University of Florence

International Doctorate in Structural Biology

Cycle XXIII (2008-2010)

The investigation of metalloproteomes through bioinformatics in a comparative and evolutionary perspective

Ph.D. thesis of

Leonardo Decaria

Tutor	Coordinator
Prof. Antonio Rosato	Prof. Ivano Bertini

S.S.D. CHIM/03

This thesis has been approved by the University of Florence,

the University of Frankfurt and the Utrecht University

Pages

1. Introduction

Life on Earth developed in equilibrium with the hydrosphere and the lithosphere, taking from these all the elements necessary for performing essential functions (1). As a consequence, a number of metal ions have been selected during evolution to take part in many crucial biological processes and are thus essential for living organisms (2;3). In particular, many proteins require metal ions or metal-containing cofactors to carry out their physiological function (4;5). The interaction with metal ions must be controlled in any moment of life: from the uptake to the trafficking within the living organisms for their utilization, until the excretion when appropriate. Systemic and cellular homeostatic control of transition metal ions is necessarily tight so as to provide essential amounts of metal ions while preventing toxicity in the context of available binding sites in metalloproteins. Many different steps lie between metal selection of an organism and metal selection of a metalloprotein. These “selectivity filters” determine which metal ions enter the cell and reach its subcellular compartments. Eukarya have more selectivity filters than prokarya, as their cells contain more subcellular compartments. Coordination chemistry of ligands that handle the metal ions provides specificity during transit in the cytoplasm and transport through membranes. The metalloprotein provides only the last step in selecting the metal. How the correct metal ion is incorporated into a protein turns out to be a question of biology, in addition to the coordination chemistry of the metalloprotein. Usually metalloproteins are identified through biochemical studies that probe the dependence of the function of the proteins of interest on the presence of metal ions. This is typically done *in vitro* on purified native or recombinant samples. At present, the complexity of these processes and the resource demands associated with the needed experimental work make it unfeasible to perform a complete identification of metalloproteins at the level of entire metalloproteomes. Metalloproteomics includes

approaches that address the expression of metalloproteins and their changes in biological time and space. Experimental approaches to investigate metalloproteomes include structural genomics, which provides insights into the architecture of metal-binding sites in metalloproteins and establishes ligand signatures from the types and spacings of the metal ligands in the protein sequence. Theoretical approaches employ these ligand signatures as templates for homology searches in sequence databases. In this way, the number of metalloproteins in the iron, copper, and zinc metalloproteomes in various phyla of life can be estimated. So, bioinformatics methods can give valuable support to experimental ones and are especially important to obtain insights into metalloproteomes, metal by metal. Systems biology approaches require the combination of large-scale studies to catalogue genome-wide data sets to obtain as detailed as possible knowledge on the molecules and their interactions. The investigation of metalloproteomes in this framework, therefore, implies the definition of all the metalloproteins encoded by an organism in conjunction with their functional characterization. This information is essential for a comprehensive understanding of the whole of the processes occurring in living systems. In the present Ph.D. thesis, we focused out attention on the highly characterized and biologically relevant metals zinc, copper and iron-heme.

Zinc is essential for life and is the second most abundant transition metal ion in living organisms after iron (6). In contrast to other transition metal ions, such as copper and iron, zinc(II) does not undergo redox reactions due to its filled *d* shell. Much is known about its coordination in structural biology, it can perform both structural or catalytic function in proteins involved in fundamental cellular processes (7;8). For example, carbonic anhvdrase is a Zn-binding enzyme with a cardinal role in the acid-base homeostasis of living organisms by catalyzing the reversible dehydratation of carbonic acid, a process critical to the transport and elimination of carbon dioxide (9). Zinc may

also modulate signaling events, as it occurs in process(es) maintaining zinc homeostasis, e.g., through zinc-regulated protein expression (10)**.** Most transcriptional factors are Zn-binding proteins, as they contain small protein structural motifs that can coordinate one or more ions to help stabilize their folds, called zinc fingers. These motifs, or domains, can be classified into several different structural families and typically function as interaction modules that bind DNA, RNA, proteins, or small molecules. Zinc fingers coordinate ions with a combination of cysteine and histidine residues, and can be classified by the type and order of these zinc coordinating residues (e.g., Cys_2His_2, Cys_4 and others). A more systematic method classifies them into different "fold groups" based on the overall shape of the protein backbone in the folded domain.

Copper is an essential trace metal utilized as a cofactor in a variety of proteins (11). In eukaryotes, copper-dependent metalloenzymes are found in multiple cellular locations (12;13). At variance with other metal ions, the proteins involved in copper trafficking are about 50% of the entire copper proteome, i.e., about half of all cellular copper proteins and they traffic copper as copper(I) ions. In eukaryotes and in a few bacterial systems such as cyanobacteria, copper(I) ions pass through the membranes of the cells using permeases or ATPases and enter the cytoplasm. The metal is necessarily imported in the cells as copper(I) since the copper(II) ion would be reduced in the cytoplasm to copper(I) without the necessary control. Excess copper, however, is highly toxic to most organisms (14;15). Accordingly, a complex machinery of proteins that bind the metal ion controls the uptake, transport, sequestration, and efflux of copper in vivo (16;17). Indeed, the intracellular concentration of free copper ions should be maintained at an essentially negligible level, as the copper ions may catalyze the formation of radicals which can damage cell membranes. On the other hand, newly produced Cu-binding proteins need to uptake copper ions to achieve their mature, active form. This dual goal can be obtained if

systems permitting rapid and efficient metal transfer and simultaneously preventing nonspecific reactions involving copper are in place. In particular, so-called metallochaperones, which deliver copper to specific intracellular targets, lower the activation barrier for copper transfer to their specific partners, thereby circumventing the significant thermodynamic overcapacity for copper chelation of the cytoplasm.

Iron is the most abundant metal in living organisms, in particular heme is the prosthetic group of many proteins that carry out a variety of key biological functions, including oxygen transport and sensing, enzyme catalysis and electron transfer. Iron binding proteins essential for life are, for example, iron-sulphur proteins. These proteins are characterized by the presence of iron-sulfur clusters containing sulfide-linked di-, tri-, and tetra-iron centers in variable oxidation states. Iron-sulfur clusters are found in a variety of metalloproteins, such as the ferredoxins, as well as NADH dehydrogenase, hydrogenases, Coenzyme Q - cytochrome c reductase and nitrogenase. Iron-sulfur clusters are best known for their role in the oxidation-reduction reactions of mitochondrial electron transport. Both Complex I and Complex II of oxidative phosphorylation have multiple Fe-S clusters. With the exception of a few species such as *Borrelia burgdorferi* (18), the growth of microbial pathogens within the host requires iron as an essential nutrient (19). Given the scarcity of free iron in tissues and cells, which is exacerbated by further limitation imposed by the host in response to infection (20), pathogenic bacteria evolved a number of iron acquisition strategies, including the secretion of siderophores (21) and the uptake of heme from host heme proteins (22;23). Studies on *Staphylococcus aureus* indicated that heme is a preferential iron source for human pathogens, consistent with the localization of most body iron in hemoglobin and myoglobin (24;25). Therefore, bacterial systems for heme uptake are considered as important virulence factors, and represent promising targets for novel therapeutic approaches.

In the past few years, CERM developed methods for the prediction of metal-binding sites on the basis of the protein sequence only (26). Exploiting these, we are able to identify metalloproteins by searching for known metal-binding functional domains and known metal binding patterns (MBPs) in their sequences. In the reported projects, developed during the three years of Ph.D., we performed metal-specific (zinc, copper and iron-heme) calculations aimed at analyzing and characterizing the complete set of metalloproteins (the metalloproteome) encoded by the fully sequenced organisms available on the NCBI website. Once defined the metalloproteomes of the selected set of organisms, we subjected them to phylogenetic, statistical and biochemical analyses.

Finally, we produced a bioinformatic tool, called RDGB and freely available from the CERM website. It addresses the matter of the identification of metalloproteins in databases of gene sequences, exploiting the proposed research method. In a simple-to-use manner (and eventually helped by an exhaustive tutorial), the user can perform a customized research running the given scripts. The tool is self-updating, as can automatically download the requested databases, Pfam and PDB, from the respective websites. To prove its functioning, we characterized two metabolic pathways involving metalloproteins in more than 1100 prokarya.

As a long-standing goal of the biological sciences, the understanding of life at the systems level is experiencing a rekindling of interest. Metallomics and metalloproteomics are emerging fields addressing the role, uptake, transport and storage of trace metals essential for protein functions. On these basis an exhaustive information on metalloproteomes is crucial to correctly represent the role of metal ions in living organisms. Bioinformatics can usefully support experimental procedures, providing high-throughput screenings able to analyze great amounts of data in a relatively restrained time.

2. Methods

For predicting metalloproteomes we developed a functional domain-based search method (26), whose flowchart is reported in Figure 1. The method is mainly based on the Pfam library (27;28), an on-line database of profiles or hidden Markov models (HMMs).

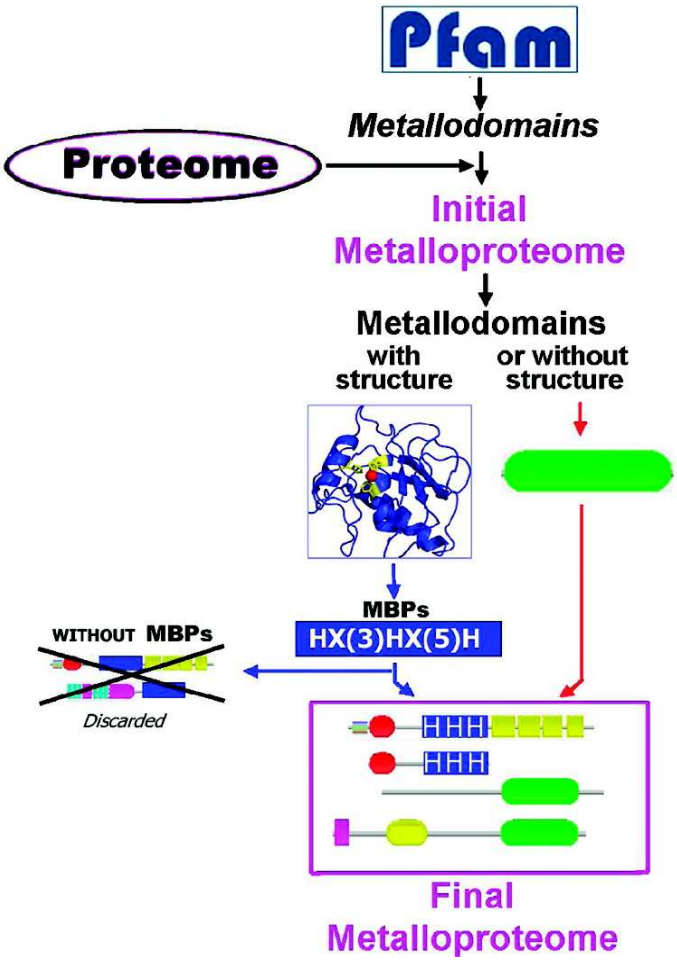


Figure 1 – The research method for metalloprotein prediction developed at CERM.

Each Pfam HMM represents a protein family or domain, and it is calculated from a multiple sequence alignment (sequence profile) by assigning to each residue in each position a value proportional to its position-specific occurrence. In this way, the

functionally relevant residues are highlighted, as they are typically conserved in the sequence alignment. The profiles in Pfam that correspond to metal-binding domains can be selected by querying the library for the domains whose annotation contains the metal name or symbol and then checking the primary literature to discard domains erroneously retrieved. In addition, the Pfam domains are identified also in the sequence of proteins of known 3D structure that are available from the PDB (29). In our experience this is quite useful when trying to collect ensembles of proteins that bind a given ligand (which can be an organic cofactor, a metal ion or a metal-containing cofactor), as sometimes not all the domains that can bind the ligand have been annotated as such in Pfam. Instead, if the ligand is present in the 3D structure of the protein, this information can be readily extracted from the PDB database together with the pattern of amino acids that are involved in the protein-ligand interaction. The latter is called the Metal Binding Pattern (MBP) and is defined by the identity and spacing of the amino acids, e.g., $CX_4CX_{20}H$, where X is any amino acid. The MBP can be usefully applied as a filter to reduce the number of false positives (i.e. of the proteins predicted to bind the cofactor but which in reality are unable to bind it) by rejecting the proteins that lack it. After collecting a list of metal-binding domains, we look for their occurrence in the protein sequences of the organisms we want to investigate and we check the occurrence of the associated MBPs (Metal Binding Patterns) in them.

The strategy has been fully implemented in a package, RDGB, which is downloadable at http://www.cerm.unifi.it/home/research/genomebrowsing.html. The use of RDGB allows the user to perform all the operations that are needed to implement the aforementioned strategy with minimal intervention and to gather all results in an ordered manner, with a tabular summary. This minimizes the (bio)informatics needed, thus facilitating non-experts. The RDGB tool can be run on computers having Linux as their operating system.

It is written in python and uses a variety of different scripts and programs, contained in the subfolder *0Tools* that is created upon installation. The tool is divided in two main python executables, *Retrieving_domains.py* and *Genome_browsing.py*, that run consecutively as the first one builds part of the input to the second script (Figure 2). The first part of the procedure is the collection of the Pfam domains of interest, which can be directly input by the user, obtained from the analysis of sequences with known 3D structure or both. In these last two cases, one or more Ligand Binding Pattern (as the tool recognizes any kind of ligand, e.g. metal ions, cofactors or others) can be associated to the retrieved Pfam domains, thanks to PDB structures analysis.

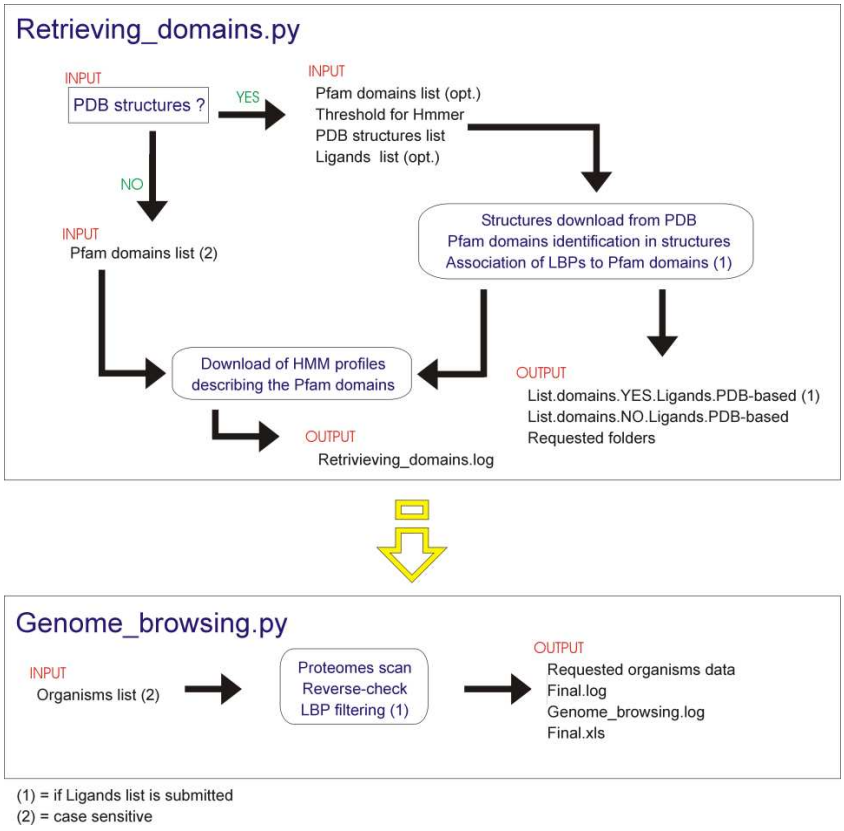


Figure 2 – Flowchart of the RDGB tool processes

The second part of the procedure is the scanning for the occurrence of proteins containing the Pfam domains in the entire proteome sequences of the organisms of interest. The

proteomes analyzed include all chromosomally encoded proteins as well as those encoded by plasmidic DNA. At the end of the whole calculation, in the main working folder the user can find one folder per each organism analyzed, with one subfolder per each Pfam domain analyzed, storing the proteins that contain at least one of the domains of interest. To reduce the rate of false positives, for domains that are associated to a LBP, the sequences are filtered by requesting that they contain the LBP (only the amino acids falling within the domain boundaries). A tolerance of 20% is applied to the spacing between amino acids in the LBP. The user can choose what type of output data and folders have to be produced by the tool in the main working folder, e.g. Pfam output data, retrieved LBPs in the submitted structures, FASTA sequences and Pfam domains composition of the submitted structures and so on.

3. Results & Discussion

The studies described in this thesis improved our understanding of the role of metalloproteins in biological processes. Here we summarize the published or submitted papers, in chronological order. As mentioned before, the research projects developed in the three years of the PhD program focused on the characterization of zinc, copper and iron-heme metalloproteomes. The first project addressed the processes of heme uptake and biosynthesis in 474 prokarya, part of them pathogens. These are fundamental processes in living organism and reliable virulence indicators for pathogens. In the second work, we re-investigated the ensemble of Zn-proteins present in a selected group of 57 organisms taken as representative for archea, bacteria and eukarya in order to extend their functional annotation. We then significantly expanded on these results to elucidate phylogenetic evolution through Zn-proteomes composition, i.e. on the basis of enzyme and transcriptional factors content in 821 organisms. The interesting outcome of this work prompted us to tackle also Cu-proteomes with the same approach and goal, to obtain a phylogenetic examination based on detailed analysis of metalloproteins. The computational strategy for the identification of metalloproteins in proteomes (see Methods and Figure 1) has been fully implemented in a bioinformatic tool, called RDGB (fifth paper), which can be downloaded from the CERM web site. As examples to illustrate the use of the tool functioning, we characterized the aromatic compound degradation process in 1136 prokarya, involving Fe_S and iron binding proteins, as well as the occurrence of proteins involved in the biosynthesis of the Cu_A cofactor.

3.1 Genome-based analysis of heme biosynthesis and uptake in prokaryotic systems

Organisms can meet their heme demands by taking it up from external sources Figures 3A for Gram-positive and 3B for Gram-negative bacteria) or by producing the cofactor through a dedicated biosynthetic pathway (Figure 4), or both. In the present work, we analyzed the distribution of proteins specifically involved in the processes of heme biosynthesis and heme uptake in 474 prokaryotic organisms (225 pathogens), retrieving nearly 12,000 protein domains.

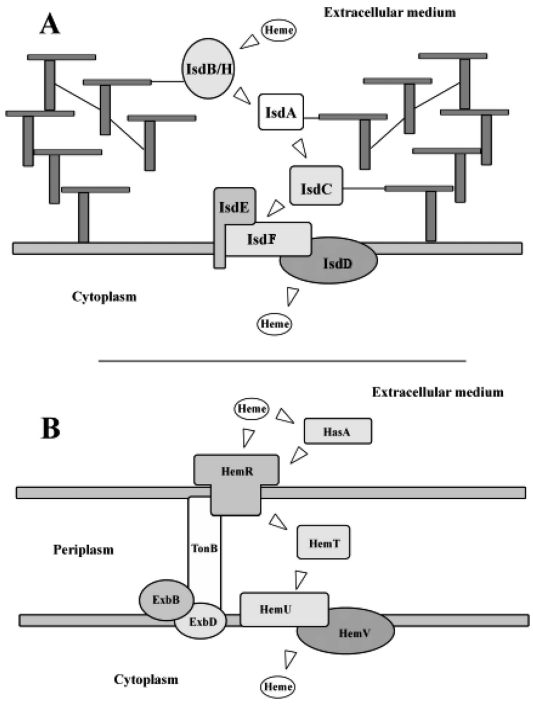


Figure 3 – Heme uptake systems in prokarya.
A) **Gram-positive B) Gram-negative bacteria**

Of these, we predicted that 168 can only synthesize heme (35 pathogens), 20 can only take up heme from an external source (19 pathogens), and 218 can perform both processes (65 pathogens), as judged on the basis of the presence or absence of corresponding protein systems similar to those described in the literature. The large majority of archea (37 in total) is able to synthesize heme, using the precorrin-2

alternative pathway, whereas only a few species lack all heme biosynthesis enzymes. No archeal organism can take up heme.

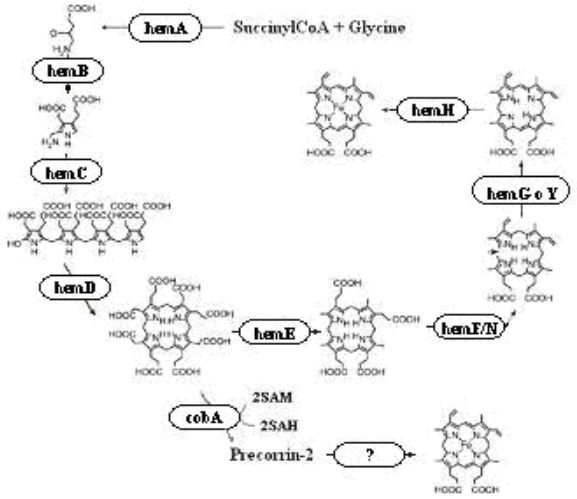


Figure 4 - The heme biosynthesis pathway (conserved in eukaryotes and prokaryotes).

Archea thus seem to be able to acquire iron from the extracellular medium only using siderophores. Bacteria display a substantially higher differentiation than archea. For example, most proteobacteria (Gram-negative) can synthesize as well as take up heme. In firmicutes (Gram-positive) only some species are able to perform heme uptake. Considering virulence, Gram-negative bacteria that can take up heme are equally distributed between pathogenic and non-pathogenic. Instead, for Gram-positive bacteria, heme uptake seems to be more related to pathogenicity.

Among all the analyzed organisms, we singled out some instances of possible variations with respect to the “canonical” pathways that may be worth of experimental investigation. In addition, we used homology modeling to build a series of structural models for two key domains in the heme uptake pathway (data not reported in the publication). The inspection of these models and the analysis of the corresponding sequence alignments

suggested that there are possible alternative modes of heme binding. Again, this is an area where future experimental work would be quite useful.

3.2 The annotation of full zinc proteomes

The zinc proteomes of 57 representative living organisms including members of archea, bacteria, and eukarya were available when we started this work (30). These 57 zinc proteomes were previously predicted to encode cumulatively 18,336 potential zinc binding proteins, which had been grouped into ensembles on the basis of sequence similarity. Functional information, either based on available experimental data or on computational biology methods, was described in the annotation of most of these protein sequences, which was relevant for all proteins in a given ensemble. Of the 18,336 zinc proteins, 1,472 did not have a defined functional annotation as no hit was retrieved from the GO database. For 1090 of them (74%) we obtained an extended functional annotation using a combination of predictive methods. The quality of functional prediction parallels the amount of information for the analyzed sequence; the more detailed and reliable is the proposed functional prediction. We exploited both sequence-based and structure-based bioinformatics tools (when available) to apply a rational functional prediction method. The coverage of functional annotation originally included about 92% of all zinc proteins, which, after our contribution, increased to 98%. Figure 5 shows the total assignment results in a pie-graph. As mentioned above, after our analysis we were able to improve significantly data the functional assignment of our previous work, and new functional annotations could be made for the predicted zinc proteomes of organisms grouped from all the three domains of life (archea, bacteria, eukarya) and of course *Homo sapiens*. In the reference work, and subsequently in this one, the release n° 36 of the human genome

was analyzed, counting about 40000 proteins. The final functional annotation that we propose for the human Zn-binding proteins retrieved, covering 9,2% of the total proteome, shows that 44% of them are involved in transcription, followed by hydrolases at 12%, in accord with the total eukarya behavior. Zn-hydrolases (light blue) cover the highest percentage in prokarya, 44% in Archea and 50% in bacteria (respectively A and B in Figure 5).

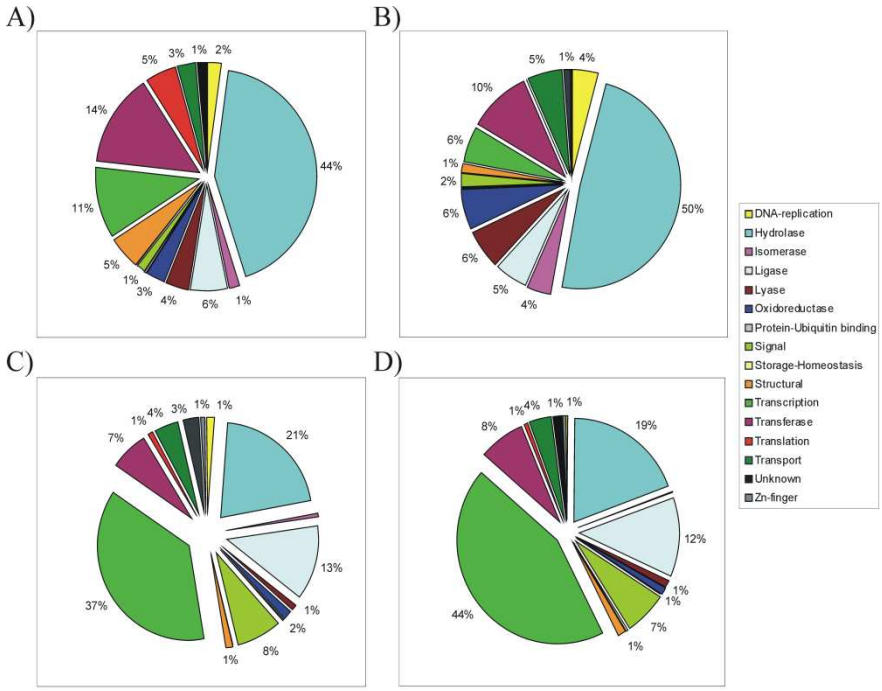


Figure 5 - Functional annotations of the 57 zinc proteomes analyzed: A) Archea, B) Bacteria, C) Eukarya, D) human.

Transcriptional factors (green) are mostly present in eukarya, where they constitute 37%. Of the Zn-proteome These proteins are often characterized by the well-known Zn-finger domain, which is able to bind DNA due to its tridimensional conformation. Gene expression regulation is one of the typical step for protein level control in eukarya. Considering the still unknown sequences (2 % of the total 18336), most of them are hypothetical, putative or predicted eukaryotic proteins. As the most of them had no hits against the Pfam database and prediction of codifying sequences from eukaryotic

genomes is highly hampered by introns, we reasonably defined these sequences as probable false positives.

3.3 Zinc proteomes, phylogenetics and evolution

To put zinc proteomes in an evolutionary perspective, we investigated 821 complete Zn-proteomes, 52 from archea, 723 from bacteria and 46 from eukarya available from the NCBI. In particular we focused our attention on the two major groups of Zn-proteins: hydrolytic enzymes and zinc fingers. There is a remarkable change in the importance of these two sets of proteins which coincides with changes in element availability in the environment (6). In order to give a comparative account of the data and their analysis, we divided bacteria into those with a small proteome of less than 1,500 proteins and those with a larger proteome. The smaller proteomes are mostly of bacteria found in animal hosts while the larger are of bacteria which in general are free. The two groups have considerable differences in their content of zinc proteins. We took the archea and the larger bacteria as representing possible early forms of life or at least life of low complexity. There is little difference in the zinc protein content of bacteria and archea and between anaerobic and aerobic prokaryotes. Amongst eukaryotes we have placed them in order of complexity and quite probable in order of evolution starting from single cell metazoans, protozoa, followed by multicellular eukaryotes in the order of single organisms *C. elegans*, *D. melanogaster*, and *Homo sapiens*. The increases in zinc finger proteins in both numbers and percentages is seen to follow the order of complexity and probably the order of evolution. Figure 6 shows a two step changes in increase of these proteins between prokaryotes and unicellular organisms, protozoa and between protozoa, that is unicellular organisms and multicellular organisms. Complexity of organisms is

associated with changes of message systems and we note that the zinc fingers are very important transcription factors. The smaller percentage and number of transcription factors in plants can be correlated with the very different complexity of them compared with *Homo sapiens*.

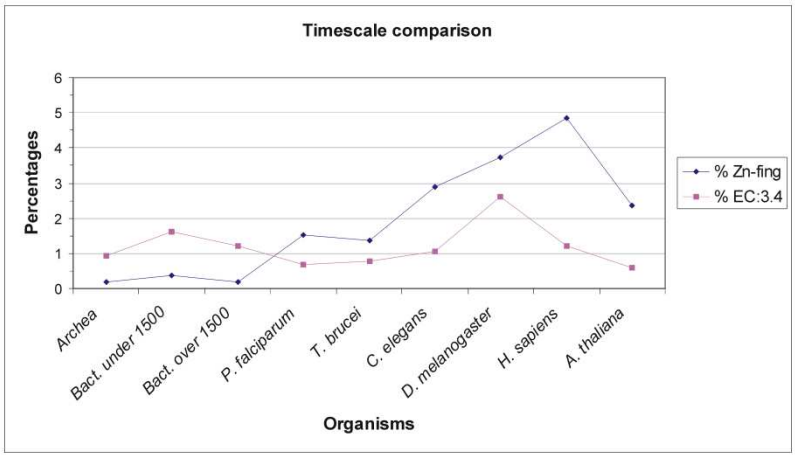


Figure 6 - A timescale comparison of small genomes in small and large prokaryotes, then in unicellular and finally in multicellular eukaryotes. It is notable that the percentages in Zn-finger content rise within this evolutionary series. The percentage value of EC:3.4 in small bacteria is higher than in the large bacteria.

We observed a greater percentage of hydrolytic zinc proteins in small bacteria and instead a low level in eukarya. In particular we see that the number of EC: 3.4 enzymes, the peptidases and proteases, has a very different pattern in the various organisms. Except for the fly, *D. melanogaster*, the percentage varies little being slightly lower in all eukaryotes than in prokaryotes. The high value in the fly could be related to its need to metamorphosing. The content of the EC:3.4 enzymes is high in all the analyzed organisms and we consider that this is a reflection in eukaryotes of the need to hydrolyze connective tissue for growth.

3.4 Copper Proteomes, phylogenetics and evolution

This paper is a continuation of the study of the connection between the changing environment and the changing use of particular elements in organisms in the course of their combined evolution (see *3.3 Zinc Proteomes, phylogenetics and evolution*). Here we treat the changes in copper proteins in historically the same increasingly oxidizing environmental conditions (11). We investigated 435 complete proteomes, 52 from archea, 337 from bacteria (247 aerobic and 90 anaerobic) and 46 from eukarya available in NCBI. Within bacteria, we considered only the organisms characterized in their oxygen request, so known aerobic or anaerobic. Our chosen example here is that of the copper proteins primarily involved in homeostasis, oxidases (EC:1.-) and electron carriers. Bacteria have been divided into aerobic and anaerobic, while eukarya were considered respecting to their complexity, using the selected organisms in the order: single-cell eukaryotes *S. cerevisiae, T. brucei* and *P. falciparum* and multicellular eukaryotes, *C. elegans, D. melanogaster, A. thaliana* and *H. sapiens*. The activities of the proteins in all the organisms have been divided using their major three separate functions: homeostatic proteins, electron transfer proteins and oxidases, treated as a sum of all such enzymes. Striking features in the eukaryotes are the rapid increase in the numbers of all three groups of copper proteins with complexity of the multicellular organisms and the even greater increase in plants, illustrated by *A. thaliana.* Figure 7 reports the percentages content of oxidases, homeostatic and electron carrier Cu-proteins, to be calibrated considering the whole genome sizes of the organisms. Our data indicate that copper was not used by the earliest anaerobic prokaryotes, as it was not an available element before there was oxidation of sulfides. Free copper ions in organisms are known to be poisonous and hence cells have always had proteins for maintaining a very low level of total copper, especially in their membranes and cytoplasm. The control is managed through storage in

homeostatic buffer proteins, such as metallothioneins in the cytoplasm and entry and exit pumps in the outer membrane. However copper became more and more valuable in cells as oxygen became more available, especially in oxidases in eukaryotic vesicles and outside cells.

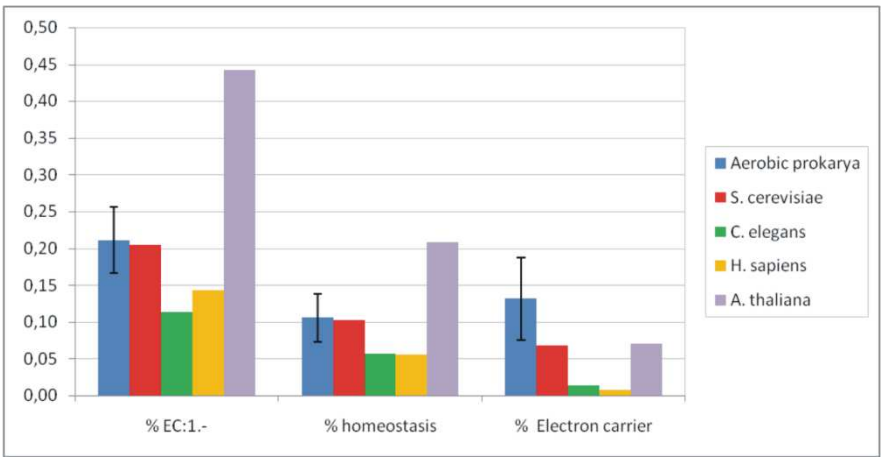


Figure 7 - The total percentages of copper proteins including oxidases, homeostatic and electron carriers.

We could observe the great difference between the content feature of copper and zinc proteins, described in 3.3. There are extremely few copper transcription factors or hydrolases in marked contrast to those of zinc, which is in virtually no oxidases, and they are largely in different cell compartments. These characteristics are indicative of the separate nature of the two metals. Copper is of use in oxidations as it can change valence but, as stated above, it presents a risk, especially in association with the cell nucleus. Zinc is more available and useful for hydrolytic reactions, it is nearly as powerful a Lewis acid as copper but unlike copper it cannot catalyze redox reactions. It can also act in signaling even to the nucleus in transcription factors as it is of low risk.

3.5 A simple protocol for the comparative analysis of the structure and occurrence of biochemical pathways across superkingdoms

In this work, we proposed a coherent, easy protocol for the identification of a set of proteins that can constitute an entire biochemical pathway on the basis of homology relationships detected through the presence of conserved domains and integrating, when available, 3D structural information. This protocol integrates all the tools that we have developed for and tested in our previous publications into a single package, which we called RDGB, Figure 8.

Figure 8 – The RDGB logo. The tool is available for downloading in the bioinformatic section of the CERM website (see *Methods* for details)

The tool not only integrates all the needed scripts and makes them easy to use for non-experts but enforces the use of a tested, internally consistent protocol in order to guarantee the reliability of the results. In addition, it provides a pre-ordered manner of storing the data which can be useful for subsequent analyses as well as further computational analyses. As an example, we analyzed two biochemical pathways, the degradation of aromatic hydrocarbons and the assembly of the Cu_A cofactor, in 1136 completely sequenced prokaryotic genomes. Aromatic hydrocarbons such as toluene or biphenyl are common contaminants of soil and groundwater and are listed as priority pollutants by the U.S. Environmental Protection Agency (31). One of the most attractive means to remove these compounds from polluted environments is through

bioremediation. Microbial cells cultivated on aromatic hydrocarbons often exhibit the induction of enzymes involved in metabolic pathways (32;33). The genes encoding these enzymes can be located either in plasmids or in chromosomal DNA. The bacterial degradation of aromatic hydrocarbons consists of many reaction steps, which have often been broadly separated into peripheral and central pathways, Figure 9.

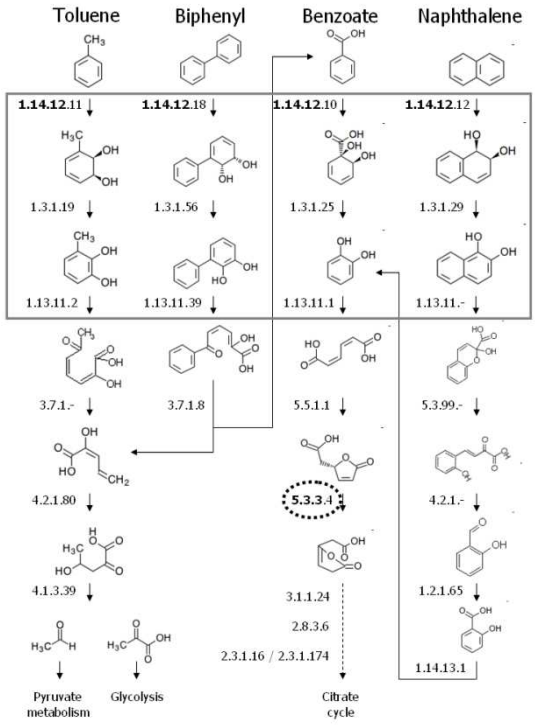


Figure 9 - Overview of the degradation of simple aromatic hydrocarbons, such as toluene, biphenyl, benzoate, naphthalene.

The substrates of interest in the present group of processes range from toluene to biphenyl and naphthalene, including various other compounds in which the aromatic ring(s) are differently substituted (34;35). Looking for the presence of all the proteins involved in the considered processes, we were able to define 919 organisms unable to aerobically degrade aromatic hydrocarbons, whereas 178 organisms were found to possess the right enzymatic portfolio. For the remaining 39 organisms, even after a manual check, we could identify only one out of the proteins needed for degradation process. It is possible

that they use an unconsidered substitute for overcome the protein deficiency, or it can be that they loss the capability of degrading aromatic compounds due to protein deletion. Cytochrome *c* oxidases use the Cu_A cofactor as the entry point of the electron that is delivered by cytochrome *c* into the enzyme (36). Cu_A is a dinuclear copper site contained in subunit II of the enzyme (Cox2) whose correct assembly is crucial for enzyme function (37). The assembly process is mediated by the soluble metallochaperone PCu_AC and the Sco1 thiol-disulfide reductase, which maintains the Cys residues in the Cu_A binding site of Cox2. We identified which of the same prokaryotic organisms of the previous section contained enzymes with a soluble Cu_A-binding domain. These were 549, corresponding to 48,3% of the ensemble investigated. The occurrence of both PCu_AC and Sco1 homologous was less frequent, corresponding respectively to 32.3% and 40.2% of the organisms analyzed. It is relevant to address the co-occurrence of these proteins. 283 organisms contained all three proteins, corresponding to 24.9% of the dataset. In 57 cases (5.0% of the organisms) a Cu_A-containing domain could not be detected but Sco1 (1.0%) or PCu_AC (0.2%) or both (3.8%) were contained in the proteome. The relatively common occurrence of the pair Sco1 and PCu_AC in the absence of any Cu_A-containing enzyme may suggest that the mechanism of formation of the Cu_A cofactor, or a close variant of it, may be relevant also for the assembly of other cuproenzymes. A Cu_A-containing enzyme co-occurs with either Sco1 or PCu_AC in respectively 120 (10.6% of all organisms) and 38 (3.3%) instances, demonstrating that none of the accessory proteins is always required for proper Cu_A assembly. Finally, it is worth noting that 108 organisms (9.5%) encode a Cu_A-containing enzyme while lacking both Sco1 and PCu_AC. This indicates that some yet uncharacterized assembly mechanisms may be operative in some organisms such as Mycobacteria (and various other Actinobacteria), δ-proteobacteria and Cyanobacteria.

4. Conclusions & Perspectives

Metalloproteins are known to be fundamental actors in biological key processes, including signal transduction, redox reactions, cellular respiration and so on. They represent one of the most diverse classes of proteins, with the intrinsic metal atoms providing catalytic, regulatory, or structural essential roles critical to protein function. With the advent of genome sequencing, a huge database of protein primary sequences has been accumulating. In parallel, a number of bioinformatic tools to investigate and expand upon this information, e.g. reconstructing and building relationships between protein families and superfamilies, have been developed. Surprisingly enough, very few of these resources are dedicated to the analysis of the interaction between metal ions and proteins, in spite of the importance of the roles that metals play in many proteins (both functional and structural).

In this framework, we analyzed entire metalloproteomes and specific metabolic pathways involving metalloproteins, for a deeper characterization of the role, uptake, transport and storage of trace metals essential for life. The reported data clearly indicate that in higher organisms zinc is essential not only to guarantee the proper functioning of a wide range of enzymatic activities, but mainly to achieve a tight control of gene expression. The recruitment of zinc in the latter group of physiological processes is a distinctive feature of eukaryota, and has been key to the development of their sophisticated mechanisms of interaction with the environment and, in multicellular organisms, of cell differentiation. Instead, copper became more and more valuable in cells as oxygen became more available, especially in oxidases in eukaryotic vesicles and outside cells. However, the value of copper therefore increased externally as seen from unicellular to multicellular eukaryotes.

As microbial pathogens require iron as an essential nutrient, they evolved a number of acquisition strategies to import heme from the extracellular medium. We obtained a comprehensive picture of the capabilities of both bacteria and archaea to carry out heme biosynthesis and uptake, providing a basis for further studies, for example, aimed at the development of molecular tools able to block them in human pathogens.

In this decade, there has been tremendous development in the fields of biology that end in "*omics*". The best-known discipline among them is *genomics*, but in the last years other well-developed disciplines raised up, as proteomics and metalloproteomics (38;39). Recent improvements in high-throughput sample separation, mass spectrometry, crystallography and NMR spectroscopy impact and positively on the proteomic treatment of proteins in systems biology. Bioinformatics is placing by side to these techniques, many research groups are providing continuously updated and improved web servers, tools and databases for proteins prediction, analysis, structural and functional characterization (40;41). The field of "computational biology" is thought to be a fundamental research landmark in the future, even if experimental works will be clearly required to further investigate the considered scenario or just confirm the available prediction. By now, to further enhance the bioinformatics development in the metalloproteomics field, we developed RDGB, a freeware bioinformatic tool that can significantly integrate the net of web resources available for the scientific community.

During my Ph.D. I produced a total of 5 papers (4 published and 1 reviewed for publication), which are listed below.

- Cavallaro G., Decaria L., Rosato A.
  "Genome-based analysis of heme biosynthesis and uptake in prokaryotic systems"
  J Proteome Res. 2008 Nov;7(11):4946-54. Epub 2008 Sep 23.
  Impact Factor: **5.132**

- Bertini I., Decaria L., Rosato A.
  "The annotation of the full zinc proteomes"
  J Biol Inorg Chem. 2010 Sep;15(7):1071-8. Epub 2010 May 5.
  Impact Factor: **3.415**

- Decaria L., Bertini I., Williams R.J.P.
  "Zinc proteomes, phylogenetics and evolution"
  Metallomics. 2010 Oct 1;2(10):706-9. Epub 2010 Aug 25.
  Impact Factor: -

- Decaria L., Bertini I., Williams R.J.P.
  "Copper proteomes, phylogenetics and evolution"
  Metallomics. 2010 Nov 1. [Epub ahead of print]
  Impact Factor: -

- Andreini C., Bertini I., Cavallaro G., Decaria L., Rosato A.
  "A simple protocol for the comparative analysis of the structure and occurrence of biochemical pathways across superkingdoms"
  Reviewed for publication by *J Chem. Inf. Model*. 2010 Dec 5.
  Impact factor: **3.882**

5. References

1. Nielsen, F. H. (2000) *Eur.J.Nutr.* 39, 62-66
2. Bertini, I., Sigel, A., and Sigel, H. (2001) *Handbook on Metalloproteins*, 1 Ed., Marcel Dekker, New York
3. Ideker, T., Galitski, T., and Hood, L. (2001) *Annu.Rev.Genomics Hum.Genet.* 2:343-72., 343-372
4. Auld, D. S. (2001) Zinc sites in metalloenzymes and related proteins. In Sigel, H., editor. *Handbook on Metalloproteins*, Marcell and Dekker, New York
5. Auld, D. S. (2001) *Biometals* 14, 271-313
6. Frausto da Silva, J. J. R. and Williams, R. J. P. (1991) *The Biological Chemistry of the Elements*, Oxford, Oxford
7. Maret, W. and Li, Y. (2009) *Chem.Rev.* 109, 4682-4707
8. (1986) *Zinc Enzymes*, Birkhauser, Boston
9. Christianson, D. W. and Fierke, C. A. (1996) *Acc.Chem.Res.* 29, 331-339
10. Gaither, L. A. and Eide, D. J. (2001) *Biometals* 14, 251-270
11. Ridge, P. G., Zhang, Y., and Gladyshev, V. N. (2008) *PLoS.ONE.* 3, e1378
12. Linder, M. C. (1991) *Biochemistry of Copper*, Plenum Press, New York
13. Linder, M. C., Wooten, L., Cerveza, P., Cotton, S., Shulze, R., and Lomeli, N. (1998) *Am.J.Clin.Nutr.* 67, 965S-971S
14. Linder, M. C. and Hazegh-Azam, M. (1996) *Am.J.Clin.Nutr.* 63, 797S-811S
15. De Freitas, J., Wintz, H., Kim, J. H., Poynton, H., Fox, T., and Vulpe, C. (2003) *Biometals* 16, 185-197
16. O'Halloran, T. V. and Culotta, V. C. (2000) *J.Biol.Chem.* 275, 25057-25060
17. Puig, S. and Thiele, D. J. (2002) *Curr.Opin.Chem.Biol.* 6, 171-180
18. Posey, J. E. and Gherardini, F. C. (2000) *Science* 288, 1651-1653
19. Ratledge, C. and Dover, L. G. (2000) *Annu.Rev.Microbiol.* 54, 881-941
20. Braun, V. (2001) *Int.J.Med.Microbiol.* 291, 67-79
21. Braun, V. and Hantke, K. (1997) Receptor-mediated bacterial iron transport. In Winkelmann, G. and Carano, C. J., editors. *Transition Metals in Microbial Metabolism.*, Harwood Acad. Publ. Amsterdam

22. Cescau, S., Cwerman, H., Letoffe, S., Delepelaire, P., Wandersman, C., and Biville, F. (2007) *Biometals* 20, 603-613

23. Wandersman, C. and Stojiljkovic, I. (2000) *Curr.Opin.Microbiol.* 3, 215-220

24. Friedman, D. B., Stauff, D. L., Pishchany, G., Whitwell, C. W., Torres, V. J., and Skaar, E. P. (2006) *PLoS.Pathog.* 2, e87

25. Skaar, E. P., Humayun, M., Bae, T., DeBord, K. L., and Schneewind, O. (2004) *Science* 305, 1626-1628

26. Andreini, C., Bertini, I., and Rosato, A. (2009) *Acc.Chem.Res.* 42, 1471-1479

27. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) *Nucleic Acids Res.* 32 Database issue, D138-D141

28. Sonnhammer, E. L., Eddy, S. R., and Durbin, R. (1997) *Proteins* 28, 405-420

29. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* 28, 235-242

30. Andreini, C., Banci, L., Bertini, I., and Rosato, A. (2006) *J.Proteome Res.* 5, 3173-3178

31. Cao, B., Nagarajan, K., and Loh, K. C. (2009) *Appl.Microbiol.Biotechnol.* 85, 207-228

32. De Vos, W. M., Van der Meer, J. R., Harayama, S., and Zehnder, A. J. B. (1992) *Microb.Rev.* 56, 677-694

33. Harayama, S. and Timmis, K. N. (1992) Aerobic Biodegradation of Aromatic Hydrocarbons by Bacteria. In Sigel, H. and Sigel, A., editors. *Metals Ions in Biological Systems*, Marcel Dekker,Inc, New York

34. Gibson, D. T. and Parales, R. E. (2000) *Curr.Opin.Biotechnol.* 11, 236-243

35. Vaillancourt, F. H., Bolin, J. T., and Eltis, L. D. (2006) *Crit Rev.Biochem.Mol.Biol.* 41, 241-267

36. Carr, H. S. and Winge, D. R. (2003) *Acc.Chem.Res.* 36, 309-316

37. Abriata, L. A., Banci, L., Bertini, I., Ciofi-Baffoni, S., Gkazonis, P., Spyroulias, G. A., Vila, A. J., and Wang, S. (2008) *Nat.Chem.Biol.* 4, 599-601

38. Shi, W. and Chance, M. R. (2008) *Cell Mol.Life Sci.* 65, 3040-3048

39. da Silva, M. A., Sussulini, A., and Arruda, M. A. (2010) *Expert.Rev.Proteomics.* 7, 387-400

40. Maret, W. (2010) *Metallomics.* 2, 117-125

41. Shi, W. and Chance, M. R. (2010) *Curr.Opin.Chem.Biol.*

6. Attachments

research articles Journal of proteome research

Genome-Based Analysis of Heme Biosynthesis and Uptake in Prokaryotic Systems

Gabriele Cavallaro,[†,‡] Leonardo Decaria,[†] and Antonio Rosato*[,†,‡]

Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, and Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Received February 26, 2008

Heme is the prosthetic group of many proteins that carry out a variety of key biological functions. In addition, for many pathogenic organisms, heme (acquired from the host) may constitute a very important source of iron. Organisms can meet their heme demands by taking it up from external sources, by producing the cofactor through a dedicated biosynthetic pathway, or both. Here we analyzed the distribution of proteins specifically involved in the processes of heme biosynthesis and heme uptake in 474 prokaryotic organisms. These data allowed us to identify which organisms are capable of performing none, one, or both processes, based on the similarity to known systems. Some specific instances where one or more proteins along the pathways had unusual modifications were singled out. For two key protein domains involved in heme uptake, we could build a series of structural models, which suggested possible alternative modes of heme binding. Future directions for experimental work are given.

Keywords: heme • heme biosynthesis • heme uptake • NEAT domain • Peripla_BP_2 domain

Introduction

Heme is the prosthetic group of many proteins that carry out a variety of key biological functions, including oxygen transport and sensing, enzyme catalysis and electron transfer.[1,2] The biosynthesis of heme occurs *via* a multistep process highly conserved across living organisms.[3,4] The pathway (see Figure 1) starts from δ-aminolevulinic acid and proceeds through the formation of porphobilinogen and hydroxymethylbilane to uroporphyrinogen III, which is the common precursor of heme and other tetrapyrroles, such as chlorophyll and vitamin B_{12}.[4] In prokaryotes, two distinct enzymes can catalyze the conversion of coproporphyrinogen III to protoporphyrinogen IX: one (called hemF) is homologous to the eukaryotic coproporphyrinogen oxidase and requires oxygen as a substrate, whereas the other (called hemN) is an oxygen-independent enzyme whose physiological oxidant is unknown.[4] Similarly, for the subsequent formation of protoporphyrin IX, prokaryotes can employ either the eukaryotic-like protoporphyrinogen oxidase hemY or the alternative enzyme hemG.[4] Previous genome analyses indicated that, while hemN and hemF can be both encoded in prokaryotic genomes, hemG and hemY appear to be mutually exclusive.[4] Heme biosynthesis in the bacterium *Desulfovibrio vulgaris*[5] as well as in the archaeon *Methanosarcina barkeri*[6] is known to deviate from the most common pathway at uroporphyrinogen III, which in these organisms is converted to the unusual intermediate precorrin-2 by two methylation reactions using *S*-adenosyl-L-methionine as the methyl donor.[6]

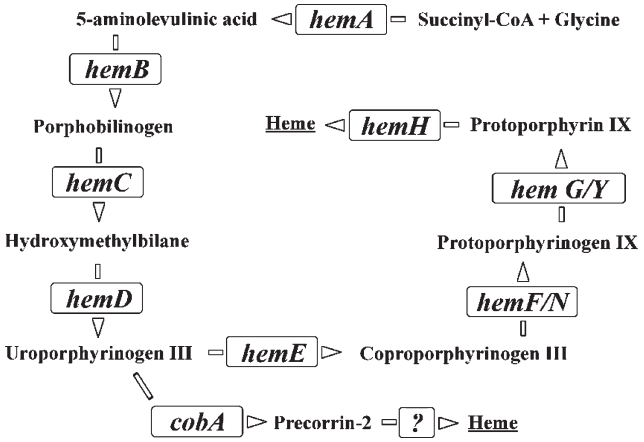


Figure 1. The heme biosynthesis pathway (conserved in eukaryotes and prokaryotes). The alternative pathway forming the Precorrin-2 intermediate for Archaea is also shown. Adapted from the KEGG database (http://www.genome.jp/kegg).

With the exception of a few species such as *Borrelia burgdorferi*,[7] the growth of microbial pathogens within the host requires iron as an essential nutrient.[8] Given the scarcity of free iron in tissues and cells, which is exacerbated by further limitation imposed by the host in response to infection,[9] pathogenic bacteria evolved a number of iron acquisition strategies, including the secretion of siderophores and the uptake of heme from host heme proteins.[10] Studies on *Staphylococcus aureus* indicated that heme is a preferential iron

* Corresponding author: Dr. Antonio Rosato, Magnetic Resonance Center, University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy. Fax, +39 055 4574253; tel, +39 055 4574267; e-mail, rosato@cerm.unifi.it.
† Magnetic Resonance Center (CERM), University of Florence.
‡ Department of Chemistry, University of Florence.

Published on Web 09/23/2008
10.1021/pr8004309 CCC: $40.75

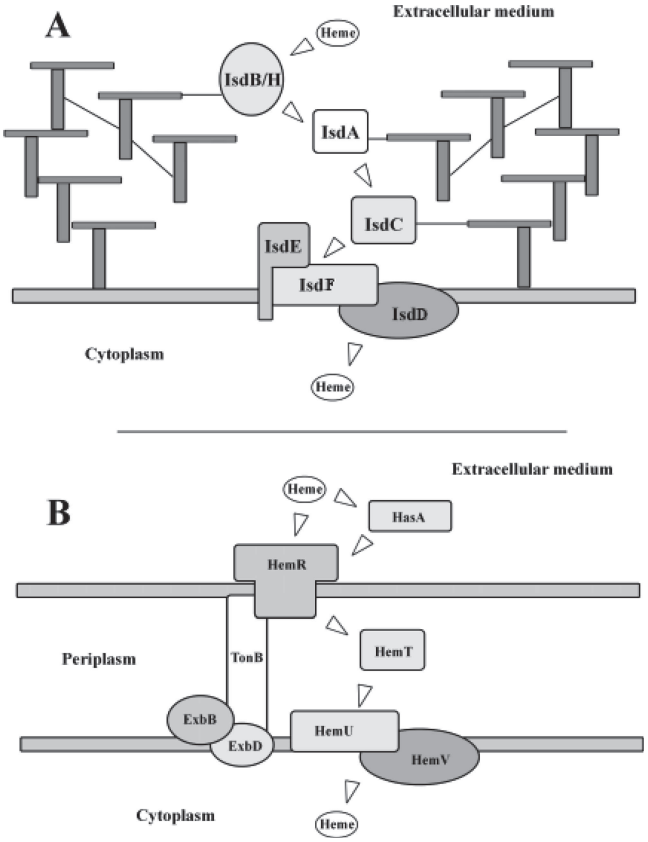


Figure 2. Heme uptake systems in bacteria. (A) Gram-positive bacteria. *Staphylococcus* has two alternative membrane heme transporters called HrtAB[64] and HtsAC,[11] *Streptococcus pyogenes* has Shr = IsdB/H,[65] Shp = IsdA/C[66] and HtsABC = IsdDEF (with HtsA = IsdE).[67] (B) Gram-negative bacteria. Note that the nomenclature varies among species, e.g., Shu in *Shigella*,[68] Phu in *Pseudomonas*,[69] Cha in *Campylobacter*,[70] Huv in *Vibrio*.[71] Some organisms may possess multiple uptake systems (e.g., Phu and Has in *Pseudomonas*[69]). In some cases HemR is associated with an auxiliary protein such as PhuW (with PhuR) in *Pseudomonas*[69] and ChaN (with ChaR) in *Campylobacter*.[70]

source for human pathogens, consistent with the localization of most body iron in hemoglobin and myoglobin.[11] Therefore, bacterial systems for heme uptake are considered as important virulence factors, and represent promising targets for novel therapeutic approaches. Several studies have provided structural and functional insights into the molecular components of these systems, as well as on their regulation mechanisms, contributing to establish models for the heme uptake machineries of both Gram-positive and Gram-negative bacteria (see Figure 2). Gram-positive bacteria (Figure 2A) use cell wall-anchored proteins to relay heme from host proteins to a specific ABC transporter, for delivery into the cytoplasm. In the well-characterized Isd system of *S. aureus*, heme transfer across the cell wall envelope involves the IsdA, IsdB, IsdC, and IsdH proteins, all of which contain one or more copies of a heme-binding domain called NEAT.[12−15] Heme is then imported through the IsdDEF complex into the cytoplasm, where it undergoes degradation by the monooxygenases IsdG and IsdI.[13] In Gram-negative bacteria (Figure 2B), heme must first be transported across the outer membrane by a cell-surface receptor, which is energized by a TonB−ExbB−ExbD system exploiting the proton motive force of the inner membrane.[16] Outer membrane receptors contain a transmembrane, heme-specific Plug domain.[17,18] Gram-negative bacteria can acquire heme directly from heme proteins of the host, or through extracellular heme-chelating proteins called hemophores, exemplified by *Serratia marcescens* HasA.[19] The subsequent transfer of heme from the periplasm to the cytoplasm is mediated by ABC transporters resembling those of Gram-positive bacteria. The understanding of this machinery relies on studies conducted on systems present in various proteobacterial species, including *Yersinia enterolitica*,[20] *Yersinia pestis*,[21] *Shigella dysenteriae*,[22] *Vibrio cholerae*,[23] *Pseudomonas aeruginosa*,[24] and *Bradyrhizobium japonicum*.[25] In the Hem system of *Yersinia* species, for instance, heme passes from the outer membrane protein HemR to HemT, a soluble periplasmic carrier (equivalent to the membrane-anchored IsdE protein of Gram-positive bacteria) which shuttles it to the HemUV complex for translocation through the inner membrane. Once in the cytoplasm, heme is bound by HemS, which is hypothesized to function in both heme storage and heme delivery to oxygenases (HemO) for degradation.[26,27]

The understanding of the homeostasis of metal ions in living systems is a current frontier of bioinorganic chemistry.[28] In this work, we used bioinformatics methods to identify genes encoding protein components of systems for heme biosynthesis and uptake in prokaryotic genomes, and we built homology models of selected proteins to integrate the currently available structural data. Our approach, which we successfully applied to the investigation of metalloprotein families such as cytochrome *c*[29] and Sco,[30] or of macromolecular machineries such as the cytochrome *c* maturation systems,[31] is based on the detection of homologues of the protein(s) of interest in complete proteomes followed by the inspection of the neighborhood of the genes encoding them. Possible homologues are detected with high sensitivity using hidden Markov models (HMMs) and then filtered on the basis of their manually curated functional annotation that is available from the COG database.[32] This methodology is conceptually similar to what is typically done in other studies of protein families and metabolic pathways across different proteomes, where homologies are most commonly detected using BLAST and context-based techniques are adopted to identify functional partners.[33−35] The present use of HMMs is motivated by their higher performance with respect to BLAST.[36,37]

The present work resulted in a comprehensive picture of the capabilities of both bacteria and archaea to carry out heme biosynthesis and uptake, providing a basis for further studies, for example, aimed at the development of molecular tools able to block them in human pathogens.

Methods

Identification of Genes Encoding Protein Domains Relevant to Heme Biosynthesis and Uptake. We used the Pfam database[38] (http://pfam.sanger.ac.uk) to identify the domains associated to the proteins involved in the processes of heme biosynthesis and uptake, based on the available literature data. Table 1 shows the domains that were taken into consideration, using the Hem system of *Yersinia* and the Isd system of *Staphylococcus* as the reference for Gram-negative and Gram-positive bacteria, respectively. Additionally, we mapped the same proteins onto the COG (Cluster of Orthologous Groups) database[32] (http://www.ncbi.nlm.nih.gov/COG), thus, linking the identified domains to one or more COG codes (Table 1). It should be noted that in both Pfam and COG databases the selected domains/codes may not be all uniquely associated to

Table 1. Functional Domains Characterizing the Heme Biosynthesis and Uptake Systems in Prokaryotes

domain	Pfam code	COG code	protein	process	no. of organisms[a]
ALAD	PF00490	COG0113	HemB	Biosynthesis	383
Porphobil_deam	PF01379	COG0181	HemC	Biosynthesis	387
Porphobil_deamC	PF03900	COG0181	HemC		371
HEM4	PF02602	COG1587	HemD	Biosynthesis	379
URO-D	PF01208	COG0407	HemE	Biosynthesis	335
Coprogen_oxidas	PF01218	COG0408	HemF	Biosynthesis	211
HemN_C	PF06969	COG0635	HemN	Biosynthesis	388
Amino_oxidase	PF01593	COG1232 COG1231	HemG	Biosynthesis	143
HemY_N	PF07219	COG3071 COG3898	HemY	Biosynthesis	174
Ferrochelatase	PF00762	COG0276	HemH	Biosynthesis	354
NEAT	PF05031	COG5386	IsdA/B/C/H	Uptake	42
Plug	PF07715	COG1629	HemR	Uptake	219[b]
TonB_dep_Rec	PF00593	COG1629	HemR		
Peripla_BP_2	PF01497	COG0614 COG4558	IsdE, HemT	Uptake	347
FecCD	PF01032	COG0609	IsdF, HemU	Uptake	324
HasA	PF06438	−	HasA	Uptake	13
Shp	[c]	−	Shp	Uptake	11
HtaA	PF04213	−	HtaA	Uptake	8

[a] Number of organisms containing at least one instance of the domain. [b] The Plug and TonB_dep_Rec domains are taken into account only when co-occurring in the same protein. [c] For this domain, there is no Pfam profile available. Consequently, a profile was built using the 11 sequences found in the 11 strains of *S. pyogenes*.

heme biosynthesis and uptake, therefore, leading to the inclusion of proteins that are involved in other processes. We discarded the domains/codes common to different processes that did not bring additional information with respect to the domains of Table 1. In particular, we did not use the ATPase component of the ABC transporter (IsdD, HemV) because it is redundant with regards to FecCD (IsdF, HemU), that is, all ABC transporters that contain FecCD contain also an ATPase component, yet the converse is not true. Similarly, we did not use the TonB−ExbB−ExbD system that energizes not only the outer membrane receptor HemR, but also other receptors. Furthermore, we discarded the *hemA* gene because not all heme biosynthesis pathways make use of it.

We used the HMMER program (http://hmmer.janelia.org)[39] to search the NCBI RefSeq sequence database[40] for matches to the hidden Markov models (HMMs) representing the selected domains. The HMMs were taken from the Pfam database without modifications. We investigated 474 complete prokaryotic genomes, 437 from Bacteria and 37 from Archaea. We applied a cutoff for the *E*-value of 1.0, which is relatively high, to minimize the possibility to miss distant homologues. The resulting hits were filtered as follows: we retained those assigned by the COGnitor program[32,41] to any of the COG numbers corresponding to the query domain as in Table 1, whereas we rejected those assigned to any other COG number. We retained also all the domains (ca. 10%) that COGnitor could not assign to any COG number. Finally, we compared all the retained hits against the whole Pfam database. We considered as true positives only those having the expected domain as the best match.

To assign an organism as being able to perform heme biosynthesis, uptake, or both, we required that its proteome encoded the majority of the proteins involved in these processes (i.e., those of Table 1). When one or a few proteins were missing, we manually checked the neighborhood of the codifying genes retrieved using SHOPS (http://bioinformatics.holstegelab.nl/shops/).[42] This approach permits the assignment for organisms having heme biosynthesis or uptake systems sufficiently similar to those already characterized (Figures 1 and 2). However, organisms assigned here as unable to perform one (or both) processes may carry it (them) out through alternative, unknown systems.

Structural Modeling. We used the structural modeling technique to obtain a deeper understanding of the mechanisms of intermolecular interaction of selected domain families (Peripla_BP_2, NEAT and Shp) with heme. For the modeling of the Peripla_BP_2 heme-binding domain, we exploited the available structure of the protein IsdE from *S. aureus* for Gram-positive organisms (Figure 3A, PDB entry 2Q8Q, GI: 157835873)[43] and the structure of the protein ShuT from *S. dysenteriae* for Gram-negative organisms (Figure 3B, PDB entry 2R7A, GI: 158430665).[44] To model NEAT domains, we exploited the available structure of the NEAT domain of the protein IsdA from *S. aureus* (Figure 3C, PDB entry 2ITF, GI: 122920748).[12] Finally, to model Shp domains, we used the structure of the homologue from *S. pyogenes* (Figure 3D, PDB entry 2Q7A, GI: 158429663).[45] Some of these structures have multiple molecules in the asymmetric unit. For modeling, only the first molecule was used. All the models were built using the program MODELER v.9.2, using default parameters.[46] The sequences of the template and target proteins were aligned with CLUSTALW.[47]

Results and Discussion

Species-Wide Distribution of the Pathways for Heme Biosynthesis and Uptake. We searched for potential homologues of the proteins known to characterize the pathways for the biosynthesis and the uptake of heme in 474 completely sequenced prokaryotic genomes (see Table S1 and Table S2 in the Supporting Information). Using the relevant Pfam sequence profiles, we initially retrieved about 33 000 domains (in some cases occurring within the same protein). As mentioned in Methods, this ensemble resulted from the application of a threshold of 1.0 for the *E*-value. Therefore, one expects that it contains several hits obtained by chance (false positives). The distribution of the *E*-values associated to the domains retrieved by HMMER is shown in Figure S1 in Supporting Information. False positives were removed through the use of both the COG and Pfam databases, yielding a total of slightly less than 12 000

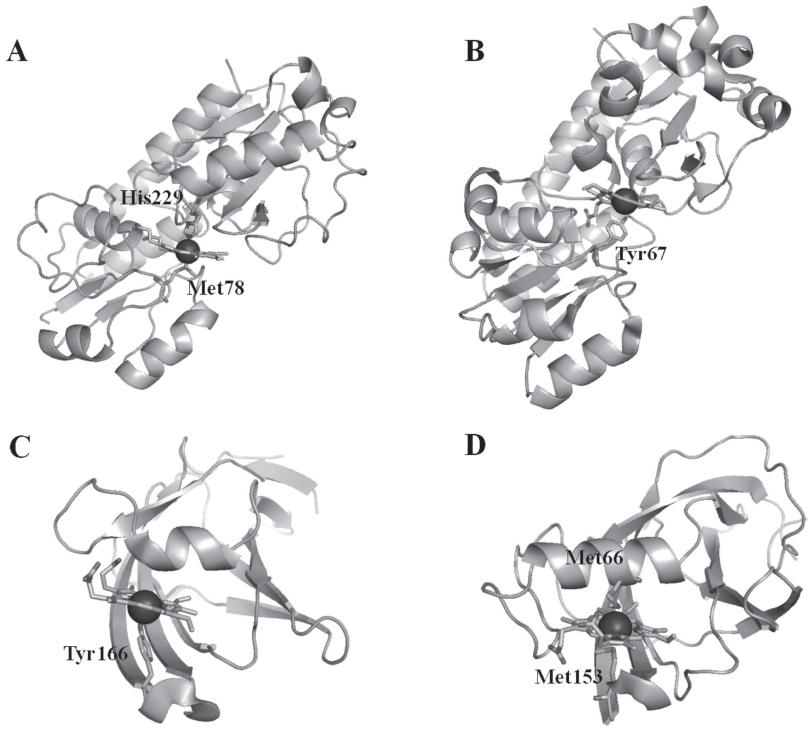


Figure 3. Ribbon representation of known structures of heme-transporters: (A) the Peripla_BP_2 domain from *S. aureus* IsdE (PDB code 2Q8Q); (B) the Peripla_BP_2 domain from *S. dysenteriae* ShuT (PDB code 2R7A); (C) the NEAT domain from *S. aureus* IsdA (PDB code 2ITF); (D) the Shp domain from the *S. pyogenes* Shp protein (PDB code 2Q7A). The axial ligands of the iron ion are shown.

domains that were considered as true positives. The distribution of the *E*-values of the 10 315 positive domains having a COG assignment indicates that, although the large majority of them (9930, 96.3%) are below 10^{-5}, there is a non-negligible fraction of higher *E*-values (Figure S1 in Supporting Information). In particular, 110 domains have *E*-values larger than 0.01, of which 66 have *E*-values larger than 0.1. Therefore, the selected threshold of 1.0 allowed us to detect a number of relatively distant homologues. Missing these hits would have led to possible errors in assigning an organism as capable of heme biosynthesis or uptake. For example, we would have missed the HemB protein of *Geobacter metallireducens* as well as the Plug-containing protein of *Chlorobium chlorochromatii*, resulting in apparently incomplete biosynthesis and uptake pathways, respectively. Our final data set included all the relevant proteins encoded in the 63 prokaryotic genomes available in the COG database. For the same 63 genomes, using only the COG system would have resulted in the loss of 3.5% of the hits. The results are summarized in Table 2. The very large number of proteins retrieved for some domains, in particular for the heme uptake process, can be accounted for by their relatively low specificity (e.g., FecCD).

The above data allowed us to characterize heme biosynthesis and uptake in all the considered prokaryotes. Out of 474 organisms, 168 can only synthesize heme, including 38 (27 Archaea and 11 Bacteria) that synthesize heme via the precorrin-2 pathway. Twenty organisms can only take up heme from an external source, and 218 can perform both processes (out of which, 12 synthesize heme via precorrin-2). Sixty-eight organisms are not able to synthesize heme or to take it up from the extracellular medium. These assignments are based on the presence or absence of heme biosynthesis or uptake systems sufficiently similar to those already characterized in the literature (Figures 1 and 2). In the remainder, we will refer to organisms as being able/unable to perform heme biosynthesis or uptake on this basis, although in principle some organisms could rely on different systems to perform one or both processes. The results of our predictions compare well with experimental evidence (Table S3 in Supporting Information).

Table 2. Number of Proteins Retrieved (top part, heme biosynthesis; middle part, uptake; the number of organisms where the proteins were retrieved is given in parenthesis) and Number of Organisms Performing the Investigated Biological Processes (bottom part)

		Bacteria (437)		
	Archaea (37)	Gram − (311)	Gram + (126)	Total
hemB	30 (30)	290 (273)	82 (80)	372 (353)
hemC	30 (30)	277 (277)	82 (80)	359 (357)
hemD	33 (30)	283 (269)	122 (80)	405 (349)
hemE	3 (3)	268 (264)	68 (68)	336 (332)
hemF	0 (0)	213 (211)	0 (0)	213 (211)
hemN	1 (1)	547 (279)	114 (108)	661 (387)
hemG	0 (0)	183 (77)	169 (66)	352 (143)
hemY	0 (0)	175 (174)	0 (0)	175 (174)
hemH	3 (3)	282 (267)	94 (84)	376 (351)
IsdA/B/C/H	0 (0)	1 (1)	108 (41)	109 (42)
HemR	0 (0)	2775 (219)	0 (0)	2775 (219)
IsdE, HemT	150 (33)	665 (208)	453 (106)	1118 (314)
IsdF, HemU	89 (30)	649 (201)	459 (93)	1108 (294)
HasA	0 (0)	16 (13)	0 (0)	16 (13)
Shp	0 (0)	0 (0)	11 (11)	11 (11)
HtaA	0 (0)	0 (0)	8 (8)	8 (8)
Biosynthesis only	30	95	43	138
Uptake only	0	8	12	20
Biosynthesis and Uptake	0	181	37	218
None	7	27	34	61

The large majority of Archaea (37 in total) is able to synthesize heme using the precorrin-2 pathway, whereas only a few species lack all heme biosynthesis enzymes. No archaeal organism can take up heme. Archaea thus seem to be able to acquire iron from the extracellular medium only using siderophores. Bacteria display a substantially higher differentiation than Archaea. For example, most Proteobacteria (Gram-negative) can synthesize as well as take up heme. HemY and HemG were not retrieved in ε-proteobacteria, which thus probably use a functionally equivalent enzyme to perform the corresponding step. Our extended data set fully confirms that HemG and HemY are mutually exclusive. In Firmicutes (Gram-positive) only some species of *Staphylococcus*, *Bacillus*, *Clostridium*, and all the *Listeria* are able to perform heme uptake. All Spirochaetes are unable to perform both heme uptake and biosynthesis, with the main exception of *Leptospira*, which synthesizes heme using the classic pathway.[48] The entire group of Chlamydiae/Verrucomicrobia can only synthesize heme. Some selected cases identified in bacterial genomes that constitute unique or unexpected results are discussed below.

