)
111	12.67 ± 5.19	1.1.1.22	UDP-glucose 6-dehydrogenase

112	14.00 ± 5.35	1.7.99.4	Nitrate reductase
113	15.00 ± 4.97	1.11.1.15	Peroxiredoxin
114	15.67 ± 3.40	1.1.1.205	IMP dehydrogenase
115	16.00 ± 1.63	1.3.99.22	Coproporphyrinogen dehydrogenase
116	20.33 ± 7.93	1.1.1.28	D-lactate dehydrogenase
117	20.67 ± 9.84	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
118	21.33 ± 8.96	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
119	21.67 ± 2.36	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
120	24.33 ± 6.60	1.8.1.8	Protein-disulfide reductase
121	24.33 ± 3.40	1.17.1.1	CDP-4-dehydro-6-deoxyglucose reductase
122	26.33 ± 2.05	1.8.1.9	Thioredoxin-disulfide reductase
123	28.00 ± 14.31	1.6.99.3	NADH dehydrogenase
124	28.33 ± 8.96	1.1.1.3	Homoserine dehydrogenase
125	28.33 ± 6.13	1.1.1.23	Histidinol dehydrogenase
126	30.33 ± 4.78	1.4.3.16	L-aspartate oxidase
127	31.00 ± 2.94	1.8.1.7	Glutathione-disulfide reductase
128	31.67 ± 21.55	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
129	32.33 ± 3.09	1.2.1.70	Glutamyl-tRNA reductase
130	34.67 ± 8.50	1.2.1.11	Aspartate-semialdehyde dehydrogenase
131	34.67 ± 15.84	1.1.3.15	(S)-2-hydroxy-acid oxidase
132	35.00 ± 4.08	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
133	35.67 ± 14.97	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
134	35.67 ± 12.50	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
135	37.33 ± 9.46	1.1.1.85	3-isopropylmalate dehydrogenase
136	40.00 ± 7.87	1.14.13.81	Magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase
137	40.67 ± 4.99	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
138	43.00 ± 12.08	1.1.5.3	Glycerol-3-phosphate dehydrogenase
139	43.67 ± 3.40	1.4.99.1	1.4.99.6
140	46.33 ± 4.11	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
141	47.00 ± 14.45	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
142	47.33 ± 3.30	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
143	48.67 ± 15.58	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
144	48.67 ± 16.54	1.2.7.3	2-oxoglutarate synthase
145	49.67 ± 9.67	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
146	50.00 ± 27.29	1.2.1.2	Formate dehydrogenase
147	50.67 ± 7.59	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
148	51.50 ± 30.50	1.8.5.2	Thiosulfate dehydrogenase (quinone)
149	52.67 ± 9.84	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)

150	54.00 ± 16.57	1.9.3.1	Cytochrome-c oxidase
151	54.33 ± 13.60	1.3.1.1	Dihydrouracil dehydrogenase (NAD(+))
152	55.33 ± 56.41	1.5.3.1	Sarcosine oxidase
153	57.00 ± 15.25	1.3.3.3	Coproporphyrinogen oxidase
154	58.33 ± 8.99	1.16.3.1	Ferroxidase
155	58.67 ± 20.85	1.1.1.1	Alcohol dehydrogenase
156	59.67 ± 9.67	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
157	60.33 ± 4.11	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
158	60.33 ± 15.76	1.3.5.1	Succinate dehydrogenase (quinone)
159	60.33 ± 5.25	1.1.1.219	Dihydrokaempferol 4-reductase
160	60.67 ± 28.55	1.3.3.4	Protoporphyrinogen oxidase
161	61.33 ± 8.22	1.4.3.19	Glycine oxidase
162	61.33 ± 12.50	1.3.5.2	Dihydroorotate dehydrogenase (quinone)
163	62.00 ± 28.08	1.4.1.1	Alanine dehydrogenase
164	63.33 ± 10.62	1.15.1.1	Superoxide dismutase
165	63.67 ± 11.32	1.14.12.17	Nitric oxide dioxygenase
166	64.67 ± 11.26	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
167	65.33 ± 23.80	1.1.1.37	Malate dehydrogenase
168	65.50 ± 15.50	1.13.11.18	Persulfide dioxygenase
169	65.67 ± 14.82	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
170	66.00 ± 13.14	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
171	66.67 ± 10.87	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
172	66.67 ± 20.53	1.1.1.47	Glucose 1-dehydrogenase (NAD(P)(+))
173	70.33 ± 15.46	1.1.1.133	dTDP-4-dehydrorhamnose reductase
174	70.67 ± 0.47	1.20.4.1	Arsenate reductase (glutaredoxin)
175	73.00 ± 1.00	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
176	74.33 ± 12.66	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
177	74.33 ± 12.71	1.5.1.2	Pyrroline-5-carboxylate reductase
178	75.67 ± 10.84	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
179	79.67 ± 9.03	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
180	80.67 ± 18.91	1.8.4.12	Peptide-methionine (R)-S-oxide reductase
181	81.33 ± 2.62	1.5.1.33	Pteridine reductase
182	83.67 ± 21.55	1.1.1.25	Shikimate dehydrogenase
183	85.00 ± 27.90	1.4.3.5	Pyridoxal 5'-phosphate synthase
184	85.33 ± 7.59	1.6.5.5	NADPH:quinone reductase
185	86.00 ± 2.16	1.14.11.33	DNA oxidative demethylase
186	86.33 ± 13.42	1.17.1.4	Xanthine dehydrogenase
187	86.67 ± 4.11	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
188	86.67 ± 12.66	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
189	87.67 ± 15.11	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
190	87.67 ± 33.16	1.6.5.2	NAD(P)H dehydrogenase (quinone)
191	90.33 ± 8.58	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))

192	91.33 ± 11.61	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
193	91.67 ± 60.25	1.13.12.3	Tryptophan 2-monooxygenase
194	92.33 ± 4.92	1.1.1.271	GDP-L-fucose synthase
195	92.33 ± 20.04	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
196	95.67 ± 30.23	1.13.11.5	Homogentisate 1,2-dioxygenase
197	96.33 ± 21.30	1.3.1.12	Prephenate dehydrogenase
198	97.33 ± 4.64	1.13.11.53	Acireductone dioxygenase (Ni(2+)-requiring)
199	98.67 ± 11.12	1.6.99.1	NADPH dehydrogenase
200	98.67 ± 6.94	1.2.1.89	D-glyceraldehyde dehydrogenase (NADP(+))

Freshwater (15 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
201	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
202	2.27 ± 1.00	1.17.4.1	Ribonucleoside-diphosphate reductase
203	3.67 ± 1.19	1.4.1.13	Glutamate synthase (NADPH)
204	4.40 ± 1.70	1.9.3.1	Cytochrome-c oxidase
205	5.60 ± 1.40	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
206	6.40 ± 1.74	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
207	9.40 ± 3.34	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
208	9.67 ± 2.36	1.3.5.1	Succinate dehydrogenase (quinone)
209	9.80 ± 2.56	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
210	11.53 ± 3.96	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
211	12.20 ± 7.33	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
212	13.67 ± 3.11	1.8.1.4	Dihydrolipoyl dehydrogenase
213	14.20 ± 6.23	1.1.1.205	IMP dehydrogenase
214	14.27 ± 3.41	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
215	15.13 ± 4.22	1.2.1.2	Formate dehydrogenase
216	15.80 ± 4.21	1.1.1.95	Phosphoglycerate dehydrogenase
217	17.40 ± 3.79	1.1.1.1	Alcohol dehydrogenase
218	17.73 ± 5.88	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
219	23.00 ± 4.59	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
220	23.47 ± 7.54	1.6.5.5	NADPH:quinone reductase
221	25.53 ± 9.03	1.1.1.22	UDP-glucose 6-dehydrogenase
222	25.93 ± 13.43	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
223	26.73 ± 5.82	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
224	27.87 ± 7.19	1.1.1.3	Homoserine dehydrogenase
225	29.47 ± 8.66	1.8.1.9	Thioredoxin-disulfide reductase
226	31.40 ± 7.64	1.3.99.22	Coproporphyrinogen dehydrogenase
227	32.07 ± 10.17	1.1.99.1	Choline dehydrogenase
228	33.87 ± 6.76	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
229	35.73 ± 18.71	1.3.99.26	All-trans-zeta-carotene desaturase
230	37.00 ± 10.03	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)

231	38.93 ± 8.39	1.1.5.3	Glycerol-3-phosphate dehydrogenase
232	39.13 ± 9.41	1.2.1.11	Aspartate-semialdehyde dehydrogenase
233	39.53 ± 15.10	1.5.3.1	Sarcosine oxidase
234	40.73 ± 28.16	1.2.1.8	Betaine-aldehyde dehydrogenase
235	41.07 ± 14.71	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
236	42.40 ± 14.80	1.11.1.15	Peroxiredoxin
237	42.53 ± 16.33	1.6.99.3	NADH dehydrogenase
238	45.53 ± 15.21	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
239	45.60 ± 34.38	1.3.99.16	Isoquinoline 1-oxidoreductase
240	45.80 ± 11.09	1.4.3.16	L-aspartate oxidase
241	48.53 ± 19.28	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
242	49.33 ± 17.46	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
243	49.53 ± 13.74	1.1.1.85	3-isopropylmalate dehydrogenase
244	49.60 ± 24.05	1.11.1.21	Catalase peroxidase
245	52.53 ± 10.52	1.1.1.23	Histidinol dehydrogenase
246	53.00 ± 17.63	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
247	53.60 ± 16.55	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
248	53.80 ± 13.42	1.1.3.15	(S)-2-hydroxy-acid oxidase
249	55.87 ± 14.16	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
250	55.87 ± 23.66	1.17.1.4	Xanthine dehydrogenase
251	58.93 ± 21.57	1.1.1.37	Malate dehydrogenase
252	59.00 ± 19.18	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
253	59.53 ± 25.71	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
254	60.53 ± 25.62	1.4.7.1	Glutamate synthase (ferredoxin)
255	61.47 ± 17.69	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
256	61.53 ± 27.47	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
257	62.47 ± 27.23	1.1.1.271	GDP-L-fucose synthase
258	63.53 ± 23.06	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
259	65.20 ± 15.22	1.20.4.1	Arsenate reductase (glutaredoxin)
260	65.73 ± 13.57	1.1.1.91	Aryl-alcohol dehydrogenase (NADP(+))
261	65.73 ± 21.14	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
262	68.13 ± 26.33	1.5.3.19	4-methylaminobutanoate oxidase (formaldehyde-forming)
263	68.67 ± 19.40	1.1.2.4	D-lactate dehydrogenase (cytochrome)
264	68.93 ± 18.92	1.8.1.8	Protein-disulfide reductase
265	69.40 ± 10.76	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
266	69.73 ± 24.76	1.1.2.3	L-lactate dehydrogenase (cytochrome)
267	70.07 ± 21.77	1.14.13.22	Cyclohexanone monooxygenase
268	70.13 ± 19.48	1.15.1.1	Superoxide dismutase
269	71.53 ± 22.69	1.4.1.1	Alanine dehydrogenase
270	71.93 ± 8.41	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase

271	72.67 ± 14.39	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
272	73.07 ± 17.88	1.1.1.25	Shikimate dehydrogenase
273	73.20 ± 43.52	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
274	73.60 ± 19.37	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
275	74.47 ± 13.18	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
276	75.87 ± 40.87	1.2.7.3	2-oxoglutarate synthase
277	78.73 ± 22.94	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
278	79.40 ± 13.08	1.2.1.70	Glutamyl-tRNA reductase
279	79.60 ± 23.47	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
280	81.73 ± 19.34	1.1.1.133	dTDP-4-dehydrorhamnose reductase
281	82.33 ± 25.30	1.14.19.1	Stearoyl-CoA 9-desaturase
282	82.60 ± 34.48	1.3.7.7	Ferredoxin:protochlorophyllide reductase (ATP-dependent)
283	83.87 ± 25.34	1.3.1.12	Prephenate dehydrogenase
284	84.07 ± 23.49	1.18.1.2	Ferredoxin--NADP(+) reductase
285	84.60 ± 26.50	1.1.1.18	Inositol 2-dehydrogenase
286	86.07 ± 16.66	1.5.1.2	Pyrroline-5-carboxylate reductase
287	87.33 ± 23.13	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
288	88.27 ± 38.76	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
289	88.60 ± 21.53	1.13.12.16	Nitronate monooxygenase
290	90.53 ± 36.54	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
291	92.93 ± 28.49	1.8.1.2	Assimilatory sulfite reductase (NADPH)
292	96.40 ± 23.27	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
293	97.73 ± 23.09	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
294	97.93 ± 16.58	1.4.3.5	Pyridoxal 5'-phosphate synthase
295	98.47 ± 17.93	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
296	100.47 ± 39.73	1.1.3.6	Cholesterol oxidase
297	102.13 ± 28.21	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
298	103.40 ± 37.35	1.3.99.35	Chlorophyllide a reductase
299	105.47 ± 14.36	1.17.99.6	Epoxyqueuosine reductase
300	108.33 ± 20.85	1.3.8.1	Short-chain acyl-CoA dehydrogenase

Hot springs (8 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
301	1.25 ± 0.43	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
302	3.50 ± 1.80	1.17.4.1	Ribonucleoside-diphosphate reductase
303	7.43 ± 6.02	1.2.7.5	Aldehyde ferredoxin oxidoreductase
304	8.00 ± 4.06	1.2.7.3	2-oxoglutarate synthase
305	8.88 ± 2.57	1.2.1.2	Formate dehydrogenase
306	8.88 ± 7.25	1.8.98.1	CoB--CoM heterodisulfide reductase

307	15.38 ± 13.02	1.4.1.13	Glutamate synthase (NADPH)
308	17.25 ± 11.39	1.8.1.4	Dihydrolipoyl dehydrogenase
309	18.00 ± 5.48	1.1.1.1	Alcohol dehydrogenase
310	18.75 ± 7.77	1.1.3.15	(S)-2-hydroxy-acid oxidase
311	18.88 ± 8.52	1.3.5.1	Succinate dehydrogenase (quinone)
312	19.12 ± 5.80	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
313	21.62 ± 20.54	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
314	22.88 ± 38.82	1.2.7.1	Pyruvate synthase
315	25.12 ± 13.49	1.8.1.9	Thioredoxin-disulfide reductase
316	25.75 ± 15.45	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
317	26.38 ± 34.63	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
318	28.75 ± 8.76	1.11.1.15	Peroxiredoxin
319	31.75 ± 26.78	1.9.3.1	Cytochrome-c oxidase
320	32.62 ± 29.47	1.7.99.4	Nitrate reductase
321	34.00 ± 21.11	1.1.1.95	Phosphoglycerate dehydrogenase
322	35.12 ± 26.72	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
323	36.12 ± 28.47	1.12.99.6	Hydrogenase (acceptor)
324	37.00 ± 15.71	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
325	37.12 ± 14.26	1.5.3.1	Sarcosine oxidase
326	37.33 ± 27.34	1.2.7.4	Carbon-monoxide dehydrogenase (ferredoxin)
327	37.50 ± 16.66	1.1.1.22	UDP-glucose 6-dehydrogenase
328	38.50 ± 20.05	1.1.1.205	IMP dehydrogenase
329	41.33 ± 24.52	1.10.3.12	Menaquinol oxidase (H(+)-transporting)
330	42.25 ± 29.19	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
331	42.88 ± 13.43	1.1.1.133	dTDP-4-dehydrorhamnose reductase
332	44.25 ± 35.24	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
333	44.29 ± 7.24	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
334	44.38 ± 23.10	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
335	44.75 ± 12.17	1.1.5.3	Glycerol-3-phosphate dehydrogenase
336	46.29 ± 28.99	1.1.1.14	L-iditol 2-dehydrogenase
337	47.67 ± 42.05	1.17.4.2	Ribonucleoside-triphosphate reductase
338	47.88 ± 69.62	1.8.5.4	Sulfide:quinone reductase
339	49.00 ± 24.94	1.2.99.5	Formylmethanofuran dehydrogenase
340	49.86 ± 26.55	1.3.99.22	Coproporphyrinogen dehydrogenase
341	50.12 ± 26.59	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
342	50.62 ± 22.93	1.4.3.16	L-aspartate oxidase
343	51.00 ± 17.49	1.1.1.3	Homoserine dehydrogenase
344	52.12 ± 22.72	1.1.1.85	3-isopropylmalate dehydrogenase
345	52.88 ± 24.71	1.1.1.23	Histidinol dehydrogenase
346	54.50 ± 13.12	1.2.1.11	Aspartate-semialdehyde dehydrogenase
347	55.38 ± 43.30	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
348	57.57 ± 20.80	1.3.1.101	2,3-bis-O-geranylgeranyl-sn-glycerol 1-phosphate reductase (NAD(P)H)

349	57.75 ± 21.76	1.4.1.1	Alanine dehydrogenase
350	62.14 ± 28.75	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
351	63.50 ± 37.88	1.8.4.8	Phosphoadenylyl-sulfate reductase (thioredoxin)
352	64.00 ± 20.01	1.5.1.2	Pyrroline-5-carboxylate reductase
353	64.57 ± 19.12	1.1.1.26	Glyoxylate reductase
354	64.57 ± 44.31	1.16.1.1	Mercury(II) reductase
355	67.38 ± 23.88	1.2.1.70	Glutamyl-tRNA reductase
356	67.88 ± 37.17	1.6.99.3	NADH dehydrogenase
357	70.57 ± 21.39	1.8.99.3	Hydrogensulfite reductase
358	71.12 ± 23.72	1.15.1.1	Superoxide dismutase
359	71.25 ± 34.43	1.17.1.4	Xanthine dehydrogenase
360	71.25 ± 20.55	1.1.1.359	Aldose 1-dehydrogenase (NAD(P)(+))
361	74.12 ± 32.95	1.7.1.15	Nitrite reductase (NADH)
362	74.50 ± 19.95	1.12.7.2	Ferredoxin hydrogenase
363	74.62 ± 30.97	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
364	75.12 ± 55.14	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
365	75.20 ± 46.07	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
366	75.29 ± 16.66	1.3.1.12	Prephenate dehydrogenase
367	75.33 ± 67.06	1.2.1.89	D-glyceraldehyde dehydrogenase (NADP(+))
368	77.17 ± 23.70	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
369	77.38 ± 22.51	1.6.3.3	NADH oxidase (H(2)O(2)-forming)
370	77.50 ± 28.32	1.1.1.37	Malate dehydrogenase
371	77.75 ± 55.87	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
372	80.57 ± 14.31	1.1.1.25	Shikimate dehydrogenase
373	80.57 ± 33.70	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
374	81.80 ± 48.05	1.13.11.55	Sulfur oxygenase/reductase
375	82.14 ± 15.55	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
376	82.29 ± 54.85	1.1.1.18	Inositol 2-dehydrogenase
377	82.57 ± 38.25	1.14.13.81	Magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase
378	83.20 ± 46.95	1.3.7.8	Benzoyl-CoA reductase
379	84.33 ± 40.20	1.12.98.1	Coenzyme F420 hydrogenase
380	84.71 ± 17.11	1.8.99.2	Adenylyl-sulfate reductase
381	85.67 ± 64.34	1.12.1.2	Hydrogen dehydrogenase
382	86.33 ± 26.80	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
383	86.50 ± 12.07	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
384	87.33 ± 12.28	1.12.98.4	Sulfhydrogenase
385	88.62 ± 48.95	1.1.1.261	sn-glycerol-1-phosphate dehydrogenase
386	88.71 ± 21.68	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
387	90.33 ± 46.99	1.2.1.76	Succinate-semialdehyde dehydrogenase (acetylating)
388	90.50 ± 28.73	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))

389	90.60 ± 60.85	1.2.7.6	Glyceraldehyde-3-phosphate dehydrogenase (ferredoxin)
390	91.83 ± 41.24	1.1.1.88	Hydroxymethylglutaryl-CoA reductase
391	92.83 ± 12.59	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
392	94.29 ± 60.04	1.1.1.374	UDP-N-acetylglucosamine 3-dehydrogenase
393	94.88 ± 30.02	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
394	96.00 ± 10.61	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
395	96.25 ± 25.93	1.1.1.335	UDP-N-acetyl-2-amino-2-deoxyglucuronate dehydrogenase
396	96.50 ± 59.26	1.3.1.14	Dihydroorotate dehydrogenase (NAD(+))
397	96.80 ± 74.63	1.12.1.3	Hydrogen dehydrogenase (NADP(+))
398	97.33 ± 23.11	1.1.1.136	UDP-N-acetylglucosamine 6-dehydrogenase
399	98.00 ± 31.43	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
400	99.25 ± 60.92	1.2.1.43	Formate dehydrogenase (NADP(+))

Marine aphotic zone (7 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
401	4.14 ± 4.09	1.17.4.1	Ribonucleoside-diphosphate reductase
402	4.86 ± 0.99	1.4.1.13	Glutamate synthase (NADPH)
403	6.14 ± 5.54	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
404	6.14 ± 3.48	1.9.3.1	Cytochrome-c oxidase
405	9.14 ± 2.03	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
406	10.43 ± 3.20	1.1.1.1	Alcohol dehydrogenase
407	11.00 ± 5.01	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
408	13.14 ± 5.79	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
409	13.57 ± 4.69	1.17.1.4	Xanthine dehydrogenase
410	13.71 ± 5.50	1.3.5.1	Succinate dehydrogenase (quinone)
411	13.86 ± 2.23	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
412	18.86 ± 10.52	1.2.1.2	Formate dehydrogenase
413	20.14 ± 2.53	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
414	20.43 ± 2.61	1.8.1.4	Dihydrolipoyl dehydrogenase
415	20.57 ± 47.94	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
416	20.71 ± 4.43	1.6.5.5	NADPH:quinone reductase
417	21.00 ± 3.38	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
418	22.43 ± 8.16	1.1.99.1	Choline dehydrogenase
419	26.29 ± 16.77	1.1.1.95	Phosphoglycerate dehydrogenase
420	26.71 ± 13.16	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
421	27.14 ± 55.89	1.5.3.1	Sarcosine oxidase
422	27.86 ± 1.81	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
423	31.43 ± 58.76	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
424	31.57 ± 9.29	1.1.1.205	IMP dehydrogenase

425	31.71 ± 4.80	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
426	32.43 ± 5.97	1.1.1.22	UDP-glucose 6-dehydrogenase
427	34.86 ± 12.69	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
428	35.43 ± 5.58	1.1.1.3	Homoserine dehydrogenase
429	36.86 ± 11.46	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
430	37.29 ± 7.28	1.11.1.15	Peroxiredoxin
431	38.57 ± 58.73	1.5.8.4	Dimethylglycine dehydrogenase
432	38.71 ± 9.97	1.3.99.16	Isoquinoline 1-oxidoreductase
433	39.43 ± 9.12	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
434	39.86 ± 34.04	1.8.1.9	Thioredoxin-disulfide reductase
435	42.43 ± 6.02	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
436	44.43 ± 23.95	1.1.1.85	3-isopropylmalate dehydrogenase
437	44.86 ± 11.92	1.2.1.11	Aspartate-semialdehyde dehydrogenase
438	45.00 ± 21.97	1.11.1.21	Catalase peroxidase
439	45.29 ± 8.28	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
440	45.29 ± 15.42	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
441	45.29 ± 65.29	1.2.7.3	2-oxoglutarate synthase
442	45.57 ± 17.31	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
443	47.86 ± 10.06	1.6.99.3	NADH dehydrogenase
444	50.71 ± 12.23	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
445	51.43 ± 10.22	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
446	51.71 ± 16.86	1.1.1.37	Malate dehydrogenase
447	53.71 ± 12.53	1.2.1.70	Glutamyl-tRNA reductase
448	54.86 ± 24.91	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
449	56.00 ± 17.46	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
450	59.29 ± 16.23	1.1.1.23	Histidinol dehydrogenase
451	60.00 ± 23.20	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
452	63.00 ± 58.30	1.5.3.19	4-methylaminobutanoate oxidase (formaldehyde-forming)
453	63.86 ± 28.83	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
454	66.29 ± 6.78	1.4.7.1	Glutamate synthase (ferredoxin)
455	67.00 ± 14.48	1.14.11.17	Taurine dioxygenase
456	69.43 ± 19.40	1.3.99.22	Coproporphyrinogen dehydrogenase
457	70.14 ± 6.92	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
458	70.57 ± 47.38	1.1.3.15	(S)-2-hydroxy-acid oxidase
459	75.00 ± 28.21	1.1.5.3	Glycerol-3-phosphate dehydrogenase
460	75.57 ± 14.85	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
461	75.86 ± 21.35	1.8.1.8	Protein-disulfide reductase
462	76.14 ± 18.21	1.4.1.4	Glutamate dehydrogenase (NADP(+))