The Gram-positive bacteria *Lactococcus*, some species of *Lactobacillus* and *Streptococcus*, *Leuconostoc mesenteroides*, and *Enterococcus faecalis* do not use the systems described above for the synthesis or the uptake of heme. Instead, they only have ferrochelatase (encoded by the gene *hemH*) and an O_2-independent coproporphyrinogen III oxidase (encoded by the gene *hemN*). These two enzymes could actually function in dismantling heme (e.g., the putative ferrochelatase could act as a ferrolyase)[49] acquired through an uncharacterised uptake system.

Among Proteobacteria, *Zymomonas mobilis* (α-proteobacteria) and *Anaeromyxobacter dehalogenans* (δ-proteobacteria) are anaerobic organisms that are able to synthesize heme. Their genomes encode an O_2-independent coproporphyrinogen III oxidase, but also the O_2-dependent enzyme (encoded by the gene *hemF*). Whether the latter enzyme is actually ever produced by the cell remains to be ascertained. In general, we noted that HemN is much more widespread than HemF (Table 1). Thirteen organisms encode only the O_2-dependent enzyme HemF. Among aerobic organisms, 29 encode only HemN.

All organisms belonging to the Chlamydiae/Verrucomicrobia group lack the C-terminal part of porphobilinogen deaminase (encoded by the gene *hemC*), resulting in a polypeptide of 240 residues, whereas the most common form of this enzyme comprises about 320 amino acids. It has been proposed that the C-terminal domain that is missing in Chlamydiae/Verrucomicrobia is involved only in interactions with the cell membrane and not in the catalytic activity.[50] Porphobilinogen deaminase is also modified in *Leptospira*, where it additionally performs the function generally carried out by the product of the gene *hemD* (uroporphyrinogen III synthetase). The C-terminal domain of the enzyme is replaced by the Pfam domain B_83352, which comprises about 270 residues. This domain is apparently responsible for conferring the additional functionality to *Leptospira* porphobilinogen deaminase.[48] An analogous situation was detected also in *Bdellovibrio bacteriovorus* (δ-proteobacteria). Unfortunately, it was not possible to build reliable models for the B_83352 domain, for which an experimental structural characterization is thus needed. It is important to note that the enzymes analyzed are relatively often able to catalyze more than one step along the heme biosynthesis pathway,[51] also through unusual intermediates.[52] This consideration can justify the apparent lack of some enzymes in the pathway of as many as 38 organisms. In these cases, the ability to synthesize heme could be maintained thanks to an extension of the catalytic capabilities of some enzymes.

The pathway for heme uptake is generally more strictly conserved than that for heme biosynthesis. We detected exceptions in *Sodalis glossinidius morsitans* (γ-proteobacteria), which is known to be able to take up heme,[53] but for which we could not identify the Plug domain in its HuvA receptor. In Actinobacteria (Gram-positive), heme uptake is performed only by a few species, such as *Corynebacterium diptheriae* or *Corynebacterium jeikeium* but not *Corynebacterium efficiens*, through the extracellullar HtaA heme-binding domain.[54] This domain was found only in some Actinobacteria, which, on the other hand, lack proteins containing the NEAT domain.

Considerations for the Treatment of Pathogenic Organisms. A total of 189 Gram-negative bacteria out of 311 analyzed are able to acquire heme from the extracellular medium, while 122 are not. Regarding Gram-positive bacteria, 49 out of 126 can take up heme. Among all the 474 organisms analyzed, 225 (155 Gram-negative bacteria, and 70 Gram-positive bacteria) are reported to be pathogenic for humans, animals or plants.

Eighty-eight Gram-negative pathogens are able to perform heme uptake, such as *Porphyromonas gingivalis*, *Campylobacter jejuni*, *Haemophilus influenzae*, *Salmonella enterica*, *S. dysenteriae*, and *Y. pestis*. Also plant pathogens such as *Ralstonia solanacearum*, *Pseudomonas syringae* pv *tomato* and *Erwinia carotovora atroseptica* are in this group. The remaining 67 Gram-negative pathogens cannot acquire heme from the extracellular medium, including *Legionella*, *Leptospira*, *Chlamydophyla pneumoniae*, *Francisella* and *Xylella*. Out of 70 Gram-positive pathogenic bacteria that have been investigated, 39 are able to perform heme uptake. Together with the highly characterized *S. aureus*, this group comprises *Bacillus anthracis*, *Clostridium tetani*, *Listeria* and *S. pyogenes*. The Gram-positive pathogens that cannot take up heme comprise *Mycobacteria*, and *E. faecalis*. The virulence of some strains of *Streptococcus pneumoniae* has been connected to hemin/iron transport.[55] We identified the two corresponding ABC transporters,[56] but we did not detect any associated NEAT domain. Therefore, we assigned *S. pneumoniae* as unable to take up heme. This is an example where the existence of an uncharacterized component of the uptake system needs to be ascertained experimentally.

Gram-negative bacteria that can take up heme are equally distributed between pathogenic and nonpathogenic. Instead, for Gram-positive bacteria, heme uptake seems to be more related to pathogenicity: 39 out of 49 bacteria taking up heme are pathogenic. Wide-spectrum treatment of bacterial infections through inibition of heme uptake thus appears more likely for Gram-positive rather than Gram-negative bacteria, while organism-specific treatment is certainly viable for bacteria of both groups, due to the importance of the heme uptake process as a source of iron.[8]

Structural Modeling of Domains Involved in Heme Uptake. As mentioned, inhibition of the heme uptake process may constitute a successful approach to the treatment of pathogens. For this reason, we wanted to analyze in further detail the mechanisms by which the cofactor is recognized and seized by the domains that are specifically devoted to this task. This can be achieved through the analysis of their three-dimensional structure, which can be obtained by the homology modeling approach.[57,58] Out of those listed in Table 1, we focused on the NEAT, Peripla_BP_2 and Shp domains, as they are strictly required components of the heme transport path-

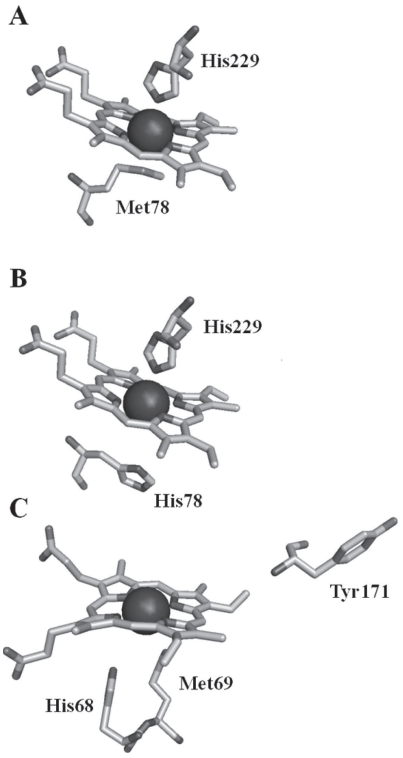


Figure 4. Iron-binding residues in Peripla_BP_2 domains. (A) Model for Gram-positive bacteria (some *Bacillus*, *Clostridium*, *Lactobacillus*, *Listeria*, *Staphylococcus*, and *Streptococcus*). This model is of *B. anthracis* GI: 47778325. (B) Model for all the Gram-positive bacteria not in panel A, where Met78 is substituted by His (some *Bacillus* and *Listeria*). This model is of *Listeria monocytogenes* GI: 16804469. (C) Model for Gram-negative bacteria, where the possible ligand may be not a Tyr residue. Both a His and a Met residue are present in the binding pocket. Also shown is a Tyr residue located in a spatially close loop. This model is of *D. vulgaris* GI: 46579061. Residue numbering is as in 2Q8Q in panels A and B, and as in 2R7A in panel C.

way but at the same time feature a significant sequence variability. It is important to note that the Peripla_BP_2 domain is known to be involved in the transport of molecules other than heme, for example, vitamin B_{12},[59] while the NEAT[12] and Shp[45,60] domains are heme-specific. The Shp domain is a modification of the NEAT domain that has been identified in *S. pyogenes*.[45] These domains are probably evolutionarily linked, but their sequences have diverged significantly (see later).

We analyzed the sequences of proteins containing the Peripla_BP_2 domain in Gram-positive bacteria that were predicted to be capable of heme uptake, with respect to the known structure 2Q8Q. Forty-two sequences, from *Bacillus*, *Clostridium*, *Lactobacillus*, *Listeria*, *Staphylococcus*, and *Streptococcus* conserve both heme iron ligands, Met78 and His229 (residue numbering of 2Q8Q). In 14 sequences, from *Bacillus* and *Listeria*, His229 is conserved, whereas Met78 is substituted by His. Domains from *Corynebacterium* species, on the other hand, lack both residues. These organisms are different also in that HtaA functionally replaces NEAT.[54] We can thus propose that, in Gram-positive bacteria, except *Corynebacterium* species, the axial ligands to the heme iron ion can only be Met/His or His/His (Figure 4A,B).

The known heme-binding Peripla_BP_2 domains from Gram-negative bacteria do not contain the Met78 and His229 residues. Instead, the available structures of the PhuT and ShuT proteins[44] demonstrate that a Tyr serves as the iron axial ligand, with the sixth coordination position being vacant. Multiple sequence alignments showed that this Tyr ligand is conserved only in 85 out of the 189 Gram-negative organisms that are predicted to be capable of taking up heme. For the other organisms, there is no obvious indication of the identity of the fifth ligand, which can be either another Tyr that does not align with that of PhuT and ShuT (e.g., being located in a loop facing the opposite side of the heme) or even a different amino acid. For example, structural modeling of *Desulfovibrio* Peripla_BP_2 domains suggested that the heme iron ion may be bound to either a His or a Met residue relatively close in sequence to the Tyr of ShuT, or, alternatively, to a Tyr residue located within another loop (Figure 4C). The latter Tyr residue is in fact quite far away from the iron ion in our model; however, the loop where it is located may be sufficiently flexible to allow the side chain to coordinate the iron ion.

The mechanism of function of Peripla_BP_2 and NEAT domains are expected to be quite different; nevertheless, in a large number of instances, they presumably adopt the same Tyr/- mode of iron coordination (see also below). This can be due to Tyr/- guaranteeing a low barrier for heme release along the transport pathway. However, Tyr/- is not the only mode of heme binding in Peripla_BP_2 domains from both Gram-positive and Gram-negative bacteria.

All NEAT domain-containing proteins, which were found only in Gram-positive bacteria, are cell-wall anchored, and many of them contain multiple NEAT domains. The NEAT domain was originally identified through bioinformatic methods and proposed to be involved in iron uptake as a receptor of an iron-siderophore complex.[15] Table 3 reports the number of proteins containing one or more NEAT domains identified in this work, and the number of NEAT domains in each protein. The three-dimensional structure of the NEAT domain of *S. aureus* IsdA has been solved in both the apo- (PDB code 2ITE) and holo-forms (PDB code 2ITF).[12] The structure of the holo-form of IsdC from the same organism is also available (PDB

Table 3. Number of Proteins Containing at Least one NEAT Domain Identified in This Work

organism	no. of proteins containing NEAT	no. of NEAT domains in each protein
Bacillus anthracis str. Ames	4	1-1-1-5
Bacillus anthracis str. Ames_0581	5	1-1-1-1-5
Bacillus anthracis str. Sterne	5	1-1-1-1-5
Bacillus cereus (3 strains)	4	1-1-1-5
Bacillus clausii	2	1-4
Bacillus halodurans	2	1-5
Bacillus thuringiensis (2 strains)	4	1-1-1-5
Clostridium novyi	3	1-2-3
Clostridium perfringens (3 strains)	2	1-4
Clostridium tetani	2	1-7
Lactobacillus brevis	2	1-1
Listeria (3 species)	2	1-3
Staphylococcus aureus (7 strains)[a]	4	1-1-2(1)-3(1)
Staphylococcus aureus str. RF122[a]	5	1(0)-1-1-2(1)-2
Staphylococcus aureus str. MRSA252[a]	3	1-1-2(1)
Streptococcus pyogenes (11 strains)	1	2

[a] Among the NEAT domains identified in the various strains of *S. aureus*, some lack the single ligand of the heme iron (Tyr166), and have been reported not to be able to bind heme.[62] In these cases, we reported in parenthesis the number of domains that do conserve Tyr166 and are thus predicted to be able to bind the cofactor.

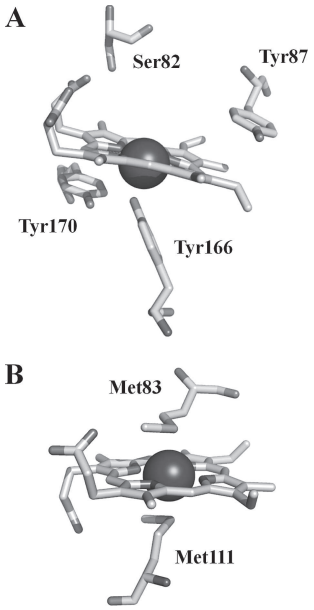


Figure 5. Iron-binding residues in NEAT domains. (A) Model for NEAT domains from *Bacillus*, *Listeria*, *Lactobacillus*, and *Staphylococcus*. Tyr166 is the iron-binding residue, Tyr87 and Tyr170 interact with heme via base-stacking, Ser82 via H-bonding. This model is of *Listeria innocua* GI: 16801354, domain 1. (B) Model for NEAT domains from *Clostridium* and *S. pyogenes*. Tyr166 is not conserved, the iron-binding residues are Met83 (residue numbering as in 2ITF, corresponding to Met66 in 2Q7A) and Met111 (corresponding to Met153 in 2Q7A). This model is of *S. pyogenes* GI: 15675635, domain 1.

code 2O6P).[61] These structures show that the heme iron is coordinated by a single amino acid, Tyr166 (numbering as in 2ITF). This key residue is not conserved in all the NEAT-containing proteins that we identified in the various strains of *S. aureus*, as already noted in the literature.[62] Although it is possible that the domains lacking Tyr166 bind heme in a different manner than in the 2ITF structure, the literature indicates that at least some of them do not interact appreciably with the cofactor.[62] Therefore, in Table 3, we report in parenthesis also the number of NEAT domains that are (predicted to be) competent for binding. The majority of the strains of *S. aureus* contain four proteins harboring NEAT domains, all of which contain a single domain conserving Tyr166. *Staphylococcus aureus str. RF122* and *Staphylococcus aureus str. MRSA252* constitute exceptions. In *Staphylococcus aureus str. RF122*, there is one additional protein, which however lacks Tyr166 and therefore may not be able to bind heme. In the same organism, IsdH contains two NEAT domains instead of three, as found in other strains, and both are predicted to be heme-binding (Table 3). In *Staphylococcus aureus str. MRSA252*, IsdH is lacking altogether. It therefore appears that the role of IsdH can be flexible and, importantly, is not strictly required.

All NEAT domains from *Bacillus*, *Listeria*, and *Lactobacillus* conserve Tyr166, and are therefore likely to bind heme similarly to what is observed in the 2ITF structure. These domains in fact contain also the Tyr87 and Tyr170 residues that have been shown to stabilize the adduct with base-stacking interaction with the pyrrolic rings, as well as Ser82, which acts as a H-donor in the hydrogen bond with a carboxylic group of the heme (Figure 5A). The number of NEAT-containing proteins in the above organisms ranges from 2 to 5, of which one is a multi-NEAT-domain protein and all others are single-domain, except for *Lactobacillus brevis*, which has two single-domain proteins (Table 3). The total number of heme-binding NEAT domains in a single organism can thus be as large as 9, versus 4 that is typical for *S. aureus*. A sequential pathway for heme delivery from methemoglobin to the cytoplasm of *S. aureus*, via the NEAT and then the Peripla_BP_2 domains, has been recently characterized *in vitro*.[63] In *S. aureus*, the NEAT domains not binding heme have been proposed to be involved in recognition of the heme-binding proteins of the host, such as hemoglobin.[62] The same function could be carried out also by the *Bacillus* or *Listeria* multidomain proteins. This would be in agreement with experimental indications that hemoglobin forms an adduct with the first domain of *S. aureus* IsdH with an approximate molar ratio of 1:4.[62] In addition, it is tempting to speculate that the multi-NEAT-domain proteins can also act as heme reservoirs from which the single-domain proteins can extract heme when there is limitation of an external source. In any case, due to the fact that three to five heme-binding domains are present in the same polypeptide, one would expect an increased efficiency in extracting heme from the host source.

When analyzing the NEAT domains identified in *Clostridium* species as well as in *S. pyogenes*, different possible heme-binding mechanisms became apparent, similarly to the case of the Peripla_BP_2 domain. Indeed, all the clostridial NEAT domains lack Tyr166. However, our structural models suggest that Met83, which is conserved in these organisms, can act as an iron axial ligand (Figure 5B). The side of the heme plane where both Tyr166 and Tyr170 lie is relatively exposed in the model structures, possibly permitting some structural rearrangements. Met/Met coordination of the heme iron has been observed in the structure (PDB code 2Q7A)[45] of a modified NEAT domain, called Shp, that is encoded in *S. pyogenes* in the same operon of a protein containing two classical NEAT domains (Table 3). These two Met residues are Met66 and Met153 (Figure 3D). The superposition of the structures of the NEAT domain (2ITF) and of the Shp domain (2Q7A) showed that there is a correspondence between residue 83 of the former and Met66 of the latter, and between the axial ligand Tyr166 of the former and Met153 of the latter. The structural superposition is of good quality, with a backbone rmsd of ca. 2.8 Å. High structure similarity and the correspondence between Tyr166 and Met153 ligands has been highlighted also in the comparison of 2O6P (the NEAT domain of IsdC) to 2Q7A.[45] Sequence alignments of the NEAT domains of *Clostridia* and of the various strains of *S. pyogenes*, and the Shp domains showed that Met83, which corresponds to the axial ligand Met66 of Shp, is conserved in the large majority of sequences. In addition, a second Met residue in the aforementioned NEAT domains (Met111) aligns well with the second axial ligand of the Shp domains (Met153) and is also quite conserved. Thus, it can be proposed that the majority of the NEAT domains of *Clostridia* are capable of binding heme, with the iron ion featuring Met/Met coordination (Figure 5B) as seen in Shp domains, rather than Tyr/- coordination as proposed for most Gram-positive bacteria. A few domains are probably not able to bind heme due to the absence of both Met83 and Met111, and also of Tyr166, and may play a role analogous to that of the NEAT domains of *S. aureus* that lack Tyr166. For the NEAT domains of *S. pyogenes*, sequence alignments to the NEAT domains that use Tyr166 to coordinate the heme iron allowed us to identify also a Tyr potentially corresponding to the axial ligand Tyr166. In the structural models of *S. pyogenes* NEAT

domains calculated using as template either a NEAT structure (2ITF), or the Shp structure (2Q7A), Met/Met coordination appears more likely on considerations of steric hindrance, with the Tyr corresponding to Tyr166 involved in base-stacking interactions with the heme moiety. However, experimental studies are needed to ascertain whether the NEAT domains of *S. pyogenes* feature Met/Met or Tyr/- iron coordination, as well as to validate our proposition of Met/Met coordination in *Clostridia*.

Concluding Remarks

We identified nearly 12 000 protein domains potentially involved in heme uptake or biosynthesis in 474 prokaryotic organisms. Of these, we predicted that 168 can only synthesize heme, 20 can only take up heme from an external source, and 218 can perform both processes, as judged on the basis of the presence or absence of corresponding protein systems similar to those described in the literature. Among these organisms, we singled out some instances of possible variations with respect to the "canonical" pathways that may be worth experimental investigation. Many pathogenic Gram-positive bacteria can take up heme, whereas this feature is less frequent among Gram-negative pathogens.

We used homology modeling to build a series of structural models for two key domains in the heme uptake pathway. The inspection of these models and the analysis of the corresponding sequence alignments suggested that there are possible alternative modes of heme binding. Again, this is an area where future experimental work would be quite useful.

Acknowledgment. This work has been supported by Ministero Italiano dell'Università e della Ricerca, project FIRB n. RBLA032ZM7.

Supporting Information Available: Table S1, list of the investigated organisms; Table S2, summary of results per organism; Table S3, comparison between predictions and experimental data for known systems; Figure S1, distribution of HMMER *E*-values for all hits and for only the hits assigned by COG. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) *Handbook of Metalloproteins*; Messerschmidt, A.; Wiley: Chichester, U.K., 2001; pp 1−1248.
(2) *The Porphyrin Handbook*; Kadish, K. M., Smith, K. M., Guilard, R.; Academic Press: Burlington, MA, 1999.
(3) Obornik, M.; Green, B. R. *Mol. Biol. Evol.* **2005**, *22*, 2343–2353.
(4) Panek, H.; O'Brian, M. R. *Microbiology* **2002**, *148*, 2273–2282.
(5) Ishida, T.; Yu, L.; Akutsu, H.; Ozawa, K.; Kawanishi, S.; Seto, A.; Inubushi, T.; Sano, S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4853–4858.
(6) Buchenau, B.; Kahnt, J.; Heinemann, I. U.; Jahn, D.; Thauer, R. K. *J. Bacteriol.* **2006**, *188*, 8666–8668.
(7) Posey, J. E.; Gherardini, F. C. *Science* **2000**, *288*, 1651–1653.
(8) Ratledge, C.; Dover, L. G. *Annu. Rev. Microbiol.* **2000**, *54*, 881–941.
(9) Braun, V. *Int. J. Med. Microbiol.* **2001**, *291*, 67–79.
(10) Wandersman, C.; Delepelaire, P. *Annu. Rev. Microbiol.* **2004**, *58*, 611–647.
(11) Skaar, E. P.; Humayun, M.; Bae, T.; DeBord, K. L.; Schneewind, O. *Science* **2004**, *305*, 1626–1628.
(12) Grigg, J. C.; Vermeiren, C. L.; Heinrichs, D. E.; Murphy, M. E. *Mol. Microbiol.* **2007**, *63*, 139–149.
(13) Maresso, A. W.; Schneewind, O. *Biometals* **2006**, *19*, 193–203.
(14) Pluym, M.; Muryoi, N.; Heinrichs, D. E.; Stillman, M. J. *J. Inorg. Biochem.* **2008**, *102*, 480–488.
(15) Andrade, M. A.; Ciccarelli, F. D.; Perez-Iratxeta, C.; Bork, P. *Genome Biology* **2002**, *3*, RESEARCH0047.1–0047.5.
(16) Higgs, P. I.; Larsen, R. A.; Postle, K. *Mol. Microbiol.* **2002**, *44*, 271–281.
(17) Letoffe, S.; Wecker, K.; Delepierre, M.; Delepelaire, P.; Wandersman, C. *J. Bacteriol.* **2005**, *187*, 4637–4645.
(18) Oke, M.; Sarra, R.; Ghirlando, R.; Farnaud, S.; Gorringe, A. R.; Evans, R. W.; Buchanan, S. K. *FEBS Lett.* **2004**, *564*, 294–300.
(19) Izadi, N.; Henry, Y.; Haladjian, J.; Goldberg, M. E.; Wandersman, C.; Delepierre, M.; Lecroisey, A. *Biochemistry* **1997**, *36*, 7050–7057.
(20) Stojiljkovic, I.; Hantke, K. *EMBO J.* **1992**, *11*, 4359–4367.
(21) Thompson, J. M.; Jones, H. A.; Perry, R. D. *Infect. Immun.* **1999**, *67*, 3879–3892.
(22) Wyckoff, E. E.; Duncan, D.; Torres, A. G.; Mills, M.; Maase, K.; Payne, S. M. *Mol. Microbiol.* **1998**, *28*, 1139–1152.
(23) Occhino, D. A.; Wyckoff, E. E.; Henderson, D. P.; Wrona, T. J.; Payne, S. M. *Mol. Microbiol.* **1998**, *29*, 1493–1507.
(24) Ochsner, U. A.; Johnson, Z.; Vasil, M. L. *Microbiology* **2000**, *146* (Pt 1), 185–198.
(25) Nienaber, A.; Hennecke, H.; Fischer, H. M. *Mol. Microbiol.* **2001**, *41*, 787–800.
(26) Hirotsu, S.; Chu, G. C.; Unno, M.; Lee, D. S.; Yoshida, T.; Park, S. Y.; Shiro, Y.; Ikeda-Saito, M. *J. Biol. Chem.* **2004**, *279*, 11937–11947.
(27) Genco, C. A.; Dixon, D. W. *Mol. Microbiol.* **2001**, *39*, 1–11.
(28) Bertini, I.; Rosato, A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3601–3604.
(29) Bertini, I.; Cavallaro, G.; Rosato, A. *Chem. Rev.* **2006**, *106*, 90–115.
(30) Banci, L.; Bertini, I.; Cavallaro, G.; Rosato, A. *J. Proteome Res.* **2007**, *6*, 1568–1579.
(31) Bertini, I.; Cavallaro, G.; Rosato, A. *J. Inorg. Biochem.* **2007**, *101*, 1798–1811.
(32) Tatusov, R. L.; Galperin, M. Y.; Natale, D. A.; Koonin, E. V. *Nucleic Acids Res.* **2000**, *28*, 33–36.
(33) Gabaldon, T.; Huynen, M. A. *Cell. Mol. Life Sci.* **2004**, *61*, 930–944.
(34) Zambelli, B.; Musiani, F.; Savini, M.; Tucker, P.; Ciurli, S. *Biochemistry* **2007**, *46*, 3171–3182.
(35) Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5879–5884.
(36) Park, J.; Karplus, K.; Barrett, C.; Hughey, R.; Haussler, D.; Hubbard, T.; Chothia, C. *J. Mol. Biol.* **1998**, *284*, 1201–1210.
(37) Gough, J.; Karplus, K.; Hughey, R.; Chothia, C. *J. Mol. Biol.* **2001**, *313*, 903–919.
(38) Bateman, A.; Coin, L.; Durbin, R.; Finn, R. D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E. L.; Studholme, D. J.; Yeats, C.; Eddy, S. R. *Nucleic Acids Res.* **2004**, *32*.
(39) Eddy, S. R. *Bioinformatics* **1998**, *14*, 755–763.
(40) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. *Nucleic Acids Res.* **2007**, *35*, D61–D65.
(41) Tatusov, R. L.; Natale, D. A.; Garkavtsev, I. V.; Tatusova, T. A.; Shankavaram, U. T.; Rao, B. S.; Kiryutin, B.; Galperin, M. Y.; Fedorova, R. D.; Koonin, E. V. *Nucleic Acids Res.* **2001**, *29*, 22–28.
(42) van Bakel, H.; Huynen, M.; Wijmenga, C. *Bioinformatics.* **2004**, *20*, 2644–2655.
(43) Grigg, J. C.; Vermeiren, C. L.; Heinrichs, D. E.; Murphy, M. E. *J. Biol. Chem.* **2007**, *282*, 28815–28822.
(44) Ho, W. W.; Li, H.; Eakanunkul, S.; Tong, Y.; Wilks, A.; Guo, M.; Poulos, T. L. *J. Biol. Chem.* **2007**, *282*, 35796–35802.
(45) Aranda, R.; Worley, C. E.; Liu, M.; Bitto, E.; Cates, M. S.; Olson, J. S.; Lei, B.; Phillips, G. N., Jr. *J. Mol. Biol.* **2007**, *374*, 374–383.
(46) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. *Current Protocols in Protein Science*; John Wiley & Sons: Brooklyn, NY, Chapter 2, Unit, 2007.
(47) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
(48) Guégan, R.; Camadro, J. M.; Saint Girons, I.; Picardeau, M. *Mol. Microbiol.* **2003**, *49*, 745–754.
(49) Taketani, S.; Ishigaki, M.; Mizutani, A.; Uebayashi, M.; Numata, M.; Ohgari, Y.; Kitajima, S. *Biochemistry* **2007**, *46*, 15054–15061.
(50) Helliwell, J. R.; Nieh, Y. P.; Habash, J.; Faulder, P. F.; Raftery, J.; Cianci, M.; Wulff, M.; Hadener, A. *Faraday Discuss.* **2003**, *122*, 131–144.
(51) Hansson, M.; Hederstedt, L. *J. Bacteriol.* **1994**, *176*, 5962–5970.
(52) Hansson, M.; Gustafsson, M. C.; Kannangara, C. G.; Hederstedt, L. *Biochim. Biophys. Acta* **1997**, *1340*, 97–104.
(53) Toh, H.; Weiss, B. L.; Perkin, S. A.; Yamashita, A.; Oshima, K.; Hattori, M.; Aksoy, S. *Genome Res.* **2006**, *16*, 149–156.
(54) Kunkle, C. A.; Schmitt, M. P. *J. Bacteriol.* **2003**, *185*, 6826–6840.
(55) Tai, S. S.; Lee, C. J.; Winter, R. E. *Infect. Immun.* **1993**, *61*, 5401–5405.

(56) Brown, J. S.; Gilliland, S. M.; Holden, D. W. *Mol. Microbiol.* **2001**, *40*, 572–585.
(57) Sanchez, R.; Sali, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 206–214.
(58) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. *Annu. Rev. Biophys.Bbiomol. Struct.* **2000**, *29*, 291–325.
(59) Karpowich, N. K.; Huang, H. H.; Smith, P. C.; Hunt, J. F. *J. Biol. Chem.* **2003**, *278*, 8429–8434.
(60) Ran, Y.; Zhu, H.; Liu, M.; Fabian, M.; Olson, J. S.; Aranda, R.; Phillips, G. N., Jr.; Dooley, D. M.; Lei, B. *J. Biol. Chem.* **2007**, *282*, 31380–31388.
(61) Sharp, K. H.; Schneider, S.; Cockayne, A.; Paoli, M. *J. Biol. Chem.* **2007**, *282*, 10625–10631.
(62) Pilpa, R. M.; Fadeev, E. A.; Villareal, V. A.; Wong, M. L.; Phillips, M.; Clubb, R. T. *J. Mol. Biol.* **2006**, *360*, 435–447.
(63) Zhu, H.; Xie, G.; Liu, M.; Olson, J. S.; Fabian, M.; Dooley, D. M.; Lei, B. *J. Biol. Chem.* **2008**.
(64) Friedman, D. B.; Stauff, D. L.; Pishchany, G.; Whitwell, C. W.; Torres, V. J.; Skaar, E. P. *PLoS Pathog.* **2006**, *2*, e87.
(65) Bates, C. S.; Montanez, G. E.; Woods, C. R.; Vincent, R. M.; Eichenbaum, Z. *Infect. Immun.* **2003**, *71*, 1042–1055.
(66) Lei, B.; Smoot, L. M.; Menning, H. M.; Voyich, J. M.; Kala, S. V.; Deleo, F. R.; Reid, S. D.; Musser, J. M. *Infect. Immun.* **2002**, *70*, 4494–4500.
(67) Lei, B.; Liu, M.; Voyich, J. M.; Prater, C. I.; Kala, S. V.; Deleo, F. R.; Musser, J. M. *Infect. Immun.* **2003**, *71*, 5962–5969.
(68) Mills, M.; Payne, S. M. *J. Bacteriol.* **1995**, *177*, 3004–3009.
(69) Ochsner, U. A.; Johnson, Z.; Vasil, M. L. *Microbiology* **2000**, *146* (Pt 1), 185–198.
(70) van Vliet, A. H.; Ketley, J. M.; Park, S. F.; Penn, C. W. *FEMS Microbiol. Rev.* **2002**, *26*, 173–186.
(71) Mourino, S.; Osorio, C. R.; Lemos, M. L. *J. Bacteriol.* **2004**, *186*, 6159–6167.

PR8004309

Table S1. List of the investigated organisms.

Acidobacteria bacterium Ellin345
Acidothermus cellulolyticus 11B
Acidovorax JS42
Acidovorax avenae citrulli AAC00-1
Acinetobacter baumannii ATCC 17978
Acinetobacter sp ADP1
Actinobacillus pleuropneumoniae L20
Aeromonas hydrophila ATCC 7966
Aeropyrum pernix
Agrobacterium tumefaciens C58 Cereon
Agrobacterium tumefaciens C58 UWash
Alcanivorax borkumensis SK2
Alkalilimnicola ehrlichei MLHE-1
Anabaena variabilis ATCC 29413
Anaeromyxobacter dehalogenans 2CP-C
Anaplasma marginale St Maries
Anaplasma phagocytophilum HZ
Aquifex aeolicus
Archaeoglobus fulgidus
Arthrobacter FB24
Arthrobacter aurescens TC1
Aster yellows witches-broom phytoplasma AYWB
Azoarcus BH72
Azoarcus sp EbN1
Bacillus anthracis Ames
Bacillus anthracis Ames 0581
Bacillus anthracis str Sterne
Bacillus cereus ATCC14579
Bacillus cereus ATCC 10987
Bacillus cereus ZK
Bacillus clausii KSM-K16
Bacillus halodurans
Bacillus licheniformis ATCC 14580
Bacillus licheniformis DSM 13
Bacillus subtilis
Bacillus thuringiensis Al Hakam
Bacillus thuringiensis konkukian
Bacteroides fragilis NCTC 9434
Bacteroides fragilis YCH46
Bacteroides thetaiotaomicron VPI-5482
Bartonella bacilliformis KC583
Bartonella henselae Houston-1
Bartonella quintana Toulouse
Baumannia cicadellinicola Homalodisca coagulata
Bdellovibrio bacteriovorus
Bifidobacterium adolescentis ATCC 15703
Bifidobacterium longum
Bordetella bronchiseptica
Bordetella parapertussis

Bordetella pertussis
Borrelia afzelii PKo
Borrelia burgdorferi
Borrelia garinii PBi
Bradyrhizobium japonicum
Brucella abortus 9-941
Brucella melitensis
Brucella melitensis biovar Abortus
Brucella suis 1330
Buchnera aphidicola
Buchnera aphidicola Cc Cinara cedri
Buchnera aphidicola Sg
Buchnera sp
Burkholderia 383
Burkholderia cenocepacia AU 1054
Burkholderia cenocepacia HI2424
Burkholderia cepacia AMMD
Burkholderia mallei ATCC 23344
Burkholderia mallei NCTC 10229
Burkholderia mallei NCTC 10247
Burkholderia mallei SAVP1
Burkholderia pseudomallei 1106a
Burkholderia pseudomallei 1710b
Burkholderia pseudomallei 668
Burkholderia pseudomallei K96243
Burkholderia thailandensis E264
Burkholderia xenovorans LB400
Campylobacter fetus 82-40
Campylobacter jejuni
Campylobacter jejuni 81-176
Campylobacter jejuni RM1221
Candidatus Blochmannia floridanus
Candidatus Blochmannia pennsylvanicus BPEN
Candidatus Carsonella ruddii
Candidatus Pelagibacter ubique HTCC1062
Candidatus Ruthia magnifica Cm Calyptogena magnifica
Carboxydothermus hydrogenoformans Z-2901
Caulobacter crescentus
Chlamydia muridarum
Chlamydia trachomatis
Chlamydia trachomatis A HAR-13
Chlamydophila abortus S26 3
Chlamydophila caviae
Chlamydophila felis Fe C-56
Chlamydophila pneumoniae AR39
Chlamydophila pneumoniae CWL029
Chlamydophila pneumoniae J138
Chlamydophila pneumoniae TW 183
Chlorobium chlorochromatii CaD3
Chlorobium phaeobacteroides DSM 266
Chlorobium tepidum TLS

Chromobacterium violaceum
Chromohalobacter salexigens DSM 3043
Clostridium acetobutylicum
Clostridium difficile 630
Clostridium novyi NT
Clostridium perfringens
Clostridium perfringens ATCC 13124
Clostridium perfringens SM101
Clostridium tetani E88
Clostridium thermocellum ATCC 27405
Colwellia psychrerythraea 34H
Corynebacterium diphtheriae
Corynebacterium efficiens YS-314
Corynebacterium glutamicum ATCC 13032 Bielefeld
Corynebacterium glutamicum ATCC 13032 Kitasato
Corynebacterium jeikeium K411
Coxiella burnetii
Cyanobacteria bacterium Yellowstone A-Prime
Cyanobacteria bacterium Yellowstone B-Prime
Cytophaga hutchinsonii ATCC 33406
Dechloromonas aromatica RCB
Dehalococcoides CBDB1
Dehalococcoides ethenogenes 195
Deinococcus geothermalis DSM 11300
Deinococcus radiodurans
Desulfitobacterium hafniense Y51
Desulfotalea psychrophila LSv54
Desulfovibrio desulfuricans G20
Desulfovibrio vulgaris DP4
Desulfovibrio vulgaris Hildenborough
Ehrlichia canis Jake
Ehrlichia chaffeensis Arkansas
Ehrlichia ruminantium Gardel
Ehrlichia ruminantium Welgevonden
Ehrlichia ruminantium str. Welgevonden
Enterococcus faecalis V583
Erwinia carotovora atroseptica SCRI1043
Erythrobacter litoralis HTCC2594
Escherichia coli 536
Escherichia coli APEC O1
Escherichia coli CFT073
Escherichia coli K12
Escherichia coli O157H7
Escherichia coli O157H7 EDL933
Escherichia coli UTI89
Escherichia coli W3110
Francisella tularensis FSC 198
Francisella tularensis holarctica
Francisella tularensis holarctica OSU18
Francisella tularensis novicida U112
Francisella tularensis tularensis

Frankia CcI3
Frankia alni ACN14a
Fusobacterium nucleatum
Geobacillus kaustophilus HTA426
Geobacter metallireducens GS-15
Geobacter sulfurreducens
Gloeobacter violaceus
Gluconobacter oxydans 621H
Gramella forsetii KT0803
Granulobacter bethesdensis CGDNIH1
Haemophilus ducreyi 35000HP
Haemophilus influenzae
Haemophilus influenzae 86 028NP
Haemophilus somnus 129PT
Hahella chejuensis KCTC 2396
Haloarcula marismortui ATCC 43049
Halobacterium sp
Haloquadratum walsbyi
Halorhodospira halophila SL1
Helicobacter acinonychis Sheeba
Helicobacter hepaticus
Helicobacter pylori 26695
Helicobacter pylori HPAG1
Helicobacter pylori J99
Herminiimonas arsenicoxydans
Hyperthermus butylicus
Hyphomonas neptunium ATCC 15444
Idiomarina loihiensis L2TR
Jannaschia CCS1
Lactobacillus acidophilus NCFM
Lactobacillus brevis ATCC 367
Lactobacillus casei ATCC 334
Lactobacillus delbrueckii bulgaricus
Lactobacillus delbrueckii bulgaricus ATCC BAA-365
Lactobacillus gasseri ATCC 33323
Lactobacillus johnsonii NCC 533
Lactobacillus plantarum
Lactobacillus sakei 23K
Lactobacillus salivarius UCC118
Lactococcus lactis
Lactococcus lactis cremoris MG1363
Lactococcus lactis cremoris SK11
Lawsonia intracellularis PHE MN1-00
Legionella pneumophila Lens
Legionella pneumophila Paris
Legionella pneumophila Philadelphia 1
Leifsonia xyli xyli CTCB0
Leptospira borgpetersenii serovar Hardjo-bovis JB197
Leptospira borgpetersenii serovar Hardjo-bovis L550
Leptospira interrogans serovar Copenhageni
Leptospira interrogans serovar Lai

Leuconostoc mesenteroides ATCC 8293
Listeria innocua
Listeria monocytogenes
Listeria monocytogenes 4b F2365
Listeria welshimeri serovar 6b SLCC5334
Magnetococcus MC-1
Magnetospirillum magneticum AMB-1
Mannheimia succiniciproducens MBEL55E
Maricaulis maris MCS10
Marinobacter aquaeolei VT8
Mesoplasma florum L1
Mesorhizobium BNC1
Mesorhizobium loti
Methanobacterium thermoautotrophicum
Methanococcoides burtonii DSM 6242
Methanococcus jannaschii
Methanococcus maripaludis C5
Methanococcus maripaludis S2
Methanocorpusculum labreanum Z
Methanoculleus marisnigri JR1
Methanopyrus kandleri
Methanosaeta thermophila PT
Methanosarcina acetivorans
Methanosarcina barkeri fusaro
Methanosarcina mazei
Methanosphaera stadtmanae
Methanospirillum hungatei JF-1
Methylibium petroleiphilum PM1
Methylobacillus flagellatus KT
Methylococcus capsulatus Bath
Moorella thermoacetica ATCC 39073
Mycobacterium JLS
Mycobacterium KMS
Mycobacterium MCS
Mycobacterium avium 104
Mycobacterium avium paratuberculosis
Mycobacterium bovis
Mycobacterium bovis BCG Pasteur 1173P2
Mycobacterium leprae
Mycobacterium smegmatis MC2 155
Mycobacterium tuberculosis CDC1551
Mycobacterium tuberculosis H37Rv
Mycobacterium ulcerans Agy99
Mycobacterium vanbaalenii PYR-1
Mycoplasma capricolum ATCC 27343
Mycoplasma gallisepticum
Mycoplasma genitalium
Mycoplasma hyopneumoniae 232
Mycoplasma hyopneumoniae 7448
Mycoplasma hyopneumoniae J
Mycoplasma mobile 163K

Mycoplasma mycoides
Mycoplasma penetrans
Mycoplasma pneumoniae
Mycoplasma pulmonis
Mycoplasma synoviae 53
Myxococcus xanthus DK 1622
Nanoarchaeum equitans
Natronomonas pharaonis
Neisseria gonorrhoeae FA 1090
Neisseria meningitidis FAM18
Neisseria meningitidis MC58
Neisseria meningitidis Z2491
Neorickettsia sennetsu Miyayama
Nitrobacter hamburgensis X14
Nitrobacter winogradskyi Nb-255
Nitrosococcus oceani ATCC 19707
Nitrosomonas europaea
Nitrosomonas eutropha C71
Nitrosospira multiformis ATCC 25196
Nocardia farcinica IFM10152
Nocardioides JS614
Nostoc sp
Novosphingobium aromaticivorans DSM 12444
Oceanobacillus iheyensis
Oenococcus oeni PSU-1
Onion yellows phytoplasma
Parachlamydia sp UWE25
Paracoccus denitrificans PD1222
Pasteurella multocida
Pediococcus pentosaceus ATCC 25745
Pelobacter carbinolicus
Pelobacter propionicus DSM 2379
Pelodictyon luteolum DSM 273
Photobacterium profundum SS9
Photorhabdus luminescens
Picrophilus torridus DSM 9790
Pirellula sp
Polaromonas JS666
Polaromonas naphthalenivorans CJ2
Porphyromonas gingivalis W83
Prochlorococcus marinus AS9601
Prochlorococcus marinus CCMP1375
Prochlorococcus marinus MED4
Prochlorococcus marinus MIT9313
Prochlorococcus marinus MIT 9301
Prochlorococcus marinus MIT 9303
Prochlorococcus marinus MIT 9312
Prochlorococcus marinus MIT 9515
Prochlorococcus marinus NATL1A
Prochlorococcus marinus NATL2A
Propionibacterium acnes KPA171202

Pseudoalteromonas atlantica T6c
Pseudoalteromonas haloplanktis TAC125
Pseudomonas aeruginosa
Pseudomonas aeruginosa UCBPP-PA14
Pseudomonas entomophila L48
Pseudomonas fluorescens Pf-5
Pseudomonas fluorescens PfO-1
Pseudomonas putida KT2440
Pseudomonas syringae phaseolicola 1448A
Pseudomonas syringae pv B728a
Pseudomonas syringae tomato DC3000
Psychrobacter arcticum 273-4
Psychrobacter cryohalolentis K5
Psychromonas ingrahamii 37
Pyrobaculum aerophilum
Pyrobaculum calidifontis JCM 11548
Pyrobaculum islandicum DSM 4184
Pyrococcus abyssi
Pyrococcus furiosus
Pyrococcus horikoshii
Ralstonia eutropha H16
Ralstonia eutropha JMP134
Ralstonia metallidurans CH34
Ralstonia solanacearum
Rhizobium etli CFN 42
Rhizobium leguminosarum bv viciae 3841
Rhodobacter sphaeroides 2 4 1
Rhodobacter sphaeroides ATCC 17029
Rhodococcus RHA1
Rhodoferax ferrireducens T118
Rhodopseudomonas palustris BisA53
Rhodopseudomonas palustris BisB18
Rhodopseudomonas palustris BisB5
Rhodopseudomonas palustris CGA009
Rhodopseudomonas palustris HaA2
Rhodospirillum rubrum ATCC 11170
Rickettsia bellii RML369-C
Rickettsia conorii
Rickettsia felis URRWXCal2
Rickettsia prowazekii
Rickettsia typhi wilmington
Roseobacter denitrificans OCh 114
Rubrobacter xylanophilus DSM 9941
Saccharophagus degradans 2-40
Saccharopolyspora erythraea NRRL 2338
Salinibacter ruber DSM 13855
Salmonella enterica Choleraesuis
Salmonella enterica Paratypi ATCC 9150
Salmonella typhi
Salmonella typhi Ty2
Salmonella typhimurium LT2

Shewanella ANA-3
Shewanella MR-4
Shewanella MR-7
Shewanella W3-18-1
Shewanella amazonensis SB2B
Shewanella baltica OS155
Shewanella denitrificans OS217
Shewanella frigidimarina NCIMB 400
Shewanella loihica PV-4
Shewanella oneidensis
Shigella boydii Sb227
Shigella dysenteriae
Shigella flexneri 2a
Shigella flexneri 2a 2457T
Shigella flexneri 5 8401
Shigella sonnei Ss046
Silicibacter TM1040
Silicibacter pomeroyi DSS-3
Sinorhizobium meliloti
Sodalis glossinidius morsitans
Solibacter usitatus Ellin6076
Sphingopyxis alaskensis RB2256
Staphylococcus aureus COL
Staphylococcus aureus MW2
Staphylococcus aureus Mu50
Staphylococcus aureus N315
Staphylococcus aureus NCTC 8325
Staphylococcus aureus RF122
Staphylococcus aureus USA300
Staphylococcus aureus aureus MRSA252
Staphylococcus aureus aureus MSSA476
Staphylococcus epidermidis ATCC 12228
Staphylococcus epidermidis RP62A
Staphylococcus haemolyticus
Staphylococcus saprophyticus
Staphylothermus marinus F1
Streptococcus agalactiae 2603
Streptococcus agalactiae A909
Streptococcus agalactiae NEM316
Streptococcus mutans
Streptococcus pneumoniae D39
Streptococcus pneumoniae R6
Streptococcus pneumoniae TIGR4
Streptococcus pyogenes M1 GAS
Streptococcus pyogenes MGAS10270
Streptococcus pyogenes MGAS10394
Streptococcus pyogenes MGAS10750
Streptococcus pyogenes MGAS2096
Streptococcus pyogenes MGAS315
Streptococcus pyogenes MGAS5005
Streptococcus pyogenes MGAS6180

Streptococcus pyogenes MGAS8232
Streptococcus pyogenes MGAS9429
Streptococcus pyogenes SSI-1
Streptococcus sanguinis SK36
Streptococcus thermophilus CNRZ1066
Streptococcus thermophilus LMD-9
Streptococcus thermophilus LMG 18311
Streptomyces avermitilis
Streptomyces coelicolor
Sulfolobus acidocaldarius DSM 639
Sulfolobus solfataricus
Sulfolobus tokodaii
Symbiobacterium thermophilum IAM14863
Synechococcus CC9311
Synechococcus CC9605
Synechococcus CC9902
Synechococcus elongatus PCC 6301
Synechococcus elongatus PCC 7942
Synechococcus sp WH8102
Synechocystis PCC6803
Syntrophobacter fumaroxidans MPOB
Syntrophomonas wolfei Goettingen
Syntrophus aciditrophicus SB
Thermoanaerobacter tengcongensis
Thermobifida fusca YX
Thermococcus kodakaraensis KOD1
Thermofilum pendens Hrk 5
Thermoplasma acidophilum
Thermoplasma volcanium
Thermosynechococcus elongatus
Thermotoga maritima
Thermus thermophilus HB27
Thermus thermophilus HB8
Thiobacillus denitrificans ATCC 25259
Thiomicrospira crunogena XCL-2
Thiomicrospira denitrificans ATCC 33889
Treponema denticola ATCC 35405
Treponema pallidum
Trichodesmium erythraeum IMS101
Tropheryma whipplei TW08 27
Tropheryma whipplei Twist
Ureaplasma urealyticum
Verminephrobacter eiseniae EF01-2
Vibrio cholerae
Vibrio fischeri ES114
Vibrio parahaemolyticus
Vibrio vulnificus CMCP6
Vibrio vulnificus YJ016
Wigglesworthia brevipalpis
Wolbachia endosymbiont of Brugia malayi TRS
Wolbachia endosymbiont of Drosophila melanogaster

Wolinella succinogenes
Xanthomonas campestris
Xanthomonas campestris 8004
Xanthomonas campestris vesicatoria 85-10
Xanthomonas citri
Xanthomonas oryzae KACC10331
Xanthomonas oryzae MAFF 311018
Xylella fastidiosa
Xylella fastidiosa Temecula1
Yersinia enterocolitica 8081
Yersinia pestis Antiqua
Yersinia pestis CO92
Yersinia pestis KIM
Yersinia pestis Nepal516
Yersinia pestis biovar Mediaevails
Yersinia pseudotuberculosis IP32953
Zymomonas mobilis ZM4

Name	S. K.	Group	hemB	hemC	hemC_C-term	hemD	hemE	hemF	HemN	hemG	hemY	hemH	Biosynth.	Precorrin-2	FecCD	NEAT	Peripla_BP_2	Plug+Rec	HasA	HtaA	Shp	Uptake	Gram	Pathogenicity
Aeropyrum_pernix	Archea	Crenarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	1	0	2	0	0	0	0	NO		No
Hyperthermus_butylicus	Archea	Crenarchaeota	1	1	1	1	0	0	1	0	0	0	NO	YES	1	0	1	0	0	0	0	NO		No
Pyrobaculum_aerophilum	Archea	Crenarchaeota	1	1	1	3	0	0	0	0	0	0	NO	YES	1	0	2	0	0	0	0	NO		No
Pyrobaculum_calidifontis_JCM_11548	Archea	Crenarchaeota	1	1	1	2	0	0	0	0	0	0	NO	YES	1	0	2	0	0	0	0	NO		
Pyrobaculum_islandicum_DSM_4184	Archea	Crenarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	3	0	6	0	0	0	0	NO		
Staphylothermus_marinus_F1	Archea	Crenarchaeota	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	1	0	0	0	0	NO		
Sulfolobus_acidocaldarius_DSM_639	Archea	Crenarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	0	0	2	0	0	0	0	NO		No
Sulfolobus_solfataricus	Archea	Crenarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	0	0	2	0	0	0	0	NO		No
Sulfolobus_tokodaii	Archea	Crenarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	0	0	2	0	0	0	0	NO		No
Thermophilum_pendens_Hrk_5	Archea	Crenarchaeota	0	0	0	0	0	0	0	0	0	0	NO	NO	1	0	1	0	0	0	0	NO		
Archaeoglobus_fulgidus	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	2	0	2	0	0	0	0	NO		No
Haloarcula_marismortui_ATCC_43049	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	3	0	8	0	0	0	0	NO		No
Halobacterium_sp	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	3	0	4	0	0	0	0	NO		No
Haloquadratum_walsbyi	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	1	0	3	0	0	0	0	NO		No
Methanobacterium_thermoautotrophicum	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	0	0	0	0	0	0	0	NO		No
Methanococcoides_burtonii_DSM_6242	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	4	0	4	0	0	0	0	NO		No
Methanococcus_jannaschii	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	3	0	4	0	0	0	0	NO		No
Methanococcus_maripaludis_C5	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	1	0	3	0	0	0	0	NO		
Methanococcus_maripaludis_S2	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	2	0	3	0	0	0	0	NO		No
Methanocorpusculum_labreanum_Z	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	6	0	10	0	0	0	0	NO		
Methanoculleus_marisnigri_JR1	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	4	0	7	0	0	0	0	NO		
Methanopyrus_kandleri	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	0	0	0	0	0	0	0	NO		No
Methanosaeta_thermophila_PT	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	6	0	14	0	0	0	0	NO		
Methanosarcina_acetivorans	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	14	0	21	0	0	0	0	NO		No
Methanosarcina_barkeri_fusaro	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	8	0	16	0	0	0	0	NO		No
Methanosarcina_mazei	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	1	0	1	0	0	0	0	NO		No
Methanosphaera_stadtmanae	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	0	0	0	0	0	0	0	NO		
Methanospirillum_hungatei_JF-1	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	8	0	11	0	0	0	0	NO		No
Natronomonas_pharaonis	Archea	Euryarchaeota	1	1	1	1	0	0	0	0	0	0	NO	YES	1	0	2	0	0	0	0	NO		
Picrophilus_torridus_DSM_9790	Archea	Euryarchaeota	1	1	1	1	1	0	0	0	0	1	YES		1	0	3	0	0	0	0	NO		
Pyrococcus_abyssi	Archea	Euryarchaeota	0	0	0	0	0	0	0	0	0	0	NO	NO	2	0	2	0	0	0	0	NO		No
Pyrococcus_furiosus	Archea	Euryarchaeota	0	0	0	0	0	0	0	0	0	0	NO	NO	2	0	2	0	0	0	0	NO		No
Pyrococcus_horikoshii	Archea	Euryarchaeota	0	0	0	0	0	0	0	0	0	0	NO	NO	2	0	2	0	0	0	0	NO		No
Thermococcus_kodakaraensis_KOD1	Archea	Euryarchaeota	0	0	0	0	0	0	0	0	0	0	NO	NO	4	0	3	0	0	0	0	NO		
Thermoplasma_acidophilum	Archea	Euryarchaeota	1	1	1	1	1	0	0	0	0	1	YES		1	0	2	0	0	0	0	NO		No
Thermoplasma_volcanium	Archea	Euryarchaeota	1	1	1	1	1	0	0	0	0	1	YES		1	0	2	0	0	0	0	NO		No
Nanoarchaeum_equitans	Archea	Nanoarchaeota	0	0	0	0	0	0	0	0	0	0	NO	NO	1	0	0	0	0	0	0	NO		No
Acidobacteria_bacterium_Ellin345	Bacteria	Acidobacteria	1	1	1	1	1	0	1	3	0	1	YES		0	0	1	11	0	0	0	NO	-	
Solibacter_usitatus_Ellin6076	Bacteria	Acidobacteria	1	1	1	3	2	0	1	4	0	2	YES		0	0	1	54	0	0	0	NO	-	No
Acidothermus_cellulolyticus_11B	Bacteria	Actinobacteria	1	1	1	2	1	0	1	2	0	1	YES		0	0	0	0	0	0	0	NO	+	
Arthrobacter_aurescens_TC1	Bacteria	Actinobacteria	1	1	1	1	1	0	1	3	0	1	YES		7	0	3	0	0	1	0	YES	+	No
Arthrobacter_FB24	Bacteria	Actinobacteria	1	1	1	2	1	0	1	5	0	1	YES		2	0	1	0	0	0	0	NO	+	No
Bifidobacterium_adolescentis_ATCC_15703	Bacteria	Actinobacteria	0	0	0	0	0	0	1	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	No
Bifidobacterium_longum	Bacteria	Actinobacteria	0	0	0	0	0	0	1	0	0	0	NO	NO	1	0	0	0	0	0	0	NO	+	No
Corynebacterium_diphtheriae	Bacteria	Actinobacteria	1	1	1	1	1	0	1	1	0	1	YES		7	0	6	0	0	1	0	YES	+	Human
Corynebacterium_efficiens_YS-314	Bacteria	Actinobacteria	1	1	1	1	1	0	1	1	0	1	YES		5	0	4	0	0	0	0	NO	+	No
Corynebacterium_glutamicum_ATCC_13032_Bielefeld	Bacteria	Actinobacteria	1	1	1	1	1	0	1	1	0	1	YES		10	0	11	0	0	1	0	YES	+	No
Corynebacterium_glutamicum_ATCC_13032_Kitasato	Bacteria	Actinobacteria	1	1	1	1	1	0	1	1	0	1	YES		10	0	10	0	0	1	0	YES	+	No
Corynebacterium_jeikeium_K411	Bacteria	Actinobacteria	1	1	1	1	1	0	1	1	0	1	YES		11	0	7	0	0	1	0	YES	+	Human
Frankia_alni_ACN14a	Bacteria	Actinobacteria	1	1	1	2	1	0	2	4	0	1	YES		4	0	6	0	0	0	0	NO	+	No
Frankia_CcI3	Bacteria	Actinobacteria	1	1	1	2	1	0	1	4	0	1	YES		1	0	2	0	0	0	0	NO	+	
Leifsonia_xyli_xyli_CTCB0	Bacteria	Actinobacteria	1	1	1	1	1	0	1	1	0	1	YES		1	0	2	0	0	0	0	NO	+	Plant
Mycobacterium_avium_104	Bacteria	Actinobacteria	1	1	1	2	1	0	1	2	0	1	YES		1	0	2	0	0	0	0	NO	+	Mammal
Mycobacterium_avium_paratuberculosis	Bacteria	Actinobacteria	1	1	1	2	1	0	1	3	0	1	YES		2	0	3	0	0	0	0	NO	+	Mammal
Mycobacterium_bovis	Bacteria	Actinobacteria	1	1	1	2	1	0	1	4	0	1	YES		1	0	2	0	0	0	0	NO	+	Human, Animal
Mycobacterium_bovis_BCG_Pasteur_1173P2	Bacteria	Actinobacteria	1	1	1	2	1	0	1	4	0	1	YES		1	0	2	0	0	0	0	NO	+	Bovine
Mycobacterium_JLS	Bacteria	Actinobacteria	1	1	1	2	1	0	1	7	0	1	YES		0	0	2	0	0	0	0	NO	+	No
Mycobacterium_KMS	Bacteria	Actinobacteria	1	1	1	2	1	0	1	6	0	1	YES		0	0	2	0	0	0	0	NO	+	No
Mycobacterium_leprae	Bacteria	Actinobacteria	1	1	1	1	1	0	0	1	0	1	YES		0	0	1	0	0	0	0	NO	+	Human
Mycobacterium_MCS	Bacteria	Actinobacteria	1	1	1	2	1	0	1	6	0	1	YES		0	0	2	0	0	0	0	NO	+	No
Mycobacterium_smegmatis_MC2_155	Bacteria	Actinobacteria	1	1	1	2	1	0	1	3	0	1	YES		4	0	7	0	0	0	0	NO	+	Human
Mycobacterium_tuberculosis_CDC1551	Bacteria	Actinobacteria	1	1	1	2	1	0	1	4	0	1	YES		1	0	2	0	0	0	0	NO	+	Human
Mycobacterium_tuberculosis_H37Rv	Bacteria	Actinobacteria	1	1	1	2	1	0	1	4	0	1	YES		1	0	2	0	0	0	0	NO	+	Human
Mycobacterium_ulcerans_Agy99	Bacteria	Actinobacteria	1	1	1	2	1	0	1	4	0	1	YES		0	0	2	0	0	0	0	NO	+	Human
Mycobacterium_vanbaalenii_PYR-1	Bacteria	Actinobacteria	1	1	1	2	1	0	1	4	0	1	YES		1	0	3	0	0	0	0	NO	+	
Nocardia_farcinica_IFM10152	Bacteria	Actinobacteria	1	1	1	2	1	0	1	4	0	1	YES		4	0	5	0	0	0	0	NO	+	Human, Animal
Nocardioides_JS614	Bacteria	Actinobacteria	1	1	1	2	1	0	1	5	0	1	YES		0	0	1	0	0	0	0	NO	+	No
Propionibacterium_acnes_KPA171202	Bacteria	Actinobacteria	1	1	1	2	1	0	1	3	0	1	YES		8	0	10	0	0	1	0	YES	+	Human
Rhodococcus_RHA1	Bacteria	Actinobacteria	1	1	1	3	1	0	1	9	0	1	YES		4	0	11	0	0	0	0	NO	+	
Rubrobacter_xylanophilus_DSM_9941	Bacteria	Actinobacteria	1	1	1	1	0	0	0	0	0	0	NO	YES	3	0	6	0	0	0	0	NO	+	No
Saccharopolyspora_erythraea_NRRL_2338	Bacteria	Actinobacteria	1	1	1	3	1	0	2	9	0	1	YES		3	0	8	0	0	0	0	NO	+	No

Streptomyces_avermitilis	Bacteria	Actinobacteria	1	2	2	2	1	0	1	3	0	1	YES		3	0	3	0	0	0	0	NO	+	No
Streptomyces_coelicolor	Bacteria	Actinobacteria	1	2	2	2	1	0	1	3	0	1	YES		6	0	7	0	0	1	0	YES	+	No
Symbiobacterium_thermophilum_IAM14863	Bacteria	Actinobacteria	0	0	0	0	0	0	1	0	0	1	NO	NO	0	0	4	0	0	0	0	NO	+	No
Thermobifida_fusca_YX	Bacteria	Actinobacteria	1	1	1	1	1	0	2	2	0	1	YES		6	0	5	0	0	0	0	NO	+	No
Tropheryma_whipplei_TW08_27	Bacteria	Actinobacteria	1	1	1	1	1	0	0	0	0	1	YES		2	0	2	0	0	1	0	YES	+	Human
Tropheryma_whipplei_Twist	Bacteria	Actinobacteria	1	1	1	1	1	0	0	0	0	1	YES		2	0	2	0	0	0	0	NO	+	Human
Agrobacterium_tumefaciens_C58_Cereon	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		8	0	7	7	0	0	0	YES	-	Plant
Agrobacterium_tumefaciens_C58_UWash	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		8	0	7	7	0	0	0	YES	-	Plant
Anaplasma_marginale_St_Maries	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Cattle
Anaplasma_phagocytophilum_HZ	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Bartonella_bacilliformis_KC583	Bacteria	Alphaproteobacteria	0	0	0	0	0	0	1	0	0	0	NO	NO	1	0	1	1	0	0	0	YES	-	Human
Bartonella_henselae_Houston-1	Bacteria	Alphaproteobacteria	0	0	0	0	0	0	1	0	0	0	NO	NO	1	0	1	1	0	0	0	YES	-	Human, Animal
Bartonella_quintana_Toulouse	Bacteria	Alphaproteobacteria	0	0	0	0	0	0	1	0	0	0	NO	NO	1	0	1	1	0	0	0	YES	-	Human
Bradyrhizobium_japonicum	Bacteria	Alphaproteobacteria	1	1	1	2	1	1	4	0	1	1	YES		1	0	2	10	0	0	0	YES	-	No
Brucella_abortus_9-941	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		2	0	2	2	0	0	0	YES	-	Human, Cattle
Brucella_melitensis	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		2	0	3	2	0	0	0	YES	-	Human, Animal
Brucella_melitensis_biovar_Abortus	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		2	0	2	2	0	0	0	YES	-	
Brucella_suis_1330	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		2	0	2	2	0	0	0	YES	-	Human, Animal
Candidatus_Pelagibacter_ubique_HTCC1062	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	1	0	1	YES		1	0	0	0	0	0	0	NO	-	
Caulobacter_crescentus	Bacteria	Alphaproteobacteria	1	1	1	0	1	1	2	0	1	1	YES		1	0	1	60	0	0	0	YES	-	No
Ehrlichia_canis_Jake	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Dog
Ehrlichia_chaffeensis_Arkansas	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Ehrlichia_ruminantium_Gardel	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Ruminant
Ehrlichia_ruminantium_str._Welgevonden	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Ruminant
Ehrlichia_ruminantium_Welgevonden	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Ruminant
Erythrobacter_litoralis_HTCC2594	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	1	0	1	YES		1	0	2	18	0	0	0	YES	-	No
Gluconobacter_oxydans_621H	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	1	0	1	YES		2	0	2	24	0	0	0	YES	-	
Granulobacter_bethesdensis_CGDNIH1	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	8	0	0	0	NO	-	Human
Hyphomonas_neptunium_ATCC_15444	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	40	0	0	0	YES	-	No
Jannaschia_CCS1	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	3	1	0	0	0	YES	-	No
Magnetospirillum_magneticum_AMB-1	Bacteria	Alphaproteobacteria	1	1	1	1	1	0	2	0	2	1	YES		0	0	0	2	0	0	0	NO	-	No
Maricaulis_maris_MCS10	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	1	24	1	0	0	YES	-	
Mesorhizobium_BNC1	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	3	2	0	0	0	YES	-	No
Mesorhizobium_loti	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	5	0	1	1	YES		2	0	2	1	0	0	0	YES	-	No
Neorickettsia_sennetsu_Miyayama	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	0	1	YES		0	0	0	0	0	0	0	NO	-	Human
Nitrobacter_hamburgensis_X14	Bacteria	Alphaproteobacteria	1	1	1	2	1	1	1	0	1	1	YES		2	0	2	9	0	0	0	YES	-	
Nitrobacter_winogradskyi_Nb-255	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		3	0	3	13	0	0	0	YES	-	
Novosphingobium_aromaticivorans_DSM_12444	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	1	0	1	YES		1	0	1	48	0	0	0	YES	-	No
Paracoccus_denitrificans_PD1222	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		17	0	14	17	0	0	0	YES	-	No
Rhizobium_etli_CFN_42	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		5	0	4	2	0	0	0	YES	-	
Rhizobium_leguminosarum_bv_viciae_3841	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		5	0	6	3	0	0	0	YES	-	No
Rhodobacter_sphaeroides_2_4_1	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		6	0	5	4	0	0	0	YES	-	No
Rhodobacter_sphaeroides_ATCC_17029	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		6	0	5	5	0	0	0	YES	-	No
Rhodopseudomonas_palustris_BisA53	Bacteria	Alphaproteobacteria	2	1	1	2	1	1	3	0	1	1	YES		4	0	4	8	0	0	0	YES	-	No
Rhodopseudomonas_palustris_BisB18	Bacteria	Alphaproteobacteria	2	1	1	2	1	1	3	0	1	1	YES		5	0	5	10	0	0	0	YES	-	No
Rhodopseudomonas_palustris_BisB5	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	6	0	0	0	YES	-	No
Rhodopseudomonas_palustris_CGA009	Bacteria	Alphaproteobacteria	1	1	1	2	1	1	3	0	1	1	YES		4	0	5	24	0	0	0	YES	-	No
Rhodopseudomonas_palustris_HaA2	Bacteria	Alphaproteobacteria	1	1	1	2	1	1	3	0	1	1	YES		3	0	4	17	0	0	0	YES	-	No
Rhodospirillum_rubrum_ATCC_11170	Bacteria	Alphaproteobacteria	2	1	1	1	1	0	2	0	1	1	YES		3	0	3	12	0	0	0	YES	-	No
Rickettsia_bellii_RML369-C	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	0	0	0	0	NO	-	
Rickettsia_conorii	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	0	0	1	YES		0	0	0	0	0	0	0	NO	-	Human
Rickettsia_felis_URRWXCal2	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	0	0	1	YES		0	0	0	0	0	0	0	NO	-	
Rickettsia_prowazekii	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Human
Rickettsia_typhi_wilmington	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Human, Rodent
Roseobacter_denitrificans_OCh_114	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	3	1	0	0	0	YES	-	No
Silicibacter_pomeroyi_DSS-3	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		0	0	0	0	0	0	0	NO	-	No
Silicibacter_TM1040	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	1	0	0	0	YES	-	No
Sinorhizobium_meliloti	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		5	0	5	7	0	0	0	YES	-	No
Sphingopyxis_alaskensis_RB2256	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	1	0	1	YES		1	0	1	36	0	0	0	YES	-	
Wolbachia_endosymbiont_of_Brugia_malayi_TRS	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Brugia malayi
Wolbachia_endosymbiont_of_Drosophila_melanogaster	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	0	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Invertebrate
Zymomonas_mobilis_ZM4	Bacteria	Alphaproteobacteria	1	1	1	1	1	1	2	1	0	1	YES		2	0	2	16	0	0	0	YES	-	No
Aquifex_aeolicus	Bacteria	Aquificae	1	1	1	1	1	0	2	1	0	1	YES		0	0	0	1	0	0	0	NO	-	No
Bacteroides_fragilis_NCTC_9434	Bacteria	Bacteroidetes/Chlorobi	0	0	0	0	0	0	3	0	0	0	NO	NO	2	0	2	84	0	0	0	YES	-	Human
Bacteroides_fragilis_YCH46	Bacteria	Bacteroidetes/Chlorobi	0	0	0	0	0	0	3	0	0	0	NO	NO	3	1	3	83	0	0	0	YES	-	Human
Bacteroides_thetaiotaomicron_VPI-5482	Bacteria	Bacteroidetes/Chlorobi	0	0	0	0	0	0	1	0	0	0	NO	NO	2	0	2	120	0	0	0	YES	-	Mammals
Chlorobium_chlorochromatii_CaD3	Bacteria	Bacteroidetes/Chlorobi	1	1	1	1	1	0	2	3	0	1	YES		1	0	1	1	0	0	0	YES	-	
Chlorobium_phaeobacteroides_DSM_266	Bacteria	Bacteroidetes/Chlorobi	1	1	1	1	1	0	2	4	0	2	YES		2	0	2	3	0	0	0	YES	-	No
Chlorobium_tepidum_TLS	Bacteria	Bacteroidetes/Chlorobi	1	1	1	1	1	0	2	5	0	1	YES		3	0	4	2	0	0	0	YES	-	No
Cytophaga_hutchinsonii_ATCC_33406	Bacteria	Bacteroidetes/Chlorobi	1	1	1	1	1	1	2	2	0	1	YES		1	0	1	13	0	0	0	YES	-	No
Gramella_forsetii_KT0803	Bacteria	Bacteroidetes/Chlorobi	1	1	1	2	1	1	2	1	0	1	YES		1	0	2	38	0	0	0	YES	-	No
Pelodictyon_luteolum_DSM_273	Bacteria	Bacteroidetes/Chlorobi	1	1	1	1	1	0	2	4	0	2	YES		1	0	1	1	0	0	0	YES	-	No

Porphyromonas_gingivalis_W83	Bacteria	Bacteroidetes/Chlorobi	0	0	0	0	0	0	1	1	0	1	NO	NO	2	0	2	9	0	0	0	YES	-	Human
Salinibacter_ruber_DSM_13855	Bacteria	Bacteroidetes/Chlorobi	1	1	1	1	1	1	2	4	0	2	YES		2	0	5	25	0	0	0	YES	-	No
Acidovorax_avenae_citrulli_AAC00-1	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	13	0	0	0	YES	-	Fruit
Acidovorax_JS42	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	1	5	0	0	0	YES	-	No
Azoarcus_BH72	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	2	15	0	0	0	YES	-	
Azoarcus_sp_EbN1	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	5	0	0	0	YES	-	No
Bordetella_bronchiseptica	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	13	0	0	0	YES	-	Animal
Bordetella_parapertussis	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	10	0	0	0	YES	-	Human, Sheep
Bordetella_pertussis	Bacteria	Betaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		2	0	2	10	0	0	0	YES	-	Human
Burkholderia_383	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		5	0	5	22	0	0	0	YES	-	Human
Burkholderia_cenocepacia_AU_1054	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	4	16	0	0	0	YES	-	
Burkholderia_cenocepacia_HI2424	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	4	19	0	0	0	YES	-	
Burkholderia_cepacia_AMMD	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	4	10	0	0	0	YES	-	
Burkholderia_mallei_ATCC_23344	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	3	7	0	0	0	YES	-	Human, Animal
Burkholderia_mallei_NCTC_10229	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	7	0	0	0	YES	-	Human, Animal
Burkholderia_mallei_NCTC_10247	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	7	0	0	0	YES	-	Human, Animal
Burkholderia_mallei_SAVP1	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	6	0	0	0	YES	-	Human, Animal
Burkholderia_pseudomallei_1106a	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	7	0	0	0	YES	-	Animal
Burkholderia_pseudomallei_1710b	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	7	0	0	0	YES	-	Animal
Burkholderia_pseudomallei_668	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	6	0	0	0	YES	-	Animal
Burkholderia_pseudomallei_K96243	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	7	0	0	0	YES	-	Animal
Burkholderia_thailandensis_E264	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	6	0	0	0	YES	-	
Burkholderia_xenovorans_LB400	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	3	9	0	0	0	YES	-	Human, Plant
Chromobacterium_violaceum	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		6	0	3	10	0	0	0	YES	-	Human
Dechloromonas_aromatica_RCB	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	2	11	0	0	0	YES	-	No
Methylibium_petroleiphilum_PM1	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	3	13	0	0	0	YES	-	
Methylobacillus_flagellatus_KT	Bacteria	Betaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		1	0	2	19	0	0	0	YES	-	
Neisseria_gonorrhoeae_FA_1090	Bacteria	Betaproteobacteria	1	1	1	1	1	0	2	0	1	1	YES		1	0	1	6	0	0	0	YES	-	Human
Neisseria_meningitidis_FAM18	Bacteria	Betaproteobacteria	1	1	1	1	1	0	2	0	1	1	YES		1	0	1	9	0	0	0	YES	-	Human
Neisseria_meningitidis_MC58	Bacteria	Betaproteobacteria	1	1	1	1	1	0	2	0	1	1	YES		1	0	1	7	0	0	0	YES	-	Human
Neisseria_meningitidis_Z2491	Bacteria	Betaproteobacteria	1	1	1	1	1	0	2	0	1	1	YES		1	0	1	8	0	0	0	YES	-	Human
Nitrosomonas_europaea	Bacteria	Betaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	25	0	0	0	NO	-	No
Nitrosomonas_eutropha_C71	Bacteria	Betaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	10	0	0	0	NO	-	
Nitrosospira_multiformis_ATCC_25196	Bacteria	Betaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		1	0	1	9	0	0	0	YES	-	
Polaromonas_JS666	Bacteria	Betaproteobacteria	1	1	1	0	1	1	2	0	1	1	YES		4	0	4	3	0	0	0	YES	-	No
Polaromonas_naphthalenivorans_CJ2	Bacteria	Betaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		1	0	2	3	0	0	0	YES	-	
Ralstonia_eutropha_H16	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	2	11	0	0	0	YES	-	
Ralstonia_eutropha_JMP134	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	2	7	0	0	0	YES	-	
Ralstonia_metallidurans_CH34	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	3	11	0	0	0	YES	-	No
Ralstonia_solanacearum	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	13	0	0	0	YES	-	Plant
Rhodoferax_ferrireducens_T118	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	2	4	0	0	0	YES	-	No
Thiobacillus_denitrificans_ATCC_25259	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	1	5	0	0	0	YES	-	No
Verminephrobacter_eiseniae_EF01-2	Bacteria	Betaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	4	0	0	0	YES	-	No
Chlamydia_muridarum	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Chlamydia_trachomatis	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human
Chlamydia_trachomatis_A_HAR-13	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human
Chlamydophila_abortus_S26_3	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Chlamydophila_caviae	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Chlamydophila_felis_Fe_C-56	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Chlamydophila_pneumoniae_AR39	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human
Chlamydophila_pneumoniae_CWL029	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human
Chlamydophila_pneumoniae_J138	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human
Chlamydophila_pneumoniae_TW_183	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	0	2	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human
Parachlamydia_sp_UWE25	Bacteria	Chlamydiae/Verrucomicrobia	1	1	0 ; truncated hemC	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Animal
Dehalococcoides_CBDB1	Bacteria	Chloroflexi	0	0	0	0	0	0	1	0	0	0	NO	NO	1	0	2	0	0	0	0	NO	+	No
Dehalococcoides_ethenogenes_195	Bacteria	Chloroflexi	0	0	0	0	0	0	1	0	0	0	NO	NO	3	0	4	0	0	0	0	NO	+	No
Anabaena_variabilis_ATCC_29413	Bacteria	Cyanobacteria	3	1	1	2	1	2	3	4	0	2	YES		4	0	6	8	0	0	0	YES	-	No
Cyanobacteria_bacterium_Yellowstone_A-Prime	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	5	0	1	YES		1	0	2	1	0	0	0	YES	-	No
Cyanobacteria_bacterium_Yellowstone_B-Prime	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	5	0	1	YES		1	0	1	1	0	0	0	YES	-	No
Gloeobacter_violaceus	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	7	0	2	YES		1	0	2	25	0	0	0	YES	-	
Nostoc_sp	Bacteria	Cyanobacteria	2	1	1	1	1	2	2	7	0	2	YES		5	0	22	20	0	0	0	YES	-	No
Prochlorococcus_marinus_AS9601	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	2	0	1	YES		0	0	0	0	0	0	0	NO	-	No
Prochlorococcus_marinus_CCMP1375	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	
Prochlorococcus_marinus_MED4	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	2	0	1	YES		0	0	0	0	0	0	0	NO	-	
Prochlorococcus_marinus_MIT_9301	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	2	0	1	YES		0	0	0	0	0	0	0	NO	-	
Prochlorococcus_marinus_MIT_9303	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	3	0	1	YES		0	0	0	0	0	0	0	NO	-	No
Prochlorococcus_marinus_MIT_9312	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	2	0	1	YES		0	0	0	0	0	0	0	NO	-	
Prochlorococcus_marinus_MIT_9515	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	2	0	1	YES		0	0	0	0	0	0	0	NO	-	No
Prochlorococcus_marinus_MIT9313	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	2	0	1	YES		0	0	0	0	0	0	0	NO	-	
Prochlorococcus_marinus_NATL1A	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	
Prochlorococcus_marinus_NATL2A	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	
Synechococcus_CC9311	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	4	0	1	YES		0	0	0	0	0	0	0	NO	-	No

Synechococcus_CC9605	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	5	0	1	YES		0	0	0	0	0	0	0	NO	-	
Synechococcus_CC9902	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	3	0	1	YES		0	0	0	0	0	0	0	NO	-	
Synechococcus_elongatus_PCC_6301	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	3	0	1	YES		0	0	0	0	0	0	0	NO	-	No
Synechococcus_elongatus_PCC_7942	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	3	0	1	YES		0	0	0	0	0	0	0	NO	-	
Synechococcus_sp_WH8102	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	3	0	1	YES		0	0	0	0	0	0	0	NO	-	No
Synechocystis_PCC6803	Bacteria	Cyanobacteria	1	1	1	1	1	1	2	4	0	1	YES		2	0	4	4	0	0	0	YES	-	No
Thermosynechococcus_elongatus	Bacteria	Cyanobacteria	1	1	1	1	1	1	2	3	0	1	YES		0	0	1	0	0	0	0	NO	-	No
Trichodesmium_erythraeum_IMS101	Bacteria	Cyanobacteria	1	1	1	1	1	1	1	5	0	2	YES		0	0	1	0	0	0	0	NO	-	No
Deinococcus_geothermalis_DSM_11300	Bacteria	Deinococcus-Thermus	1	1	1	2	0	0	1	0	0	0	NO	YES	4	0	8	0	0	0	0	NO	+	
Deinococcus_radiodurans	Bacteria	Deinococcus-Thermus	1	1	1	1	1	0	1	2	0	1	YES		4	0	6	0	0	0	0	NO	+	No
Thermus_thermophilus_HB27	Bacteria	Deinococcus-Thermus	1	1	1	2	1	0	1	1	0	1	YES		2	0	4	0	0	0	0	NO	-	
Thermus_thermophilus_HB8	Bacteria	Deinococcus-Thermus	1	1	1	2	1	0	1	1	0	1	YES		2	0	3	0	0	0	0	NO	-	
Anaeromyxobacter_dehalogenans_2CP-C	Bacteria	Deltaproteobacteria	1	1	1	1	1	1	2	2	0	1	YES		1	0	2	1	0	0	0	YES	-	No
Bdellovibrio_bacteriovorus	Bacteria	Deltaproteobacteria	1	1	0 ; B_83352	0	1	0	2	2	0	1	YES		0	0	1	2	0	0	0	NO	-	Gram-negative bacteria
Desulfotalea_psychrophila_LSv54	Bacteria	Deltaproteobacteria	1	1	1	1	1	0	1	2	0	1	YES		2	0	2	1	0	0	0	YES	-	No
Desulfovibrio_desulfuricans_G20	Bacteria	Deltaproteobacteria	1	1	1	1	0	0	1	0	0	0	NO	YES	2	0	2	2	0	0	0	YES	-	No
Desulfovibrio_vulgaris_DP4	Bacteria	Deltaproteobacteria	1	1	1	1	0	0	1	0	0	0	NO	YES	1	0	1	1	0	0	0	YES	-	No
Desulfovibrio_vulgaris_Hildenborough	Bacteria	Deltaproteobacteria	1	1	1	1	0	0	1	0	0	0	NO	YES	1	0	1	2	0	0	0	YES	-	No
Geobacter_metallireducens_GS-15	Bacteria	Deltaproteobacteria	1	1	1	1	2	0	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	No
Geobacter_sulfurreducens	Bacteria	Deltaproteobacteria	1	1	1	1	2	0	1	2	0	1	YES		0	0	0	2	0	0	0	NO	-	No
Lawsonia_intracellularis_PHE_MN1-00	Bacteria	Deltaproteobacteria	1	1	1	1	0	0	1	0	0	0	NO	YES	0	0	0	0	0	0	0	NO	-	Animal
Myxococcus_xanthus_DK_1622	Bacteria	Deltaproteobacteria	1	1	1	1	1	1	3	4	0	1	YES		3	0	3	11	0	0	0	YES	-	No
Pelobacter_carbinolicus	Bacteria	Deltaproteobacteria	1	1	1	1	1	0	1	1	0	1	YES		3	0	4	5	0	0	0	YES	-	No
Pelobacter_propionicus_DSM_2379	Bacteria	Deltaproteobacteria	1	1	1	1	2	0	2	3	0	1	YES		4	0	5	6	0	0	0	YES	-	No
Syntrophobacter_fumaroxidans_MPOB	Bacteria	Deltaproteobacteria	1	1	1	1	0	0	1	0	0	0	NO	YES	2	0	6	1	0	0	0	YES	-	No
Syntrophus_aciditrophicus_SB	Bacteria	Deltaproteobacteria	1	1	1	1	0	0	2	0	0	0	NO	YES	2	0	4	2	0	0	0	YES	-	
Campylobacter_fetus_82-40	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	3	0	0	1	YES		3	0	3	3	0	0	0	YES	-	Human, Animal
Campylobacter_jejuni	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	3	0	0	1	YES		1	0	1	2	0	0	0	YES	-	Human
Campylobacter_jejuni_81-176	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	3	0	0	1	YES		1	0	1	2	0	0	0	YES	-	Human
Campylobacter_jejuni_RM1221	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	3	0	0	1	YES		1	0	1	2	0	0	0	YES	-	Human
Helicobacter_acinonychis_Sheeba	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	2	0	0	1	YES		1	0	1	3	0	0	0	YES	-	feline
Helicobacter_hepaticus	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	2	0	0	1	YES		0	0	0	2	0	0	0	NO	-	Animal
Helicobacter_pylori_26695	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	2	0	0	1	YES		1	0	2	2	0	0	0	YES	-	Human
Helicobacter_pylori_HPAG1	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	2	0	0	1	YES		1	0	2	3	0	0	0	YES	-	Human
Helicobacter_pylori_J99	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	2	0	0	1	YES		1	0	2	3	0	0	0	YES	-	Human
Thiomicrospira_denitrificans_ATCC_33889	Bacteria	Epsilonproteobacteria	1	1	1	2	1	0	2	0	0	1	YES		2	0	2	6	0	0	0	YES	-	No
Wolinella_succinogenes	Bacteria	Epsilonproteobacteria	1	1	1	1	1	0	3	0	0	1	YES		1	0	1	7	0	0	0	YES	-	No
Aster_yellows_witches-broom_phytoplasma_AYWB	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO		Plants
Bacillus_anthracis_Ames	Bacteria	Firmicutes	1	1	1	2	1	0	1	3	0	2	YES		13	4	9	0	0	0	0	YES	+	Animal
Bacillus_anthracis_Ames_0581	Bacteria	Firmicutes	1	1	1	2	1	0	1	3	0	2	YES		13	5	9	0	0	0	0	YES	+	Animal
Bacillus_anthracis_str_Sterne	Bacteria	Firmicutes	1	1	1	2	1	0	1	3	0	2	YES		15	5	10	0	0	0	0	YES	+	Animal
Bacillus_cereus_ATCC_10987	Bacteria	Firmicutes	1	1	1	1	1	0	1	3	0	2	YES		12	4	8	0	0	0	0	YES	+	Human
Bacillus_cereus_ATCC14579	Bacteria	Firmicutes	1	1	1	1	1	0	1	3	0	2	YES		11	4	10	0	0	0	0	YES	+	Human
Bacillus_cereus_ZK	Bacteria	Firmicutes	1	1	1	1	1	0	1	3	0	2	YES		12	4	9	0	0	0	0	YES	+	Human, Animal
Bacillus_clausii_KSM-K16	Bacteria	Firmicutes	1	1	1	3	1	0	2	1	0	1	YES		8	2	8	0	0	0	0	YES	+	
Bacillus_halodurans	Bacteria	Firmicutes	1	1	1	3	1	0	1	1	0	1	YES		8	2	8	0	0	0	0	YES	+	No
Bacillus_licheniformis_ATCC_14580	Bacteria	Firmicutes	2	1	1	2	1	0	1	1	0	2	YES		9	0	7	0	0	0	0	NO	+	Human
Bacillus_licheniformis_DSM_13	Bacteria	Firmicutes	2	1	1	2	1	0	1	2	0	2	YES		9	0	7	0	0	0	0	NO	+	Human
Bacillus_subtilis	Bacteria	Firmicutes	1	1	1	2	1	0	1	2	0	1	YES		9	0	7	0	0	0	0	NO	+	No
Bacillus_thuringiensis_Al_Hakam	Bacteria	Firmicutes	1	1	1	1	1	0	1	4	0	2	YES		14	4	10	0	0	0	0	YES	+	Insect
Bacillus_thuringiensis_konkukian	Bacteria	Firmicutes	1	1	1	1	1	0	1	3	0	2	YES		13	4	10	0	0	0	0	YES	+	Insect
Carboxydothermus_hydrogenoformans_Z-2901	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		2	0	2	0	0	0	0	NO	+	No
Clostridium_acetobutylicum	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	4	0	3	0	0	0	0	NO	+	No
Clostridium_difficile_630	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	3	0	2	0	0	0	0	NO	+	Human
Clostridium_novyi_NT	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	2	3	2	0	0	0	0	YES	+	
Clostridium_perfringens	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	7	2	5	0	0	0	0	YES	+	Human, Animal
Clostridium_perfringens_ATCC_13124	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	5	2	4	0	0	0	0	YES	+	Human, Animal
Clostridium_perfringens_SM101	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	6	2	4	0	0	0	0	YES	+	Human, Animal
Clostridium_tetani_E88	Bacteria	Firmicutes	1	1	1	1	0	0	3	0	0	0	NO	YES	5	2	5	0	0	0	0	YES	+	Human
Clostridium_thermocellum_ATCC_27405	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	2	1	1	0	0	0	0	YES	+	No
Desulfitobacterium_hafniense_Y51	Bacteria	Firmicutes	1	1	1	1	0	0	2	0	0	0	NO	YES	12	0	12	0	0	0	0	NO	-	
Enterococcus_faecalis_V583	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	3	0	2	0	0	0	0	NO	+	Human
Geobacillus_kaustophilus_HTA426	Bacteria	Firmicutes	1	1	1	2	0	0	1	1	0	1	YES		7	0	5	0	0	0	0	NO	+	
Lactobacillus_acidophilus_NCFM	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	No
Lactobacillus_brevis_ATCC_367	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	1	3	2	0	0	0	0	YES	+	No
Lactobacillus_casei_ATCC_334	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	0	0	0	0	0	0	0	NO	+	No
Lactobacillus_delbrueckii_bulgaricus	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	No
Lactobacillus_delbrueckii_bulgaricus_ATCC_BAA-365	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	No
Lactobacillus_gasseri_ATCC_33323	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	No
Lactobacillus_johnsonii_NCC_533	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	No
Lactobacillus_plantarum	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	1	NO	NO	3	0	2	0	0	0	0	NO	+	No
Lactobacillus_sakei_23K	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	2	0	1	0	0	0	0	NO	+	

Lactobacillus_salivarius_UCC118	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	
Lactococcus_lactis	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	2	0	1	0	0	0	0	NO	+	No
Lactococcus_lactis_cremoris_MG1363	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	2	0	1	0	0	0	0	NO	+	No
Lactococcus_lactis_cremoris_SK11	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	2	0	1	0	0	0	0	NO	+	No
Leuconostoc_mesenteroides_ATCC_8293	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	1	NO	NO	0	0	0	0	0	0	0	NO	+	No
Listeria_innocua	Bacteria	Firmicutes	1	1	1	2	1	0	1	1	0	1	YES		4	2	5	0	0	0	0	YES	+	Human
Listeria_monocytogenes	Bacteria	Firmicutes	1	1	1	2	1	0	1	1	0	1	YES		4	2	5	0	0	0	0	YES	+	Human
Listeria_monocytogenes_4b_F2365	Bacteria	Firmicutes	1	1	1	2	1	0	1	1	0	1	YES		4	2	5	0	0	0	0	YES	+	Human
Listeria_welshimeri_serovar_6b_SLCC5334	Bacteria	Firmicutes	1	1	1	2	1	0	1	1	0	1	YES		4	2	5	0	0	0	0	YES	+	No
Mesoplasma_florum_L1	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Mammal, Insect, Plant
Moorella_thermoacetica_ATCC_39073	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	2	0	3	0	0	0	0	NO	+	
Mycoplasma_capricolum_ATCC_27343	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Ruminant
Mycoplasma_gallisepticum	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Birds
Mycoplasma_genitalium	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Human, Animal, Insect
Mycoplasma_hyopneumoniae_232	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Swine
Mycoplasma_hyopneumoniae_7448	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Swine
Mycoplasma_hyopneumoniae_J	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Swine
Mycoplasma_mobile_163K	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Fish
Mycoplasma_mycoides	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Cattle
Mycoplasma_penetrans	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	1	0	0	0	0	0	0	NO	-	Human
Mycoplasma_pneumoniae	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Human
Mycoplasma_pulmonis	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Rodent
Mycoplasma_synoviae_53	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Chicken
Oceanobacillus_iheyensis	Bacteria	Firmicutes	1	1	1	2	1	0	1	2	0	1	YES		6	0	4	0	0	0	0	NO	+	No
Oenococcus_oeni_PSU-1	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	No
Onion_yellows_phytoplasma	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO		Onion
Pediococcus_pentosaceus_ATCC_25745	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	+	No
Staphylococcus_aureus_aureus_MRSA252	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		6	3	5	0	0	0	0	YES	+	Human
Staphylococcus_aureus_aureus_MSSA476	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		6	4	7	0	0	0	0	YES	+	Human
Staphylococcus_aureus_COL	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		5	4	7	0	0	0	0	YES	+	Human
Staphylococcus_aureus_Mu50	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		6	4	7	0	0	0	0	YES	+	Human
Staphylococcus_aureus_MW2	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		6	4	7	0	0	0	0	YES	+	Human
Staphylococcus_aureus_N315	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		5	4	6	0	0	0	0	YES	+	Human
Staphylococcus_aureus_NCTC_8325	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		6	4	6	0	0	0	0	YES	+	Human
Staphylococcus_aureus_RF122	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		6	5	6	0	0	0	0	YES	+	Human, Animal
Staphylococcus_aureus_USA300	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		6	4	6	0	0	0	0	YES	+	Human
Staphylococcus_epidermidis_ATCC_12228	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		0	0	2	0	0	0	0	NO	+	Human
Staphylococcus_epidermidis_RP62A	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		0	0	2	0	0	0	0	NO	+	Human
Staphylococcus_haemolyticus	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		0	0	3	0	0	0	0	NO	+	Human
Staphylococcus_saprophyticus	Bacteria	Firmicutes	1	1	1	1	1	0	1	1	0	1	YES		0	0	9	0	0	0	0	NO	+	Human
Streptococcus_agalactiae_2603	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	0	0	1	0	0	0	0	NO	+	Human
Streptococcus_agalactiae_A909	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	0	0	1	0	0	0	0	NO	+	Human
Streptococcus_agalactiae_NEM316	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	0	0	1	0	0	0	0	NO	+	Human
Streptococcus_mutans	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	0	0	0	0	0	0	0	NO	+	Human
Streptococcus_pneumoniae_D39	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	4	0	1	0	0	0	0	NO	+	Human
Streptococcus_pneumoniae_R6	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	4	0	1	0	0	0	0	NO	+	Human
Streptococcus_pneumoniae_TIGR4	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	4	0	1	0	0	0	0	NO	+	Human
Streptococcus_pyogenes_M1_GAS	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS10270	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS10394	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS10750	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS2096	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS315	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS5005	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS6180	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS8232	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_MGAS9429	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_pyogenes_SSI-1	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	3	1	2	0	0	0	1	YES	+	Human
Streptococcus_sanguinis_SK36	Bacteria	Firmicutes	1	1	1	1	1	0	1	0	0	1	YES		0	0	2	0	0	0	0	NO	+	Human
Streptococcus_thermophilus_CNRZ1066	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	0	0	0	0	0	0	0	NO	+	No
Streptococcus_thermophilus_LMD-9	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	0	0	0	0	0	0	0	NO	+	No
Streptococcus_thermophilus_LMG_18311	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	1	NO	NO	0	0	0	0	0	0	0	NO	+	No
Syntrophomonas_wolfei_Goettingen	Bacteria	Firmicutes	1	1	1	1	0	0	1	0	0	0	NO	YES	0	0	9	0	0	0	0	NO	-	No
Thermoanaerobacter_tengcongensis	Bacteria	Firmicutes	0	0	0	0	0	0	1	0	0	0	NO	NO	2	0	2	0	0	0	0	NO	-	No
Ureaplasma_urealyticum	Bacteria	Firmicutes	0	0	0	0	0	0	0	0	0	0	NO	NO	10	0	0	0	0	0	0	NO		Animal
Fusobacterium_nucleatum	Bacteria	Fusobacteria	1	1	1	1	0	0	2	0	0	0	NO	YES	5	0	5	5	0	0	0	YES	-	Human, Animal
Acinetobacter_baumannii_ATCC_17978	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	1	0	1	YES		1	0	1	13	0	0	0	YES	-	Yes
Acinetobacter_sp_ADP1	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	2	0	1	YES		1	0	1	22	0	0	0	YES	-	Human
Actinobacillus_pleuropneumoniae_L20	Bacteria	Gammaproteobacteria	1	1	1	1	1	0	3	0	1	1	YES		2	0	3	6	0	0	0	YES	-	Porcine
Aeromonas_hydrophila_ATCC_7966	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		5	0	4	7	0	0	0	YES	-	Animal
Alcanivorax_borkumensis_SK2	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	1	6	0	0	0	YES	-	No
Alkalilimnicola_ehrlichei_MLHE-1	Bacteria	Gammaproteobacteria	1	1	1	2	1	1	3	0	1	1	YES		1	0	1	2	0	0	0	YES	-	

Baumannia_cicadellinicola_Homalodisca_coagulata	Bacteria	Gammaproteobacteria	0	1	1	1	0	0	1	0	0	0	NO	YES	0	0	0	0	0	0	0	NO	-	
Buchnera_aphidicola	Bacteria	Gammaproteobacteria	0	0	0	0	0	0	1	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	No
Buchnera_aphidicola_Cc_Cinara_cedri	Bacteria	Gammaproteobacteria	0	0	0	0	0	0	1	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	No
Buchnera_aphidicola_Sg	Bacteria	Gammaproteobacteria	0	0	0	0	0	0	1	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	No
Buchnera_sp	Bacteria	Gammaproteobacteria	0	1	1	0	0	0	1	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	No
Candidatus_Blochmannia_floridanus	Bacteria	Gammaproteobacteria	0	1	1	1	0	0	0	0	0	0	NO	YES	0	0	0	1	0	0	0	NO	-	No
Candidatus_Blochmannia_pennsylvanicus_BPEN	Bacteria	Gammaproteobacteria	0	1	1	1	0	0	0	0	0	0	NO	YES	0	0	0	1	0	0	0	NO	-	
Candidatus_Carsonella_ruddii_PV	Bacteria	Gammaproteobacteria	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	
Candidatus_Ruthia_magnifica_Cm_Calyptogena_magnifica	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	0	0	0	1	YES		0	0	0	0	0	0	0	NO	-	
Chromohalobacter_salexigens_DSM_3043	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		6	0	5	5	0	0	0	YES	-	No
Colwellia_psychrerythraea_34H	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	1	42	0	0	0	YES	-	No
Coxiella_burnetii	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		1	0	0	0	0	0	0	NO	-	Animal
Erwinia_carotovora_atroseptica_SCRI1043	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		10	0	11	13	1	0	0	YES	-	Plant
Escherichia_coli_536	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		9	0	7	9	0	0	0	YES	-	Human
Escherichia_coli_APEC_O1	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		10	0	6	7	0	0	0	YES	-	Avian
Escherichia_coli_CFT073	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		10	0	8	11	0	0	0	YES	-	Human
Escherichia_coli_K12	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		7	0	3	4	0	0	0	YES	-	No
Escherichia_coli_O157H7	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		8	0	6	7	0	0	0	YES	-	Human
Escherichia_coli_O157H7_EDL933	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		8	0	6	7	0	0	0	YES	-	Human
Escherichia_coli_UTI89	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		10	0	7	12	0	0	0	YES	-	Human
Escherichia_coli_W3110	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		7	0	3	4	0	0	0	YES	-	
Francisella_tularensis_FSC_198	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animals
Francisella_tularensis_holarctica	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Francisella_tularensis_holarctica_OSU18	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Francisella_tularensis_novicida_U112	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	
Francisella_tularensis_tularensis	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	1	0	1	YES		0	0	0	0	0	0	0	NO	-	Human, Animal
Haemophilus_ducreyi_35000HP	Bacteria	Gammaproteobacteria	0	0	0	0	0	0	2	0	1	0	NO	NO	0	0	0	3	0	0	0	NO	-	Human
Haemophilus_influenzae	Bacteria	Gammaproteobacteria	0	0	0	0	0	0	1	0	1	1	NO	NO	1	0	1	8	0	0	0	YES	-	Human
Haemophilus_influenzae_86_028NP	Bacteria	Gammaproteobacteria	0	0	0	0	0	0	1	0	1	1	NO	NO	0	0	0	5	0	0	0	NO	-	Human
Haemophilus_somnus_129PT	Bacteria	Gammaproteobacteria	1	1	1	1	1	0	2	0	1	1	YES		0	0	0	5	0	0	0	NO	-	Human, Animal
Hahella_chejuensis_KCTC_2396	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	2	0	1	1	YES		9	0	6	6	0	0	0	YES	-	No
Halorhodospira_halophila_SL1	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		2	0	2	7	0	0	0	YES	-	
Idiomarina_loihiensis_L2TR	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	1	28	0	0	0	YES	-	
Legionella_pneumophila_Lens	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Animal
Legionella_pneumophila_Paris	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Animal
Legionella_pneumophila_Philadelphia_1	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	0	0	0	0	NO	-	Animal
Mannheimia_succiniciproducens_MBEL55E	Bacteria	Gammaproteobacteria	1	1	1	1	1	0	2	0	1	1	YES		2	0	2	3	0	0	0	YES	-	No
Marinobacter_aquaeolei_VT8	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		5	0	4	3	0	0	0	YES	-	
Methylococcus_capsulatus_Bath	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		0	0	0	4	0	0	0	NO	-	No
Nitrosococcus_oceani_ATCC_19707	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		1	0	1	8	0	0	0	YES	-	
Pasteurella_multocida	Bacteria	Gammaproteobacteria	1	1	1	1	1	0	2	0	1	1	YES		5	0	4	11	0	0	0	YES	-	Human, Animal
Photobacterium_profundum_SS9	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		6	0	4	5	0	0	0	YES	-	No
Photorhabdus_luminescens	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		7	0	5	8	0	0	0	YES	-	Insect
Pseudoalteromonas_atlantica_T6c	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	73	0	0	0	YES	-	Shellfish
Pseudoalteromonas_haloplanktis_TAC125	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	2	38	0	0	0	YES	-	No
Pseudomonas_aeruginosa	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	4	24	1	0	0	YES	-	Human
Pseudomonas_aeruginosa_UCBPP-PA14	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		6	0	5	26	1	0	0	YES	-	Human, Animal, Plant
Pseudomonas_entomophila_L48	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	3	23	1	0	0	YES	-	Insects
Pseudomonas_fluorescens_Pf-5	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	2	0	1	1	YES		5	0	6	30	1	0	0	YES	-	No
Pseudomonas_fluorescens_PfO-1	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	2	0	1	1	YES		2	0	3	19	0	0	0	YES	-	No
Pseudomonas_putida_KT2440	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	3	0	1	1	YES		3	0	3	20	0	0	0	YES	-	No
Pseudomonas_syringae_phaseolicola_1448A	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		6	0	4	16	0	0	0	YES	-	Plant
Pseudomonas_syringae_pv_B728a	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		6	0	4	15	0	0	0	YES	-	Plant
Pseudomonas_syringae_tomato_DC3000	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	3	20	0	0	0	YES	-	Plant
Psychrobacter_arcticum_273-4	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	3	0	1	YES		0	0	0	1	0	0	0	NO	-	No
Psychrobacter_cryohalolentis_K5	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	5	0	1	YES		2	0	1	3	0	0	0	YES	-	
Psychromonas_ingrahamii_37	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	1	0	0	0	0	NO	-	No
Saccharophagus_degradans_2-40	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	1	43	0	0	0	YES	-	No
Salmonella_enterica_Choleraesuis	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		5	0	4	3	0	0	0	YES	-	Human, Animal
Salmonella_enterica_Paratypi_ATCC_9150	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		5	0	3	2	0	0	0	YES	-	Human
Salmonella_typhi	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	4	2	0	0	0	YES	-	Human
Salmonella_typhi_Ty2	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		4	0	4	2	0	0	0	YES	-	Human
Salmonella_typhimurium_LT2	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		5	0	4	3	0	0	0	YES	-	Human
Shewanella_amazonensis_SB2B	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		2	0	1	25	0	0	0	YES	-	
Shewanella_ANA-3	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	3	0	1	2	YES		2	0	2	29	0	0	0	YES	-	
Shewanella_baltica_OS155	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	3	0	1	2	YES		2	0	2	25	0	0	0	YES	-	
Shewanella_denitrificans_OS217	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	2	0	1	1	YES		2	0	2	19	0	0	0	YES	-	
Shewanella_frigidimarina_NCIMB_400	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		2	0	1	25	0	0	0	YES	-	
Shewanella_loihica_PV-4	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	2	YES		3	0	1	20	0	0	0	YES	-	
Shewanella_MR-4	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	3	0	1	2	YES		2	0	1	30	0	0	0	YES	-	
Shewanella_MR-7	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	3	0	1	2	YES		2	0	1	29	0	0	0	YES	-	
Shewanella_oneidensis	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	3	0	1	2	YES		4	0	2	20	0	0	0	YES	-	Human

Shewanella_W3-18-1	Bacteria	Gammaproteobacteria	2	1	1	1	1	1	3	0	1	2	YES		3	0	2	21	0	0	0	YES	-	Human
Shigella_boydii_Sb227	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	3	5	0	0	0	YES	-	Human
Shigella_dysenteriae	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	4	4	0	0	0	YES	-	Human
Shigella_flexneri_2a	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	4	3	0	0	0	YES	-	Human
Shigella_flexneri_2a_2457T	Bacteria	Gammaproteobacteria	1	1	1	1	1	0	2	0	1	1	YES		4	0	4	3	0	0	0	YES	-	Human
Shigella_flexneri_5_8401	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		3	0	4	3	0	0	0	YES	-	Human
Shigella_sonnei_Ss046	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		2	0	3	5	0	0	0	YES	-	Human
Sodalis_glossinidius_morsitans	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	0	1	YES		5	0	2	0 ; no Plug	0	0	0	YES	-	No
Thiomicrospira_crunogena_XCL-2	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		1	0	1	4	0	0	0	NO	-	No
Vibrio_cholerae	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		5	0	4	5	0	0	0	YES	-	Human, Primate
Vibrio_fischeri_ES114	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		4	0	4	7	0	0	0	YES	-	No
Vibrio_parahaemolyticus	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		5	0	4	8	0	0	0	YES	-	Human
Vibrio_vulnificus_CMCP6	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		3	0	3	6	0	0	0	YES	-	Human
Vibrio_vulnificus_YJ016	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		3	0	3	6	0	0	0	YES	-	Human
Wigglesworthia_brevipalpis	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	0	0	0	1	YES		0	0	0	0	0	0	0	NO	-	No
Xanthomonas_campestris	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		0	0	0	61	0	0	0	NO	-	Plant
Xanthomonas_campestris_8004	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		0	0	0	62	0	0	0	NO	-	Plant
Xanthomonas_campestris_vesicatoria_85-10	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	47	0	0	0	NO	-	Plant
Xanthomonas_citri	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	62	0	0	0	NO	-	Plant
Xanthomonas_oryzae_KACC10331	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	30	0	0	0	NO	-	Rice
Xanthomonas_oryzae_MAFF_311018	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	33	0	0	0	NO	-	Rice
Xylella_fastidiosa	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	8	0	0	0	NO	-	Plant
Xylella_fastidiosa_Temecula1	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	1	0	1	1	YES		0	0	0	8	0	0	0	NO	-	Plant
Yersinia_enterocolitica_8081	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		6	0	8	9	4	0	0	YES	-	Human, Animal
Yersinia_pestis_Antiqua	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		7	0	8	8	1	0	0	YES	-	Yes
Yersinia_pestis_biovar_Mediaevails	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	2	0	1	1	YES		7	0	8	7	1	0	0	YES	-	Human, Rodent
Yersinia_pestis_CO92	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		7	0	8	8	1	0	0	YES	-	Human, Rodent
Yersinia_pestis_KIM	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		8	0	8	8	1	0	0	YES	-	Human, Rodent
Yersinia_pestis_Nepal516	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		7	0	8	7	1	0	0	YES	-	Yes
Yersinia_pseudotuberculosis_IP32953	Bacteria	Gammaproteobacteria	1	1	1	1	1	1	3	0	1	1	YES		7	0	8	8	1	0	0	YES	-	Human, Animal
Herminiimonas_arsenicoxydans	Bacteria	Other Bacteria	1	1	1	1	1	1	3	0	1	1	YES		1	0	1	6	0	0	0	YES		No
Magnetococcus_MC-1	Bacteria	Other Bacteria	1	1	1	1	1	0	2	2	0	1	YES		0	0	0	1	0	0	0	NO	-	No
Pirellula_sp	Bacteria	Planctomycetes	1	1	1	1	1	0	2	4	0	1	YES		1	0	1	0	0	0	0	NO	-	No
Borrelia_afzelii_PKo	Bacteria	Spirochaetes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Human
Borrelia_burgdorferi	Bacteria	Spirochaetes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Human
Borrelia_garinii_PBi	Bacteria	Spirochaetes	0	0	0	0	0	0	0	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Human
Leptospira_borgpetersenii_serovar_Hardjo-bovis_JB197	Bacteria	Spirochaetes	1	1	0 ; B_83352	0	1	0	2	4	0	1	YES		0	0	1	4	0	0	0	NO	-	Cattle, Human
Leptospira_borgpetersenii_serovar_Hardjo-bovis_L550	Bacteria	Spirochaetes	1	1	0 ; B_83352	0	1	0	2	4	0	1	YES		0	0	1	4	0	0	0	NO	-	Cattle, Human
Leptospira_interrogans_serovar_Copenhageni	Bacteria	Spirochaetes	1	1	0 ; B_83352	0	1	0	2	3	0	1	YES		0	0	1	8	0	0	0	NO	-	Animal
Leptospira_interrogans_serovar_Lai	Bacteria	Spirochaetes	1	1	0 ; B_83352	0	1	0	2	3	0	1	YES		0	0	1	9	0	0	0	NO	-	Animal
Treponema_denticola_ATCC_35405	Bacteria	Spirochaetes	0	0	0	0	0	0	2	0	0	0	NO	NO	5	0	5	0	0	0	0	NO	-	Human
Treponema_pallidum	Bacteria	Spirochaetes	0	0	0	0	0	0	1	0	0	0	NO	NO	0	0	0	0	0	0	0	NO	-	Human
Thermotoga_maritima	Bacteria	Thermotogae	0	0	0	0	0	0	1	0	0	0	NO	NO	2	0	2	0	0	0	0	NO	-	No

Table S3. Comparison between literature data and predictions made in the present work on organisms known to be capable of performing heme biosynthesis and/or uptake. All predictions were correct with the exception of that for *Sodalis glossidinius*. At variance with the experimental data, we assigned this organism as being unable to carry out heme uptake because no Plug domain was detected in its genome.

Biosynthesis

Organism	Reference	Experiment	Prediction
Bacillus subtilis	(1)	Yes	Yes
Borrelia burgdorferi	(2)	No	No
Desulfovibrio vulgaris	(3)	Yes (*via* precorrin-2)	Yes (*via* precorrin-2)
Escherichia coli	(4)	Yes	Yes
Leptospira spp.	(5)	Yes	Yes
Methanosarcina barkeri	(6)	Yes (*via* precorrin-2)	Yes (*via* precorrin-2)
Myxococcus xanthus	(7)	Yes	Yes
Neisseria gonorrhoeae	(8)	Yes	Yes
Salmonella typhimurium	(9)	Yes	Yes
Staphylococcus aureus	(10)	Yes	Yes

Uptake

Organism	Reference	Experiment	Prediction
Borrelia burgdorferi	(2)	No	No
Bradyrhizobium japonicum	(11)	Yes	Yes
Campylobacter jejuni	(12)	Yes	Yes
Porphyromonas gingivalis	(13)	Yes	Yes
Pseudomonas aeruginosa	(14)	Yes	Yes
Serratia marcescens	(15)	Yes	Yes
Shigella dysenteriae	(16)	Yes	Yes
Sodalis glossinidius	(17)	Yes	No
Staphylococcus aureus	(18)	Yes	Yes
Vibrio cholerae	(19)	Yes	Yes
Yersinia enterocolitica	(20)	Yes	Yes
Yersinia pestis	(21)	Yes	Yes

References

(1) Hansson, M.; Gustafsson, M. C.; Kannangara, C. G.; Hederstedt, L. *Biochim.Biophys.Acta* **1997**, *1340*, 97-104.

(2) Posey, J. E.; Gherardini, F. C. S*cience* **2000**, *288*, 1651-1653.

(3) Ishida, T.; Yu, L.; Akutsu, H.; Ozawa, K.; Kawanishi, S.; Seto, A.; Inubushi, T.; Sano, S. *Proc.Natl.Acad.Sci.U.S.A* **1998**, *95*, 4853-4858.

(4) McNicholas, P. M.; Javor, G.; Darie, S.; Gunsalus, R. P. *FEMS Microbiol.Lett.* **1997**, *146*, 143-148.

(5) Guegan, R.; Camadro, J. M.; Saint, G., I; Picardeau, M. *Mol.Microbiol.* **2003**, *49*, 745-754.

(6) Buchenau, B.; Kahnt, J.; Heinemann, I. U.; Jahn, D.; Thauer, R. K. *J.Bacteriol.* **2006**, *188*, 8666-8668.

(7) Dailey, H. A.; Dailey, T. A. *J.Biol.Chem.* **1996**, *271*, 8714-8718.

(8) Turner, P. C.; Thomas, C. E.; Elkins, C.; Clary, S.; Sparling, P. F. *Infect.Immun.* **1998**, *66*, 5215-5223.

(9) Xu, K.; Delling, J.; Elliott, T. *J.Bacteriol.* **1992**, *174*, 3953-3963.

(10) Von Eiff, C.; Heilmann, C.; Proctor, R. A.; Woltz, C.; Peters, G.; Gotz, F. *J.Bacteriol.* **1997**, *179*, 4706-4712.

(11) Nienaber, A.; Hennecke, H.; Fischer, H. M. *Mol. Microbiol.* **2001**, *41*, 787-800.

(12) van Vliet, A. H.; Ketley, J. M.; Park, S. F.; Penn, C. W. *FEMS Microbiol.Rev.* **2002**, *26*, 173-186.

(13) Liu, X.; Olczak, T.; Guo, H. C.; Dixon, D. W.; Genco, C. A. *Infect.Immun.* **2006**, *74*, 1222-1232.

(14) Ochsner, U. A.; Johnson, Z.; Vasil, M. L. *Microbiology* **2000**, *146 (Pt 1)*, 185-198.

(15) Arnoux, P.; Haser, R.; Izadi, N.; Lecroisey, A.; Delepierre, M.; Wandersman, C.; Czjzek, M. *Nat.Struct.Biol.* **1999**, *6*, 516-520.

(16) Wyckoff, E. E.; Duncan, D.; Torres, A. G.; Mills, M.; Maase, K.; Payne, S. M. *Mol.Microbiol.* **1998**, *28*, 1139-1152.

(17) Toh, H.; Weiss, B. L.; Perkin, S. A.; Yamashita, A.; Oshima, K.; Hattori, M.; Aksoy, S. *Genome Res.* **2006**, *16*, 149-156.

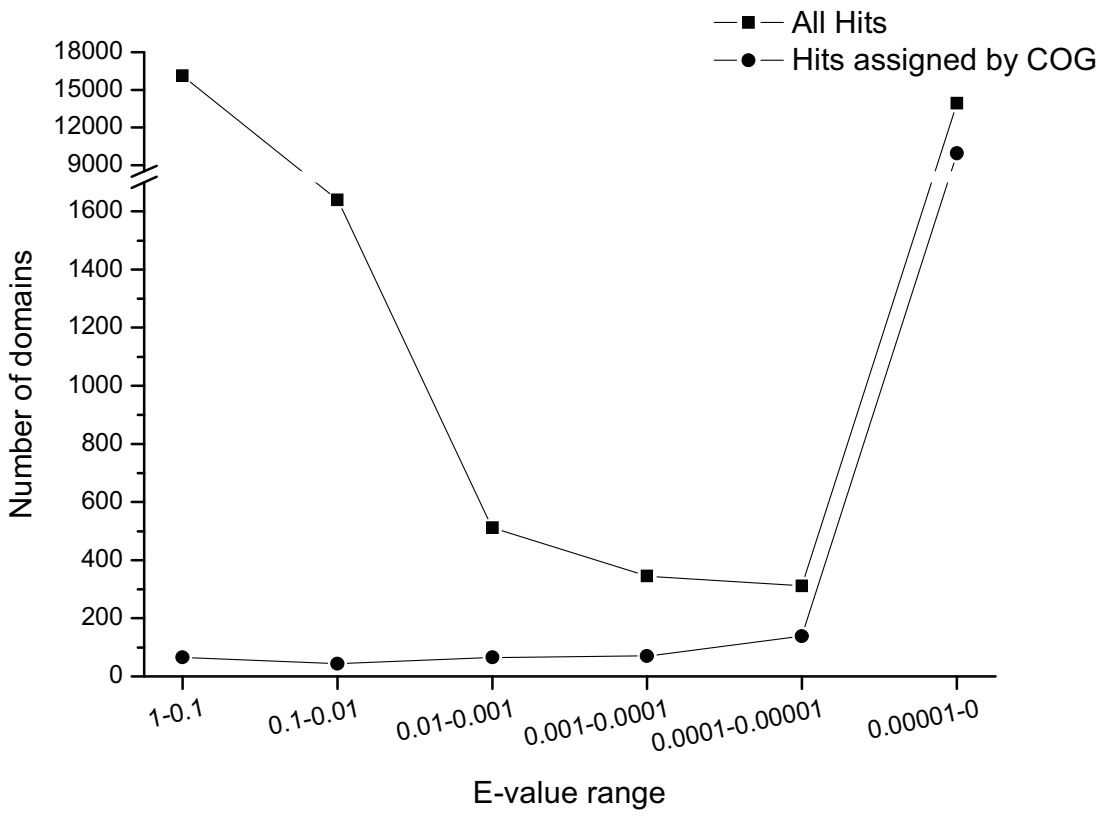
(18) Maresso, A. W.; Schneewind, O. *Biometals* **2006**, *19*, 193-203.

(19) Occhino, D. A.; Wyckoff, E. E.; Henderson, D. P.; Wrona, T. J.; Payne, S. M. *Mol.Microbiol.* **1998**, *29*, 1493-1507.

(20) Stojiljkovic, I.; Hantke, K. *EMBO J.* **1992**, *11*, 4359-4367.

(21) Thompson, J. M.; Jones, H. A.; Perry, R. D. *Infect.Immun.* **1999**, *67*, 3879-3892.

Figure S1. Distribution of HMMER E-values for all hits and for only the hits assigned by COG.



J Biol Inorg Chem
DOI 10.1007/s00775-010-0666-6

ORIGINAL PAPER

The annotation of full zinc proteomes

Ivano Bertini · Leonardo Decaria · Antonio Rosato

Received: 15 January 2010 / Accepted: 16 April 2010

Abstract We obtained an extended functional annotation of zinc proteins using a combination of bioinformatic methods. This work was performed using a number of available predicted zinc proteomes of various representative organisms, leading to the almost complete annotation of, among others, the predicted human zinc proteome. The computational tools exploited included sequence-based and, when possible, structure-based functional predictions. We assigned a hypothetical function to 74% of the 1,472 sequences analyzed that lacked annotation in the starting dataset. We also added new functional categories, not described in the reference dataset, such as ubiquitin binding and DNA replication. As a general conclusion, we can state that the quality of each functional prediction parallels the amount of information for the sequence analyzed: the larger the amount of information, the more detailed and reliable is the proposed functional prediction. Among the findings, we have propose a zinc binding site for archaeal zinc-importing proteins. Furthermore, we propose two groups of transcriptional regulators that are involved in fatty acid metabolism.

Keywords Zinc · Metalloproteomics · Metalloproteome · Zinc finger

Introduction

Zinc is essential for life and is the second most abundant transition metal ion in living organisms after iron. In contrast to other transition metal ions, such as copper and iron, zinc is present in cells in a single oxidation state, zinc(II), which does not undergo redox reactions owing to its filled *d* shell. In the present work, we want to further our understanding of the biochemical functions that underlie the requirement of organisms for zinc. The present work constitutes a bioinformatic contribution toward mapping the many cellular processes involving zinc.

The zinc proteomes of 57 representative living organisms including members of Archea, Bacteria, and Eukarya are available [1]. These 57 zinc proteomes were previously predicted to encode cumulatively 18,336 potential zinc-binding proteins, which had been grouped into ensembles on the basis of sequence similarity [1]. Functional information, either based on available experimental data or on computational biology methods, was described in the annotation of most of these protein sequences, which was relevant for all proteins in a given ensemble. Functional hints for additional protein ensembles were provided by the Gene Ontology (GO) database [2]. Of the 18,336 zinc proteins, 1,472 (784 from Eukarya, 212 from Archea, and 476 from Bacteria) did not have a defined functional annotation and no hit was retrieved from the GO database; their annotations typically described them as hypothetical,

Electronic supplementary material The online version of this article (doi:10.1007/s00775-010-0666-6) contains supplementary material, which is available to authorized users.

I. Bertini (✉) · L. Decaria · A. Rosato
Magnetic Resonance Center (CERM),
University of Florence,
Via L. Sacconi 6,
50019 Sesto Fiorentino, Italy
e-mail: bertini@cerm.unifi.it

I. Bertini · A. Rosato
Department of Chemistry,
University of Florence,
Via della Lastruccia 3,
50019 Sesto Fiorentino, Italy

Published online: 05 May 2010

putative, or predicted proteins [1]. Of these sequences, 932 grouped into 204 ensembles, whereas the remaining 540 did not cluster into any group.

In the present work, we attempted to make functional predictions for these proteins by using sequence- and/or structure-based approaches, exploiting online as well as stand-alone bioinformatics databases and software tools [3–5] (see "Materials and methods"). We assigned a hypothetical function to 1,090 sequences (74% of 1,472), of which 721 constituted 132 ensembles of homologs (65% of the 204 ensembles of putative zinc proteins). For the 382 remaining sequences there were not enough data to assign a function even at a low level of confidence. The coverage of functional annotation originally included about 92% of all zinc proteins, which, after our contribution, increased to 98%. It was found that hydrolytic activity is the most represented zinc-related function in our dataset (33% of the total), followed by transcription (24% of the total). Specifically for eukaryotic zinc proteins, a role in transcription is proposed for more than 37% of these proteins. The present data confirmed the dominant role of zinc fingers in the regulation of expression in eukaryotes, within both activators and repressors. This research shows that an essentially complete annotation of the zinc proteome can be achieved for every living organism whose genome sequence is available.

Materials and methods

The prediction of the zinc proteomes [1] from which the present work started to obtain an extended functional annotation of all zinc proteins was essentially based on the use of a list of known zinc-binding domains available in the Pfam database [6, 7], filtered [8] with the available zinc-binding patterns (ZBPs). Below, we briefly recapitulate the procedure that was used in [1] to obtain these data. The Pfam domains that had been retrieved using "Zn" and "zinc" as query keywords were analyzed manually to collect a list of physiological zinc-binding domains. In parallel, all the structures in the Protein Data Bank (PDB) [9] that physiologically bind at least one zinc ion were selected. In both cases, physiological and nonphysiological zinc binders were separated by manually checking the literature for each of the systems under analysis (protein domains or individual structures). The composition in terms of Pfam domains was determined using the HMMER 2.0 [10] program for all the proteins of known structure where zinc binding is of physiological relevance. The residues coordinating the zinc ion(s) defined the ZBP for each structure analyzed [11–13]. When a ZBP was made up of residues contained in a known Pfam domain, the latter was associated with the ZBP. This resulted in a list of zinc-binding Pfam domains that was as extended as possible; one or more ZBPs were assigned to all zinc-binding Pfam domains having a structurally characterized representative [1]. The physiological relevance of zinc binding was typically supported by the available scientific literature. Andreini et al. [1] obtained the list of predicted zinc proteins by making use of the aforementioned Pfam domains and the associated ZBPs, when available, to scan the proteomes of 57 selected organisms using HMMER 2.0. No experimental verification of these predictions has been carried out. The overall procedure was recently reviewed in [8].

Sequence-based methods

We started our search by analyzing all of the sequences of interest against the entire Pfam database (release 23.0) [6, 7, 14] as we deemed that defining the composition in terms of functional domain(s) (including also those not endowed with zinc-binding capability) for each unknown sequence constitutes the most reasonable starting point for any subsequent analysis (see also our workflow for metalloprotein prediction in [8] and the diagram for comparative genomic analyses of trace elements in [15]). Unfortunately, not all the Pfam domains feature a detailed functional description. For example, in various cases only structural similarities to other proteins are reported. For this reason, we complemented the information available from Pfam with queries to the COG database [16, 17]. Each cluster of an orthologous group of proteins (COG) consists of individual proteins or groups of paralogs from at least three lineages and thus corresponds to an ancient conserved domain. In other words, each COG contains protein sequences or groups thereof that, in different species, evolved from a common ancestral gene by speciation and are thus orthologous. Usually, orthologous proteins have the same Pfam domain composition, thereby making the two methods complementary for the identification of orthology relationships. The identification of orthologs is important for reliable predictions of protein function because they normally retain the same function in the course of evolution. In contrast to the Pfam-based methods, which are mainly dependent on sequence similarity, the use of COGs, which take into account the phylogenetic distance between homologs, is more appropriate to identify relationships between distant proteins with low sequence similarity [16, 17]. To predict putative transmembrane regions in the sequences analyzed, we used TMHMM [18–21] as a stand-alone program, whereas we used pSORT [18, 22, 23], an online tool, to predict the cellular localization. For eukaryotic sequences, these were the only sequence-based tools employed. In contrast, for prokaryotic proteins we additionally exploited the STRING [24–26] and

ShOPs [27] tools. The former shows the possible functional partners of the target protein, on the basis of its COG classification, taking into account a variety of experimentally validated data (such as physical interactions, occurrence in the same metabolic pathways, gene fusions, and co-occurrence or coexpression in different organisms). ShOPs is an online server that allows the visualization of operons. The latter analysis can give indirect but very useful hints about the hypothetical function of a protein when it is codified within an operon containing genes that code for other proteins of known function [28].

Structure-based approach

For each of the 204 functionally unassigned ensembles of homologs that were contained in the starting dataset of predicted zinc proteins [1], we created a hidden Markov model (HMM) profile [10] and queried the entire PDB to identify related proteins with known structures to be used as templates in homology modeling, performed using Modeller 6v2 [29–31]. With a structural model of the target protein, the most conserved residues within the ensemble and/or the Pfam domain HMM profile (which reflects the information of a greater number of sequences) to which the target protein belongs could be mapped onto the protein surface. This, in turn, allowed us to define potential functionally important regions such as catalytic pockets or binding sites. As we are dealing with proteins predicted to bind zinc, it is also important to verify whether the proposed binding residues are close in space in the proposed model, i.e., whether they define a reasonable zinc binding site.

No HMM profile was created for the 540 sequences that did not cluster with other proteins in an ensemble in the original dataset. For these we therefore queried the PDB using each individual amino acidic sequence. When the PDB templates corresponded to proteins lacking a functional characterization (e.g., for structures determined within structural genomics projects), we used bioinformatics tools for structural analysis, such as ProFunc [32], WHISCY [33], ProMate [34], and CastP, to obtain additional functional hints. ProFunc is an online server providing clues on a protein's likely or possible function from its 3D structure. This analysis is based on the use of various databases, including the PDB and UniProt [35, 36] for the identification of structurally characterized clefts, folds, or binding motifs on the protein surface. WHISCY, ProMate, and CastP can identify active residues on the target protein surface and consequently define potential functional areas, such as protein–protein binding sites, enzyme active sites, or small-ligand binding pockets.

When homology modeling was not applicable, we performed protein threading, also known as fold recognition, using the online tool Phyre [37, 38]. Protein threading is a method to predict protein structures that aims at assigning known folds to proteins that do not have homologs with known structure. The prediction is made by aligning each amino acid of the target sequence to a position in the template structure, taken from a library of diverse folds, and evaluating how well the target fits the template, typically using a simplified potential for energy calculation. After the best-fit template has been selected, the structural model of the sequence is built on the basis of the alignment with the chosen template. The quality of the final structural prediction is measured through an expectation value (E value) [37, 38], which estimates quantitatively the number of possible errors given the size of the library used (thus, the lower the better). The functional predictions derived from the analysis of structural models obtained through threading have a lower degree of confidence with respect to the case of structural models obtained by homology modeling, because the models are inherently of lower quality.

For each ensemble we manually checked the consistency between the results given by all the tools used, and the support provided by the relevant scientific literature.

Results and discussion

Overview of functional annotations

We have complemented the already-available information on homology relationships among the predicted zinc proteins contained in our reference dataset using the COG database [16, 17], which, however, does not cover all of the organisms addressed in this work. The Pfam and COG assignments showed a good agreement. Indeed, the sequences in each of the starting ensembles of zinc proteins had the same composition in terms of Pfam domains as well as, when available, the same COG annotation. This provides significant evidence that the 204 groups of sequences defined in the original prediction of the zinc proteome [1] with which we started our analysis were indeed groups of homologs. In addition, we analyzed all the 540 individual sequences that could not be assigned to any ensemble (i.e., they did not have homologs in the organisms subjected to analysis in [1]).

Of the 1,472 sequences lacking functional information in [1], only for about 26% could we obtain structural models of any kind. These can be further separated into homology models (150), threading models encompassing the entire length of the target protein (100), and threading models encompassing only a part of the target protein (134). It is to be noted that we imposed a relatively restrictive threshold on the degree of sequence similarity between the target and template sequences (40% sequence

identity over the entire target length), to ensure that only reliable models were produced [39]. On the other hand, we used a relatively loose threshold (*E* value lower than 10.0) to select results from threading to ensure that the maximum amount of structural information could be obtained. We were thus left with 1,086 proteins for which no structural information could be gathered at all. We obtained useful functional information from the analysis of sequence features alone for 704 of these proteins (Table S1).

In total, we therefore assigned, with variable degrees of confidence, a possible function to 1,090 proteins, of which 721 belonged to 132 ensembles of homologs. This assignment is entirely based on the application of bioinformatic methods and therefore is typically not supported by specific experimental evidence (although there may be experimental evidence available for other, relatively close systems). The assignment results are shown in the pie graph in Fig. 1a. After our analysis, the functional assignment of the zinc proteomes has almost been completed (Figs. 1b, 2; the percentage distribution of the functional categories assigned is given in Table S2). In the reference work, and subsequently in this work, release number 36 of the human proteome was analyzed, counting about 40,000 proteins, of which 9.2% constituted the zinc proteome. The present functional annotation shows that 44% of human zinc proteins are involved in the regulation

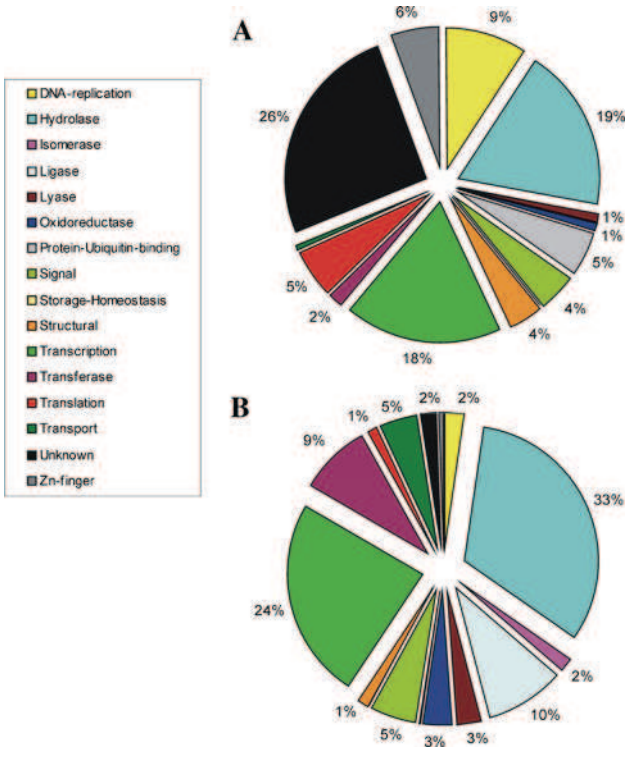


Fig. 1 The functional assignment obtained for **a** the 1,472 sequences analyzed in this work that were previously unassigned, and **b** all the 18,336 proteins in the complete 57 zinc proteomes

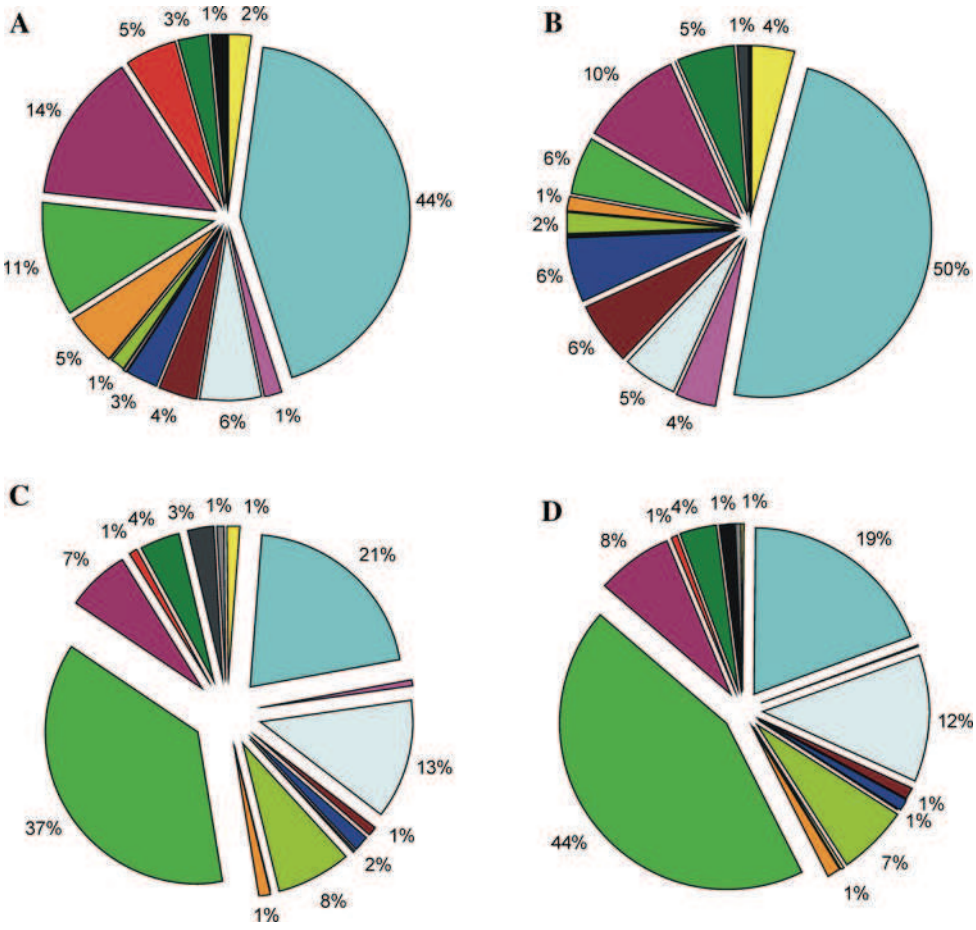
of gene expression, followed by 12% hydrolases (Fig. 2d). These figures compare well with the average distribution in Eukarya. The present work also provided some new hints on the cellular role of various families of zinc proteins, as discussed in more detail below. The portfolio of functional categories to which we could assign zinc-binding proteins was larger and finer-grained than that described in [1], thanks to a more detailed comparison of the various databases used in this work, using functional categories from GO as the reference.

It must be pointed out that about 75% of the proteins still lacking a functional assignment are hypothetical, putative, or predicted eukaryotic proteins. The average size of unassigned ensembles is 2.9, indicating that a significant fraction of them in fact contain only two/three proteins. Nearly two thirds of the proteins that we could not assign had no hits against the Pfam database. To a large extent, these sequences may actually be the result of noncodifying regions of the genome sequence that were erroneously interpreted, e.g., due to wrong positioning of introns [40].

In the following, we analyze three selected case studies. These cases were chosen on the basis of the different content of information available for each of them, to exemplify the degree of insight that can be reached in each case. In the first test case, the amount of available information is extensive also at the functional level. In the second case, at the time of preparation of this manuscript, there was good structural information but hardly any even indirectly relevant functional information; nevertheless, the information obtained from the analysis of gene organization features allowed us to obtain a detailed functional prediction. Similarly, for the third case study, we could obtain a detailed functional prediction, on the basis of the analysis of potential physiological protein partners.

ZIP proteins

Four archeal sequences contained the ZIP domain, which is characteristic of various transmembrane zinc transporters found in all domains of life [41–43]. ZIP proteins move zinc to the cytoplasm from the extracellular medium or from vacuoles, in contrast with the action of cation diffusion facilitator proteins, which mediate the reverse process. Through structure threading methods, we could model part of the NP_147044.1 sequence onto the PDB structure of a Cl^- transporter with 11 transmembrane regions (PDB code 1KPL [44]) (Fig. 3). In particular, although the *E* value (see "Materials and methods") is quite poor (4.0), the model can be used to interpret the common sequence properties of this ensemble of proteins. On the basis of the high conservation in sequence of the residues involved (Fig. 3b), we propose that Hx(3)Ex(29)H is the putative ZBP; note that the Glu ligand is either conserved or



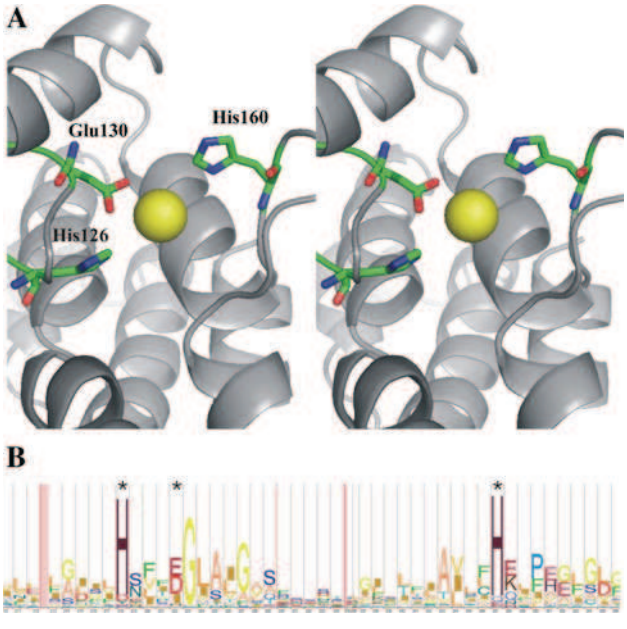


Fig. 3 **a** Protein structure threading onto 1KPL for the representative sequence NP_147044.1. The side chains of the amino acids in the proposed zinc-binding pattern Hx(3)Ex(29)H are shown. **b** The hidden Markov model (HMM) logo of the ZIP domain; the three residues are highlighted with *asterisks*

conservatively substituted by Asp. The present proposition is reinforced by the fact that these residues are close in space in the structural model (Fig. 3a). Note that although zinc binding by ZIP proteins has been established, the mode by which this is accomplished is still not fully supported by structural evidence; the present data therefore provide novel insight into the atomic-level features of the archaeal system.

A putative regulator of the metabolism of fatty acid

When we collected our data, for an ensemble of 27 archeal sequences we identified a homolog of known structure with PDB code 2GNR (from *Sulfolobus solfataricus*). This was the structure of a protein dimer solved at the Joint Center of Structural Genomics. The protein was described as having unknown function. Each subunit binds a zinc ion with the known ZBP Cx(2)Cx(10)Cx(2)C. Using STRING, we found a functional correlation between the target protein and acetyl-CoA acetyltransferase, the first enzyme in the fatty acid biosynthetic pathway. Other putative functional partners were 3-hydroxy-3-methylglutaryl-CoA synthetase, 3-hydroxyacyl-CoA dehydrogenase, 3-ketoacid-CoA

transferase, hydroxymethylglutaryl-CoA reductase, and pyruvate/ferredoxin oxidoreductase, which are all involved in the fatty acid anabolism. These enzymes are coded by genes in well-defined operons, upstream of which is located the codifying sequence of the protein under analysis here. Finally, the zinc-binding knuckle contained in the structure corresponds to a known zinc-finger fold, potentially able to bind DNA. Figure 4a shows the structure of the protein, with the two subunits in light blue and green. As the sequence logo shows, the four Cys in each subunit are either completely or very highly conserved (Fig. 4b). CastP [45] recognized a putative binding pocket; the residues involved are shown in red in Fig. 4a. The size and shape of the pocket were compatible with acetyl-CoA, the starting molecule for biosynthesis of fatty acids. All these hints allowed us to propose that these proteins are putative transcriptional factors regulating fatty acid metabolism, possibly responding to acetyl-CoA concentration. It has to be noted that, after the completion of this work, the reference PDB structure 2GNR was superseded by 3IRB (doi:10.2210/pdb3irb/pdb), classified as acyl-CoA binding protein. We considered the recent data as a validation of the proposed functional predictions.