463	77.14 ± 14.10	1.4.1.1	Alanine dehydrogenase
464	78.57 ± 46.56	1.1.2.8	Alcohol dehydrogenase (cytochrome c)
465	79.71 ± 19.43	1.1.1.25	Shikimate dehydrogenase
466	80.29 ± 16.77	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
467	80.67 ± 19.35	1.1.1.14	L-iditol 2-dehydrogenase
468	82.57 ± 12.89	1.1.1.169	2-dehydropantoate 2-reductase
469	82.86 ± 24.09	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
470	84.00 ± 19.72	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
471	84.43 ± 38.56	1.14.15.7	Choline monooxygenase
472	84.86 ± 16.09	1.3.1.12	Prephenate dehydrogenase
473	85.29 ± 48.25	1.14.13.22	Cyclohexanone monooxygenase
474	85.86 ± 44.96	1.4.99.1	1.4.99.6
475	86.43 ± 19.03	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
476	86.71 ± 70.18	1.8.99.2	Adenylyl-sulfate reductase
477	87.43 ± 48.47	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
478	88.71 ± 17.96	1.1.1.133	dTDP-4-dehydrorhamnose reductase
479	88.86 ± 26.99	1.3.8.1	Short-chain acyl-CoA dehydrogenase
480	93.71 ± 31.71	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
481	94.14 ± 47.44	1.8.1.2	Assimilatory sulfite reductase (NADPH)
482	95.00 ± 23.59	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
483	95.29 ± 50.81	1.4.1.2	Glutamate dehydrogenase
484	97.29 ± 43.09	1.7.2.1	Nitrite reductase (NO-forming)
485	97.71 ± 25.14	1.17.4.2	Ribonucleoside-triphosphate reductase
486	98.14 ± 11.52	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
487	98.14 ± 14.02	1.17.99.6	Epoxyqueuosine reductase
488	102.71 ± 51.47	1.1.1.308	Sulfopropanediol 3-dehydrogenase
489	103.86 ± 16.07	1.13.12.16	Nitronate monooxygenase
490	105.00 ± 53.32	1.7.99.4	Nitrate reductase
491	105.71 ± 26.34	1.1.1.271	GDP-L-fucose synthase
492	106.43 ± 16.22	1.2.1.8	Betaine-aldehyde dehydrogenase
493	107.57 ± 17.99	1.1.1.125	2-deoxy-D-gluconate 3-dehydrogenase
494	108.50 ± 44.20	1.3.1.101	2,3-bis-O-geranylgeranyl-sn-glycerol 1-phosphate reductase (NAD(P)H)
495	109.71 ± 23.67	1.20.4.1	Arsenate reductase (glutaredoxin)
496	111.43 ± 18.95	1.1.1.69	Gluconate 5-dehydrogenase
497	111.57 ± 26.93	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
498	112.29 ± 48.71	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
499	113.14 ± 57.30	1.8.7.1	Assimilatory sulfite reductase (ferredoxin)
500	113.43 ± 64.18	1.1.1.18	Inositol 2-dehydrogenase

Benthic zone and subsea floor (26 metagenomes)			
row	**Avg.rank**	**EC number**	**Oxidoreductase**

501	1.58 ± 1.92	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
502	2.85 ± 1.70	1.8.98.1	CoB--CoM heterodisulfide reductase
503	6.00 ± 4.70	1.2.7.3	2-oxoglutarate synthase
504	6.19 ± 6.01	1.2.7.5	Aldehyde ferredoxin oxidoreductase
505	7.12 ± 3.68	1.17.4.1	Ribonucleoside-diphosphate reductase
506	7.27 ± 7.47	1.2.1.2	Formate dehydrogenase
507	8.50 ± 2.65	1.4.1.13	Glutamate synthase (NADPH)
508	8.58 ± 4.48	1.2.7.1	Pyruvate synthase
509	10.31 ± 6.46	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
510	13.19 ± 11.67	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
511	15.46 ± 8.47	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
512	16.81 ± 11.24	1.1.1.1	Alcohol dehydrogenase
513	17.77 ± 7.30	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
514	18.31 ± 9.08	1.1.1.95	Phosphoglycerate dehydrogenase
515	20.23 ± 7.45	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
516	22.62 ± 11.26	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
517	23.12 ± 17.27	1.3.5.1	Succinate dehydrogenase (quinone)
518	29.62 ± 8.58	1.1.1.205	IMP dehydrogenase
519	31.08 ± 24.17	1.2.7.4	Carbon-monoxide dehydrogenase (ferredoxin)
520	32.96 ± 11.90	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
521	33.88 ± 11.60	1.8.1.9	Thioredoxin-disulfide reductase
522	35.08 ± 13.84	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
523	35.35 ± 36.25	1.8.1.4	Dihydrolipoyl dehydrogenase
524	36.69 ± 16.71	1.1.1.22	UDP-glucose 6-dehydrogenase
525	37.27 ± 44.69	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
526	38.77 ± 23.86	1.1.1.14	L-iditol 2-dehydrogenase
527	39.31 ± 23.86	1.1.3.15	(S)-2-hydroxy-acid oxidase
528	39.73 ± 21.17	1.1.1.85	3-isopropylmalate dehydrogenase
529	40.32 ± 32.62	1.17.1.4	Xanthine dehydrogenase
530	41.92 ± 26.86	1.2.1.43	Formate dehydrogenase (NADP(+))
531	43.81 ± 26.64	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
532	45.38 ± 16.92	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
533	46.38 ± 11.87	1.1.1.3	Homoserine dehydrogenase
534	46.85 ± 18.87	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
535	47.81 ± 21.10	1.4.3.16	L-aspartate oxidase
536	50.73 ± 55.21	1.9.3.1	Cytochrome-c oxidase
537	51.38 ± 43.14	1.17.4.2	Ribonucleoside-triphosphate reductase
538	52.42 ± 21.89	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
539	52.65 ± 23.76	1.2.1.11	Aspartate-semialdehyde dehydrogenase
540	52.92 ± 31.89	1.5.3.1	Sarcosine oxidase
541	53.40 ± 18.76	1.11.1.15	Peroxiredoxin
542	53.62 ± 39.35	1.7.99.4	Nitrate reductase

543	54.23 ± 24.39	1.3.8.1	Short-chain acyl-CoA dehydrogenase
544	54.42 ± 25.59	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
545	55.12 ± 24.49	1.1.5.3	Glycerol-3-phosphate dehydrogenase
546	56.42 ± 15.78	1.4.1.1	Alanine dehydrogenase
547	56.54 ± 33.48	1.6.5.5	NADPH:quinone reductase
548	57.12 ± 33.59	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
549	58.96 ± 29.17	1.12.1.3	Hydrogen dehydrogenase (NADP(+))
550	59.92 ± 30.74	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
551	59.92 ± 33.35	1.1.1.23	Histidinol dehydrogenase
552	62.31 ± 26.78	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
553	62.76 ± 31.28	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
554	64.29 ± 31.27	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
555	64.38 ± 23.31	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
556	64.62 ± 38.49	1.2.99.5	Formylmethanofuran dehydrogenase
557	64.65 ± 41.06	1.3.7.8	Benzoyl-CoA reductase
558	64.92 ± 22.70	1.1.1.133	dTDP-4-dehydrorhamnose reductase
559	68.33 ± 28.58	1.3.99.22	Coproporphyrinogen dehydrogenase
560	69.42 ± 34.45	1.1.1.271	GDP-L-fucose synthase
561	69.50 ± 37.89	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
562	71.46 ± 41.87	1.18.1.2	Ferredoxin--NADP(+) reductase
563	72.31 ± 31.21	1.3.1.14	Dihydroorotate dehydrogenase (NAD(+))
564	73.31 ± 22.93	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
565	74.00 ± 26.56	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
566	74.60 ± 21.66	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
567	75.12 ± 17.30	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
568	75.46 ± 22.05	1.13.12.16	Nitronate monooxygenase
569	75.77 ± 33.63	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
570	76.20 ± 24.24	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
571	78.24 ± 28.81	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
572	80.96 ± 20.44	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
573	81.12 ± 18.19	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
574	81.68 ± 28.36	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
575	82.46 ± 36.29	1.12.99.6	Hydrogenase (acceptor)
576	84.77 ± 24.59	1.1.1.136	UDP-N-acetylglucosamine 6-dehydrogenase
577	84.92 ± 18.69	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
578	84.96 ± 20.11	1.1.1.25	Shikimate dehydrogenase
579	86.00 ± 29.76	1.17.99.6	Epoxyqueuosine reductase
580	88.48 ± 38.12	1.1.1.18	Inositol 2-dehydrogenase
581	88.60 ± 37.80	1.4.1.4	Glutamate dehydrogenase (NADP(+))
582	89.42 ± 23.83	1.1.1.37	Malate dehydrogenase

583	90.31 ± 43.53	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
584	90.95 ± 41.25	1.3.99.16	Isoquinoline 1-oxidoreductase
585	92.70 ± 27.16	1.2.1.70	Glutamyl-tRNA reductase
586	93.83 ± 44.60	1.6.99.3	NADH dehydrogenase
587	95.58 ± 24.45	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
588	95.84 ± 32.42	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
589	96.18 ± 45.00	1.5.3.19	4-methylaminobutanoate oxidase (formaldehyde-forming)
590	96.38 ± 69.29	1.11.1.6	Catalase
591	96.50 ± 43.93	1.12.98.1	Coenzyme F420 hydrogenase
592	97.62 ± 42.28	1.7.99.1	Hydroxylamine reductase
593	98.00 ± 41.33	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
594	98.23 ± 31.92	1.15.1.1	Superoxide dismutase
595	98.32 ± 35.88	1.12.1.2	Hydrogen dehydrogenase
596	98.58 ± 24.23	1.5.1.2	Pyrroline-5-carboxylate reductase
597	99.96 ± 41.31	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
598	100.46 ± 29.49	1.6.5.2	NAD(P)H dehydrogenase (quinone)
599	100.77 ± 30.99	1.1.1.26	Glyoxylate reductase
600	100.85 ± 34.91	1.1.1.103	L-threonine 3-dehydrogenase

Marine cold seeps (7 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
601	1.57 ± 1.05	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
602	3.43 ± 3.92	1.4.1.13	Glutamate synthase (NADPH)
603	5.00 ± 3.74	1.17.4.1	Ribonucleoside-diphosphate reductase
604	13.86 ± 5.54	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
605	15.14 ± 3.00	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
606	17.57 ± 19.83	1.3.5.1	Succinate dehydrogenase (quinone)
607	19.86 ± 35.65	1.7.99.4	Nitrate reductase
608	21.29 ± 11.70	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
609	22.00 ± 15.37	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
610	22.29 ± 22.40	1.2.1.2	Formate dehydrogenase
611	22.57 ± 13.53	1.11.1.15	Peroxiredoxin
612	23.43 ± 8.40	1.1.1.22	UDP-glucose 6-dehydrogenase
613	24.86 ± 49.88	1.9.3.1	Cytochrome-c oxidase
614	25.43 ± 10.22	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
615	25.57 ± 9.16	1.1.1.205	IMP dehydrogenase
616	26.14 ± 7.77	1.1.1.95	Phosphoglycerate dehydrogenase
617	27.57 ± 16.04	1.3.99.22	Coproporphyrinogen dehydrogenase
618	29.00 ± 31.49	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
619	29.00 ± 22.40	1.3.8.7	Medium-chain acyl-CoA dehydrogenase

620	29.14 ± 7.16	1.1.1.3	Homoserine dehydrogenase
621	30.14 ± 41.70	1.2.7.1	Pyruvate synthase
622	32.86 ± 7.02	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
623	33.43 ± 13.63	1.1.1.1	Alcohol dehydrogenase
624	33.86 ± 34.83	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
625	36.57 ± 15.64	1.8.1.4	Dihydrolipoyl dehydrogenase
626	37.43 ± 46.86	1.2.7.3	2-oxoglutarate synthase
627	37.43 ± 6.63	1.2.1.11	Aspartate-semialdehyde dehydrogenase
628	38.00 ± 14.97	1.8.1.9	Thioredoxin-disulfide reductase
629	43.14 ± 14.71	1.4.3.16	L-aspartate oxidase
630	43.14 ± 29.34	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
631	43.14 ± 8.85	1.1.1.23	Histidinol dehydrogenase
632	43.71 ± 32.64	1.17.4.2	Ribonucleoside-triphosphate reductase
633	44.86 ± 15.16	1.1.1.85	3-isopropylmalate dehydrogenase
634	45.17 ± 22.19	1.1.99.1	Choline dehydrogenase
635	45.57 ± 11.60	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
636	49.14 ± 29.77	1.1.3.15	(S)-2-hydroxy-acid oxidase
637	50.29 ± 55.16	1.4.1.2	Glutamate dehydrogenase
638	52.86 ± 25.75	1.8.1.8	Protein-disulfide reductase
639	53.00 ± 27.11	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
640	53.00 ± 19.33	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
641	56.29 ± 49.77	1.8.5.4	Sulfide:quinone reductase
642	57.43 ± 45.27	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
643	57.86 ± 24.30	1.13.12.16	Nitronate monooxygenase
644	58.29 ± 22.58	1.17.1.4	Xanthine dehydrogenase
645	58.71 ± 46.18	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
646	60.33 ± 26.30	1.7.2.4	Nitrous-oxide reductase
647	61.14 ± 21.91	1.4.1.4	Glutamate dehydrogenase (NADP(+))
648	62.14 ± 14.86	1.2.1.70	Glutamyl-tRNA reductase
649	62.57 ± 13.44	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
650	64.29 ± 16.29	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
651	64.57 ± 37.14	1.7.1.15	Nitrite reductase (NADH)
652	65.57 ± 42.01	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
653	66.00 ± 48.23	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
654	66.43 ± 15.97	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
655	68.71 ± 38.93	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
656	70.29 ± 55.63	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
657	71.43 ± 8.72	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
658	71.71 ± 19.95	1.3.8.1	Short-chain acyl-CoA dehydrogenase
659	72.71 ± 25.54	1.1.5.3	Glycerol-3-phosphate dehydrogenase
660	73.71 ± 23.02	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase

661	73.71 ± 45.62	1.11.1.5	Cytochrome-c peroxidase
662	74.86 ± 25.45	1.6.99.3	NADH dehydrogenase
663	75.86 ± 9.34	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
664	76.86 ± 9.45	1.3.1.12	Prephenate dehydrogenase
665	77.00 ± 62.05	1.3.99.16	Isoquinoline 1-oxidoreductase
666	77.00 ± 37.57	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
667	78.00 ± 32.72	1.1.1.37	Malate dehydrogenase
668	78.57 ± 35.27	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
669	78.71 ± 14.14	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
670	78.86 ± 23.01	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
671	79.29 ± 12.67	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
672	79.71 ± 15.21	1.1.1.25	Shikimate dehydrogenase
673	83.43 ± 10.72	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
674	87.67 ± 50.14	1.1.5.4	Malate dehydrogenase (quinone)
675	88.57 ± 28.90	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
676	88.86 ± 52.94	1.5.3.1	Sarcosine oxidase
677	92.86 ± 35.74	1.6.5.5	NADPH:quinone reductase
678	93.14 ± 50.40	1.11.1.21	Catalase peroxidase
679	94.14 ± 18.80	1.1.1.133	dTDP-4-dehydrorhamnose reductase
680	94.71 ± 106.44	1.8.1.2	Assimilatory sulfite reductase (NADPH)
681	97.29 ± 15.84	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
682	97.43 ± 25.80	1.15.1.1	Superoxide dismutase
683	98.14 ± 43.30	1.1.1.28	D-lactate dehydrogenase
684	100.43 ± 53.25	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
685	100.43 ± 79.47	1.8.98.1	CoB--CoM heterodisulfide reductase
686	100.86 ± 18.67	1.17.99.6	Epoxyqueuosine reductase
687	101.29 ± 38.60	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
688	101.71 ± 59.47	1.3.5.2	Dihydroorotate dehydrogenase (quinone)
689	101.86 ± 13.86	1.3.1.76	Precorrin-2 dehydrogenase
690	103.43 ± 23.82	1.5.1.2	Pyrroline-5-carboxylate reductase
691	103.86 ± 43.69	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
692	104.86 ± 37.50	1.4.1.1	Alanine dehydrogenase
693	105.86 ± 67.61	1.2.1.43	Formate dehydrogenase (NADP(+))
694	106.57 ± 52.20	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
695	107.14 ± 55.98	1.21.98.1	Cyclic dehypoxanthinyl futalosine synthase
696	107.57 ± 22.36	1.7.1.13	PreQ(1) synthase
697	108.29 ± 40.64	1.2.1.8	Betaine-aldehyde dehydrogenase
698	110.14 ± 40.03	1.1.1.271	GDP-L-fucose synthase
699	110.33 ± 119.44	1.12.5.1	Hydrogen:quinone oxidoreductase
700	111.14 ± 21.52	1.18.1.2	Ferredoxin--NADP(+) reductase

Hydrothermal vents (20 metagenomes)			
row	**Avg.rank**	**EC number**	**Oxidoreductase**
701	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
702	3.50 ± 1.36	1.4.1.13	Glutamate synthase (NADPH)
703	4.85 ± 2.73	1.17.4.1	Ribonucleoside-diphosphate reductase
704	9.85 ± 9.13	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
705	12.15 ± 21.24	1.9.3.1	Cytochrome-c oxidase
706	14.55 ± 15.36	1.3.5.1	Succinate dehydrogenase (quinone)
707	17.00 ± 10.48	1.8.1.4	Dihydrolipoyl dehydrogenase
708	17.55 ± 15.17	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
709	17.70 ± 7.46	1.1.1.205	IMP dehydrogenase
710	21.80 ± 16.48	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
711	22.05 ± 11.71	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
712	22.65 ± 7.98	1.8.1.9	Thioredoxin-disulfide reductase
713	25.25 ± 20.47	1.1.1.1	Alcohol dehydrogenase
714	25.90 ± 20.76	1.5.3.1	Sarcosine oxidase
715	26.95 ± 17.02	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
716	27.10 ± 20.21	1.2.1.2	Formate dehydrogenase
717	30.00 ± 10.41	1.1.1.85	3-isopropylmalate dehydrogenase
718	32.75 ± 16.20	1.1.1.95	Phosphoglycerate dehydrogenase
719	33.65 ± 35.71	1.2.7.3	2-oxoglutarate synthase
720	34.45 ± 19.02	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
721	35.55 ± 27.56	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
722	35.65 ± 18.38	1.2.1.11	Aspartate-semialdehyde dehydrogenase
723	36.75 ± 9.48	1.1.1.3	Homoserine dehydrogenase
724	37.80 ± 24.83	1.11.1.15	Peroxiredoxin
725	37.95 ± 18.25	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
726	38.10 ± 28.57	1.7.99.4	Nitrate reductase
727	38.15 ± 13.37	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
728	38.60 ± 42.53	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
729	38.75 ± 13.37	1.1.1.22	UDP-glucose 6-dehydrogenase
730	38.80 ± 18.42	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
731	40.40 ± 17.75	1.1.1.23	Histidinol dehydrogenase
732	41.00 ± 14.69	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
733	47.25 ± 34.41	1.1.3.15	(S)-2-hydroxy-acid oxidase
734	47.85 ± 26.99	1.4.3.16	L-aspartate oxidase
735	48.10 ± 40.44	1.5.8.4	Dimethylglycine dehydrogenase
736	48.15 ± 29.83	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
737	48.80 ± 27.93	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)

738	49.50 ± 22.13	1.1.1.37	Malate dehydrogenase
739	54.25 ± 32.99	1.3.99.22	Coproporphyrinogen dehydrogenase
740	55.95 ± 21.70	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
741	56.75 ± 31.64	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
742	56.75 ± 16.26	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
743	57.40 ± 41.66	1.8.1.8	Protein-disulfide reductase
744	58.15 ± 21.68	1.4.1.1	Alanine dehydrogenase
745	58.35 ± 24.39	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
746	59.25 ± 26.82	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
747	60.55 ± 55.39	1.8.99.2	Adenylyl-sulfate reductase
748	60.75 ± 21.61	1.2.1.70	Glutamyl-tRNA reductase
749	61.25 ± 51.74	1.8.5.4	Sulfide:quinone reductase
750	61.75 ± 19.42	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
751	61.75 ± 64.80	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
752	62.80 ± 33.74	1.17.1.4	Xanthine dehydrogenase
753	65.40 ± 29.17	1.6.5.5	NADPH:quinone reductase
754	66.40 ± 41.32	1.11.1.21	Catalase peroxidase
755	68.05 ± 18.14	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
756	69.20 ± 26.89	1.1.99.1	Choline dehydrogenase
757	70.70 ± 36.72	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
758	71.05 ± 18.02	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
759	72.45 ± 41.43	1.3.1.14	Dihydroorotate dehydrogenase (NAD(+))
760	73.90 ± 42.86	1.5.3.19	4-methylaminobutanoate oxidase (formaldehyde-forming)
761	74.75 ± 16.12	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
762	74.80 ± 20.49	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
763	75.40 ± 20.13	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
764	76.80 ± 14.51	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
765	77.85 ± 24.77	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
766	78.25 ± 48.77	1.7.1.15	Nitrite reductase (NADH)
767	78.35 ± 45.88	1.3.1.76	Precorrin-2 dehydrogenase
768	78.65 ± 33.66	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
769	78.80 ± 50.73	1.8.1.7	Glutathione-disulfide reductase
770	78.80 ± 30.51	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
771	80.45 ± 19.32	1.1.1.25	Shikimate dehydrogenase
772	82.55 ± 45.28	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
773	82.95 ± 24.22	1.5.1.2	Pyrroline-5-carboxylate reductase
774	85.15 ± 23.07	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
775	86.50 ± 26.38	1.3.1.12	Prephenate dehydrogenase
776	86.90 ± 42.24	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
777	90.65 ± 29.41	1.3.1.98	UDP-N-acetylmuramate dehydrogenase

778	91.80 ± 69.13	1.2.7.1	Pyruvate synthase
779	92.75 ± 26.54	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
780	93.30 ± 28.64	1.6.99.3	NADH dehydrogenase
781	94.10 ± 31.83	1.4.1.4	Glutamate dehydrogenase (NADP(+))
782	95.47 ± 46.12	1.3.5.4	Fumarate reductase (quinol)
783	95.50 ± 63.56	1.7.2.1	Nitrite reductase (NO-forming)
784	95.55 ± 51.66	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
785	95.70 ± 83.65	1.12.99.6	Hydrogenase (acceptor)
786	96.30 ± 32.94	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
787	97.05 ± 22.23	1.1.1.271	GDP-L-fucose synthase
788	97.35 ± 40.29	1.3.5.2	Dihydroorotate dehydrogenase (quinone)
789	98.05 ± 24.57	1.8.4.12	Peptide-methionine (R)-S-oxide reductase
790	98.50 ± 45.00	1.4.1.2	Glutamate dehydrogenase
791	100.10 ± 54.24	1.3.99.16	Isoquinoline 1-oxidoreductase
792	100.21 ± 49.61	1.10.2.2	Quinol--cytochrome-c reductase
793	101.20 ± 40.19	1.1.5.3	Glycerol-3-phosphate dehydrogenase
794	102.05 ± 50.95	1.18.1.2	Ferredoxin--NADP(+) reductase
795	103.60 ± 36.82	1.3.3.3	Coproporphyrinogen oxidase
796	105.75 ± 47.67	1.11.1.5	Cytochrome-c peroxidase
797	106.00 ± 95.01	1.8.99.3	Hydrogensulfite reductase
798	106.28 ± 60.86	1.8.7.1	Assimilatory sulfite reductase (ferredoxin)
799	109.05 ± 27.79	1.15.1.1	Superoxide dismutase
800	109.47 ± 40.65	1.1.1.14	L-iditol 2-dehydrogenase

Oxygen minimum zone (18 metagenomes)			
row	**Avg.rank**	**EC number**	**Oxidoreductase**
801	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
802	3.67 ± 2.08	1.17.4.1	Ribonucleoside-diphosphate reductase
803	3.83 ± 0.96	1.4.1.13	Glutamate synthase (NADPH)
804	4.17 ± 1.67	1.5.3.1	Sarcosine oxidase
805	4.78 ± 1.08	1.9.3.1	Cytochrome-c oxidase
806	7.78 ± 2.74	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
807	8.78 ± 3.98	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
808	9.78 ± 3.34	1.5.8.4	Dimethylglycine dehydrogenase
809	10.11 ± 3.05	1.3.5.1	Succinate dehydrogenase (quinone)
810	11.44 ± 10.04	1.7.99.4	Nitrate reductase
811	11.50 ± 2.03	1.2.1.2	Formate dehydrogenase
812	13.61 ± 2.93	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
813	14.28 ± 2.53	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
814	14.56 ± 4.74	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
815	14.61 ± 4.68	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
816	15.11 ± 4.32	1.1.1.1	Alcohol dehydrogenase

817	19.50 ± 10.58	1.1.99.1	Choline dehydrogenase
818	19.56 ± 3.30	1.1.1.95	Phosphoglycerate dehydrogenase
819	20.78 ± 5.98	1.17.1.4	Xanthine dehydrogenase
820	21.44 ± 2.61	1.8.1.4	Dihydrolipoyl dehydrogenase
821	21.50 ± 11.71	1.2.7.3	2-oxoglutarate synthase
822	25.56 ± 8.53	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
823	25.78 ± 3.05	1.1.1.205	IMP dehydrogenase
824	26.94 ± 4.58	1.5.3.19	4-methylaminobutanoate oxidase (formaldehyde-forming)
825	27.78 ± 18.23	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
826	28.17 ± 4.81	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
827	29.67 ± 20.10	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
828	30.72 ± 8.89	1.8.1.9	Thioredoxin-disulfide reductase
829	31.83 ± 10.07	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
830	33.94 ± 13.68	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
831	36.78 ± 14.45	1.6.5.5	NADPH:quinone reductase
832	37.33 ± 8.57	1.1.1.22	UDP-glucose 6-dehydrogenase
833	38.00 ± 7.53	1.1.1.85	3-isopropylmalate dehydrogenase
834	39.33 ± 6.72	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
835	40.50 ± 12.91	1.8.99.2	Adenylyl-sulfate reductase
836	40.89 ± 7.80	1.11.1.15	Peroxiredoxin
837	41.56 ± 7.27	1.1.1.3	Homoserine dehydrogenase
838	42.17 ± 5.59	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
839	42.33 ± 31.44	1.2.7.1	Pyruvate synthase
840	44.39 ± 7.34	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
841	44.72 ± 9.42	1.2.1.11	Aspartate-semialdehyde dehydrogenase
842	45.00 ± 13.15	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
843	45.56 ± 9.64	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
844	46.83 ± 11.19	1.1.1.37	Malate dehydrogenase
845	48.28 ± 12.62	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
846	50.89 ± 8.69	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
847	52.56 ± 8.05	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
848	52.83 ± 11.44	1.1.1.23	Histidinol dehydrogenase
849	53.39 ± 16.38	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
850	54.39 ± 22.64	1.3.99.22	Coproporphyrinogen dehydrogenase
851	55.67 ± 21.17	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
852	57.22 ± 9.66	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
853	57.56 ± 24.99	1.14.15.7	Choline monooxygenase
854	59.39 ± 12.31	1.4.99.1	1.4.99.6
855	59.56 ± 9.43	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase