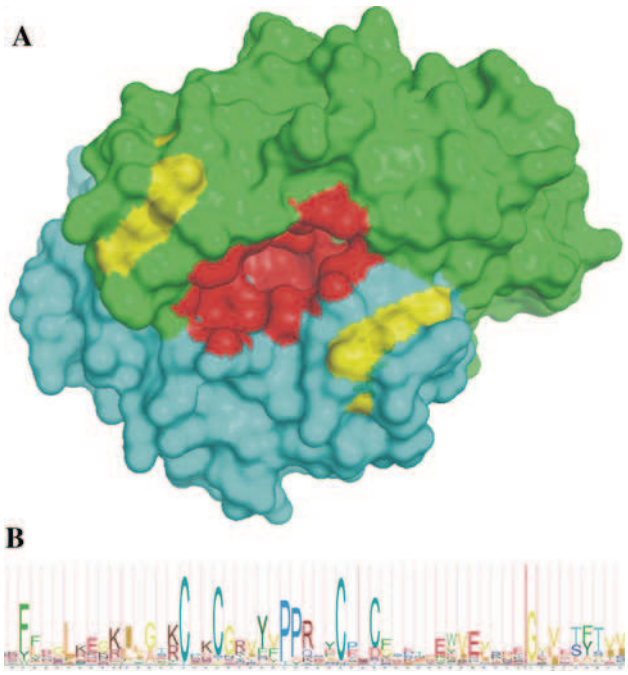


Fig. 4 **a** The homodimeric Protein Data Bank structure of 3IRB. The two monomers are in *light blue* and *green*, the eight Cys constituting the zinc-binding pattern are in *yellow*. The residues constituting the putative acetyl-CoA binding-pocket are reported in *red*. **b** HMM logo of the corresponding ensemble of proteins (only one monomer is shown)

A putative transcriptional factor regulating pilin biosynthesis

Twelve bacterial sequences had a homolog of known structure corresponding to PDB entry 1LV3 [46], which is a monomeric zinc-binding protein (YacG) with unknown function whose structure was solved by the Northeast Structural Genomics Consortium. In these sequences, we identified a DNA-binding zinc-finger domain that is present in known transcriptional factors. We now additionally propose that they are involved in the type II secretion system. Using the STRING tool, we identified various putative functional partners belonging to this system, such as PilA (pre-pilin), a precursor of type IV fimbrial pilin, PilB, and PilC, ATPases for PilA maturation and assembly, and PilD, a peptidase processing the N-terminal region of

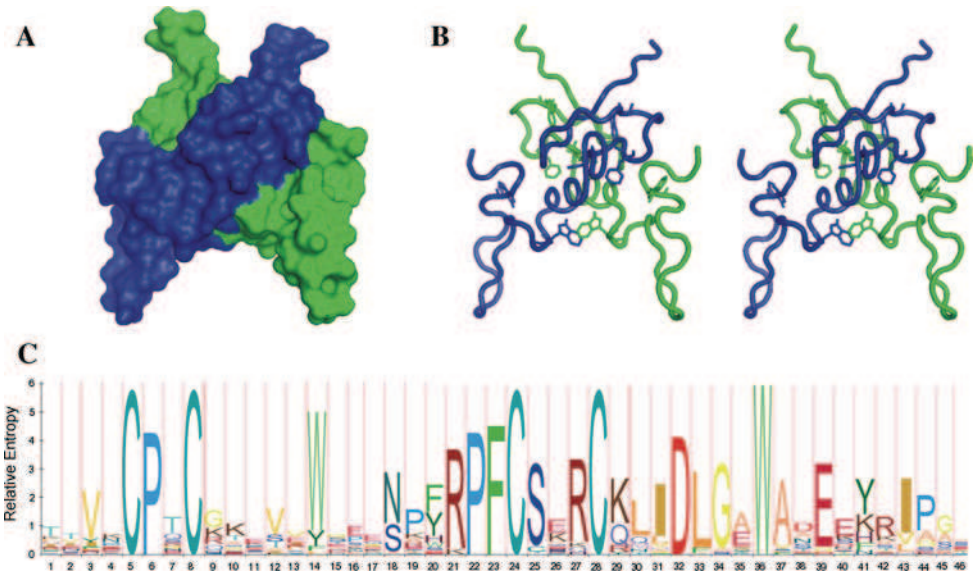


Fig. 5 Proposed model for dimerization for YacG based on the 1LV3 structure. **a** Protein surface, **b** protein backbone representation, showing the side chains of the highly conserved residues (stereoview), and **c** HMM logo of the corresponding ensemble of proteins

PilA. In archea and bacteria, polymers of type IV fimbrian pilin form flagella, for twitching motility, and f-pilus, for DNA transfer in processes such as conjugation, infection, and transformation. All the reported codifying sequences are contained in well-characterized operons, having the target sequence downstream. A model for the possibly functionally active YacG homodimer, which is commonly the oligomerization state for transcriptional factors, could be successfully built (Fig. 5). The HMM-logo in Fig. 5c shows that the zinc-binding residues are completely conserved.

Conclusions

Genome sequencing projects are continuously making new DNA sequences and potential protein sequences available. Several of these are not experimentally characterized, and thus a hypothetical function can be proposed only by functional prediction [47, 48]. Functional prediction can be a powerful and relatively high confidence method to this end, exploiting sequence and/or structural features [49, 50]. The composition of the analyzed sequences in terms of functional domain(s), their cellular localization, hints about functional partners, and conservation of residues among homologs are all important information allowing computational biologists to figure out hypothetical functions. The analysis of protein structures and protein surfaces can provide even more reliable and detailed hints [51].

There are computational approaches reported in the literature that can provide the metal proteome for each completely sequenced genome. We showed that this information can be complemented for the zinc proteome by an essentially complete functional annotation, again as the result of the systematic application of computational prediction methods. An experimental verification of these predictions is thus generally warranted, at least for selected cases of particular interest, where the purpose would be beyond the validation of the functional predictions proposed in this work. The annotations of the zinc proteomes of the 57 organisms analyzed are available at http://www.cerm.unifi.it/home/research/genomebrowsing.html.

References

1. Andreini C, Banci L, Bertini I, Rosato A (2006) J Proteome Res 5:3173–3178
2. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Nat Genet 25:25–29
3. Dobson PD, Cai YD, Stapley BJ, Doig AJ (2004) Curr Med Chem 11:2135–2142
4. Baker EN, Arcus VL, Lott JS (2003) Appl Bioinformatics 2:S3–10
5. Lee D, Redfern O, Orengo C (2007) Nat Rev Mol Cell Biol 8:995–1005
6. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR (2004) Nucleic Acids Res 32(Database issue):D138–D141
7. Sonnhammer EL, Eddy SR, Durbin R (1997) Proteins 28:405–420
8. Andreini C, Bertini I, Rosato A (2009) Acc Chem Res 42:1471–1479
9. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235–242
10. Eddy SR (1998) Bioinformatics 14:755–763
11. Andreini C, Bertini I, Rosato A (2004) Bioinformatics 20:1373–1380
12. Andreini C, Banci L, Bertini I, Rosato A (2006) J Proteome Res 5:196–201
13. Castagnetto JM, Hennessy SW, Roberts VA, Getzoff ED, Tainer JA, Piquet ME (2002) Nucleic Acids Res 30:379–382
14. Coggill P, Finn RD, Bateman A (2008) Curr Protoc Bioinformatics 23:2.5.1–2.5.17
15. Zhang Y, Gladyshev VN (2009) Chem Rev 109:4828–4861
16. Tatusov RL, Koonin EV, Lipman DJ (1997) Science 278:631–637
17. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova RD, Koonin EV (2001) Nucleic Acids Res 29:22–28
18. Chen Y, Yu P, Luo J, Jiang Y (2003) Mamm Genome 14:859–865
19. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) J Mol Biol 305:567–580
20. Moller S, Croning MD, Apweiler R (2001) Bioinformatics 17:646–653
21. Sonnhammer EL, von Heijne G, Krogh A (1998) Proc Int Conf Intell Syst Mol Biol 6:175–182
22. Sprenger J, Fink JL, Teasdale RD (2006) BMC Bioinformatics 7(Suppl 5):S3
23. Liu J, Kang S, Tang C, Ellis LB, Li T (2007) Nucleic Acids Res 35:e96
24. Snel B, Lehmann G, Bork P, Huynen MA (2000) Nucleic Acids Res 28:3442–3444
25. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B (2003) Nucleic Acids Res 31:258–261
26. von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P (2007) Nucleic Acids Res 35:D358–D362
27. van Bakel H, Huynen M, Wijmenga C (2004) Bioinformatics 20:2644–2655
28. Galperin MY, Koonin EV (2000) Nat Biotechnol 18:609–613
29. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A (2007) Curr Protoc Protein Sci 50:2.9.1–2.9.31
30. Sali A, Potterton L, Yuan F, Van Vlijmen H, Karplus M (1995) Proteins Struct Funct Genet 23:318–326
31. Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Methods Mol Biol 426:145–159
32. Laskowski RA, Watson JD, Thornton JM (2005) Nucleic Acids Res 33:W89–W93
33. de Vries SJ, van Dijk AD, Bonvin AM (2006) Proteins 63:479–489
34. Neuvirth H, Raz R, Schreiber G (2004) J Mol Biol 338:181–199
35. Consortium The Uniprot (2007) Nucleic Acids Res 35:D193–D197

36. Leinonen R, Diez FG, Binns D, Fleischmann W, Lopez R, Apweiler R (2004) Bioinformatics 20:3236–3237
37. Bennett-Lovsey RM, Herbert AD, Sternberg MJ, Kelley LA (2008) Proteins 70:611–625
38. Kelley LA, Sternberg MJ (2009) Nat Protoc 4:363–371
39. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Annu Rev Biophys Biomol Struct 29:291–325
40. Brent MR, Guigo R (2004) Curr Opin Struct Biol 14:264–272
41. Grotz N, Fox T, Connolly E, Park W, Guerinot ML, Eide D (1998) Proc Natl Acad Sci USA 95:7220–7224
42. Eide DJ (2006) Biochim Biophys Acta 1763:711–722
43. Gaither LA, Eide DJ (2001) Biometals 14:251–270
44. Dutzler R, Campbell EB, Cadene M, Chait BT, MacKinnon R (2002) Nature 415:287–294
45. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) Nucleic Acids Res 34:W116–W118
46. Ramelot TA, Cort JR, Yee AA, Semesi A, Edwards AM, Arrowsmith CH, Kennedy MA (2002) Proteins 49:289–293
47. Godzik A, Jambon M, Friedberg I (2007) Cell Mol Life Sci 64:2505–2511
48. Baker D, Sali A (2001) Science 294:93–96
49. Pandit SB, Bhadra R, Gowri VS, Balaji S, Anand B, Srinivasan N (2004) BMC Bioinformatics 5:28
50. Madera M (2008) Bioinformatics 24:2630–2631
51. Serres MH, Riley M (2004) OMICS 8:306–321

Seq. ID	Pfam	COG	hmmtop	pSORT	Structural data
NP_011304.2	ATP_bind_3	COG0037	0	C	Homology modeling
NP_013797.1	No hits	COG1112	0	C	Homology modeling
NP_054758.2	No hits	No hits	1	T	Homology modeling
NP_068860.1	DUF35	COG1545	0	C	Homology modeling
NP_068874.1	ACP_syn_III ACP_syn_III_C DUF35	COG1545	0	C	Homology modeling
NP_068971.1	DUF35	COG1545	0	C	Homology modeling
NP_068974.1	DUF35	COG1545	0	C	Homology modeling
NP_069041.1	DUF35	COG1545	0	C	Homology modeling
NP_069044.1	No hits	COG1096	0	C	Homology modeling
NP_069122.1	DUF35	COG1545	0	C	Homology modeling
NP_069362.1	Nop10p	COG2260	0	C	Homology modeling
NP_069799.1	DUF35	COG1545	0	C	Homology modeling
NP_070121.1	DUF35	COG1545	0	C	Homology modeling
NP_070150.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_070424.1	ATP_bind_3	COG0037	1	T	Homology modeling
NP_071239.1	DUF35	COG1545	0	C	Homology modeling
NP_104365.1	Glyoxalase	COG0346	0	C	Homology modeling
NP_104942.1	Metallophos	COG1692	0	C	Homology modeling
NP_106959.1	DUF329	COG3024	0	C	Homology modeling
NP_125745.1	No hits	COG1096	0	C	Homology modeling
NP_126184.1	ATP_bind_3 ExsB	COG0037	0	C	Homology modeling
NP_126361.1	Mov34	COG1310	0	C	Homology modeling
NP_126530.1	Nop10p	COG2260	0	C	Homology modeling
NP_127340.1	ThiI tRNA_Me_trans	COG0037	0	C	Homology modeling
NP_147226.1	Nop10p	COG2260	0	C	Homology modeling
NP_147238.1	No hits	COG1096	0	C	Homology modeling
NP_147309.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_147414.1	No hits	COG1310	0	C	Homology modeling
NP_148226.1	DUF35	COG1545	0	C	Homology modeling
NP_148374.1	ATP_bind_3 RecR	COG0037	0	C	Homology modeling
NP_148579.1	DUF35	COG1545	0	C	Homology modeling
NP_177744.1	ATP_bind_3	COG0037	1	T	Homology modeling
NP_181956.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_212639.1	Metallophos	COG1692	0	C	Homology modeling
NP_213713.1	FdhE	COG3058	0	C	Homology modeling
NP_213916.1	ATP_bind_3 PAPS_reduct	COG0037	0	C	Homology modeling
NP_214163.1	Mov34	COG1310	0	C	Homology modeling
NP_214229.1	Metallophos	COG1692	0	C	Homology modeling
NP_228012.1	ThiI ATP_bind_3 ExsB	COG0037	0	C	Homology modeling
NP_247081.1	Nop10p	COG2260	0	C	Homology modeling
NP_248152.1	ATP_bind_3	COG0037	1	T	Homology modeling
NP_248483.1	ThiI ATP_bind_3 ExsB	COG0037	0	C	Homology modeling
NP_248560.1	DUF35	COG1545	0	C	Homology modeling
NP_249330.1	NLPC_P60	COG1310	0	C	Homology modeling
NP_253220.1	DUF329	COG3024	0	C	Homology modeling
NP_253497.1	FdhE	COG3058	1	T	Homology modeling
NP_273379.1	DUF329	COG3024	0	C	Homology modeling
NP_276314.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_276424.1	Nop10p	COG2260	0	C	Homology modeling
NP_276435.1	No hits	COG1096	0	C	Homology modeling
NP_276514.1	DUF35	COG1545	0	C	Homology modeling
NP_276852.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_279323.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_294125.1	No hits	COG1310	0	C	Homology modeling
NP_295005.1	Metallophos	COG1692	0	C	Homology modeling

NP_347552.1	Glyco_hydro_9	No hits	0	C	Homology modeling
NP_347553.1	Glyco_hydro_9	No hits	0	C	Homology modeling
NP_375928.1	DUF35	COG1545	0	C	Homology modeling
NP_376178.1	Pox_D5	COG3378	0	C	Homology modeling
NP_376519.1	ATP_bind_3 ExsB	COG0037	0	C	Homology modeling
NP_376861.1	Pox_D5	COG3378	0	C	Homology modeling
NP_377035.1	DUF35	COG1545	0	C	Homology modeling
NP_377305.1	DUF35	COG1545	0	C	Homology modeling
NP_377416.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_377780.1	DUF35	COG1545	0	C	Homology modeling
NP_377781.1	DUF35	COG1545	0	C	Homology modeling
NP_378262.1	No hits	COG1096	0	C	Homology modeling
NP_378420.1	DUF35	COG1545	0	C	Homology modeling
NP_389579.1	Metallophos	COG1692	0	C	Homology modeling
NP_390560.1	No hits	COG3443	1	T	Homology modeling
NP_393510.1	DUF35	COG1545	0	C	Homology modeling
NP_393831.1	DUF35	COG1545	0	C	Homology modeling
NP_394387.1	No hits	COG1096	0	C	Homology modeling
NP_394578.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_394658.1	Nop10p	COG2260	0	C	Homology modeling
NP_394910.1	DUF35	COG1545	0	C	Homology modeling
NP_414643.1	DUF329	COG3024	0	C	Homology modeling
NP_416482.1	No hits	COG3443	1	T	Homology modeling
NP_418327.1	FdhE	COG3058	1	T	Homology modeling
NP_421143.1	DUF329	COG3024	0	C	Homology modeling
NP_421756.1	Glyoxalase	COG0346	0	C	Homology modeling
NP_422038.1	Metallophos	COG1692	0	C	Homology modeling
NP_438182.1	FdhE	COG3058	0	C	Homology modeling
NP_439052.1	DUF329	COG3024	0	C	Homology modeling
NP_440942.1	No hits	COG1310	0	C	Homology modeling
NP_444227.1	Mov34	COG1310	0	C	Homology modeling
NP_484459.1	Phosphodiest	COG1524	0	C	Homology modeling
NP_486947.1	No hits	COG1310	0	C	Homology modeling
NP_490829.1	ResIII	COG1112	0	C	Homology modeling
NP_497144.1	zf-TAZ	No hits	0	C	Homology modeling
NP_497145.1	zf-TAZ	No hits	0	C	Homology modeling
NP_499197.1	zf-TAZ	No hits	0	C	Homology modeling
NP_499201.1	zf-TAZ	No hits	0	C	Homology modeling
NP_499865.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_505382.2	No hits	No hits	0	C	Homology modeling
NP_510006.1	zf-TAZ	No hits	1	T	Homology modeling
NP_519816.1	NLPC_P60	COG1310	0	C	Homology modeling
NP_520951.1	DUF329	COG3024	0	C	Homology modeling
NP_559262.1	DUF35	COG1545	0	C	Homology modeling
NP_559672.1	DNA_ligase_ZBD	COG1096	0	C	Homology modeling
NP_559708.1	No hits	COG1310	0	C	Homology modeling
NP_559910.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_560037.1	DUF35	COG1545	0	C	Homology modeling
NP_560050.1	DUF35	COG1545	0	C	Homology modeling
NP_560441.1	Nop10p	COG2260	0	C	Homology modeling
NP_560888.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_572767.1	ResIII	COG1112	0	C	Homology modeling
NP_602425.1	Metallophos	COG1692	1	T	Homology modeling
NP_610451.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_610920.2	No hits	COG1112	0	C	Homology modeling
NP_612081.1	No hits	No hits	1	T	Homology modeling
NP_613347.1	No hits	COG1096	0	C	Homology modeling
NP_613705.1	Nop10p	COG2260	0	C	Homology modeling

NP_614386.1	tRNA_Me_trans ATP_bind_3 ExsB	COG0037	0	C	Homology modeling
NP_614660.1	DUF35	COG1545	0	C	Homology modeling
NP_615605.1	Nop10p	COG2260	0	C	Homology modeling
NP_615680.1	No hits	COG1096	0	C	Homology modeling
NP_616897.1	ATP_bind_3 PAPS_reduct ExsB	COG0037	0	C	Homology modeling
NP_618911.1	DUF35	COG1545	0	C	Homology modeling
NP_621776.1	ThiI tRNA_Me_trans	COG0037	0	C	Homology modeling
NP_622989.1	Metallophos	COG1692	0	C	Homology modeling
NP_627139.1	Mov34	COG1310	0	C	Homology modeling
NP_627449.1	Phosphodiest	COG1524	1	T	Homology modeling
NP_630659.1	Phosphodiest	COG1524	1	T	Homology modeling
NP_660275.1	ATP_bind_3	COG0037	0	C	Homology modeling
NP_661261.1	Metallophos	COG1692	1	T	Homology modeling
NP_662556.1	Mov34	COG1310	0	C	Homology modeling
NP_715746.1	FdhE	COG3058	1	T	Homology modeling
NP_716049.1	DUF329	COG3024	0	C	Homology modeling
NP_718517.1	NLPC_P60	COG1310	0	C	Homology modeling
NP_728575.1	No hits	No hits	0	C	Homology modeling
NP_798116.1	Glyoxalase	COG0346	0	C	Homology modeling
NP_798908.1	DUF329	COG3024	0	C	Homology modeling
NP_842349.1	Mov34	COG1310	0	C	Homology modeling
NP_864703.1	Metallophos	COG1692	0	C	Homology modeling
NP_865141.1	DUF329	COG3024	0	C	Homology modeling
NP_869560.1	Phosphodiest	COG1524	1	T	Homology modeling
NP_952191.1	Metallophos	COG1692	0	C	Homology modeling
NP_952271.1	DUF329	COG3024	0	C	Homology modeling
NP_963572.1	ATP_bind_3	COG0037	3	T	Homology modeling
NP_963802.1	ATP_bind_3	COG0037	0	C	Homology modeling
YP_004081.1	ATP_bind_3	COG0037	0	C	Homology modeling
YP_004234.1	Metallophos	COG1692	0	C	Homology modeling
YP_005102.1	No hits	COG1310	0	C	Homology modeling
YP_009360.1	ATP_bind_3	COG0037	0	C	Homology modeling
YP_009799.1	FdhE	COG3058	1	T	Homology modeling
YP_010173.1	Metallophos	COG1692	0	C	Homology modeling
YP_053480.1	Metallophos	COG1692	0	C	Homology modeling
YP_112633.1	Mov34	COG1310	0	C	Homology modeling
YP_114520.1	DUF329	COG3024	0	C	Homology modeling
NP_001001677.1	No hits	No hits	0	C	N/A
NP_001005404.1	Yippee	No hits	0	N	N/A
NP_001006658.1	No hits	No hits	0	C	N/A
NP_001012733.1	No hits	No hits	0	C	N/A
NP_001014610.1	FLYWCH FLYWCH GST_N GST_C	No hits	0	C	N/A
NP_001019764.1	zf-MYND	No hits	1	T	N/A
NP_001021187.1	No hits	No hits	0	C	N/A
NP_001021478.1	No hits	No hits	0	C	N/A
NP_001021504.1	No hits	No hits	0	C	N/A
NP_001021725.1	FKBP_C FKBP_C	No hits	0	C	N/A
NP_001021749.1	zf-C2H2	No hits	0	N	N/A
NP_001021922.1	No hits	No hits	1	T	N/A
NP_001021991.1	No hits	No hits	1	T	N/A
NP_001022097.1	No hits	No hits	1	T	N/A
NP_001022098.1	No hits	No hits	1	T	N/A
NP_001022244.1	No hits	No hits	0	C	N/A
NP_001022535.1	No hits	COG1196	0	C	N/A

NP_001022588.1	zf-C2H2	No hits	0	N	N/A
NP_001022897.1	No hits	No hits	0	C	N/A
NP_001023041.1	No hits	No hits	0	C	N/A
NP_001023042.1	No hits	No hits	0	C	N/A
NP_001023059.1	No hits	No hits	0	C	N/A
NP_001023559.1	zf-C2H2	No hits	0	N	N/A
NP_001023796.1	No hits	No hits	0	C	N/A
NP_001023832.1	zf-MYND	No hits	0	N	N/A
NP_001023834.1	zf-MYND 2OG-FeII_Oxy	No hits	0	N	N/A
NP_001024083.1	No hits	No hits	1	T	N/A
NP_001024423.1	No hits	No hits	1	T	N/A
NP_001024904.1	zf-C3HC4	No hits	1	T	N/A
NP_001027416.1	RWD zf-C3HC4	No hits	0	C	N/A
NP_001031397.1	No hits	COG5109	0	N	N/A
NP_001031643.1	SCRL	No hits	1	T	N/A
NP_001031733.1	Yippee	No hits	0	N	N/A
NP_001032059.1	zf-CCCH	No hits	0	C	N/A
NP_001032972.1	No hits	No hits	0	C	N/A
NP_001032973.1	No hits	No hits	1	T	N/A
NP_001032975.1	zf-CCHC	No hits	1	T	N/A
NP_001033473.1	No hits	No hits	0	C	N/A
NP_001033841.1	Yippee	No hits	0	N	N/A
NP_001034030.1	No hits	No hits	0	C	N/A
NP_002589.2	zf-MYND PDCD2_C	No hits	0	N	N/A
NP_002753.2	DNA_pol_B_exo DNA_pol_B	No hits	0	N	N/A
NP_005416.1	TP2	No hits	0	C	N/A
NP_005822.1	CoCoA	No hits	0	C	N/A
NP_009504.1	Yippee	No hits	0	N	N/A
NP_010023.1	No hits	No hits	0	C	N/A
NP_010541.1	No hits	COG5109	2	T	N/A
NP_011063.1	No hits	No hits	0	C	N/A
NP_011585.1	zf-BED	No hits	2	T	N/A
NP_012040.1	NMD3	COG1499	1	T	N/A
NP_012622.1	BIR BIR	No hits	0	C	N/A
NP_013028.1	PA14	No hits	1	T	N/A
NP_013671.1	No hits	No hits	0	C	N/A
NP_013818.1	zf-MYND	No hits	2	T	N/A
NP_013912.1	No hits	No hits	5	T	N/A
NP_014695.1	zf-AN1	No hits	1	T	N/A
NP_027420.1	zf-CCCH	No hits	1	T	N/A
NP_036477.1	zf-C2H2	No hits	0	N	N/A
NP_037445.1	Yippee	No hits	0	N	N/A
NP_054781.1	No hits	COG1439	0	C	N/A
NP_054890.1	No hits	No hits	0	C	N/A
NP_055959.1	SAM_1 zf-CCHC	No hits	1	T	N/A
NP_056464.1	No hits	No hits	0	N	N/A
NP_056980.2	zf-MYND	No hits	0	N	N/A
NP_057022.2	NMD3	COG1499	0	C	N/A
NP_057094.1	No hits	No hits	0	N	N/A
NP_057145.1	Yippee	No hits	0	N	N/A
NP_057305.1	Dynactin_p62	No hits	0	C	N/A
NP_057589.2	S1 zf-CCHC	No hits	1	T	N/A
NP_059993.2	zf-TRAF	No hits	0	N	N/A
NP_060212.3	zf-CCHC	No hits	0	N	N/A
NP_061154.1	No hits	COG1196	0	C	N/A
NP_061180.1	No hits	No hits	2	T	N/A
NP_061935.1	No hits	No hits	3	T	N/A

NP_064525.1	SAM_2 SAM_1 SAM_2 SAM_1 PID	No hits	1	T	N/A
NP_064582.1	SET zf-MYND	No hits	1	T	N/A
NP_065746.2	Exonuc_X-T	No hits	1	T	N/A
NP_065830.2	No hits	No hits	0	C	N/A
NP_065965.3	zf-C3HC4	No hits	0	N	N/A
NP_068752.2	No hits	No hits	0	C	N/A
NP_068958.1	Radical_SAM	No hits	0	C	N/A
NP_069319.1	TFIIB_Zn_Ribbon	No hits	0	C	N/A
NP_069758.1	Peptidase_M48	No hits	6	T	N/A
NP_069831.1	HTH_5 MarR	No hits	0	C	N/A
NP_069938.1	SWIM	No hits	0	C	N/A
NP_070018.1	No hits	No hits	0	C	N/A
NP_070134.1	OMPdecase Ribul_P_3_epim	No hits	0	C	N/A
NP_070796.1	NMD3	COG1499	0	C	N/A
NP_071192.1	Auto_anti-p27	No hits	2	T	N/A
NP_071334.1	zf-MYND 2OG-FeII_Oxy	No hits	1	T	N/A
NP_071918.1	WD40 WD40 WD40 WD40 WD40 zf-C2H2	No hits	2	T	N/A
NP_073580.1	zf-MYND SET	No hits	0	N	N/A
NP_073599.2	No hits	COG5109	0	N	N/A
NP_073617.1	No hits	COG5109	0	N	N/A
NP_077722.1	Opiods_neuropep	No hits	1	T	N/A
NP_078772.1	No hits	No hits	0	N	N/A
NP_079408.3	No hits	No hits	0	C	N/A
NP_103353.1	DUF899	COG4312	0	C	N/A
NP_104033.1	AstE_AspA	No hits	0	C	N/A
NP_104286.1	DUF1037	No hits	0	C	N/A
NP_104405.1	Polysacc_deac_1	No hits	1	T	N/A
NP_104422.1	Glycos_transf_2 NodS CheR Methyltransf_11 Methyltransf_12	No hits	0	C	N/A
NP_104575.1	Polysacc_deac_1 adh_short ADH_zinc_N Epimerase	No hits	1	T	N/A
NP_104585.1	SBP_bac_1	No hits	1	T	N/A
NP_104792.1	DUF899	COG4312	0	C	N/A
NP_105030.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_105261.1	Abi	COG1266	8	T	N/A
NP_105701.1	DUF1272	COG3813	0	C	N/A
NP_108456.1	No hits	No hits	0	C	N/A
NP_112504.1	zf-RanBP	No hits	0	N	N/A
NP_112564.1	zf-RanBP	No hits	0	N	N/A
NP_112587.1	No hits	No hits	1	T	N/A
NP_113665.2	Yippee	No hits	0	N	N/A
NP_115633.2	zf-MYND	No hits	0	N	N/A
NP_115635.1	WD40 WD40	No hits	0	C	N/A
NP_115711.2	No hits	No hits	1	T	N/A
NP_116575.1	DEAD_2	No hits	1	T	N/A
NP_126099.1	Abi	No hits	6	T	N/A
NP_126666.1	NMD3	COG1499	0	C	N/A
NP_147076.1	No hits	No hits	0	C	N/A
NP_147288.1	TFIIB_Zn_Ribbon	No hits	0	C	N/A
NP_147394.1	DUF1743 tRNA_anti DUF1610	No hits	1	T	N/A
NP_147399.1	SWIM	No hits	1	T	N/A

NP_147465.1	NMD3	COG1499	1	T	N/A
NP_147924.1	No hits	No hits	1	T	N/A
NP_148201.1	UPF0020	No hits	1	T	N/A
NP_148286.1	No hits	No hits	0	C	N/A
NP_148394.1	No hits	No hits	0	C	N/A
NP_148547.1	NMD3	COG1499	0	C	N/A
NP_171623.1	No hits	No hits	0	C	N/A
NP_172776.1	PH Oxysterol_BP	No hits	2	T	N/A
NP_172920.1	No hits	No hits	0	C	N/A
NP_173258.1	Arf Miro MMR_HSR1 Ras GCK	No hits	1	T	N/A
NP_174498.1	Abhydrolase_2	No hits	0	C	N/A
NP_174611.1	zf-GRF	No hits	1	T	N/A
NP_175087.1	No hits	No hits	0	C	N/A
NP_175414.1	MuDR SWIM	No hits	0	C	N/A
NP_175480.1	Stig1	No hits	1	T	N/A
NP_175487.1	Stig1	No hits	1	T	N/A
NP_175721.1	Stig1	No hits	1	T	N/A
NP_176256.2	SWIM	No hits	0	C	N/A
NP_176608.2	MuDR SWIM	No hits	0	C	N/A
NP_176823.1	No hits	No hits	0	N	N/A
NP_176970.1	PHD	No hits	0	N	N/A
NP_177039.1	dCMP_cyt_deam_1	No hits	1	C	N/A
NP_177172.2	zf-MYND	No hits	0	N	N/A
NP_177854.3	zf-CW SET	No hits	0	C	N/A
NP_178413.1	Abi	No hits	8	T	N/A
NP_178422.1	GCK	No hits	0	C	N/A
NP_178476.1	NMD3	COG1499	0	C	N/A
NP_178725.1	zf-GRF	No hits	1	T	N/A
NP_178742.1	SWIM DUF223	No hits	0	C	N/A
NP_179063.1	SWIM	No hits	0	N	N/A
NP_179239.2	GYF zf-CCCH	No hits	1	T	N/A
NP_181540.1	Yippee	No hits	0	N	N/A
NP_187511.2	Yippee	No hits	0	N	N/A
NP_187801.2	zf-TRAF UbiA	No hits	10	T	N/A
NP_188372.1	PHD	No hits	1	T	N/A
NP_188819.2	zf-MYND SET	No hits	0	N	N/A
NP_189669.1	zf-GRF	No hits	1	T	N/A
NP_189910.2	ARID PHD	No hits	0	N	N/A
NP_189935.1	zf-CCHC	No hits	0	N	N/A
NP_190393.1	Response_reg	No hits	0	C	N/A
NP_190732.1	No hits	No hits	3	T	N/A
NP_190733.1	TolA	No hits	3	T	N/A
NP_191148.1	Yippee	No hits	0	N	N/A
NP_191611.1	No hits	No hits	0	N	N/A
NP_191792.1	zf-CCHC	No hits	0	N	N/A
NP_191800.1	CP12	No hits	0	C	N/A
NP_191869.1	No hits	No hits	0	C	N/A
NP_192387.1	zf-GRF	No hits	1	T	N/A
NP_192445.1	zf-CCHC	No hits	0	N	N/A
NP_192594.2	No hits	No hits	0	C	N/A
NP_192933.2	No hits	No hits	3	T	N/A
NP_192957.1	zf-CCHC	No hits	0	N	N/A
NP_193133.3	SWIM	No hits	0	C	N/A
NP_193471.1	No hits	No hits	1	T	N/A
NP_194274.2	zf-CCCH	No hits	0	C	N/A
NP_194418.1	Stig1	No hits	1	T	N/A
NP_194504.2	Yippee	No hits	0	N	N/A

NP_194628.1	zf-RanBP	No hits	0	N	N/A
NP_194895.1	zf-HIT	No hits	2	T	N/A
NP_195501.1	No hits	COG5109	0	N	N/A
NP_195616.2	Kinesin	No hits	0	C	N/A
NP_195841.1	GCK	No hits	0	C	N/A
NP_195948.1	Nramp	No hits	12	T	N/A
NP_196259.2	DnaJ HSCB_C	No hits	0	N	N/A
NP_196525.1	No hits	COG5109	0	N	N/A
NP_197073.1	zf-GRF	No hits	0	N	N/A
NP_197147.1	C1_3	No hits	0	C	N/A
NP_197327.1	zf-C3HC4	No hits	0	N	N/A
NP_197702.1	zf-C3HC4	No hits	0	N	N/A
NP_198209.1	zf-GRF	No hits	1	T	N/A
NP_198904.1	zf-CCCH	No hits	0	C	N/A
NP_198935.1	No hits	COG1439	2	T	N/A
NP_199175.1	No hits	No hits	0	N	N/A
NP_199554.1	U-box	No hits	1	T	N/A
NP_199731.1	zf-CCCH	No hits	0	C	N/A
NP_199856.1	Sel1 zf-MYND	No hits	0	N	N/A
NP_200011.1	zf-CCCH	No hits	0	C	N/A
NP_200205.1	Yippee	No hits	0	N	N/A
NP_200322.1	Stig1	No hits	1	T	N/A
NP_200565.1	GCK	No hits	0	C	N/A
NP_200572.1	GCK	No hits	0	C	N/A
NP_200742.2	Homeobox	No hits	0	N	N/A
NP_201348.1	zf-MYND UCH	No hits	1	T	N/A
NP_207185.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_207750.1	DUF164	COG1579	1	T	N/A
NP_208197.1	Radical_SAM	No hits	0	C	N/A
NP_212148.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_212633.1	Ribosomal_L36	No hits	0	C	N/A
NP_212795.1	No hits	No hits	1	T	N/A
NP_214073.1	No hits	No hits	3	T	N/A
NP_214206.1	No hits	COG1198	0	C	N/A
NP_219908.1	DUF164	COG1579	0	C	N/A
NP_219956.1	Chlam_OMP3	No hits	1	T	N/A
NP_220297.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_220913.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_227959.1	No hits	COG2331	0	C	N/A
NP_227993.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_228409.1	PHP	No hits	1	T	N/A
NP_228435.1	Phosphodiest	No hits	2	T	N/A
NP_228904.1	Abi	No hits	7	T	N/A
NP_229276.1	Ribosomal_L36	No hits	0	C	N/A
NP_229442.1	OTCace_N OTCace Pyrl Pyrl_C	No hits	0	C	N/A
NP_247390.1	No hits	No hits	0	C	N/A
NP_247569.1	zf-C2H2	COG4049	0	C	N/A
NP_247603.1	Radical_SAM	No hits	0	C	N/A
NP_247759.1	Arch_ATPase	No hits	1	T	N/A
NP_248030.1	Zip	COG0428	8	T	N/A
NP_248086.1	Radical_SAM	No hits	0	C	N/A
NP_248133.1	DUF116	No hits	2	T	N/A

NP_248399.1	Lectin_legB	No hits	1	T	N/A
NP_248500.1	SWIM	No hits	0	C	N/A
NP_248660.1	NMD3	COG1499	0	C	N/A
NP_249217.1	No hits	No hits	1	T	N/A
NP_249725.1	SEC-C	No hits	0	C	N/A
NP_249841.1	Pyocin_S HNH	No hits	0	C	N/A
NP_249880.1	DUF335	COG3091	0	C	N/A
NP_250041.1	DUF899	COG4312	1	T	N/A
NP_250540.1	DUF1272	COG3813	0	C	N/A
NP_250584.1	Penicil_amidase	No hits	1	T	N/A
NP_250830.1	Metallothio_Pro	No hits	0	C	N/A
NP_250950.1	AP_endonuc_2	No hits	0	C	N/A
NP_251817.1	Glyoxalase	No hits	0	C	N/A
NP_251941.1	Phosphodiest	No hits	0	C	N/A
NP_251973.1	DUF692	COG3220	0	C	N/A
NP_252795.1	DUF692	COG3220	0	C	N/A
NP_252926.1	Catalase	No hits	0	C	N/A
NP_253737.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_268809.1	AP_endonuc_2	No hits	0	C	N/A
NP_268843.1	DUF335	COG3091	0	C	N/A
NP_269679.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_273222.1	Ribosomal_L36	No hits	0	C	N/A
NP_273596.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_274087.1	OpcA	No hits	0	C	N/A
NP_274104.1	SEC-C	No hits	0	C	N/A
NP_274675.1	PqiA PqiA	COG2995	9	T	N/A
NP_275127.1	DUF692	COG3220	0	C	N/A
NP_275751.1	Rib_5-P_isom_A	No hits	0	C	N/A
NP_276409.1	PadR	No hits	0	C	N/A
NP_276461.1	FwdE	No hits	0	C	N/A
NP_276468.1	Rubredoxin	No hits	1	T	N/A
NP_276820.1	Prismane	No hits	0	C	N/A
NP_276874.1	NMD3	COG1499	0	C	N/A
NP_277011.1	PHP	No hits	0	C	N/A
NP_279326.1	No hits	No hits	0	C	N/A
NP_279412.1	No hits	No hits	0	C	N/A
NP_279481.1	Peptidase_M48	No hits	7	T	N/A
NP_279864.1	No hits	No hits	1	T	N/A
NP_279923.1	SWIM	No hits	0	C	N/A
NP_280285.1	No hits	No hits	0	C	N/A
NP_280419.1	Ribonuc_red_lgN Ribonuc_red_lgC	No hits	0	C	N/A
NP_280495.1	Zip	COG0428	8	T	N/A
NP_280642.1	Radical_SAM	No hits	0	C	N/A
NP_280836.1	MCM	No hits	0	C	N/A
NP_280951.1	Abi	No hits	4	T	N/A
NP_280952.1	Abi	No hits	2	T	N/A
NP_281035.1	No hits	No hits	0	C	N/A
NP_281068.1	NMD3	COG1499	0	C	N/A
NP_295064.1	Glyoxalase	No hits	1	C	N/A
NP_295303.1	Glyoxalase	No hits	0	C	N/A
NP_295604.1	No hits	No hits	0	C	N/A
NP_296325.1	No hits	COG1198	0	C	N/A
NP_346953.1	No hits	No hits	0	C	N/A
NP_347114.1	No hits	No hits	0	C	N/A

NP_347655.1	Zn_dep_PLPC	No hits	0	C	N/A
NP_348347.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_349893.1	GerA	No hits	5	T	N/A
NP_375843.1	No hits	No hits	0	C	N/A
NP_376053.1	Ribosomal_L40e	No hits	0	C	N/A
NP_376108.1	SWIM	No hits	0	C	N/A
NP_376168.1	No hits	No hits	1	T	N/A
NP_376393.1	No hits	No hits	0	C	N/A
NP_376404.1	NMD3	COG1499	0	C	N/A
NP_376732.1	zf-RanBP	COG1716	0	C	N/A
NP_377039.1	HTH_5 AsnC_trans_reg	No hits	2	T	N/A
NP_377088.1	No hits	No hits	1	T	N/A
NP_377648.1	No hits	COG1716	4	T	N/A
NP_377723.1	SWIM	COG4715	0	C	N/A
NP_377761.1	No hits	No hits	0	C	N/A
NP_377839.1	SWIM	COG4715	0	C	N/A
NP_377930.1	MerR Peptidase_M48	No hits	0	C	N/A
NP_378614.1	zf-RanBP	COG1716	4	T	N/A
NP_388076.1	Abi	No hits	6	T	N/A
NP_388135.1	No hits	No hits	0	C	N/A
NP_388360.1	DUF335	COG3091	0	C	N/A
NP_388420.1	HTH_5	No hits	0	C	N/A
NP_388500.1	TFIIB_Zn_Ribbon	No hits	2	T	N/A
NP_388900.1	No hits	No hits	0	C	N/A
NP_388920.1	DAO Rieske	No hits	0	C	N/A
NP_389243.1	Put_Phosphatase	No hits	0	C	N/A
NP_389453.1	ResIII DEAD Helicase_C	COG1198	1	T	N/A
NP_389563.1	GntR UTRA	No hits	0	C	N/A
NP_389875.1	Metallophos	No hits	0	C	N/A
NP_390312.1	NusB	No hits	0	C	N/A
NP_391268.1	No hits	No hits	1	T	N/A
NP_391419.1	No hits	No hits	0	C	N/A
NP_391508.1	SWIM	No hits	0	C	N/A
NP_391923.1	Glyoxalase	No hits	0	C	N/A
NP_393503.1	No hits	No hits	0	C	N/A
NP_393654.1	Rpr2	No hits	0	C	N/A
NP_394448.1	NMD3	COG1499	0	C	N/A
NP_394467.1	No hits	No hits	1	T	N/A
NP_414944.1	HNH	No hits	0	C	N/A
NP_415470.1	PqiA PqiA	COG2995	8	T	N/A
NP_415749.1	SEC-C	No hits	0	C	N/A
NP_416347.1	PqiA PqiA	COG2995	8	T	N/A
NP_416469.1	Vsr	No hits	0	C	N/A
NP_416476.1	DJ-1_PfpI	No hits	0	C	N/A
NP_416626.3	SWIM	No hits	0	C	N/A
NP_416922.1	HTH_6 SIS	No hits	0	C	N/A
NP_417419.1	DUF335	COG3091	0	C	N/A
NP_418370.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_419048.1	Abi	COG1266	7	T	N/A
NP_419801.1	Peptidase_M14	No hits	1	T	N/A
NP_421528.1	No hits	No hits	0	C	N/A
NP_421629.1	DUF1272	COG3813	0	C	N/A
NP_421700.1	DUF692	COG3220	0	C	N/A
NP_421868.1	Abi	COG4449	3	T	N/A
NP_421945.1	Peptidase_M56	No hits	3	T	N/A

NP_422049.1	DUF692	COG3220	0	C	N/A
NP_422276.1	SWIM	No hits	0	C	N/A
NP_422384.1	No hits	No hits	0	C	N/A
NP_428268.1	FCH DEP RhoGAP	No hits	0	C	N/A
NP_438503.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_439331.1	DUF335	COG3091	0	C	N/A
NP_439444.1	No hits	No hits	0	C	N/A
NP_439492.1	No hits	No hits	0	C	N/A
NP_439742.1	DUF692	COG3220	0	C	N/A
NP_439953.1	Metallophos	No hits	0	C	N/A
NP_441019.1	CP12	No hits	0	C	N/A
NP_441277.1	No hits	No hits	1	T	N/A
NP_441614.1	Abi	No hits	5	T	N/A
NP_441677.1	HypA	No hits	0	C	N/A
NP_441816.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_441935.1	Peptidase_M10	No hits	1	T	N/A
NP_442594.1	FHA	No hits	0	C	N/A
NP_442605.1	No hits	No hits	4	T	N/A
NP_442782.1	AAA DUF815	No hits	0	C	N/A
NP_442832.1	SWIM	COG4279	0	C	N/A
NP_442915.1	Abi	COG4449	6	T	N/A
NP_443160.1	zf-MYND	No hits	0	N	N/A
NP_443189.1	DUF335	No hits	1	T	N/A
NP_476672.1	PHD	No hits	1	T	N/A
NP_476737.3	zf-RanBP	No hits	0	N	N/A
NP_476917.1	No hits	No hits	0	C	N/A
NP_477317.1	PDZ C1_1	No hits	0	C	N/A
NP_477332.1	NMD3	COG1499	0	C	N/A
NP_477493.2	No hits	No hits	1	T	N/A
NP_484460.1	No hits	No hits	0	C	N/A
NP_484591.1	PP2C	No hits	0	C	N/A
NP_484772.1	Response_reg GAF PAS PAS_4 PAS_3 GAF	No hits	0	C	N/A
NP_484789.1	Glyoxalase	No hits	1	C	N/A
NP_484830.1	Abi	No hits	4	T	N/A
NP_484948.1	CP12	No hits	0	C	N/A
NP_485643.1	FHA	No hits	1	T	N/A
NP_485752.1	Abi	COG4449	6	T	N/A
NP_486113.1	ArgJ	No hits	0	C	N/A
NP_486117.1	No hits	No hits	2	T	N/A
NP_486181.1	Abi	COG1266	8	T	N/A
NP_486531.1	HTH_5 HTH_11 HTH_DeoR	No hits	0	C	N/A
NP_486890.1	CP12	No hits	0	C	N/A
NP_488195.1	SWIM	COG4279	0	C	N/A
NP_488288.1	LexA_DNA_bind ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_488475.1	No hits	No hits	5	T	N/A
NP_488784.1	SWIM	COG4715	0	C	N/A
NP_489246.1	RVT_1 GIIM HNH	No hits	0	C	N/A
NP_489375.1	No hits	No hits	0	C	N/A
NP_490701.1	No hits	No hits	1	T	N/A
NP_490719.1	Ank zf-MYND	No hits	0	N	N/A
NP_490824.1	zf-C2H2	No hits	0	N	N/A
NP_491022.2	DUF1399	No hits	0	C	N/A

NP_491090.1	No hits	COG1439	0	C	N/A
NP_491110.1	No hits	No hits	0	C	N/A
NP_491253.1	No hits	No hits	0	C	N/A
NP_491414.2	No hits	No hits	0	C	N/A
NP_491646.3	No hits	No hits	1	T	N/A
NP_491764.1	C1_1	No hits	0	C	N/A
NP_491838.1	No hits	No hits	0	C	N/A
NP_491976.2	RRM_1	No hits	0	C	N/A
NP_491988.1	No hits	No hits	2	T	N/A
NP_492047.1	No hits	No hits	1	T	N/A
NP_492114.2	NMD3	COG1499	0	C	N/A
NP_492122.2	No hits	No hits	0	C	N/A
NP_492183.2	Activin_recp	No hits	1	T	N/A
NP_492199.2	zf-C3HC4	No hits	2	T	N/A
NP_492391.2	No hits	No hits	0	C	N/A
NP_492553.1	Nicastrin	No hits	3	T	N/A
NP_492632.1	No hits	No hits	0	C	N/A
NP_492772.1	zf-MYND	No hits	0	N	N/A
NP_492879.2	zf-C2H2	No hits	0	N	N/A
NP_492899.1	No hits	No hits	0	C	N/A
NP_493018.1	DUF1280	No hits	1	T	N/A
NP_493224.1	No hits	No hits	0	C	N/A
NP_493359.1	No hits	No hits	0	C	N/A
NP_493432.1	No hits	No hits	0	C	N/A
NP_493433.1	No hits	No hits	0	C	N/A
NP_493620.2	zf-MYND	No hits	0	N	N/A
NP_493629.2	No hits	No hits	0	C	N/A
NP_494124.1	MATH BTB	No hits	0	C	N/A
NP_494346.1	No hits	No hits	1	T	N/A
NP_494369.1	No hits	No hits	0	C	N/A
NP_494509.1	No hits	No hits	1	T	N/A
NP_494701.3	zf-C2H2	No hits	0	N	N/A
NP_494874.1	No hits	No hits	1	T	N/A
NP_494936.1	No hits	No hits	1	T	N/A
NP_494943.1	No hits	No hits	1	T	N/A
NP_494944.1	No hits	No hits	1	T	N/A
NP_495429.2	zf-C3HC4	No hits	0	N	N/A
NP_495656.1	No hits	No hits	1	T	N/A
NP_495679.1	ARID	No hits	1	T	N/A
NP_495859.1	No hits	No hits	1	T	N/A
NP_496101.1	Extensin_2 zf-nanos	No hits	0	C	N/A
NP_496102.1	Extensin_2 zf-nanos	No hits	0	C	N/A
NP_496274.1	No hits	No hits	1	T	N/A
NP_496323.1	zf-MYND	No hits	0	N	N/A
NP_496368.1	No hits	No hits	2	T	N/A
NP_496395.1	No hits	No hits	0	C	N/A
NP_496539.1	No hits	COG1196	1	T	N/A
NP_496806.2	No hits	No hits	0	C	N/A
NP_496808.1	Peptidase_A17 rve	No hits	0	N	N/A
NP_496811.1	No hits	No hits	0	C	N/A
NP_496835.1	No hits	No hits	0	C	N/A
NP_496852.1	No hits	No hits	1	T	N/A
NP_496911.1	No hits	No hits	0	C	N/A
NP_497090.1	No hits	No hits	0	C	N/A
NP_497136.1	AT_hook	No hits	1	T	N/A
NP_497410.1	No hits	No hits	1	T	N/A
NP_497624.1	No hits	No hits	2	T	N/A
NP_497688.2	RWD zf-C3HC4	No hits	0	C	N/A

NP_497785.1	SAM_2 SAM_1 SAM_2 TIR	No hits	0	C	N/A
NP_497796.1	Yippee	No hits	0	N	N/A
NP_497863.1	No hits	No hits	0	C	N/A
NP_497887.1	No hits	No hits	0	C	N/A
NP_497896.1	zf-MYND PDCD2_C	No hits	0	N	N/A
NP_497992.1	No hits	No hits	0	C	N/A
NP_498049.2	FLYWCH	No hits	0	N	N/A
NP_498067.2	Polyketide_cyc	No hits	0	C	N/A
NP_498124.2	zf-C2H2	No hits	0	N	N/A
NP_498497.1	GATA	No hits	0	N	N/A
NP_498747.2	No hits	No hits	2	T	N/A
NP_498827.1	No hits	No hits	0	C	N/A
NP_498846.1	No hits	No hits	0	C	N/A
NP_499137.1	No hits	No hits	0	C	N/A
NP_499237.1	No hits	No hits	0	C	N/A
NP_499330.2	No hits	No hits	0	C	N/A
NP_499430.1	No hits	No hits	0	C	N/A
NP_499432.1	No hits	No hits	0	C	N/A
NP_499474.1	zf-BED	No hits	0	N	N/A
NP_499558.1	Toprim Topoisom_bac zf-GRF	No hits	1	T	N/A
NP_499727.1	No hits	No hits	0	C	N/A
NP_499772.1	Kunitz_BPTI	No hits	1	T	N/A
NP_499873.1	No hits	No hits	0	C	N/A
NP_500320.1	No hits	No hits	0	C	N/A
NP_500335.1	Yippee	No hits	0	N	N/A
NP_500488.1	ShTK ShTK ShTK ShTK	No hits	1	T	N/A
NP_500903.1	No hits	No hits	0	C	N/A
NP_501162.1	No hits	No hits	0	C	N/A
NP_501222.1	No hits	No hits	0	C	N/A
NP_501344.1	Dynactin_p62	No hits	0	C	N/A
NP_501499.1	No hits	No hits	0	C	N/A
NP_501546.1	zf-CCHC	No hits	0	N	N/A
NP_501597.1	No hits	No hits	0	C	N/A
NP_501618.2	FLYWCH	No hits	1	T	N/A
NP_501619.2	FLYWCH	No hits	0	N	N/A
NP_501697.1	Peptidase_A17 rve	No hits	1	T	N/A
NP_501785.1	zf-BED	No hits	1	T	N/A
NP_501814.2	No hits	No hits	0	C	N/A
NP_501957.1	No hits	No hits	1	T	N/A
NP_502173.2	No hits	No hits	1	T	N/A
NP_502290.1	No hits	No hits	0	C	N/A
NP_502429.1	No hits	No hits	1	T	N/A
NP_502562.1	No hits	No hits	0	C	N/A
NP_502597.1	No hits	No hits	0	C	N/A
NP_502750.1	No hits	No hits	1	T	N/A
NP_502841.1	No hits	No hits	2	T	N/A
NP_502868.1	No hits	No hits	0	C	N/A
NP_503022.1	PHD	No hits	2	T	N/A
NP_503139.1	No hits	No hits	0	C	N/A
NP_503172.1	PHD	No hits	0	N	N/A
NP_504367.1	Srb	No hits	7	T	N/A
NP_504391.1	No hits	No hits	0	N	N/A
NP_504696.1	Metallothio_2	No hits	0	C	N/A

NP_504806.1	LysM LysM LysM Self-incomp_S1 Glyco_hydro_18	COG3979	0	C	N/A
NP_504862.1	LysM LysM LysM LysM LysM LysM LysM LysM LysM LysM LysM LysM Self-incomp_S1 Glyco_hydro_18	COG3979	0	C	N/A
NP_504916.1	DUF316 Trypsin	No hits	0	C	N/A
NP_505031.1	No hits	No hits	1	T	N/A
NP_505182.1	PHD	No hits	0	C	N/A
NP_505184.1	Peptidase_A17 rve	No hits	0	N	N/A
NP_505271.1	No hits	No hits	0	C	N/A
NP_505273.1	Peptidase_A17 rve	No hits	0	N	N/A
NP_505285.1	No hits	No hits	0	C	N/A
NP_505492.1	No hits	No hits	0	C	N/A
NP_505663.2	TTL	No hits	1	T	N/A
NP_505769.2	ELM2 Myb_DNA-binding zf-C2H2	No hits	0	N	N/A
NP_505802.1	7tm_1	No hits	8	T	N/A
NP_505973.1	zf-C3HC4	No hits	2	T	N/A
NP_506163.1	zf-CCCH	No hits	0	N	N/A
NP_506288.1	DM	No hits	0	N	N/A
NP_506289.1	DM	No hits	0	N	N/A
NP_506316.1	No hits	No hits	0	N	N/A
NP_506317.2	No hits	No hits	0	N	N/A
NP_506320.3	zf-C2H2	No hits	0	N	N/A
NP_506516.1	No hits	No hits	0	C	N/A
NP_506647.1	No hits	No hits	1	T	N/A
NP_506653.1	No hits	No hits	2	T	N/A
NP_506675.1	No hits	No hits	0	C	N/A
NP_506703.1	No hits	No hits	2	T	N/A
NP_506835.1	TIL	No hits	0	C	N/A
NP_507339.1	No hits	No hits	0	C	N/A
NP_507450.1	No hits	No hits	0	C	N/A
NP_507451.1	No hits	No hits	2	T	N/A
NP_507558.1	ShTK	No hits	0	C	N/A
NP_507610.1	No hits	No hits	0	C	N/A
NP_507643.2	TIL TIL TIL TIL	No hits	0	C	N/A
NP_507645.2	TIL TIL TIL	No hits	0	C	N/A
NP_507780.1	DNA_pol_B_2	No hits	0	C	N/A
NP_507936.1	No hits	No hits	0	C	N/A
NP_507938.1	No hits	No hits	0	C	N/A
NP_507998.1	SAM_1	No hits	0	C	N/A
NP_508158.2	No hits	No hits	1	T	N/A
NP_508305.1	No hits	No hits	2	T	N/A
NP_508353.1	Endonuclease_7	No hits	2	T	N/A
NP_508472.1	zf-CCHC	No hits	0	N	N/A
NP_508531.1	No hits	No hits	0	C	N/A
NP_508532.1	No hits	No hits	0	C	N/A
NP_508646.1	RVT_1 Peptidase_A17 RnaseH rve	No hits	1	T	N/A
NP_508850.2	zf-MYND	No hits	1	T	N/A
NP_508879.1	zf-C2H2	No hits	1	T	N/A
NP_508940.3	No hits	No hits	0	C	N/A
NP_508967.1	zf-RanBP	No hits	0	N	N/A

NP_508983.2	THAP 2-Hacid_dh 2-Hacid_dh_C	No hits	2	T	N/A
NP_509216.1	No hits	No hits	0	C	N/A
NP_509552.1	Peptidase_A17	No hits	0	N	N/A
NP_509661.1	PHD	No hits	0	N	N/A
NP_509662.1	No hits	No hits	0	C	N/A
NP_510374.1	No hits	No hits	0	C	N/A
NP_510587.1	THAP	No hits	1	T	N/A
NP_510645.1	No hits	No hits	0	C	N/A
NP_510794.1	No hits	No hits	0	C	N/A
NP_510821.1	TSP_1	No hits	1	T	N/A
NP_518370.1	Peptidase_M23	No hits	0	C	N/A
NP_518585.1	No hits	COG2331	0	C	N/A
NP_519060.1	No hits	No hits	0	C	N/A
NP_519234.1	DUF899	COG4312	1	T	N/A
NP_520613.1	Metallophos	No hits	1	T	N/A
NP_521300.1	DUF746 DUF746	No hits	0	C	N/A
NP_521421.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_523665.2	zf-CCCH	No hits	0	N	N/A
NP_523726.2	zf-C3HC4	No hits	1	T	N/A
NP_523747.1	zf-BED	No hits	0	N	N/A
NP_523766.1	zf-C2H2	No hits	0	N	N/A
NP_524224.2	THAP	No hits	0	N	N/A
NP_524253.1	MSSP	No hits	0	C	N/A
NP_524254.1	MSSP	No hits	0	C	N/A
NP_524768.2	zf-MYND	No hits	0	N	N/A
NP_525081.1	Glyco_hydro_20b Glyco_hydro_20	No hits	1	T	N/A
NP_536786.1	Tsg	No hits	1	T	N/A
NP_558746.1	zf-C2H2	No hits	0	C	N/A
NP_558854.1	TFIIB_Zn_Ribbon	No hits	0	C	N/A
NP_558868.1	zf-RanBP	COG1716	0	C	N/A
NP_559059.1	Transposase_35	No hits	0	C	N/A
NP_559166.1	SWIM	No hits	0	C	N/A
NP_559282.1	SWIM	No hits	0	C	N/A
NP_559336.1	DUF1610	No hits	0	C	N/A
NP_559451.1	NMD3	COG1499	0	C	N/A
NP_559778.1	TFIIB_Zn_Ribbon	No hits	1	T	N/A
NP_560012.1	Abi	No hits	3	T	N/A
NP_560328.1	No hits	No hits	0	C	N/A
NP_560568.1	Transposase_35	No hits	1	T	N/A
NP_560885.1	TFIIB_Zn_Ribbon	No hits	3	T	N/A
NP_563644.1	No hits	No hits	0	N	N/A
NP_563718.1	UBA	No hits	0	C	N/A
NP_563840.1	FSH1	COG1054	0	C	N/A
NP_563856.1	Acid_phosphat_A	No hits	0	C	N/A
NP_564704.2	WLM zf-RanBP zf-RanBP	No hits	0	C	N/A
NP_564894.1	zf-MYND	No hits	0	N	N/A
NP_565108.2	zf-CCCH	No hits	0	N	N/A
NP_565134.1	CP12	No hits	0	C	N/A
NP_565541.1	No hits	COG5109	0	N	N/A
NP_565576.1	zf-MYND UCH	No hits	2	T	N/A
NP_565578.1	No hits	No hits	0	C	N/A
NP_565848.1	DUF618	No hits	0	N	N/A
NP_565858.1	PLAC8	No hits	1	T	N/A
NP_566100.2	CP12	No hits	0	C	N/A

NP_566389.1	Yippee	No hits	0	N	N/A
NP_566691.1	zf-CCCH	No hits	0	N	N/A
NP_567225.1	zf-MYND PDCD2_C	No hits	0	N	N/A
NP_567672.1	No hits	No hits	2	T	N/A
NP_568119.1	No hits	No hits	0	C	N/A
NP_569080.1	IGFBP	No hits	0	C	N/A
NP_569840.1	No hits	No hits	0	C	N/A
NP_569947.2	DUF335	No hits	0	C	N/A
NP_570037.1	No hits	No hits	1	T	N/A
NP_570852.1	zf-C2H2	COG4049	0	C	N/A
NP_572344.1	zf-AD	No hits	0	N	N/A
NP_572348.1	zf-C2H2	No hits	1	T	N/A
NP_572479.1	DUF753	No hits	2	T	N/A
NP_572539.2	No hits	No hits	1	T	N/A
NP_572603.1	No hits	COG1439	1	T	N/A
NP_572609.1	Yippee	No hits	0	N	N/A
NP_572623.1	No hits	No hits	1	T	N/A
NP_572685.2	zf-C2H2	No hits	0	N	N/A
NP_572692.1	PHD	No hits	1	T	N/A
NP_572809.1	No hits	No hits	0	N	N/A
NP_572823.1	BTB	No hits	1	T	N/A
NP_572882.1	Yippee	No hits	0	N	N/A
NP_572888.2	SET WW	No hits	0	C	N/A
NP_573014.2	DUF1740	No hits	1	T	N/A
NP_573181.1	No hits	COG1196	0	C	N/A
NP_573209.1	No hits	No hits	0	C	N/A
NP_573304.1	No hits	No hits	0	C	N/A
NP_573387.1	PWWP	No hits	1	T	N/A
NP_603650.1	GatB_N	No hits	0	C	N/A
NP_604053.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_608409.1	Acyltransferase	No hits	2	T	N/A
NP_608511.1	No hits	No hits	0	C	N/A
NP_608582.2	RRM_1 zf-RanBP	No hits	1	T	N/A
NP_608644.1	No hits	No hits	0	C	N/A
NP_608744.1	zf-C3HC4	No hits	0	N	N/A
NP_608838.1	zf-AD	No hits	0	N	N/A
NP_608971.1	No hits	No hits	0	C	N/A
NP_608988.1	PPI_Ypi1	No hits	0	C	N/A
NP_609051.1	DPPIV_N	No hits	1	T	N/A
NP_609072.1	IBR	No hits	2	T	N/A
NP_609317.1	No hits	No hits	0	C	N/A
NP_609360.2	No hits	No hits	0	C	N/A
NP_609485.2	WD40 WD40 WD40	No hits	0	C	N/A
NP_609495.1	Mov34	No hits	1	T	N/A
NP_609568.3	OATP	No hits	12	T	N/A
NP_609652.2	No hits	No hits	0	C	N/A
NP_609889.1	C1_3	No hits	0	C	N/A
NP_609998.1	zf-C2H2	No hits	0	C	N/A
NP_610136.1	zf-MYND	No hits	0	N	N/A
NP_610202.3	zf-MYND	No hits	0	N	N/A
NP_610311.1	Dynactin_p62	No hits	0	C	N/A
NP_610529.1	No hits	No hits	0	C	N/A
NP_610532.1	No hits	No hits	0	C	N/A
NP_610684.1	zf-MYND	No hits	0	N	N/A
NP_610730.1	zf-MYND	No hits	0	N	N/A
NP_610915.2	RWD zf-C3HC4	No hits	0	C	N/A
NP_610944.1	zf-MYND	No hits	0	N	N/A

NP_611012.1	No hits	No hits	0	C	N/A
NP_611181.1	No hits	No hits	0	N	N/A
NP_611408.2	CUE zf-RanBP	No hits	0	N	N/A
NP_611426.1	No hits	No hits	0	C	N/A
NP_611536.1	No hits	COG5109	0	N	N/A
NP_611557.2	DUF618 RRM_1	No hits	2	T	N/A
NP_611612.1	DUF753 Activin_recp DUF753 DUF753 DUF753 DUF753 Activin_recp DUF753	No hits	0	C	N/A
NP_611703.1	No hits	No hits	0	C	N/A
NP_611889.1	No hits	No hits	1	T	N/A
NP_611890.1	zf-MYND PDCD2_C	No hits	0	N	N/A
NP_612000.2	DC_STAMP	No hits	5	T	N/A
NP_612007.1	PHD	No hits	2	T	N/A
NP_612054.1	THAP	No hits	1	T	N/A
NP_612471.1	zf-MYND	No hits	0	N	N/A
NP_612569.1	No hits	No hits	0	C	N/A
NP_613386.1	No hits	No hits	0	C	N/A
NP_613802.1	NMD3	COG1499	0	C	N/A
NP_613862.1	zf-C2H2	COG4049	0	C	N/A
NP_613884.1	Met_10	No hits	0	C	N/A
NP_614209.1	Mur_ligase_M	No hits	0	C	N/A
NP_614218.1	dCMP_cyt_deam_1	No hits	0	C	N/A
NP_614557.1	HTH_5	No hits	0	C	N/A
NP_615163.1	SWIM	COG4279	0	C	N/A
NP_615164.1	HTH_5 DUF1724	COG4742	0	C	N/A
NP_615315.1	Aminotran_1_2	No hits	0	C	N/A
NP_615427.1	SWIM	COG4715	0	C	N/A
NP_615430.1	SWIM	COG4715	0	C	N/A
NP_615682.1	APH	No hits	0	C	N/A
NP_615704.1	DnaJ_CXXCXGXG	No hits	0	C	N/A
NP_615870.1	Peptidase_M50	No hits	0	C	N/A
NP_617110.1	AAA AAA_3 AAA_5	No hits	6	T	N/A
NP_617116.1	zf-AN1	No hits	0	C	N/A
NP_617511.1	Abi	No hits	0	C	N/A
NP_617550.1	Phosphodiest	No hits	8	T	N/A
NP_617812.1	DUF434	No hits	1	T	N/A
NP_618144.1	Metalloenzyme	No hits	0	C	N/A
NP_618254.1	No hits	No hits	1	T	N/A
NP_618608.1	HTH_5	No hits	0	C	N/A
NP_618740.1	NMD3	COG1499	0	C	N/A
NP_618796.1	YHS Fer4 FrhB_FdhB_N FrhB_FdhB_C	No hits	0	C	N/A
NP_618881.1	DUF1724	COG4742	0	C	N/A
NP_619245.1	DUF1724	COG4742	0	C	N/A
NP_619380.1	NTP_transf_2	No hits	0	C	N/A
NP_621774.1	Transposase_35	No hits	0	C	N/A
NP_621860.1	Transposase_35	No hits	0	C	N/A
NP_622200.1	Transposase_35	No hits	0	C	N/A
NP_622239.1	Transposase_35	No hits	0	C	N/A
NP_622254.1	Transposase_35	No hits	0	C	N/A
NP_622343.1	Transposase_35	No hits	0	C	N/A
NP_622370.1	HTH_5	No hits	0	C	N/A
NP_622399.1	Transposase_35	No hits	0	C	N/A
NP_622508.1	Transposase_35	No hits	0	C	N/A
NP_622743.1	Transposase_35	No hits	0	C	N/A

NP_622819.1	Transposase_35	No hits	0	C	N/A
NP_623122.1	ResIII DEAD PhnA_Zn_Ribbon	COG1198	0	C	N/A
NP_623167.1	SWIM	COG4715	0	C	N/A
NP_623266.1	Transposase_35	No hits	0	C	N/A
NP_623310.1	Transposase_35	No hits	0	C	N/A
NP_623537.1	Transposase_35	No hits	0	C	N/A
NP_623559.1	Transposase_35	No hits	0	C	N/A
NP_623735.1	No hits	No hits	2	T	N/A
NP_624108.1	Peptidase_M28	No hits	0	C	N/A
NP_624718.1	zf-C3HC4	No hits	1	T	N/A
NP_624949.1	Peptidase_M48 Peptidase_M56	COG0501	4	T	N/A
NP_625154.1	Endonuclease_7	No hits	0	C	N/A
NP_625395.1	Peptidase_M48	COG0501	3	T	N/A
NP_625756.1	No hits	COG1198	0	C	N/A
NP_625920.1	No hits	No hits	0	C	N/A
NP_626902.1	Collagen	No hits	0	C	N/A
NP_626904.1	SEC-C	No hits	0	C	N/A
NP_627084.1	No hits	No hits	0	C	N/A
NP_627401.1	No hits	COG2331	0	C	N/A
NP_627422.1	Clp_N Clp_N	No hits	1	T	N/A
NP_627450.1	AP_endonuc_2	No hits	0	C	N/A
NP_627639.1	No hits	No hits	0	C	N/A
NP_627734.1	No hits	No hits	0	C	N/A
NP_628157.1	SEC-C	No hits	0	C	N/A
NP_628416.1	MarR HTH_5	No hits	1	T	N/A
NP_628612.1	Glyoxalase	No hits	0	C	N/A
NP_628741.1	Peptidase_M48	COG0501	4	T	N/A
NP_629008.1	Peptidase_M14	No hits	0	C	N/A
NP_629221.1	Trypsin	No hits	0	C	N/A
NP_629415.1	Meth_synt_2	No hits	0	C	N/A
NP_629628.1	PIG-L	No hits	0	C	N/A
NP_630136.1	Phosphodiest	No hits	0	C	N/A
NP_630155.1	DUF692	COG3220	0	C	N/A
NP_630451.1	Peptidase_M48 Peptidase_M56	COG0501	3	T	N/A
NP_630608.1	Metallophos	No hits	4	T	N/A
NP_630658.1	AP_endonuc_2	No hits	0	C	N/A
NP_630747.1	No hits	No hits	0	C	N/A
NP_631057.1	HTH_5	No hits	0	C	N/A
NP_631205.1	Abi	No hits	0	C	N/A
NP_636418.1	No hits	No hits	8	T	N/A
NP_638247.2	PDZ	No hits	4	T	N/A
NP_638380.1	No hits	COG2331	0	C	N/A
NP_639110.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_640339.3	Rhodanese	COG1054	0	C	N/A
NP_647653.1	WD40	No hits	0	C	N/A
NP_647733.1	No hits	No hits	1	T	N/A
NP_647742.1	zf-CXXC	No hits	0	N	N/A
NP_647752.1	THAP	No hits	0	N	N/A
NP_647774.1	No hits	No hits	1	T	N/A
NP_648169.2	CBM_14 CBM_14	No hits	1	T	N/A
NP_648303.1	No hits	No hits	1	T	N/A
NP_648372.1	No hits	No hits	3	T	N/A
NP_648392.1	IBR	No hits	0	N	N/A
NP_648527.1	DUF706	No hits	1	T	N/A

NP_648556.2	No hits	No hits	0	C	N/A
NP_648574.1	No hits	No hits	3	T	N/A
NP_648619.1	No hits	No hits	0	C	N/A
NP_648633.1	DC_STAMP	No hits	5	T	N/A
NP_648748.1	DUF1692	No hits	1	T	N/A
NP_648758.1	zf-CCHC	No hits	2	T	N/A
NP_648912.2	LIM	No hits	0	C	N/A
NP_648926.3	No hits	No hits	1	T	N/A
NP_649058.1	No hits	No hits	0	N	N/A
NP_649084.1	zf-MYND	No hits	0	N	N/A
NP_649166.1	No hits	No hits	0	C	N/A
NP_649226.1	Ldl_recept_a	No hits	2	T	N/A
NP_649426.3	Chromo	No hits	0	N	N/A
NP_649593.1	No hits	No hits	0	N	N/A
NP_649714.1	zf-MYND TUDOR TUDOR	No hits	0	N	N/A
NP_649797.2	zf-C3HC4	No hits	0	N	N/A
NP_649941.3	zf-C2H2	No hits	0	C	N/A
NP_650059.1	Mucin	No hits	1	T	N/A
NP_650403.1	THAP THAP	No hits	0	C	N/A
NP_650419.1	zf-CCHC	No hits	1	T	N/A
NP_650515.1	zf-BED	No hits	0	N	N/A
NP_650589.1	zf-MYND	No hits	0	N	N/A
NP_650591.1	No hits	No hits	0	C	N/A
NP_650706.1	No hits	No hits	0	C	N/A
NP_650805.1	HIT	No hits	0	N	N/A
NP_650839.1	FLYWCH	No hits	0	C	N/A
NP_650856.2	ADH_N	COG1063	3	T	N/A
NP_650955.1	No hits	No hits	0	N	N/A
NP_651047.1	PDZ	No hits	0	C	N/A
NP_651110.1	Lectin_leg-like	No hits	0	C	N/A
NP_651224.1	zf-C2H2	No hits	1	T	N/A
NP_651263.1	zf-BED	No hits	1	T	N/A
NP_651739.1	PHD	No hits	0	N	N/A
NP_651872.1	zf-CCHC	No hits	0	N	N/A
NP_652607.1	DUF1671	No hits	1	T	N/A
NP_653169.1	No hits	No hits	2	T	N/A
NP_653265.2	No hits	No hits	0	C	N/A
NP_659005.1	zf-MYND	No hits	0	N	N/A
NP_659445.1	Yippee	No hits	0	N	N/A
NP_659499.1	Churchill	No hits	0	C	N/A
NP_660148.2	DUF134	No hits	0	C	N/A
NP_661843.1	No hits	No hits	1	T	N/A
NP_662078.1	ResIII DEAD Helicase_C	COG1198	1	T	N/A
NP_662971.1	No hits	COG2331	0	C	N/A
NP_671735.1	No hits	No hits	0	C	N/A
NP_671791.1	No hits	No hits	2	T	N/A
NP_671838.1	zf-GRF	No hits	0	N	N/A
NP_680175.2	zf-CCHC	No hits	1	T	N/A
NP_680282.1	No hits	No hits	0	C	N/A
NP_680587.1	zf-GRF	No hits	1	T	N/A
NP_680598.2	No hits	No hits	0	N	N/A
NP_680664.1	zf-GRF	No hits	1	T	N/A
NP_683273.1	No hits	No hits	0	C	N/A
NP_683301.1	zf-GRF	No hits	1	T	N/A
NP_683506.1	zf-GRF	No hits	1	T	N/A
NP_683614.1	zf-GRF	No hits	0	N	N/A