856	60.17 ± 17.45	1.1.3.15	(S)-2-hydroxy-acid oxidase
857	60.89 ± 17.99	1.2.1.70	Glutamyl-tRNA reductase
858	62.89 ± 10.76	1.4.1.1	Alanine dehydrogenase
859	63.89 ± 16.58	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
860	64.00 ± 16.62	1.1.1.308	Sulfopropanediol 3-dehydrogenase
861	64.22 ± 20.17	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
862	64.28 ± 9.53	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
863	66.50 ± 26.83	1.8.1.8	Protein-disulfide reductase
864	67.50 ± 8.78	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
865	69.17 ± 10.95	1.1.1.25	Shikimate dehydrogenase
866	70.11 ± 11.55	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
867	72.61 ± 34.43	1.6.99.3	NADH dehydrogenase
868	72.67 ± 11.58	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
869	73.56 ± 12.50	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
870	74.11 ± 26.61	1.5.8.1	Dimethylamine dehydrogenase
871	74.17 ± 32.54	1.7.2.1	Nitrite reductase (NO-forming)
872	75.22 ± 19.07	1.1.5.3	Glycerol-3-phosphate dehydrogenase
873	77.56 ± 14.83	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
874	77.72 ± 10.50	1.1.1.133	dTDP-4-dehydrorhamnose reductase
875	77.78 ± 28.42	1.8.1.2	Assimilatory sulfite reductase (NADPH)
876	78.28 ± 18.48	1.1.2.3	L-lactate dehydrogenase (cytochrome)
877	79.06 ± 13.96	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
878	80.61 ± 15.11	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
879	80.72 ± 13.48	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
880	81.78 ± 11.20	1.3.1.12	Prephenate dehydrogenase
881	82.17 ± 31.65	1.14.11.17	Taurine dioxygenase
882	83.56 ± 25.55	1.4.3.16	L-aspartate oxidase
883	85.00 ± 27.68	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
884	86.33 ± 14.62	1.4.7.1	Glutamate synthase (ferredoxin)
885	88.67 ± 17.53	1.1.1.271	GDP-L-fucose synthase
886	91.39 ± 19.35	1.3.1.14	Dihydroorotate dehydrogenase (NAD(+))
887	93.39 ± 19.92	1.13.12.16	Nitronate monooxygenase
888	94.06 ± 27.67	1.1.1.108	Carnitine 3-dehydrogenase
889	94.56 ± 10.49	1.1.2.4	D-lactate dehydrogenase (cytochrome)
890	94.61 ± 13.04	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
891	95.67 ± 28.03	1.18.1.2	Ferredoxin--NADP(+) reductase
892	97.28 ± 13.44	1.5.1.2	Pyrroline-5-carboxylate reductase
893	97.94 ± 22.91	1.3.1.76	Precorrin-2 dehydrogenase
894	100.33 ± 36.45	1.14.13.22	Cyclohexanone monooxygenase
895	102.06 ± 16.54	1.3.8.1	Short-chain acyl-CoA dehydrogenase
896	102.06 ± 26.00	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))

897	103.39 ± 23.17	1.14.11.18	Phytanoyl-CoA dioxygenase
898	104.17 ± 20.15	1.4.1.4	Glutamate dehydrogenase (NADP(+))
899	105.83 ± 28.89	1.14.13.148	Trimethylamine monooxygenase
900	108.67 ± 21.63	1.14.11.1	Gamma-butyrobetaine dioxygenase

Marine photic zone (18 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
901	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
902	2.39 ± 1.01	1.17.4.1	Ribonucleoside-diphosphate reductase
903	3.17 ± 0.60	1.4.1.13	Glutamate synthase (NADPH)
904	4.33 ± 1.45	1.5.3.1	Sarcosine oxidase
905	4.67 ± 1.00	1.9.3.1	Cytochrome-c oxidase
906	7.50 ± 2.01	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
907	7.50 ± 1.67	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
908	8.72 ± 2.21	1.3.5.1	Succinate dehydrogenase (quinone)
909	10.39 ± 1.95	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
910	10.50 ± 2.59	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
911	11.06 ± 4.94	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
912	11.89 ± 3.63	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
913	14.50 ± 2.95	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
914	16.28 ± 6.09	1.1.99.1	Choline dehydrogenase
915	16.44 ± 3.40	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
916	16.61 ± 4.47	1.5.8.4	Dimethylglycine dehydrogenase
917	17.17 ± 2.22	1.8.1.4	Dihydrolipoyl dehydrogenase
918	17.78 ± 4.87	1.2.1.2	Formate dehydrogenase
919	19.00 ± 6.89	1.8.1.9	Thioredoxin-disulfide reductase
920	20.44 ± 2.67	1.1.1.95	Phosphoglycerate dehydrogenase
921	20.67 ± 5.35	1.1.1.1	Alcohol dehydrogenase
922	23.72 ± 3.56	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
923	25.17 ± 4.97	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
924	27.56 ± 3.50	1.1.1.205	IMP dehydrogenase
925	28.61 ± 5.06	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
926	28.94 ± 7.46	1.6.5.5	NADPH:quinone reductase
927	29.89 ± 3.98	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
928	29.94 ± 6.33	1.1.1.22	UDP-glucose 6-dehydrogenase
929	33.67 ± 4.89	1.1.1.85	3-isopropylmalate dehydrogenase
930	34.33 ± 4.52	1.1.1.3	Homoserine dehydrogenase
931	34.50 ± 13.76	1.1.2.3	L-lactate dehydrogenase (cytochrome)
932	36.11 ± 10.17	1.5.3.19	4-methylaminobutanoate oxidase (formaldehyde-forming)

933	36.22 ± 8.19	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
934	37.44 ± 15.32	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
935	37.50 ± 20.54	1.11.1.21	Catalase peroxidase
936	38.89 ± 6.14	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
937	39.94 ± 10.76	1.11.1.15	Peroxiredoxin
938	40.94 ± 4.72	1.2.1.11	Aspartate-semialdehyde dehydrogenase
939	42.11 ± 6.51	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
940	43.56 ± 6.11	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
941	43.83 ± 3.67	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
942	46.56 ± 12.54	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
943	46.89 ± 23.75	1.4.7.1	Glutamate synthase (ferredoxin)
944	48.22 ± 10.87	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
945	49.22 ± 6.85	1.1.1.23	Histidinol dehydrogenase
946	49.33 ± 7.34	1.1.1.37	Malate dehydrogenase
947	49.72 ± 17.86	1.17.1.4	Xanthine dehydrogenase
948	50.39 ± 12.60	1.3.99.26	All-trans-zeta-carotene desaturase
949	50.67 ± 7.72	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
950	50.89 ± 9.10	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
951	51.00 ± 15.17	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
952	54.22 ± 9.72	1.3.5.2	Dihydroorotate dehydrogenase (quinone)
953	56.28 ± 9.92	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
954	58.06 ± 16.48	1.6.99.3	NADH dehydrogenase
955	59.00 ± 25.02	1.2.7.3	2-oxoglutarate synthase
956	60.44 ± 23.15	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
957	61.39 ± 23.07	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
958	61.94 ± 28.82	1.8.99.2	Adenylyl-sulfate reductase
959	62.17 ± 11.88	1.4.1.1	Alanine dehydrogenase
960	63.44 ± 10.18	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
961	63.67 ± 8.46	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
962	64.78 ± 9.83	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
963	65.28 ± 12.40	1.4.99.1	1.4.99.6
964	65.83 ± 13.47	1.3.3.3	Coproporphyrinogen oxidase
965	66.56 ± 10.95	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
966	67.17 ± 10.30	1.1.1.25	Shikimate dehydrogenase
967	67.61 ± 13.80	1.14.11.17	Taurine dioxygenase
968	70.94 ± 13.55	1.1.1.308	Sulfopropanediol 3-dehydrogenase
969	71.50 ± 27.07	1.1.5.3	Glycerol-3-phosphate dehydrogenase
970	71.83 ± 17.31	1.1.2.4	D-lactate dehydrogenase (cytochrome)
971	71.94 ± 11.40	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
972	73.17 ± 14.91	1.3.99.22	Coproporphyrinogen dehydrogenase
973	74.33 ± 11.65	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase

974	75.50 ± 10.37	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
975	76.44 ± 17.65	1.13.12.16	Nitronate monooxygenase
976	79.50 ± 13.15	1.4.3.16	L-aspartate oxidase
977	79.61 ± 19.34	1.14.13.22	Cyclohexanone monooxygenase
978	80.67 ± 27.29	1.14.15.7	Choline monooxygenase
979	81.06 ± 19.29	1.1.1.271	GDP-L-fucose synthase
980	83.78 ± 10.05	1.5.1.2	Pyrroline-5-carboxylate reductase
981	84.06 ± 16.21	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
982	85.50 ± 17.91	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
983	87.83 ± 54.69	1.3.7.7	Ferredoxin:protochlorophyllide reductase (ATP-dependent)
984	87.83 ± 11.10	1.3.1.12	Prephenate dehydrogenase
985	89.17 ± 11.79	1.11.1.9	Glutathione peroxidase
986	90.67 ± 21.97	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
987	91.17 ± 13.24	1.4.3.5	Pyridoxal 5'-phosphate synthase
988	92.28 ± 22.75	1.8.1.8	Protein-disulfide reductase
989	93.50 ± 20.51	1.8.4.12	Peptide-methionine (R)-S-oxide reductase
990	93.83 ± 22.03	1.1.3.15	(S)-2-hydroxy-acid oxidase
991	93.83 ± 12.68	1.2.1.70	Glutamyl-tRNA reductase
992	95.00 ± 24.78	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
993	96.94 ± 12.15	1.1.1.26	Glyoxylate reductase
994	98.22 ± 18.88	1.15.1.1	Superoxide dismutase
995	99.39 ± 19.82	1.20.4.1	Arsenate reductase (glutaredoxin)
996	101.11 ± 15.44	1.17.99.6	Epoxyqueuosine reductase
997	103.28 ± 25.51	1.14.11.1	Gamma-butyrobetaine dioxygenase
998	103.78 ± 10.60	1.1.1.133	dTDP-4-dehydrorhamnose reductase
999	104.00 ± 26.26	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
1000	106.72 ± 28.14	1.2.1.8	Betaine-aldehyde dehydrogenase

Mangrove sediment (10 metagenomes)			
row	**Avg.rank**	**EC number**	**Oxidoreductase**
1001	1.10 ± 0.30	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
1002	4.40 ± 6.55	1.8.98.1	CoB--CoM heterodisulfide reductase
1003	5.40 ± 3.83	1.17.4.1	Ribonucleoside-diphosphate reductase
1004	5.50 ± 2.20	1.4.1.13	Glutamate synthase (NADPH)
1005	6.00 ± 1.67	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
1006	6.80 ± 2.60	1.2.7.3	2-oxoglutarate synthase
1007	8.00 ± 3.52	1.2.1.2	Formate dehydrogenase
1008	8.10 ± 2.84	1.2.7.1	Pyruvate synthase
1009	9.60 ± 12.69	1.2.7.5	Aldehyde ferredoxin oxidoreductase

1010	10.70 ± 1.85	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
1011	10.70 ± 1.95	1.1.1.1	Alcohol dehydrogenase
1012	13.00 ± 8.58	1.9.3.1	Cytochrome-c oxidase
1013	14.20 ± 3.74	1.8.1.4	Dihydrolipoyl dehydrogenase
1014	14.30 ± 1.27	1.3.5.1	Succinate dehydrogenase (quinone)
1015	16.10 ± 2.81	1.1.1.95	Phosphoglycerate dehydrogenase
1016	16.80 ± 9.81	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
1017	18.40 ± 4.94	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
1018	18.60 ± 4.57	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
1019	18.70 ± 4.15	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
1020	20.90 ± 2.84	1.8.1.9	Thioredoxin-disulfide reductase
1021	22.30 ± 7.11	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
1022	24.00 ± 7.21	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
1023	25.90 ± 8.97	1.17.1.4	Xanthine dehydrogenase
1024	28.60 ± 6.12	1.1.1.205	IMP dehydrogenase
1025	31.00 ± 13.02	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
1026	32.50 ± 8.59	1.11.1.21	Catalase peroxidase
1027	32.90 ± 6.86	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
1028	33.70 ± 6.63	1.6.5.5	NADPH:quinone reductase
1029	33.70 ± 6.03	1.1.5.3	Glycerol-3-phosphate dehydrogenase
1030	35.50 ± 14.10	1.5.3.1	Sarcosine oxidase
1031	36.40 ± 8.78	1.1.1.22	UDP-glucose 6-dehydrogenase
1032	36.40 ± 12.69	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
1033	37.10 ± 18.17	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
1034	37.70 ± 14.48	1.1.3.15	(S)-2-hydroxy-acid oxidase
1035	37.80 ± 6.71	1.11.1.15	Peroxiredoxin
1036	39.80 ± 7.67	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
1037	40.70 ± 17.46	1.3.99.16	Isoquinoline 1-oxidoreductase
1038	41.00 ± 17.74	1.7.99.4	Nitrate reductase
1039	41.10 ± 16.43	1.3.99.22	Coproporphyrinogen dehydrogenase
1040	41.80 ± 10.32	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
1041	45.60 ± 15.39	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
1042	48.20 ± 18.39	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
1043	50.30 ± 10.28	1.1.1.85	3-isopropylmalate dehydrogenase
1044	50.60 ± 8.98	1.1.1.3	Homoserine dehydrogenase
1045	54.00 ± 11.32	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
1046	55.00 ± 12.30	1.4.1.1	Alanine dehydrogenase
1047	55.60 ± 10.13	1.4.3.16	L-aspartate oxidase
1048	55.70 ± 43.06	1.2.7.4	Carbon-monoxide dehydrogenase (ferredoxin)
1049	56.90 ± 16.52	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
1050	57.50 ± 17.10	1.3.8.1	Short-chain acyl-CoA dehydrogenase
1051	59.20 ± 24.53	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating)

			(NADP(+))
1052	59.90 ± 26.82	1.12.99.6	Hydrogenase (acceptor)
1053	61.40 ± 15.33	1.4.1.4	Glutamate dehydrogenase (NADP(+))
1054	62.40 ± 32.93	1.17.4.2	Ribonucleoside-triphosphate reductase
1055	62.70 ± 11.51	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
1056	63.00 ± 32.80	1.1.1.14	L-iditol 2-dehydrogenase
1057	63.50 ± 35.10	1.8.99.2	Adenylyl-sulfate reductase
1058	64.40 ± 33.82	1.2.1.43	Formate dehydrogenase (NADP(+))
1059	64.50 ± 17.43	1.1.1.23	Histidinol dehydrogenase
1060	64.70 ± 8.63	1.2.1.11	Aspartate-semialdehyde dehydrogenase
1061	67.10 ± 39.49	1.3.7.8	Benzoyl-CoA reductase
1062	67.30 ± 14.28	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
1063	67.40 ± 23.99	1.12.1.2	Hydrogen dehydrogenase
1064	67.40 ± 11.72	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
1065	68.00 ± 20.74	1.8.1.8	Protein-disulfide reductase
1066	69.40 ± 6.34	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
1067	70.70 ± 10.15	1.1.1.133	dTDP-4-dehydrorhamnose reductase
1068	72.90 ± 19.23	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
1069	74.10 ± 26.12	1.4.1.2	Glutamate dehydrogenase
1070	74.70 ± 17.46	1.12.1.3	Hydrogen dehydrogenase (NADP(+))
1071	76.00 ± 28.15	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
1072	76.90 ± 13.56	1.1.1.18	Inositol 2-dehydrogenase
1073	77.60 ± 15.03	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
1074	79.20 ± 26.63	1.6.99.3	NADH dehydrogenase
1075	79.90 ± 31.62	1.8.99.3	Hydrogensulfite reductase
1076	80.20 ± 12.16	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
1077	80.70 ± 16.03	1.1.1.169	2-dehydropantoate 2-reductase
1078	81.60 ± 30.76	1.11.1.5	Cytochrome-c peroxidase
1079	81.70 ± 20.50	1.13.12.16	Nitronate monooxygenase
1080	82.30 ± 28.01	1.4.7.1	Glutamate synthase (ferredoxin)
1081	82.30 ± 27.46	1.14.13.81	Magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase
1082	84.10 ± 12.15	1.1.1.37	Malate dehydrogenase
1083	84.60 ± 14.46	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
1084	85.00 ± 12.50	1.1.1.25	Shikimate dehydrogenase
1085	85.40 ± 16.44	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
1086	86.90 ± 17.17	1.1.1.271	GDP-L-fucose synthase
1087	88.70 ± 13.83	1.17.99.6	Epoxyqueuosine reductase
1088	89.70 ± 21.50	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
1089	90.20 ± 21.36	1.8.5.4	Sulfide:quinone reductase
1090	90.80 ± 32.62	1.3.5.4	Fumarate reductase (quinol)

1091	90.90 ± 15.86	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
1092	92.00 ± 8.52	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
1093	92.70 ± 16.39	1.18.1.2	Ferredoxin--NADP(+) reductase
1094	94.30 ± 11.03	1.1.1.91	Aryl-alcohol dehydrogenase (NADP(+))
1095	94.30 ± 13.10	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
1096	96.60 ± 14.58	1.7.1.15	Nitrite reductase (NADH)
1097	96.80 ± 17.20	1.15.1.1	Superoxide dismutase
1098	96.80 ± 42.16	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
1099	97.50 ± 15.17	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
1100	97.70 ± 39.55	1.2.99.5	Formylmethanofuran dehydrogenase

Forest soil (9 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
1101	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
1102	2.67 ± 0.67	1.9.3.1	Cytochrome-c oxidase
1103	3.33 ± 1.70	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
1104	5.11 ± 1.73	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
1105	5.44 ± 2.01	1.2.1.2	Formate dehydrogenase
1106	7.56 ± 2.50	1.4.1.13	Glutamate synthase (NADPH)
1107	9.00 ± 4.57	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
1108	9.11 ± 3.96	1.17.4.1	Ribonucleoside-diphosphate reductase
1109	10.00 ± 3.46	1.3.99.16	Isoquinoline 1-oxidoreductase
1110	11.22 ± 6.49	1.3.5.1	Succinate dehydrogenase (quinone)
1111	11.67 ± 2.54	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
1112	11.78 ± 4.05	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
1113	12.33 ± 3.13	1.1.1.1	Alcohol dehydrogenase
1114	14.67 ± 4.11	1.2.7.3	2-oxoglutarate synthase
1115	15.78 ± 2.20	1.8.1.9	Thioredoxin-disulfide reductase
1116	19.89 ± 4.79	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
1117	20.00 ± 7.90	1.17.1.4	Xanthine dehydrogenase
1118	20.89 ± 1.97	1.8.1.4	Dihydrolipoyl dehydrogenase
1119	21.11 ± 8.01	1.1.2.8	Alcohol dehydrogenase (cytochrome c)
1120	21.56 ± 6.45	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
1121	21.67 ± 6.57	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
1122	22.78 ± 4.57	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
1123	23.00 ± 9.76	1.6.5.5	NADPH:quinone reductase
1124	25.11 ± 3.87	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
1125	25.67 ± 5.54	1.1.1.95	Phosphoglycerate dehydrogenase
1126	28.44 ± 5.40	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
1127	29.11 ± 5.74	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
1128	30.22 ± 9.69	1.1.99.1	Choline dehydrogenase

1129	34.00 ± 5.54	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
1130	34.33 ± 9.29	1.1.1.205	IMP dehydrogenase
1131	34.44 ± 9.08	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
1132	36.89 ± 11.86	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
1133	38.00 ± 18.20	1.11.1.21	Catalase peroxidase
1134	38.78 ± 5.43	1.11.1.15	Peroxiredoxin
1135	38.89 ± 19.33	1.6.99.3	NADH dehydrogenase
1136	40.22 ± 14.31	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
1137	40.89 ± 2.88	1.1.1.22	UDP-glucose 6-dehydrogenase
1138	41.67 ± 5.72	1.1.3.15	(S)-2-hydroxy-acid oxidase
1139	44.78 ± 13.10	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
1140	46.00 ± 12.41	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
1141	46.89 ± 20.15	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
1142	49.00 ± 20.40	1.5.3.1	Sarcosine oxidase
1143	52.89 ± 31.54	1.7.99.4	Nitrate reductase
1144	54.56 ± 9.31	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
1145	55.44 ± 16.26	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
1146	55.56 ± 8.80	1.1.1.85	3-isopropylmalate dehydrogenase
1147	56.78 ± 22.58	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
1148	57.89 ± 35.69	1.2.7.1	Pyruvate synthase
1149	57.89 ± 19.13	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
1150	58.44 ± 16.96	1.15.1.1	Superoxide dismutase
1151	58.56 ± 10.40	1.4.1.2	Glutamate dehydrogenase
1152	58.89 ± 20.78	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
1153	60.78 ± 12.80	1.12.99.6	Hydrogenase (acceptor)
1154	60.89 ± 25.09	1.11.1.10	Chloride peroxidase
1155	61.22 ± 12.84	1.2.1.11	Aspartate-semialdehyde dehydrogenase
1156	62.44 ± 14.89	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
1157	62.67 ± 6.18	1.8.1.2	Assimilatory sulfite reductase (NADPH)
1158	63.11 ± 18.07	1.14.13.22	Cyclohexanone monooxygenase
1159	65.44 ± 12.44	1.1.5.3	Glycerol-3-phosphate dehydrogenase
1160	66.44 ± 28.93	1.14.14.5	Alkanesulfonate monooxygenase
1161	67.67 ± 7.57	1.3.8.1	Short-chain acyl-CoA dehydrogenase
1162	68.00 ± 29.28	1.1.1.37	Malate dehydrogenase
1163	70.44 ± 23.01	1.3.99.22	Coproporphyrinogen dehydrogenase
1164	72.00 ± 4.88	1.1.1.3	Homoserine dehydrogenase
1165	72.11 ± 20.12	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
1166	73.44 ± 22.89	1.1.1.91	Aryl-alcohol dehydrogenase (NADP(+))
1167	73.44 ± 23.04	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-

			specific)
1168	74.22 ± 24.26	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
1169	74.67 ± 12.97	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
1170	74.89 ± 24.78	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
1171	74.89 ± 18.24	1.13.12.16	Nitronate monooxygenase
1172	75.89 ± 26.48	1.2.1.8	Betaine-aldehyde dehydrogenase
1173	77.22 ± 29.32	1.14.19.1	Stearoyl-CoA 9-desaturase
1174	79.00 ± 22.53	1.4.7.1	Glutamate synthase (ferredoxin)
1175	80.44 ± 46.98	1.11.1.6	Catalase
1176	80.67 ± 16.27	1.11.1.5	Cytochrome-c peroxidase
1177	81.78 ± 36.61	1.1.99.3	Gluconate 2-dehydrogenase (acceptor)
1178	82.89 ± 30.64	1.13.11.27	4-hydroxyphenylpyruvate dioxygenase
1179	83.56 ± 13.90	1.1.1.136	UDP-N-acetylglucosamine 6-dehydrogenase
1180	84.78 ± 15.08	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
1181	88.33 ± 25.76	1.14.11.17	Taurine dioxygenase
1182	89.11 ± 24.55	1.14.14.9	4-hydroxyphenylacetate 3-monooxygenase
1183	89.78 ± 32.24	1.17.2.1	Nicotinate dehydrogenase (cytochrome)
1184	89.89 ± 31.51	1.1.1.14	L-iditol 2-dehydrogenase
1185	90.44 ± 60.40	1.2.5.1	Pyruvate dehydrogenase (quinone)
1186	91.67 ± 23.82	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
1187	95.11 ± 23.80	1.4.1.1	Alanine dehydrogenase
1188	95.22 ± 47.41	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
1189	96.00 ± 21.18	1.1.1.23	Histidinol dehydrogenase
1190	96.33 ± 27.62	1.8.4.12	Peptide-methionine (R)-S-oxide reductase
1191	96.67 ± 30.86	1.14.13.40	Anthraniloyl-CoA monooxygenase
1192	97.00 ± 31.92	1.4.3.16	L-aspartate oxidase
1193	97.89 ± 32.83	1.7.1.15	Nitrite reductase (NADH)
1194	98.22 ± 29.36	1.13.11.5	Homogentisate 1,2-dioxygenase
1195	98.78 ± 15.19	1.1.2.4	D-lactate dehydrogenase (cytochrome)
1196	99.56 ± 28.06	1.1.1.169	2-dehydropantoate 2-reductase
1197	101.00 ± 17.35	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
1198	104.00 ± 21.87	1.14.12.21	Benzoyl-CoA 2,3-dioxygenase
1199	104.22 ± 35.25	1.1.1.271	GDP-L-fucose synthase
1200	104.56 ± 18.39	1.1.2.3	L-lactate dehydrogenase (cytochrome)

Grassland soil (18 metagenomes)			
row	**Avg.rank**	**EC number**	**Oxidoreductase**
1201	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
1202	2.28 ± 0.45	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
1203	3.17 ± 1.01	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
1204	4.00 ± 1.00	1.9.3.1	Cytochrome-c oxidase
1205	5.22 ± 1.03	1.1.1.1	Alcohol dehydrogenase
1206	7.11 ± 1.20	1.6.5.5	NADPH:quinone reductase

1207	7.33 ± 2.83	1.2.1.2	Formate dehydrogenase
1208	10.11 ± 3.03	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
1209	10.17 ± 2.59	1.8.1.9	Thioredoxin-disulfide reductase
1210	12.67 ± 4.01	1.3.99.16	Isoquinoline 1-oxidoreductase
1211	13.06 ± 4.10	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
1212	13.22 ± 3.94	1.17.1.4	Xanthine dehydrogenase
1213	13.50 ± 4.57	1.4.1.13	Glutamate synthase (NADPH)
1214	13.72 ± 4.43	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
1215	14.39 ± 4.98	1.6.99.3	NADH dehydrogenase
1216	17.06 ± 4.54	1.17.4.1	Ribonucleoside-diphosphate reductase
1217	17.83 ± 4.00	1.1.2.8	Alcohol dehydrogenase (cytochrome c)
1218	17.89 ± 5.28	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
1219	18.00 ± 5.11	1.8.1.4	Dihydrolipoyl dehydrogenase
1220	18.22 ± 3.71	1.3.5.1	Succinate dehydrogenase (quinone)
1221	21.00 ± 4.77	1.2.7.3	2-oxoglutarate synthase
1222	21.44 ± 3.39	1.1.1.95	Phosphoglycerate dehydrogenase
1223	23.06 ± 4.22	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
1224	26.00 ± 7.59	1.2.7.1	Pyruvate synthase
1225	26.11 ± 5.29	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
1226	30.56 ± 7.91	1.5.3.1	Sarcosine oxidase
1227	31.33 ± 8.50	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
1228	33.22 ± 17.25	1.1.99.1	Choline dehydrogenase
1229	34.17 ± 10.69	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
1230	34.33 ± 10.11	1.1.1.205	IMP dehydrogenase
1231	34.83 ± 12.51	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
1232	35.50 ± 12.76	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
1233	37.50 ± 9.42	1.11.1.15	Peroxiredoxin
1234	39.72 ± 13.96	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
1235	42.67 ± 15.91	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
1236	44.78 ± 23.41	1.14.14.5	Alkanesulfonate monooxygenase
1237	44.83 ± 16.14	1.1.1.22	UDP-glucose 6-dehydrogenase
1238	46.00 ± 16.50	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
1239	47.72 ± 18.03	1.11.1.10	Chloride peroxidase
1240	47.78 ± 12.43	1.17.2.1	Nicotinate dehydrogenase (cytochrome)
1241	48.11 ± 17.98	1.1.99.3	Gluconate 2-dehydrogenase (acceptor)
1242	49.39 ± 13.54	1.1.1.91	Aryl-alcohol dehydrogenase (NADP(+))
1243	51.06 ± 15.15	1.7.99.4	Nitrate reductase
1244	51.39 ± 17.79	1.3.99.22	Coproporphyrinogen dehydrogenase
1245	51.89 ± 16.30	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
1246	53.00 ± 25.62	1.1.3.15	(S)-2-hydroxy-acid oxidase
1247	53.06 ± 13.20	1.1.1.18	Inositol 2-dehydrogenase

1248	54.78 ± 26.43	1.18.1.3	Ferredoxin--NAD(+) reductase
1249	58.06 ± 14.49	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
1250	58.11 ± 28.74	1.11.1.5	Cytochrome-c peroxidase
1251	61.06 ± 12.09	1.1.1.14	L-iditol 2-dehydrogenase
1252	61.28 ± 25.75	1.1.5.3	Glycerol-3-phosphate dehydrogenase
1253	62.28 ± 15.23	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
1254	62.94 ± 18.56	1.13.12.16	Nitronate monooxygenase
1255	65.33 ± 21.66	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
1256	66.89 ± 32.51	1.4.1.2	Glutamate dehydrogenase
1257	67.61 ± 23.94	1.1.1.3	Homoserine dehydrogenase
1258	67.89 ± 19.61	1.1.1.169	2-dehydropantoate 2-reductase
1259	67.94 ± 22.91	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
1260	68.28 ± 27.74	1.3.8.1	Short-chain acyl-CoA dehydrogenase
1261	68.44 ± 28.95	1.14.13.22	Cyclohexanone monooxygenase
1262	69.61 ± 25.36	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
1263	72.06 ± 21.40	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
1264	72.78 ± 19.68	1.1.1.65	Pyridoxine 4-dehydrogenase
1265	75.00 ± 23.25	1.14.19.1	Stearoyl-CoA 9-desaturase
1266	76.17 ± 23.95	1.15.1.1	Superoxide dismutase
1267	76.17 ± 25.08	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
1268	77.00 ± 21.77	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
1269	77.06 ± 24.52	1.8.1.2	Assimilatory sulfite reductase (NADPH)
1270	79.17 ± 25.08	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
1271	80.33 ± 31.87	1.12.99.6	Hydrogenase (acceptor)
1272	81.17 ± 36.73	1.1.1.85	3-isopropylmalate dehydrogenase
1273	83.44 ± 34.48	1.11.1.21	Catalase peroxidase
1274	84.72 ± 28.35	1.13.12.3	Tryptophan 2-monooxygenase
1275	84.72 ± 21.36	1.1.1.25	Shikimate dehydrogenase
1276	85.89 ± 25.54	1.20.4.1	Arsenate reductase (glutaredoxin)
1277	88.00 ± 23.48	1.8.1.8	Protein-disulfide reductase
1278	88.06 ± 30.45	1.2.1.11	Aspartate-semialdehyde dehydrogenase
1279	88.22 ± 22.55	1.6.5.2	NAD(P)H dehydrogenase (quinone)
1280	88.50 ± 30.67	1.14.11.17	Taurine dioxygenase
1281	88.67 ± 32.77	1.4.3.4	Monoamine oxidase
1282	89.83 ± 24.95	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
1283	93.44 ± 31.31	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
1284	93.44 ± 35.28	1.11.1.6	Catalase
1285	95.28 ± 47.04	1.2.1.8	Betaine-aldehyde dehydrogenase
1286	96.11 ± 21.37	1.1.2.6	Polyvinyl alcohol dehydrogenase (cytochrome)
1287	96.28 ± 23.79	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase

1288	96.72 ± 35.13	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
1289	96.83 ± 27.55	1.1.1.219	Dihydrokaempferol 4-reductase
1290	97.72 ± 31.43	1.3.1.34	2,4-dienoyl-CoA reductase (NADPH)
1291	97.78 ± 30.67	1.1.1.133	dTDP-4-dehydrorhamnose reductase
1292	97.89 ± 27.21	1.1.1.69	Gluconate 5-dehydrogenase
1293	98.61 ± 34.85	1.5.1.3	Dihydrofolate reductase
1294	99.72 ± 40.43	1.1.1.136	UDP-N-acetylglucosamine 6-dehydrogenase
1295	101.22 ± 26.69	1.1.1.37	Malate dehydrogenase
1296	101.94 ± 43.91	1.13.11.27	4-hydroxyphenylpyruvate dioxygenase
1297	102.06 ± 28.18	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
1298	103.00 ± 23.82	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
1299	104.78 ± 45.47	1.14.13.1	Salicylate 1-monooxygenase
1300	105.83 ± 39.99	1.1.2.3	L-lactate dehydrogenase (cytochrome)

Hot desert (3 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
1301	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
1302	2.00 ± 0.00	1.9.3.1	Cytochrome-c oxidase
1303	3.00 ± 0.00	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
1304	4.33 ± 0.47	1.17.4.1	Ribonucleoside-diphosphate reductase
1305	5.00 ± 0.82	1.3.5.1	Succinate dehydrogenase (quinone)
1306	5.67 ± 0.47	1.4.1.13	Glutamate synthase (NADPH)
1307	7.33 ± 0.47	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
1308	7.67 ± 0.47	1.2.1.2	Formate dehydrogenase
1309	10.00 ± 1.41	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
1310	10.33 ± 0.47	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
1311	11.00 ± 1.41	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
1312	11.33 ± 1.25	1.1.1.1	Alcohol dehydrogenase
1313	12.33 ± 0.94	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
1314	14.67 ± 0.94	1.8.1.9	Thioredoxin-disulfide reductase
1315	15.33 ± 0.47	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
1316	15.67 ± 1.25	1.8.1.4	Dihydrolipoyl dehydrogenase
1317	16.33 ± 0.94	1.2.7.3	2-oxoglutarate synthase
1318	18.33 ± 0.47	1.1.1.205	IMP dehydrogenase
1319	19.00 ± 0.82	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
1320	20.00 ± 0.82	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
1321	22.33 ± 2.62	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
1322	23.00 ± 1.41	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
1323	23.33 ± 0.47	1.1.1.95	Phosphoglycerate dehydrogenase
1324	24.00 ± 1.41	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
1325	24.33 ± 3.30	1.17.1.4	Xanthine dehydrogenase

1326	26.00 ± 2.16	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
1327	26.67 ± 0.94	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
1328	28.33 ± 2.36	1.1.1.22	UDP-glucose 6-dehydrogenase
1329	29.33 ± 2.62	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
1330	30.33 ± 3.40	1.4.1.2	Glutamate dehydrogenase
1331	32.33 ± 0.47	1.5.3.1	Sarcosine oxidase
1332	33.00 ± 2.83	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
1333	35.00 ± 0.82	1.1.1.37	Malate dehydrogenase
1334	36.00 ± 7.12	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
1335	36.67 ± 4.92	1.6.99.3	NADH dehydrogenase
1336	38.00 ± 4.97	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
1337	38.33 ± 0.47	1.1.3.15	(S)-2-hydroxy-acid oxidase
1338	40.33 ± 2.87	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
1339	40.33 ± 1.25	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
1340	41.33 ± 4.50	1.11.1.21	Catalase peroxidase
1341	41.67 ± 1.25	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
1342	43.00 ± 7.48	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
1343	45.33 ± 7.59	1.11.1.15	Peroxiredoxin
1344	45.33 ± 2.87	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
1345	45.33 ± 13.96	1.12.99.6	Hydrogenase (acceptor)
1346	47.67 ± 1.70	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
1347	47.67 ± 5.31	1.2.1.11	Aspartate-semialdehyde dehydrogenase
1348	49.67 ± 1.25	1.1.1.85	3-isopropylmalate dehydrogenase
1349	50.67 ± 2.62	1.4.1.1	Alanine dehydrogenase
1350	51.00 ± 2.16	1.6.5.5	NADPH:quinone reductase
1351	51.33 ± 8.99	1.1.1.133	dTDP-4-dehydrorhamnose reductase
1352	53.00 ± 1.63	1.15.1.1	Superoxide dismutase
1353	53.00 ± 14.17	1.3.99.16	Isoquinoline 1-oxidoreductase
1354	53.67 ± 7.41	1.13.11.27	4-hydroxyphenylpyruvate dioxygenase
1355	54.67 ± 9.03	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
1356	55.33 ± 21.23	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
1357	56.67 ± 4.50	1.2.1.8	Betaine-aldehyde dehydrogenase
1358	57.67 ± 0.47	1.7.1.15	Nitrite reductase (NADH)
1359	57.67 ± 4.03	1.3.8.1	Short-chain acyl-CoA dehydrogenase
1360	58.00 ± 7.48	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
1361	60.33 ± 1.70	1.11.1.6	Catalase
1362	61.33 ± 3.30	1.7.99.4	Nitrate reductase
1363	61.33 ± 4.78	1.14.19.1	Stearoyl-CoA 9-desaturase
1364	62.33 ± 6.13	1.1.1.3	Homoserine dehydrogenase

1365	64.33 ± 10.66	1.4.7.1	Glutamate synthase (ferredoxin)
1366	65.00 ± 6.53	1.1.2.8	Alcohol dehydrogenase (cytochrome c)
1367	66.33 ± 3.30	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
1368	66.33 ± 2.49	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
1369	68.00 ± 1.41	1.1.99.1	Choline dehydrogenase
1370	68.33 ± 6.60	1.1.5.3	Glycerol-3-phosphate dehydrogenase
1371	68.67 ± 6.34	1.13.11.5	Homogentisate 1,2-dioxygenase
1372	70.33 ± 2.05	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
1373	71.67 ± 0.47	1.8.4.12	Peptide-methionine (R)-S-oxide reductase
1374	74.00 ± 6.16	1.1.1.41	Isocitrate dehydrogenase (NAD(+))
1375	74.67 ± 4.78	1.1.1.136	UDP-N-acetylglucosamine 6-dehydrogenase
1376	76.00 ± 4.90	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
1377	76.33 ± 1.70	1.4.3.16	L-aspartate oxidase
1378	78.67 ± 3.68	1.8.1.2	Assimilatory sulfite reductase (NADPH)
1379	79.00 ± 4.97	1.2.5.1	Pyruvate dehydrogenase (quinone)
1380	79.33 ± 2.49	1.1.1.23	Histidinol dehydrogenase
1381	80.67 ± 4.50	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
1382	81.33 ± 5.25	1.14.13.149	Phenylacetyl-CoA 1,2-epoxidase
1383	82.33 ± 4.78	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
1384	84.33 ± 3.30	1.1.1.91	Aryl-alcohol dehydrogenase (NADP(+))
1385	85.00 ± 4.90	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
1386	85.67 ± 3.40	1.3.99.26	All-trans-zeta-carotene desaturase
1387	88.00 ± 5.72	1.14.13.40	Anthraniloyl-CoA monooxygenase
1388	89.33 ± 1.70	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
1389	89.67 ± 6.24	1.14.13.22	Cyclohexanone monooxygenase
1390	91.00 ± 1.41	1.8.7.1	Assimilatory sulfite reductase (ferredoxin)
1391	91.33 ± 4.03	1.3.99.22	Coproporphyrinogen dehydrogenase
1392	93.00 ± 1.63	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
1393	93.67 ± 6.94	1.13.11.2	Catechol 2,3-dioxygenase
1394	94.00 ± 7.26	1.1.1.271	GDP-L-fucose synthase
1395	95.00 ± 8.04	1.14.13.127	3-(3-hydroxy-phenyl)propanoic acid hydroxylase
1396	96.00 ± 1.63	1.1.1.83	D-malate dehydrogenase (decarboxylating)
1397	96.00 ± 1.63	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
1398	96.67 ± 3.30	1.21.98.1	Cyclic dehypoxanthinyl futalosine synthase
1399	99.33 ± 3.30	1.18.1.2	Ferredoxin--NADP(+) reductase
1400	101.33 ± 7.41	1.8.4.8	Phosphoadenylyl-sulfate reductase (thioredoxin)

Polar desert (8 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
1401	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
1402	2.12 ± 0.33	1.9.3.1	Cytochrome-c oxidase
1403	3.75 ± 0.83	1.17.4.1	Ribonucleoside-diphosphate reductase

1404	4.88 ± 1.45	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
1405	5.12 ± 1.96	1.3.5.1	Succinate dehydrogenase (quinone)
1406	6.25 ± 2.63	1.4.1.13	Glutamate synthase (NADPH)
1407	9.00 ± 2.40	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
1408	9.88 ± 2.85	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
1409	11.50 ± 3.54	1.8.1.4	Dihydrolipoyl dehydrogenase
1410	12.38 ± 5.27	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
1411	14.12 ± 5.16	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
1412	14.50 ± 5.94	1.2.1.2	Formate dehydrogenase
1413	17.75 ± 11.89	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
1414	17.75 ± 5.87	1.8.1.9	Thioredoxin-disulfide reductase
1415	17.88 ± 1.83	1.1.1.205	IMP dehydrogenase
1416	19.88 ± 4.99	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
1417	20.12 ± 5.90	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
1418	22.25 ± 3.27	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
1419	23.38 ± 5.83	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
1420	23.62 ± 23.23	1.11.1.6	Catalase
1421	23.75 ± 4.68	1.1.1.95	Phosphoglycerate dehydrogenase
1422	24.38 ± 19.15	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
1423	24.50 ± 6.80	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
1424	26.00 ± 6.89	1.1.1.1	Alcohol dehydrogenase
1425	26.25 ± 5.54	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
1426	27.38 ± 10.50	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
1427	29.38 ± 4.50	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
1428	29.50 ± 2.06	1.1.1.22	UDP-glucose 6-dehydrogenase
1429	29.75 ± 46.78	1.12.99.6	Hydrogenase (acceptor)
1430	30.88 ± 11.07	1.17.1.4	Xanthine dehydrogenase
1431	31.75 ± 11.46	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
1432	35.50 ± 29.26	1.2.7.3	2-oxoglutarate synthase
1433	36.12 ± 5.64	1.1.1.37	Malate dehydrogenase
1434	37.50 ± 7.47	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
1435	40.12 ± 16.51	1.11.1.21	Catalase peroxidase
1436	40.50 ± 12.88	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
1437	42.62 ± 10.15	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
1438	43.12 ± 9.02	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
1439	43.62 ± 6.96	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
1440	44.00 ± 4.72	1.4.1.1	Alanine dehydrogenase
1441	44.25 ± 3.80	1.2.1.11	Aspartate-semialdehyde dehydrogenase
1442	45.50 ± 16.05	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))

1443	46.00 ± 11.42	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
1444	48.00 ± 10.00	1.15.1.1	Superoxide dismutase
1445	48.50 ± 15.91	1.4.1.2	Glutamate dehydrogenase
1446	50.62 ± 6.82	1.1.1.3	Homoserine dehydrogenase
1447	52.00 ± 13.88	1.11.1.15	Peroxiredoxin
1448	53.62 ± 13.55	1.1.3.15	(S)-2-hydroxy-acid oxidase
1449	54.00 ± 14.86	1.6.99.3	NADH dehydrogenase
1450	55.25 ± 7.05	1.1.1.85	3-isopropylmalate dehydrogenase
1451	55.75 ± 8.88	1.8.4.12	Peptide-methionine (R)-S-oxide reductase
1452	56.12 ± 8.80	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
1453	56.75 ± 21.78	1.5.3.1	Sarcosine oxidase
1454	56.88 ± 11.36	1.13.11.27	4-hydroxyphenylpyruvate dioxygenase
1455	58.25 ± 14.69	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
1456	58.62 ± 6.82	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
1457	59.50 ± 13.68	1.3.8.1	Short-chain acyl-CoA dehydrogenase
1458	60.75 ± 20.36	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
1459	61.75 ± 14.33	1.14.19.1	Stearoyl-CoA 9-desaturase
1460	62.88 ± 18.61	1.3.99.26	All-trans-zeta-carotene desaturase
1461	63.25 ± 37.80	1.2.5.1	Pyruvate dehydrogenase (quinone)
1462	63.25 ± 2.95	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
1463	63.50 ± 14.27	1.1.5.3	Glycerol-3-phosphate dehydrogenase
1464	63.62 ± 12.61	1.13.11.5	Homogentisate 1,2-dioxygenase
1465	63.75 ± 20.57	1.4.7.1	Glutamate synthase (ferredoxin)
1466	63.88 ± 21.96	1.4.3.16	L-aspartate oxidase
1467	64.00 ± 28.46	1.7.1.15	Nitrite reductase (NADH)
1468	70.12 ± 9.75	1.1.1.23	Histidinol dehydrogenase
1469	71.12 ± 18.24	1.7.99.4	Nitrate reductase
1470	71.25 ± 17.79	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
1471	72.38 ± 16.48	1.1.1.133	dTDP-4-dehydrorhamnose reductase
1472	73.12 ± 21.69	1.6.5.5	NADPH:quinone reductase
1473	74.62 ± 12.55	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
1474	75.00 ± 13.69	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
1475	75.12 ± 30.02	1.2.1.8	Betaine-aldehyde dehydrogenase
1476	75.38 ± 10.25	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
1477	76.25 ± 12.36	1.14.13.149	Phenylacetyl-CoA 1,2-epoxidase
1478	78.88 ± 12.62	1.3.99.22	Coproporphyrinogen dehydrogenase
1479	80.88 ± 23.98	1.1.1.136	UDP-N-acetylglucosamine 6-dehydrogenase
1480	83.88 ± 7.24	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
1481	84.50 ± 8.49	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
1482	86.00 ± 4.03	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
1483	86.25 ± 25.18	1.8.1.7	Glutathione-disulfide reductase

1484	87.75 ± 11.87	1.14.13.40	Anthraniloyl-CoA monooxygenase
1485	88.88 ± 24.19	1.8.1.2	Assimilatory sulfite reductase (NADPH)
1486	89.88 ± 24.89	1.1.1.271	GDP-L-fucose synthase
1487	90.25 ± 30.66	1.14.19.3	Acyl-CoA 6-desaturase
1488	90.25 ± 17.41	1.17.99.6	Epoxyqueuosine reductase
1489	92.50 ± 7.52	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
1490	93.00 ± 22.36	1.1.99.1	Choline dehydrogenase
1491	93.25 ± 16.70	1.18.1.2	Ferredoxin--NADP(+) reductase
1492	93.62 ± 7.55	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
1493	97.88 ± 30.72	1.1.1.41	Isocitrate dehydrogenase (NAD(+))
1494	98.25 ± 13.40	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
1495	99.25 ± 19.27	1.8.7.1	Assimilatory sulfite reductase (ferredoxin)
1496	101.00 ± 5.83	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
1497	102.75 ± 37.82	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
1498	103.88 ± 27.51	1.3.99.16	Isoquinoline 1-oxidoreductase
1499	105.50 ± 12.62	1.8.4.8	Phosphoadenylyl-sulfate reductase (thioredoxin)
1500	106.50 ± 21.37	1.16.3.1	Ferroxidase

Rhizosphere (13 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
1501	1.00 ± 0.00	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
1502	3.69 ± 2.61	1.9.3.1	Cytochrome-c oxidase
1503	4.15 ± 1.41	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
1504	5.38 ± 3.10	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
1505	5.77 ± 1.48	1.4.1.13	Glutamate synthase (NADPH)
1506	5.92 ± 3.77	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
1507	7.85 ± 2.68	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
1508	8.31 ± 1.32	1.1.1.1	Alcohol dehydrogenase
1509	8.31 ± 2.43	1.2.1.2	Formate dehydrogenase
1510	10.15 ± 4.42	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
1511	13.23 ± 13.95	1.17.4.1	Ribonucleoside-diphosphate reductase
1512	13.31 ± 2.70	1.3.5.1	Succinate dehydrogenase (quinone)
1513	15.31 ± 3.20	1.17.1.4	Xanthine dehydrogenase
1514	15.54 ± 3.91	1.8.1.4	Dihydrolipoyl dehydrogenase
1515	18.69 ± 10.62	1.7.99.4	Nitrate reductase
1516	18.69 ± 2.70	1.8.1.9	Thioredoxin-disulfide reductase
1517	19.00 ± 5.64	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
1518	20.15 ± 4.79	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
1519	21.00 ± 4.26	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
1520	21.54 ± 5.94	1.6.5.5	NADPH:quinone reductase
1521	23.23 ± 9.74	1.1.1.95	Phosphoglycerate dehydrogenase
1522	24.31 ± 6.72	1.2.7.3	2-oxoglutarate synthase
1523	25.85 ± 8.60	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase

1524	27.00 ± 6.95	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
1525	27.46 ± 13.95	1.3.99.16	Isoquinoline 1-oxidoreductase
1526	27.92 ± 5.85	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
1527	29.00 ± 5.70	1.1.1.205	IMP dehydrogenase
1528	29.46 ± 3.25	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
1529	32.69 ± 11.42	1.5.3.1	Sarcosine oxidase
1530	33.31 ± 19.08	1.1.99.1	Choline dehydrogenase
1531	34.85 ± 7.61	1.6.99.3	NADH dehydrogenase
1532	34.92 ± 8.39	1.4.1.2	Glutamate dehydrogenase
1533	36.77 ± 4.89	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
1534	40.46 ± 11.41	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
1535	42.46 ± 6.46	1.1.1.22	UDP-glucose 6-dehydrogenase
1536	43.62 ± 8.72	1.1.5.3	Glycerol-3-phosphate dehydrogenase
1537	44.25 ± 27.05	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
1538	45.31 ± 9.08	1.1.3.15	(S)-2-hydroxy-acid oxidase
1539	46.08 ± 24.04	1.1.2.8	Alcohol dehydrogenase (cytochrome c)
1540	46.69 ± 7.82	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
1541	46.77 ± 9.98	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
1542	47.38 ± 10.59	1.11.1.21	Catalase peroxidase
1543	48.46 ± 10.81	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
1544	48.77 ± 23.80	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
1545	49.77 ± 16.04	1.2.1.8	Betaine-aldehyde dehydrogenase
1546	52.08 ± 11.60	1.14.14.5	Alkanesulfonate monooxygenase
1547	56.00 ± 6.77	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
1548	56.77 ± 16.22	1.1.1.91	Aryl-alcohol dehydrogenase (NADP(+))
1549	57.08 ± 15.40	1.3.99.22	Coproporphyrinogen dehydrogenase
1550	58.00 ± 6.75	1.1.1.85	3-isopropylmalate dehydrogenase
1551	58.15 ± 8.53	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
1552	58.69 ± 18.80	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
1553	59.00 ± 12.94	1.1.1.3	Homoserine dehydrogenase
1554	59.23 ± 27.36	1.11.1.15	Peroxiredoxin
1555	59.38 ± 30.71	1.7.1.15	Nitrite reductase (NADH)
1556	60.08 ± 12.33	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
1557	60.62 ± 15.94	1.8.1.2	Assimilatory sulfite reductase (NADPH)
1558	61.92 ± 16.59	1.14.13.22	Cyclohexanone monooxygenase
1559	65.15 ± 20.25	1.11.1.6	Catalase
1560	65.69 ± 12.60	1.13.12.16	Nitronate monooxygenase
1561	67.15 ± 15.27	1.1.1.37	Malate dehydrogenase
1562	67.46 ± 18.11	1.13.11.27	4-hydroxyphenylpyruvate dioxygenase