NP_689808.2	No hits	No hits	0	C	N/A
NP_689873.1	No hits	No hits	0	C	N/A
NP_694585.1	DapB_N Peptidase_M9_N	No hits	1	T	N/A
NP_696014.1	Peptidase_M9 PPC PPC PPC	No hits	0	C	N/A
NP_696938.1	No hits	COG1198	0	C	N/A
NP_716272.1	Hpt	No hits	0	C	N/A
NP_716465.1	DUF335	COG3091	0	C	N/A
NP_716947.1	SRP54_N SRP54	No hits	0	C	N/A
NP_716976.1	Polysacc_deac_1	No hits	0	C	N/A
NP_717222.1	No hits	No hits	0	C	N/A
NP_717348.1	HNH	No hits	0	C	N/A
NP_717614.1	DUF692	COG3220	0	C	N/A
NP_718671.1	No hits	No hits	0	C	N/A
NP_718734.1	7tm_1	No hits	7	T	N/A
NP_723123.2	zf-C2H2	No hits	5	T	N/A
NP_723214.1	IBR	No hits	2	T	N/A
NP_723535.1	No hits	No hits	0	C	N/A
NP_723663.1	C1_3 PHD	No hits	0	C	N/A
NP_725094.1	zf-AD	No hits	1	T	N/A
NP_725394.1	zf-AN1	No hits	0	N	N/A
NP_725535.1	No hits	No hits	0	C	N/A
NP_725563.1	RhoGEF	No hits	0	C	N/A
NP_725655.1	No hits	No hits	1	T	N/A
NP_726121.2	DUF753 Activin_recp DUF753 DUF753 DUF753 DUF753 Activin_recp DUF753	No hits	0	C	N/A
NP_726426.1	No hits	No hits	1	T	N/A
NP_726509.1	No hits	No hits	0	C	N/A
NP_726770.2	DUF335	No hits	0	C	N/A
NP_727123.1	zf-C2H2	No hits	1	T	N/A
NP_727617.1	No hits	No hits	1	T	N/A
NP_727620.1	No hits	No hits	0	N	N/A
NP_727839.2	DUF1740	No hits	0	C	N/A
NP_727889.1	IPPT	No hits	0	C	N/A
NP_728761.1	No hits	No hits	1	T	N/A
NP_728762.1	No hits	No hits	1	T	N/A
NP_728763.1	No hits	No hits	1	T	N/A
NP_728764.1	No hits	No hits	1	T	N/A
NP_728765.1	No hits	No hits	1	T	N/A
NP_729575.1	No hits	No hits	3	T	N/A
NP_729576.1	No hits	No hits	3	T	N/A
NP_729606.1	IBR	No hits	0	N	N/A
NP_730192.2	LIM	No hits	0	C	N/A
NP_730762.2	Chromo	No hits	0	N	N/A
NP_730763.2	Chromo	No hits	0	N	N/A
NP_730906.1	zf-MYND 2OG-FeII_Oxy	No hits	0	N	N/A
NP_731012.1	No hits	No hits	0	N	N/A
NP_731014.1	No hits	No hits	0	N	N/A
NP_731402.2	zf-C2H2	No hits	0	C	N/A
NP_732421.1	HIT	No hits	0	N	N/A
NP_732476.2	ADH_N	COG1063	3	T	N/A
NP_733023.2	PHD	No hits	0	C	N/A
NP_733057.1	zf-AD	No hits	0	N	N/A
NP_733073.1	No hits	No hits	0	C	N/A

NP_733543.1	PG_binding_1 Peptidase_M15_3	No hits	1	T	N/A
NP_733665.1	zf-C2H2	No hits	1	T	N/A
NP_740974.1	No hits	No hits	0	C	N/A
NP_741256.1	zf-C2H2	No hits	0	N	N/A
NP_741257.1	No hits	No hits	0	N	N/A
NP_741258.1	No hits	No hits	0	N	N/A
NP_741535.1	DM	No hits	0	N	N/A
NP_741552.1	No hits	No hits	0	C	N/A
NP_741621.1	zf-MYND 2OG-FeII_Oxy	No hits	0	N	N/A
NP_741651.1	No hits	No hits	0	C	N/A
NP_741698.1	No hits	No hits	0	C	N/A
NP_741795.1	EB EB EB EB EB EB EB EB Polysacc_deac_1	No hits	0	C	N/A
NP_741841.1	Polysacc_deac_1	No hits	0	C	N/A
NP_741866.1	GATA	No hits	0	N	N/A
NP_741888.1	No hits	No hits	0	C	N/A
NP_741897.2	MBT MBT MBT zf-C2HC	No hits	0	C	N/A
NP_775735.1	ZU5 Death	No hits	0	N	N/A
NP_775832.1	No hits	No hits	1	T	N/A
NP_783321.2	No hits	No hits	0	C	N/A
NP_788620.1	CH LIM	No hits	1	T	N/A
NP_788621.1	CH LIM	No hits	1	T	N/A
NP_788622.1	CH LIM	No hits	1	T	N/A
NP_788623.1	CH LIM	No hits	1	T	N/A
NP_788624.1	CH LIM	No hits	1	T	N/A
NP_788625.1	CH LIM	No hits	1	T	N/A
NP_788626.1	CH LIM	No hits	1	T	N/A
NP_788735.1	PHD	No hits	0	C	N/A
NP_796632.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_797082.1	No hits	No hits	0	C	N/A
NP_797405.1	No hits	No hits	0	C	N/A
NP_797989.1	PqiA PqiA	COG2995	9	T	N/A
NP_798203.1	No hits	No hits	1	T	N/A
NP_798231.1	No hits	No hits	1	T	N/A
NP_798396.1	PqiA PqiA	COG2995	8	T	N/A
NP_798770.1	HTH_5	No hits	0	C	N/A
NP_798987.1	DUF335	COG3091	0	C	N/A
NP_799394.1	DUF692	COG3220	0	C	N/A
NP_809034.1	FtsX	No hits	0	C	N/A
NP_810128.1	Metallophos	No hits	4	T	N/A
NP_811258.1	Abi	No hits	5	T	N/A
NP_811664.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_811942.1	HTH_AraC	No hits	1	T	N/A
NP_811970.1	SWIM	No hits	0	C	N/A
NP_812817.1	Glyoxalase	No hits	0	C	N/A
NP_813417.1	VWC	No hits	0	C	N/A
NP_817088.1	AAA AAA_5	No hits	0	C	N/A
NP_820411.1	tRNA_U5-meth_tr	No hits	0	C	N/A
NP_820541.1	No hits	COG2331	0	C	N/A
NP_820794.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_840260.1	No hits	No hits	0	C	N/A

NP_841546.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_841879.1	No hits	No hits	0	C	N/A
NP_842253.1	No hits	COG2331	0	C	N/A
NP_848515.1	No hits	No hits	0	C	N/A
NP_849422.1	No hits	No hits	2	T	N/A
NP_849423.1	No hits	No hits	2	T	N/A
NP_849624.1	FSH1	COG1054	0	C	N/A
NP_849659.1	GCK	No hits	0	C	N/A
NP_849969.1	SET zf-MYND	No hits	0	N	N/A
NP_850052.1	zf-CCCH	No hits	0	N	N/A
NP_850059.1	zf-GRF	No hits	1	T	N/A
NP_850221.1	No hits	No hits	0	C	N/A
NP_850430.1	AP_endonuc_2	No hits	0	C	N/A
NP_863965.1	No hits	No hits	0	C	N/A
NP_864538.1	Peptidase_M50	No hits	10	T	N/A
NP_865070.1	No hits	No hits	0	C	N/A
NP_865075.1	Peptidase_M50	No hits	8	T	N/A
NP_865181.1	Abi	No hits	6	T	N/A
NP_865212.1	Metallophos	No hits	1	T	N/A
NP_865278.1	DUF74	COG1716	2	T	N/A
NP_866702.1	Metallophos	No hits	0	C	N/A
NP_866923.1	Glyoxalase	No hits	1	C	N/A
NP_866946.1	TPR_1 TPR_2	No hits	0	C	N/A
NP_867122.1	No hits	No hits	0	C	N/A
NP_867821.1	No hits	No hits	0	C	N/A
NP_867863.1	No hits	No hits	0	C	N/A
NP_868354.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_868565.1	No hits	No hits	5	T	N/A
NP_868647.1	No hits	COG2331	0	C	N/A
NP_868816.1	Peptidase_M50	No hits	8	T	N/A
NP_868966.1	SWIM	COG4715	0	C	N/A
NP_869082.1	SEC-C	No hits	2	T	N/A
NP_869448.1	Peptidase_M50	No hits	7	T	N/A
NP_869978.1	No hits	No hits	0	C	N/A
NP_870173.1	AP_endonuc_2	No hits	0	C	N/A
NP_870350.1	AP_endonuc_2	No hits	0	C	N/A
NP_870560.1	Peptidase_M50	No hits	6	T	N/A
NP_870768.1	No hits	No hits	1	T	N/A
NP_870815.1	No hits	No hits	3	T	N/A
NP_871849.1	zf-MYND	No hits	0	N	N/A
NP_871969.1	No hits	No hits	0	C	N/A
NP_872097.1	Yippee	No hits	0	N	N/A
NP_872161.2	zf-C2H2	No hits	0	N	N/A
NP_872304.2	SAM_2 RhoGAP START	No hits	0	C	N/A
NP_872584.1	No hits	No hits	1	T	N/A
NP_892017.1	No hits	No hits	0	N	N/A
NP_892615.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
NP_938015.1	SET zf-MYND	No hits	0	N	N/A
NP_940988.2	No hits	No hits	1	T	N/A
NP_942148.1	AA_kinase	No hits	0	C	N/A
NP_945352.1	No hits	No hits	0	C	N/A
NP_950406.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A

NP_950419.1	Fer2 Fer4 Molybdop_Fe4S4 Molybdopterin Molydop_binding	No hits	3	T	N/A
NP_951190.1	ResIII DEAD Helicase_C	COG1198	2	T	N/A
NP_951567.1	Glyoxalase	No hits	1	C	N/A
NP_951622.1	No hits	No hits	4	T	N/A
NP_951920.1	No hits	COG2331	0	C	N/A
NP_952264.1	HNH	No hits	0	C	N/A
NP_952376.1	Abi	No hits	0	C	N/A
NP_953040.1	No hits	No hits	6	T	N/A
NP_953215.1	Aldolase_II	No hits	0	C	N/A
NP_953239.1	No hits	No hits	1	T	N/A
NP_954230.1	No hits	No hits	1	T	N/A
NP_954587.1	Peptidase_M10	No hits	0	C	N/A
NP_954590.1	zf-TRAF	No hits	0	N	N/A
NP_963527.1	NMD3	COG1499	0	C	N/A
NP_967280.1	A_deaminase	No hits	0	C	N/A
NP_967336.1	Astacin	No hits	0	C	N/A
NP_967580.1	No hits	No hits	0	C	N/A
NP_968065.1	tRNA-synt_2b	No hits	1	T	N/A
NP_968069.1	No hits	No hits	0	C	N/A
NP_968778.1	No hits	No hits	0	C	N/A
NP_969859.1	DUF692	COG3220	0	C	N/A
NP_970483.1	ResIII DEAD Zot Helicase_C	COG1198	1	T	N/A
NP_970546.1	zf-CCCH	No hits	0	C	N/A
NP_973645.1	Yippee	No hits	0	N	N/A
NP_974088.1	MuDR SWIM	No hits	0	C	N/A
NP_974945.1	No hits	No hits	0	C	N/A
NP_995682.1	WD40 WD40 WD40	No hits	0	C	N/A
NP_995839.1	No hits	No hits	0	C	N/A
NP_995868.1	PDZ C1_1	No hits	0	C	N/A
NP_995942.1	DC_STAMP	No hits	3	T	N/A
NP_995944.1	PHD	No hits	2	T	N/A
XP_039676.5	AAA_2 AAA AAA_5 Bromodomain	COG0464	0	C	N/A
XP_048592.5	No hits	No hits	0	N	N/A
XP_066484.1	No hits	No hits	1	T	N/A
XP_086761.2	I-set	No hits	0	C	N/A
XP_371354.1	Kinesin	No hits	3	T	N/A
XP_371590.3	zf-DBF	No hits	1	T	N/A
XP_931977.1	No hits	No hits	1	T	N/A
XP_931983.1	No hits	No hits	0	C	N/A
XP_933750.1	No hits	No hits	0	N	N/A
XP_934863.1	zf-GRF	No hits	0	C	N/A
XP_935128.1	No hits	No hits	1	T	N/A
XP_935557.1	No hits	No hits	0	C	N/A
XP_936209.1	No hits	No hits	0	C	N/A
XP_937038.1	No hits	No hits	0	C	N/A
XP_937253.1	No hits	No hits	0	C	N/A
XP_937892.1	No hits	No hits	0	C	N/A
XP_941873.1	Kinesin	No hits	3	T	N/A
XP_941897.1	S_100 efhand	No hits	0	C	N/A
XP_942010.1	zf-GRF	No hits	0	C	N/A
XP_942264.1	LIM	No hits	1	T	N/A
XP_942327.1	zf-DBF	No hits	1	T	N/A

XP_942456.1	zf-MYND	No hits	0	N	N/A
XP_942845.1	No hits	No hits	1	T	N/A
XP_943286.1	No hits	No hits	1	T	N/A
XP_943415.1	zf-B_box	No hits	0	N	N/A
XP_943760.1	zf-MYND	No hits	1	T	N/A
XP_944000.1	AAA_2 AAA AAA_5 Bromodomain	COG0464	0	C	N/A
XP_944050.1	C1_1	No hits	0	C	N/A
XP_944714.1	No hits	No hits	0	N	N/A
XP_945726.1	zf-CCCH	No hits	0	N	N/A
XP_945923.1	C1_4	No hits	0	C	N/A
XP_946007.1	ADH_N	No hits	0	C	N/A
XP_946358.1	No hits	No hits	0	N	N/A
XP_946440.1	No hits	No hits	0	C	N/A
XP_946874.1	zf-C2H2	No hits	0	N	N/A
XP_947332.1	I-set	No hits	0	C	N/A
XP_947390.1	zf-C4	No hits	0	N	N/A
XP_948594.1	No hits	No hits	0	C	N/A
XP_948758.1	No hits	No hits	0	C	N/A
XP_948907.1	No hits	No hits	0	C	N/A
XP_949920.1	zf-C3HC4	No hits	0	N	N/A
XP_950904.1	No hits	No hits	0	C	N/A
XP_950975.1	ATP-sulfurylase	No hits	1	T	N/A
YP_004282.1	Abi	No hits	0	C	N/A
YP_004467.1	No hits	No hits	4	T	N/A
YP_005315.1	SWIM	No hits	0	C	N/A
YP_005375.1	PHP Glycos_transf_1	No hits	1	T	N/A
YP_005911.1	No hits	COG1198	0	C	N/A
YP_009318.1	Cytochrom_CIII	No hits	0	C	N/A
YP_009487.1	Rhodanese	No hits	1	T	N/A
YP_009790.1	No hits	No hits	0	C	N/A
YP_009923.1	No hits	No hits	1	T	N/A
YP_010503.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
YP_010699.1	No hits	No hits	1	T	N/A
YP_010744.1	No hits	No hits	0	C	N/A
YP_011424.1	Transketolase_N	No hits	3	T	N/A
YP_011742.1	VanY	No hits	1	T	N/A
YP_012153.1	No hits	COG2331	0	C	N/A
YP_012406.1	No hits	No hits	1	T	N/A
YP_012528.1	Ribosomal_L36	No hits	1	T	N/A
YP_064425.1	PqiA PqiA	COG2995	9	T	N/A
YP_064884.1	No hits	No hits	0	C	N/A
YP_064930.1	Radical_SAM	No hits	0	C	N/A
YP_065942.1	ResIII DEAD DUF1610	COG1198	2	T	N/A
YP_066112.1	No hits	No hits	0	C	N/A
YP_066156.1	Ald_Xan_dh_C Ald_Xan_dh_C2	No hits	1	T	N/A
YP_066448.1	No hits	No hits	4	T	N/A
YP_066580.1	No hits	COG2331	0	C	N/A
YP_066802.1	CMAS	No hits	0	C	N/A
YP_112622.1	Abi	No hits	0	C	N/A
YP_113046.1	No hits	No hits	8	T	N/A
YP_113722.1	No hits	COG2331	0	C	N/A
YP_114215.1	No hits	COG2331	0	C	N/A
YP_115000.1	ResIII DEAD Helicase_C	COG1198	0	C	N/A
YP_115220.1	PHP	No hits	0	C	N/A

YP_115397.1	No hits	No hits	1	T	N/A
YP_115495.1	DUF692	COG3220	0	C	N/A
NP_001014491.1	PB1 ZZ	No hits	0	C	Partial length threading
NP_001021009.1	HMG_box	No hits	0	N	Partial length threading
NP_001021502.1	zf-C2H2	No hits	0	N	Partial length threading
NP_001021503.1	zf-BED zf-C2H2	No hits	0	N	Partial length threading
NP_001023028.1	ZZ	No hits	0	N	Partial length threading
NP_001024227.1	zf-BED	No hits	0	N	Partial length threading
NP_001024228.1	zf-BED	No hits	0	N	Partial length threading
NP_001024910.1	FLYWCH	No hits	0	N	Partial length threading
NP_001024911.1	FLYWCH	No hits	0	N	Partial length threading
NP_001024942.1	zf-CCCH	No hits	0	N	Partial length threading
NP_001030761.1	DUF1644	No hits	0	N	Partial length threading
NP_001031250.1	DUF1644	No hits	0	N	Partial length threading
NP_001031741.1	zf-HIT	No hits	0	C	Partial length threading
NP_004764.1	zf-HIT	No hits	0	C	Partial length threading
NP_005890.2	PB1 ZZ	No hits	1	T	Partial length threading
NP_009729.1	zf-NPL4 NPL4	COG5100	0	C	Partial length threading
NP_009798.1	Zn_clus	No hits	0	N	Partial length threading
NP_012432.1	Zn_clus	No hits	0	N	Partial length threading
NP_012589.1	zf-HIT	No hits	0	C	Partial length threading
NP_012839.1	zf-CHY	COG4357	0	N	Partial length threading
NP_014734.2	zf-CCCH	COG5252	0	N	Partial length threading
NP_014815.1	Zn_clus	No hits	2	T	Partial length threading
NP_015192.1	Zn_clus	No hits	0	N	Partial length threading
NP_060391.1	zf-NPL4 NPL4	COG5100	0	C	Partial length threading
NP_060941.1	zf-CCCH	COG5252	0	N	Partial length threading
NP_070634.1	DUF24	COG1733	0	C	Partial length threading
NP_071226.1	HTH_5	COG1733	0	C	Partial length threading
NP_105467.1	No hits	No hits	0	C	Partial length threading
NP_114064.1	PB1 ZZ	No hits	1	T	Partial length threading
NP_114068.1	PB1 ZZ	No hits	1	T	Partial length threading
NP_147044.1	Zip	COG0428	8	T	Partial length threading
NP_176981.1	DUF1644	No hits	1	T	Partial length threading
NP_177900.2	DUF1644	No hits	0	N	Partial length threading
NP_178139.1	DUF1644	No hits	0	N	Partial length threading
NP_179618.1	zf-CCCH	COG5252	0	N	Partial length threading
NP_180175.1	DUF1644	No hits	0	N	Partial length threading
NP_189118.1	DUF1644	No hits	0	N	Partial length threading
NP_192586.1	DUF1644	No hits	0	N	Partial length threading
NP_194611.2	zf-HIT	No hits	0	C	Partial length threading
NP_200628.2	No hits	No hits	0	C	Partial length threading
NP_229381.1	Phosphodiest	COG1524	3	T	Partial length threading
NP_268578.1	zf-CHY	COG4357	0	C	Partial length threading
NP_273388.1	No hits	COG2956	1	T	Partial length threading
NP_376169.1	DUF701	COG4888	0	C	Partial length threading
NP_376438.1	Phosphodiest	COG1524	0	C	Partial length threading
NP_394805.1	Phosphodiest	COG1524	0	C	Partial length threading
NP_415796.1	TPR_1 TPR_2	COG2956	0	C	Partial length threading
NP_439379.1	TPR_1 TPR_2	COG2956	0	C	Partial length threading
NP_442311.1	Metallophos	No hits	0	C	Partial length threading
NP_476700.1	PB1 ZZ	No hits	0	C	Partial length threading
NP_488714.1	Metallophos	No hits	1	T	Partial length threading
NP_490860.1	ZZ	No hits	0	N	Partial length threading
NP_492148.2	zf-CCCH	COG5252	0	N	Partial length threading
NP_495093.1	zf-NPL4 NPL4 zf-RanBP	COG5100	0	C	Partial length threading
NP_495094.1	zf-NPL4 NPL4 zf-RanBP	COG5100	0	C	Partial length threading
NP_495096.1	zf-NPL4 NPL4 zf-RanBP	COG5100	0	C	Partial length threading

NP_495097.1	zf-NPL4 NPL4 zf-RanBP	COG5100	0	C	Partial length threading
NP_495452.1	zf-nanos	No hits	0	C	Partial length threading
NP_495460.2	zf-nanos	No hits	0	C	Partial length threading
NP_496175.1	zf-C2H2 zf-C2H2	No hits	0	N	Partial length threading
NP_497502.1	zf-C2H2	No hits	0	N	Partial length threading
NP_497850.2	zf-HIT	No hits	0	C	Partial length threading
NP_498227.1	THAP	No hits	0	N	Partial length threading
NP_499579.1	zf-CXXC	No hits	0	N	Partial length threading
NP_499886.1	zf-C2H2	No hits	0	N	Partial length threading
NP_500269.1	zf-C2H2	No hits	0	N	Partial length threading
NP_500489.2	zf-CCHC	No hits	0	N	Partial length threading
NP_502940.1	zf-CCHC	No hits	1	T	Partial length threading
NP_503271.1	C1_1	No hits	0	C	Partial length threading
NP_503273.1	C1_1	No hits	0	C	Partial length threading
NP_505144.2	zf-C2H2	No hits	0	N	Partial length threading
NP_506132.1	zf-C2H2	No hits	0	N	Partial length threading
NP_506487.1	zf-C2H2	No hits	1	T	Partial length threading
NP_507173.1	ZZ	No hits	0	N	Partial length threading
NP_507486.1	C1_1	No hits	0	C	Partial length threading
NP_507535.1	Retrotrans_gag zf-CCHC	No hits	1	T	Partial length threading
NP_507948.1	Retrotrans_gag zf-CCHC	No hits	3	T	Partial length threading
NP_510727.2	zf-C2H2	No hits	0	N	Partial length threading
NP_519033.1	TPR_1 TPR_2	COG2956	1	T	Partial length threading
NP_558704.1	Zip	COG0428	8	T	Partial length threading
NP_560453.1	DUF701	COG4888	1	T	Partial length threading
NP_563977.1	DUF1644	No hits	0	N	Partial length threading
NP_566784.1	DUF1644	No hits	0	N	Partial length threading
NP_567874.1	DUF1644	No hits	0	N	Partial length threading
NP_572387.1	zf-CCCH	No hits	0	N	Partial length threading
NP_610401.1	zf-CCCH	COG5252	0	N	Partial length threading
NP_610955.1	Retrotrans_gag	No hits	1	T	Partial length threading
NP_611037.1	C1_1	No hits	0	C	Partial length threading
NP_611050.1	zf-HIT	No hits	0	C	Partial length threading
NP_611123.2	Trehalase	COG1626	0	C	Partial length threading
NP_617583.1	zf-UBP	No hits	0	C	Partial length threading
NP_629391.1	No hits	No hits	0	C	Partial length threading
NP_629967.1	Phosphodiest	COG1524	0	C	Partial length threading
NP_631351.1	zf-UBP	No hits	0	C	Partial length threading
NP_636460.1	No hits	No hits	0	C	Partial length threading
NP_637547.1	No hits	COG2956	1	T	Partial length threading
NP_648982.1	zf-C2H2	No hits	2	T	Partial length threading
NP_649864.1	JmjC zf-CXXC	No hits	0	N	Partial length threading
NP_651288.1	Bromodomain Bromodomain Bromodomain Bromodomain Bromodomain BAH BAH HMG_box	No hits	0	N	Partial length threading
NP_651406.1	zf-C3HC4	No hits	0	N	Partial length threading
NP_651407.3	zf-NPL4 NPL4	COG5100	0	C	Partial length threading
NP_653205.2	zf-CCCH	No hits	0	N	Partial length threading
NP_663307.1	Retrotrans_gag	No hits	0	N	Partial length threading
NP_696596.1	Phosphodiest	COG1524	0	C	Partial length threading
NP_717989.1	No hits	COG2956	0	C	Partial length threading
NP_725574.2	Trehalase	COG1626	0	C	Partial length threading
NP_725575.1	No hits	COG1626	0	C	Partial length threading
NP_733110.2	zf-NPL4 NPL4	COG5100	0	C	Partial length threading
NP_733111.1	zf-C3HC4	No hits	0	N	Partial length threading

NP_741719.2	zf-CCCH	No hits	0	N	Partial length threading
NP_741720.2	zf-CCCH	No hits	0	N	Partial length threading
NP_798406.1	TPR_2	COG2956	0	C	Partial length threading
NP_819562.1	No hits	COG2956	1	T	Partial length threading
NP_849480.1	DUF1644	No hits	0	N	Partial length threading
NP_850606.1	Rhomboid zf-RanBP	COG0705	3	T	Partial length threading
NP_871624.2	THAP	No hits	0	N	Partial length threading
NP_892797.1	Metallophos	No hits	0	C	Partial length threading
NP_963423.1	Phosphodiest	COG1524	0	C	Partial length threading
NP_973835.1	DUF1644	No hits	0	N	Partial length threading
NP_974163.1	DUF1644	No hits	0	N	Partial length threading
NP_974329.1	Rhomboid zf-RanBP	COG0705	3	T	Partial length threading
NP_974521.1	DUF1644	No hits	0	N	Partial length threading
XP_496034.2	ZZ	No hits	0	N	Partial length threading
XP_931099.1	zf-CCHC	No hits	0	N	Partial length threading
XP_931712.1	zf-CCHC	No hits	0	N	Partial length threading
XP_933214.1	zf-CCHC	No hits	0	N	Partial length threading
XP_940391.1	ZZ	No hits	0	N	Partial length threading
XP_940392.1	ZZ	No hits	1	T	Partial length threading
XP_940393.1	ZZ	No hits	1	T	Partial length threading
XP_941532.1	ZZ	No hits	0	N	Partial length threading
XP_942099.1	zf-CCHC	No hits	0	N	Partial length threading
XP_943743.1	zf-CCHC	No hits	0	N	Partial length threading
YP_010868.1	No hits	COG2956	0	C	Partial length threading
YP_113868.1	TPR_2	COG2956	0	C	Partial length threading
NP_069340.1	Lactamase_B	COG2248	0	C	Total length threading
NP_102843.1	Glyoxalase	COG2764	0	C	Total length threading
NP_103294.1	HNH	No hits	0	C	Total length threading
NP_103841.1	Metallophos	COG0639	0	C	Total length threading
NP_104206.1	Peptidase_M15_2	COG3108	0	C	Total length threading
NP_105877.1	Glyoxalase	COG0346	1	C	Total length threading
NP_106993.1	Peptidase_M15_2	COG3108	1	T	Total length threading
NP_107825.1	Glyoxalase	COG3324	0	C	Total length threading
NP_108186.1	HNH	COG3183	2	T	Total length threading
NP_108396.1	Glyoxalase	No hits	0	C	Total length threading
NP_126655.1	No hits	COG0639	0	C	Total length threading
NP_126657.1	No hits	COG2248	0	C	Total length threading
NP_147282.1	No hits	COG2248	0	C	Total length threading
NP_147659.1	PA Peptidase_M28	COG2234	1	T	Total length threading
NP_214148.1	Metallophos	COG4186	0	C	Total length threading
NP_248449.1	Metallophos	COG4186	0	C	Total length threading
NP_248639.1	No hits	COG2248	0	C	Total length threading
NP_249235.1	Metallophos	No hits	0	C	Total length threading
NP_249511.1	HNH	COG3183	0	C	Total length threading
NP_250044.1	3-dmu-9_3-mt	COG2764	0	C	Total length threading
NP_274729.1	Peptidase_M61	COG3975	0	C	Total length threading
NP_293768.1	Metallophos	No hits	0	C	Total length threading
NP_295449.1	HNH	COG3183	0	C	Total length threading
NP_295827.1	Glyoxalase	COG0346	0	C	Total length threading
NP_296130.1	HNH	COG3183	0	C	Total length threading
NP_376315.1	Peptidase_M28	COG2234	0	C	Total length threading
NP_376530.1	Peptidase_M61	COG3975	0	C	Total length threading
NP_377887.1	Peptidase_M28	COG2234	0	C	Total length threading
NP_378324.1	Lactamase_B	COG2248	1	T	Total length threading
NP_394449.1	Metallophos	COG0639	0	C	Total length threading
NP_418531.1	3-dmu-9_3-mt	COG2764	0	C	Total length threading
NP_419117.1	Glyoxalase	COG0346	0	C	Total length threading
NP_419948.1	PHP	No hits	0	C	Total length threading

NP_420092.1	Glyoxalase	COG2764	1	T	Total length threading
NP_421342.1	Peptidase_M61	COG3975	1	T	Total length threading
NP_422081.1	Peptidase_M61	COG3975	0	C	Total length threading
NP_422218.1	Glyoxalase	COG3324	0	C	Total length threading
NP_440495.1	Glyoxalase	No hits	1	T	Total length threading
NP_440868.1	Glyoxalase	No hits	0	C	Total length threading
NP_442720.1	Peptidase_M61 PDZ	COG3975	0	C	Total length threading
NP_485065.1	Glyoxalase	No hits	1	T	Total length threading
NP_485147.1	Glyoxalase	COG0346	0	C	Total length threading
NP_485241.1	Metallophos	COG0639	0	C	Total length threading
NP_485242.1	No hits	COG0639	0	C	Total length threading
NP_485915.1	Metallophos MMR_HSR1 NACHT HEAT HEAT_PBS HEAT_PBS HEAT_PBS HEAT HEAT_PBS HEAT_PBS HEAT_PBS Adaptin_N HEAT_PBS HEAT HEAT_PBS	No hits	0	C	Total length threading
NP_485943.1	HEAT_PBS HEAT HEAT_PBS HEAT_PBS HEAT HEAT_PBS HEAT_PBS HEAT HEAT_PBS HEAT_PBS HEAT HEAT_PBS HEAT_PBS HEAT	COG1413	2	T	Total length threading
NP_486366.1	Peptidase_M61	COG3975	0	C	Total length threading
NP_486497.1	HNH	No hits	0	C	Total length threading
NP_488987.1	Metallophos	No hits	0	C	Total length threading
NP_499660.2	Recep_L_domain Recep_L_domain	No hits	0	C	Total length threading
NP_500348.1	Peptidase_M13	No hits	0	C	Total length threading
NP_503636.1	Peptidase_M13	No hits	0	C	Total length threading
NP_503898.2	Peptidase_M13	No hits	0	C	Total length threading
NP_503902.2	Peptidase_M13	No hits	0	C	Total length threading
NP_503921.1	Peptidase_M13	No hits	0	C	Total length threading
NP_503927.1	Peptidase_M13	No hits	0	C	Total length threading
NP_509680.1	Peptidase_M13	No hits	0	C	Total length threading
NP_509681.2	Peptidase_M13	No hits	0	C	Total length threading
NP_521011.1	Peptidase_M61	COG3975	1	T	Total length threading
NP_558919.1	Peptidase_M28	COG2234	0	C	Total length threading
NP_560908.1	Lactamase_B	COG2248	0	C	Total length threading
NP_602758.1	AP_endonuc_2	No hits	0	C	Total length threading
NP_613500.1	No hits	COG2248	0	C	Total length threading
NP_618379.1	No hits	COG3183	0	C	Total length threading
NP_625081.1	No hits	No hits	0	C	Total length threading
NP_626109.1	Glyoxalase	COG2764	0	C	Total length threading
NP_626703.1	No hits	No hits	0	C	Total length threading
NP_627276.1	Glyoxalase	No hits	0	C	Total length threading
NP_627277.1	No hits	No hits	0	C	Total length threading
NP_627285.1	No hits	No hits	1	T	Total length threading
NP_628395.1	No hits	No hits	0	C	Total length threading
NP_628397.1	Glyoxalase	No hits	0	C	Total length threading
NP_628752.1	No hits	No hits	0	C	Total length threading
NP_628834.1	Glyoxalase	No hits	0	C	Total length threading
NP_630250.1	PHP	No hits	1	T	Total length threading
NP_630882.1	Glyoxalase	COG0346	0	C	Total length threading

NP_631048.1	3-dmu-9_3-mt	COG2764	0	C	Total length threading
NP_631649.1	Glyoxalase	No hits	0	C	Total length threading
NP_635629.1	Glyoxalase	COG3324	1	T	Total length threading
NP_636537.1	TPR_3 Peptidase_M61	COG3975	0	C	Total length threading
NP_638980.1	No hits	No hits	0	C	Total length threading
NP_661320.1	AP_endonuc_2	No hits	0	C	Total length threading
NP_661620.1	Metallophos	No hits	1	T	Total length threading
NP_695983.1	Metallophos	COG4186	0	C	Total length threading
NP_696099.1	No hits	COG4186	0	C	Total length threading
NP_717822.1	Peptidase_M14	No hits	0	C	Total length threading
NP_719288.1	Peptidase_M14	No hits	0	C	Total length threading
NP_796761.1	Metallophos	No hits	0	C	Total length threading
NP_813437.1	Metallophos	No hits	0	C	Total length threading
NP_813439.1	Metallophos	COG4186	0	C	Total length threading
NP_841044.1	3-dmu-9_3-mt	COG2764	0	C	Total length threading
NP_864680.1	3-dmu-9_3-mt	COG2764	0	C	Total length threading
NP_953812.1	Metallophos	No hits	1	T	Total length threading
NP_963361.1	No hits	COG2248	0	C	Total length threading
YP_005891.1	Metallophos	No hits	0	C	Total length threading
YP_064605.1	Metallophos	COG1413	0	C	Total length threading
YP_065769.1	HNH	COG3183	0	C	Total length threading
YP_113022.1	No hits	COG3183	0	C	Total length threading
YP_114683.1	Glyoxalase	COG0346	0	C	Total length threading
YP_115203.1	Glyoxalase	COG2764	0	C	Total length threading

Supplementary Table 2. Functional annotations (percentage) of the 57 zinc proteomes: A) archaea; B) bacteria. C) eukarya; D) human.

	Archaea	**Bacteria**	**Eukarya**	***Homo sapiens***
DNA-replication	2.30	4.05	1.42	0.40
Hydrolase	42.81	48.59	20.89	18.98
Isomerase	1.43	3.95	0.34	0.26
Ligase	5.85	5.28	12.81	12.32
Lyase	3.73	6.21	1.00	1.00
Oxidoreductase	3.05	5.98	1.65	1.08
Protein-Ubiquitin-binding	0.31	0.46	0.36	0.13
Signal	1.43	1.94	7.55	6.56
Storage-Homeostasis	0.00	0.03	0.22	0.48
Structural	4.85	1.15	1.07	1.27
Transcription	10.77	5.58	37.11	43.81
Transferase	14.37	10.13	6.97	7.59
Translation	5.04	0.11	0.79	0.58
Transport	2.68	5.31	4.32	3.65
Unknown	1.37	1.22	2.71	1.22
Zn-fingers with unknown role	0.00	0.00	0.78	0.66

Zinc proteomes, phylogenetics and evolution†

Leonardo Decaria,[a] Ivano Bertini[ab] and Robert J. P. Williams*[c]

Received 30th June 2010, Accepted 4th August 2010
DOI: 10.1039/c0mt00024h

Evolution has not been studied in detail with reference to the changing environment. This requires a study of the inorganic chemistry of organisms, especially metalloproteins. The evolution of organisms has been analysed many times previously using comparative studies, fossils, and molecular sequences of proteins, DNA and 16s rRNA (Zhang and Gladyshev, *Chem. Rev.*, 2009, **109**, 4828). These methods have led to the confirmation of Darwin's original proposal that evolution followed from natural selection in a changing environment often pictured as a tree. In all cases, the main tree in its upper later reaches has been well studied but its lower earlier parts are not so well defined. To approach this topic we have treated evolution as due to the intimate combination of the effect of chemical changes in the environment and in the organisms (Williams and da Silva, *The Chemistry of Evolution*, 2006, Elsevier). The best chemicals to examine are inorganic ions as they are common to both. As a more detailed example of the chemical study of organisms we report in this paper a bioinformatic approach to the characterization of the zinc proteomes. We deduce them from the 821 totally sequenced DNA of organisms available on NCBI, exploiting a published method developed by one of us (Andreini, Bertini and Rosato, *Acc. Chem. Res.*, 2009, **42**, 1471). Comparing the derived zinc-finger-containing proteins and zinc hydrolytic enzymes in organisms of different complexity there is a correlation in their changes during evolution related to environmental change.

Introduction

Evolution has a well-mapped outline especially from the Cambrian period to today. Many authors have described it in diagrams with a tree-like structure using substantially two methods: comparison of organisms while exploiting also fossil evidence as Darwin did, or, more recently, using sequence studies of DNA, RNA and proteins.[1] In all cases the recent upper reaches of the tree have been well studied but the earlier trunk and lower branches have not been so easy to study. In an effort to overcome some of the difficulties, we have begun a study of the environmental organisms and changes from the earliest possible dates since it is changes in the environment, which have had the greatest effect on early evolution.[2] The environment can be examined and dated in sediments and organisms are considered to have arisen in large groups from bacteria to higher animals relative to these datings. The dating of organisms is aided by knowledge of the ages of fossils. The best chemicals for a comparative study during evolution are the inorganic elements as they are common to both. This paper is the first example of a detailed examination based on bioinformatic analysis of one element, zinc, from organism DNA sequences, devised by one of us.[3]

Methods

We obtained 821 complete proteomes, 52 from archaea, 723 from bacteria and 46 from eukarya available on NCBI. We selected zinc proteins from them as an example of inorganic element chemistry involvement in organisms. We were able to do so from our ability to recognize zinc-binding domains in sequences of proteins. We recognized separately two major groups of zinc proteins, those of the hydrolytic enzymes and those of the zinc fingers as there are clear identifying structural data on both.[4,5] In the case of the zinc fingers the metal sites are recognized by the coordination of the zinc by a combination of four residues of His or Cys in a particular sequence arrangement, giving rise to a tetrahedral geometry. The hydrolytic enzymes have zinc coordinated by a particular sequence of three residues selected from His, Glu, Asp and Arg with one or two water molecules. The knowledge of the site structures allows us to recognise the zinc binding domains in a protein or an enzyme as published.[3] As reported in that paper, we used the HMMER program to search the NCBI refseq proteins database for matches to the hidden Markov models (HMMs) representing the selected domains. The HMMs were taken from the Pfam database without modification. We selected 10^{-3} as an Evalue cut-off. We thereby obtained a set of 271 zinc-binding proteins and divided them by their functions as mentioned above. The specific activities of the enzymes are known and we divided the hydrolases further as proposed by

[a] *Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy*
[b] *Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy*
[c] *Department of Inorganic Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QR, United Kingdom. E-mail: bob.williams@chem.ox.ac.uk*

† Electronic supplementary information (ESI) available: Additional data. See DOI: 10.1039/c0mt00024h

the Enzyme Commission, once their sequences were recognized. Further information about the methods is included in the ESI.†

Results

In order to give a comparative account of the zinc protein data and their analyses, we have considered organisms in the following ways. The prokaryotes were divided into archaea and eubacteria, and the eubacteria were further divided into those with a small proteome of less than 1500 domains and those with a larger domain, see the Table 1. The smaller proteins are mostly of invasive bacteria found in animal hosts while the larger are of bacteria which in general are free-living. The two have considerable differences in the numbers of their zinc proteins. We shall take the archaea and the larger eubacteria as representing possible early forms of life or at least life of low complexity. There is little difference in the zinc protein content of these eubacteria and archaea and between anaerobic and aerobic species of both. Amongst eukaryotes we have placed multicellular metazoans in order of complexity, and quite probably in the order of their evolution, as is conventional; *C. elegans*, *D. melanogaster*, and *Homo sapiens*. We do not have data on a single-cell free-living eukaryote organisms and have used the small parasites, *P. falsiparsium* and *T. brucei* as examples. We have looked separately at the zinc proteins of one plant, an *Arabidopsis* and the yeast, *Saccharomyces cerevisiae* as an example of a fungus. In order that the reader can appreciate our approach to the common features of changes in the environment and this description of the evolution of organisms, we note that the prokaryotes are always considered as arising even before there was oxygen in the atmosphere and that the single cell eukaryotes arose about two billion years ago with the first step rise in oxygen, followed by the multicellular organisms with the second rise in oxygen after about one billion years ago.[1,2] Accompanying the steps in oxygen were steps in trace metals in the environment including zinc.[2] The data on the times of evolution of organisms can therefore be tied in time to those of the environment which enables us to compare zinc analyses of both. We turn to the examination of the zinc data on organisms in a comparative format.

Table 1 Characterisation of zinc proteins in organisms

	Total proteome	%Zn-finger	%EC:3.4
Archea (52)	2176[a]	0.180	0.923
Bact. under 1500 (93)	940[a]	0.383	1.611
Bact. over 1500 (630)	3671[a]	0.177	1.227
P. falciparum	6265	1.538	0.675
T. brucei	9279	1.369	0.787
C. elegans	22 844	2.889	1.064
D. melanogaster	20 513	3.734	2.613
H. sapiens	37 742	4.849	1.200
A. thaliana	32 615	2.370	0.584

[a] Average Value: Numbers in the first column refer to groupings of proteome sizes and the numbers in brackets refer to the numbers of proteomes examined. The average of which is given in the second column.

Discussion

We shall refer to the total content of the zinc proteins in the organisms, see the Table 1, but to make a better comparison, we shall also give the zinc proteins as a percentage in their proteomes, Fig. 1. The increases in zinc finger proteins in both numbers and percentages there are seen to follow the order of complexity of organisms and probably the order of their evolution, Fig. 2. While the average number in all prokaryotes is less than six the number in *Homo sapiens* has increased to 1800. Both in numbers and percentages there are to be two step changes in the increase of these proteins between prokaryotes and unicellular metazoan organisms and between these unicellular organisms and the three multicellular organisms.[6,7] There is a larger increase in the percentage in *Homo sapiens*, Fig. 3. Of great interest is to note that as mentioned above the zinc in the environment increased with oxygen increase in two steps which are close to these steps of organism evolution at 2.5 to 2.0 billion years ago and 1.0 to 0.5 billion years ago. The greater increase is in the second period when multicellular organisms arose. Increases in complexity of organisms is paralleled by the need for increases of message systems and we note that the zinc fingers are very important transcription factors for hormonal messengers. The smaller percentage and number of transcription factors in plants, Fig. 3 can be correlated with the much lower complexity of them compared with *Homo sapiens*. The value of percentage zinc fingers in unicellular yeast of 1.9% is similar to that in the unicellular metazoans.

Turning to the other enzymes and proteins we have studied there are no metallothioneins in prokaryotes (not shown), plants and yeast but small numbers in metazoans. Greatest interest centres on the large numbers of hydrolytic zinc enzymes. We noted above the greater percentage of them in small eubacteria and their low level in eukarya, Fig. 2. In particular we see in Fig. 3 that the percentage of EC:3.4 enzymes, that is the peptidases and proteases, has a very different pattern from that of zinc fingers. Except for the fly, *D. melanogaster*, the percentage varies little being slightly lower in all the other eukaryotes than in prokaryotes. The high value in the fly could be related to its need to metamorphose. This will be examined in a wider range of organisms. The content of the zinc hydrolytic enzymes is high in all the organisms and we consider that this is a reflection of the need to hydrolyse proteins for food in all organisms and to hydrolyse connective tissue for growth in eukaryotes. It is noticeably lower in the multicellular plant and in yeast (0.8%) than in multicellular animals.

The changes in the EC:6 enzymes are of considerable interest as they include those for the hydrolytic reactions of phosphates. In particular zinc is associated in the earliest forms of life with the activities of RNA enzymes. In Fig. 2 we observe that the average values for all eukaryotes is much lower, less than 0.5%, than for all prokaryotes, greater than 0.6%. The parasytic eubacteria have the high value of above 2.0% on average. Notice however that they have very small genomes indicative of a loss of many genes but not of EC:4 and EC:6 enzymes. Unlike most of the other zinc proteins it appears that these enzymes have not evolved greatly from their initial functions.

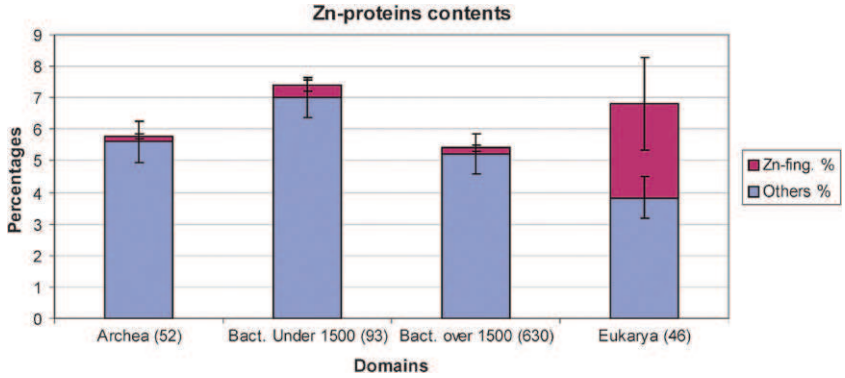


Fig. 1 The average zinc protein contents for archaea, small and large bacteria and eukarya. Archaea and large bacteria have averages near to 0.2% of Zn-finger proteins in their proteomes, while small bacteria have about 0.4%. Eukarya Zn-finger content rises up to 3%.

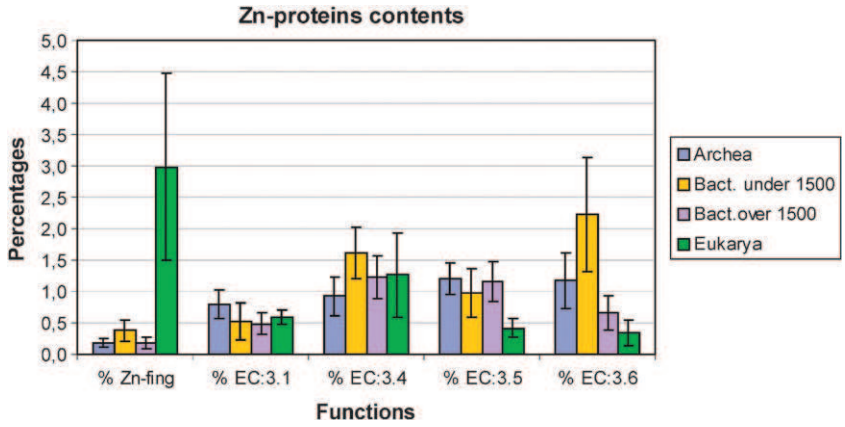


Fig. 2 Zn–protein distribution in the four groups archaea, small and large bacteria and eukarya. EC:3.4 = protease/peptidase; EC:3.5 = hydrolases of C–N bonds other than in peptides; EC:3.6 = acid anhydride hydrolases.

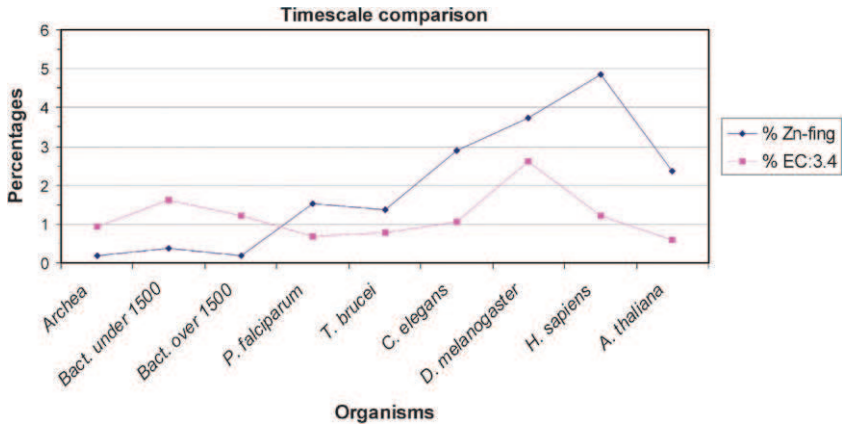


Fig. 3 A timescale comparison of small genomes in small and large prokaryotes, then in unicellular and finally in multicellular eukaryotes. It is notable that the percentages in Zn-finger content rise within this evolutionary series. The percentage value of EC:3.4 in small bacteria is higher than in the large bacteria. Small bacteria are usually parasites, they need a bigger pool of proteases/peptidases to break down extracellular proteins for food.

The conclusion of this paper is that much light is thrown on the development of complexity and probably of evolution of organisms from a comparison of the zinc in the environment with that in organisms. It will be possible to examine our conclusions more closely as we acquire more data. It is certainly necessary to repeat the analysis with data on other metal ions. We stress the great advantage of such examinations of the inorganic content of organisms with their

co-existing environment particularly in the study of early evolution.[2] In this paper we have not attempted to trace special recent features of the evolution of zinc use such as the loss of cobalt enzymes, using vitamin B12, and their replacement by zinc enzymes in higher plants. Again zinc is required in the synthesis of shikimic acid, an essential part of the pathway to all amino acids carrying aromatic side-chains but the zinc enzyme is absent in higher animals. When did these gains or losses of zinc genes occur? The bioinformatic approach to metallomics of one of us[3] as used here should be able to provide such information.

References

1 Y. Zhang and V. N. Gladyshev, *Chem. Rev.*, 2009, **109**, 4828–4861.
2 R. J. P. Williams and J. J. R. Fraùsto da Silva, *The Chemistry of Evolution*, Elsevier, Chichester, 2006.
3 C. Andreini, I. Bertini and A. Rosato, *Acc. Chem. Res.*, 2009, **42**, 1471–1479.
4 W. Maret and Y. Li, *Chem. Rev.*, 2009, **109**, 4682–4707.
5 A. Messerschmidt, W. Bode and M. Cygler, *Handbook of Metalloproteins*, Elsevier, Chichester, 2004.
6 A. D. Anbar and A. H. Knoll, *Science*, 2002, **297**, 1137–1142.
7 M. A. Saito, D. M. Sigman and F. M. M. Morel, *Inorg. Chim. Acta*, 2003, **356**, 308–318.

Table S1 – The starting data set, which consists of 271 proposed Zn-binding Pfam domains, some of them with one or more metal binding pattern (MBP) associated. In the zinc fingers the zinc ion is bound by a combination of four residues of His or Cys in a tetrahedral geometry. The hydrolytic enzymes have zinc coordinated by three residues among His, Glu, Asp and Arg with one or two water molecules.

Pfam domain	Associated pattern(s)
3_5_exonuc	'DX(1)EX(124)D', 'DX(1)EX(131)D', 'DX(1)EX(143)D', 'DX(56)D'
5_3_exonuc_N	'DX(22)D'
5_nucleotid_C	No pattern
AAL_decarboxy	'HX(1)HX(10)H'
ADH_N	'CX(0)T', 'CX(2)CX(2)CX(7)C', 'CX(2)CX(2)CX(8)C', 'CX(20)H', 'CX(20)HX(0)E', 'CX(21)H', 'CX(21)HX(0)E', 'CX(22)H', 'CX(23)H', 'CX(24)H', 'CX(24)HX(0)E', 'CX(26)HX(0)E', 'CX(29)H', 'CX(29)HX(0)E', 'CX(95)H', 'DX(1)H', 'DX(2)CX(2)CX(7)C', 'EX(2)CX(2)CX(7)C'
ADH_zinc_N	'DX(1)H'
ADK_lid	'CX(16)CX(2)D', 'CX(2)CX(16)CX(2)C', 'CX(2)CX(16)CX(2)D'
ALAD	'CX(1)C', 'CX(1)CX(7)C', 'CX(1)DX(7)C', 'CX(7)CX(36)D', 'DX(7)C', 'EX(43)H'
APOBEC_C	No pattern
APOBEC_N	'EX(37)HX(29)CX(2)C', 'HX(29)CX(2)C', 'HX(30)CX(2)C'
ATP-sulfurylase	'CX(2)CX(8)CX(3)H'
A_deamin	'HX(56)CX(64)C'
A_deaminase	'HX(1)HX(196)HX(80)D', 'HX(1)HX(196)HX(80)E', 'HX(1)HX(277)D', 'HX(76)H', 'HX(83)H'
Abhydrolase_2	'HX(25)DX(2)CX(27)H'
Ada_Zn_binding	'CX(26)C', 'CX(26)CX(2)C', 'CX(3)CX(26)CX(2)C'
Adenine_glyco	'CX(157)HX(3)H'
Alk_phosphatase	'DX(101)HX(1)TX(166)E', 'DX(103)TX(166)E', 'DX(111)HX(1)TX(158)E', 'DX(216)SX(0)D', 'DX(3)H', 'DX(3)HX(111)H', 'DX(3)HX(112)H', 'DX(3)HX(73)H', 'DX(3)HX(80)H', 'DX(317)DX(0)H', 'DX(42)H', 'DX(48)SX(269)DX(0)H', 'DX(49)SX(264)DX(0)H', 'DX(50)CX(266)DX(0)H', 'DX(50)SX(224)DX(41)DX(0)H', 'DX(50)SX(266)DX(0)H', 'DX(84)H', 'NX(50)SX(266)DX(0)H'
Alpha_kinase	'HX(56)HX(1)CX(3)C'
Amidase_2	'HX(104)HX(7)C', 'HX(109)HX(7)C', 'HX(115)HX(9)D', 'HX(119)HX(9)D', 'HX(99)HX(7)C'
Amidase_3	'HX(15)EX(53)H', 'HX(55)E'
Amidinotransf	'HX(100)C'
Amidohydro_1	'CX(123)HX(29)H', 'DX(26)HX(58)H', 'DX(36)H', 'EX(3)E', 'EX(63)HX(20)H', 'HX(1)HX(142)HX(78)D', 'HX(1)HX(152)HX(91)D', 'HX(1)HX(154)HX(85)D', 'HX(1)HX(166)HX(74)D', 'HX(1)HX(169)HX(36)HX(50)D', 'HX(1)HX(209)D', 'HX(1)HX(214)D', 'HX(1)HX(214)N', 'HX(1)HX(231)D', 'HX(1)HX(249)D', 'HX(1)HX(253)D', 'HX(1)HX(254)D',

	'HX(1)HX(256)D', 'HX(1)HX(258)D', 'HX(1)HX(26)C', 'HX(1)HX(293)D', 'HX(1)HX(89)DX(158)D', 'HX(19)H', 'HX(214)D', 'HX(22)H', 'HX(231)H', 'HX(269)H', 'HX(27)H', 'HX(28)H', 'HX(37)H', 'HX(39)E', 'HX(55)H', 'HX(57)H', 'HX(89)H', 'KX(32)HX(55)H', 'QX(1)HX(247)D'
Amidohydro_2	'DX(26)HX(3)H', 'EX(122)H', 'EX(28)HX(28)H', 'EX(35)HX(23)H', 'HX(1)HX(110)EX(105)D', 'HX(1)HX(165)HX(116)D', 'HX(1)HX(169)HX(105)E', 'HX(1)HX(85)EX(109)D', 'HX(35)H'
Amidohydro_3	'CX(123)HX(29)H', 'DX(26)HX(58)H', 'DX(36)H', 'EX(3)E', 'HX(1)HX(142)HX(78)D', 'HX(1)HX(152)HX(91)D', 'HX(1)HX(154)HX(85)D', 'HX(1)HX(166)HX(74)D', 'HX(1)HX(169)HX(36)HX(50)D', 'HX(1)HX(214)D', 'HX(1)HX(214)N', 'HX(1)HX(249)D', 'HX(1)HX(253)D', 'HX(1)HX(254)D', 'HX(1)HX(256)D', 'HX(1)HX(258)D', 'HX(1)HX(26)C', 'HX(1)HX(89)DX(158)D', 'HX(19)H', 'HX(214)D', 'HX(269)H', 'HX(27)H', 'HX(28)H', 'HX(39)E', 'HX(55)H', 'KX(32)HX(55)H'
Asp	'DX(181)D', 'HX(25)DX(63)D', 'HX(5)HX(141)DX(3)D'
AstE_AspA	'HX(2)E', 'HX(2)EX(88)H', 'HX(2)EX(89)H', 'HX(2)EX(91)H'
Astacin	'HX(3)HX(5)H', 'HX(3)HX(5)HX(46)Y'
Beta-lactamase	No pattern
COX2_TM	'HX(3)H'
COX3	'HX(83)HX(3)E'
COX5B	'CX(1)CX(19)CX(2)C', 'CX(7)HX(14)CX(2)C', 'DX(7)E'
COX7C	'HX(2)E'
Carb_anhydrase	'CX(1)HX(22)H', 'DX(1)HX(22)H', 'HX(1)H', 'HX(1)HX(16)H', 'HX(1)HX(17)H', 'HX(1)HX(22)C', 'HX(1)HX(22)D', 'HX(1)HX(22)H', 'HX(1)HX(22)HX(79)C', 'HX(1)HX(22)HX(79)D', 'HX(1)HX(22)HX(79)E', 'HX(1)HX(22)N', 'HX(1)HX(22)Q', 'HX(2)RX(132)H', 'HX(22)H', 'HX(24)H', 'NX(1)H', 'NX(1)HX(22)H', 'QX(68)C'
D-aminoacyl_C	No pattern
DFF40	'CX(8)CX(3)HX(64)C'
DHH	'DX(61)DX(21)H', 'HX(3)DX(63)D'
DHQ_synthase	'DX(3)HX(62)HX(13)H', 'DX(78)HX(15)H', 'DX(82)HX(16)H', 'DX(82)HX(17)H', 'EX(2)H', 'EX(63)HX(13)H', 'EX(76)HX(15)H', 'EX(77)H'
DNA_ligase_A_M	'DX(129)E'
DNA_ligase_ZBD	'CX(2)CX(12)CX(4)C', 'CX(2)CX(14)CX(5)C'
DNA_pol_A	'EX(171)D', 'HX(3)E'
DOPA_dioxygen	'HX(69)H'
DUF1907	'HX(1)HX(9)H'
DUF258	'CX(1)HX(5)C', 'CX(4)CX(1)HX(5)C'
Endonuclease_7	'CX(2)CX(31)CX(2)C'
FTP	No pattern
F_bP_aldolase	'DX(29)EX(6)E', 'DX(62)EX(91)HX(19)K', 'HX(115)HX(37)H', 'HX(37)H', 'HX(63)EX(51)HX(37)H', 'HX(93)HX(31)H', 'HX(96)HX(27)H'
Fe-ADH	'DX(3)HX(62)HX(13)H', 'DX(3)HX(68)HX(13)H', 'DX(78)HX(15)H', 'DX(82)HX(16)H',

	'DX(82)HX(17)H', 'EX(2)H'
Flavi_NS5	'EX(3)HX(4)CX(2)C', 'HX(1)HX(13)CX(118)C', 'HX(15)CX(118)C'
Flavodoxin_2	'HX(3)H', 'HX(3)HX(44)C'
FmdA_AmdA	'DX(1)HX(13)E', 'NX(1)DX(24)D'
GDPD	'EX(1)DX(79)E'
GFA	'CX(1)CX(2)C', 'CX(2)C'
GPI	'DX(0)E', 'HX(1)HX(6)EX(38)H'
GTP_EFTU	'CX(2)CX(8)CX(2)C'
GTP_cyclohydro2	'CX(10)CX(1)C'
GTP_cyclohydroI	'CX(2)HX(67)C'
GalP_UDP_tr_C	'CX(54)C', 'CX(54)H'
GalP_UDP_transf	'CX(2)CX(59)HX(48)H'
GatB_N	'CX(1)CX(50)CX(2)E'
Glyco_hydro_26	'EX(3)H', 'HX(31)DX(9)E', 'RX(2)HX(0)EX(70)DX(36)E', 'RX(2)HX(71)DX(36)E'
Glyco_hydro_38	'HX(1)DX(111)D', 'HX(1)DX(121)D'
Glyoxalase	'EX(65)E', 'HX(45)E', 'HX(45)Q', 'HX(46)EX(22)HX(48)E', 'HX(47)E', 'HX(47)H', 'HX(50)E', 'QX(65)E'
Guanylate_cyc	'DX(43)D'
HCV_NS5a_1a	'CX(17)CX(1)CX(20)C'
Hist_deacetyl	'DX(1)HX(85)D', 'DX(1)HX(86)D', 'DX(1)HX(87)D', 'DX(1)HX(91)D'
Histidinol_dh	'HX(97)D', 'QX(2)HX(97)D'
HtpX_N	No pattern
HycI	No pattern
IPT	'DX(39)EX(0)H'
Iso_dh	'DX(3)D', 'HX(1)E', 'HX(4)G'
KduI	'HX(1)HX(4)EX(41)H'
Lactamase_B	'CX(0)H', 'DX(0)H', 'DX(0)HX(102)D', 'DX(0)HX(102)DX(46)H', 'DX(0)HX(112)DX(25)H', 'DX(0)HX(124)EX(44)H', 'DX(0)HX(135)H', 'DX(0)HX(136)DX(53)H', 'DX(0)HX(142)D', 'DX(0)HX(142)DX(57)H', 'DX(0)HX(74)DX(38)H', 'DX(0)HX(76)D', 'DX(0)HX(81)DX(43)H', 'DX(0)HX(82)DX(45)H', 'DX(0)HX(90)DX(55)H', 'DX(0)HX(91)D', 'DX(0)HX(97)D', 'DX(0)RX(99)C', 'DX(100)C', 'DX(76)CX(38)H', 'DX(77)CX(41)H', 'DX(79)CX(41)H', 'EX(13)HX(56)K', 'EX(76)CX(38)H', 'HX(1)HX(100)D', 'HX(1)HX(108)EX(18)E', 'HX(1)HX(127)E', 'HX(1)HX(53)HX(23)D', 'HX(1)HX(55)HX(18)D', 'HX(1)HX(59)H', 'HX(1)HX(60)H', 'HX(1)HX(60)HX(24)D', 'HX(1)HX(62)H', 'HX(1)HX(62)HX(21)D', 'HX(1)HX(72)HX(20)D', 'HX(1)HX(72)HX(21)D', 'HX(1)HX(73)H', 'HX(1)HX(74)HX(70)D', 'HX(1)HX(77)H', 'HX(1)HX(77)HX(21)D', 'HX(1)HX(79)H', 'HX(1)HX(79)HX(20)D', 'HX(1)HX(83)HX(55)D', 'HX(1)HX(84)HX(20)D', 'HX(1)HX(95)NX(19)D', 'HX(164)E', 'HX(74)HX(70)D', 'HX(96)DX(45)H'
Ldh_1_C	'DX(41)KX(8)E'
Ldh_1_N	'DX(2)E', 'NX(1)C'

LigB	'HX(176)H'
Metallophos	'DX(1)HX(39)DX(146)H', 'DX(1)HX(40)DX(169)Q', 'DX(1)QX(44)DX(206)H', 'DX(169)H', 'DX(29)NX(75)HX(38)H', 'DX(31)NX(100)HX(34)H', 'DX(31)NX(122)H', 'DX(31)NX(48)HX(81)H', 'DX(33)NX(102)HX(31)H', 'DX(35)NX(131)HX(36)H', 'DX(35)NX(31)HX(21)H', 'DX(36)NX(84)HX(36)H', 'HX(8)E'
Metallothio	'CX(3)CX(4)CX(4)C', 'CX(5)CX(1)CX(10)C'
Metallothio_5	No pattern
Metallothio_PEC	No pattern
Metallothio_Pro	'CX(15)CX(14)CX(1)H', 'CX(24)CX(3)HX(13)C', 'CX(32)CX(4)CX(1)C', 'CX(4)CX(17)CX(3)C'
Metallothionein	No pattern
Meth_synt_1	No pattern
Meth_synt_2	'HX(1)CX(21)EX(61)C', 'HX(1)CX(83)C', 'HX(3)D'
Methyltransf_1N	'CX(4)CX(55)H'
Monooxygenase_B	'DX(1)D'
NAD_binding_1	'EX(34)H'
NUDIX-like	'CX(2)C'
OMPdecase	'HX(1)DX(30)HX(108)D'
PADR1	'CX(2)CX(12)CX(9)C'
PARP	'CX(2)HX(4)CX(2)C'
PDEase_I	'HX(0)HX(108)D', 'HX(0)HX(109)D', 'HX(0)HX(112)D', 'HX(0)HX(116)D', 'HX(110)H', 'HX(33)HX(0)DX(109)D', 'HX(33)HX(0)NX(109)D', 'HX(35)HX(0)D', 'HX(35)HX(0)DX(105)D', 'HX(35)HX(0)DX(108)D', 'HX(35)HX(0)DX(109)D', 'HX(35)HX(0)DX(110)D', 'HX(35)HX(0)DX(116)D', 'HX(35)HX(0)NX(116)D'
PMI_typeI	'HX(17)EX(56)H', 'QX(1)HX(24)EX(146)H'
PTE	'DX(20)H', 'EX(32)HX(27)H', 'EX(74)H', 'HX(1)HX(110)EX(117)D', 'HX(1)HX(243)D', 'HX(22)D', 'HX(28)H', 'RX(28)DX(1)S'
PTPS	'HX(1)H', 'HX(1)HX(76)H', 'HX(24)HX(1)H'
PdxA	'HX(99)H'
Pep_deformylase	No pattern
Peptidase_A25	No pattern
Peptidase_C2	No pattern
Peptidase_C78	No pattern
Peptidase_M1	'HX(3)HX(18)E'
Peptidase_M10	'CX(3)HX(5)H', 'HX(0)KX(2)HX(5)H', 'HX(1)DX(12)H', 'HX(1)DX(12)HX(12)H', 'HX(14)HX(12)H', 'HX(3)HX(5)H', 'HX(3)HX(5)HX(29)Y', 'HX(3)HX(5)HX(7)M'
Peptidase_M11	No pattern
Peptidase_M13	'HX(3)HX(58)E'
Peptidase_M13_N	No pattern
Peptidase_M14	'EX(205)E', 'HX(2)EX(103)H', 'HX(2)EX(123)H', 'HX(2)EX(124)H', 'HX(2)EX(131)H'

Peptidase_M15	'HX(6)DX(60)H'
Peptidase_M15_3	'HX(6)DX(35)H'
Peptidase_M16	'HX(3)HX(75)E', 'HX(3)HX(76)E'
Peptidase_M16_C	No pattern
Peptidase_M17	'DX(76)DX(1)E', 'KX(4)DX(17)DX(60)E', 'LX(0)MX(1)TX(97)R'
Peptidase_M18	'DX(32)EX(108)H', 'HX(103)DX(54)D'
Peptidase_M19	'EX(69)HX(20)H', 'EX(72)HX(20)H', 'HX(1)DX(102)E'
Peptidase_M2	'HX(3)HX(23)E'
Peptidase_M20	'DX(260)E', 'DX(3)D', 'DX(30)EX(109)H', 'DX(30)EX(311)H', 'DX(32)EX(108)H', 'DX(32)EX(97)H', 'DX(33)EX(204)H', 'DX(34)E', 'DX(34)EX(103)H', 'DX(34)EX(208)H', 'DX(34)EX(254)H', 'DX(34)EX(277)H', 'DX(34)EX(284)H', 'HX(10)DX(97)H', 'HX(11)DX(60)D', 'HX(11)DX(62)D', 'HX(19)DX(0)DX(60)D', 'HX(19)DX(33)EX(27)D', 'HX(19)DX(61)D', 'HX(28)DX(58)E', 'HX(31)DX(57)D', 'HX(32)DX(62)D', 'HX(52)D', 'HX(53)D', 'HX(54)D', 'HX(55)D', 'HX(61)D'
Peptidase_M22	No pattern
Peptidase_M23	'HX(1)D', 'HX(3)DX(78)H', 'HX(76)D', 'HX(78)D'
Peptidase_M24	No pattern
Peptidase_M26_C	No pattern
Peptidase_M26_N	No pattern
Peptidase_M27	'DX(172)H', 'HX(3)H', 'HX(3)HX(33)E', 'HX(3)HX(34)E', 'HX(3)HX(35)E', 'HX(38)E'
Peptidase_M28	'DX(3)D', 'DX(32)EX(97)H', 'DX(34)EX(103)H', 'DX(34)EX(114)H', 'DX(34)EX(208)H', 'DX(37)EX(127)H', 'DX(42)EX(127)H', 'HX(11)DX(60)D', 'HX(11)DX(62)D', 'HX(19)DX(0)DX(60)D', 'HX(19)DX(33)EX(27)D', 'HX(19)DX(61)D', 'HX(28)DX(58)E', 'HX(36)DX(28)E', 'HX(65)D', 'HX(9)DX(65)D'
Peptidase_M29	No pattern
Peptidase_M3	'DX(1)H', 'HX(3)H', 'HX(3)HX(23)E', 'HX(3)HX(24)E'
Peptidase_M30	No pattern
Peptidase_M32	'HX(3)HX(25)E'
Peptidase_M35	'HX(3)HX(10)D'
Peptidase_M36	No pattern
Peptidase_M3_N	No pattern
Peptidase_M4	'HX(3)H'
Peptidase_M41	'HX(3)HX(72)D'
Peptidase_M42	'DX(30)EX(109)H', 'DX(32)EX(108)H', 'HX(103)DX(54)D', 'HX(113)DX(52)D', 'HX(113)DX(54)D'
Peptidase_M43	No pattern
Peptidase_M44	No pattern
Peptidase_M48	'HX(3)HX(51)EX(45)H'
Peptidase_M49	'HX(4)HX(51)E'
Peptidase_M4_C	'YX(8)EX(64)H'