1563	70.00 ± 16.98	1.3.8.1	Short-chain acyl-CoA dehydrogenase
1564	70.23 ± 37.39	1.2.1.11	Aspartate-semialdehyde dehydrogenase
1565	71.77 ± 46.52	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
1566	71.92 ± 18.85	1.4.1.1	Alanine dehydrogenase
1567	72.67 ± 12.36	1.18.1.3	Ferredoxin--NAD(+) reductase
1568	74.31 ± 19.24	1.1.1.31	3-hydroxyisobutyrate dehydrogenase
1569	76.23 ± 17.02	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
1570	78.46 ± 16.27	1.13.11.5	Homogentisate 1,2-dioxygenase
1571	80.15 ± 53.81	1.2.5.1	Pyruvate dehydrogenase (quinone)
1572	80.15 ± 23.18	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
1573	81.85 ± 12.95	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
1574	82.31 ± 22.19	1.14.13.40	Anthraniloyl-CoA monooxygenase
1575	82.77 ± 18.75	1.1.2.3	L-lactate dehydrogenase (cytochrome)
1576	83.77 ± 20.05	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
1577	84.92 ± 8.63	1.1.1.23	Histidinol dehydrogenase
1578	85.08 ± 31.48	1.2.7.1	Pyruvate synthase
1579	86.31 ± 19.12	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase
1580	86.54 ± 16.87	1.4.3.16	L-aspartate oxidase
1581	86.77 ± 33.00	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
1582	88.46 ± 14.70	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
1583	88.69 ± 21.99	1.4.7.1	Glutamate synthase (ferredoxin)
1584	88.85 ± 36.10	1.15.1.1	Superoxide dismutase
1585	89.62 ± 13.99	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
1586	90.00 ± 17.18	1.1.2.4	D-lactate dehydrogenase (cytochrome)
1587	91.08 ± 22.32	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
1588	92.69 ± 21.24	1.14.13.127	3-(3-hydroxy-phenyl)propanoic acid hydroxylase
1589	94.54 ± 58.43	1.14.19.1	Stearoyl-CoA 9-desaturase
1590	94.62 ± 9.62	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
1591	94.69 ± 33.48	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
1592	95.38 ± 17.26	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
1593	96.85 ± 32.75	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
1594	96.85 ± 62.99	1.12.99.6	Hydrogenase (acceptor)
1595	97.31 ± 34.97	1.1.1.25	Shikimate dehydrogenase
1596	98.85 ± 17.65	1.1.1.169	2-dehydropantoate 2-reductase
1597	99.23 ± 12.56	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
1598	103.14 ± 137.36	1.14.13.124	Phenylalanine N-monooxygenase
1599	103.23 ± 17.57	1.13.11.2	Catechol 2,3-dioxygenase
1600	103.62 ± 15.32	1.3.1.98	UDP-N-acetylmuramate dehydrogenase

Solar saltern and salt sarsh

(12 metagenomes)			
row	**Avg.rank**	**EC number**	**Oxidoreductase**
1601	1.17 ± 0.37	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
1602	2.42 ± 1.04	1.17.4.1	Ribonucleoside-diphosphate reductase
1603	3.08 ± 1.61	1.9.3.1	Cytochrome-c oxidase
1604	5.75 ± 1.53	1.4.1.13	Glutamate synthase (NADPH)
1605	8.42 ± 2.06	1.3.5.1	Succinate dehydrogenase (quinone)
1606	12.67 ± 5.86	1.8.1.4	Dihydrolipoyl dehydrogenase
1607	12.83 ± 5.79	1.1.1.205	IMP dehydrogenase
1608	13.00 ± 4.20	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
1609	13.33 ± 6.28	1.1.1.1	Alcohol dehydrogenase
1610	13.58 ± 9.35	1.2.1.2	Formate dehydrogenase
1611	14.00 ± 7.45	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
1612	14.00 ± 16.63	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
1613	14.25 ± 8.68	1.2.7.3	2-oxoglutarate synthase
1614	17.58 ± 11.43	1.2.7.1	Pyruvate synthase
1615	19.83 ± 3.44	1.8.1.9	Thioredoxin-disulfide reductase
1616	21.42 ± 12.80	1.1.5.3	Glycerol-3-phosphate dehydrogenase
1617	21.50 ± 5.09	1.1.1.95	Phosphoglycerate dehydrogenase
1618	24.67 ± 5.04	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
1619	25.08 ± 7.50	1.1.1.22	UDP-glucose 6-dehydrogenase
1620	25.58 ± 8.18	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
1621	26.58 ± 25.82	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)
1622	29.67 ± 17.26	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
1623	30.42 ± 25.10	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
1624	31.25 ± 10.69	1.6.99.3	NADH dehydrogenase
1625	31.25 ± 20.76	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
1626	32.25 ± 12.46	1.5.3.1	Sarcosine oxidase
1627	32.67 ± 19.10	1.5.5.1	Electron-transferring-flavoprotein dehydrogenase
1628	34.33 ± 33.65	1.2.7.5	Aldehyde ferredoxin oxidoreductase
1629	36.67 ± 7.34	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
1630	37.25 ± 9.05	1.11.1.15	Peroxiredoxin
1631	37.50 ± 7.35	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
1632	38.75 ± 19.71	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
1633	39.08 ± 28.13	1.3.99.26	All-trans-zeta-carotene desaturase
1634	40.08 ± 12.30	1.4.1.1	Alanine dehydrogenase
1635	40.25 ± 24.31	1.11.1.21	Catalase peroxidase
1636	40.67 ± 9.76	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
1637	46.00 ± 42.07	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
1638	47.17 ± 8.27	1.2.1.11	Aspartate-semialdehyde dehydrogenase
1639	47.67 ± 12.13	1.6.5.5	NADPH:quinone reductase

1640	48.50 ± 54.83	1.7.99.4	Nitrate reductase
1641	48.58 ± 31.00	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
1642	48.75 ± 8.53	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
1643	50.58 ± 5.35	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
1644	53.50 ± 22.25	1.1.1.157	3-hydroxybutyryl-CoA dehydrogenase
1645	53.58 ± 34.99	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
1646	53.92 ± 5.06	1.15.1.1	Superoxide dismutase
1647	54.92 ± 8.60	1.1.1.3	Homoserine dehydrogenase
1648	57.42 ± 6.97	1.1.1.37	Malate dehydrogenase
1649	58.25 ± 27.38	1.8.4.8	Phosphoadenylyl-sulfate reductase (thioredoxin)
1650	58.25 ± 48.63	1.2.4.4	3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring)
1651	58.83 ± 15.47	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
1652	61.33 ± 10.92	1.1.1.85	3-isopropylmalate dehydrogenase
1653	61.75 ± 37.33	1.18.1.3	Ferredoxin--NAD(+) reductase
1654	66.75 ± 13.14	1.1.1.23	Histidinol dehydrogenase
1655	68.83 ± 12.29	1.1.1.343	Phosphogluconate dehydrogenase (NAD(+)-dependent, decarboxylating)
1656	71.17 ± 24.74	1.3.8.6	Glutaryl-CoA dehydrogenase (ETF)
1657	72.42 ± 40.29	1.4.1.2	Glutamate dehydrogenase
1658	72.50 ± 28.09	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
1659	73.00 ± 23.12	1.2.1.70	Glutamyl-tRNA reductase
1660	73.33 ± 26.54	1.1.1.346	2,5-didehydrogluconate reductase (2-dehydro-L-gulonate-forming)
1661	78.08 ± 19.47	1.8.4.12	Peptide-methionine (R)-S-oxide reductase
1662	78.42 ± 29.71	1.1.5.2	Quinoprotein glucose dehydrogenase (PQQ, quinone)
1663	79.00 ± 29.14	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
1664	79.00 ± 33.52	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
1665	79.25 ± 22.00	1.8.5.4	Sulfide:quinone reductase
1666	83.83 ± 46.70	1.3.99.22	Coproporphyrinogen dehydrogenase
1667	83.92 ± 25.25	1.1.99.1	Choline dehydrogenase
1668	84.67 ± 52.97	1.7.7.1	Ferredoxin--nitrite reductase
1669	85.92 ± 30.40	1.4.1.4	Glutamate dehydrogenase (NADP(+))
1670	87.00 ± 22.73	1.1.1.136	UDP-N-acetylglucosamine 6-dehydrogenase
1671	87.83 ± 23.94	1.1.1.91	Aryl-alcohol dehydrogenase (NADP(+))
1672	90.08 ± 47.38	1.14.13.149	Phenylacetyl-CoA 1,2-epoxidase
1673	90.25 ± 34.22	1.4.3.16	L-aspartate oxidase
1674	90.25 ± 39.35	1.3.5.2	Dihydroorotate dehydrogenase (quinone)
1675	91.50 ± 46.29	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
1676	92.08 ± 23.76	1.2.1.18	Malonate-semialdehyde dehydrogenase (acetylating)
1677	94.83 ± 49.48	1.13.11.49	Chlorite O(2)-lyase
1678	95.08 ± 23.63	1.1.1.25	Shikimate dehydrogenase

1679	96.00 ± 27.67	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
1680	98.67 ± 21.24	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
1681	99.00 ± 41.09	1.5.1.3	Dihydrofolate reductase
1682	103.25 ± 38.09	1.20.4.1	Arsenate reductase (glutaredoxin)
1683	104.25 ± 24.60	1.1.1.38	Malate dehydrogenase (oxaloacetate-decarboxylating)
1684	104.92 ± 33.81	1.1.1.271	GDP-L-fucose synthase
1685	105.33 ± 47.90	1.1.3.15	(S)-2-hydroxy-acid oxidase
1686	106.25 ± 30.53	1.1.2.4	D-lactate dehydrogenase (cytochrome)
1687	106.58 ± 28.30	1.5.1.2	Pyrroline-5-carboxylate reductase
1688	106.83 ± 25.77	1.8.1.8	Protein-disulfide reductase
1689	107.08 ± 15.05	1.1.1.133	dTDP-4-dehydrorhamnose reductase
1690	107.33 ± 30.50	1.3.1.12	Prephenate dehydrogenase
1691	110.92 ± 36.39	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
1692	111.92 ± 31.32	1.11.1.5	Cytochrome-c peroxidase
1693	112.08 ± 67.91	1.12.99.6	Hydrogenase (acceptor)
1694	112.73 ± 22.83	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
1695	113.10 ± 88.43	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
1696	114.75 ± 42.31	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
1697	116.73 ± 26.88	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
1698	116.75 ± 65.04	1.2.1.59	Glyceraldehyde-3-phosphate dehydrogenase (NAD(P)(+)) (phosphorylating)
1699	116.82 ± 99.39	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
1700	117.58 ± 83.04	1.7.2.4	Nitrous-oxide reductase

Subterraneum (8 metagenomes)

row	Avg.rank	EC number	Oxidoreductase
1701	1.12 ± 0.33	1.6.5.3	NADH:ubiquinone reductase (H(+)-translocating)
1702	4.25 ± 2.38	1.17.4.1	Ribonucleoside-diphosphate reductase
1703	4.62 ± 3.04	1.4.1.13	Glutamate synthase (NADPH)
1704	5.62 ± 3.28	1.2.7.1	Pyruvate synthase
1705	13.38 ± 5.52	1.2.7.3	2-oxoglutarate synthase
1706	15.25 ± 13.12	1.3.5.1	Succinate dehydrogenase (quinone)
1707	16.50 ± 7.07	1.2.1.2	Formate dehydrogenase
1708	19.00 ± 6.32	1.1.1.205	IMP dehydrogenase
1709	20.88 ± 7.82	1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)
1710	21.62 ± 24.83	1.12.99.6	Hydrogenase (acceptor)
1711	21.88 ± 21.89	1.9.3.1	Cytochrome-c oxidase
1712	22.00 ± 10.58	1.3.99.22	Coproporphyrinogen dehydrogenase
1713	24.38 ± 26.49	1.7.99.4	Nitrate reductase
1714	24.75 ± 18.86	1.1.1.42	Isocitrate dehydrogenase (NADP(+))
1715	26.00 ± 5.61	1.4.3.16	L-aspartate oxidase

1716	27.88 ± 32.46	1.18.6.1	Nitrogenase
1717	30.12 ± 8.24	1.8.1.9	Thioredoxin-disulfide reductase
1718	31.75 ± 8.60	1.2.1.11	Aspartate-semialdehyde dehydrogenase
1719	33.00 ± 6.56	1.1.1.85	3-isopropylmalate dehydrogenase
1720	33.12 ± 14.99	1.1.1.95	Phosphoglycerate dehydrogenase
1721	33.12 ± 7.27	1.1.1.3	Homoserine dehydrogenase
1722	33.38 ± 7.61	1.1.1.86	Ketol-acid reductoisomerase (NADP(+))
1723	33.62 ± 11.78	1.1.1.22	UDP-glucose 6-dehydrogenase
1724	33.75 ± 9.05	1.1.1.23	Histidinol dehydrogenase
1725	35.00 ± 22.99	1.12.1.2	Hydrogen dehydrogenase
1726	36.00 ± 28.92	1.1.1.40	Malate dehydrogenase (oxaloacetate-decarboxylating) (NADP(+))
1727	37.62 ± 33.62	1.11.1.15	Peroxiredoxin
1728	38.00 ± 27.56	1.8.1.4	Dihydrolipoyl dehydrogenase
1729	38.38 ± 35.18	1.8.98.1	CoB--CoM heterodisulfide reductase
1730	38.38 ± 25.17	1.4.4.2	Glycine dehydrogenase (aminomethyl-transferring)
1731	38.75 ± 32.85	1.2.4.1	Pyruvate dehydrogenase (acetyl-transferring)
1732	38.88 ± 7.99	1.17.7.1	(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ferredoxin)
1733	40.12 ± 3.89	1.2.1.41	Glutamate-5-semialdehyde dehydrogenase
1734	42.12 ± 47.21	1.97.1.4	[Formate-C-acetyltransferase]-activating enzyme
1735	42.50 ± 9.85	1.2.1.38	N-acetyl-gamma-glutamyl-phosphate reductase
1736	43.25 ± 10.63	1.1.1.100	3-oxoacyl-[acyl-carrier-protein] reductase
1737	44.75 ± 6.81	1.5.1.5	Methylenetetrahydrofolate dehydrogenase (NADP(+))
1738	46.50 ± 49.89	1.8.99.2	Adenylyl-sulfate reductase
1739	47.38 ± 33.82	1.2.4.2	Oxoglutarate dehydrogenase (succinyl-transferring)
1740	49.38 ± 9.87	1.1.1.267	1-deoxy-D-xylulose-5-phosphate reductoisomerase
1741	49.88 ± 33.74	1.8.1.8	Protein-disulfide reductase
1742	50.25 ± 46.52	1.8.5.4	Sulfide:quinone reductase
1743	51.50 ± 14.45	1.2.1.70	Glutamyl-tRNA reductase
1744	51.75 ± 18.86	1.1.1.133	dTDP-4-dehydrorhamnose reductase
1745	53.25 ± 12.35	1.17.1.2	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
1746	53.75 ± 11.21	1.13.12.16	Nitronate monooxygenase
1747	54.00 ± 47.69	1.8.99.3	Hydrogensulfite reductase
1748	54.25 ± 28.86	1.1.1.1	Alcohol dehydrogenase
1749	54.75 ± 45.73	1.2.7.5	Aldehyde ferredoxin oxidoreductase
1750	55.62 ± 44.50	1.1.3.15	(S)-2-hydroxy-acid oxidase
1751	55.88 ± 41.75	1.2.7.8	Indolepyruvate ferredoxin oxidoreductase
1752	56.62 ± 61.69	1.2.1.43	Formate dehydrogenase (NADP(+))
1753	56.62 ± 65.63	1.7.1.15	Nitrite reductase (NADH)
1754	57.38 ± 22.53	1.1.1.37	Malate dehydrogenase
1755	62.25 ± 41.61	1.17.4.2	Ribonucleoside-triphosphate reductase
1756	62.38 ± 9.10	1.1.1.193	5-amino-6-(5-phosphoribosylamino)uracil reductase

1757	63.00 ± 19.09	1.17.1.8	4-hydroxy-tetrahydrodipicolinate reductase
1758	64.50 ± 24.80	1.1.1.271	GDP-L-fucose synthase
1759	64.62 ± 24.87	1.5.1.20	Methylenetetrahydrofolate reductase (NAD(P)H)
1760	67.62 ± 14.76	1.1.1.94	Glycerol-3-phosphate dehydrogenase (NAD(P)(+))
1761	68.00 ± 10.56	1.3.1.98	UDP-N-acetylmuramate dehydrogenase
1762	72.50 ± 20.27	1.3.1.12	Prephenate dehydrogenase
1763	77.38 ± 17.03	1.1.1.25	Shikimate dehydrogenase
1764	79.38 ± 46.46	1.11.1.5	Cytochrome-c peroxidase
1765	80.12 ± 14.06	1.15.1.1	Superoxide dismutase
1766	80.25 ± 46.35	1.11.1.21	Catalase peroxidase
1767	82.38 ± 24.24	1.1.1.262	4-hydroxythreonine-4-phosphate dehydrogenase
1768	83.62 ± 24.18	1.2.1.16	Succinate-semialdehyde dehydrogenase (NAD(P)(+))
1769	83.62 ± 52.73	1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)
1770	84.75 ± 46.41	1.3.1.10	Enoyl-[acyl-carrier-protein] reductase (NADPH, Si-specific)
1771	86.38 ± 66.10	1.2.7.4	Carbon-monoxide dehydrogenase (ferredoxin)
1772	86.88 ± 32.52	1.3.5.2	Dihydroorotate dehydrogenase (quinone)
1773	87.00 ± 20.07	1.20.4.1	Arsenate reductase (glutaredoxin)
1774	87.12 ± 19.60	1.6.99.3	NADH dehydrogenase
1775	87.12 ± 52.92	1.21.98.1	Cyclic dehypoxanthinyl futalosine synthase
1776	87.50 ± 43.55	1.12.1.3	Hydrogen dehydrogenase (NADP(+))
1777	87.88 ± 26.04	1.18.1.2	Ferredoxin--NADP(+) reductase
1778	88.00 ± 45.58	1.1.1.35	3-hydroxyacyl-CoA dehydrogenase
1779	88.50 ± 17.40	1.5.1.2	Pyrroline-5-carboxylate reductase
1780	89.88 ± 32.46	1.8.4.11	Peptide-methionine (S)-S-oxide reductase
1781	92.00 ± 21.53	1.17.99.6	Epoxyqueuosine reductase
1782	96.50 ± 69.60	1.4.1.4	Glutamate dehydrogenase (NADP(+))
1783	97.25 ± 29.61	1.3.3.3	Coproporphyrinogen oxidase
1784	99.00 ± 27.08	1.3.1.76	Precorrin-2 dehydrogenase
1785	100.50 ± 17.42	1.6.5.2	NAD(P)H dehydrogenase (quinone)
1786	102.75 ± 37.94	1.1.1.49	Glucose-6-phosphate dehydrogenase (NADP(+))
1787	104.38 ± 31.67	1.16.3.1	Ferroxidase
1788	105.25 ± 48.95	1.3.8.7	Medium-chain acyl-CoA dehydrogenase
1789	107.12 ± 25.82	1.2.1.3	Aldehyde dehydrogenase (NAD(+))
1790	107.50 ± 29.51	1.4.1.1	Alanine dehydrogenase
1791	108.38 ± 32.11	1.2.1.88	L-glutamate gamma-semialdehyde dehydrogenase
1792	108.50 ± 39.65	1.7.1.13	PreQ(1) synthase
1793	108.62 ± 42.35	1.3.1.14	Dihydroorotate dehydrogenase (NAD(+))
1794	108.75 ± 28.18	1.1.5.3	Glycerol-3-phosphate dehydrogenase
1795	109.00 ± 38.68	1.17.1.1	CDP-4-dehydro-6-deoxyglucose reductase
1796	109.00 ± 30.34	1.4.1.3	Glutamate dehydrogenase (NAD(P)(+))
1797	109.50 ± 43.30	1.10.2.2	Quinol--cytochrome-c reductase
1798	110.62 ± 51.51	1.7.2.1	Nitrite reductase (NO-forming)

1799	111.62 ± 55.35	1.7.99.1	Hydroxylamine reductase
1800	115.00 ± 47.75	1.2.99.2	Carbon-monoxide dehydrogenase (acceptor)

REFERENCES

1. C. Luo, D. Tsementzi, N. Kyrpides, T. Read, K. T. Konstantinidis, Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*. **7**, e30087 (2012).

2. M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40** (2012), doi:10.1093/nar/gkr988.

3. M. H. Saier *et al.*, The transporter classification database (TCDB): recent advances. *Nucleic Acids Res.* **44**, D372-9 (2016).

4. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*. **12**, 59–60 (2014).

5. L. Philippot *et al.*, The ecological coherence of high bacterial taxonomic ranks. *Nat. Rev. Microbiol.* **8**, 523–529 (2010).

6. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

7. P. J. McMurdie, S. Holmes, Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **10**, e1003531 (2014).

8. G. B. Gloor, J. M. Macklaim, V. Pawlowsky-Glahn, J. J. Egozcue, Microbiome datasets are compositional: And this is not optional. *Front. Microbiol.* **8**, 1–6 (2017).

9. C. Zhu *et al.*, Functional sequencing read annotation for high precision microbiome analysis. *Nucleic Acids Res.* (2018), doi:10.1093/nar/gkx1209.

10. D. N. Reshef *et al.*, Detecting novel associations in large data sets. *Science (80-.).* **334**, 1518–1524 (2011).

11. D. Tang, M. Wang, W. Zheng, H. Wang, RapidMic: Rapid computation of the maximal information coefficient. *Evol. Bioinforma.* **10**, 11–16 (2013).

12. M. Leonard, S. Graham, D. Bonacum, The human factor: The critical importance of effective teamwork and communication in providing safe care. *Qual. Saf. Heal. Care*. **13**, 361–362 (2004).

13. R. Suzuki, H. Shimodaira, pvclust : An R package for hierarchical clustering with p-values. *Bioinformatics*. **22**, 1–7 (2013).

14. K. R. Clarke, Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–143 (1993).

15. D. I. Warton, S. T. Wright, Y. Wang, Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol. Evol.* **3**, 89–101 (2012).

16. T. Lumley, P. Diehr, S. Emerson, L. Chen, The importance of the normality assumption in large public health data sets. *Annu. Rev. Public Health* (2002), doi:10.1146/annurev.publhealth.23.100901.140546.

17. A. Chao, L. Jost, Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*. **93**, 2533–2547 (2012).

18. A. Chao *et al.*, Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).

19. H. Tuomisto, A consistent terminology for quantifying species diversity? Yes, it does exist. *Oecologia*. **164**, 853–860 (2010).

20. C.-H. Chiu, A. Chao, Estimating and comparing microbial diversity in the presence of sequencing errors. *PeerJ*. **4**, e1634 (2016).

21. D. S. Jones, I. Schaperdoth, J. L. Macalady, Metagenomic evidence for sulfide oxidation in extremely acidic cave biofilms. *Geomicrobiol. J.* **31**, 194–204 (2014).

22. D. S. Jones *et al.*, Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. *ISME J.* **6** (2012), pp. 158–170.

23. P. J. Turnbaugh *et al.*, A core gut microbiome in obese and lean twins. *Nature*. **457**, 480–484 (2009).

24. K. Kurokawa *et al.*, Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14**, 169–181 (2007).

25. T. Yatsunenko *et al.*, Human gut microbiome viewed across age and geography. *Nature*. **486** (2012), pp. 222–227.

26. J. Lloyd-Price *et al.*, Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. **550**, 61–66 (2017).

27. K. S. Swanson *et al.*, Phylogenetic and gene-centric metagenomics of the canine intestinal microbiome reveals similarities with humans and mice. *ISME J.* **5**, 639–649 (2011).

28. B. M. Satinsky *et al.*, Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome*. **3**, 39–39 (2015).

29. R. Ghai *et al.*, Metagenomics of the water column in the pristine upper course of the Amazon river. *PLoS One*. **6**, e23785 (2011).

30. S. Oh *et al.*, Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl. Environ. Microbiol.* **77**, 6000–6011 (2011).

31. C. Simon, A. Wiezer, A. W. Strittmatter, R. Daniel, Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Appl. Environ. Microbiol.* **75**, 7519–7526 (2009).

32. P. Menzel *et al.*, Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* **70**, 411–424 (2015).

33. C. Lüke, D. R. Speth, M. A. R. Kox, L. Villanueva, M. S. M. Jetten, Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone. *PeerJ*. **4**, e1924 (2016).

34. F. J. Stewart, O. Ulloa, E. F. Delong, Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ. Microbiol.* **14**, 23–40 (2012).

35. J. F. Biddle, S. Fitz-Gibbon, S. C. Schuster, J. E. Brenchley, C. H. House, Metagenomic signatures of the Peru margin subseafloor biosphere show a genetically distinct environment. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10583–10588 (2008).

36. J. F. Biddle, J. R. White, A. P. Teske, C. H. House, Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. *ISME J.* **5**, 1038–1047 (2011).

37. A. Quaiser, Y. Zivanovic, D. Moreira, P. López-García, Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *ISME J.* **5**, 285–304 (2011).

38. O. E. Håvelsrud *et al.*, Metagenomic and geochemical characterization of pockmarked sediments overlaying the Troll petroleum reservoir in the North Sea. *BMC Microbiol.* **12**, 203 (2012).

39. O. Håvelsrud, T. H. Haverkamp, T. Kristensen, K. S. Jakobsen, A. Rike, A metagenomic study of methanotrophic microorganisms in coal oil point seep sediments. *BMC Microbiol.* **11**, 221 (2011).

40. I. P. G. Marshall, S. M. Karst, P. H. Nielsen, B. B. Jørgensen, Metagenomes from deep Baltic Sea sediments reveal how past and present environmental conditions determine microbial community composition. *Mar. Genomics* (2017), , doi:10.1016/j.margen.2017.08.004.

41. W. Zhang *et al.*, Synchronized dynamics of bacterial niche-specific functions during biofilm development in a cold seep brine pool. *Environ. Microbiol.* **17**, 4089–4104 (2015).

42. R. Stokke, I. Roalkvam, A. Lanzen, H. Haflidason, I. H. Steen, Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. *Environ. Microbiol.* **14**, 1333–1346 (2012).

43. F. Smedile *et al.*, Metagenomic analysis of hadopelagic microbial assemblages thriving at the deepest part of Mediterranean Sea, Matapan-Vavilov Deep. *Environ. Microbiol.* **15**, 167–182 (2013).

44. E. A. Eloe *et al.*, Going deeper: Metagenome of a hadopelagic microbial community. *PLoS One*. **6**, e20388 (2011).

45. K. T. Konstantinidis, J. Braff, D. M. Karl, E. F. DeLong, Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific Subtropical Gyre. *Appl. Environ. Microbiol.* **75**, 5345–5355 (2009).

46. S. Sunagawa *et al.*, Structure and function of the global ocean microbiome. *Science (80-.).* **348**, 1261359–1261359 (2015).

47. K. Anantharaman, J. A. Breier, G. J. Dick, Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. *ISME J.* **10**, 225–239 (2015).

48. K. Tang, K. Liu, N. Jiao, Y. Zhang, C. T. A. Chen, Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system. *PLoS One*. **8**, 1–11 (2013).

49. T. Urich *et al.*, Microbial community structure and functioning in marine sediments associated with diffuse hydrothermal venting assessed by integrated meta-omics. *Environ. Microbiol.* **16**, 2699–2710 (2014).

50. F. D. Andreote *et al.*, The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS One*. **7**, e38600 (2012).

51. H. Alzubaidy *et al.*, Rhizosphere microbiome metagenomics of gray mangroves (*Avicennia marina*) in the Red Sea. *Gene*. **576**, 626–636 (2016).

52. N. Fierer *et al.*, Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci.* **109**, 21390–21395 (2012).

53. P. T. Le *et al.*, Comparative metagenomic analysis reveals mechanisms for stress response in hypoliths from extreme hyperarid deserts. *Genome Biol. Evol.* **8**, 2737–2747 (2016).

54. F. J. Stewart, A. K. Sharma, J. A. Bryant, J. M. Eppley, E. F. DeLong, Community transcriptomics reveals universal patterns of protein sequence conservation in natural microbial communities. *Genome Biol.* **12**, R26 (2011).

55. P. Kanokratana *et al.*, Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. *Microb. Ecol.* **61**, 518–528 (2011).

56. T. O. Delmont *et al.*, Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J.* **6**, 1677–1687 (2012).

57. Y. Pan *et al.*, Impact of long-term N, P, K, and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil. *FEMS Microbiol. Ecol.* **90**, 195–205 (2014).

58. D. Bulgarelli *et al.*, Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe*. **17**, 392–403 (2015).

59. R. G. Taketani, V. N. Kavamura, R. Mendes, I. S. Melo, Functional congruence of rhizosphere microbial communities associated to leguminous tree from Brazilian semiarid region. *Environ. Microbiol. Rep.* **7**, 95–101 (2015).

60. A. M. Plominsky *et al.*, Metagenome sequencing of the microbial community of a solar saltern crystallizer pond at Cáhuil lagoon, Chile. *Genome Announc.* **2**, 1–2 (2014).

61. A. S. Pandit *et al.*, A snapshot of microbial communities from the Kutch: one of the largest salt deserts in the World. *Extremophiles*. **19**, 973–987 (2015).

62. R. Ghai *et al.*, New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* **1**, 135 (2011).

63. K. Rossmassler, T. E. Hanson, B. J. Campbell, Diverse sulfur metabolisms from two subterranean sulfidic spring systems. *FEMS Microbiol. Lett.* **363** (2016), doi:10.1093/femsle/fnw162.

64. C. Magnabosco *et al.*, A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J.* **10**, 730–741 (2016).

CHAPTER 3. PROTEIN CODING POTENTIAL OF NUCLEOTIDE SEQUENCES BASED ON KMERS

MANUSCRIPT TEXT

ABSTRACT

Transcription of unexpected genomic segments is a generalised phenomenon that has been reported over the past few years. Recent studies have also revealed that part of these transcripts, previously declared as non-coding RNA, can actually be translated. These genetic elements are normally neglected by standard genome annotations. Whereas, current bioinformatic techniques for detecting these elements typically define a protein-coding potential based on evolutionary sequence conservation, or on the assumption that RNA sequences can be exclusively divided into protein coding and non-protein coding classes. Here, we describe a protein-coding measure based solely on the occurrence of in-frame Kmers more frequently found in protein-coding gene sequence databases. First, we evaluate different K values, finding best results with Kmer sizes from 9 to 12. Then, we use 9-mers to compare predicted high protein-coding potential regions with genes from some annotated genomes. Our results not only reproduce those from sophisticated gene-finders, but also reveal additional potential protein coding segments, which in many cases bear high homology with sequences from popular protein sequence databases. The simplicity of this method should imply a broad applicability, and thus it is expected that it can be used to explore and improve the understanding of complex genomic contexts.

INTRODUCTION

Historically, it has been assumed that nucleotide sequences can be separated into two classes: protein coding and non-protein coding sequences (hereinafter, coding and noncoding sequences). This assumption comes in part from the fact that protein and RNA folding are two complex and very different processes, thus their associated sequences should somehow encode these differences. Early attempts to detect coding segments within genomic sequences include analyses for the discrimination of introns from exons in eukaryotic sequences (1). Subsequent statistical analyses of amino acid sequences, and coding and noncoding nucleotide sequences showed that, in general, it is not possible to recognise amino acid sequences in proper proteins from randomly generated amino acid sequences; or coding from noncoding nucleotide sequences (2–4). One of the reasons for this may be that some protein-coding genes can simultaneously encode regulatory genes (e.g. riboswitches), or even alternative protein-coding genes in the antisense strand (5–8). However, proteomic and sequence homology research have allowed associating a protein-coding characteristic to many genes in previous years; whereas a lack of associated protein information, or RNA secondary structure validation has

been used to populate databases of noncoding genes. However, recent techniques, such as ribosome profiling, have revealed that translation is much more pervasive than previously thought, and that many sequences classified as noncoding can actually be translated (9,10). Thus, according to common practice, a single nucleotide sequence can be coding and noncoding at the same time, blurring the line between these supposed different classes of sequences (10,11). On the other hand, the increasing evidence that many transcripts do not correspond to annotated genes indicates that transcription is also more widespread in the genomes than anticipated (10–13). To elucidate the function of this so-called transcribed 'dark matter' (14) a number of methods have been designed to measure a protein-coding potential for transcripts, as a first step in the annotation of these sequences. These methods include alignment-based techniques that depend on the availability of conserved sequences in representative genomes (15–17). Other methods involve models that impose to coding RNA sequences gene structure constraints, some of which are lineage-specific (18). Most of these recent developments are implemented with machine learning algorithms to discriminate between coding and noncoding classes of sequences (16,19,20). These algorithms demand training datasets strictly classified for the learning process, which can be difficult to establish in the aforedescribed context. All these implementation designs can restrict the scope in which these tools can be used. For example, in a recent study of μ-proteins (short proteins with less than 80 amino acids) in cyanobacteria, the authors had to use comparative genomic and transcriptomic methods to predict some of these elements, which was only possible by the availability of closely related genomic sequences in those particular lineages (21). Moreover, current analyses of metagenomic and metatranscriptomic data are typically carried out only by normal sequence homology searches in protein and RNA databases; leaving a high number of sequences out of posterior analyses. Metagenomic and metatranscriptomic data from microbial populations are currently abundant in databases, and it is expected to further expand as part of several high impact initiatives such as the Human Microbiome (22) and the Earth Microbiome projects (23).

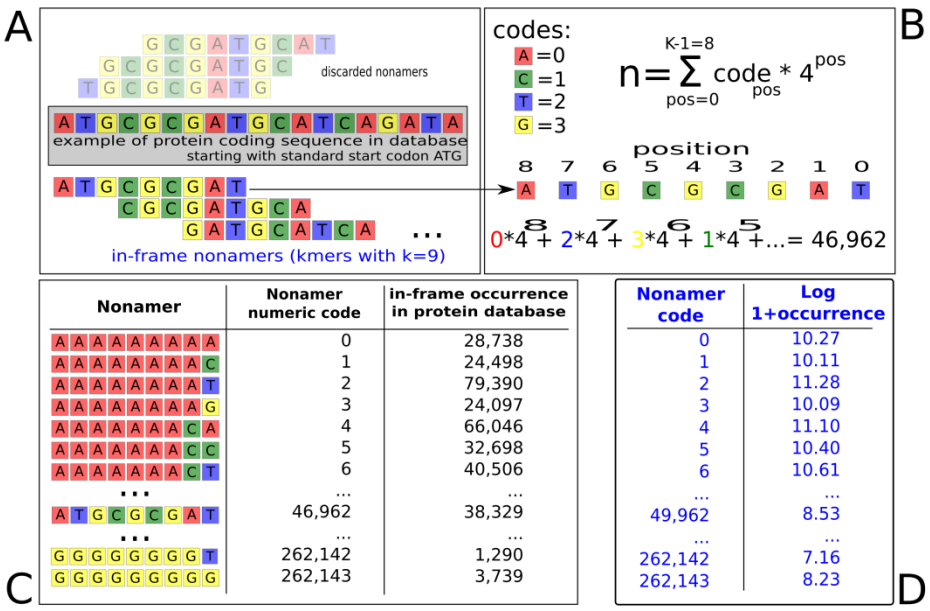
Combined, this suggests that a measure of protein-coding potential applicable to a wide range of nucleotide sequences has to be: a) a numeric value that somehow represents a probability of protein-coding, avoiding exclusive categorisation of sequences into coding and noncoding; b) simple, in order to avoid lineage-specific constraints; and c) based on features present only in reliable protein-coding gene sequences, since the noncoding characteristic have shown to be rather volatile in many cases. In previous works, biases in the hexamers (6-mers, six contiguous nucleotides) usage have been regarded as one of the most discriminatory features for classifying coding and noncoding sequences (1,20,24). Here, we describe a simple method for measuring the protein-coding potential of nucleotide sequences based solely on the occurrence of in-frame Kmers more frequently found in protein-coding gene sequence databases. To this end we first evaluated different Kmers (from 5- to 13-mers), and found significant evidence that

best results are obtained with 9- to 12-mers. Accordingly, we used 9-mers (or nonamers) for simplicity in subsequent analyses. We compared predicted high protein-coding potential regions with genes from some annotated genomes. Our results in most cases coincided with those from current state-of-the-art gene-finders, and also revealed some protein-coding segments missed by these tools, but still possessing high homology with sequences from protein sequence databases.

MATERIALS AND METHODS

Protein coding-potential measure based on in-frame Kmers. Given a database of protein-coding sequences, it is possible to compute the frequencies for each kmer occurring as in-frame in those sequences (Section 1, Supplementary Material). Then, the logarithm of these frequencies is considered in order to reduce the effect of possible uneven representation of some type of protein sequences in the database (Fig. 1). The protein-coding database used in this study was constructed from all CDS (Coding Data Sequence) in FASTA files downloaded from the NCBI FTP site of bacterial genomes (more details in Supplementary Material). To detect high protein-coding potential regions within genomic sequences, we defined an algorithm that considers the differences between the log frequency of each Kmer (from Log Count Tables, LCT_k tables, Fig. 1) and an arbitrary zero potential. A simple value for this parameter can be the average of all non-zero frequencies (Table S1). The sum of these differences is then stored into a buffer potential variable with a maximum value (buffer size), recording the position of the Kmer that started a positive sum and the number of Kmers that keep this sum positive. When this sum reaches zero (because of negative differences coming from Kmer frequencies below the zero potential) then the high protein-coding potential region ends, and the region is only considered if the number of Kmers exceeds a predefined minimum number of Kmers. A simple sketch of this algorithm is presented below:

```
// INPUT: sequence, zero_potential, min_kmers, buffer_size, K
region_ini = region_fin = -1;
num_kmers = buffer = 0;
for each in-frame kmer at position pos within sequence {
        buffer = min(buffer_size, buffer + (LCTk(kmer) - zero_potential));
        if ( buffer <= 0 ) {
                if ( num_kmers >= min_kmers )
                        print("region detected from region_ini to pos")
                buffer = num_kmers = 0;
                region_ini = -1;
        } else {
                if ( region_ini < 0 ) region_ini = pos;
                num_kmers = num_kmers + 1;
        }
}
```



Nonamer	Nonamer numeric code	in-frame occurrence in protein database
AAAAAAAAA	0	28,738
AAAAAAAAC	1	24,498
AAAAAAAAT	2	79,390
AAAAAAAAG	3	24,097
AAAAAAACA	4	66,046
AAAAAAACC	5	32,698
AAAAAAACT	6	40,506
...	...	...
ATGCGCGAT	46,962	38,329
...	...	...
GGGGGGGGT	262,142	1,290
GGGGGGGGG	262,143	3,739

Nonamer code	Log 1+occurrence
0	10.27
1	10.11
2	11.28
3	10.09
4	11.10
5	10.40
6	10.61
...	...
49,962	8.53
...	...
262,142	7.16
262,143	8.23

Figure. 1. Construction of a table of log frequencies of in-frame nonamers for use in coding-potential estimation. A) Given a database of protein-coding gene sequences (CDS or coding data sequence), the process starts by analysing each sequence in terms of its compounding in-frame nonamers. The first nonamer is the subsequence starting at position 1 and ending at position 9. The second nonamer starts at position 4, ending at position 12, and so on. The rest of out-frame nonamers are simply discarded. B) Each nonamer is assigned a number making use of the well-known formula for numeric base conversion. In this case, each nucleotide is given a number (numeric code from 0 to 3), and we consider nonamers as numbers in base 4, converting them to classical decimal base and using this number as a perfect hash number in a table. C) Occurrences of each nonamer are counted over the entire CDS database and stored in a table, indexed by the previously described nonamer numeric code (hash number). D) The number of occurrences for each nonamer (described by hash number from 0 to 4^9=262,143) is then applied the conversion 1 + Log to reduce undesirable effects of uneven sequencing efforts on certain taxa (and consequently in some recurrent proteins encoded by their genomes). A final table (Log Count Table, LCT_k) associating a numeric value (log count of occurrences, a number between 1 and ~12 in the figure) to each nonamer (represented by the hash number described above) is then ready to use to compute coding-potential to input transcriptomic or genomic sequences.

All genomic analyses presented here used a buffer size of 10. Computer programs written in C were developed to carry out these procedures (Section 2, Supplementary Material). Free access to this software and online analyses can be found at http://bioinf.udec.cl/kodpot.

Evaluation of different sets of Kmers for measuring the protein-coding potential. Biases in hexamer (6-mer) usage in nucleotide sequences have been extensively used as a discriminant feature between coding and noncoding sequences (1,20,24). Since no explicit justification for this particular Kmer size choice has been reported, we first evaluated different Kmer sizes (from k=5 to 13) as features for calculating a protein-coding potential measure. To compare the different sets of Kmer frequencies, each of these sets were used to predict the frame in which the BLASTX algorithm found the best homology in short metagenomic sequence reads. The datasets used in this study were 164 metagenomes from different biomes (Table S2), ensuring a high diversity both functionally and taxonomically, which allows an unbiased Kmer evaluation. Given LCT_k tables for *k=1..13*, constructed as in Fig.1 for *k=9* the following formula:

$$AvgCP_i = \frac{(\sum_{j=1}^{N_i} kmer\_lc_j)_i}{N_i},\ i = 1,\ldots,6$$

was used to compute a coding potential for each of the six possible frames for every sequence read. $AvgCP_i$ represents the respective average coding-potentials for the six frames ($i=1..6$); $kmer\_kc_j$ are the log counts (from LCT_k tables) for each of the N_i in-frame $kmers_j$ ($j=1..N_i$) observed in each frame of each metagenomic sequence read. The frame i, in which $AvgCP_i$ is maximal was considered the frame prediction for protein-coding potential. This frame was then compared to the frame in which Diamond/BLASTX (25) found the best hit in the Pfam database (26), with a cut-off bitscore of 50.

If we consider an uninformed or random predictor for the frame in which a nucleotide sequence encodes a protein, we have to expect a success ratio of 1/7 (six possible frames plus one option for noncoding sequences, i.e. ~14% of success). Our prediction of protein-coding frame of metagenomic sequence reads were above 80% of success when using Kmers from 6- to 12-mers, with best results in the range of 9- to 12-mers (Fig. S1).

Plotting signals of protein coding-potential of nucleotide sequences. To plot protein-coding signals for each frame in genomic sequences, the same procedure described above was applied to moving averages of regions along the sequences. The window length considered was varied depending on the length of the sequence to analyse. For each of those moving windows, an average was computed using the above formula, and plotted in different colours for each of the six possible frames. A zero-potential (described above) line was drawn for each strand in all of the plots.

RESULTS

Given the proposed method for calculating a protein-coding potential, which requires a choice in the Kmers size to use, we found significant differences between K = 5 to 13. The best results were achieved with K = 9 to 12 (Fig. S1). Thus, for simplicity, we used 9-mers (K=9, nonamers) in all subsequent analyses to evaluate the applicability of this method for analysing generic genomic data. In particular, we used nonamer log counts to estimate variations in the coding-potential of genomic regions, and graphically see this as signals for each genomic frame. To this end, moving averages of nonamers log counts were computed for each frame. Figure 2 shows the genomic region encoding the ammonia monooxygenase enzyme (*amoA*, *amoB* and *amoC* subunit genes) of *Nitrosopumilus maritimus* SCM1. It can be appreciated that whenever the genome encodes a protein, only one frame shows values consistently above a zero potential (horizontal lines within the signals), while the others display, in general, a noise-like behaviour. As the genomic region including multiple genes was taken as a unit, the frames in which the different genes are encoded, changes from one to another. For example, the genes with locus tag Nmar_1498 and Nmar_1499 are both encoded in the frame displayed as blue, while Nmar_1503 (*amoB*) is encoded in the red frame, relative to the start coordinate of this region (1,365,300) (Fig. 2A). A similar situation occurs in the reverse strand, which encodes *amoA* and

amoC. It is also worth noting that most intergenic regions display a noise-like behaviour in all the frames (Fig. 2A). Additionally, this analysis revealed other small protein-coding segments in regions commonly regarded as noncoding. This can be visualised in Fig. 2B, within the large ribosomal subunit 23S gene (between coordinates 893,640 and 894,042 in the reverse strand). To check if the high protein-coding potential found for this region can actually correspond to a protein sequence homology (or simply be an artefact of the method), we translated this region and used BLASTP against the NCBI-nr database and detected an archaeal rRNA intron-encoded endonuclease among the highest scored hits (bitscore=52, e-value=$2x10^{-5}$). The existence of this type of endonuclease within rRNA sequences has been previously reported (27), but it was missed in the official annotation of this genome (28). Note how the rest of this sequence displays noise-like signals in the coding-potential for all six frames (Fig. 2B), indicating that this measure is controlled in undesirable false positive assessments of high coding-potential regions.

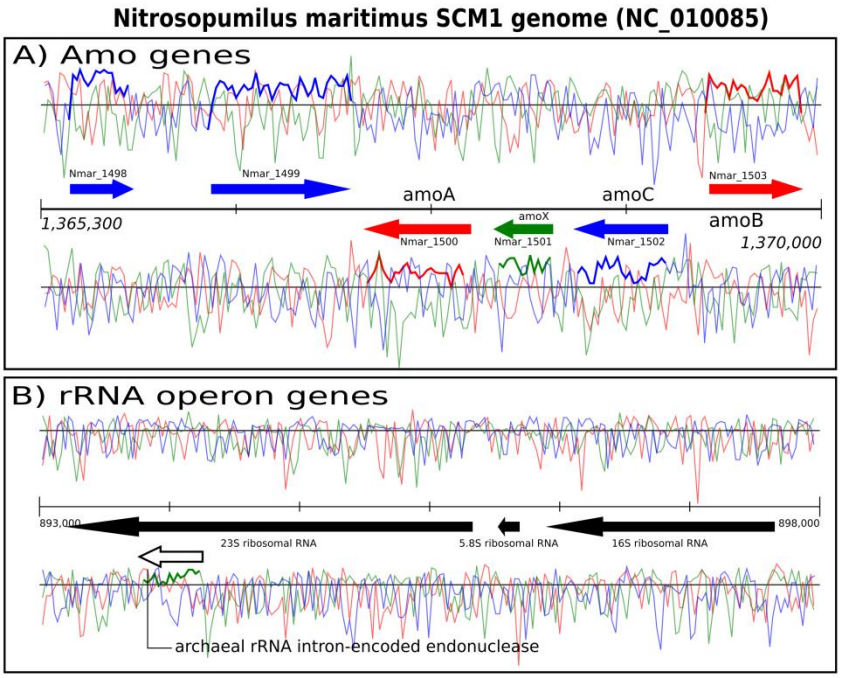


Figure. 2. Nonamer-based coding analysis of two genomic regions of *Nitrosopumilus maritimus* SCM1 archaeal genome. A) Genomic region encoding ammonia oxidation genes (*amoACB*) and other neighbouring genes. In the centre of the diagram is shown the annotated genes and both above and below this annotation are three frame signals (in red, blue and green), corresponding to frames in the forward and the complement reverse strand, respectively. The lines around which the frame signals oscillate represent a zero-potential value. Note how the protein-coding potential for a frame is consistently above this zero-potential wherever a gene is encoded. B) Genomic region associated to rRNA genes. Unlike the above case, these genes are not protein-coding, thus their associated frames (both forward and reverse) do not get above zero-potential, except for one small region within the 23S gene. This protein-coding region was missed by gene-finders in the official genome release at RefSeq (NC_010085). However, a BLASTX of the corresponding sequence reveals a significant hit to an archaeal rRNA intron-encoded endonuclease.

To show the behaviour of this coding-potential method in genomic regions whose genes have not been used in the computation of the log counts of nonamers (LCT_k tables), we analysed a recent draft genome of the cyanobacterium *Hassalia byssoidea* VB51217 (29). The result of this analysis for a region of this sequence is shown in Figure 3. As can be noted, all annotated CDSs within this region were predicted to have a high coding potential. However, a number of additional regions were predicted with high coding-potential as well, most of them with confirmed significant protein homology in current databases (e.g. filamentous hemagglutinin

outer membrane protein genes, Fig. 3). A closer inspection of the official annotation (carried out by the NCBI Prokaryotic Genome Annotation Pipeline) showed that these regions were indeed annotated as *pseudo-genes* without CDS, indicating that the gene-finder software GeneMarkS (30) was unable to find an expected gene signal in these regions. However, a direct sequence alignment search actually found a protein sequence homology, as confirmed by our high coding-potential assessment and the subsequent manual search performed on them to verify the prediction.

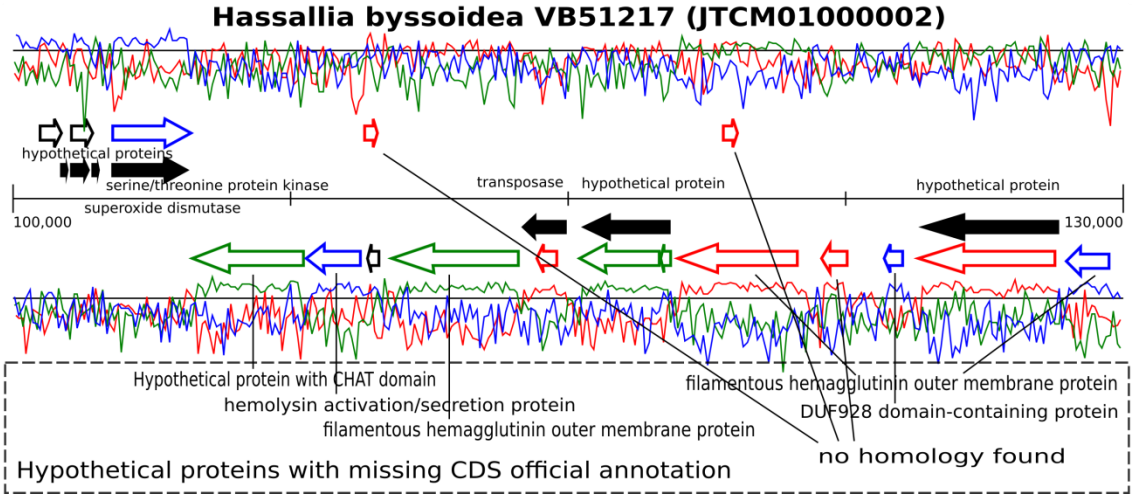


Figure. 3. Nonamer-based coding potential analysis of a region from the draft genome of the cyanobacterium *Hassallia byssoidea* VB51217. Filled arrows represent genes in the annotation carried out by the NCBI Prokaryotic Genome Annotation Pipeline. Coloured empty arrows, on the other hand, represent regions in which the protein-coding potential by nonamers gets consistently higher than zero-potential. All annotated genes were predicted by this potential, plus a number of other regions, called pseudo-genes by the above mentioned pipeline because the underlying gene-finder software was unable to detect them. All these regions have significant homologies in protein databases, except for a few small regions with no known homology.