Peptidase_M50	'HX(3)HX(89)D'
Peptidase_M54	No pattern
Peptidase_M55	'DX(1)EX(49)H', 'DX(95)HX(28)E'
Peptidase_M56	No pattern
Peptidase_M6	No pattern
Peptidase_M61	No pattern
Peptidase_M7	'HX(3)HX(5)D'
Peptidase_M74	'DX(2)H', 'HX(2)HX(6)DX(90)H'
Peptidase_M8	'HX(3)HX(65)H'
Peptidase_M9	No pattern
Peptidase_M9_N	No pattern
Peptidase_S29	'CX(0)TX(0)CX(45)C', 'CX(0)TX(46)C', 'CX(1)CX(45)C', 'CX(47)C'
PhnA_Zn_Ribbon	No pattern
Phosphodiest	'DX(3)HX(148)H', 'DX(3)HX(161)H', 'DX(35)TX(166)DX(0)H', 'DX(38)TX(176)DX(0)H'
Pico_P2A	'CX(1)CX(57)CX(1)H'
Pkinase_Tyr	'CX(11)CX(2)C', 'CX(3)CX(7)CX(2)C'
Polysacc_deac_1	'DX(3)H', 'DX(48)HX(3)H', 'DX(49)HX(3)H', 'HX(3)H'
Pre-SET	'CX(1)CX(2)CX(4)C', 'CX(1)CX(3)CX(4)C', 'CX(1)CX(4)CX(4)C', 'CX(1)CX(5)CX(4)C', 'CX(1)CX(5)CX(7)C', 'CX(10)CX(26)CX(3)C', 'CX(12)CX(29)CX(3)C', 'CX(13)CX(28)CX(3)C', 'CX(14)CX(46)CX(3)C', 'CX(17)CX(28)CX(3)C', 'CX(32)CX(5)CX(3)C', 'CX(35)CX(5)CX(3)C', 'CX(36)CX(5)CX(3)C', 'CX(38)CX(5)CX(3)C', 'CX(53)CX(5)CX(3)C'
ProRS-C_1	'CX(2)C', 'CX(4)CX(25)CX(2)C'
ProRS-C_2	No pattern
Pro_CA	'CX(1)DX(50)HX(2)C', 'CX(1)DX(53)HX(2)C', 'CX(52)HX(2)C', 'CX(54)HX(2)C', 'CX(59)HX(2)C'
PseudoU_synth_1	No pattern
Put_Phosphatase	'CX(3)CX(3)CX(2)C'
QRPTase_C	'HX(30)DX(25)D', 'HX(56)D'
RNA_POL_M_15KD	'CX(0)RX(1)C', 'CX(18)CX(2)C', 'CX(2)CX(18)C', 'CX(2)CX(18)CX(2)C'
RNA_pol_A_bac	'CX(1)CX(3)C', 'CX(1)CX(3)CX(2)C'
RNA_pol_L	'CX(1)CX(3)C', 'CX(1)CX(3)CX(2)C'
RNA_pol_N	'CX(2)CX(33)CX(0)C', 'CX(2)CX(34)CX(0)C', 'CX(33)CX(0)C'
RNA_pol_Rpb1_1	'AX(16)C', 'CX(0)MX(1)CX(37)CX(18)C', 'CX(0)QX(1)CX(6)CX(2)H', 'CX(1)CX(12)CX(2)C', 'CX(1)CX(15)C', 'CX(12)H', 'CX(18)C', 'CX(2)CX(37)CX(18)C', 'CX(2)CX(56)C', 'CX(2)CX(6)C', 'CX(2)CX(6)CX(2)H', 'CX(2)CX(9)H', 'CX(37)C', 'CX(37)CX(18)C', 'CX(4)NX(4)CX(2)H', 'CX(40)C', 'CX(56)C', 'CX(6)C', 'CX(6)CX(2)H', 'MX(1)CX(56)C'
RNA_pol_Rpb1_5	'CX(6)CX(2)C', 'CX(81)CX(6)C', 'CX(81)CX(6)CX(2)C'
RNA_pol_Rpb2_7	'CX(15)CX(2)C', 'CX(18)C', 'CX(2)CX(14)C', 'CX(2)CX(15)C', 'CX(2)CX(15)CX(2)C', 'CX(2)CX(18)C'
RdRP_1	'HX(1)HX(8)C'

RdRP_4	'HX(1)HX(8)C'
Reprolysin	'DX(0)FX(45)N', 'DX(32)HX(3)HX(5)H', 'HX(3)HX(5)H'
RhaA	'EX(32)DX(26)HX(39)D', 'EX(34)DX(26)HX(45)D', 'HX(31)D'
Ribosomal_L36	'CX(2)CX(12)CX(4)H', 'CX(2)CX(12)CX(5)H'
Ribosomal_L37ae	No pattern
Ribosomal_L37e	No pattern
Ribosomal_L40e	'CX(2)CX(10)CX(2)C'
Ribosomal_S14	'CX(12)CX(2)C', 'CX(2)CX(12)CX(2)C', 'CX(2)CX(15)C', 'RX(0)CX(1)RX(13)C'
Ribosomal_S27	No pattern
Ribosomal_S27e	No pattern
Ribul_P_3_epim	'HX(1)DX(30)HX(108)D', 'HX(1)DX(31)HX(108)D'
S-methyl_trans	'CX(26)NX(37)CX(0)C', 'CX(64)CX(0)C', 'CX(81)CX(0)C', 'YX(56)CX(81)CX(0)C'
SelR	'CX(2)CX(45)CX(2)C'
SoxD	'CX(2)CX(49)HX(3)C'
Sulfatase	'CX(0)C'
TGT	'CX(1)CX(2)CX(22)H', 'CX(1)CX(2)CX(25)H'
TK	'CX(2)CX(28)CX(2)C', 'CX(2)CX(29)CX(2)C', 'CX(2)CX(34)CX(2)C', 'CX(2)CX(34)CX(2)H'
TRM13	No pattern
TatD_DNase	'DX(26)HX(3)H', 'EX(35)HX(23)H', 'EX(35)HX(24)H', 'HX(1)HX(84)EX(110)D', 'HX(1)HX(85)EX(109)D'
Toxin_trans	No pattern
Trypsin	'HX(137)S', 'HX(143)S', 'HX(152)G', 'HX(152)GX(0)D', 'HX(2)H', 'HX(5)E', 'HX(51)E', 'HX(91)D'
U-box	'CX(1)HX(22)CX(4)C'
UCH	'CX(2)CX(22)HX(7)H', 'CX(2)CX(25)CX(2)C', 'CX(2)CX(43)CX(2)C', 'CX(2)CX(45)CX(2)C', 'CX(2)CX(47)CX(2)C'
VanY	'HX(6)DX(45)H'
Viral_protease	'CX(2)CX(31)CX(1)C'
YgbB	'DX(1)H', 'DX(1)HX(31)H'
Zn_peptidase	No pattern
Zn_peptidase_2	No pattern
dCMP_cyt_deam_1	'CX(29)CX(44)HX(7)E', 'CX(32)CX(2)C', 'CX(33)CX(2)C', 'CX(34)CX(2)C', 'CX(4)CX(6)H', 'CX(8)C', 'HX(24)CX(8)C', 'HX(26)CX(2)C', 'HX(27)CX(2)C', 'HX(27)CX(8)C', 'HX(28)CX(2)C', 'HX(29)CX(2)C', 'HX(32)CX(2)C', 'HX(34)CX(2)C'
dCMP_cyt_deam_2	'CX(32)CX(2)C', 'HX(32)CX(2)C'
malic	'KX(41)EX(0)DX(24)D'
tRNA-synt_1	'CX(2)C', 'CX(2)CX(13)CX(2)H', 'CX(2)CX(164)CX(2)C', 'CX(2)CX(204)CX(2)C', 'CX(2)CX(37)CX(1)C', 'CX(2)CX(41)CX(2)C', 'CX(2)CX(9)CX(2)C', 'CX(2)DX(42)CX(2)C', 'CX(20)CX(2)C', 'CX(40)CX(1)C', 'GX(36)C'
tRNA-synt_1c	'CX(1)CX(11)YX(3)C'

tRNA-synt_1e	'CX(180)CX(24)H', 'CX(180)CX(24)HX(3)E'
tRNA-synt_1f	'CX(2)HX(18)C', 'DX(3)CX(0)HX(5)H'
tRNA-synt_1g	'CX(180)CX(24)H', 'CX(180)CX(24)HX(3)E', 'CX(2)C', 'CX(2)CX(13)CX(2)H', 'CX(2)CX(164)CX(2)C', 'CX(2)CX(204)CX(2)C', 'CX(2)CX(37)CX(1)C', 'CX(2)CX(41)CX(2)C', 'CX(2)CX(9)CX(2)C', 'CX(2)DX(42)CX(2)C', 'CX(20)CX(2)C', 'CX(40)CX(1)C', 'GX(36)C'
tRNA-synt_2b	'CX(48)E', 'CX(50)H'
tRNA-synt_2d	'CX(2)CX(4)CX(2)C'
tRNA_SAD	'HX(3)H'
zf-4CXXC_R1	No pattern
zf-A20	'CX(2)C', 'CX(4)CX(11)CX(2)C'
zf-AD	'CX(2)CX(45)CX(2)C'
zf-AN1	'CX(1)CX(13)HX(1)C', 'CX(17)CX(2)C', 'CX(2)C', 'CX(2)CX(13)HX(1)C', 'CX(2)CX(17)CX(2)H', 'CX(2)CX(18)CX(2)H', 'CX(2)CX(20)CX(2)H', 'CX(2)CX(26)CX(2)C', 'CX(4)CX(17)CX(2)H'
zf-BED	'CX(2)CX(15)HX(4)H', 'CX(2)CX(19)HX(4)H'
zf-B_box	'CX(2)CX(12)HX(2)H', 'CX(2)CX(13)HX(2)C', 'CX(2)DX(11)HX(4)H', 'CX(2)DX(12)HX(2)H', 'CX(2)DX(13)HX(2)H', 'CX(2)HX(15)CX(2)C', 'CX(2)HX(16)CX(2)C', 'CX(2)HX(18)CX(5)H'
zf-C2H2	'CX(2)CX(11)H', 'CX(2)CX(12)HX(3)C', 'CX(2)CX(12)HX(3)H', 'CX(2)CX(12)HX(4)H', 'CX(2)CX(12)HX(5)H', 'CX(3)C', 'CX(3)H', 'CX(4)C', 'CX(4)CX(12)HX(3)H', 'CX(4)CX(12)HX(4)H', 'CX(4)CX(15)HX(3)H', 'CX(5)CX(12)HX(4)H'
zf-C2HC	'CX(4)CX(12)HX(5)C', 'CX(5)C'
zf-C2HC5	No pattern
zf-C3H1	No pattern
zf-C3HC	No pattern
zf-C3HC4	'CX(0)CX(20)CX(2)C', 'CX(1)CX(23)CX(2)C', 'CX(1)CX(24)CX(4)C', 'CX(1)H', 'CX(1)HX(13)CX(2)C', 'CX(1)HX(16)CX(2)C', 'CX(1)HX(17)CX(2)C', 'CX(1)HX(17)CX(2)D', 'CX(1)HX(19)CX(2)C', 'CX(1)HX(21)CX(2)C', 'CX(1)HX(22)CX(2)C', 'CX(1)HX(23)CX(2)C', 'CX(2)C', 'CX(2)CX(15)CX(2)C', 'CX(2)CX(16)CX(2)C', 'CX(2)CX(17)CX(2)C', 'CX(2)CX(17)HX(2)C', 'CX(2)CX(19)CX(2)C', 'CX(2)CX(19)HX(2)C', 'CX(2)CX(20)CX(2)C', 'CX(2)CX(20)HX(2)C', 'CX(2)CX(21)CX(2)C', 'CX(2)CX(21)HX(2)C', 'CX(2)CX(22)CX(2)C', 'CX(2)CX(23)CX(2)C', 'CX(2)CX(23)HX(2)C', 'CX(2)CX(31)HX(2)C', 'CX(2)CX(34)HX(2)C', 'CX(2)CX(8)CX(13)C', 'CX(2)H', 'CX(5)HX(21)CX(5)C'
zf-C4	'CX(13)CX(2)C', 'CX(2)C', 'CX(2)CX(13)CX(2)C', 'CX(3)CX(9)CX(2)C', 'CX(5)CX(9)CX(2)C'
zf-C4H2	No pattern
zf-C4_ClpX	'CX(2)C'
zf-C4_Topoisom	No pattern
zf-C5HC2	No pattern
zf-CCCH	'CX(7)CX(4)CX(3)H', 'CX(7)CX(5)CX(3)H', 'CX(8)CX(4)CX(3)H', 'CX(8)CX(5)CX(3)H'
zf-CCHC	'CX(2)CX(4)HX(4)C'
zf-CCHH	No pattern

zf-CHC2	'CX(2)HX(17)CX(2)C'
zf-CHCC	No pattern
zf-CHY	'CX(0)CX(9)HX(5)H', 'CX(1)HX(17)CX(2)C', 'CX(1)NX(0)CX(9)CX(1)EX(0)C', 'CX(2)C', 'CX(2)CX(9)CX(2)C'
zf-CSL	'CX(1)CX(19)CX(2)C'
zf-CW	'CX(4)CX(20)CX(10)C'
zf-CXXC	'CX(2)CX(2)CX(15)C', 'CX(2)CX(2)CX(32)C'
zf-DBF	No pattern
zf-DHHC	No pattern
zf-DNA_Pol	No pattern
zf-DNL	'CX(2)CX(21)CX(2)C'
zf-Dof	No pattern
zf-FCS	No pattern
zf-FPG_IleRS	'CX(19)CX(2)C', 'CX(2)CX(16)CX(2)C'
zf-GRF	No pattern
zf-H2C2	No pattern
zf-HIT	'CX(2)C', 'CX(2)CX(15)CX(3)C', 'CX(2)CX(19)CX(3)C'
zf-HYPF	No pattern
zf-LITAF-like	No pattern
zf-LSD1	No pattern
zf-LYAR	'CX(2)CX(11)HX(2)C'
zf-MIZ	'CX(1)HX(17)CX(2)C', 'CX(1)HX(19)CX(2)C', 'CX(1)HX(22)CX(2)C', 'CX(2)C', 'CX(2)CX(16)CX(2)C', 'CX(2)CX(17)CX(2)C'
zf-MYND	'CX(17)CX(3)H', 'CX(2)CX(12)HX(3)C', 'CX(2)CX(15)CX(3)C', 'CX(2)CX(16)CX(3)C', 'CX(2)CX(17)HX(3)C', 'CX(3)C'
zf-NADH-PPase	'CX(2)C', 'CX(2)CX(14)CX(2)C', 'CX(2)CX(3)H'
zf-NF-X1	No pattern
zf-NPL4	No pattern
zf-P11	No pattern
zf-PARP	'CX(2)CX(28)HX(2)C', 'CX(2)CX(30)HX(2)C'
zf-RAG1	No pattern
zf-RING-like	'CX(2)CX(17)HX(2)C', 'CX(2)CX(20)CX(2)C'
zf-RNPHF	No pattern
zf-RanBP	'CX(2)CX(10)CX(2)C', 'CX(4)CX(10)CX(2)C'
zf-Sec23_Sec24	'CX(17)CX(2)C', 'CX(18)CX(2)C', 'CX(2)C', 'CX(4)CX(18)CX(2)C'
zf-TAZ	'HX(3)CX(12)CX(4)C', 'HX(3)CX(4)CX(2)C', 'HX(3)CX(4)CX(4)C', 'HX(3)CX(5)CX(4)C', 'HX(3)CX(7)CX(2)C'
zf-TRAF	'CX(2)CX(11)H', 'CX(3)C', 'CX(3)CX(11)HX(3)C', 'CX(3)CX(11)HX(4)C', 'CX(3)CX(12)HX(4)C', 'CX(6)CX(11)H', 'HX(15)C'

zf-Tim10_DDP	No pattern
zf-U1	'CX(2)CX(14)HX(5)H'
zf-UBP	'CX(12)C', 'CX(16)CX(6)C', 'CX(19)CX(6)C', 'CX(2)CX(15)HX(5)H', 'CX(2)CX(16)HX(8)H', 'CX(2)H', 'CX(4)CX(15)HX(5)H', 'CX(7)C'
zf-UBR	No pattern
zf-XS	No pattern
zf-ZPR1	'CX(2)C'
zf-dskA_traR	'CX(2)CX(17)CX(2)C'
zf-nanos	No pattern
zf-piccolo	No pattern
zf-primase	No pattern

Copper proteomes, phylogenetics and evolution†

Leonardo Decaria,[a] Ivano Bertini[ab] and Robert J. P. Williams*[c]

Received 9th September 2010, Accepted 15th October 2010
DOI: 10.1039/c0mt00045k

This paper is a continuation of our study of the connection between the changing environment and the changing use of particular elements in organisms in the course of their combined evolution (Decaria, Bertini and Williams, *Metallomics*, 2010, **2**, 706). Here we treat the changes in copper proteins in historically the same increasingly oxidising environmental conditions. The study is a bioinformatic analysis of the types and the numbers of copper domains of proteins from 435 DNA sequences of a wide range of organisms available in NCBI, using the method developed by Andreini, Bertini and Rosato in *Accounts of Chemical Research* 2009, **42**, 1471. The copper domains of greatest interest are found predominantly in copper chaperones, homeostatic proteins and redox enzymes mainly used outside the cytoplasm which are in themselves somewhat diverse. The multiplicity of these proteins is strongly marked. The contrasting use of the iron and heme iron proteins in oxidations, mostly in the cytoplasm, is compared with them and with activity of zinc fingers during evolution. It is shown that evolution is a coordinated development of the chemistry of elements with use of novel and multiple copies of their proteins as their availability rises in the environment.

Introduction

It is conventional today to analyse evolution both by comparative studies of organisms following Darwin or of DNA sequences from organisms present today using mathematical methods to deduce their history. Both methods are aided by the dating of fossils. They are very effective in tracing development following the Cambrian Explosion 0.54 billion years ago. As we stated in our previous paper[1] the procedures do not describe well the evolution of organisms before this time and there are few helpful fossils of dates before 0.54 billion years ago. The fossils of this earlier period are largely imprints of soft-bodied organisms difficult to classify and to relate with certainty to today's organisms and of uncertain DNA content. It is our belief that in such circumstances the most revealing evidence of evolution lies in the changing nature of the chemical environment, largely of inorganic ions, together with the deduced evidence of organism inorganic chemistry, especially the metallome, in the agreed evolutionary order of anaerobic and then aerobic prokaryotes followed by single-cell and then multicellular eukaryotes from 3.5 to 0.5 or 0.4 billion years ago (Ga).[2,3] From the quantitative evidence of the amounts of trace elements, of various element ratios and of isotope distribution in sediments it has proved to be possible to give a record of the likely availability of elements in the sea as they changed with time. The principal effect is due to the gradual rise in atmospheric oxygen giving rise to more oxidising conditions in the sea. The redox potential has risen from about −0.4 (anaerobic) at 3.5 Ga in the original oceans to +0.4 volts (aerobic) today with the solubilisation of elements from sulfide minerals. In parallel with these analytical studies we and others have examined the general uses of the elements in proteins in organisms, that is their metallomes, using bioinformatic analysis of organisms extant today, *e.g.* modern prokaryotes, plants and animals while judging their times of evolution from general biochemical studies.[4–8] This paper will give details of the presence of copper proteins in organisms looking especially at the duplication of the copper proteins which have evolved with related properties much as we did in the study of zinc.[1]

Methods

We have investigated 435 complete proteomes, 52 from archaea, 337 from bacteria (247 aerobic and 90 anaerobic) and 46 from eukarya available in NCBI. Our chosen example here is that of the copper proteins, primarily involved in homostasic, carrier (chaperone) functions, redox reactions and electron transfer. Knowledge of the site structures allows us to recognise the copper binding domains in a protein. To obtain our starting data set, which consists of 44 proposed Cu-binding domains, some of them with one or more metal associated binding patterns (MBP) (Table S1, ESI†), we used the prediction method published by the group of one of us.[7] As reported in that reference paper, we applied the HMMER program to search the NCBI refseq proteins database for matches to the hidden Markov models (HMMs) representing the selected domains. The HMMs were taken from the Pfam database without modification. We selected 10^{-3} as Evalue cut-off. For multi-domain proteins (*e.g.* ammonia monooxygenase made from 3 distinct Pfam

[a] *Magnetic Resonance Center (CERM) – University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy*
[b] *Department of Chemistry – University of Florence, Via della Lastruccia 3, 50010 Sesto Fiorentino, Italy*
[c] *Department of Inorganic Chemistry – University of Oxford, South Parks Road, Oxford, UK OX1 3QR. E-mail: bob.williams@chem.ox.ac.uk*

† Electronic supplementary information (ESI) available: Additional data—the 44 retrieved Cu-binding domains with the eventual Metal Binding Pattern. See DOI: 10.1039/c0mt00045k

Table 1 Pfam domains composition of the analyzed Cu–protein

Pfam domain	No. of bound ions	Function
Ald_Xan_dh_C2	1	Electron_carrier
Monooxygenase_B	3	Ammonia monooxygenase
Biopterin_H	1	Aromatic-AA hydroxylase
CcoS	1	Copper_chaperone
CdhC	1	Carbon monoxide dehydrogenase
Cmc1	1	Copper_chaperone
CopB	1	Copper_homeostasis
CopC	1	Copper_homeostasis
CopD	1	Copper_homeostasis
Copper-bind	1	Electron_carrier
Copper-fist	1	Transcription
COX1	1	Electron_carrier
COX17	1	Copper_chaperone
COX2	1	COX 2
CtaG_Cox11	1	Copper_chaperone
Ctr	1	Copper_homeostasis
Cu_amine_oxid	1	Amine oxidase
Cu_bind_like	1	Electron_carrier
Cu2_monoox_C	1	Ascorbate dep. Monooxygenase
Cu2_monooxygen	1	
Cu-oxidase	1	Laccase-like
Cu-oxidase_2	1	
Cu-oxidase_3	2	
Cu-oxidase_4	4	Laccase-like
CutA1	1	Copper_homeostasis
CutC	1	Copper_homeostasis
Glyco_hydro_10	1	Glycosyl hydrolase
Hemocyanin_M	1	Copper_homeostasis
HMA	1	Copper_homeostasis
Lysyl_oxidase	1	Lysyl_oxidase
Metallothio	1	Copper_homeostasis
Metallothio_11	1	Copper_homeostasis
Metallothio_5	1	Copper_homeostasis
Metallothio_Pro	1	Copper_homeostasis
Metallothionein	1	Copper_homeostasis
NlpE	1	Copper_homeostasis
NosD	1	Copper_chaperone
NosL	1	Copper_chaperone
Sod_Cu	1	Superoxide dismutase
Tyrosinase	1	Tyrosinase
Uricase	1	Uricase

domains) we considered as true positives only the retrieved sequences containing all the reference Pfam domains. Table 1 reports the Pfam domain composition of all the related Cu-binding proteins and their functions, and the number of ions bound.

Results

In order to give a comparative account of the data and their analysis we have considered organisms in the following ways: the prokaryotes have been divided into the major groups of archaea and eubacteria, and both have been further divided into aerobic and anaerobic. The average content of copper proteins of each prokaryotic group has been used for comparative purposes. Amongst eukaryotes we have divided them into single-cell and multicellular examples and then considered them with respect to their complexity, using selective organisms in the order: single-cell eukaryotes *S. cerevisiae*, *T. brucei* and *P. falciparine* and multicell eukaryotes, *C. elegans*, *D. melanogaster*, *A. thaliana* and *H. sapiens*. Each chosen organism has been observed to be similar in DNA sequences and numbers of duplications to those in several other organisms in the group to which it belongs. The particular organism described can therefore be taken as indicative of the nature of a group.

The activities of the proteins in all the organisms have been divided using their major four separate functions: homeostatic proteins, chaperones, electron transfer proteins and oxidases. The homeostatic proteins include the metallothioneins and the copper pumps. The oxidases are treated at first as a sum of all such enzymes but later we shall discuss their further functional divisions together with the superoxide dismutases. We shall not refer to either transcription factors or hydrolytic enzymes which were the major groups in our analysis of zinc proteins.

We turn now to a more detailed description of the copper oxidases which can be divided in three ways: by the number of copper sites, by the structural nature and domains in one protein, and by their organic substrates, Table 1. The numbers of copper atoms vary from 1 to 4 (and perhaps one or two more in caeruloplasmin) and the enzymes are grouped under the Enzyme Commission EC.1 label. The types of copper are also described structurally as Type I (electron transfer proteins with one copper), Type II (a single copper) with Type III (a pair of linked copper atoms) where Types II

and III ions form the site of reaction of oxygen in the complicated oxidases such as lactase, EC.1.10.3, and ascorbate oxidase, EC.1.14.17. Here oxygen goes directly to water and oxidation of substrate is at a remote site. Of the other enzymes some such as galactose oxidase and amine oxidases, EC.1.4.3, have but one copper while tyrosinase, also known as catechol oxidase, EC. 1.14.18, has two linked coppers. Finally Superoxide Dismutase EC.1.15 has a copper close to a zinc site. Now these oxidases can be separately recognised in the genome by the way that copper atoms are chelated or their cofactors bind, see Table S1, ESI† and methods above. Some of the copper-dependent hydroxylases are dependent on an initial reduction of the organic substrate with release of one water molecule as they introduce one atom of oxygen only into the organic substrate much as does cytochrome P-450. We know of at least three reducing cofactors, NADH or NADPH, pteridine and ascorbate. We present the data either in terms of total numbers of copper proteins as in the Tables or as percentages of the genome as in the Figures.

There are extremely few, perhaps no, copper proteins in all the anaerobic archaea or eubacteria. There are only a few copper proteins in any of the four classes in aerobic archaea or eubacteria, one with a total genome less than 1500 and the other with a greater number. The data on EC.1 oxidases are given in Table 2. In fact there no noticeable differences in copper proteins between the bacteria of low gene and those of high gene content (not shown). Aerobic prokaryotes and all eukaryotes, animals and plants, have a copper domain in cytochrome oxidase but it is coded in the mitochondrial DNA in eukaryotes and is not included in our search. Chloroplasts in plants also have a copper electron transfer protein, plastocyanin, and it too is not included in our analysis.

Striking features in the eukaryotes are the rapid increase in the numbers of all four groups of copper proteins with complexity of the multicellular organisms and the even greater increase in plants, illustrated by Arabidopsis. The data for EC.1 oxidases are given in Table 2 but they do not show in the percentages, Fig. 1. The fungi form a group with a relatively steady number of all four kinds though in larger numbers than in the animals (not shown).

Discussion

The description and analysis of copper proteins and their probable evolution of them all jointly has been described in several previous publications.[2–10] The main conclusions are that copper was not used by the earliest anaerobic prokaryotes, as it was not an available element before there was oxidation of sulfides. Free copper ions in organisms are known to be poisonous and hence cells have always had proteins for maintaining a very low level of total copper, especially in their membranes and cytoplasm. The control is managed through storage in homeostatic buffer proteins, such as metallothioneins in the cytoplasm and entry and exit pumps in the outer membrane. However copper became more and more valuable in cells as oxygen became more available, especially in oxidases in eukaryotic vesicles and outside cells. Here iron cannot be used since the ferrous ion is readily dissociated from proteins, oxidised and loses its function. Even the porphyrin of heme iron is susceptible to oxidation. The value of copper therefore increased externally as seen from unicellular to multicellular eukaryotes, Table 2. Its enzymes are used in the synthesis of extracellular matrices, oxidases for cross-linking phenolic units in plants and lysine oxidase in animals, absent in plants, for cross-linking of collagens. Many oxidative processes are required, especially for the production of messenger organic molecules such as adrenaline and amidated peptides in vesicles more notably in animals than in plants, Table 3. The presence of copper generally also raises risks such as chemical oxidative stress, possibly associated with Alzheimer's Disease for example, due to partial reduction oxygen to superoxide. Superoxide is removed rapidly by Cu/Zn superoxide dismutases. In this paper we have approached the problem of the evolution and these uses of copper in organisms in a different way from the above general descriptions by examining the duplication of the enzymes.

We shall follow the ideas which Ohno pointed out that while mutation can improve individual protein function it can hardly provide new functions without impairing the existing function of a protein.[11] Duplication is therefore essential for novel functions prior to mutation.[12] Tables 2–4 and Fig. 1–3 show that duplication is very extensive amongst certain copper proteins as it was amongst particular but different zinc proteins. We observe first the great difference between the copper and zinc proteins described previously. There are extremely few copper transcription factors or hydrolases in

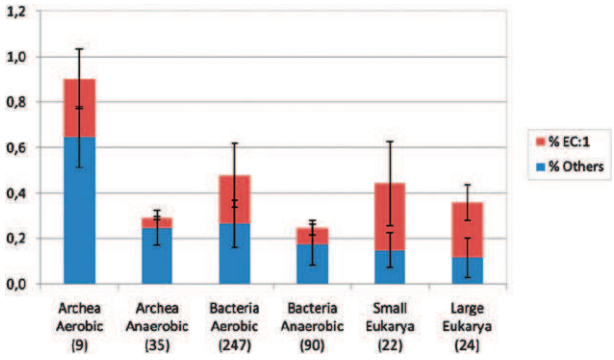


Fig. 1 The total percentages of copper proteins including oxidases EC.1 in prokaryotes and eukaryotes. The percentages must be taken together with the total numbers so that the diversity of copper proteins, much greater in eukaryotes, Table 2, can be appreciated. The numbers in brackets refer to the toal number of organisms in each group.

Table 2 The numbers of total and EC : 1 (oxidoreductases) Cu–proteins for the analyzed groups of organisms. * = average value for archea, aerobic and anaerobic bacteria

	Proteome	No. Total	No. EC:1
Archea (*)	2176	8	1
Bacteria Anaerobic (*)	2749	6	1
Bacteria Aerobic (*)	3792	18	8
S. cerevisiae	5880	29	12
P. falciparum	6265	7	1
T. brucei	9279	5	2
C. elegans	22 844	46	26
D. melanogaster	20 513	70	47
H. sapiens	37 742	82	54
A. thaliana	32 165	245	144

Table 3 *H. sapiens* and *A. thaliana* copper proteins content comparison

Function	*H. Sapiens*	*A. thaliana*
Copper_chaperone	4	6
Copper_homeostasis	21	68
Laccase	5	39
Monooxygenase	0	0
Aromatic-AA hydroxylase	6	0
Ascorbate dep. Monooxygenase	12	0
Tyrosinase	3	0
Ammonia monooxygenase	0	0
Superoxide dismutase	3	8
Carbon monoxide dehydrogenase	0	0
Amine oxidase	4	10
Lysyl_oxidase	5	0
Uricase	0	1
COX2	2	1
Multicopper-oxidase	2	2
Glycosyl hydrolase	0	4
Electron_carrier	3	23
Transcription	0	0

marked contrast to those of zinc,[1] which is in virtually no oxidases, and they are largely in different cell compartments. These features are indicative of the separate nature of the two metals. Copper is of use in oxidations as it can change valence but, as stated above, it presents a risk, especially in association with the cell nucleus. Zinc is more available and useful for hydrolytic reactions, it is nearly as powerful a Lewis acid as copper but unlike copper it cannot catalyse redox reactions. It can also act in signalling even to the nucleus in transcription factors as it is of low risk. The data show that duplication is therefore very selective to both different metal ions, proteins and enzymes and is characteristic of particular groups of organisms. For example oxidases are in greater numbers in plants but hydroxylases and transcription factors are more numerous in animals, Table 3. We therefore have to consider that the multiplied functions are for selected purpose—copper in certain oxidases, different in different organisms, and zinc in certain hydrolases and transcription factors. By far the greatest multiplications are seen in enzymes required for either the management of connective tissue and of messenger systems, copper for transmitters, zinc for hormones—both for external products. Moreover as we show in Table 4 there are large increases in the heme iron cytochrome P-450, also valuable in hormone synthesis, and the ferrous oxy-glutamate-dependent oxidases in parallel with the increases of copper oxidases. It

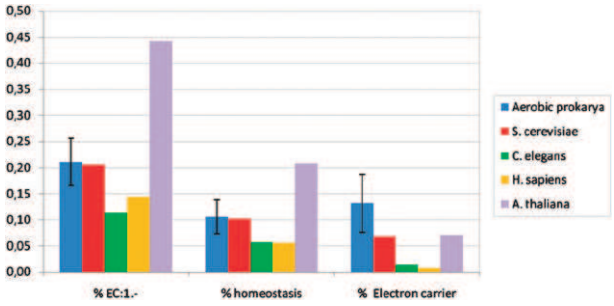


Fig. 2 The percentage of three kinds of copper domains from five different organisms, see Table 1 for numbers.

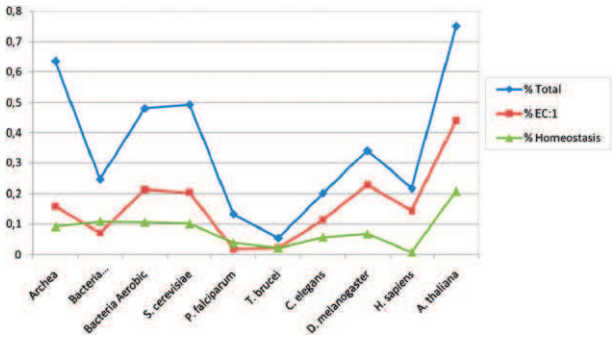


Fig. 3 A graphical presentation of the percentage of domains of some proteins in a more extensive list of organisms. Tables 2 and 3 give numbers for these organisms and their total domain size as well as data for some proteins of very low multiplicity.

would appear that duplication is not random though subsequent mutation may be but is perhaps preferentially in the duplicated proteins.[12,13]

In our previous paper[1] we drew attention to the parasitic organisms *plasmodia* and *trypanosomes* but they were not outstandingly different from other single-cell eukaryotes in zinc protein content. In the case of the copper enzymes we observe that the parasites have very few if any oxidases or other copper proteins except two or three for homeostasis or which act as chaperones. They then behave as single-cell eukaryotes with little oxygen chemistry. Thus they show loss of particular enzymes much as did higher eukaryotes when they became dependent on lower organisms for synthesis of many coenzymes and so require vitamins. We must ask how

Table 4 Comparison among Cu-oxidoreductases, Fe-dependent oxygenases, Fe-binding p450 proteins and heme-binding peroxidases contains for the analyzed groups of organisms. * = average value for archea, aerobic anaerobic bacteria

	No. Cu EC:1 (oxidoreductases)	No. Fe-dependent oxygenases	No. Fe p450	No. Heme peroxidases
Archea (*)	1	0	0	0
Bacteria Anaerobic (*)	1	0	0	0
Bacteria Aerobic (*)	8	1	5	1
S. cerevisiae	12	1	3	1
P. falciparum	1	0	0	0
T. brucei	2	9	2	0
C. elegans	26	8	76	14
D. melanogaster	47	26	97	14
H. sapiens	54	9	70	16
A. thaliana	144	116	268	194

Note. No. Cu EC:1 is the number of copper domains and some proteins have three or four domains, see Table 1.

these developments of genes occur during evolution and at particular times such as the gain of copper enzymes with the rise in oxygen and copper and the losses of some of these enzymes with symbiosis.

Why are the oxidases of copper, cytochrome P-450 or the Fe(II)OG types all so greatly multiplied in plants relative to the numbers in animals, Table 4? These oxidases have a protective value as well as one in synthesis. It is very likely that the seed of a plant as it forms will be more exposed to adverse chemicals than the highly protected reproduction modes of animals. In particular the plants produce the oxygen used in these enzymes and accidentally produce both superoxide and hydrogen provide as well as the adventitious erroneous oxidation of the organic substrates and products of them. Oxygen and these products in cells as well as the copper and other metal ions are hazards, particularly to the enzymes themselves in plants. This gives a reason for the generally higher numbers of oxidases in plants together with those of homeostatic and chaperone proteins since the supply of copper must be kept from damaging the cytoplasm. The free copper is reduced to 10^{-15} in the cytoplasm of all cells while zinc is held at 10^{-10} M. The possible explanation of the duplication of oxygenases is that oxygen itself is the cause by two means. It can damage DNA directly but this does not explain selectivity of duplication or it can stress the production of proteins by damaging them. The most likely proteins to be damaged are those which use oxygen and when damage occurs production of them must increase. The stress affects the DNA in that increased production of a protein requires greater local exposure of its coded DNA and can lead to mismatching of DNA strands at this site during reproduction. Mismatching is a know cause of duplication. Different stresses of many kinds can affect proteins of the external matrices and also messenger systems associated with production of messengers and hormones, through damage to their copper and iron oxidases, and of messenger receptors, zinc fingers.

The most obvious simple stress due to oxygen is that of increase of both copper and zinc in the environment requiring multiplication of homeostatic and chaperone proteins. Any such possible sensitivity to stress has to be tested experimentally as a possible explanation of the particular multiplication and appearance of these useful products of oxygen generation which are also causes of stress. Is stress a major cause of evolution in the sequence

novel chemicals (from oxygen) → stress → multiplication of protective proteins with mutation of the proteins (which bind or are affected by the stress) → further multiplication followed by further mutation to give novel organisms?

Acknowledgements

We wish to thank Dr R. E. M. Rickaby and Dr L. Dupont for many valuable exchanges of views.

References

1 L. Decaria, I. Bertini and R. J. P. Williams, *Metallomics*, 2010, **2**, 706–709.
2 R. J. P. Williams and J. J. R. Fraùsto da Silva, *The Chemistry of Evolution*, Elsevier, Chichester, 2006.
3 R. J. P. Williams and R. E. M. Rickaby, submitted for publication.
4 J. J. R. Fraùsto da Silva and R. J. P. Williams, *The Biological Chemistry of the Elements*, Oxford University Press, Oxford, 2nd edn, 2001.
5 C. L. Dupont, A. Butcher, R. E. Valas, P. E. Brown and G. Caetano-Anolles, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 10567–10572.
6 Y. Zhang and V. N. Gladychev, *Chem. Rev.*, 2009, **109**, 4828–4861.
7 C. Andreini, Lbunci, I. Bertini and A. Rosato, *J. Proteome Res.*, 2008, **7**, 209–216.
8 A. D. Anbar and A. H. Knoll, *Science*, 2002, **297**, 1137–1142.
9 M. A. Saito, D. M. Sigman and F. M. M. Morel, *Inorg. Chim. Acta*, 2003, **356**, 308–318.
10 D. Magnani and M. Sohoz, in *Bacterial Transition Metal Homeostasis*, ed. D. H. Nies and S. Silver, Springer, Heidelberg, 2007, pp. 259–285.
11 S. Ohno, *Evolution by Gene Duplication*, Springer, Heidelberg, 1970.
12 E. V. Kooning, *Nucleic Acids Res.*, 2009, **37**, 1011–1034.
13 M. H. Serves, A. R. Kerr, T. J. McCormack and M. Riley, *Biol. Direct*, 2009, **4**, 46–54.

Supplementary Material (ESI) for Metallomics

Supplementary Data

Table S1 – The 44 retrieved Cu-binding domains with the eventual Metal Binding Pattern.

Pfam domain	*Cu-binding pattern*
Ald_Xan_dh_C2	CX(0)S
AMO	No pattern
AmoC	No pattern
Biopterin_H	No pattern
CCoS	No pattern
CdhC	CX(85)CX(1)C
Cmc1	No pattern
CopB	No pattern
CopC	MX(10)M
CopD	No pattern
Copper-bind	HX(37)CX(2)HX(2)M
Copper-fist	No pattern
COX1	HX(49)HX(0)H
COX17	CX(2)C
COX2	HX(34)CX(1)QX(1)CX(3)HX(2)M
COX2_TM	No pattern
CtaG_Cox11	No pattern
Ctr	No pattern
Cu_amine_oxid	YX(48)HX(1)HX(158)H
Cu_bind_like	HX(42)CX(4)HX(4)Q
Cu2_monoox_C	HX(1)HX(69)M
Cu2_monooxygen	HX(0)HX(63)H
Cu-oxidase	HX(42)CX(4)H
Cu-oxidase_2	HX(40)CX(8)HX(4)M
Cu-oxidase_3	HX(40)CX(7)HX(4)M
Cu-oxidase_4	No pattern
CutA1	DX(0)K
CutC	No pattern
Glyco_hydro_10	DX(3)EX(59)H
Hemocyanin_M	HX(3)HX(25)H
HMA	CX(0)AX(1)C
Lysyl_oxidase	No pattern
Metallothio	No pattern
Metallothio_11	No pattern
Metallothio_5	No pattern
Metallothio_Pro	No pattern
Metallothionein	No pattern
Monooxygenase_B	HX(23)HX(331)Q

NlpE	No pattern
NosD	No pattern
NosL	No pattern
Sod_Cu	HX(1)HX(22)HX(54)H
Tyrosinase	HX(18)HX(8)HX(99)HX(3)HX(26)H
Uricase	No pattern

A simple protocol for the comparative analysis of the structure and occurrence of biochemical pathways across superkingdoms

Claudia Andreini[1,2], Ivano Bertini[1,2,*], Gabriele Cavallaro[1], Leonardo Decaria[1], Antonio Rosato[1,2]

[1]Magnetic Resonance Center (CERM) – University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

[2]Department of Chemistry – University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Keywords: proteome; metalloproteins; protein domains

Running title: Searching genomes for biological pathways

Corresponding Author:

Prof. Ivano Bertini

Magnetic Resonance Center

University of Florence

Via Luigi Sacconi 6

50019 Sesto Fiorentino (Italy)

e-mail: ivanobertini@cerm.unifi.it

Summary

A biochemical pathway can be viewed as an ensemble of a number of different proteins, each of which typically contributes a step along a biochemical process within the cell. These processes can be of a very different nature, including for example biosynthesis or catabolism of compounds, and substrate transport. Computational methods can be applied to assess whether one organism is able to perform a biochemical process of interest by checking whether its genome encodes all the protein components that are known to be necessary for the task. Here we present a simple strategy for collecting the above data that is based on, but not limited to, our experience on processes involving metal ions and metal-binding cofactors. The strategy is fully implemented in a bioinformatics package, RDGB, which is available from http://www.cerm.unifi.it/home/research/genomebrowsing.html. The use of RDGB allows users to perform all the operations that are needed to implement the aforementioned strategy with minimal intervention and to gather all results in an ordered manner, with a tabular summary. This minimizes the (bio)informatics needed, thus facilitating non-experts. As examples, we analysed over more than a thousand prokaryotes the pathways for the degradation of organic compounds containing one or two aromatic rings as well as the distribution of some proteins involved in Cu_A assembly.

Introduction

Thanks to the success of genome sequencing projects, it is possible to perform experimental and computational studies at the whole genome and/or whole proteome level, which leverage on the availability of a potentially complete list of the proteins codified by a living organism. In particular, there has been a great deal of interest in the identification of so-called functional modules, i.e. groups of proteins working together for the same cellular function. A typical case is that of enzymatic pathways in metabolic networks. Over the years, a huge portfolio of tools has been developed by a great number of different bioinformaticians all over the world to reconstruct such modules by computationally predicting functional relationships among the proteins encoded by genomes. Three main methodologies have been deployed to this aim: the occurrence of gene fusion events (Rosetta-stone) (1;2); the conservation of gene order (3-5); the similarity of phylogenetic profiles (6;7). Combinations of the above (and other) methodologies have also been developed (8-10). The intended use of the results output by these tools, which typically include the identity of the partners of each protein along a pathway in one or more organisms and their functional linkages, is to drive experimental studies aimed at e.g., defining the role of uncharacterized proteins in the pathway (11-13) or supporting more complex computational studies (14-16). Alternatively, it is possible to obtain information on already characterized pathways thanks to specialized databases, such as KEGG (17;18) or BioCyc (19). These databases contain information on metabolic pathways extracted from the relevant literature with manual curation. Within the frame described here, they are useful to identify which enzymes are known to be part of a given pathway.

Despite this wealth of available resources, it is not entirely obvious how to use the knowledge of the protein components that make up a biochemical pathway in an individual organism or set of organisms for tasks such as the investigation of how widespread is a process throughout the domains of Life, or the identification of its possible variants. Whereas there are again many computational tools available to facilitate these tasks, setting up a consistent strategy to use them against hundreds of genomes and extract reliable data may not be trivial. We have extensively faced this difficulty when trying to investigate pathways in the biosynthesis, assembly, or transport of different metal-containing cofactors (20;21) or to characterize the occurrence in proteomes of some metalloproteins (20-24). To tackle this kind of studies, we have developed a number of scripts and programs that e.g. automate the download and interrogation of databases or the identification of specific amino acidic patterns such as metal-binding patterns (25). Commonly, the proteome-level analysis of the occurrence of a biochemical

pathway is based on the identification of homologues of all the involved proteins. Again, various computational methods can be used to this end. The detection of bidirectional best hits, often done with the BLAST program (26), is one of the most widely used approaches (3;27). Another approach is based on the identification of conserved domains through the use of profiles (28).

In this work, we describe a coherent, easy protocol for the identification of a set of proteins that can constitute an entire biochemical pathway on the basis of homology relationships detected through the presence of conserved domains and integrating, when available, 3D structural information. This protocol integrates all the tools that we have developed and tested in our previous publications (21) into a single package, which we called RDGB (Retrieval of Domains and Genome Browsing). RDGB not only integrates all the needed scripts and makes them easy to use for non-experts but enforces the use of a tested, internally consistent protocol in order to guarantee the reliability of the results. In addition, it provides a pre-ordered manner of storing the data which can be useful for subsequent analyses as well as further computational analyses.

As an example, we analyzed the degradation of aromatic hydrocarbons in 1136 completely sequenced prokaryotic genomes. Aromatic hydrocarbons such as toluene or biphenyl are common contaminants of soil and groundwater and are listed as priority pollutants by the U.S. Environmental Protection Agency (http://www.epa.gov/waterscience/criteria/wqcriteria.html), either as single compounds or in mixtures. One of the most attractive means to remove these compounds from polluted environments is through bioremediation (29;30). Numerous bacterial strains have been isolated for the ability to aerobically degrade a variety of aromatic hydrocarbons (31). The genes encoding the enzymes needed for the biodegradation of aromatic hydrocarbons can be located either in plasmids or in chromosomal DNA. The bacterial degradation of aromatic hydrocarbons consists of many reaction steps, which have often been broadly separated into peripheral and central pathways. Peripheral pathways convert a large proportion of different aromatic hydrocarbons into a limited number of key central intermediates, such as catechol and protocatechuate. The aerobic degradation of aromatic compounds is frequently initiated by ring-hydroxylating oxygenases (32), which catalyze the incorporation of two oxygen atoms into the aromatic ring to form arene cis-diols, followed by a dehydrogenation reaction catalyzed by a cis-dihydrodiol dehydrogenase to give catechol or substituted catechols which serve as substrates for oxygenolytic aromatic ring cleavage (33).

As a further example, we investigated the occurrence of proteins involved in the assembly of the Cu_A cofactor. Cu_A is a redox-active cofactor that contains two copper ions; in the reduced state both ions are in the +1 state. Upon one-electron oxidation of the cofactor, a mixed-valence species forms

where the two copper ions are formally in the +1.5 oxidation state. The Cu_A cofactor is contained within the soluble domain of subunit II (Cox2) of prokaryotic and eukaryotic cytochrome *c* oxidases or within a homologous C-terminal domain of prokaryotic nitrous oxide reductase. Its physiological function is to shuttle the electron that it acquires from cytochrome *c* (either soluble or membrane-anchored) to the catalytic core of the enzyme, where it is used to reduce dioxygen or nitrous oxide, respecitvely. The correct assembly of Cu_A is crucial for enzyme's function. The assembly process is relatively complex, and a number of ancillary proteins have been implicated in it (34). NMR studies have demonstrated that in *Thermus thermophilus* copper(I) ions are delivered to the Cu_A binding site of Cox2 by a periplasmic metallochaperone called PCu_AC, while a second protein, Sco1, is responsible for maintaining the correct oxidation state of the Cys ligands in the Cu_A binding site by acting as a thiol-disulfide oxidoreductase (35). Interestingly, Sco1 can also bind copper(I) or copper(II) ions but this ability does not seem important for the assembly of Cu_A; an interplay between the oxidoreductase and metallochaperone activities of Sco1 proteins has been proposed based on computational studies (23).

Methods

Overview of the computational approach

The RDGB (Retrieval of Domains and Genome Browsing) tool can be run on computers having Linux as their operating system. It is written in python and uses a variety of different scripts and programs, which we have developed in the past few years (20-24), contained in the subfolder *0Tools* that is created upon installation. From the user's point of view, it is important to note that RDGB is divided in two main python scripts: *Retrieving_domains.py* and *Genome_browsing.py*, which are described in detail below. The two scripts are run consecutively as the first one builds part of the input to the second script. Python version 2.4.3 or higher is needed, with the following modules installed: *Bio, decimal, ftplib, math, os, pickle, re, string, sys, time, urllib*.

In the present strategy, we use the protein domains defined in the Pfam library (36;37) to identify putative homologues of the proteins involved in the pathway in any desired genome or list of genomes. When not already known, the domains can be initially identified in the sequence of proteins of known 3D structure that are available from the PDB (38). In our experience this is quite useful when trying to collect ensembles of proteins that can bind the same ligand, as sometimes not all the domains that can do this have been annotated as such in Pfam. Instead, if the ligand is present in the 3D structure of the protein, this information can be readily extracted from the PDB database together with the

pattern of amino acids that are involved in the interaction of the protein with it. The latter is called the Ligand Binding Pattern (LBP) and is defined by the identity and spacing of the amino acids, e.g., $CX_4CX_{20}H$, where X is any amino acid. As discussed in more detail in the next sections, this pattern can be usefully applied as a filter to reduce the number of false positives (i.e. of the proteins predicted to bind the cofactor but which in reality are unable to bind it) by rejecting the proteins that lack the LBP. The script *Retrieving_domains.py* performs the identification of domains in PDB structures and of the corresponding LBP's, and downloads the relevant Hidden Markov Models (HMM's) (39) that describe the domain from the Pfam database for the subsequent proteome searches. These data can be used independently.

The *Genome_browsing.py* script, which should be run after *Retrieving_domains.py*, downloads the proteomes of the organisms of interest from the ftp site of the NCBI (40), and then identifies the sequences containing the domains previously retrieved executing *Retrieving_domains.py*. For the latter step it uses the HMMER 3.0 program (http://hmmer.janelia.org)(39). The retrieved sequences are subjected to two filters: i) for the presence of Pfam domains not included in the user's selection that match the same region of a selected domain with a better (i.e. lower) HMMER E-value; ii) for the presence of the LBP, if available.

The computational flow chart is shown in Figure 1.

Retrieving_domains.py

The main purpose of this script is to download the HMM's that describe the Pfam domains to be identified in the entire proteome sequences by the next script. These can be supplemented with LBP's, when relevant. This part of the procedure must thus start with assembling a list of the domains of interest. These can be i) directly input by the user or ii) obtained from the analysis of sequences with known 3D structure or iii) both (Figure 1). Only in the case in which an user wants to extract the domains from the sequence of a protein of unknown structure, s/he should independently scan the sequence for Pfam domains using the interface at the Pfam web site (http://pfam.sanger.ac.uk/search). In i) the user is asked to provide a list of Pfam domains, for which the script downloads the corresponding HMM's from the Pfam library. In ii), the user inputs a list of PDB codes, whose protein sequences are downloaded from the PDB and scanned, using the HMMsearch function of HMMER 3.0, against the entire Pfam database to identify the domains they contain. If a PDB entry contains multiple chains, all the chains that are different in sequence are analyzed. Optionally, a list of ligands (identified by the three-letter chemical component identifier in the PDB database, corresponding to the

HET field; see http://deposit.pdb.org/cc_dict_tut.html#PDBformat) can be input, in which case the script will also identify the LBP and associate it to the Pfam domain within whose boundaries the amino acids of the LBP are (at least two amino acids must be within a domain to create an association; only the amino acids that are within the domain are then taken into account as the LBP). In iii), the input data and results of i) and ii) are joined. Note that in ii) and iii), the user is asked to provide a threshold to decide whether the identification of a domain within a sequence is meaningful or not. This is done by setting an upper limit for the expectation value (E-value), which is a measure of the expected rate of errors in the identification of domains in protein sequences. Typical values are in the range 10^{-3}-10^{-5}. The results are optionally stored in separate subfolders (Figure 1). In all cases the *0Profiles* subfolder is created, which contains the HMM models downloaded, as well as the log file of the script.

Genome_browsing.py

The second main script, *Genome_browsing.py*, asks the user for a list of organisms of interest (Figure 1). The corresponding proteomes are downloaded from the NCBI ftp site (ftp.ncbi.nih.gov/genomes/) and then scanned for the occurrence of proteins containing the Pfam domains from the previous step. The proteomes retrieved include all chromosomally encoded proteins as well as those encoded by plasmidic DNA. This is done by using the HMM models stored in the *0Profiles* subfolder with the HMMsearch function of HMMER 3.0. For the successful download of the proteomes, it is important that the name of the organisms is written exactly as listed at the NCBI, including, when relevant, the subspecies information (e.g. *Burkholderia xenovorans LB400*). A script is provided to obtain the lists of all prokaryotic and eukaryotic organisms whose full proteome is available, from which the names can be pasted (*Retrieving_Organisms.py*).

This script creates one folder per each organism in the list, in which all the sequences (in FASTA format) of the proteins that contain at least one of the domains of interest are saved. To reduce the rate of false positives, each sequence retrieved is compared against the whole Pfam database (Figure 1). This allows the user to determine whether the domain of interest that has been identified in each retrieved sequence actually constitutes the best domain assignment for that region of the sequence. In other words, if a Pfam domain not in the list of the domains of interest matches better than any of the domains of the list a given region of the sequence (even though one of the domains of interest did match with an E-value better than the threshold), the assignment of the protein as one of the pathway becomes dubious and the sequence is therefore segregated for a possible further inspection. For

domains that are associated to a LBP, the sequences are additionally filtered by requesting that they contain the LBP (only the amino acids falling within the domain boundaries). A tolerance of 20% is applied to the spacing between amino acids in the LBP. Rejected files are moved in separate subfolders (see also Supporting text).

A separate log file is created by this script, in addition to one which recapitulates all the results. A tutorial for the use of the RDGB tool is given in the supplementary text and is included in the RDGB download file.

Results & Discussion

The strategy

The identification of biochemical pathways using computational methods has been the focus of a great deal of interest, especially since a large amount of sequence information for a variety of different organisms has been accumulating in genomic databases. Developments in the field have included identification of missing enzymes in otherwise complete pathways, to lead experimental efforts for the discovery of new enzymes and gene functions, and the annotation of the entire metabolic network of organisms, in a systems biology perspective. The latter is generally a quite complex task, requiring the application of sophisticated bioinformatics methods by highly skilled researchers. Another specific application is the comparison of the occurrence and distribution of a biochemical process in different organisms. This endeavour, which is computationally much less demanding than the aforementioned metabolic reconstructions, has a value e.g. to determine how widespread a pathway for the acquisition of nutrients is or what pathways are shared by a group of pathogens (see for example our work on heme uptake as a source of iron for prokaryotes (20)). Although not computationally intensive, when performed on hundreds of organisms this kind of investigation generates a considerable amount of data, preventing manual inspection of all the results and therefore creating the need for a stable strategy that minimizes errors and is prone to automation.

In this work, we present a simple protocol to tackle the task mentioned above. The protocol relies on the identification of proteins on the basis of their domain content. This allows one to identify the possible homologues of all the proteins involved in a biochemical pathway of interest though a systematic scanning of the proteome (including the proteins that are encoded by both chromosomal and plasmidic DNA) of an organism. By inspecting the presence of homologues of all the proteins of the pathway (or only of some selected key ones), it is possible to readily identify which organisms encode a pathway. At the same time, hints on variations on the composition of the pathway can also be

obtained by analyzing more closely the organisms that lack only a small (with respect to the number of components in the pathway) number of proteins. Finally, the sequences identified can be inspected at the per-residue or, when the 3D structure of at least one representative of the family is available, at the atomic level, e.g. by taking advantage of homology modelling, to ascertain possible differences in the mechanisms of substrate or intermolecular recognition.

The present procedure thus starts with defining the list of domains that characterize the biochemical process of interest. We propose to use the Pfam library of domains as the annotation of the domains in the library is normally sufficiently detailed to allow users to evaluate the actual relevance of a domain to the biochemistry under investigation. The domains can practically be identified by scanning the sequence of one (or more) representative of each protein in the pathway against the full Pfam database with a reasonable E-value threshold (in the range 10^{-3}-10^{-5}), using the service at the Pfam web site (http://pfam.sanger.ac.uk/search). When a protein is binding a ligand/cofactor (e.g. organic ligands, metal ions, metal-containing cofactors such as heme) and the 3D structure of the bound form is available, it is possible to take advantage of the information on the protein-ligand mode of interaction to filter the results of domain-based searches. This information is condensed in the Ligand Binding Pattern, LBP. The LBP defines the identity and the spacing of the amino acids in direct contact with or bound to the ligand; LBP's can be represented in the form AX_nBX_mC…, where A, B, C, … are the metal-binding amino acids, and n, m, … the number of amino acids in between two subsequent ligands.

After the list of domains (and associated LBP's) is compiled, it can be used to scan any complete proteome to identify the proteins that contain one (or more) of them. For domains associated with an LBP, the latter can be used to filter the results and improve the precision of the method. The filter is applied by imposing that the predicted protein contains all the ligands of the LBP with a spacing in sequence that it is maintained within ± 20% (or ± 1 amino acid for short spacing). This procedure leads to a significant reduction of the number of false positives (proteins wrongly predicted to be homologues of the ones of interest), as extensively documented for metal-binding proteins (25). A further refinement to improve the precision is to check whether the region of the sequence corresponding to the domain of interest in each retrieved protein matches to another domain, not in the list, with a better (i.e. lower) E-value; if yes, the protein is removed from the list of positives. This can happen because in the initial scan of the proteomes we only search for the domains of the list, in order to save time. By scanning the retrieved proteins, which typically are a very small fraction of each proteome, against the entire Pfam database, we can identify in each sequence other domains, not in the

list, that overlap with the region spanned by the domain of interest (domains that correspond to protein regions not in overlap do not pose a problem and actually define multi-domain proteins). Because both domains would match the sequence at an E-value lower than the threshold, it is useful not to discard the sequence immediately but rather to further inspect it. Useful guides are the extent of overlap between the two domains and the ratio of the corresponding E-values.

An example application to the degradation of aromatic hydrocarbons

To demonstrate an application of the methodology described here (Figure 1), we characterized the pathways for the aerobic degradation of aromatic hydrocarbons that start with cis-dihydroxylation of the substrate in 1136 prokaryotic proteomes available from the NCBI database. As a further difficulty of analysis, the enzymes in these pathways can be encoded both by chromosomal and plasmidic genes. This does not pose a problem with RDGB, as our tool analyses all the proteins of the organism regardless of their genetic origin. The substrates of interest in the present group of processes range from toluene to biphenyl and naphthalene, including various other compounds in which the aromatic ring(s) are differently substituted. Figure 2 presents a general overview of these pathways, as it can be derived from the information in the KEGG database (17). Note that many variants to these pathways can exist in nature, e.g. regarding the regiochemistry of some reactions or the involvement of mono-oxygenation reactions (30). It can be seen that only the upper part of the pathways is common to all substrates, i.e. the initial dihydroxylation, followed by dehydrogenation and then by the opening of one aromatic ring through the cleavage of a carbon-carbon double bond (Figure 2). Among these common enzymes (and their ancillary proteins, such as electron-transporting ferredoxins), there are only two domains that are specific to the pathways of interest, namely the Ring_hydroxyl_A and Ring_hydroxyl_B domains. These are contained respectively in the α- and β-subunits of the dihydroxylating dioxygenase performing the first step in the process of Figure 2 (Table 1). The other proteins instead contain relatively common functional domains that are present also in enzymes involved in other metabolic processes. Here we are dealing with three of the four known families of hydroxylating dioxygenases, namely with the so-called toluene/biphenyl, naphthalene and benzoate dioxygenases (32) (the substrate specificity of these enzymes is much broader than the names suggest), which are heteromultimers consisting of α- and β-subunits. A fourth class of dioxygenases exists, namely phthalate dioxygenases (which include also carbazole and 2-oxo-1,2-dihydroquinoline among their substrates), which are instead homomultimers with an α_n quaternary structure. Phthalate

dioxygenases differ significantly in sequence from the members of the other three families (32;41) and are actually associated to a different Pfam domain. Further along the degradation pathways of Figure 2, another pathway-specific domain is muconolactone delta-isomerase, which plays a role within the degradation of catechol. The latter can be generated during the degradation of either benzoate or naphthalene, or can be present itself in the environment. With these three specific domains (Table 1), we retrieved a total of 1099 proteins (Table 2).

The proteome of an organism able to aerobically degrade aromatic hydrocarbons must encode at least one homologue for all the proteins in the toluene pathway of Figure 2, which is the simplest of those analyzed here. In particular, because all proteins but those contained in the first enzyme of the pathway are common to various metabolic processes, it is the presence of both Ring_hydroxyl_A- and Ring_hydroxyl_B-containing proteins that can be used as the indicator of the ability to aerobically degrade toluene. With this simple rule, 919 organisms were found to be unable to aerobically degrade aromatic hydrocarbons, whereas 178 organisms were found to possess the right enzymatic portfolio. In 14 out of 53 cases where either a Ring_hydroxyl_A- or a Ring_hydroxyl_B-containing protein was missing, we found out that a suitable protein was detected with an E-value just above the chosen threshold, so these organisms were included in the list of putative degraders (shown as yellow cells in Supplementary Table 1). Instead, two organisms missed one or more unspecific Pfam domain, so we assigned them as unable to perform the process. The organisms where we could identify only one out of the two subunits of the initial ring-hydroxylating dioxygenase (Figure 2) were marked with "?" in Supplementary Table 1.

The detection of all other domains involved in the pathway, regardless of their specificity, can be taken as an useful countercheck that there are no major problems with the above assumption. To this end, we checked as an example the toluene pathway of Figure 2, including the additional relevant domains (Table 2). Overall, we thus retrieved a total of 65196 proteins, corresponding to 65762 hits to the selected Pfam domains (Table 1). All the hits are reported in the Supplementary material (Supplementary Table 2). In all cases but one, the proteome of an organism encoding Ring_hydroxyl_A- and Ring_hydroxyl_B-containing proteins contained also all other domains (Supplementary Table 1).

Finally, the detection of the MIase domain, which uniquely identifies muconolactone delta-isomerase (EC 5.3.3.4) in the benzoate pathway (Figure 2), permits the identification of organisms that can potentially degrade also biphenyls, benzoate and naphthalene (all of them or a combination thereof). This domain has been detected in 140 organisms. Of these, 23 corresponded to organisms that

were assigned as unable to degrade hydrocarbons and 5 to organisms marked with "?" in Supplementary Table 1. The corresponding genes are in the neighbourhood of proteins annotated as hypothetical, suggesting the presence of uncharacterised mechanisms for the degradation of the substrates of interest here or, alternatively, that the MIase domain can play other uncharacterised roles.

An example application to assembly of the Cu_A cofactor

Cytochrome *c* oxidases use the Cu_A cofactor as the entry point of the electron that is delivered by cytochrome *c* into the enzyme. Cu_A is a dinuclear copper site contained in subunit II of the enzyme (Cox2) whose correct assembly is crucial for enzyme function. It has been shown by NMR that in *Thermus thermophilus* the assembly process is mediated by the soluble metallochaperone PCu_AC and the Sco1 thiol-disulfide reductase, which maintains the Cys residues in the Cu_A binding site of Cox2 in the reduced state (35)(Figure 3). We used this relatively small ensemble of proteins for a further demonstration of an application of RDGB. Also nitrous oxide reductases (NosZ) contain a Cu_A-binding domain that is homologus to that of Cox2; the assembly of the Cu_A cofactor in NosZ has not been studied in detail, but it is likely that it involves a mechanism similar to Cox2. The data are reported in Supplementary Table 3.

We identified which of the same prokaryotic organisms of the previous section contained enzymes with a soluble Cu_A-binding domain. These were 548, corresponding to 48.3% of the ensemble investigated. The occurrence of PCu_AC and Sco1 homologous was less frequent, corresponding respectively to 32.3% and 40.2% of the organisms analysed. It is relevant to address the co-occurrence of these proteins. 283 organisms contained all three proteins, corresponding to 24.9% of the dataset. In other words, one quarter of the prokaryotic organisms investigated encoded in their proteomes a Cu_A-binding domain, most likely in Cox2 as this is much more widespread than NosZ (23), and the two accessory proteins that have been demonstrated to be active in the assembly of the cofactor in *T. thermophilus*. This corresponds to 51.5% of the organisms that encode at least one Cu_A-binding domain. In 57 cases (5.0% of the organisms) a Cu_A-containing domain could not be detected but Sco1 (1.0%) or PCu_AC (0.2%) or both (3.8%) were contained in the proteome. The occurrence of Sco1 in the absence of any Cu_A-containing enzyme had been noted before and was proposed to be justified by its possible activity as a thiol-disulfide oxidoreductase (23). However, the relatively common occurrence of the pair Sco1 and PCu_AC in the absence of any Cu_A-containing enzyme may suggest that the mechanism of formation of the Cu_A cofactor, or a close variant of it, may be relevant also for the assembly of other cuproenzymes. Finally, it is worth noting that 108 organisms (9.5%) encode a Cu_A-

containing enzyme while lacking both Sco1 and PCu_AC. Thus, some yet uncharacterised assembly mechanisms may be operative in organisms such as Mycobacteria (and various other Actinobacteria), δ-proteobacteria and Cyanobacteria.

Conclusions

Following the strategy of Figure 1, the computational tool described here allows users to characterize known metabolic pathways in a wide range of organisms, exploiting three main databases: Pfam, PDB and NCBI. The functional (Pfam) domains are taken as characterizing elements of the proteins of interest and can be obtained also from the analysis of 3D structures available from the PDB. In addition, by investigating the PDB for the presence of ligands, it if possible to obtain one or more ligand binding patterns (LBP's) that are associated to the Pfam functional domain. This can be used as a filter to reduce the number of false positives. The sequences of all the proteins encoded by both the chromosomal and plasmidic DNA of an organism with fully sequenced genome are obtained from the NCBI. In addition to characterizing the distribution of a given pathway across the biological world, the data output by RDGB can also be used to identify possible variants of known pathways, such as the occurrence of alternative steps or, at the atomic level, of different modes of intermolecular interaction with the substrate.

The overall approach is demonstrated through two examples. In one, we analyzed the processes for aerobic degradation of mono-and poli-cyclic aromatic hydrocarbons, which include toluene, naphthalene and biphenyls. Out of 1136 organisms analysed, only 178 were able to degrade at least one of the above compounds. 112 could potentially degrade all of them. 39 organisms missed only one requested protein/subunit along the pathway, preventing us from assigning them as degraders or non-degraders. In the second example, we investigated the distribution of some accessory proteins that are involved in the assembly of the Cu_A cofactor. The data indicate that one or more Cu_A-binding domains can be detected in nearly half of the organisms analyzed. Among the organisms encoding Cu_A-binding domains, 51.5% contain both Sco1 and PCu_AC, demonstrating that this mechanism of Cu_A biogenesis is quite widespread in prokaryotes.

Supplementary material

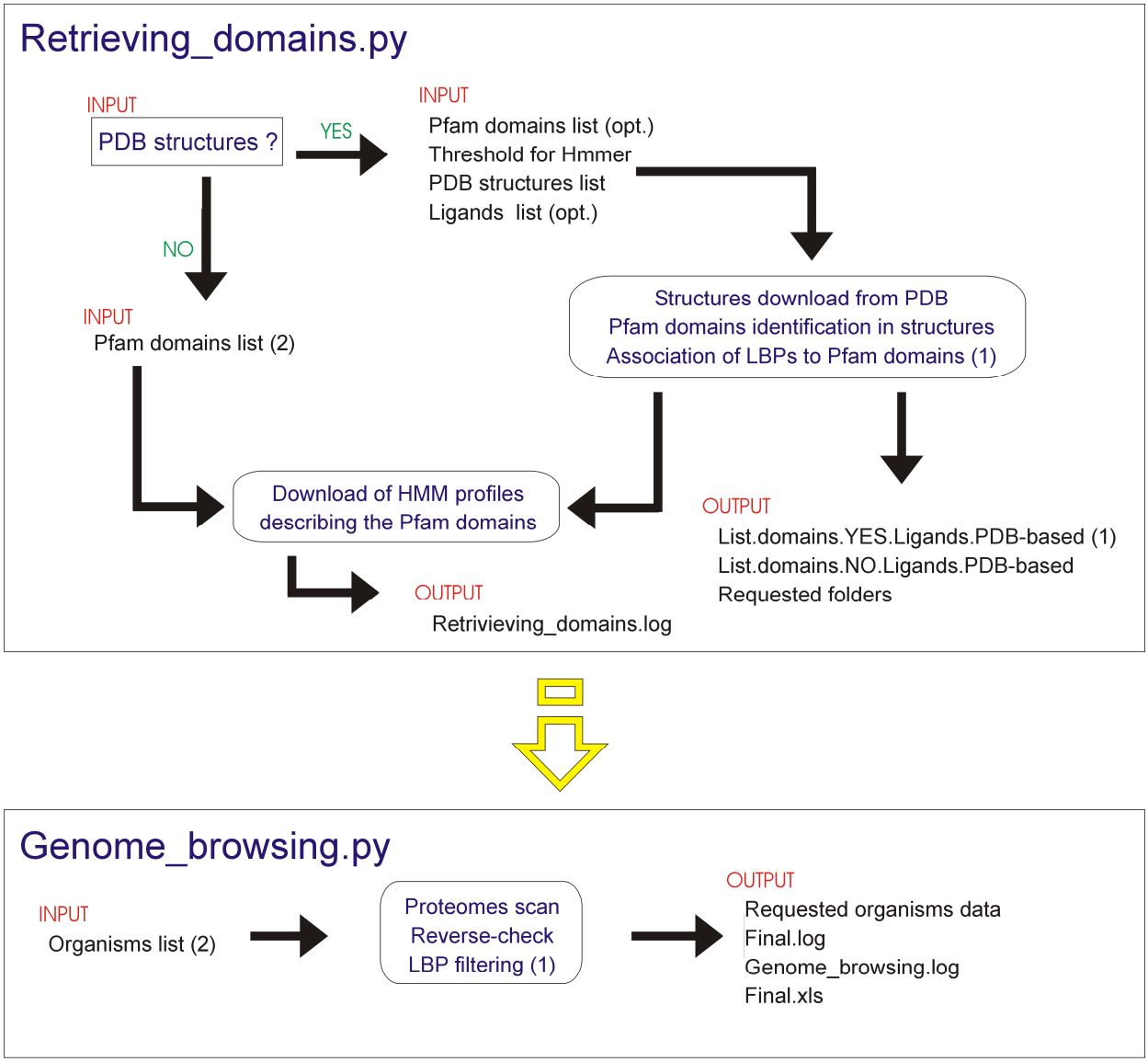
Table S1. Table summarizing the results obtained by investigating the per-organism distribution of the proteins involved in the biodegradation of aromatic hydrocarbons. This table has been output by RDGB and edited for colours, titles and the right-most column.