Taking into account that the log counts of Kmers (nonamers in particular) were computed from a database of bacterial genes (see above), we examined the performance of our method on genomic eukaryotic sequences, which represent a different and more challenging test, as normal sequence repetitions and the splitting of genes into spatially separated exons add a layer of complexity to the detection of related protein-coding regions. Results of the analysis we carried out of the protein-coding potential of region [31,873,100 - 31,925,300] from chromosome 21 of the human genome sequence GRCh38 are shown in Figure 4 as an example. The official annotation of this region currently includes only two exons of the protein ENSG00000142149, and has been previously analysed to demonstrate the good prediction of a previous protein-coding potential tool (31). In our analysis these two exons are detected as high-potential regions with simultaneous signals in multiple frames (displayed in the upper left and right corners in Figure 4). These parallel high coding potential signals may represent different stages in the evolution of overlapping genes (32,33). In addition, a number of other regions were also predicted with high protein-coding potential, some of them with noteworthy homology to protein sequences in Swiss-Prot (34) and Pfam (26) databases. Three of them,

encoded in the reverse strand, scored highly to the same protein sequence FLJ38264 in Swiss-Prot database (as determined by the best BLAST hit with e-value < 10^{-5} and bitscore > 50), displayed in the lower left corner in Fig. 4). Moreover, one of these regions shared the same Pfam domain (activator of Hsp90 ATPase, PF08327) with another region located ca. 30,000 bp upstream in the same strand. More extensive analysis of these sequences is beyond the scope of this article, but although only 4 out of 16 detected regions presented confirmed protein homology in current databases, the result of this simple analysis may have a significance worth of more in-depth analyses. For example, eukaryotic protein-coding sequences may be used to compute the LCT_k tables (Material and Methods) to see if the same or new regions with high coding potential are detected.

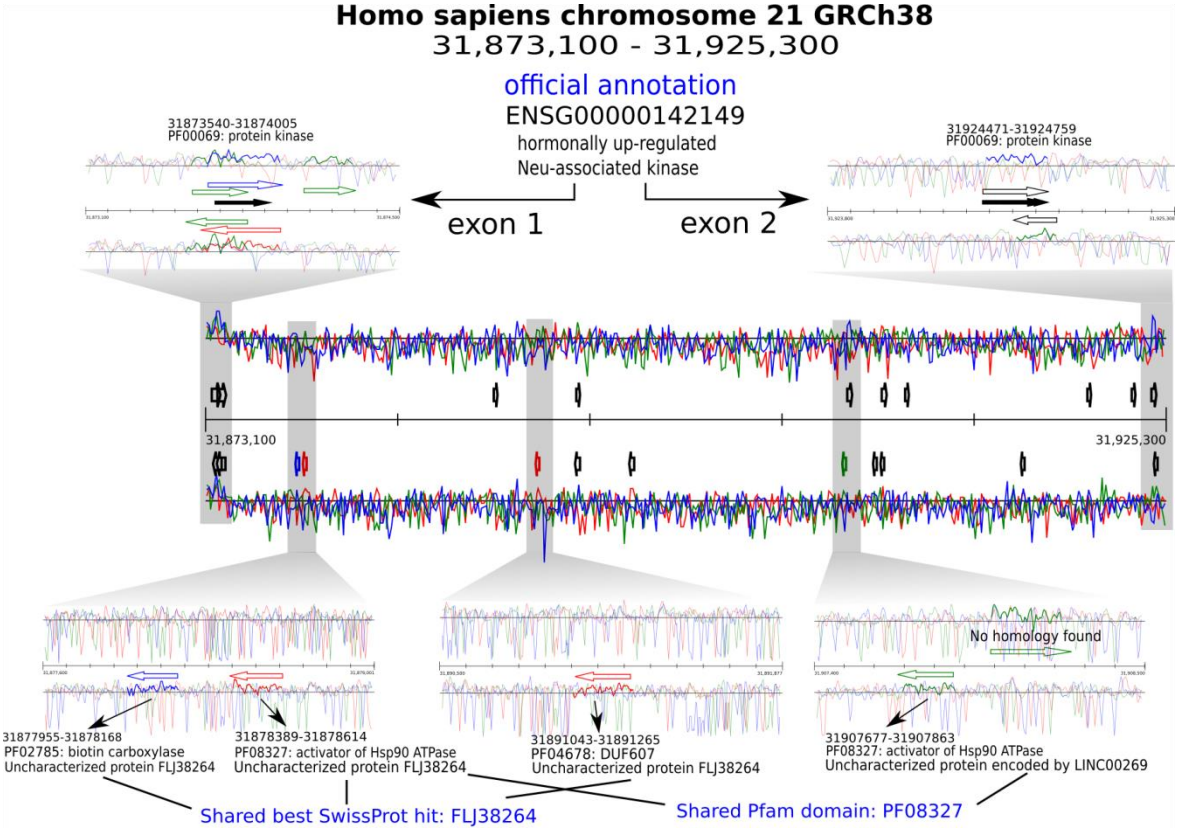


Figure. 4. Nonamer-based coding potential analysis of a section of chromosome 21 from the human genome. The sequence region GRCh38 encodes two separated exons from protein ENSG00000142149 and has previously been used to compare the performance of translation initiation site predictors (31). Here, we show that the coding-potential measure can also detect the annotated exons (upper left and right). A number of other protein-coding regions were also predicted, apparently with a coherent context of homologies (lower). All of these regions are missed in the official annotation, despite their homology significance in protein databases and the apparent relation among them. Small arrows in black represent predicted regions with coding-potential without homology in current protein databases. The detection of high protein-coding regions in parallel regions may represent an early stage in the formation of overlapping genes (33).

DISCUSSION

An effective estimation of the protein-coding potential of nucleotide sequences can be applied to several challenging and important bioinformatic tasks, such as genomic gene prediction,

annotation of high numbers of unexpected transcribed 'dark matter', and protein homology search of short metagenomic and metatranscriptomic sequences. A robust gene calling assessment process should avoid two undesirable extreme behaviours: calling every genomic region a gene and missing important genes that may hinder our understanding of the genomic complexity developed by biological systems during evolution. Along this line of thought, it has been previously demonstrated that in many cases simpler models normally achieve better results for predicting genes in general data (31). Current coding-potential prediction developments and *ab-initio* gene finders rely on a series of assumptions for their models, typically considering multiple features in the sequences (e.g. specific start and stop codons, ORF size, Fickett score and hexamer score). Here, we focused on determining the best Kmer option for estimating a protein-coding potential, and establishing whether this set of Kmers alone is able to provide an efficient protein-coding potential measure. Our results show that a measure of protein-coding potential based on log frequencies of in-frame Kmers from a protein database can, in general, predict the protein coding frame of metagenomic sequences, and that among these Kmers sets, Kmers from 9- to 11-mers are the best predictors of this characteristic (Figure S1). Based on this result, we further explored the use of this measure to predict protein-coding regions within genomic regions. The result of these analyses shows a coincidence with the CDS annotated by state-of-the-art annotation pipelines, plus a number of other predictions, like *pseudo-genes* or genes encoding µ-proteins that current gene-finders are unable to detect. These elements are normally either missed or annotated only by direct searches against protein databases, a process whose success depends on the existence of close homologues (Fig. 2 and 3).

The use of hexamers for discriminating protein-coding and noncoding sequences is a technique that has been used for decades (1,20,24). In order to assign numerical values to these sequence features, these methods rely on the existence of pre-classified protein-coding and noncoding sequences. Historically, the term noncoding has been used to refer to sequences with a biological functionality as RNA (e.g. regulation), without being first translated. However, RNA catalytic functionality does not necessarily exclude subsequent translation (11). It has been reported that the manually curated Swiss-Prot database of proteins would include up to 10% of erroneously translated noncoding RNA (18). Although no database can be totally free of errors, this claim was based on the application of noncoding classifier programs implementing a number of debatable constraints (e.g. that protein-coding genes cannot overlap), or trained with some arbitrarily pre-classified noncoding sequences. On the other hand, databases of sequences commonly regarded as noncoding, apparently include even higher percentages of protein-coding sequences. For example, the high-confident subset of LNCipedia (35), a database of human long noncoding RNA, includes more than 30% of sequences with a BLASTX result (e-value < 10^{-5}, bitscore > 50) against Pfam or Uniref90 protein databases (Tables S3 and S4). This view is apparently more congruent with recent experimental studies that have

demonstrated that sequences previously regarded as noncoding can actually encode proteins, or at least be translated as such (9,10). Thus, it is seemingly not possible to establish strictly separated datasets of coding and noncoding sequences, required for training of binary classifier models used in many modern methods for estimating protein-coding potential. This problem, together with the aforementioned use of complex models with many assumptions constitute the ingredients for the so-called overfitting problem in machine learning techniques, which may prevent the applicability of these methods to more general genomic contexts. For example, it has been reported that current gene-finding methods have low success in some underrepresented protist genomes, mainly because of the presence of a high number of overlapping genes (36). Here, we first showed that nonamers are better than hexamers recognising general protein-coding nucleotide sequences; when a value for these features is computed as log counts of their in-frame occurrences in a dataset comprised solely of protein-coding sequences. In this way, only in-frame Kmers count for protein-coding features, keeping low counts for recurrent off-frame Kmers assumed to be randomly distributed in sequences devoid of translation activity. Our results suggest that a protein-coding potential solely based on nonamers can be useful in the prediction of protein-coding genes both in short metagenomic/metatranscriptomic sequences and in general genomic data. In the former case, a prior coding-potential analysis of input sequences can allow DNA-to-protein aligner programs to reduce the number of searches by predicting the frame (or at least reducing their number) in which a sequence is most likely encoding a protein. In the latter case, the parallel determination of a protein-coding potential for each genomic frame avoiding the use of lineage-specific constraints, can provide a more complete interpretation of general and complex genomic sequences.

REFERENCES

1. Claverie JM, Bougueleret L. Heuristic informational analysis of sequences. Nucleic Acids Res. 1986;14(1):179–96.

2. Weiss O, Jiménez-Montaño M a, Herzel H. Information content of protein sequences. J Theor Biol. 2000;206(3):379–86.

3. Mantegna RN, Buldyrev S V., Goldberger AL, Havlin S, Peng CK, Simons M, et al. Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics. Phys Rev E. 1995;52(3):2939–50.

4. Lavelle DT, Pearson WR. Globally, unrelated protein sequences appear random. Bioinformatics [Internet]. 2010 Feb 1 [cited 2014 Dec 25];26(3):310–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2852211&tool=pmcentrez&rendertype=abstract

5. Fahey ME, Moore TF, Higgins DG. Overlapping antisense transcription in the human genome. Comp Funct Genomics. 2002;3(3):244–53.

6. Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G. In search of antisense. Trends Biochem Sci. 2004;29(2):88–94.

7. Georg J, Hess WR. Natural antisense transcripts in Bacteria. Regul RNAs [Internet]. 2012;95–108. Available from: http://dx.doi.org/10.1007/978-3-7091-0218-3_5

8. Serganov A, Patel DJ. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet. 2007;8(10):776–90.

9. Ingolia NT, Brar GA, Stern-Ginossar N, Harris MS, Talhouarne GJS, Jackson SE, et al. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. Cell Rep. 2014;1365–79.

10. Pauli A, Valen E, Schier AF. Identifying (non-)coding RNAs and small peptides: Challenges and opportunities. BioEssays [Internet]. 2015;37(1):103–12. Available from: http://doi.wiley.com/10.1002/bies.201400103

11. Dinger ME, Pang KC, Mercer TR, Mattick JS. Differentiating protein-coding and noncoding RNA: Challenges and ambiguities. PLoS Comput Biol. 2008;4(11):e1000176.

12. Kapranov P, St Laurent G, Raz T, Ozsolak F, Reynolds CP, Sorensen PHB, et al. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. BMC Biol. 2010;8:149.

13. Mercer TR, Gerhardt DJ, Dinger ME, Crawford J, Trapnell C, Jeddeloh J a, et al. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nat Biotechnol [Internet]. 2011;30(1):99–104. Available from: http://dx.doi.org/10.1038/nbt.2024

14. Ponting CP, Grant Belgard T. Transcribed dark matter: Meaning or myth? Hum Mol Genet. 2010;19(R2):162–8.

15. Washietl S, Findeiss S, Müller SA, Kalkhof S, von Bergen M, Hofacker IL, et al. RNAcode: robust discrimination of coding and noncoding regions in comparative sequence data. RNA [Internet]. 2011;17(4):578–94. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3062170&tool=pmcentrez&rendertype=abstract

16. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35(SUPPL.2):345–9.

17. Lin MF, Jungreis I, Kellis M. PhyloCSF: A comparative genomics method to distinguish protein coding and non-coding regions. Bioinformatics. 2011;27(13):i275–82.

18. Frith MC, Bailey TL, Kasukawa T, Mignone F, Kummerfeld SK, Madera M, et al. Discrimination of non-protein-coding transcripts from protein-coding mRNA. RNA Biol. 2006;3(1):40–8.

19. Liu J, Gough J, Rost B. Distinguishing protein-coding from non-coding RNAs through support vector machines. PLoS Genet. 2006;2(4):529–36.

20. Wang L, Park HJ, Dasari S, Wang S, Kocher J-PP, Li W. CPAT: Coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41(6):1–7.

21. Baumgartner D, Kopf M, Klähn S, Steglich C, Hess WR. Small proteins in cyanobacteria provide a paradigm for the functional analysis of the bacterial micro-proteome. BMC Microbiol [Internet]. 2016;16(1):285. Available from: http://bmcmicrobiol.biomedcentral.com/articles/10.1186/s12866-016-0896-z%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/27894276

22. Integrative T. The integrative Human Microbiome Project: Dynamic analysis of Microbiome-Host omics profiles during periods of human health and disease. Cell Host Microbe [Internet]. 2014;16(3):276–89. Available from: http://dx.doi.org/10.1016/j.chom.2014.08.014

23. Gilbert JA, Meyer F, Antonopoulos D, Balaji P, Brown CT, Brown CT, et al. Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. Stand Genomic Sci. 2010;3(3):243.

24. Fickett JW, Tung CS. Assessment of protein coding measures. Nucleic Acids Res. 1992;20(24):6441–50.

25. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods [Internet]. 2014 Jan;12(1):59–60. Available from: http://dx.doi.org/10.1038/nmeth.3176

26. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. Vol. 42, Nucleic Acids Res. 2014. p. D222–30.

27. Lambowitz AM, Belfort M. Introns as mobile genetic elements. Annu Rev Biochem. 1993;62:587–622.

28. Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, et al. *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally

distributed marine crenarchaea. Proc Natl Acad Sci USA [Internet]. 2010;107(19):8818–23. Available from: http://www.pnas.org/content/107/19/8818.long

29. Singh D, Chandrababunaidu M, Panda A, Sen D, Bhattacharyya S, Adhikary P. Draft genome sequence of cyanobacterium *Hassallia byssoidea* strain VB512170, isolated from monuments in India. J Bacteriol. 2015;3(2):2014–5.

30. Borodovsky M, Lomsadze A. Gene identification in prokaryotic genomes, phages, metagenomes, and EST sequences with GeneMarkS suite. Curr Protoc Microbiol. 2014;(SUPPL.32).

31. Saeys Y, Abeel T, Degroeve S, Van de Peer Y. Translation initiation site prediction on a genomic scale: Beauty in simplicity. Bioinformatics. 2007;23(13):418–23.

32. Veeramachaneni V. Mammalian overlapping genes: The comparative perspective. Genome Res [Internet]. 2004;14(2):280–6. Available from: http://www.genome.org/cgi/doi/10.1101/gr.1590904

33. Fonseca MM, Harris DJ, Posada D. Origin and length distribution of unidirectional prokaryotic overlapping genes. G3 (Bethesda) [Internet]. 2014;4(1):19–27. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3887535&tool=pmcentrez&rendertype=abstract

34. Consortium TU. UniProt: a hub for protein information. Nucleic Acids Res [Internet]. 2014;43(D1):D204–12. Available from: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku989

35. Volders P-J, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res [Internet]. 2014;43(D1):D174–80. Available from: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1060

36. Caron D a, Worden AZ, Countway PD, Demir E, Heidelberg KB. Protists are microbes too: a perspective. ISME J. 2009;3(1):4–12.

SUPPLEMENTARY INFORMATION APPENDIX (SI APPENDIX)

SECTION 1. IN-FRAME KMERS OF CODING GENES IN SEQUENCED BACTERIAL GENOMES

Here we consider a kmer as a subsequence of length k of a nucleotide sequence. An in-frame kmer is a kmer that starts in a position multiple of 3 from a protein-coding gene sequence. If, for example, we have the sequence AGCTGATAGCTTAGATAA, then it contains 4 nonamers (kmer with k=9) in the forward direction, namely AGCTGATAG, TGATAGCTT, TAGCTTAGA and CTTAGATAA. In-frame k-mers from coding gene sequences were retrieved and counted for calculating their occurrences in currently sequenced bacterial genomes (*_cds_from_genomic.fna.gz files retrieved from complete genomes listed in ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/bacteria/assembly_summary.txt as of January 2017). As these data are expected to be biased by uneven sequencing efforts in different taxonomical groups, the logarithm of these frequencies values were used to reduce this effect. The results of the calculations described in Materials and Methods are presented in Table S1.

SECTION 2. C PROGRAM TO PROCESS IN-FRAME KMERS FROM GENE SEQUENCES

This program computes the log count values for kmers present in the sequences in the FASTA files within a subdirectory. This program can be compiled in any Unix system (e.g. Linux) with the command:

gcc -O3 -o output_program_name source_c_file.c

For easy access the source code of this program is displayed here and is also provided as a separate C file.

```c
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <ctype.h>
#include <math.h>
#include <dirent.h>
#include <unistd.h>
#include <sys/types.h>

#define MAXSEQNAMELEN     256
#define MAXSEQLEN        8192
#define  MAXLINELEN       8192

struct Seq {
        char name[MAXSEQNAMELEN];
        char seq[MAXSEQLEN];
};

int getNextSeqFromFastaFile( struct Seq *seq, FILE *f );
int ACTG2int( char *actg, int len );

int main( int argc, char *argv[] ) {
struct Seq seq;
FILE *f;
int K, maxidx, *kmers_arr, num, i, j, l, slen;
float avg, var, logsum, logsum2;
char *kmer, fpath[8192];
DIR *dir;
struct dirent *ent;

        if( argc != 3 ) {
                fprintf(stderr, "Usage: %s <K> <ffn_dir>\n", argv[0]);
                return 1;
```

```
        }

        dir = opendir( argv[2] );
        if( !dir ) {
                fprintf(stderr, "Error: can't open ffn_files directory\n");
                return 1;
        }
        K = atoi( argv[1] );
        maxidx = pow( 4, K );
        kmers_arr = (int *)malloc( (maxidx + 1) * sizeof(int) );
        if( !kmers_arr ) { fprintf(stderr, "malloc failed\n"); return 1; }
        memset( kmers_arr, 0, sizeof(int) * maxidx );
        kmer = (char *)malloc( (K+1) * sizeof(char) );
        if( !kmer ) { fprintf(stderr, "malloc failed\n"); return 1; }
        while( (ent = readdir(dir)) != NULL ) {
                if( ent->d_type != DT_REG ) continue;
                sprintf( fpath, "%s/%s", argv[2], ent->d_name );
                f = fopen( fpath, "r" );
                if( !f ) { fprintf(stderr, "fopen(%s) failed\n", fpath); continue; }
                while( getNextSeqFromFastaFile( &seq, f ) ) {
                        slen = strlen( seq.seq );
                        for( i = 0; i < slen - K; i += 3 ) {
                                for( j = 0, l = i; j < K; j++, l++ )
                                        kmer[j] = seq.seq[l];
                                kmer[j] = 0;
                                num = ACTG2int(  kmer, K );
                                if( num >= 0 ) kmers_arr[num] += 1;
                        }
                }
                fclose( f );
        }
        printf("\n");

        sprintf( fpath, "cdskmerf_%d.db", K );
        f = fopen( fpath, "w" );
        if( !f ) { printf("Can't open cdskmerf.db\n"); return 1; }

        logsum = logsum2 = 0;
        for( i = 0; i < maxidx; i++ ) {
                float logkcount;
                logkcount = log( 1 + kmers_arr[i] );
                fwrite( &logkcount, sizeof(float), 1, f );
                logsum += logkcount;
                logsum2 += logkcount * logkcount;
        }
        fclose(f);
        avg = logsum/maxidx;
        var = logsum2/maxidx - avg * avg;
        printf("Average : %.2f\n", avg);
        printf("Std dev : %.2f\n", sqrt(var));
        free( kmer );
        free( kmers_arr );
        return 0;
}

// -------------------------------------------------
int ACTG2int( char *actg, int len ) {
        int POWERS4[] = {1,4,16,64,256,1024,4096,16384,65536,262144,1048576,4194304,16777216};
        int ret = 0, i = len - 1;
        char *ptr = actg;
        while( *ptr ) {
                // A=0, C=1, T=2, G=3
                switch( *ptr ) {
                        case 'A': break;
                        case 'N': break;
                        case 'C': ret += POWERS4[i]; break;
                        case 'T': ret += 2 * POWERS4[i]; break;
                        case 'G': ret += 3 * POWERS4[i]; break;
                        default: return -1;
                }
                i--; ptr++;
        }
        return ret;
}

int getNextSeqFromFastaFile( struct Seq *seq, FILE *f ) {
        char *ptr, buf[MAXLINELEN];
        int namei, seqi;
        long fpos;

        namei = seqi = 0;
```

```
        seq->name[0] = seq->seq[0] = 0;
        fpos = ftell( f );
        while( fgets( buf, MAXLINELEN, f ) != NULL ) {
                ptr = buf;
                while( isspace(*ptr) ) ptr++;
                if( *ptr == '>' ) {
                        if( namei > 0 && seqi > 0 ) {
                                fseek( f, fpos, SEEK_SET );
                                return( 1 );
                        }
                        ptr++; /* ignore > */
                        for( ; *ptr && namei < MAXSEQNAMELEN ; ptr++ )
                                if( isprint( *ptr ) )
                                        seq->name[namei++] = *ptr;
                } else {
                        for( ; *ptr && seqi < MAXSEQLEN ; ptr++ )
                                if( isalnum( *ptr ) )
                                        seq->seq[seqi++] = *ptr;
                }
                seq->name[namei] = 0;
                seq->seq[seqi] = 0;
                fpos = ftell( f );
        }
        if( namei > 0 && seqi > 0 )
                return( 1 );
        return( 0 );
}
```

FIGURES

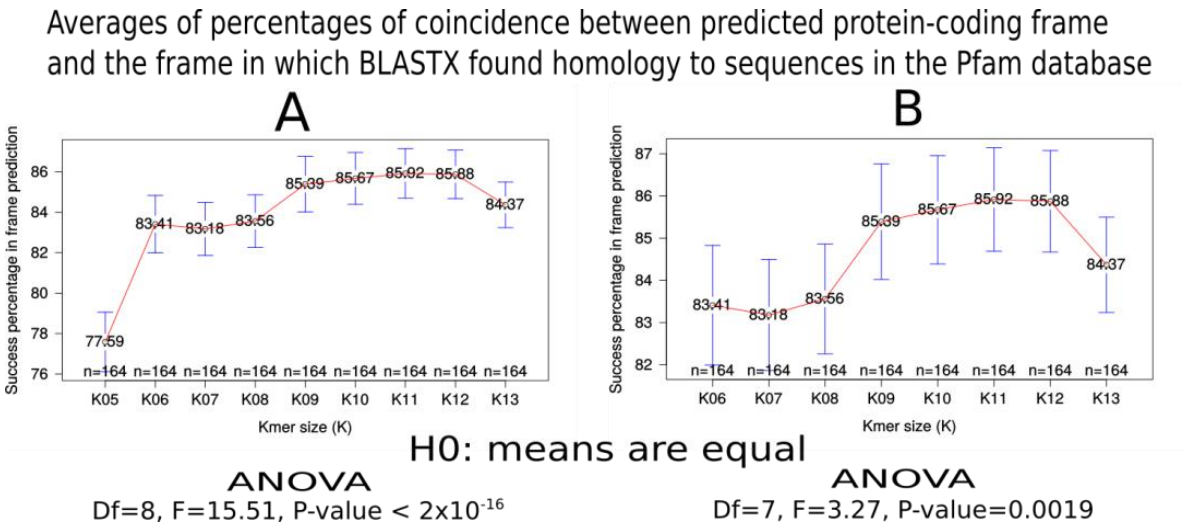


Figure. S1. Averages of percentages of coincidence between predicted protein-coding frame and the frame in which BLASTX found homology to sequences in the Pfam database. The protein-coding potential was estimated for each of the six frames of every sequence from the metagenomes listed in Table S1. The plots show the percentage of coincidence of the frame prediction (highest potential among the six frames) with the best BLASTX hit, using Pfam as the reference protein database and a bitscore cut-off of 50. The best results were obtained when using K=11, i.e. undecamers, as features in the protein-coding sequences to estimate a protein-coding potential. The database used for estimating the Kmers (as described in Fig. 1) log frequencies was the set of protein-coding genes from sequenced bacterial genomes in the NCBI site. Plot A considers Kmers from K=5 to 13. The ANOVA test for the differences of means indicates that there is strong evidence (P-value < $2x10^{-16}$) to reject the null hypothesis. As K=5 is an outlier, we repeated the test excluding K=5, resulting in the analysis shown in part B of the figure. In this case the evidence is less strong (P-value=0.0019) but we can equally reject the null hypothesis at a confidence of 95%.

TABLES

K	Maximum number of Kmers	Average Log count	Standard deviation	Coefficient of variation
5	1,024	14.25	1.73	0.12
6	4,096	12.50	2.33	0.19
7	16,384	11.05	2.38	0.22
8	65,536	9.59	2.42	0.25
9	262,144	7.87	2.74	0.35
10	1,048,576	6.51	2.53	0.39
11	4,194,304	5.18	2.23	0.43
12	16,777,216	3.64	2.18	0.60
13	67,108,864	1.96	1.89	0.96

Table S1. Statistics of in-frame kmers (k=5..13) of the genes from the CDS database of sequenced genomes used in this study (Material and Methods).