Table S2. Output log file produced by RDGB for all the proteins involved in the biodegradation of aromatic hydrocarbons in all the investigated organisms. This file contains the list of all the proteins retrieved.

Table S3. Table summarizing the results obtained by investigating the per-organism distribution of the proteins involved in the assembly of the Cu_A cofactor.

Supplementary text: RDGB tutorial

Figure 1 – Flow chart describing the RDGB tool. (opt.)= optional; (1)= if Ligand list is submitted; (2)= case sensitive



(1) = if Ligands list is submitted
(2) = case sensitive

Figure 2 Overview of the pathways for the degradation of simple aromatic hydrocarbons, such as toluene, biphenyl, benzoate, naphthalene. Note that naphthalene and biphenyl give rise to the formation of catechol and benzoate which are further degraded in the benzoate pathway. Alternative pathways may exist, e.g. involving mono-oxygenation reactions. The regiochemistry of the ring-opening reaction generally depends on both the organism under consideration and the nature of the ring substituents. Sub-products such as acetate are not shown. This figure has been adapted from the KEGG pathway database. The benzoate degradation-specific enzyme muconolactone delta-isomerase is highlighted by a dotted ellipse.

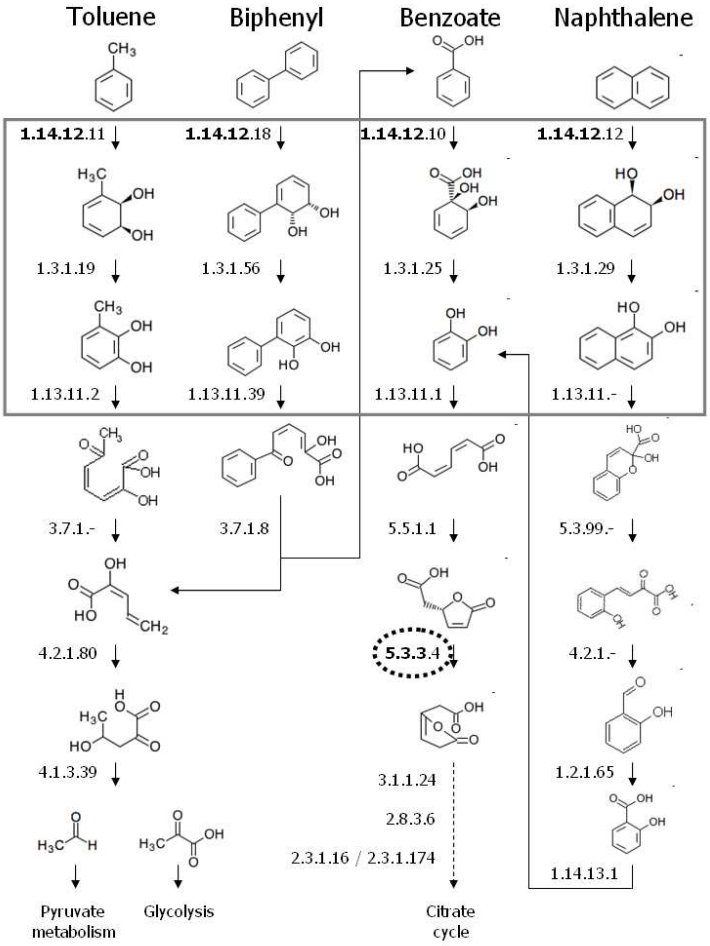


Figure 3. The mechanism of assembly of the CuA cofactor, as described in (35) (Ox = oxidized; Red = reduced). The sulphur atoms of redox-active cysteines are shown as circles. Metal ligands are shown as sticks. Metal ions are shown as spheres.

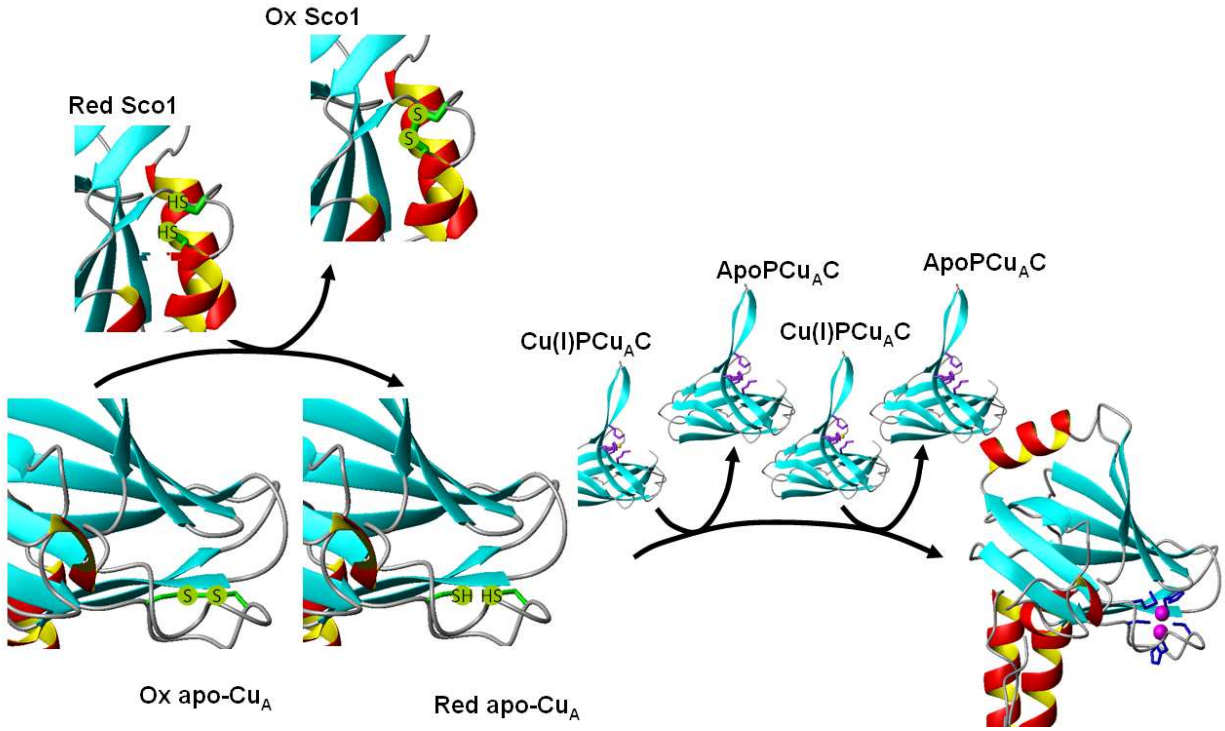


Table 1 –Proteins involved in the aerobic biodegradation of aromatic hydrocarbons (Figure 2). The proteins in the top part of the table contain domains specific to the pathways of interest, whereas the proteins in the bottom part (separated by a double line) are not specific. For each protein, we report the EC number (only three levels are given, as the fourth depends on the identity of the substrate), the composition in Pfam domains and the corresponding PDB structures (only if a ligand is present and thus a LBP is available).

Protein	**EC number**	**Pfam**	**PDB**	**Ligand**
Dioxygenase	1.14.12	**Ring_hydoxyl_A** Rieske	1ULI; 2B1X 2BMO; 2GBW; 2HMJ 3EN1;	Fe FeS
	1.14.12	**Ring_hydroxyl_B**	-	-
Isomerase	5.3.3	**MIase**	-	-
Dehydrogenase	1.3.1	Adh_short	-	-
Dioxygenase	1.13.11	Glyoxalase	1EIQ; 1HAN; 1KMY; 2EHZ; 2EI2 2ZI8; 3HPV	Fe
Hydrolase	3.7.1	Abhydrolase_1	-	-
Dehydratase	4.2.1	FAA_hydrolase	-	-
Aldolase	4.1.3	HMGL_like	-	-

Table 2 – Number of hits and number of organisms where each Pfam domain of Table 1 has been identified.

Pfam domain	Protein	N° of hits	N° of organisms
Ring_hydroxyl_A	Dioxygenase	482	200
Ring_hydroxyl_B		461	185
Rieske		3217	726
Adh_short	Dehydrogenase	32539	1108
Glyoxalase	Dioxygenase	5015	887
Abhydrolase_1	Hydrolase	17847	1098
FAA_hydrolase	Dehydratase	2854	788
HMGL-like	Aldolase	3191	988
MIase	Isomerase	156	140

Table 3 – Number of hits and number of organisms where domains related to the mechanism of assembly of the Cu_A cofactor (Figure 3) were deteced.

Pfam domain	Protein	N° of hits	N° of organisms
COX2	Cox2	899	548
SCO1-SenC	Sco1	790	456
DUF461	PCu_AC	454	366

Reference List

1. Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, C. A. (1999) *Nature* **402,** 86-90

2. Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999) *Science* **285,** 751-753

3. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D., and Maltsev, N. (1999) *Proc.Natl.Acad.Sci.USA* **96,** 2896-2901

4. Wolf, Y. I., Rogozin, I. B., Kondrashov, A. S., and Koonin, E. V. (2001) *Genome Res.* **11,** 356-372

5. Snel, B., Bork, P., and Huynen, M. A. (2002) *Proc.Natl.Acad.Sci.U.S.A* **99,** 5890-5895

6. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999) *Proc.Natl.Acad.Sci.U.S.A* **96,** 4285-4288

7. Pagel, P., Wong, P., and Frishman, D. (2004) *J.Mol.Biol.* **344,** 1331-1346

8. Snel, B., Lehmann, G., Bork, P., and Huynen, M. A. (2000) *Nucl.Acids Res.* **28,** 3442-3444

9. von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003) *Nucl.Acids Res.* **31,** 258-261

10. Lee, I., Date, S. V., Adai, A. T., and Marcotte, E. M. (2004) *Science* **306,** 1555-1558

11. Date, S. V. and Marcotte, E. M. (2003) *Nat.Biotechnol.* **21,** 1055-1062

12. Osterman, A. and Overbeek, R. (2003) *Curr.Opin.Chem.Biol.* **7,** 238-251

13. Cordwell, S. J. (1999) *Arch.Microbiol.* **172,** 269-279

14. Gianchandani, E. P., Brautigan, D. L., and Papin, J. A. (2006) *Trends Biochem.Sci.* **31,** 284-291

15. Price, N. D., Reed, J. L., and Palsson, B. O. (2004) *Nat.Rev.Microbiol.* **2,** 886-897

16. Tyson, J. J., Chen, K., and Novak, B. (2001) *Nat.Rev.Mol.Cell Biol.* **2,** 908-916

17. Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008) *Nucleic Acids Res.* **36,** D480-D484

18. Kanehisa, M. and Goto, S. (2000) *Nucleic Acids Res.* **28,** 27-30

19. Karp, P. D., Ouzounis, C. A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahren, D., Tsoka, S., Darzentas, N., Kunin, V., and Lopez-Bigas, N. (2005) *Nucleic Acids Res.* **33,** 6083-6089

20. Cavallaro, G., Decaria, L., and Rosato, A. (2008) *J.Proteome.Res.* **7,** 4946-4954

21. Bertini, I., Cavallaro, G., and Rosato, A. (2007) *J.Inorg.Biochem.* **101,** 1798-1811

22. Sharma, S., Cavallaro, G., and Rosato, A. (2010) *J.Biol.Inorg.Chem.* **15,** 559-571

23. Banci, L., Bertini, I., Cavallaro, G., and Rosato, A. (2007) *J.Proteome Res.* **6,** 1568-1579

24. Bertini, I., Cavallaro, G., and Rosato, A. (2006) *Chem.Rev.* **106,** 90-115

25. Andreini, C., Bertini, I., and Rosato, A. (2009) *Acc.Chem.Res.* **42,** 1471-1479

26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J.Mol.Biol.* **215,** 403-410

27. Hulsen, T., Huynen, M. A., de Vlieg, J., and Groenen, P. M. (2006) *Genome Biol.* **7,** R31

28. Claudel-Renard, C., Chevalet, C., Faraut, T., and Kahn, D. (2003) *Nucleic Acids Res.* **31,** 6633-6639

29. Ramos, J. L., Diaz, E., Dowling, D., de Lorenzo, V., Molin, S., O'Gara, F., Ramos, C., and Timmis, K. N. (1994) *Biotechnology (N.Y.).* **12,** 1349-1356

30. Cao, B., Nagarajan, K., and Loh, K. C. (2009) *Appl.Microbiol.Biotechnol.* **85,** 207-228

31. Yakimov, M. M., Timmis, K. N., and Golyshin, P. N. (2007) *Curr Opin Biotechnol* **18,** 257-266

32. Gibson, D. T. and Parales, R. E. (2000) *Curr.Opin.Biotechnol.* **11,** 236-243

33. Vaillancourt, F. H., Bolin, J. T., and Eltis, L. D. (2006) *Crit Rev.Biochem.Mol.Biol.* **41,** 241-267

34. Carr, H. S. and Winge, D. R. (2003) *Acc.Chem.Res.* **36,** 309-316

35. Abriata, L. A., Banci, L., Bertini, I., Ciofi-Baffoni, S., Gkazonis, P., Spyroulias, G. A., Vila, A. J., and Wang, S. (2008) *Nat.Chem.Biol.* **4,** 599-601

36. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C., and Eddy, S. R. (2004) *Nucleic Acids Res.* **32 Database issue,** D138-D141

37. Sonnhammer, E. L., Eddy, S. R., and Durbin, R. (1997) *Proteins* **28,** 405-420

38. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* **28,** 235-242

39. Eddy, S. R. (1998) *Bioinformatics* **14,** 755-763

40. Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005) *Nucleic Acids Res.* **33,** D501-D504

41. Nam, J. W., Nojiri, H., Yoshida, T., Habe, H., Yamane, H., and Omori, T. (2001) *Biosci.Biotechnol.Biochem.* **65,** 254-263

A simple protocol for the comparative analysis of the structure and occurrence of biochemical pathways across superkingdoms

Claudia Andreini[1,2], Ivano Bertini[1,2,*], Gabriele Cavallaro[1], Leonardo Decaria[1], Antonio Rosato[1,2]

[1]Magnetic Resonance Center (CERM) – University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

[2]Department of Chemistry – University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Supplementary text: RDGB tutorial

Installation

To install the RDGB tool, copy the RDGB.tar.gz file in your work folder and type the following commands:

> uncompress RDGB.zip
> python 0Tools/START.py

RDGB will automatically download and install HMMER. This will requires a few minutes.

Inputs (RDGB accepts both lowercase and uppercase letters**):**

- Y or N questions: Other inputs will be interpreted as "N".
- Databases download: Wrong inputs will be interpreted as Yes.
- Chosen threshold: both floating point (e.g. 0.001) and scientific E notation formats (e.g. 1.0e-3) are accepted
- Input files: type the name of the requested files.
- A or O case: if a new run is performed in a folder that already contains the results of a previous calculation, the tool asks whether to append (A) or overwrite (O) the data.

Outputs:

All output files are plain text files except otherwise specified.

- USER.INPUT_FILES: this file contains the list of the input files submitted by the user.
- List.domains.YES.ligands.PDB-based: this file is created when a list of ligands has been input. It reports the analyzed PDB structures, the associated Pfam domains, the associated LBPs and the bound ligands.
- List.domains.NO.ligands.PDB-based: this file is created when no list of ligands has been input. It reports the analyzed PDB structures and the associated Pfam domains.
- Retrieving_domains.log & Genome_browsing.log: each file reports the steps performed by the programs. Note that the "No pattern" entry under the "Pattern" column in the *Genome_browsing.log* file indicates that there is no LBP associated to the domain. The date, time, any error and other information are also reported.
- 0Profiles/Retrieved_Domains.list: this file is a list of all the retrieved Pfam domains.
- Final.log: this file contains both *Retrieving_domains.log* and *Genome_browsing.log* files. Note that the "No pattern" entry under the "Pattern" column indicates that there is no LBP associated to the domain.

- Final.xls: this file, readable by every release of Microsoft Excel, shows the final results in a domain/organism matrix.

Utilities

By typing:

> python 0Tools/Retrieving_Organisms.py

the lists of all prokaryotic and eukaryotic organisms with complete genome sequences available are created (*Prokarya.list* and *Eukarya.list*, respectively).

By typing:

> python 0Tools/Cleaner.py

The data of previous runs (also if incomplete) are removed. It is possible to remove all data or only the output of *Genome_Browsing.py*.

Tutorial

The user can perform a sample run following the tutorial below, which uses two PDB structures and five selected organisms taken from an application example described in the main text.

> cp 0Tools/*.example .

> python Retrieving_Domains.py

…Downloading Pfam database… (if first run, this will take several minutes; the installation of the Pfam database requires approx. 2.0 Gb of disk space)

PDB structures to submit? Y/N: > Y

…Downloading PDB database… (if first run, this will take several minutes)

Insert list of Pfam profiles, else type NO: > DOMS.example
Insert list of PDB codes: > STRS.example
Insert list of ligands codes, else type NO: > LIGS.example
Insert Chosen.Threshold or type D for default (1.0e-5) > D
Create 0PDB folder containing .pdb files, Y/N: > Y
Create 0Fasta folder containing .fasta files, Y/N: > Y
Create 0Pfam_output folder containing .Pfam_output files, Y/N: > Y
Create 0Domains folder containing .domains files, Y/N: > Y
Create 0Patterns folder containing .pattern files, Y/N: > Y

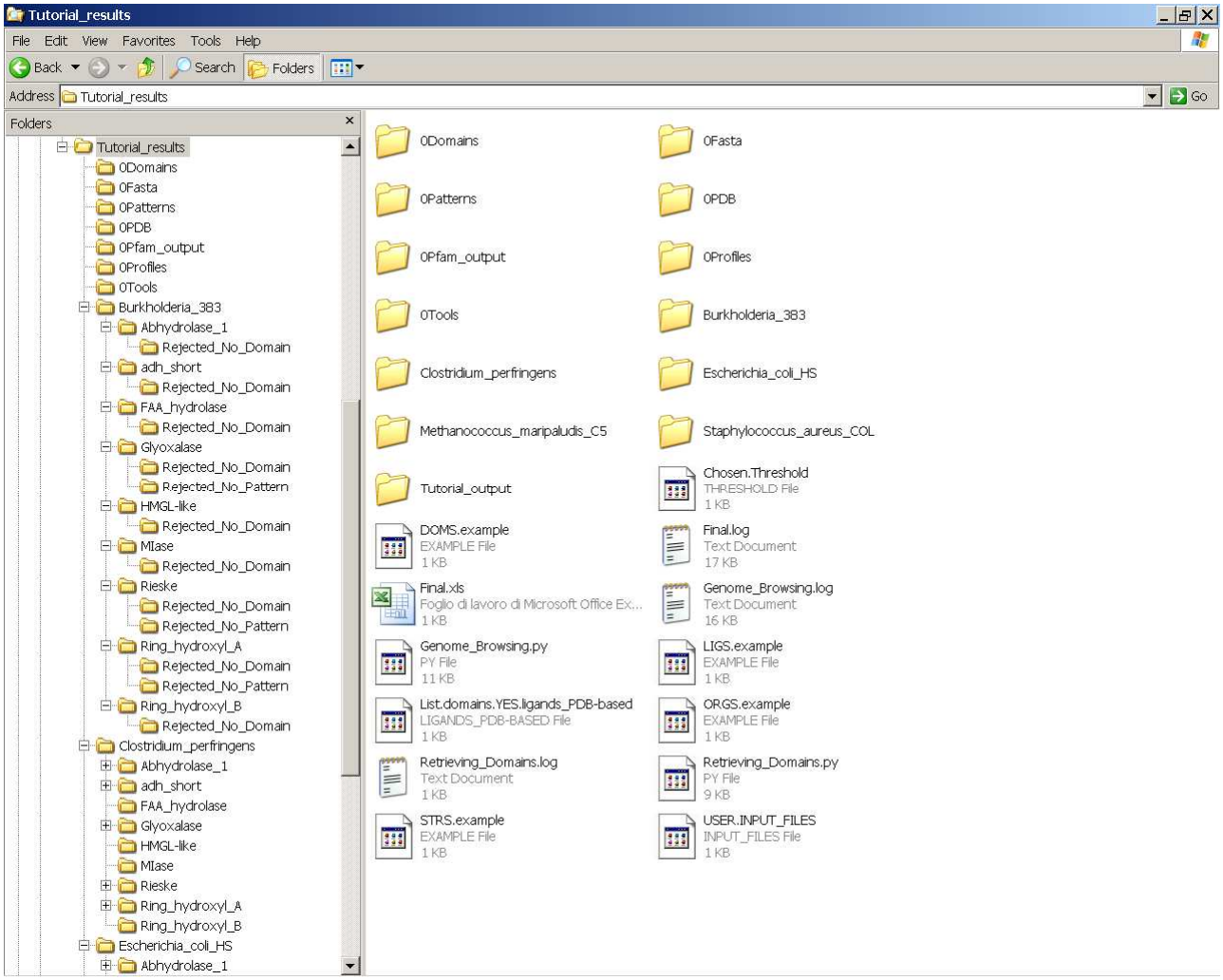
… program running…

> python Genome_Browsing.py
Pfam database already present, overwrite? Y/N: > N

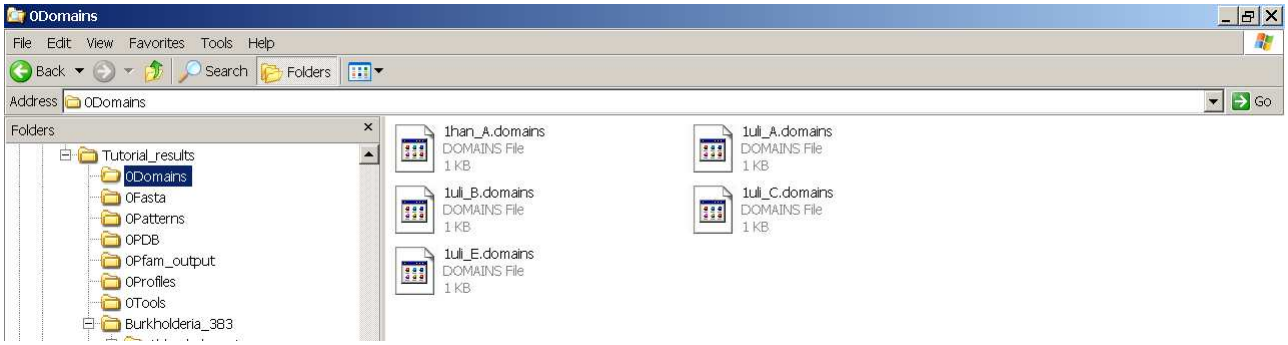
Insert list of organisms: > ORGS.example
Do you want to compress the output folders (tar.gz)? Y/N: > Y
Using previous Chosen.Threshold of 1.0e-5? Type Y or submit new value: > Y

… program running…

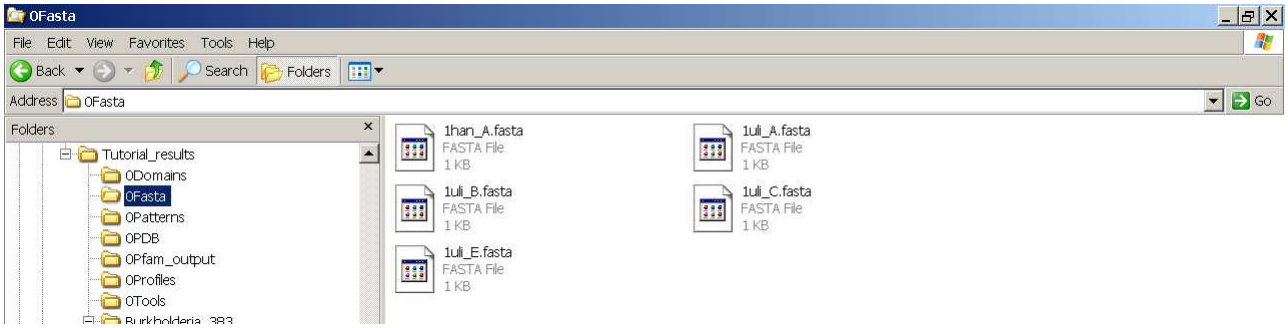
Output folders scheme:



- The folder **0Domains** contains files listing the Pfam domains contained in each protein chain of the input PDB structures



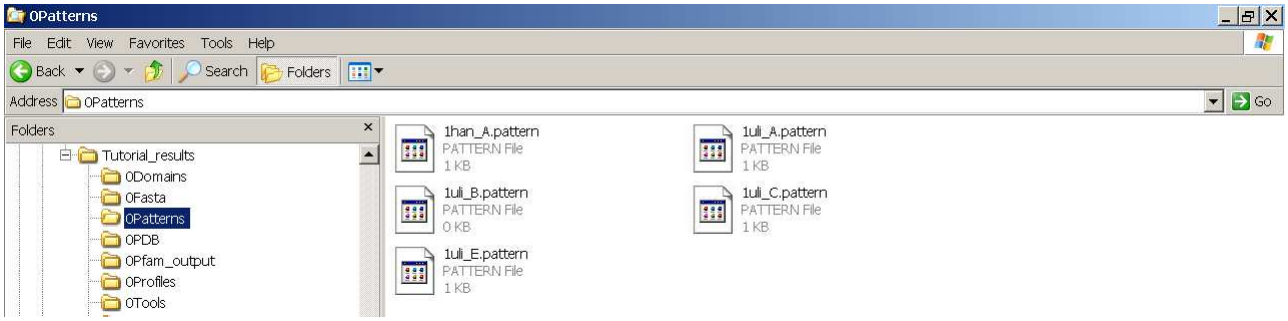
- The folder **0Fasta** contains the sequences of each protein chain of the input PDB structures



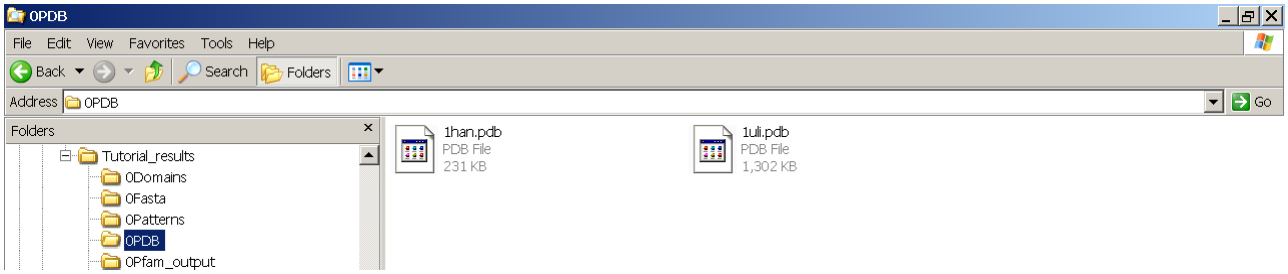
- The folder **0Pattern** contains files describing the LBPs found in each protein of the input PDB structures, with the following format:

`3.9e-30 Glyoxalase 1 142-260 3/3 HX(63)HX(49)E FE A 500`

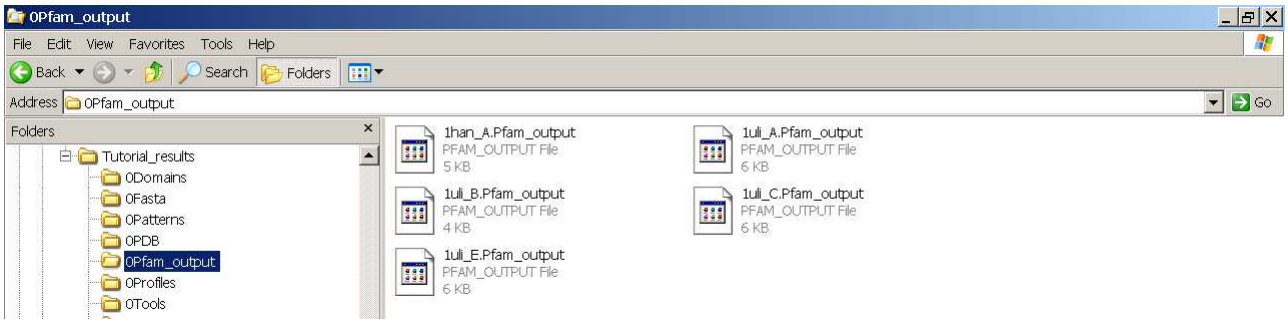
where “3.9e-30” is the E-value of the matched domain, “Glyoxalase” is the domain containing the LBP, “1” is the position of the domain in the domain list for the protein, “142-260” is the sequence region spanned by the domain, “3/3” is the fraction of amino acids of the LBP within the domain, “HX(63)HX(49)E” is the LBP, “FE” is the ligand bound, “A 500” is the ligand residue number



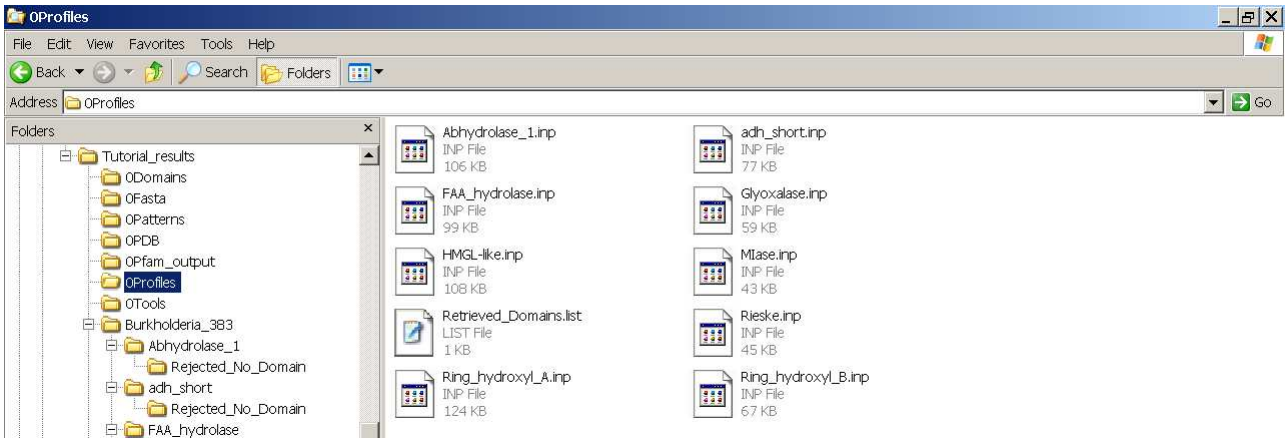
- The folder **0PDB** contains the input PDB structures



- The folder **0Pfam_output** contains files with the full results of the analysis of each protein chain of the input PDB structures against the Pfam database



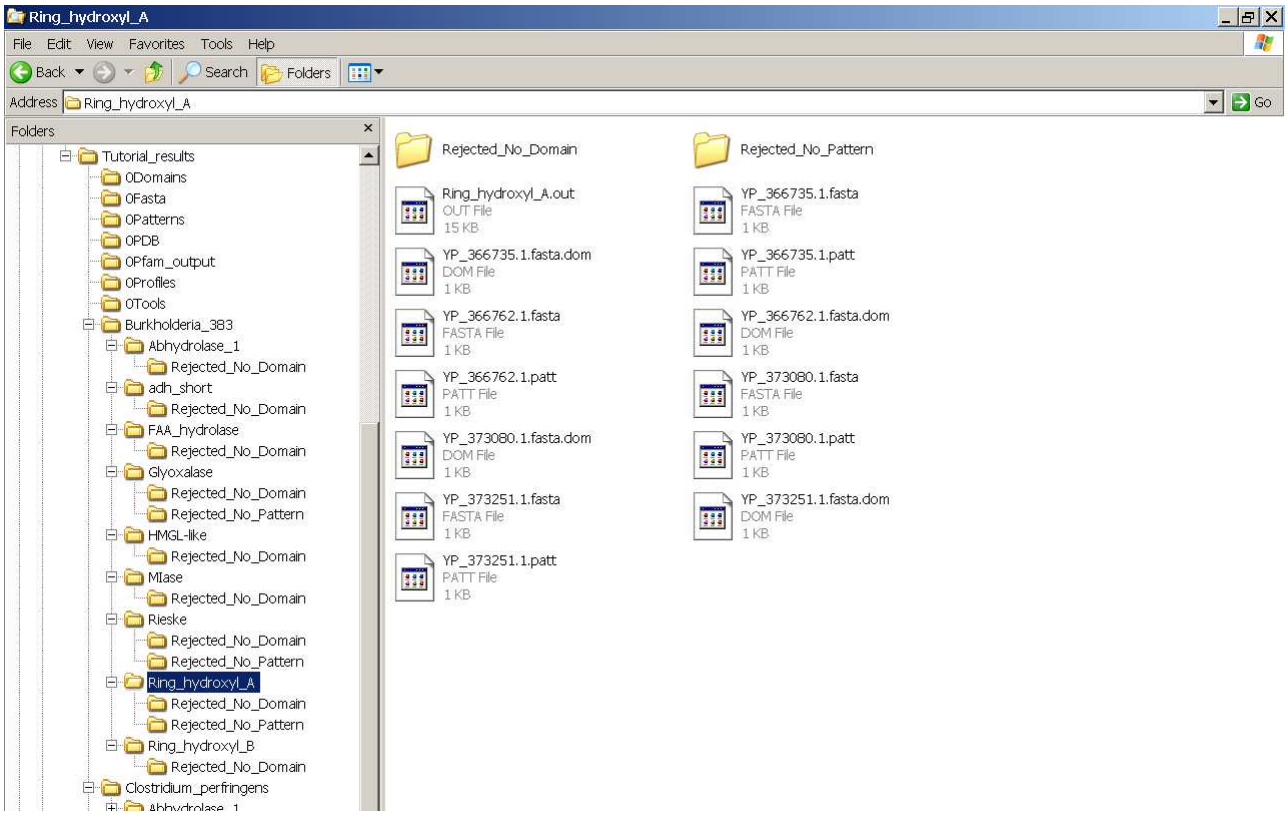
- The folder **0Profiles** contains the HMM's of all domains



A folder is created by *Genome_Browsing.py* for each analyzed organism. This folder contains the sequence of the proteome of the organism as well as one subfolder for each domain searched within the proteome.

The latter domain subfolders include all the sequences of the proteins containing that domain (*.fasta* files), the list of the corresonding LBP's (*.patt* files), and the list of all the Pfam domains (i.e. not limited to the selected domains) found in each sequence (*.dom* files). In summary, three files are generated for each protein if there is a LBP associated to the domain; otherwise, the *.patt* file is not

created. Additionally, sequences discarded because they did not pass either the check against the entire Pfam database or the filter for the presence of the LBP are respectively stored in the subfolders **Rejected_No_Domain** and **Rejected_No_Pattern**. Finally, the *.out* file contains the output of the HMMER search of the domain within the proteome.



Enzymes for aromatic hydrocarbons degradation		1.14.12	1.14.12	1.14.12	1.3.1	1.13.11	3.7.1	4.2.1	4.1.3	5.3.3	**RESULT**
Organisms - Pfam domains	***GROUP***	Ring_hydroxyl_A	Ring_hydroxyl_B	Rieske	adh_short	Glyoxalase	Abhydrolase_1	FAA_hydrolase	HMGL-like	MIase	-
Acaryochloris_marina_MBIC11017	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Acetobacter_pasteurianus_IFO_3283_01	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	NO	YES	NO	**?**
Acholeplasma_laidlawii_PG_8A	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Acidaminococcus_fermentans_DSM_20731_uid43471	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Acidimicrobium_ferrooxidans_DSM_10331	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Acidiphilium_cryptum_JF-5	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Acidithiobacillus_ferrooxidans_ATCC_23270	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Acidithiobacillus_ferrooxidans_ATCC_53993	Other_Bacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Acidobacterium_capsulatum_ATCC_51196	Acidobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Acidothermus_cellulolyticus_11B	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Acidovorax_avenae_citrulli_AAC00-1	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Acidovorax_JS42	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Aciduliprofundum_boonei_T469_uid43333	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Acinetobacter_baumannii_AB0057	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Acinetobacter_baumannii_AB307_0294	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Acinetobacter_baumannii_ACICU	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Acinetobacter_baumannii_ATCC_17978	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Acinetobacter_baumannii_AYE	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Acinetobacter_baumannii_SDF	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Acinetobacter_DR1_uid50119	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Acinetobacter_sp_ADP1	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Actinobacillus_pleuropneumoniae_L20	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Actinobacillus_pleuropneumoniae_serovar_3_JL03	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Actinobacillus_pleuropneumoniae_serovar_7_AP76	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Actinobacillus_succinogenes_130Z	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Actinosynnema_mirum_DSM_43827	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Aeromonas_hydrophila_ATCC_7966	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Aeromonas_salmonicida_A449	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Aeropyrum_pernix	Crenarchaeota	NO	NO	YES	YES	YES	YES	YES	NO	NO	**NO**
Aggregatibacter_actinomycetemcomitans_D11S_1	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Aggregatibacter_aphrophilus_NJ8700	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Agrobacterium_radiobacter_K84	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Agrobacterium_tumefaciens_C58_Cereon	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Agrobacterium_vitis_S4	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Akkermansia_muciniphila_ATCC_BAA_835	Other_Bacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Alcanivorax_borkumensis_SK2	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Alicyclobacillus_acidocaldarius_DSM_446	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Aliivibrio_salmonicida_LFI1238	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Alkalilimnicola_ehrlichei_MLHE-1	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Alkaliphilus_metalliredigens_QYMF	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Alkaliphilus_oremlandii_OhILAs	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Allochromatium_vinosum_DSM_180_uid46083	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Alteromonas_macleodii__Deep_ecotype_	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Aminobacterium_colombiense_DSM_12261_uid47083	Other_Bacteria	NO	NO	NO	YES	YES	NO	YES	YES	NO	**NO**
Ammonifex_degensii_KC4	Firmicutes	NO	NO	NO	YES	NO	NO	NO	YES	NO	**NO**
Amycolatopsis_mediterranei_U32_uid50565	Other_Bacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Anabaena_variabilis_ATCC_29413	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Anaerocellum_thermophilum_DSM_6725	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Anaerococcus_prevotii_DSM_20548	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Anaeromyxobacter_dehalogenans_2CP_1	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Anaeromyxobacter_dehalogenans_2CP-C	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Anaeromyxobacter_Fw109-5	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Anaeromyxobacter_K	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Anaplasma_centrale_Israel	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Anaplasma_marginale_Florida	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Anaplasma_marginale_St_Maries	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Anaplasma_phagocytophilum_HZ	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Anoxybacillus_flavithermus_WK1	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Aquifex_aeolicus	Aquificae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Arcanobacterium_haemolyticum_DSM_20595_uid49489	Actinobacteria	NO	NO	YES	YES	NO	YES	YES	NO	NO	**NO**
Archaeoglobus_fulgidus	Euryarchaeota	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Archaeoglobus_profundus_DSM_5631_uid43493	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Arcobacter_butzleri_RM4018	Epsilonproteobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Arcobacter_nitrofigilis_DSM_7299_uid49001	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Aromatoleum_aromaticum_EbN1	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Arthrobacter_aurescens_TC1	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Arthrobacter_chlorophenolicus_A6	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Arthrobacter_FB24	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Aster_yellows_witches-broom_phytoplasma_AYWB	Firmicutes	NO	NO	NO	NO	NO	NO	NO	NO	NO	**NO**
Atopobium_parvulum_DSM_20469	Actinobacteria	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Azoarcus_BH72	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Azorhizobium_caulinodans_ORS_571	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Azospirillum_B510_uid46085	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Azotobacter_vinelandii_DJ	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Bacillus_amyloliquefaciens_FZB42	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_anthracis_A0248	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_anthracis_Ames	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_anthracis_Ames_0581	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_anthracis_CDC_684	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_anthracis_CI_uid50615	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_anthracis_str_Sterne	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_03BB102	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_AH187	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_AH820	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_ATCC_10987	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_ATCC14579	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_B4264	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_cytotoxis_NVH_391-98	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_G9842	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_Q1	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_cereus_ZK	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_clausii_KSM-K16	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_halodurans	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_licheniformis_ATCC_14580	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_licheniformis_DSM_13	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_megaterium_DSM319_uid48371	Firmicutes	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Bacillus_megaterium_QM_B1551_uid15862	Firmicutes	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Bacillus_pseudofirmus_OF4_uid45847	Firmicutes	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Bacillus_pumilus_SAFR-032	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_selenitireducens_MLS10_uid49513	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_subtilis	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_thuringiensis_Al_Hakam	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Bacillus_thuringiensis_BMB171_uid49135	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_thuringiensis_konkukian	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacillus_tusciae_DSM_2912_uid48361	Firmicutes	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Bacillus_weihenstephanensis_KBAB4	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bacteroides_fragilis_NCTC_9434	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Bacteroides_fragilis_YCH46	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Bacteroides_thetaiotaomicron_VPI-5482	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Bacteroides_vulgatus_ATCC_8482	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Bartonella_bacilliformis_KC583	Alphaproteobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Bartonella_grahamii_as4aup	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	YES	NO	NO	**NO**
Bartonella_henselae_Houston-1	Alphaproteobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Bartonella_quintana_Toulouse	Alphaproteobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Bartonella_tribocorum_CIP_105476	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Baumannia_cicadellinicola_Homalodisca_coagulata	Gammaproteobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Bdellovibrio_bacteriovorus	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Beijerinckia_indica_ATCC_9039	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Beutenbergia_cavernae_DSM_12333	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bifidobacterium_adolescentis_ATCC_15703	Actinobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Bifidobacterium_animalis_lactis_AD011	Actinobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Bifidobacterium_animalis_lactis_Bl_04	Actinobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Bifidobacterium_animalis_lactis_DSM_10140	Actinobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Bifidobacterium_dentium_Bd1	Actinobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Bifidobacterium_longum	Actinobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Bifidobacterium_longum_DJO10A	Actinobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Bifidobacterium_longum_infantis_ATCC_15697	Actinobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Bifidobacterium_longum_JDM301_uid49131	Actinobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Blattabacterium__Blattella_germanica__Bge	Bacteroidetes/Chlorobi	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Blattabacterium__Periplaneta_americana__BPLAN	Bacteroidetes/Chlorobi	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Bordetella_avium_197N	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Bordetella_bronchiseptica	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Bordetella_parapertussis	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Bordetella_pertussis	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Bordetella_petrii	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Borrelia_afzelii_PKo	Spirochaetes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Borrelia_burgdorferi	Spirochaetes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Borrelia_burgdorferi_ZS7	Spirochaetes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Borrelia_duttonii_Ly	Spirochaetes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Borrelia_garinii_PBi	Spirochaetes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Borrelia_hermsii_DAH	Spirochaetes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Borrelia_recurrentis_A1	Spirochaetes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Borrelia_turicatae_91E135	Spirochaetes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Brachybacterium_faecium_DSM_4810	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brachyspira_hyodysenteriae_WA1	Spirochaetes	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Brachyspira_murdochii_DSM_12563_uid48819	Spirochaetes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brachyspira_pilosicoli_95_1000_uid50609	Spirochaetes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Bradyrhizobium_BTAi1	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Bradyrhizobium_japonicum	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Bradyrhizobium_ORS278	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brevibacillus_brevis_NBRC_100599	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_abortus_bv__1_9_941	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_abortus_S19	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_canis_ATCC_23365	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Brucella_melitensis	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_melitensis_ATCC_23457	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_melitensis_biovar_Abortus	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_microti_CCM_4915	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_ovis	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_suis_1330	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Brucella_suis_ATCC_23445	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Buchnera_aphidicola	Gammaproteobacteria	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Buchnera_aphidicola_5A__Acyrthosiphon_pisum_	Gammaproteobacteria	NO	NO	NO	YES	NO	NO	NO	NO	NO	**NO**
Buchnera_aphidicola_Cc_Cinara_cedri	Gammaproteobacteria	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Buchnera_aphidicola_Sg	Gammaproteobacteria	NO	NO	NO	YES	NO	NO	NO	NO	NO	**NO**
Buchnera_aphidicola_Tuc7__Acyrthosiphon_pisum_	Gammaproteobacteria	NO	NO	NO	YES	NO	NO	NO	NO	NO	**NO**
Buchnera_sp	Gammaproteobacteria	NO	NO	NO	YES	NO	NO	NO	YES	NO	**NO**
Burkholderia_383	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_ambifaria_AMMD	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_ambifaria_MC40_6	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_CCGE1002_uid42523	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_cenocepacia_AU_1054	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_cenocepacia_HI2424	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_cenocepacia_J2315	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_cenocepacia_MC0_3	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_glumae_BGR1	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_mallei_ATCC_23344	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_mallei_NCTC_10229	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_mallei_NCTC_10247	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_mallei_SAVP1	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_multivorans_ATCC_17616_JGI	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_multivorans_ATCC_17616_Tohoku	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_phymatum_STM815	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_phytofirmans_PsJN	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_pseudomallei_1106a	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_pseudomallei_1710b	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_pseudomallei_668	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_pseudomallei_K96243	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_pseudomallei_MSHR346	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Burkholderia_thailandensis_E264	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_vietnamiensis_G4	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Burkholderia_xenovorans_LB400	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Caldicellulosiruptor_saccharolyticus_DSM_8903	Other_Bacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Caldivirga_maquilingensis_IC-167	Crenarchaeota	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Campylobacter_concisus_13826	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Campylobacter_curvus_525_92	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Campylobacter_fetus_82-40	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Campylobacter_hominis_ATCC_BAA-381	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Campylobacter_jejuni	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Campylobacter_jejuni_81116	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Campylobacter_jejuni_81-176	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Campylobacter_jejuni_doylei_269_97	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Campylobacter_jejuni_RM1221	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Campylobacter_lari_RM2100	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Candidatus_Accumulibacter_phosphatis_clade_IIA_UW_1	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Candidatus_Amoebophilus_asiaticus_5a2	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**

Candidatus_Azobacteroides_pseudotrichonymphae_genomovar__(	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	NO	YES	YES	NO	**NO**
Candidatus_Blochmannia_floridanus	Gammaproteobacteria	NO	NO	NO	YES	NO	NO	NO	YES	NO	**NO**
Candidatus_Blochmannia_pennsylvanicus_BPEN	Bacteroidetes/Chlorobi	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Candidatus_Carsonella_ruddii_PV	Gammaproteobacteria	NO	NO	NO	NO	NO	NO	NO	YES	NO	**NO**
Candidatus_Desulfococcus_oleovorans_Hxd3	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Candidatus_Desulforudis_audaxviator_MP104C	Firmicutes	NO	NO	YES	YES	NO	NO	YES	YES	NO	**NO**
Candidatus_Hamiltonella_defensa_5AT__Acyrthosiphon_pisum_	Gammaproteobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Candidatus_Hodgkinia_cicadicola_Dsem	Alphaproteobacteria	NO	NO	NO	NO	NO	NO	NO	NO	NO	**NO**
Candidatus_Korarchaeum_cryptofilum_OPF8	Other_Archaea	NO	NO	YES	YES	NO	NO	YES	YES	NO	**NO**
Candidatus_Koribacter_versatilis_Ellin345	Acidobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Candidatus_Liberibacter_asiaticus_psy62	Alphaproteobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Candidatus_Methanoregula_boonei_6A8	Euryarchaeota	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Candidatus_Methanosphaerula_palustris_E1_9c	Euryarchaeota	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Candidatus_Pelagibacter_ubique_HTCC1062	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Candidatus_Phytoplasma_australiense	Firmicutes	NO	NO	NO	NO	NO	NO	NO	NO	NO	**NO**
Candidatus_Phytoplasma_mali	Other_Bacteria	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Candidatus_Puniceispirillum_marinum_IMCC1322_uid47081	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Candidatus_Riesia_pediculicola_USDA_uid46841	Gammaproteobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Candidatus_Ruthia_magnifica_Cm_Calyptogena_magnifica_	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Candidatus_Sulcia_muelleri_DMIN_uid47075	Bacteroidetes/Chlorobi	NO	NO	NO	NO	NO	NO	NO	YES	NO	**NO**
Candidatus_Sulcia_muelleri_GWSS	Bacteroidetes/Chlorobi	NO	NO	NO	NO	NO	NO	NO	YES	NO	**NO**
Candidatus_Sulcia_muelleri_SMDSEM	Bacteroidetes/Chlorobi	NO	NO	NO	NO	NO	NO	NO	YES	NO	**NO**
Candidatus_Vesicomyosocius_okutanii_HA	Gammaproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Capnocytophaga_ochracea_DSM_7271	Bacteroidetes/Chlorobi	NO	NO	NO	YES	NO	YES	YES	NO	NO	**NO**
Carboxydothermus_hydrogenoformans_Z-2901	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Catenulispora_acidiphila_DSM_44928	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Caulobacter_crescentus	Alphaproteobacteria	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Caulobacter_crescentus_NA1000	Alphaproteobacteria	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Caulobacter_K31	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Caulobacter_segnis_ATCC_21756_uid41709	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Cellulomonas_flavigena_DSM_20109_uid48821	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Cellvibrio_japonicus_Ueda107	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chitinophaga_pinensis_DSM_2588	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chlamydia_muridarum	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydia_trachomatis_434_Bu	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydia_trachomatis_A_HAR-13	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydia_trachomatis_B_Jali20_OT	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydia_trachomatis_B_TZ1A828_OT	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydia_trachomatis_D_UW_3_CX	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydia_trachomatis_L2b_UCH_1_proctitis	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydophila_abortus_S26_3	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydophila_caviae	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydophila_felis_Fe_C-56	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Chlamydophila_pneumoniae_AR39	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	NO	NO	NO	NO	**NO**
Chlamydophila_pneumoniae_CWL029	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	NO	NO	NO	NO	**NO**
Chlamydophila_pneumoniae_J138	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	NO	NO	NO	NO	**NO**
Chlamydophila_pneumoniae_TW_183	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	NO	NO	NO	NO	**NO**
Chlorobaculum_parvum_NCIB_8327	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chlorobium_chlorochromatii_CaD3	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Chlorobium_limicola_DSM_245	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chlorobium_luteolum_DSM_273	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chlorobium_phaeobacteroides_BS1	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Chlorobium_phaeobacteroides_DSM_266	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chlorobium_tepidum_TLS	Bacteroidetes/Chlorobi	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Chloroflexus_aggregans_DSM_9485	Chloroflexi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chloroflexus_aurantiacus_J_10_fl	Chloroflexi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chloroflexus_Y_400_fl	Chloroflexi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chloroherpeton_thalassium_ATCC_35110	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chromobacterium_violaceum	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Chromohalobacter_salexigens_DSM_3043	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Citrobacter_koseri_ATCC_BAA-895	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Citrobacter_rodentium_ICC168	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Clavibacter_michiganensis_NCPPB_382	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Clavibacter_michiganensis_sepedonicus	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Clostridiales_genomosp__BVAB3_UPII9_5_uid46219	Firmicutes	NO	NO	NO	YES	YES	YES	YES	NO	NO	**NO**
Clostridium_acetobutylicum	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_beijerinckii_NCIMB_8052	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Clostridium_botulinum_A	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_A_ATCC_19397	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_A_Hall	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_A2_Kyoto	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_A3_Loch_Maree	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_B_Eklund_17B	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_B1_Okra	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_Ba4_657	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_E3_Alaska_E43	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_botulinum_F_Langeland	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_cellulolyticum_H10	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_difficile_630	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Clostridium_difficile_CD196	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Clostridium_difficile_R20291	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Clostridium_kluyveri_DSM_555	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Clostridium_kluyveri_NBRC_12016	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Clostridium_ljungdahlii_ATCC_49587_uid50583	Firmicutes	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Clostridium_novyi_NT	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_perfringens	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Clostridium_perfringens_ATCC_13124	Firmicutes	NO	NO	YES	YES	YES	YES	NO	NO	NO	**NO**
Clostridium_perfringens_SM101_uid12521	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Clostridium_phytofermentans_ISDg	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_tetani_E88	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Clostridium_thermocellum_ATCC_27405	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Colwellia_psychrerythraea_34H	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Comamonas_testosteroni_CNB_1_uid29203	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Conexibacter_woesei_DSM_14684_uid43467	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Coprothermobacter_proteolyticus_DSM_5265	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Coraliomargarita_akajimensis_DSM_45221_uid47079	Other_Bacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Corynebacterium_aurimucosum_ATCC_700975	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Corynebacterium_diphtheriae	Actinobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Corynebacterium_efficiens_YS-314	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Corynebacterium_glutamicum_ATCC_13032_Bielefeld	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Corynebacterium_glutamicum_ATCC_13032_Kitasato	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Corynebacterium_glutamicum_R	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Corynebacterium_jeikeium_K411	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Corynebacterium_kroppenstedtii_DSM_44385	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Corynebacterium_pseudotuberculosis_uid50585	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Corynebacterium_urealyticum_DSM_7109	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Coxiella_burnetii	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Coxiella_burnetii_CbuG_Q212	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Coxiella_burnetii_CbuK_Q154	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Coxiella_burnetii_Dugway_7E9-12	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Coxiella_burnetii_RSA_331	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Croceibacter_atlanticus_HTCC2559_uid49661	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Cronobacter_turicensis_z3032_uid40821	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Cryptobacterium_curtum_DSM_15641	Actinobacteria	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Cupriavidus_metallidurans_CH34_uid250	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Cupriavidus_taiwanensis	Betaproteobacteria	NO	YES	YES	YES	YES	YES	YES	YES	YES	**?**
Cyanobacteria_bacterium_Yellowstone_A-Prime	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Cyanobacteria_bacterium_Yellowstone_B-Prime	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Cyanothece_ATCC_51142	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Cyanothece_PCC_7424	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Cyanothece_PCC_7425	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Cyanothece_PCC_8801	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Cyanothece_PCC_8802	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Cytophaga_hutchinsonii_ATCC_33406	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Dechloromonas_aromatica_RCB	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Deferribacter_desulfuricans_SSM1_uid46653	Other_Bacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Dehalococcoides_BAV1	Chloroflexi	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Dehalococcoides_CBDB1	Chloroflexi	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Dehalococcoides_ethenogenes_195	Chloroflexi	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Dehalococcoides_GT_uid42115	Chloroflexi	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Dehalococcoides_VS	Chloroflexi	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Dehalogenimonas_lykanthroporepellens_BL_DC_9_uid48131	Chloroflexi	NO	NO	YES	YES	NO	YES	YES	YES	YES	**NO**
Deinococcus_deserti_VCD115	Deinococcus-Thermus	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Deinococcus_geothermalis_DSM_11300	Deinococcus-Thermus	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Deinococcus_radiodurans	Deinococcus-Thermus	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Delftia_acidovorans_SPH-1	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Denitrovibrio_acetiphilus_DSM_12809_uid46657	Other_Bacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfatibacillum_alkenivorans_AK_01	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfitobacterium_hafniense_DCB_2	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Desulfitobacterium_hafniense_Y51	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Desulfobacterium_autotrophicum_HRM2	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfohalobium_retbaense_DSM_5692	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfomicrobium_baculatum_DSM_4028	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Desulfotalea_psychrophila_LSv54	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Desulfotomaculum_acetoxidans_DSM_771	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfotomaculum_reducens_MI-1	Firmicutes	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Desulfovibrio_desulfuricans_ATCC_27774	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Desulfovibrio_desulfuricans_G20	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfovibrio_magneticus_RS_1	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfovibrio_salexigens_DSM_2638	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfovibrio_vulgaris__Miyazaki_F_	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfovibrio_vulgaris_DP4	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfovibrio_vulgaris_Hildenborough	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Desulfurivibrio_alkaliphilus_AHT2_uid49487	Deltaproteobacteria	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Desulfurococcus_kamchatkensis_1221n	Crenarchaeota	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Diaphorobacter_TPSY_uid29975	Other_Bacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**

Dichelobacter_nodosus_VCS1703A	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Dickeya_dadantii_Ech586	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Dickeya_dadantii_Ech703	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Dickeya_zeae_Ech1591	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Dictyoglomus_thermophilum_H_6_12	Other_Bacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Dictyoglomus_turgidum_DSM_6724	Other_Bacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Dinoroseobacter_shibae_DFL_12	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Dyadobacter_fermentans_DSM_18053	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Edwardsiella_ictaluri_93_146	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Edwardsiella_tarda_EIB202	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Eggerthella_lenta_DSM_2243	Actinobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Ehrlichia_canis_Jake	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Ehrlichia_chaffeensis_Arkansas	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Ehrlichia_ruminantium_Gardel	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Ehrlichia_ruminantium_str._Welgevonden_CIRAD	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Ehrlichia_ruminantium_Welgevonden_UPSA	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Elusimicrobium_minutum_Pei191	Other_Bacteria	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Enterobacter_638	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Enterobacter_cloacae_ATCC_13047_uid48363	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Enterobacter_sakazakii_ATCC_BAA-894	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Enterococcus_faecalis_V583	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Erwinia_amylovora_ATCC_49946_uid46943	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Erwinia_amylovora_CFBP1430_uid46839	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Erwinia_billingiae_Eb661_uid50547	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Erwinia_carotovora_atroseptica_SCRI1043	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Erwinia_pyrifoliae_Ep1_96_uid40659	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Erwinia_tasmaniensis	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Erythrobacter_litoralis_HTCC2594	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Escherichia_coli__BL21_Gold_DE3_pLysS_AG__uid30681	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_0127_H6_E2348_69	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Escherichia_coli_536	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Escherichia_coli_55989	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_APEC_O1	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Escherichia_coli_B_REL606	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_BW2952	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_C_ATCC_8739	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_CFT073	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Escherichia_coli_E24377A	Gammaproteobacteria	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Escherichia_coli_ED1a	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Escherichia_coli_HS	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_IAI1	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_IAI39	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_K_12_substr__DH10B	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_K_12_substr__MG1655	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_K_12_substr__W3110	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_O103_H2_12009	Gammaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Escherichia_coli_O111_H__11128	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_O157_H7_EC4115	Gammaproteobacteria	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Escherichia_coli_O157_H7_TW14359	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_O157H7	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_O157H7_EDL933	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_O26_H11_11368	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**

Escherichia_coli_O55_H7_CB9615_uid46655	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_S88	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Escherichia_coli_SE11	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_SMS_3_5	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_UMN026	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Escherichia_coli_UTI89	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Escherichia_fergusonii_ATCC_35469	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Eubacterium_eligens_ATCC_27750	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Eubacterium_rectale_ATCC_33656	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Exiguobacterium_AT1b	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Exiguobacterium_sibiricum_255_15	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Ferroglobus_placidus_DSM_10642_uid40863	Euryarchaeota	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Fervidobacterium_nodosum_Rt17-B1	Thermotogae	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Fibrobacter_succinogenes_S85	Other_Bacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Finegoldia_magna_ATCC_29328	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Flavobacteriaceae_bacterium_3519_10	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Flavobacterium_johnsoniae_UW101	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Flavobacterium_psychrophilum_JIP02_86	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Francisella_philomiragia_ATCC_25017	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Francisella_tularensis_FSC_198	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Francisella_tularensis_holarctica	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Francisella_tularensis_holarctica_FTNF002_00	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	NO	NO	**NO**
Francisella_tularensis_holarctica_OSU18	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	NO	NO	**NO**
Francisella_tularensis_mediasiatica_FSC147	Gammaproteobacteria	NO	NO	YES	YES	NO	YES	YES	NO	NO	**NO**
Francisella_tularensis_novicida_U112	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Francisella_tularensis_tularensis	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Francisella_tularensis_WY96-3418	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Frankia_alni_ACN14a	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Frankia_CcI3	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Frankia_EAN1pec	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Fusobacterium_nucleatum	Fusobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Gardnerella_vaginalis_409_05_uid43211	Actinobacteria	NO	NO	NO	YES	NO	YES	YES	NO	NO	**NO**
Gemmatimonas_aurantiaca_T_27	Other_Bacteria	NO	NO	YES	YES	YES	YES	YES	NO	NO	**NO**
Geobacillus_C56_T3_uid49467	Firmicutes	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Geobacillus_kaustophilus_HTA426	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacillus_thermodenitrificans_NG80-2	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacillus_WCH70	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacillus_Y412MC10	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacillus_Y412MC61	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacter_bemidjiensis_Bem	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacter_FRC_32	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacter_lovleyi_SZ	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacter_M21	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacter_metallireducens_GS-15	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacter_sulfurreducens	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Geobacter_uraniumreducens_Rf4	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Geodermatophilus_obscurus_DSM_43160_uid43725	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Gloeobacter_violaceus	Cyanobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Gluconacetobacter_diazotrophicus_PAl_5_FAPERJ	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Gluconacetobacter_diazotrophicus_PAl_5_JGI	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Gluconobacter_oxydans_621H	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Gordonia_bronchialis_DSM_43247	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**

Gramella_forsetii_KT0803	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Granulobacter_bethesdensis_CGDNIH1	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Haemophilus_ducreyi_35000HP	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Haemophilus_influenzae	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Haemophilus_influenzae_86_028NP	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Haemophilus_influenzae_PittEE	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Haemophilus_influenzae_PittGG	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Haemophilus_parasuis_SH0165	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Haemophilus_somnus_129PT	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Haemophilus_somnus_2336	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Hahella_chejuensis_KCTC_2396	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Halalkalicoccus_jeotgali_B3_uid50305	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Haliangium_ochraceum_DSM_14365	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Haloarcula_marismortui_ATCC_43049	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Halobacterium_salinarum_R1	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	NO	NO	**NO**
Halobacterium_sp	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	NO	NO	**NO**
Haloferax_volcanii_DS2_uid46845	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Halomicrobium_mukohataei_DSM_12286	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Haloquadratum_walsbyi	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Halorhabdus_utahensis_DSM_12940	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Halorhodospira_halophila_SL1	Gammaproteobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Halorubrum_lacusprofundi_ATCC_49239	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Haloterrigena_turkmenica_DSM_5511_uid43501	Euryarchaeota	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Halothermothrix_orenii_H_168	Firmicutes	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Halothiobacillus_neapolitanus_c2	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Helicobacter_acinonychis_Sheeba	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Helicobacter_hepaticus	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Helicobacter_mustelae_12198_uid46647	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Helicobacter_pylori_26695	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Helicobacter_pylori_B38_uid39685	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Helicobacter_pylori_B8_uid49873	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Helicobacter_pylori_G27	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Helicobacter_pylori_HPAG1	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Helicobacter_pylori_J99	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Helicobacter_pylori_P12	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Helicobacter_pylori_Shi470	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Heliobacterium_modesticaldum_Ice1	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Herbaspirillum_seropedicae_SmR1_uid50427	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Herminiimonas_arsenicoxydans	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Herpetosiphon_aurantiacus_ATCC_23779	Chloroflexi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Hirschia_baltica_ATCC_49814	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Hydrogenobacter_thermophilus_TK_6_uid45927	Aquificae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Hydrogenobaculum_Y04AAS1	Aquificae	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Hyperthermus_butylicus	Crenarchaeota	NO	NO	NO	YES	NO	YES	YES	NO	NO	**NO**
Hyphomicrobium_denitrificans_ATCC_51888_uid50325	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Hyphomonas_neptunium_ATCC_15444	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Idiomarina_loihiensis_L2TR	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Ignicoccus_hospitalis_KIN4_I	Crenarchaeota	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Jannaschia_CCS1	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Janthinobacterium_Marseille	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Jonesia_denitrificans_DSM_20603	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Kangiella_koreensis_DSM_16069	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Kineococcus_radiotolerans_SRS30216	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Klebsiella_pneumoniae_342	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Klebsiella_pneumoniae_MGH_78578	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Klebsiella_pneumoniae_NTUH_K2044	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Klebsiella_variicola_At_22_uid42113	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Kocuria_rhizophila_DC2201	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Kosmotoga_olearia_TBF_19_5_1	Thermotogae	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Kribbella_flavida_DSM_17836_uid43465	Actinobacteria	NO	YES	YES	YES	YES	YES	YES	YES	YES	**?**
Kytococcus_sedentarius_DSM_20547	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Lactobacillus_acidophilus_NCFM	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Lactobacillus_brevis_ATCC_367	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Lactobacillus_casei	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Lactobacillus_casei_ATCC_334	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Lactobacillus_casei_Zhang_uid50673	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Lactobacillus_crispatus_ST1_uid48359	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Lactobacillus_delbrueckii_bulgaricus	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Lactobacillus_delbrueckii_bulgaricus_ATCC_BAA-365	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Lactobacillus_fermentum_IFO_3956	Firmicutes	NO	NO	NO	YES	YES	YES	YES	NO	NO	**NO**
Lactobacillus_gasseri_ATCC_33323	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Lactobacillus_helveticus_DPC_4571	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Lactobacillus_johnsonii_FI9785	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Lactobacillus_johnsonii_NCC_533	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Lactobacillus_plantarum	Firmicutes	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Lactobacillus_plantarum_JDM1	Firmicutes	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Lactobacillus_reuteri_DSM_20016	Firmicutes	NO	NO	NO	YES	YES	YES	YES	NO	NO	**NO**
Lactobacillus_reuteri_F275_Kitasato	Firmicutes	NO	NO	NO	YES	YES	YES	YES	NO	NO	**NO**
Lactobacillus_rhamnosus_GG	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Lactobacillus_rhamnosus_Lc_705	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Lactobacillus_sakei_23K	Firmicutes	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Lactobacillus_salivarius_UCC118	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Lactococcus_lactis	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Lactococcus_lactis_cremoris_MG1363	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Lactococcus_lactis_cremoris_SK11	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Lactococcus_lactis_KF147	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Laribacter_hongkongensis_HLHK9	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Lawsonia_intracellularis_PHE_MN1-00	Deltaproteobacteria	NO	NO	NO	YES	NO	NO	YES	NO	NO	**NO**
Legionella_longbeachae_NSW150_uid46099	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Legionella_pneumophila_2300_99_Alcoy_uid48801	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Legionella_pneumophila_Corby	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Legionella_pneumophila_Lens	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Legionella_pneumophila_Paris	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Legionella_pneumophila_Philadelphia_1	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Leifsonia_xyli_xyli_CTCB0	Actinobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Leptospira_biflexa_serovar_Patoc__Patoc_1__Ames_	Spirochaetes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Leptospira_biflexa_serovar_Patoc__Patoc_1__Paris_	Spirochaetes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Leptospira_borgpetersenii_serovar_Hardjo-bovis_JB197	Spirochaetes	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Leptospira_borgpetersenii_serovar_Hardjo-bovis_L550	Spirochaetes	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Leptospira_interrogans_serovar_Copenhageni	Spirochaetes	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Leptospira_interrogans_serovar_Lai	Spirochaetes	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Leptothrix_cholodnii_SP_6	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Leptotrichia_buccalis_DSM_1135	Fusobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Leuconostoc_citreum_KM20	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**

Leuconostoc_gasicomitatum_LMG_18811_uid50385	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Leuconostoc_kimchii_IMSNU_11154_uid48589	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Leuconostoc_mesenteroides_ATCC_8293	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Listeria_innocua	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Listeria_monocytogenes	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Listeria_monocytogenes_08_5923_uid43727	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Listeria_monocytogenes_4b_F2365	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Listeria_monocytogenes_Clip81459	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Listeria_monocytogenes_HCC23	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Listeria_monocytogenes_uid43671	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Listeria_seeligeri_serovar_1_2b_SLCC3954_uid46215	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Listeria_welshimeri_serovar_6b_SLCC5334	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Lysinibacillus_sphaericus_C3_41	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Macrococcus_caseolyticus_JCSC5402	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Magnetococcus_MC-1	Other_Bacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Magnetospirillum_magneticum_AMB-1	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Mannheimia_succiniciproducens_MBEL55E	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Maricaulis_maris_MCS10	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Marinobacter_aquaeolei_VT8	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Marinomonas_MWYL1	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Meiothermus_ruber_DSM_1279_uid46661	Deinococcus-Thermus	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Meiothermus_silvanus_DSM_9946_uid49485	Deinococcus-Thermus	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Mesoplasma_florum_L1	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mesorhizobium_BNC1	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Mesorhizobium_loti	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Metallosphaera_sedula_DSM_5348	Crenarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Methanobacterium_thermoautotrophicum	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Methanobrevibacter_ruminantium_M1_uid45857	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Methanobrevibacter_smithii_ATCC_35061	Euryarchaeota	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Methanocaldococcus_fervens_AG86	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanocaldococcus_FS406_22_uid42499	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanocaldococcus_infernus_ME_uid48803	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanocaldococcus_vulcanius_M7	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanocella_paludicola_SANAE	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Methanococcoides_burtonii_DSM_6242	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Methanococcus_aeolicus_Nankai-3	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanococcus_jannaschii	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanococcus_maripaludis_C5	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanococcus_maripaludis_C6	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanococcus_maripaludis_C7	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanococcus_maripaludis_S2	Euryarchaeota	NO	NO	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanococcus_vannielii_SB	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Methanococcus_voltae_A3_uid49529	Euryarchaeota	NO	NO	NO	YES	NO	NO	NO	YES	NO	**NO**
Methanocorpusculum_labreanum_Z	Euryarchaeota	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Methanoculleus_marisnigri_JR1	Euryarchaeota	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Methanohalobium_evestigatum_Z_7303_uid49857	Euryarchaeota	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Methanohalophilus_mahii_DSM_5219_uid47313	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Methanopyrus_kandleri	Euryarchaeota	NO	YES	NO	YES	NO	NO	YES	YES	NO	**NO**
Methanosaeta_thermophila_PT	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Methanosarcina_acetivorans	Euryarchaeota	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Methanosarcina_barkeri_fusaro	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Methanosarcina_mazei	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Methanosphaera_stadtmanae	Euryarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Methanospirillum_hungatei_JF-1	Euryarchaeota	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Methylacidiphilum_infernorum_V4	Other_Bacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Methylibium_petroleiphilum_PM1	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Methylobacillus_flagellatus_KT	Betaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Methylobacterium_4_46	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Methylobacterium_chloromethanicum_CM4	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Methylobacterium_extorquens_AM1	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Methylobacterium_extorquens_DM4	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Methylobacterium_extorquens_PA1	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Methylobacterium_nodulans_ORS_2060	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Methylobacterium_populi_BJ001	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Methylobacterium_radiotolerans_JCM_2831	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Methylocella_silvestris_BL2	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Methylococcus_capsulatus_Bath	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Methylotenera_301_uid49469	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Methylotenera_mobilis_JLW8	Betaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Methylovorus_SIP3_4	Betaproteobacteria	YES	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Micrococcus_luteus_NCTC_2665	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Microcystis_aeruginosa_NIES_843	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Mobiluncus_curtisii_ATCC_43063_uid49695	Actinobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Moorella_thermoacetica_ATCC_39073	Firmicutes	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Moraxella_catarrhalis_RH4_uid48809	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Mycobacterium_abscessus_ATCC_19977	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Mycobacterium_avium_104	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_avium_paratuberculosis	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_bovis	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Mycobacterium_bovis_BCG_Pasteur_1173P2	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Mycobacterium_bovis_BCG_Tokyo_172	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Mycobacterium_gilvum_PYR-GCK	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_JLS	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_KMS	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_leprae	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Mycobacterium_leprae_Br4923	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Mycobacterium_marinum_M	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_MCS	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_smegmatis_MC2_155	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_tuberculosis_CDC1551	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Mycobacterium_tuberculosis_F11	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Mycobacterium_tuberculosis_H37Ra	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Mycobacterium_tuberculosis_H37Rv	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Mycobacterium_tuberculosis_KZN_1435	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Mycobacterium_ulcerans_Agy99	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Mycobacterium_vanbaalenii_PYR-1	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Mycoplasma_agalactiae_PG2	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_agalactiae_uid46679	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_arthritidis_158L3_1	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_capricolum_ATCC_27343	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_conjunctivae_HRC_581_uid32285	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Mycoplasma_crocodyli_MP145_uid47087	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_gallisepticum	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Mycoplasma_genitalium	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**