Table S2

Number of metagenomes	Biome	Publication or electronic resource
1	Acidic cave	Metagenomic evidence for sulfide oxidation in extremely acidic cave biofilms (1)
1	Canine gut	MG-RAST id 4444703.3
1	Chicken cecum	MG-RAST id 4440283.3
1	Cow rumen	MG-RAST id 4441679.3
1	Termite gut	MG-RAST id 4442701.3
18	Human gut	A core gut microbiome in obese and lean twins (2)
1	Acid salt lake	Insights from the metagenome of an acid salt lake: the role of biology in an extreme depositional environment (3)
8	River	Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011 (4)
1	Lake	Metagenomic Insights into the evolution, function, and complexity of the planktonic microbial community of lake Lanier, a temperate freshwater ecosystem (5)
8	Hot spring	Comparative metagenomics of eight geographically remote terrestrial hot springs (6)
16	Seawater	CAMERA project CAM_P_000692
2	Seawater	Metagenomic analysis of nitrogen and methane cycling in the Arabian sea Oxygen Minimum Zone (OMZ) (7)
8	Seawater	Microbial metatranscriptomics in a permanent marine oxygen minimum zone (8)
4	Subseafloor	Metagenomic signatures of the Peru margin subseafloor biosphere show a genetically distinct environment (9)

1	Subseafloor	Metagenomics of the subsurface Brazos-Trinity basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes (10)
2	Marine sediment	Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara (11)
7	Marine sediment	Metagenomic and geochemical characterization of pockmarked sediments overlaying the Troll petroleum reservoir in the North Sea (12)
2	Marine sediment	A metagenomic study of methanotrophic microorganisms in coal oil point seep sediments (13)
6	Marine cold seep	Synchronized dynamics of bacterial niche-specific functions during biofilm development in a cold seep brine pool (14)
1	Marine cold seep	Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments (15)
1	Deep ocean	Metagenomic analysis of hadopelagic microbial assemblages thriving at the deepest part of Mediterranean sea, Matapan-Vavilov Deep (16)
1	Deep ocean	Going deeper: metagenome of a hadopelagic microbial community (17)
12	Hydrothermal vent	Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau spreading center (18)
2	Hydrothermal vent	Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system (19)
4	Mangrove sediment	The microbiome of brazilian mangrove sediments as revealed by metagenomics (20)
2	Mangrove sediment	Rhizosphere microbiome metagenomics of gray mangroves (avicennia marina) in the red sea (21)
4	Mangrove rhizosphere	Rhizosphere microbiome metagenomics of gray mangroves (avicennia marina) in the red sea (21)
6	Polar desert	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (22)
3	Hot desert	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (22)
3	Tropical forest	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (22)
1	Boreal forest	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (22)
1	Temperate deciduous forest	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (22)
1	Temperate coniferous forest	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (22)
1	Temperate grassland	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (22)
1	Arctic tundra	Cross-biome metagenomic analyses of soil microbial communities and their functional attributes (22)
1	Temperate forest	Community transcriptomics reveals universal patterns of protein sequence conservation in natural microbial communities (23)

1	Tropical forest	Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis (24)
1	Tropical forest	NCBI SRA id SRP001743
3	Grassland	Structure, fluctuation and magnitude of a natural grassland soil metagenome (25)
1	Rhizosphere	Structure, fluctuation and magnitude of a natural grassland soil metagenome (25)
6	Rhizosphere	Structure and function of the bacterial root microbiota in wild and domesticated barley (26)
6	Rhizosphere	Functional congruence of rhizosphere microbial communities associated to leguminous tree from Brazilian semiarid region (27)
1	Solar saltern	Metagenome sequencing of the microbial community of a solar saltern crystallizer pond at Cahuil lagoon, Chile (28)
7	Salt desert	A snapshot of microbial communities from the Kutch, one of the largest salt deserts in the worlds (29)
4	Solar saltern	New abundant microbial groups in aquatic hypersaline environments (30)

Table S2. Metagenomes used in this study for evaluating the protein-coding potential for the different sets of kmers.

REFERENCES

1. Jones,D.S., Schaperdoth,I. and Macalady,J.L. (2014) Metagenomic evidence for sulfide oxidation in extremely acidic cave biofilms. *Geomicrobiol. J.*, **31**, 194–204.

2. Turnbaugh,P.J., Hamady,M., Yatsunenko,T., Cantarel,B.L., Duncan,A., Ley,R.E., Sogin,M.L., Jones,W.J., Roe,B.A., Affourtit,J.P., *et al.* (2009) A core gut microbiome in obese and lean twins. *Nature*, **457**, 480–484.

3. Johnson,S.S., Chevrette,M.G., Ehlmann,B.L. and Benison,K.C. (2015) Insights from the metagenome of an acid salt lake: The role of Biology in an extreme depositional environment. *PLoS One*, **10**, e0122869.

4. Satinsky,B.M., Fortunato,C.S., Doherty,M., Smith,C.B., Sharma,S., Ward,N.D., Krusche,A. V, Yager,P.L., Richey,J.E., Moran,M.A., *et al.* (2015) Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. *Microbiome*, **3**, 39.

5. Oh,S., Caro-Quintero,A., Tsementzi,D., DeLeon-Rodriguez,N., Luo,C., Poretsky,R. and Konstantinidis,K.T. (2011) Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl. Environ. Microbiol.*, **77**, 6000–6011.

6. Menzel,P., Gudbergsdóttir,S.R., Rike,A.G., Lin,L., Zhang,Q., Contursi,P., Moracci,M., Kristjansson,J.K., Bolduc,B., Gavrilov,S., *et al.* (2015) Comparative Metagenomics of Eight Geographically Remote Terrestrial Hot Springs. *Microb. Ecol.*, **70**, 411–424.

7. Lüke,C., Speth,D.R., Kox,M.A.R., Villanueva,L. and Jetten,M.S.M. (2016) Metagenomic analysis of nitrogen and methane cycling in the Arabian Sea oxygen minimum zone. *PeerJ*, **4**, e1924.

8. Stewart,F.J., Ulloa,O. and Delong,E.F. (2012) Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ. Microbiol.*, **14**, 23–40.

9. Biddle,J.F., Fitz-Gibbon,S., Schuster,S.C., Brenchley,J.E. and House,C.H. (2008) Metagenomic signatures of the Peru margin subseafloor biosphere show a genetically distinct environment. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 10583–10588.

10. Biddle,J.F., White,J.R., Teske,A.P. and House,C.H. (2011) Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes. *ISME J.*, **5**, 1038–1047.

11. Quaiser,A., Zivanovic,Y., Moreira,D. and López-García,P. (2011) Comparative metagenomics of bathypelagic plankton and bottom sediment from the Sea of Marmara. *ISME J.*, **5**, 285–304.

12. Håvelsrud,O., Haverkamp,T.H., Kristensen,T., Jakobsen,K.S. and Rike,A. (2012) Metagenomic and geochemical characterization of pockmarked sediments overlaying the Troll petroleum reservoir in the North Sea. *BMC Microbiol.*, **12**, 203.

13. Håvelsrud,O., Haverkamp,T.H., Kristensen,T., Jakobsen,K.S. and Rike,A. (2011) A metagenomic study of methanotrophic microorganisms in coal oil point seep sediments. *BMC Microbiol.*, **11**, 221.

14. Zhang,W., Wang,Y., Bougouffa,S., Tian,R., Cao,H., Li,Y., Cai,L., Wong,Y.H., Zhang,G., Zhou,G., *et al.* (2015) Synchronized dynamics of bacterial niche-specific functions during biofilm development in a cold seep brine pool. *Environ. Microbiol.*, **17**, 4089–4104.

15. Stokke,R., Roalkvam,I., Lanzen,A., Haflidason,H. and Steen,I.H. (2012) Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. *Environ. Microbiol.*, **14**, 1333–1346.

16. Smedile,F., Messina,E., La Cono,V., Tsoy,O., Monticelli,L.S., Borghini,M., Giuliano,L., Golyshin,P.N., Mushegian,A. and Yakimov,M.M. (2013) Metagenomic analysis of hadopelagic microbial assemblages thriving at the deepest part of Mediterranean Sea, Matapan-Vavilov Deep. *Environ. Microbiol.*, **15**, 167–182.

17. Eloe,E.A., Fadrosh,D.W., Novotny,M., Zeigler Allen,L., Kim,M., Lombardo,M.-J., Yee-Greenbaum,J., Yooseph,S., Allen,E.E., Lasken,R., *et al.* (2011) Going Deeper: Metagenome of a Hadopelagic Microbial Community. *PLoS One*, **6**, e20388.

18. Anantharaman,K., Breier,J.A. and Dick,G.J. (2015) Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. *ISME J.*, 10.1038/ismej.2015.81.

19. Tang,K., Liu,K., Jiao,N., Zhang,Y. and Chen,C.T.A. (2013) Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system. *PLoS One*, **8**, 1–11.

20. Andreote,F.D., Jiménez,D.J., Chaves,D., Dias,A.C.F., Luvizotto,D.M., Dini-Andreote,F., Fasanella,C.C., Lopez,M.V., Baena,S., Taketani,R.G., *et al.* (2012) The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS One*, **7**.

21. Alzubaidy,H., Essack,M., Malas,T.B., Bokhari,A., Motwalli,O., Kamanu,F.K., Jamhor,S.A., Mokhtar,N.A., Antunes,A., Simões,M.F., *et al.* (2016) Rhizosphere microbiome metagenomics of gray mangroves (*Avicennia marina*) in the Red Sea. *Gene*, **576**, 626–636.

22. Fierer,N., Leff,J.W., Adams,B.J., Nielsen,U.N., Bates,S.T., Lauber,C.L., Owens,S., Gilbert,J.A., Wall,D.H. and Caporaso,J.G. (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 21390–21395.

23. Stewart,F.J., Sharma,A.K., Bryant,J.A., Eppley,J.M. and DeLong,E.F. (2011) Community transcriptomics reveals universal patterns of protein sequence conservation in natural microbial communities. *Genome Biol.*, **12**, R26.

24. Kanokratana,P., Uengwetwanit,T., Rattanachomsri,U., Bunterngsook,B., Nimchua,T., Tangphatsornruang,S., Plengvidhya,V., Champreda,V. and Eurwilaichitr,L. (2011) Insights into the phylogeny and metabolic potential of a primary tropical peat swamp forest microbial community by metagenomic analysis. *Microb. Ecol.*, **61**, 518–528.

25. Delmont,T.O., Prestat,E., Keegan,K.P., Faubladier,M., Robe,P., Clark,I.M., Pelletier,E., Hirsch,P.R., Meyer,F., Gilbert,J. a, *et al.* (2012) Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J.*, **6**, 1677–1687.

26. Bulgarelli,D., Garrido-Oter,R., Münch,P.C., Weiman,A., Dröge,J., Pan,Y., McHardy,A.C. and Schulze-Lefert,P. (2015) Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe*, **17**, 392–403.

27. Taketani,R.G., Kavamura,V.N., Mendes,R. and Melo,I.S. (2015) Functional congruence of rhizosphere microbial communities associated to leguminous tree from Brazilian semiarid region. *Environ. Microbiol. Rep.*, **7**, 95–101.

28. Plominsky AM, Delherbe N, Ugalde JA, Allen EE, Blanchet M, et al. (2014) Metagenome Sequencing of the Microbial Community of a Solar Saltern Crystallizer Pond at Cáhuil Lagoon, Chile. *Genome Announc.*, **2**, 1–2.

29. Pandit,A.S., Joshi,M.N., Bhargava,P., Shaikh,I., Ayachit,G.N., Raj,S.R., Saxena,A.K. and Bagatharia,S.B. (2015) A snapshot of microbial communities from the Kutch: one of the largest salt deserts in the World. *Extremophiles*, **19**, 973–987.

30. Ghai,R., Pašić,L., Fernández,A.B., Martin-Cuadrado,A.-B., Mizuno,C.M., McMahon,K.D., Papke,R.T., Stepanauskas,R., Rodriguez-Brito,B., Rohwer,F., *et al.* (2011) New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.*, **1**.

31. Volders,P.-J., Verheggen,K., Menschaert,G., Vandepoele,K., Martens,L., Vandesompele,J. and Mestdagh,P. (2014) An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.*, **43**, D174–D180.

32. Consortium,T.U. (2014) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.

33. Buchfink,B., Xie,C. and Huson,D.H. (2014) Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, **12**, 59–60.

34. Finn,R.D., Bateman,A., Clements,J., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Heger,A., Hetherington,K., Holm,L., Mistry,J., *et al.* (2014) Pfam: The protein families database. *Nucleic Acids Res.*, **42**, D222–D230.

CHAPTER 4. CONCLUSION

In this thesis, the topic of the characterization of the microbial communities is addressed. The main result of this work is the recognition that the set of genes encoding oxidoreductases characterize the microbial communities better than other gene categories, including taxonomic gene markers, transporter genes, and other protein-coding genes. In recent years, the trend towards the use of functional genetic traits has grown in microbial ecology as an alternative to the taxonomic approach, which has been proven to be often unable to resolve traits or producing useful ecological patterns (3-5, 39). However, there were no compelling arguments for selecting one particular gene category over others in a global context (9, 35, 36). Noteworthy, the oxidoreductase structure of microbial communities also has the advantage of directly describing the energetic matrix and biogeochemical links of the microbial ecosystems.

This result is, however, limited by the current coverage of known functions within metagenomes. Many yet-to-discover functional genes, including genes encoding unknown oxidoreductases, might be present within the unknown functional dark matter of metagenomes. As a first step towards the elucidation of this known-unknown, a method for computing a protein-coding potential of nucleotide sequences is proposed. The emphasis of this development is to define a method as general as possible, allowing future refinements for its application to diverse types of analyses, such as protein-coding gene annotations, analysis of transcriptomic sequences and massive nucleotide sequence alignment to protein-coding sequences databases. The latter application can be accomplished by reducing the number of analyzed open reading frames for aligning (ORFs, which are six in total) to a single ORF with the highest protein-coding potential. Another possible application of this method includes the detection of novel sequences belonging to already known functional categories, such as oxidoreductase genes.

The relevance of these results relay on the recognized need for better predictors of the microbial processes on the different ecosystems currently threatened both by the unbalanced conservation and conversion of biomes and also by the global warming. On this matter, the taxonomic approach has been proved to be often unable to predict ecosystem functioning, and thus the need for trait-based approaches on microbial ecology has been claimed for years. The mentioned environmental threats demand the ability to predict the effects of a changing environment on the biosphere, and approaches that ignore the environment or focus on a few species at a time cannot address this question (5). In future developments, the results of this thesis should allow a better assessment and evaluation of the impact of the environmental stressors on the ecosystem services from the different environments of our planet. This improved diagnostic of ecosystems should be possible by focusing directly on the diversity of the redox functions encoded in the metagenomes of microbial communities, rather than on

their taxonomic structures. Thus, this approach should help in developing better management and conservation policies that effectively include not only iconic species or colonies, such as polar bears or coral reefs but also microorganisms, which underpin the nutrient recycling and biogeochemical cycles on this planet.

REFERENCES

1. Maynard-Smith J (1974) Models in Ecology (Cambridge Univ Press, Cambridge, UK).
2. Godfrey-Smith P (2013) Philosophy of biology (Princeton University Press, US).
3. Lawton JH (1999) Are there general laws in ecology? *Oikos*. doi:10.2307/3546712.
4. Roughgarden J (2009) Is there a general theory of community ecology? *Biol Philos*. doi:10.1007/s10539-009-9164-z.
5. McGill BJ, Enquist BJ, Weiher E, Westoby M (2006) Rebuilding community ecology from functional traits. *Trends Ecol Evol* 21:178–185.
6. Cohan FM (2002) What are bacterial species? *Annu Rev Microbiol* 56:457–487.
7. Doolittle WF, Papke RT (2006) Genomics and the bacterial species problem. *Genome Biol* 7:116.
8. Doolittle WF, Zhaxybayeva O (2009) On the origin of prokaryotic species. *Genome Res* 19:744–756.
9. Boon E, et al. (2014) Interactions in the microbiome: communities of organisms and communities of genes. *FEMS Microbiol Rev* 38:90–118.
10. Amann RI, Ludwig W, Schleifer K (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143–169.
11. Hug LA, et al. (2016) A new view of the tree of life. *Nat Microbiol*. doi:10.1038/nmicrobiol.2016.48.
12. Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17:589–596.
13. Giovannoni SJ, et al. (2005) Genetics: Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309:1242–1245.
14. Darmon E, Leach DRF (2014) Bacterial genome instability. *Microbiol Mol Biol Rev* 78:1–39.
15. Polz MF, Alm EJ, Hanage WP (2013) Horizontal gene transfer and the evolution of

bacterial and archaeal population structure. *Trends Genet* 29:170–175.

16. 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* 87:4576–4579.

17. Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol*. doi:0021-9193/98/$04.00+0.

18. Tyson GW, et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. doi:10.1038/nature02340.

19. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5:R245–R249.

20. Fuhrman JA, et al. (2006) Annually reoccurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci U S A* 103:13104–13109.

21. Faust K, Raes J (2012) Microbial interactions: from networks to models. *Nat Rev Microbiol* 10:538–550.

22. Fierer N, Bradford MA, Jackson RB (2007) Toward an ecological classification of soil bacteria. *Ecology* 88:1354–1364.

23. Philippot L, et al. (2009) Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. *Environ Microbiol* 11:3096–3104.

24. Lennon JT, Aanderud ZT, Lehmkuhl BK, Schoolmaster DR (2012) Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology*. doi:10.1890/11-1745.1.

25. Philippot L, et al. (2010) The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol* 8:523–529.

26. 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.

27. Nemergut DR, et al. (2016) Decreases in average bacterial community rRNA operon copy number during succession. *ISME J* 10:1147–1156.

28. Louca S, Parfrey LW, Doebeli M (2016) Decoupling function and taxonomy in the global ocean microbiome. *Science* 353:1272–1277.

29. Bletz MC, et al. (2016) Amphibian gut microbiota shifts differentially in community structure but converges on habitat-specific predicted functions. *Nat Commun* 7:13699.

30. Dinsdale E, et al. (2008) Functional metagenomic profiling of nine biomes. *Nature* 452:629–632.

31. Ma X, et al. (2019) Microbial functional traits are sensitive indicators of mild disturbance by lamb grazing. *ISME J*. doi:10.1038/s41396-019-0354-7.

32. Doolittle WF, Booth A (2017) It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biol Philos* 32:5–24.

33. Handelsman J (2005) Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev*. doi:10.1128/MMBR.69.1.195.2005.

34. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N (2017) Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35:833–844.

35. Green JL, Bohannan BJM, Whitaker RJ (2008) Microbial biogeography: from taxonomy to traits. *Science* 320:1039–1043.

36. Nemergut DR, et al. (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* 77:342–356.

37. Heintz-Buschart A, Wilmes P (2018) Human Gut Microbiome: Function Matters. *Trends Microbiol* 26:563–574.

38. Sunagawa S, et al. (2015) Structure and function of the global ocean microbiome. *Science* 348:1261359–1261359.

39. Krause S, et al. (2014) Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front Microbiol* 5:1–10.

PUBLISHED PAPERS WITH AFFILIATION TO THIS PROGRAM

Although only the first paper listed below is associated with this thesis, here is the complete list of papers published with affiliation to this postgraduate program.

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)

Ramírez-Flandes, S., González, B., & Ulloa, O. (2019). Redox traits characterize the organization of global microbial communities. *Proceedings of the National Academy of Sciences*, *116*(9), 3630-3635.

2015. Block invariance in elementary cellular automata

Goles, E., Montalva-Medel, M., Mortveit, H., & Ramirez-Flandes, S. (2015). Block invariance in elementary cellular automata. *Journal of Cellular Automata*, *10*(1-2), 119-135.

frontiers in MARINE SCIENCE

ORIGINAL RESEARCH ARTICLE
published: 07 July 2014
doi: 10.3389/fmars.2014.00018

Enhanced metabolic versatility of planktonic sulfur-oxidizing γ-proteobacteria in an oxygen-deficient coastal ecosystem

Alejandro A. Murillo[1], Salvador Ramírez-Flandes[1,2], Edward F. DeLong[3] and Osvaldo Ulloa[1]*

[1] *Laboratorio de Oceanografía Microbiana, Departmento de Oceanografía & Instituto Milenio de Oceanografía, Universidad de Concepción, Concepción, Chile*
[2] *Programa de Doctorado en Ingeniería de Sistemas Complejos, Universidad Adolfo Ibáñez, Santiago, Chile*
[3] *Department of Biological Engineering and Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA*

Murillo, A. A., Ramírez-Flandes, S., DeLong, E. F., & Ulloa, O. (2014). Enhanced metabolic versatility of planktonic sulfur-oxidizing γ-proteobacteria in an oxygen-deficient coastal ecosystem. *Frontiers in Marine Science*, *1*, 18.

The ISME Journal (2015) 9, 1264–1267

www.nature.com/ismej

Genomic potential for nitrogen assimilation in uncultivated members of *Prochlorococcus* from an anoxic marine zone

Marcia Astorga-Eló[1,2], Salvador Ramírez-Flandes[1,3], Edward F DeLong[4,5,6] and Osvaldo Ulloa[1]

[1]*Departamento de Oceanografía and Instituto Milenio de Oceanografía, Universidad de Concepción, Concepción, Chile;* [2]*Programa de Magíster en Bioquímica y Bioinformática, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción, Chile;* [3]*Programa de Doctorado en Ingeniería de Sistemas Complejos, Universidad Adolfo Ibáñez, Santiago, Chile;* [4]*Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA;* [5]*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA and* [6]*Daniel K. Inouye Center for Microbial Oceanography: Research and Education, University of Hawaii, Honolulu, HI, USA*

Astorga-Eló, M., Ramírez-Flandes, S., DeLong, E. F., & Ulloa, O. (2015). Genomic potential for nitrogen assimilation in uncultivated members of Prochlorococcus from an anoxic marine zone. *The ISME journal*, *9*(5), 1264-1267.

Distinctive Archaeal Composition of an Artisanal Crystallizer Pond and Functional Insights Into Salt-Saturated Hypersaline Environment Adaptation

Alvaro M. Plominsky[1,2†‡], Carlos Henríquez-Castillo[1,2†], Nathalie Delherbe[3], Sheila Podell[4], Salvador Ramirez-Flandes[2,5], Juan A. Ugalde[6,7], Juan F. Santibañez[1], Ger van den Engh[8], Kurt Hanselmann[9], Osvaldo Ulloa[1,2], Rodrigo De la Iglesia[10], Eric E. Allen[4,11] and Nicole Trefault[12]**

[1] Department of Oceanography, Faculty of Natural and Oceanographic Sciences, Universidad de Concepción, Concepción, Chile, [2] Instituto Milenio de Oceanografía, Concepción, Chile, [3] Biology Department, Cell and Molecular Biology Joint Doctoral Program with UC San Diego, San Diego State University, San Diego, CA, United States, [4] Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, United States, [5] Programa de Doctorado en Ingeniería de Sistemas Complejos, Universidad Adolfo Ibáñez, Santiago, Chile, [6] uBiome, Inc., San Francisco, CA, United States, [7] Center for Bioinformatics and Integrative Biology, Facultad de Ciencias Biológicas, Universidad Andrés Bello, Santiago, Chile, [8] Center for Marine Cytometry, Concrete, WA, United States, [9] Department of Earth Sciences, ETH Zürich, Zurich, Switzerland, [10] Department of Molecular Genetics and Microbiology, Faculty of Biological Sciences, Pontificia Universidad Católica de Chile, Santiago, Chile, [11] Division of Biological Sciences, University of California, San Diego, La Jolla, CA, United States, [12] GEMA Center for Genomics, Ecology and Environment, Faculty of Sciences, Universidad Mayor, Santiago, Chile*

Plominsky, A. M., Henríquez-Castillo, C., Delherbe, N., Podell, S., Ramirez-Flandes, S., Ugalde, J. A., ... & De la Iglesia, R. (2018). Distinctive Archaeal Composition of an Artisanal Crystallizer Pond and Functional Insights into Salt-Saturated Hypersaline Environment Adaptation. *Frontiers in microbiology*, *9*, 1800

Ostreococcus tauri Luminescent Reporter Lines as Biosensors for Detecting Pollution From Copper-Mine Tailing Effluents in Coastal Environments

Carlos Henríquez-Castillo [1,2,3], Hugo Botebol [4], Adelaide Mouton [4], Salvador Ramírez-Flandes [2,3,5], Jean-Claude Lozano [4], Gaelle Lelandais [6], Santiago Andrade [7,8], Nicole Trefault [8], Rodrigo De la Iglesia [1] and François-Yves Bouget [4]**

[1] Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile, [2] Departamento de Oceanografía, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción, Concepción, Chile, [3] Instituto Milenio de Oceanografía, Concepción, Chile, [4] Sorbonne Universités, UPMC Univ Paris 06 & Centre National pour la Recherche Scientifique, UMR 7621, Laboratoire d'Océanographie Microbienne, Observatoire Océanologique, Banyuls-sur-Mer, France, [5] Programa de Doctorado en Ingeniería de Sistemas Complejos, Universidad Adolfo Ibáñez, Santiago, Chile, [6] Centre National de la Recherche Scientifique, Institut Jacques Monod, Université Paris Diderot (Paris 07), Paris, France, [7] Departamento de Ecología, Center of Applied Ecology and Sustainability, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile, [8] Centro GEMA- Genómica, Ecología y Medio Ambiente, Facultad de Ciencias, Universidad Mayor, Santiago, Chile

Henríquez-Castillo, C., Botebol, H., Mouton, A., Ramírez-Flandes, S., Lozano, J. C., Lelandais, G., ... & Bouget, F. Y. (2018). Ostreococcus tauri luminescent reporter lines as biosensors for detecting pollution from copper-mine tailing effluents in coastal environments. *Frontiers in Environmental Science*, *6*.