Mycoplasma_hominis	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Mycoplasma_hyopneumoniae_232	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Mycoplasma_hyopneumoniae_7448	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Mycoplasma_hyopneumoniae_J	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Mycoplasma_mobile_163K	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_mycoides	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_penetrans	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_pneumoniae	Firmicutes	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Mycoplasma_pulmonis	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Mycoplasma_synoviae_53	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Myxococcus_xanthus_DK_1622	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nakamurella_multipartita_DSM_44233	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nanoarchaeum_equitans	Nanoarchaeota	NO	NO	NO	YES	NO	NO	NO	NO	NO	**NO**
Natranaerobius_thermophilus_JW_NM_WN_LF	Firmicutes	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Natrialba_magadii_ATCC_43099_uid46245	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Natronomonas_pharaonis	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nautilia_profundicola_AmH	Epsilonproteobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Neisseria_gonorrhoeae_FA_1090	Betaproteobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Neisseria_gonorrhoeae_NCCP11945	Betaproteobacteria	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Neisseria_meningitidis_053442	Betaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Neisseria_meningitidis_alpha14	Betaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Neisseria_meningitidis_FAM18	Betaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Neisseria_meningitidis_MC58	Betaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Neisseria_meningitidis_Z2491	Betaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Neorickettsia_risticii_Illinois	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Neorickettsia_sennetsu_Miyayama	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Nitratiruptor_SB155-2	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Nitrobacter_hamburgensis_X14	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nitrobacter_winogradskyi_Nb-255	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nitrosococcus_halophilus_Nc4_uid46803	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nitrosococcus_oceani_ATCC_19707	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nitrosococcus_watsoni_C_113_uid50331	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nitrosomonas_europaea	Betaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Nitrosomonas_eutropha_C71	Betaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Nitrosopumilus_maritimus_SCM1	Crenarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Nitrosospira_multiformis_ATCC_25196	Betaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Nocardia_farcinica_IFM10152	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Nocardioides_JS614	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Nocardiopsis_dassonvillei_DSM_43111_uid49483	Actinobacteria	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Nostoc_punctiforme_PCC_73102	Cyanobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Nostoc_sp	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Novosphingobium_aromaticivorans_DSM_12444	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Oceanobacillus_iheyensis	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Ochrobactrum_anthropi_ATCC_49188	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Oenococcus_oeni_PSU-1	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Oligotropha_carboxidovorans_OM5	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Onion_yellows_phytoplasma	Firmicutes	NO	NO	NO	NO	NO	NO	NO	NO	NO	**NO**
Opitutus_terrae_PB90_1	Other_Bacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Orientia_tsutsugamushi_Boryong	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Orientia_tsutsugamushi_Ikeda	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Paenibacillus_JDR_2	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Pantoea_ananatis_LMG_20103_uid46807	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Parabacteroides_distasonis_ATCC_8503	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Parachlamydia_sp_UWE25	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Paracoccus_denitrificans_PD1222	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Parvibaculum_lavamentivorans_DS-1	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Pasteurella_multocida	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Pectobacterium_carotovorum_PC1	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Pectobacterium_wasabiae_WPP163	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Pediococcus_pentosaceus_ATCC_25745	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Pedobacter_heparinus_DSM_2366	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Pelobacter_carbinolicus	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Pelobacter_propionicus_DSM_2379	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Pelodictyon_phaeoclathratiforme_BU_1	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Pelotomaculum_thermopropionicum_SI	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Persephonella_marina_EX_H1	Aquificae	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Petrotoga_mobilis_SJ95	Thermotogae	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Phenylobacterium_zucineum_HLK1	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Photobacterium_profundum_SS9	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Photorhabdus_asymbiotica	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Photorhabdus_luminescens	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Picrophilus_torridus_DSM_9790	Euryarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Pirellula_sp	Planctomycetes	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Pirellula_staleyi_DSM_6068_uid43209	Planctomycetes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Planctomyces_limnophilus_DSM_3776_uid48643	Planctomycetes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Polaromonas_JS666	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Polaromonas_naphthalenivorans_CJ2	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Polynucleobacter_necessarius_asymbioticus_QLW_P1DMWA_1	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Polynucleobacter_necessarius_STIR1	Betaproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Porphyromonas_gingivalis_ATCC_33277	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Porphyromonas_gingivalis_W83	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Prevotella_ruminicola_23_uid47507	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Prochlorococcus_marinus_AS9601	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_CCMP1375	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_MED4	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_MIT_9211	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_MIT_9215	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_MIT_9301	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_MIT_9312	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_MIT_9515	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_MIT9313	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_NATL1A	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Prochlorococcus_marinus_NATL2A	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Propionibacterium_acnes_KPA171202	Actinobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Propionibacterium_acnes_SK137_uid48071	Actinobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Propionibacterium_freudenreichii_shermanii_CIRM_BIA1_uid4953	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Prosthecochloris_aestuarii_DSM_271	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Prosthecochloris_vibrioformis_DSM_265	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Proteus_mirabilis	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Pseudoalteromonas_atlantica_T6c	Gammaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	YES	**?**
Pseudoalteromonas_haloplanktis_TAC125	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Pseudomonas_aeruginosa	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_aeruginosa_LESB58	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_aeruginosa_PA7	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**

Pseudomonas_aeruginosa_UCBPP-PA14	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_entomophila_L48	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_fluorescens_Pf0_1	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_fluorescens_Pf-5	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_fluorescens_SBW25	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_mendocina_ymp	Gammaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Pseudomonas_putida_F1	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_putida_GB_1	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_putida_KT2440	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_putida_W619	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_stutzeri_A1501	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Pseudomonas_syringae_phaseolicola_1448A	Gammaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Pseudomonas_syringae_pv_B728a	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Pseudomonas_syringae_tomato_DC3000	Gammaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Psychrobacter_arcticum_273-4	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Psychrobacter_cryohalolentis_K5	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Psychrobacter_PRwf-1	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Psychromonas_ingrahamii_37	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Pyrobaculum_aerophilum	Crenarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Pyrobaculum_arsenaticum_DSM_13514	Crenarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Pyrobaculum_calidifontis_JCM_11548	Crenarchaeota	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Pyrobaculum_islandicum_DSM_4184	Crenarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Pyrococcus_abyssi	Euryarchaeota	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Pyrococcus_furiosus	Euryarchaeota	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Pyrococcus_horikoshii	Euryarchaeota	NO	NO	NO	YES	YES	NO	YES	YES	NO	**NO**
Ralstonia_eutropha_H16	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Ralstonia_eutropha_JMP134	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Ralstonia_pickettii_12D	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Ralstonia_pickettii_12J	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Ralstonia_solanacearum	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Ralstonia_solanacearum_CFBP2957_uid50545	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Renibacterium_salmoninarum_ATCC_33209	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhizobium_etli_CFN_42	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Rhizobium_etli_CIAT_652	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Rhizobium_leguminosarum_bv__trifolii_WSM1325	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Rhizobium_leguminosarum_bv__trifolii_WSM2304	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Rhizobium_leguminosarum_bv_viciae_3841	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhizobium_NGR234	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Rhodobacter_capsulatus_SB_1003_uid47509	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Rhodobacter_sphaeroides_2_4_1	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodobacter_sphaeroides_ATCC_17025	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodobacter_sphaeroides_ATCC_17029	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodobacter_sphaeroides_KD131	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodococcus_erythropolis_PR4	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Rhodococcus_jostii_RHA1	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Rhodococcus_opacus_B4_uid13791	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Rhodoferax_ferrireducens_T118	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodopseudomonas_palustris_BisA53	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodopseudomonas_palustris_BisB18	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodopseudomonas_palustris_BisB5	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodopseudomonas_palustris_CGA009	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodopseudomonas_palustris_HaA2	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**

Rhodopseudomonas_palustris_TIE_1	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodospirillum_centenum_SW	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodospirillum_rubrum_ATCC_11170	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rhodothermus_marinus_DSM_4252	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Rickettsia_africae_ESF_5	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_akari_Hartford	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_bellii_OSU_85-389	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_bellii_RML369-C	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_canadensis_McKiel	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_conorii	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_felis_URRWXCal2	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_massiliae_MTU5	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_peacockii_Rustic	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_prowazekii	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_rickettsii_Iowa	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_rickettsii_Sheila_Smith	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Rickettsia_typhi_wilmington	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Robiginitalea_biformata_HTCC2501	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Roseiflexus_castenholzii_DSM_13941	Chloroflexi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Roseiflexus_RS-1	Chloroflexi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Roseobacter_denitrificans_OCh_114	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Rothia_mucilaginosa	Actinobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Rubrobacter_xylanophilus_DSM_9941	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Ruegeria_pomeroyi_DSS_3	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Saccharomonospora_viridis_DSM_43017	Actinobacteria	YES	NO	YES	YES	YES	YES	YES	YES	YES	**?**
Saccharophagus_degradans_2-40	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Saccharopolyspora_erythraea_NRRL_2338	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Salinibacter_ruber_DSM_13855	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salinibacter_ruber_uid47323	Bacteroidetes/Chlorobi	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salinispora_arenicola_CNS-205	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Salinispora_tropica_CNB-440	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_arizonae_serovar_62_z4_z23__	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_Choleraesuis	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_Paratypi_ATCC_9150	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Agona_SL483	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Dublin_CT_02021853	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Enteritidis_P125109	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Gallinarum_287_91	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Heidelberg_SL476	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Newport_SL254	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Paratyphi_A_AKU_12601	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Paratyphi_B_SPB7	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Paratyphi_C_RKS4594	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Schwarzengrund_CVM19633	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_enterica_serovar_Typhi_Ty2	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Salmonella_typhi	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Sanguibacter_keddieii_DSM_10542_uid40845	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Sebaldella_termitidis_ATCC_33386	Fusobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Segniliparus_rotundus_DSM_44985_uid49049	Actinobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Serratia_proteamaculans_568	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_amazonensis_SB2B	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_ANA-3	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**

Shewanella_baltica_OS155	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_baltica_OS185	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_baltica_OS195	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_baltica_OS223	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_denitrificans_OS217	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_frigidimarina_NCIMB_400	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_halifaxensis_HAW_EB4	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_loihica_PV-4	Gammaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Shewanella_MR-4	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_MR-7	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_oneidensis	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_pealeana_ATCC_700345	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_piezotolerans_WP3	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_putrefaciens_CN-32	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_sediminis_HAW-EB3	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_violacea_DSS12_uid47085	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_W3-18-1	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Shewanella_woodyi_ATCC_51908	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Shigella_boydii_CDC_3083_94	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Shigella_boydii_Sb227	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Shigella_dysenteriae	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Shigella_flexneri_2a	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Shigella_flexneri_2a_2457T	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Shigella_flexneri_5_8401	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Shigella_sonnei_Ss046	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Sideroxydans_lithotrophicus_ES_1_uid46801	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Silicibacter_TM1040	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Sinorhizobium_medicae_WSM419	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Sinorhizobium_meliloti	Alphaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Slackia_heliotrinireducens_DSM_20476	Actinobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Sodalis_glossinidius_morsitans	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Solibacter_usitatus_Ellin6076	Acidobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Sorangium_cellulosum__So_ce_56_	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Sphaerobacter_thermophilus_DSM_20745	Chloroflexi	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Sphingobium_japonicum_UT26S_uid47077	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sphingomonas_wittichii_RW1	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sphingopyxis_alaskensis_RB2256	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Spirosoma_linguale_DSM_74_uid43413	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Stackebrandtia_nassauensis_DSM_44728_uid46663	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Staphylococcus_aureus_aureus_MRSA252	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_aureus_MSSA476	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_COL	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_ED98	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_JH1	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_JH9	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_Mu3	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_Mu50	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_MW2	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_N315	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_NCTC_8325	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_Newman	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_RF122	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Staphylococcus_aureus_USA300_FPR3757	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_aureus_USA300_TCH1516	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_carnosus_TM300	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_epidermidis_ATCC_12228	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_epidermidis_RP62A	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_haemolyticus	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_lugdunensis_HKU09_01_uid46233	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Staphylococcus_saprophyticus	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Staphylothermus_hellenicus_DSM_12710_uid45893	Crenarchaeota	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Staphylothermus_marinus_F1	Crenarchaeota	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Starkeya_novella_DSM_506_uid48815	Alphaproteobacteria	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Stenotrophomonas_maltophilia_K279a	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Stenotrophomonas_maltophilia_R551_3	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Streptobacillus_moniliformis_DSM_12112	Fusobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Streptococcus_agalactiae_2603	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_agalactiae_A909	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_agalactiae_NEM316	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_dysgalactiae_equisimilis_GGS_124	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_equi_4047	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_equi_zooepidemicus	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_equi_zooepidemicus_MGCS10565	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_gallolyticus_UCN34_uid46061	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_gordonii_Challis_substr_CH1	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_mitis_B6_uid46097	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_mutans	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_mutans_NN2025_uid46353	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pneumoniae_70585	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pneumoniae_ATCC_700669	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_pneumoniae_CGSP14	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pneumoniae_D39	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_pneumoniae_G54	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pneumoniae_Hungary19A_6	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pneumoniae_JJA	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_pneumoniae_P1031	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_pneumoniae_R6	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pneumoniae_Taiwan19F_14	Firmicutes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Streptococcus_pneumoniae_TCH8431_19A_uid49735	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pneumoniae_TIGR4	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_M1_GAS	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_Manfredo	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_MGAS10270	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_MGAS10394	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_MGAS10750	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_MGAS2096	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_MGAS315	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_MGAS5005	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_MGAS6180	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_MGAS8232	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_NZ131	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_pyogenes_SSI-1	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_sanguinis_SK36	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_suis_05ZYH33	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**

Streptococcus_suis_98HAH33	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_suis_BM407	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_suis_P1_7	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_suis_SC84	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_thermophilus_CNRZ1066	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_thermophilus_LMD-9	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_thermophilus_LMG_18311	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptococcus_uberis_0140J	Firmicutes	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Streptomyces_avermitilis	Actinobacteria	YES	NO	YES	YES	YES	YES	YES	YES	YES	**?**
Streptomyces_coelicolor	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Streptomyces_griseus_NBRC_13350	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Streptomyces_scabiei_87_22_uid46531	Actinobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Streptosporangium_roseum_DSM_43021	Actinobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**
Sulfolobus_acidocaldarius_DSM_639	Crenarchaeota	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sulfolobus_islandicus_L_D_8_5_uid43679	Crenarchaeota	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sulfolobus_islandicus_L_S_2_15	Crenarchaeota	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sulfolobus_islandicus_M_14_25	Crenarchaeota	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sulfolobus_islandicus_M_16_27	Crenarchaeota	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sulfolobus_islandicus_M_16_4	Crenarchaeota	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sulfolobus_islandicus_Y_G_57_14	Crenarchaeota	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Sulfolobus_islandicus_Y_N_15_51	Crenarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Sulfolobus_solfataricus	Crenarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Sulfolobus_tokodaii	Crenarchaeota	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Sulfurihydrogenibium_azorense_Az_Fu1	Aquificae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Sulfurihydrogenibium_YO3AOP1	Aquificae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Sulfurospirillum_deleyianum_DSM_6946	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Sulfurovum_NBC37-1	Epsilonproteobacteria	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Symbiobacterium_thermophilum_IAM14863	Firmicutes	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Synechococcus_CC9311	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Synechococcus_CC9605	Cyanobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Synechococcus_CC9902	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Synechococcus_elongatus_PCC_6301	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Synechococcus_elongatus_PCC_7942	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Synechococcus_PCC_7002	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Synechococcus_RCC307	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Synechococcus_sp_WH8102	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Synechococcus_WH_7803	Cyanobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Syntrophobacter_fumaroxidans_MPOB	Deltaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	YES	**NO**
Syntrophomonas_wolfei_Goettingen	Other_Bacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Syntrophothermus_lipocalidus_DSM_12680_uid49527	Firmicutes	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Syntrophus_aciditrophicus_SB	Deltaproteobacteria	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Teredinibacter_turnerae_T7901	Gammaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Thauera_MZ1T	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Thermanaerovibrio_acidaminovorans_DSM_6589	Other_Bacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Thermincola_JR_uid48823	Firmicutes	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Thermoanaerobacter_italicus_Ab9_uid46241	Firmicutes	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Thermoanaerobacter_mathranii_A3_uid49481	Firmicutes	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Thermoanaerobacter_pseudethanolicus_ATCC_33223	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Thermoanaerobacter_tengcongensis	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Thermoanaerobacter_X514	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Thermobaculum_terrenum_ATCC_BAA_798	Other_Bacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Thermobifida_fusca_YX	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**

Thermobispora_bispora_DSM_43833_uid48999	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Thermococcus_gammatolerans_EJ3	Euryarchaeota	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Thermococcus_kodakaraensis_KOD1	Euryarchaeota	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Thermococcus_onnurineus_NA1	Euryarchaeota	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Thermococcus_sibiricus_MM_739	Euryarchaeota	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Thermocrinis_albus_DSM_14484_uid46231	Aquificae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thermodesulfovibrio_yellowstonii_DSM_11347	Other_Bacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thermofilum_pendens_Hrk_5	Crenarchaeota	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Thermomicrobium_roseum_DSM_5159	Chloroflexi	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Thermomonospora_curvata_DSM_43183	Actinobacteria	NO	YES	YES	YES	YES	YES	YES	YES	NO	**?**
Thermoplasma_acidophilum	Euryarchaeota	NO	NO	YES	YES	NO	YES	YES	NO	NO	**NO**
Thermoplasma_volcanium	Euryarchaeota	NO	NO	YES	YES	NO	YES	YES	YES	NO	**NO**
Thermoproteus_neutrophilus_V24Sta	Crenarchaeota	NO	NO	NO	YES	NO	YES	YES	YES	NO	**NO**
Thermosipho_africanus_TCF52B	Thermotogae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thermosipho_melanesiensis_BI429	Thermotogae	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Thermosphaera_aggregans_DSM_11486_uid48993	Crenarchaeota	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Thermosynechococcus_elongatus	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Thermotoga_lettingae_TMO	Thermotogae	NO	NO	NO	YES	YES	YES	NO	YES	NO	**NO**
Thermotoga_maritima	Thermotogae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thermotoga_naphthophila_RKU_10	Thermotogae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thermotoga_neapolitana_DSM_4359	Thermotogae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thermotoga_petrophila_RKU-1	Thermotogae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thermotoga_RQ2	Thermotogae	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thermus_thermophilus_HB27	Deinococcus-Thermus	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Thermus_thermophilus_HB8	Deinococcus-Thermus	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Thioalkalivibrio_HL_EbGR7	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Thioalkalivibrio_K90mix_uid46181	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Thiobacillus_denitrificans_ATCC_25259	Betaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Thiomicrospira_crunogena_XCL-2	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Thiomicrospira_denitrificans_ATCC_33889	Gammaproteobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Thiomonas_intermedia_K12_uid48825	Betaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Tolumonas_auensis_DSM_9187	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Treponema_denticola_ATCC_35405	Spirochaetes	NO	NO	NO	YES	YES	YES	NO	NO	NO	**NO**
Treponema_pallidum	Spirochaetes	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Treponema_pallidum_SS14	Spirochaetes	NO	NO	NO	YES	NO	YES	NO	YES	NO	**NO**
Trichodesmium_erythraeum_IMS101	Cyanobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Tropheryma_whipplei_TW08_27	Actinobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Tropheryma_whipplei_Twist	Actinobacteria	NO	NO	YES	YES	NO	YES	NO	YES	NO	**NO**
Truepera_radiovictrix_DSM_17093_uid49533	Deinococcus-Thermus	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Tsukamurella_paurometabola_DSM_20162_uid48829	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
uncultured_Termite_group_1_bacterium_phylotype_Rs_D17	Other_Bacteria	NO	NO	NO	YES	NO	NO	NO	YES	NO	**NO**
Ureaplasma_parvum_serovar_3_ATCC_27815	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Ureaplasma_urealyticum	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Ureaplasma_urealyticum_serovar_10_ATCC_33699	Firmicutes	NO	NO	NO	NO	NO	YES	NO	NO	NO	**NO**
Variovorax_paradoxus_S110	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Veillonella_parvula_DSM_2008	Firmicutes	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Verminephrobacter_eiseniae_EF01-2	Betaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	NO	**Toluene**
Vibrio_cholerae	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_cholerae_M66_2	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_cholerae_MJ_1236	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_cholerae_O395	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_Ex25	Gammaproteobacteria	YES	NO	YES	YES	YES	YES	YES	YES	NO	**?**

Vibrio_fischeri_ES114	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_fischeri_MJ11	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_harveyi_ATCC_BAA-1116	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_parahaemolyticus	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_splendidus_LGP32	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_vulnificus_CMCP6	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Vibrio_vulnificus_YJ016	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Waddlia_chondrophila_WSU_86_1044_uid49531	Chlamydiae/Verrucomicrobia	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Wigglesworthia_brevipalpis	Gammaproteobacteria	NO	NO	NO	YES	NO	YES	NO	NO	NO	**NO**
Wolbachia_endosymbiont_of_Brugia_malayi_TRS	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Wolbachia_endosymbiont_of_Culex_quinquefasciatus_Pel	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Wolbachia_endosymbiont_of_Drosophila_melanogaster	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Wolbachia_wRi	Alphaproteobacteria	NO	NO	YES	YES	NO	YES	NO	NO	NO	**NO**
Wolinella_succinogenes	Epsilonproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Xanthobacter_autotrophicus_Py2	Alphaproteobacteria	YES	YES	YES	YES	YES	YES	YES	YES	YES	**All**
Xanthomonas_albilineans	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xanthomonas_campestris_8004	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xanthomonas_campestris_ATCC_33913	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xanthomonas_campestris_B100	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xanthomonas_campestris_vesicatoria_85-10	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xanthomonas_citri	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xanthomonas_oryzae_KACC10331	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xanthomonas_oryzae_MAFF_311018	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xanthomonas_oryzae_PXO99A	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xenorhabdus_bovienii_SS_2004_uid46345	Gammaproteobacteria	NO	NO	NO	YES	YES	YES	YES	YES	NO	**NO**
Xenorhabdus_nematophila_ATCC_19061_uid49133	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xylanimonas_cellulosilytica_DSM_15894	Actinobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Xylella_fastidiosa	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Xylella_fastidiosa_M12	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Xylella_fastidiosa_M23	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Xylella_fastidiosa_Temecula1	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Yersinia_enterocolitica_8081	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pestis_Antiqua	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pestis_biovar_Microtus_91001	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pestis_CO92	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pestis_KIM_10_uid288	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pestis_Nepal516	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pestis_Pestoides_F	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pestis_Z176003_uid47317	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pseudotuberculosis_IP_31758	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pseudotuberculosis_IP32953	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pseudotuberculosis_PB1_	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Yersinia_pseudotuberculosis_YPIII	Gammaproteobacteria	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Zunongwangia_profunda_SM_A87_uid48073	Bacteroidetes/Chlorobi	NO	NO	YES	YES	YES	YES	YES	YES	NO	**NO**
Zymomonas_mobilis_NCIMB_11163	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**
Zymomonas_mobilis_ZM4	Alphaproteobacteria	NO	NO	YES	YES	YES	YES	NO	YES	NO	**NO**

ORGANISM	*GROUP*	*Cox2*	*PCuAC*	*Sco1*	*AcoP*
Acaryochloris_marina_MBIC11017	Cyanobacteria	YES	NO	NO	YES
Acetobacter_pasteurianus_IFO_3283_01	Alphaproteobacteria	NO	YES	YES	YES
Acholeplasma_laidlawii_PG_8A	Firmicutes	NO	NO	NO	NO
Acidaminococcus_fermentans_DSM_20731_uid43471	Firmicutes	NO	NO	NO	NO
Acidimicrobium_ferrooxidans_DSM_10331	Actinobacteria	YES	NO	NO	NO
Acidiphilium_cryptum_JF-5	Alphaproteobacteria	YES	YES	YES	NO
Acidithiobacillus_ferrooxidans_ATCC_23270	Gammaproteobacteria	YES	NO	NO	YES
Acidithiobacillus_ferrooxidans_ATCC_53993	Other_Bacteria	YES	NO	NO	YES
Acidobacterium_capsulatum_ATCC_51196	Acidobacteria	YES	NO	NO	NO
Acidothermus_cellulolyticus_11B	Actinobacteria	YES	NO	YES	YES
Acidovorax_avenae_citrulli_AAC00-1	Betaproteobacteria	YES	YES	YES	NO
Acidovorax_JS42	Betaproteobacteria	YES	YES	YES	YES
Aciduliprofundum_boonei_T469_uid43333	Gammaproteobacteria	NO	NO	NO	NO
Acinetobacter_baumannii_AB0057	Gammaproteobacteria	NO	NO	NO	NO
Acinetobacter_baumannii_AB307_0294	Gammaproteobacteria	NO	NO	NO	NO
Acinetobacter_baumannii_ACICU	Gammaproteobacteria	NO	NO	NO	NO
Acinetobacter_baumannii_ATCC_17978	Gammaproteobacteria	NO	NO	NO	NO
Acinetobacter_baumannii_AYE	Gammaproteobacteria	NO	NO	NO	NO
Acinetobacter_baumannii_SDF	Gammaproteobacteria	NO	NO	NO	NO
Acinetobacter_DR1_uid50119	Gammaproteobacteria	NO	NO	NO	NO
Acinetobacter_sp_ADP1	Gammaproteobacteria	NO	NO	NO	NO
Actinobacillus_pleuropneumoniae_L20	Gammaproteobacteria	NO	NO	NO	NO
Actinobacillus_pleuropneumoniae_serovar_3_JL03	Gammaproteobacteria	NO	NO	NO	NO
Actinobacillus_pleuropneumoniae_serovar_7_AP76	Gammaproteobacteria	NO	NO	NO	NO
Actinobacillus_succinogenes_130Z	Gammaproteobacteria	NO	NO	NO	NO
Actinosynnema_mirum_DSM_43827	Actinobacteria	YES	YES	NO	YES
Aeromonas_hydrophila_ATCC_7966	Gammaproteobacteria	NO	YES	YES	NO
Aeromonas_salmonicida_A449	Gammaproteobacteria	NO	YES	YES	NO
Aeropyrum_pernix	Crenarchaeota	YES	YES	YES	NO
Aggregatibacter_actinomycetemcomitans_D11S_1	Gammaproteobacteria	NO	YES	NO	NO
Aggregatibacter_aphrophilus_NJ8700	Gammaproteobacteria	NO	YES	NO	NO
Agrobacterium_radiobacter_K84	Alphaproteobacteria	YES	YES	YES	YES
Agrobacterium_tumefaciens_C58_Cereon	Alphaproteobacteria	YES	YES	YES	NO
Agrobacterium_vitis_S4	Alphaproteobacteria	YES	YES	YES	YES
Akkermansia_muciniphila_ATCC_BAA_835	Other_Bacteria	NO	NO	NO	NO
Alcanivorax_borkumensis_SK2	Gammaproteobacteria	YES	YES	YES	YES
Alicyclobacillus_acidocaldarius_DSM_446	Firmicutes	NO	NO	NO	YES
Aliivibrio_salmonicida_LFI1238	Gammaproteobacteria	NO	YES	YES	NO
Alkalilimnicola_ehrlichei_MLHE-1	Gammaproteobacteria	YES	YES	YES	NO
Alkaliphilus_metalliredigens_QYMF	Firmicutes	NO	NO	NO	NO
Alkaliphilus_oremlandii_OhILAs	Firmicutes	NO	NO	NO	NO
Allochromatium_vinosum_DSM_180_uid46083	Gammaproteobacteria	NO	YES	YES	NO
Alteromonas_macleodii__Deep_ecotype_	Gammaproteobacteria	YES	YES	YES	YES
Aminobacterium_colombiense_DSM_12261_uid47083	Other_Bacteria	NO	NO	NO	NO

Ammonifex_degensii_KC4	Firmicutes	NO	NO	NO	NO
Amycolatopsis_mediterranei_U32_uid50565	Other_Bacteria	YES	YES	NO	NO
Anabaena_variabilis_ATCC_29413	Cyanobacteria	YES	NO	NO	YES
Anaerocellum_thermophilum_DSM_6725	Firmicutes	NO	NO	NO	NO
Anaerococcus_prevotii_DSM_20548	Firmicutes	NO	NO	NO	NO
Anaeromyxobacter_dehalogenans_2CP_1	Deltaproteobacteria	YES	NO	YES	YES
Anaeromyxobacter_dehalogenans_2CP-C	Deltaproteobacteria	YES	NO	YES	YES
Anaeromyxobacter_Fw109-5	Deltaproteobacteria	YES	NO	YES	YES
Anaeromyxobacter_K	Deltaproteobacteria	YES	NO	YES	YES
Anaplasma_centrale_Israel	Alphaproteobacteria	YES	NO	YES	NO
Anaplasma_marginale_Florida	Alphaproteobacteria	YES	NO	YES	NO
Anaplasma_marginale_St_Maries	Alphaproteobacteria	YES	NO	YES	NO
Anaplasma_phagocytophilum_HZ	Alphaproteobacteria	YES	NO	YES	NO
Anoxybacillus_flavithermus_WK1	Firmicutes	YES	NO	YES	NO
Aquifex_aeolicus	Aquificae	YES	YES	YES	NO
Arcanobacterium_haemolyticum_DSM_20595_uid49489	Actinobacteria	NO	NO	NO	NO
Archaeoglobus_fulgidus	Euryarchaeota	YES	NO	NO	NO
Archaeoglobus_profundus_DSM_5631_uid43493	Euryarchaeota	NO	NO	NO	NO
Arcobacter_butzleri_RM4018	Epsilonproteobacteria	NO	NO	NO	NO
Arcobacter_nitrofigilis_DSM_7299_uid49001	Epsilonproteobacteria	NO	YES	YES	NO
Aromatoleum_aromaticum_EbN1	Betaproteobacteria	YES	YES	YES	YES
Arthrobacter_aurescens_TC1	Actinobacteria	YES	NO	NO	NO
Arthrobacter_chlorophenolicus_A6	Actinobacteria	YES	NO	NO	NO
Arthrobacter_FB24	Actinobacteria	YES	NO	NO	NO
Aster_yellows_witches-broom_phytoplasma_AYWB	Firmicutes	NO	NO	NO	NO
Atopobium_parvulum_DSM_20469	Actinobacteria	NO	NO	NO	NO
Azoarcus_BH72	Betaproteobacteria	YES	YES	YES	YES
Azorhizobium_caulinodans_ORS_571	Alphaproteobacteria	YES	YES	YES	NO
Azospirillum_B510_uid46085	Alphaproteobacteria	YES	YES	YES	YES
Azotobacter_vinelandii_DJ	Gammaproteobacteria	YES	YES	YES	NO
Bacillus_amyloliquefaciens_FZB42	Firmicutes	YES	NO	YES	NO
Bacillus_anthracis_A0248	Firmicutes	YES	NO	YES	YES
Bacillus_anthracis_Ames	Firmicutes	YES	NO	YES	YES
Bacillus_anthracis_Ames_0581	Firmicutes	YES	NO	YES	YES
Bacillus_anthracis_CDC_684	Firmicutes	YES	NO	YES	YES
Bacillus_anthracis_CI_uid50615	Firmicutes	YES	NO	YES	YES
Bacillus_anthracis_str_Sterne	Firmicutes	YES	NO	YES	YES
Bacillus_cereus_03BB102	Firmicutes	YES	NO	YES	YES
Bacillus_cereus_AH187	Firmicutes	YES	NO	YES	YES
Bacillus_cereus_AH820	Firmicutes	YES	NO	YES	YES
Bacillus_cereus_ATCC_10987	Firmicutes	YES	NO	YES	YES
Bacillus_cereus_ATCC14579	Firmicutes	YES	NO	YES	YES
Bacillus_cereus_B4264	Firmicutes	YES	NO	YES	YES
Bacillus_cereus_cytotoxis_NVH_391-98	Firmicutes	YES	NO	YES	NO
Bacillus_cereus_G9842	Firmicutes	YES	NO	YES	YES

Bacillus_cereus_Q1	Firmicutes	YES	NO	YES	NO
Bacillus_cereus_ZK	Firmicutes	YES	NO	YES	YES
Bacillus_clausii_KSM-K16	Firmicutes	YES	NO	YES	NO
Bacillus_halodurans	Firmicutes	YES	NO	YES	NO
Bacillus_licheniformis_ATCC_14580	Firmicutes	YES	NO	YES	NO
Bacillus_licheniformis_DSM_13	Firmicutes	YES	NO	YES	NO
Bacillus_megaterium_DSM319_uid48371	Firmicutes	YES	NO	YES	YES
Bacillus_megaterium_QM_B1551_uid15862	Firmicutes	YES	NO	YES	YES
Bacillus_pseudofirmus_OF4_uid45847	Firmicutes	YES	NO	YES	YES
Bacillus_pumilus_SAFR-032	Firmicutes	YES	NO	YES	NO
Bacillus_selenitireducens_MLS10_uid49513	Firmicutes	YES	NO	YES	NO
Bacillus_subtilis	Firmicutes	YES	NO	YES	NO
Bacillus_thuringiensis_Al_Hakam	Firmicutes	YES	NO	YES	NO
Bacillus_thuringiensis_BMB171_uid49135	Firmicutes	YES	NO	YES	NO
Bacillus_thuringiensis_konkukian	Firmicutes	YES	NO	YES	YES
Bacillus_tusciae_DSM_2912_uid48361	Firmicutes	YES	NO	YES	NO
Bacillus_weihenstephanensis_KBAB4	Firmicutes	YES	NO	YES	YES
Bacteroides_fragilis_NCTC_9434	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Bacteroides_fragilis_YCH46	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Bacteroides_thetaiotaomicron_VPI-5482	Bacteroidetes/Chlorobi	NO	NO	YES	NO
Bacteroides_vulgatus_ATCC_8482	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Bartonella_bacilliformis_KC583	Alphaproteobacteria	NO	YES	YES	NO
Bartonella_grahamii_as4aup	Alphaproteobacteria	NO	YES	YES	NO
Bartonella_henselae_Houston-1	Alphaproteobacteria	NO	YES	YES	NO
Bartonella_quintana_Toulouse	Alphaproteobacteria	NO	YES	YES	NO
Bartonella_tribocorum_CIP_105476	Alphaproteobacteria	NO	YES	YES	NO
Baumannia_cicadellinicola_Homalodisca_coagulata	Gammaproteobacteria	NO	NO	NO	NO
Bdellovibrio_bacteriovorus	Deltaproteobacteria	YES	YES	YES	NO
Beijerinckia_indica_ATCC_9039	Alphaproteobacteria	YES	YES	YES	NO
Beutenbergia_cavernae_DSM_12333	Actinobacteria	YES	NO	NO	NO
Bifidobacterium_adolescentis_ATCC_15703	Actinobacteria	NO	NO	NO	NO
Bifidobacterium_animalis_lactis_AD011	Actinobacteria	NO	NO	NO	NO
Bifidobacterium_animalis_lactis_Bl_04	Actinobacteria	NO	NO	NO	NO
Bifidobacterium_animalis_lactis_DSM_10140	Actinobacteria	NO	NO	NO	NO
Bifidobacterium_dentium_Bd1	Actinobacteria	NO	NO	NO	NO
Bifidobacterium_longum	Actinobacteria	NO	NO	NO	NO
Bifidobacterium_longum_DJO10A	Actinobacteria	NO	NO	NO	NO
Bifidobacterium_longum_infantis_ATCC_15697	Actinobacteria	NO	NO	NO	NO
Bifidobacterium_longum_JDM301_uid49131	Actinobacteria	NO	NO	NO	NO
Blattabacterium__Blattella_germanica__Bge	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Blattabacterium__Periplaneta_americana__BPLAN	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Bordetella_avium_197N	Betaproteobacteria	NO	YES	YES	NO
Bordetella_bronchiseptica	Betaproteobacteria	YES	YES	YES	YES
Bordetella_parapertussis	Betaproteobacteria	YES	YES	YES	YES
Bordetella_pertussis	Betaproteobacteria	YES	YES	YES	YES

Bordetella_petrii	Betaproteobacteria	YES	YES	YES	YES
Borrelia_afzelii_PKo	Spirochaetes	NO	NO	NO	NO
Borrelia_burgdorferi	Spirochaetes	NO	NO	NO	NO
Borrelia_burgdorferi_ZS7	Spirochaetes	NO	NO	NO	NO
Borrelia_duttonii_Ly	Spirochaetes	NO	NO	NO	NO
Borrelia_garinii_PBi	Spirochaetes	NO	NO	NO	NO
Borrelia_hermsii_DAH	Spirochaetes	NO	NO	NO	NO
Borrelia_recurrentis_A1	Spirochaetes	NO	NO	NO	NO
Borrelia_turicatae_91E135	Spirochaetes	NO	NO	NO	NO
Brachybacterium_faecium_DSM_4810	Actinobacteria	YES	YES	NO	YES
Brachyspira_hyodysenteriae_WA1	Spirochaetes	NO	NO	NO	NO
Brachyspira_murdochii_DSM_12563_uid48819	Spirochaetes	NO	NO	NO	NO
Brachyspira_pilosicoli_95_1000_uid50609	Spirochaetes	NO	NO	NO	NO
Bradyrhizobium_BTAi1	Alphaproteobacteria	YES	YES	YES	YES
Bradyrhizobium_japonicum	Alphaproteobacteria	YES	YES	YES	YES
Bradyrhizobium_ORS278	Alphaproteobacteria	YES	YES	YES	NO
Brevibacillus_brevis_NBRC_100599	Firmicutes	YES	NO	YES	YES
Brucella_abortus_bv__1_9_941	Alphaproteobacteria	YES	YES	YES	NO
Brucella_abortus_S19	Alphaproteobacteria	YES	YES	YES	NO
Brucella_canis_ATCC_23365	Alphaproteobacteria	YES	YES	YES	NO
Brucella_melitensis	Alphaproteobacteria	YES	YES	YES	NO
Brucella_melitensis_ATCC_23457	Alphaproteobacteria	YES	YES	YES	NO
Brucella_melitensis_biovar_Abortus	Alphaproteobacteria	YES	YES	YES	NO
Brucella_microti_CCM_4915	Alphaproteobacteria	YES	YES	YES	NO
Brucella_ovis	Alphaproteobacteria	YES	YES	YES	NO
Brucella_suis_1330	Alphaproteobacteria	YES	YES	YES	NO
Brucella_suis_ATCC_23445	Alphaproteobacteria	YES	YES	YES	NO
Buchnera_aphidicola	Gammaproteobacteria	NO	NO	NO	NO
Buchnera_aphidicola_5A__Acyrthosiphon_pisum_	Gammaproteobacteria	NO	NO	NO	NO
Buchnera_aphidicola_Cc_Cinara_cedri	Gammaproteobacteria	NO	NO	NO	NO
Buchnera_aphidicola_Sg	Gammaproteobacteria	NO	NO	NO	NO
Buchnera_aphidicola_Tuc7__Acyrthosiphon_pisum_	Gammaproteobacteria	NO	NO	NO	NO
Buchnera_sp	Gammaproteobacteria	NO	NO	NO	NO
Burkholderia_383	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_ambifaria_AMMD	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_ambifaria_MC40_6	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_CCGE1002_uid42523	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_cenocepacia_AU_1054	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_cenocepacia_HI2424	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_cenocepacia_J2315	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_cenocepacia_MC0_3	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_glumae_BGR1	Betaproteobacteria	YES	YES	YES	NO
Burkholderia_mallei_ATCC_23344	Betaproteobacteria	YES	YES	YES	NO
Burkholderia_mallei_NCTC_10229	Betaproteobacteria	YES	YES	YES	NO
Burkholderia_mallei_NCTC_10247	Betaproteobacteria	YES	YES	YES	NO

Burkholderia_mallei_SAVP1	Betaproteobacteria	YES	YES	YES	NO
Burkholderia_multivorans_ATCC_17616_JGI	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_multivorans_ATCC_17616_Tohoku	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_phymatum_STM815	Betaproteobacteria	YES	YES	YES	NO
Burkholderia_phytofirmans_PsJN	Betaproteobacteria	YES	YES	YES	NO
Burkholderia_pseudomallei_1106a	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_pseudomallei_1710b	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_pseudomallei_668	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_pseudomallei_K96243	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_pseudomallei_MSHR346	Betaproteobacteria	YES	YES	YES	NO
Burkholderia_thailandensis_E264	Betaproteobacteria	YES	YES	YES	NO
Burkholderia_vietnamiensis_G4	Betaproteobacteria	YES	YES	YES	YES
Burkholderia_xenovorans_LB400	Betaproteobacteria	YES	YES	YES	NO
Caldicellulosiruptor_saccharolyticus_DSM_8903	Other_Bacteria	NO	NO	NO	NO
Caldivirga_maquilingensis_IC-167	Crenarchaeota	YES	NO	NO	NO
Campylobacter_concisus_13826	Epsilonproteobacteria	YES	YES	YES	NO
Campylobacter_curvus_525_92	Epsilonproteobacteria	YES	YES	YES	NO
Campylobacter_fetus_82-40	Epsilonproteobacteria	YES	YES	YES	NO
Campylobacter_hominis_ATCC_BAA-381	Epsilonproteobacteria	NO	NO	NO	NO
Campylobacter_jejuni	Epsilonproteobacteria	NO	YES	YES	NO
Campylobacter_jejuni_81116	Epsilonproteobacteria	NO	YES	YES	NO
Campylobacter_jejuni_81-176	Epsilonproteobacteria	NO	YES	YES	NO
Campylobacter_jejuni_doylei_269_97	Epsilonproteobacteria	NO	YES	YES	NO
Campylobacter_jejuni_RM1221	Epsilonproteobacteria	NO	YES	YES	NO
Campylobacter_lari_RM2100	Epsilonproteobacteria	NO	YES	YES	NO
Candidatus_Accumulibacter_phosphatis_clade_IIA_UW_1	Betaproteobacteria	YES	YES	YES	NO
Candidatus_Amoebophilus_asiaticus_5a2	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Candidatus_Azobacteroides_pseudotrichonymphae_genomova	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Candidatus_Blochmannia_floridanus	Gammaproteobacteria	NO	NO	NO	NO
Candidatus_Blochmannia_pennsylvanicus_BPEN	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Candidatus_Carsonella_ruddii_PV	Gammaproteobacteria	NO	NO	NO	NO
Candidatus_Desulfococcus_oleovorans_Hxd3	Deltaproteobacteria	NO	NO	NO	NO
Candidatus_Desulforudis_audaxviator_MP104C	Firmicutes	NO	NO	NO	NO
Candidatus_Hamiltonella_defensa_5AT__Acyrthosiphon_pisum	Gammaproteobacteria	NO	NO	NO	NO
Candidatus_Hodgkinia_cicadicola_Dsem	Alphaproteobacteria	YES	NO	NO	NO
Candidatus_Korarchaeum_cryptofilum_OPF8	Other_Archaea	NO	NO	NO	NO
Candidatus_Koribacter_versatilis_Ellin345	Acidobacteria	YES	NO	YES	NO
Candidatus_Liberibacter_asiaticus_psy62	Alphaproteobacteria	NO	NO	YES	NO
Candidatus_Methanoregula_boonei_6A8	Euryarchaeota	NO	NO	NO	NO
Candidatus_Methanosphaerula_palustris_E1_9c	Euryarchaeota	NO	NO	NO	YES
Candidatus_Pelagibacter_ubique_HTCC1062	Alphaproteobacteria	YES	NO	NO	NO
Candidatus_Phytoplasma_australiense	Firmicutes	NO	NO	NO	NO
Candidatus_Phytoplasma_mali	Other_Bacteria	NO	NO	NO	NO
Candidatus_Puniceispirillum_marinum_IMCC1322_uid47081	Alphaproteobacteria	YES	YES	YES	NO
Candidatus_Riesia_pediculicola_USDA_uid46841	Gammaproteobacteria	NO	NO	NO	NO

Candidatus_Ruthia_magnifica_Cm_Calyptogena_magnifica_	Gammaproteobacteria	YES	YES	YES	NO
Candidatus_Sulcia_muelleri_DMIN_uid47075	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Candidatus_Sulcia_muelleri_GWSS	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Candidatus_Sulcia_muelleri_SMDSEM	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Candidatus_Vesicomyosocius_okutanii_HA	Gammaproteobacteria	YES	YES	YES	NO
Capnocytophaga_ochracea_DSM_7271	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Carboxydothermus_hydrogenoformans_Z-2901	Firmicutes	NO	NO	NO	NO
Catenulispora_acidiphila_DSM_44928	Actinobacteria	YES	YES	YES	NO
Caulobacter_crescentus	Alphaproteobacteria	YES	YES	YES	NO
Caulobacter_crescentus_NA1000	Alphaproteobacteria	YES	YES	YES	NO
Caulobacter_K31	Alphaproteobacteria	YES	YES	YES	NO
Caulobacter_segnis_ATCC_21756_uid41709	Alphaproteobacteria	YES	YES	YES	NO
Cellulomonas_flavigena_DSM_20109_uid48821	Actinobacteria	YES	NO	NO	YES
Cellvibrio_japonicus_Ueda107	Gammaproteobacteria	YES	YES	YES	YES
Chitinophaga_pinensis_DSM_2588	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Chlamydia_muridarum	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydia_trachomatis_434_Bu	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydia_trachomatis_A_HAR-13	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydia_trachomatis_B_Jali20_OT	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydia_trachomatis_B_TZ1A828_OT	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydia_trachomatis_D_UW_3_CX	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydia_trachomatis_L2b_UCH_1_proctitis	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydophila_abortus_S26_3	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydophila_caviae	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydophila_felis_Fe_C-56	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydophila_pneumoniae_AR39	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydophila_pneumoniae_CWL029	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydophila_pneumoniae_J138	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlamydophila_pneumoniae_TW_183	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Chlorobaculum_parvum_NCIB_8327	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Chlorobium_chlorochromatii_CaD3	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Chlorobium_limicola_DSM_245	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Chlorobium_luteolum_DSM_273	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Chlorobium_phaeobacteroides_BS1	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Chlorobium_phaeobacteroides_DSM_266	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Chlorobium_tepidum_TLS	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Chloroflexus_aggregans_DSM_9485	Chloroflexi	YES	YES	YES	YES
Chloroflexus_aurantiacus_J_10_fl	Chloroflexi	YES	YES	YES	YES
Chloroflexus_Y_400_fl	Chloroflexi	YES	YES	YES	YES
Chloroherpeton_thalassium_ATCC_35110	Bacteroidetes/Chlorobi	NO	NO	YES	NO
Chromobacterium_violaceum	Betaproteobacteria	YES	YES	YES	NO
Chromohalobacter_salexigens_DSM_3043	Gammaproteobacteria	NO	NO	NO	NO
Citrobacter_koseri_ATCC_BAA-895	Gammaproteobacteria	NO	NO	NO	NO
Citrobacter_rodentium_ICC168	Gammaproteobacteria	NO	NO	NO	NO
Clavibacter_michiganensis_NCPPB_382	Actinobacteria	YES	NO	NO	NO

Clavibacter_michiganensis_sepedonicus	Actinobacteria	YES	NO	NO	NO
Clostridiales_genomosp__BVAB3_UPII9_5_uid46219	Firmicutes	NO	NO	NO	NO
Clostridium_acetobutylicum	Firmicutes	NO	NO	NO	NO
Clostridium_beijerinckii_NCIMB_8052	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_A	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_A_ATCC_19397	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_A_Hall	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_A2_Kyoto	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_A3_Loch_Maree	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_B_Eklund_17B	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_B1_Okra	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_Ba4_657	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_E3_Alaska_E43	Firmicutes	NO	NO	NO	NO
Clostridium_botulinum_F_Langeland	Firmicutes	NO	NO	NO	NO
Clostridium_cellulolyticum_H10	Firmicutes	NO	NO	NO	NO
Clostridium_difficile_630	Firmicutes	NO	NO	NO	NO
Clostridium_difficile_CD196	Firmicutes	NO	NO	NO	NO
Clostridium_difficile_R20291	Firmicutes	NO	NO	NO	NO
Clostridium_kluyveri_DSM_555	Firmicutes	NO	NO	NO	NO
Clostridium_kluyveri_NBRC_12016	Firmicutes	NO	NO	NO	NO
Clostridium_ljungdahlii_ATCC_49587_uid50583	Firmicutes	NO	NO	NO	NO
Clostridium_novyi_NT	Firmicutes	NO	NO	NO	NO
Clostridium_perfringens	Firmicutes	NO	NO	NO	NO
Clostridium_perfringens_ATCC_13124	Firmicutes	NO	NO	NO	NO
Clostridium_perfringens_SM101_uid12521	Firmicutes	NO	NO	NO	NO
Clostridium_phytofermentans_ISDg	Firmicutes	NO	NO	NO	YES
Clostridium_tetani_E88	Firmicutes	NO	NO	NO	NO
Clostridium_thermocellum_ATCC_27405	Firmicutes	NO	NO	NO	NO
Colwellia_psychrerythraea_34H	Gammaproteobacteria	YES	YES	YES	NO
Comamonas_testosteroni_CNB_1_uid29203	Betaproteobacteria	YES	YES	YES	YES
Conexibacter_woesei_DSM_14684_uid43467	Actinobacteria	YES	NO	YES	YES
Coprothermobacter_proteolyticus_DSM_5265	Firmicutes	NO	NO	NO	NO
Coraliomargarita_akajimensis_DSM_45221_uid47079	Other_Bacteria	YES	NO	YES	NO
Corynebacterium_aurimucosum_ATCC_700975	Actinobacteria	YES	YES	NO	YES
Corynebacterium_diphtheriae	Actinobacteria	YES	YES	NO	YES
Corynebacterium_efficiens_YS-314	Actinobacteria	YES	YES	NO	YES
Corynebacterium_glutamicum_ATCC_13032_Bielefeld	Actinobacteria	YES	YES	NO	YES
Corynebacterium_glutamicum_ATCC_13032_Kitasato	Actinobacteria	YES	YES	NO	YES
Corynebacterium_glutamicum_R	Actinobacteria	YES	NO	NO	YES
Corynebacterium_jeikeium_K411	Actinobacteria	YES	YES	NO	NO
Corynebacterium_kroppenstedtii_DSM_44385	Actinobacteria	YES	YES	NO	YES
Corynebacterium_pseudotuberculosis_uid50585	Actinobacteria	YES	YES	NO	YES
Corynebacterium_urealyticum_DSM_7109	Actinobacteria	YES	YES	NO	YES
Coxiella_burnetii	Gammaproteobacteria	NO	NO	NO	NO
Coxiella_burnetii_CbuG_Q212	Gammaproteobacteria	NO	NO	NO	NO

Coxiella_burnetii_CbuK_Q154	Gammaproteobacteria	NO	NO	NO	NO
Coxiella_burnetii_Dugway_7E9-12	Gammaproteobacteria	NO	NO	NO	NO
Coxiella_burnetii_RSA_331	Gammaproteobacteria	NO	NO	NO	NO
Croceibacter_atlanticus_HTCC2559_uid49661	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Cronobacter_turicensis_z3032_uid40821	Gammaproteobacteria	NO	NO	NO	NO
Cryptobacterium_curtum_DSM_15641	Actinobacteria	NO	NO	NO	NO
Cupriavidus_metallidurans_CH34_uid250	Betaproteobacteria	YES	YES	YES	YES
Cupriavidus_taiwanensis	Betaproteobacteria	YES	YES	YES	NO
Cyanobacteria_bacterium_Yellowstone_A-Prime	Cyanobacteria	YES	NO	NO	YES
Cyanobacteria_bacterium_Yellowstone_B-Prime	Cyanobacteria	YES	NO	NO	YES
Cyanothece_ATCC_51142	Cyanobacteria	YES	NO	NO	YES
Cyanothece_PCC_7424	Cyanobacteria	YES	NO	NO	YES
Cyanothece_PCC_7425	Cyanobacteria	YES	NO	NO	YES
Cyanothece_PCC_8801	Cyanobacteria	YES	NO	NO	NO
Cyanothece_PCC_8802	Cyanobacteria	YES	NO	NO	NO
Cytophaga_hutchinsonii_ATCC_33406	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Dechloromonas_aromatica_RCB	Betaproteobacteria	YES	YES	YES	NO
Deferribacter_desulfuricans_SSM1_uid46653	Other_Bacteria	NO	NO	NO	NO
Dehalococcoides_BAV1	Chloroflexi	NO	NO	NO	NO
Dehalococcoides_CBDB1	Chloroflexi	NO	NO	NO	NO
Dehalococcoides_ethenogenes_195	Chloroflexi	NO	NO	NO	NO
Dehalococcoides_GT_uid42115	Chloroflexi	NO	NO	NO	NO
Dehalococcoides_VS	Chloroflexi	NO	NO	NO	NO
Dehalogenimonas_lykanthroporepellens_BL_DC_9_uid48131	Chloroflexi	NO	NO	NO	NO
Deinococcus_deserti_VCD115	Deinococcus-Thermus	YES	YES	YES	YES
Deinococcus_geothermalis_DSM_11300	Deinococcus-Thermus	YES	YES	YES	NO
Deinococcus_radiodurans	Deinococcus-Thermus	YES	YES	YES	NO
Delftia_acidovorans_SPH-1	Betaproteobacteria	YES	YES	YES	YES
Denitrovibrio_acetiphilus_DSM_12809_uid46657	Other_Bacteria	YES	NO	NO	NO
Desulfatibacillum_alkenivorans_AK_01	Firmicutes	NO	NO	NO	NO
Desulfitobacterium_hafniense_DCB_2	Firmicutes	YES	NO	NO	NO
Desulfitobacterium_hafniense_Y51	Firmicutes	YES	NO	NO	NO
Desulfobacterium_autotrophicum_HRM2	Deltaproteobacteria	NO	NO	NO	NO
Desulfohalobium_retbaense_DSM_5692	Deltaproteobacteria	NO	NO	NO	NO
Desulfomicrobium_baculatum_DSM_4028	Deltaproteobacteria	YES	NO	NO	YES
Desulfotalea_psychrophila_LSv54	Deltaproteobacteria	NO	NO	NO	NO
Desulfotomaculum_acetoxidans_DSM_771	Firmicutes	NO	NO	NO	NO
Desulfotomaculum_reducens_MI-1	Firmicutes	NO	NO	NO	NO
Desulfovibrio_desulfuricans_ATCC_27774	Deltaproteobacteria	NO	NO	NO	NO
Desulfovibrio_desulfuricans_G20	Deltaproteobacteria	YES	NO	NO	NO
Desulfovibrio_magneticus_RS_1	Deltaproteobacteria	NO	NO	NO	NO
Desulfovibrio_salexigens_DSM_2638	Deltaproteobacteria	NO	NO	NO	NO
Desulfovibrio_vulgaris__Miyazaki_F_	Deltaproteobacteria	YES	NO	NO	NO
Desulfovibrio_vulgaris_DP4	Deltaproteobacteria	YES	NO	NO	NO
Desulfovibrio_vulgaris_Hildenborough	Deltaproteobacteria	YES	NO	NO	NO

Desulfurivibrio_alkaliphilus_AHT2_uid49487	Deltaproteobacteria	NO	NO	NO	YES
Desulfurococcus_kamchatkensis_1221n	Crenarchaeota	NO	NO	NO	NO
Diaphorobacter_TPSY_uid29975	Other_Bacteria	YES	YES	YES	YES
Dichelobacter_nodosus_VCS1703A	Gammaproteobacteria	NO	NO	NO	NO
Dickeya_dadantii_Ech586	Gammaproteobacteria	NO	NO	NO	NO
Dickeya_dadantii_Ech703	Gammaproteobacteria	NO	NO	NO	NO
Dickeya_zeae_Ech1591	Gammaproteobacteria	NO	NO	NO	NO
Dictyoglomus_thermophilum_H_6_12	Other_Bacteria	NO	NO	NO	NO
Dictyoglomus_turgidum_DSM_6724	Other_Bacteria	NO	NO	NO	NO
Dinoroseobacter_shibae_DFL_12	Alphaproteobacteria	YES	YES	YES	NO
Dyadobacter_fermentans_DSM_18053	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Edwardsiella_ictaluri_93_146	Gammaproteobacteria	NO	NO	NO	NO
Edwardsiella_tarda_EIB202	Gammaproteobacteria	NO	NO	NO	NO
Eggerthella_lenta_DSM_2243	Actinobacteria	NO	NO	NO	NO
Ehrlichia_canis_Jake	Alphaproteobacteria	YES	NO	YES	NO
Ehrlichia_chaffeensis_Arkansas	Alphaproteobacteria	YES	NO	YES	NO
Ehrlichia_ruminantium_Gardel	Alphaproteobacteria	YES	NO	YES	NO
Ehrlichia_ruminantium_str._Welgevonden_CIRAD	Alphaproteobacteria	YES	NO	YES	NO
Ehrlichia_ruminantium_Welgevonden_UPSA	Alphaproteobacteria	YES	NO	YES	NO
Elusimicrobium_minutum_Pei191	Other_Bacteria	NO	NO	NO	NO
Enterobacter_638	Firmicutes	NO	NO	NO	NO
Enterobacter_cloacae_ATCC_13047_uid48363	Firmicutes	NO	NO	NO	NO
Enterobacter_sakazakii_ATCC_BAA-894	Firmicutes	NO	NO	NO	NO
Enterococcus_faecalis_V583	Firmicutes	NO	NO	NO	NO
Erwinia_amylovora_ATCC_49946_uid46943	Gammaproteobacteria	NO	NO	NO	NO
Erwinia_amylovora_CFBP1430_uid46839	Gammaproteobacteria	NO	NO	NO	NO
Erwinia_billingiae_Eb661_uid50547	Gammaproteobacteria	NO	NO	NO	NO
Erwinia_carotovora_atroseptica_SCRI1043	Gammaproteobacteria	NO	NO	NO	NO
Erwinia_pyrifoliae_Ep1_96_uid40659	Gammaproteobacteria	NO	NO	NO	NO
Erwinia_tasmaniensis	Gammaproteobacteria	NO	NO	NO	NO
Erythrobacter_litoralis_HTCC2594	Alphaproteobacteria	YES	YES	YES	NO
Escherichia_coli__BL21_Gold_DE3_pLysS_AG__uid30681	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_0127_H6_E2348_69	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_536	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_55989	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_APEC_O1	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_B_REL606	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_BW2952	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_C_ATCC_8739	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_CFT073	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_E24377A	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_ED1a	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_HS	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_IAI1	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_IAI39	Gammaproteobacteria	NO	NO	NO	NO

Escherichia_coli_K_12_substr__DH10B	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_K_12_substr__MG1655	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_K_12_substr__W3110	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_O103_H2_12009	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_O111_H__11128	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_O157_H7_EC4115	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_O157_H7_TW14359	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_O157H7	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_O157H7_EDL933	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_O26_H11_11368	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_O55_H7_CB9615_uid46655	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_S88	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_SE11	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_SMS_3_5	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_UMN026	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_coli_UTI89	Gammaproteobacteria	NO	NO	NO	NO
Escherichia_fergusonii_ATCC_35469	Gammaproteobacteria	NO	NO	NO	NO
Eubacterium_eligens_ATCC_27750	Firmicutes	NO	NO	NO	NO
Eubacterium_rectale_ATCC_33656	Firmicutes	NO	NO	NO	NO
Exiguobacterium_AT1b	Firmicutes	YES	NO	YES	NO
Exiguobacterium_sibiricum_255_15	Firmicutes	YES	NO	YES	NO
Ferroglobus_placidus_DSM_10642_uid40863	Euryarchaeota	YES	NO	NO	NO
Fervidobacterium_nodosum_Rt17-B1	Thermotogae	NO	NO	NO	NO
Fibrobacter_succinogenes_S85	Other_Bacteria	NO	NO	NO	NO
Finegoldia_magna_ATCC_29328	Firmicutes	NO	NO	NO	NO
Flavobacteriaceae_bacterium_3519_10	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Flavobacterium_johnsoniae_UW101	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Flavobacterium_psychrophilum_JIP02_86	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Francisella_philomiragia_ATCC_25017	Gammaproteobacteria	NO	YES	YES	NO
Francisella_tularensis_FSC_198	Gammaproteobacteria	NO	NO	NO	NO
Francisella_tularensis_holarctica	Gammaproteobacteria	NO	NO	NO	NO
Francisella_tularensis_holarctica_FTNF002_00	Gammaproteobacteria	NO	NO	NO	NO
Francisella_tularensis_holarctica_OSU18	Gammaproteobacteria	NO	NO	NO	NO
Francisella_tularensis_mediasiatica_FSC147	Gammaproteobacteria	NO	NO	NO	NO
Francisella_tularensis_novicida_U112	Gammaproteobacteria	NO	NO	NO	NO
Francisella_tularensis_tularensis	Gammaproteobacteria	NO	NO	NO	NO
Francisella_tularensis_WY96-3418	Gammaproteobacteria	NO	NO	NO	NO
Frankia_alni_ACN14a	Actinobacteria	YES	YES	YES	NO
Frankia_CcI3	Actinobacteria	YES	YES	YES	NO
Frankia_EAN1pec	Actinobacteria	YES	YES	YES	YES
Fusobacterium_nucleatum	Fusobacteria	NO	NO	NO	NO
Gardnerella_vaginalis_409_05_uid43211	Actinobacteria	NO	NO	NO	NO
Gemmatimonas_aurantiaca_T_27	Other_Bacteria	YES	YES	YES	NO
Geobacillus_C56_T3_uid49467	Firmicutes	YES	NO	YES	NO
Geobacillus_kaustophilus_HTA426	Firmicutes	YES	NO	YES	YES

Geobacillus_thermodenitrificans_NG80-2	Firmicutes	YES	NO	YES	YES
Geobacillus_WCH70	Firmicutes	YES	NO	YES	YES
Geobacillus_Y412MC10	Firmicutes	YES	NO	YES	NO
Geobacillus_Y412MC61	Firmicutes	YES	NO	YES	NO
Geobacter_bemidjiensis_Bem	Deltaproteobacteria	YES	NO	NO	NO
Geobacter_FRC_32	Deltaproteobacteria	NO	NO	NO	YES
Geobacter_lovleyi_SZ	Deltaproteobacteria	NO	NO	YES	NO
Geobacter_M21	Deltaproteobacteria	YES	NO	NO	NO
Geobacter_metallireducens_GS-15	Deltaproteobacteria	YES	NO	NO	YES
Geobacter_sulfurreducens	Deltaproteobacteria	YES	NO	YES	YES
Geobacter_uraniumreducens_Rf4	Deltaproteobacteria	YES	NO	NO	YES
Geodermatophilus_obscurus_DSM_43160_uid43725	Actinobacteria	YES	YES	YES	NO
Gloeobacter_violaceus	Cyanobacteria	YES	NO	NO	NO
Gluconacetobacter_diazotrophicus_PAl_5_FAPERJ	Alphaproteobacteria	YES	YES	YES	YES
Gluconacetobacter_diazotrophicus_PAl_5_JGI	Alphaproteobacteria	YES	YES	YES	YES
Gluconobacter_oxydans_621H	Alphaproteobacteria	NO	YES	YES	NO
Gordonia_bronchialis_DSM_43247	Actinobacteria	YES	NO	NO	NO
Gramella_forsetii_KT0803	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Granulobacter_bethesdensis_CGDNIH1	Gammaproteobacteria	YES	YES	YES	NO
Haemophilus_ducreyi_35000HP	Gammaproteobacteria	NO	NO	NO	NO
Haemophilus_influenzae	Gammaproteobacteria	NO	NO	NO	NO
Haemophilus_influenzae_86_028NP	Gammaproteobacteria	NO	NO	NO	NO
Haemophilus_influenzae_PittEE	Gammaproteobacteria	NO	NO	NO	NO
Haemophilus_influenzae_PittGG	Gammaproteobacteria	NO	NO	NO	NO
Haemophilus_parasuis_SH0165	Gammaproteobacteria	NO	NO	NO	NO
Haemophilus_somnus_129PT	Gammaproteobacteria	NO	NO	NO	NO
Haemophilus_somnus_2336	Gammaproteobacteria	NO	NO	NO	NO
Hahella_chejuensis_KCTC_2396	Gammaproteobacteria	YES	YES	YES	YES
Halalkalicoccus_jeotgali_B3_uid50305	Euryarchaeota	YES	NO	YES	NO
Haliangium_ochraceum_DSM_14365	Deltaproteobacteria	YES	NO	YES	NO
Haloarcula_marismortui_ATCC_43049	Euryarchaeota	YES	NO	YES	NO
Halobacterium_salinarum_R1	Euryarchaeota	YES	NO	YES	NO
Halobacterium_sp	Euryarchaeota	YES	NO	YES	NO
Haloferax_volcanii_DS2_uid46845	Euryarchaeota	YES	NO	YES	NO
Halomicrobium_mukohataei_DSM_12286	Euryarchaeota	YES	NO	NO	NO
Haloquadratum_walsbyi	Euryarchaeota	YES	NO	YES	NO
Halorhabdus_utahensis_DSM_12940	Euryarchaeota	YES	NO	YES	YES
Halorhodospira_halophila_SL1	Gammaproteobacteria	NO	YES	YES	NO
Halorubrum_lacusprofundi_ATCC_49239	Euryarchaeota	YES	NO	YES	NO
Haloterrigena_turkmenica_DSM_5511_uid43501	Euryarchaeota	YES	NO	YES	NO
Halothermothrix_orenii_H_168	Firmicutes	NO	NO	NO	NO
Halothiobacillus_neapolitanus_c2	Gammaproteobacteria	YES	YES	YES	NO
Helicobacter_acinonychis_Sheeba	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_hepaticus	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_mustelae_12198_uid46647	Epsilonproteobacteria	NO	NO	NO	NO

Helicobacter_pylori_26695	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_pylori_B38_uid39685	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_pylori_B8_uid49873	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_pylori_G27	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_pylori_HPAG1	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_pylori_J99	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_pylori_P12	Epsilonproteobacteria	NO	NO	NO	NO
Helicobacter_pylori_Shi470	Epsilonproteobacteria	NO	NO	NO	NO
Heliobacterium_modesticaldum_Ice1	Firmicutes	NO	NO	NO	NO
Herbaspirillum_seropedicae_SmR1_uid50427	Betaproteobacteria	YES	YES	YES	NO
Herminiimonas_arsenicoxydans	Betaproteobacteria	YES	YES	YES	YES
Herpetosiphon_aurantiacus_ATCC_23779	Chloroflexi	YES	YES	YES	NO
Hirschia_baltica_ATCC_49814	Alphaproteobacteria	YES	YES	YES	NO
Hydrogenobacter_thermophilus_TK_6_uid45927	Aquificae	YES	YES	YES	YES
Hydrogenobaculum_Y04AAS1	Aquificae	NO	YES	YES	NO
Hyperthermus_butylicus	Crenarchaeota	NO	NO	NO	NO
Hyphomicrobium_denitrificans_ATCC_51888_uid50325	Alphaproteobacteria	YES	YES	YES	YES
Hyphomonas_neptunium_ATCC_15444	Alphaproteobacteria	YES	YES	YES	NO
Idiomarina_loihiensis_L2TR	Gammaproteobacteria	YES	YES	YES	NO
Ignicoccus_hospitalis_KIN4_I	Crenarchaeota	NO	NO	NO	NO
Jannaschia_CCS1	Alphaproteobacteria	YES	YES	YES	NO
Janthinobacterium_Marseille	Betaproteobacteria	YES	YES	YES	YES
Jonesia_denitrificans_DSM_20603	Actinobacteria	YES	YES	NO	YES
Kangiella_koreensis_DSM_16069	Gammaproteobacteria	YES	YES	YES	YES
Kineococcus_radiotolerans_SRS30216	Actinobacteria	YES	NO	NO	NO
Klebsiella_pneumoniae_342	Gammaproteobacteria	NO	NO	NO	NO
Klebsiella_pneumoniae_MGH_78578	Gammaproteobacteria	NO	NO	NO	NO
Klebsiella_pneumoniae_NTUH_K2044	Gammaproteobacteria	NO	NO	NO	NO
Klebsiella_variicola_At_22_uid42113	Gammaproteobacteria	NO	NO	NO	NO
Kocuria_rhizophila_DC2201	Actinobacteria	YES	YES	NO	NO
Kosmotoga_olearia_TBF_19_5_1	Thermotogae	NO	NO	NO	NO
Kribbella_flavida_DSM_17836_uid43465	Actinobacteria	YES	YES	YES	NO
Kytococcus_sedentarius_DSM_20547	Actinobacteria	YES	YES	NO	NO
Lactobacillus_acidophilus_NCFM	Firmicutes	NO	NO	NO	YES
Lactobacillus_brevis_ATCC_367	Firmicutes	NO	NO	NO	NO
Lactobacillus_casei	Firmicutes	NO	NO	NO	NO
Lactobacillus_casei_ATCC_334	Firmicutes	NO	NO	NO	NO
Lactobacillus_casei_Zhang_uid50673	Firmicutes	NO	NO	NO	NO
Lactobacillus_crispatus_ST1_uid48359	Firmicutes	NO	NO	NO	YES
Lactobacillus_delbrueckii_bulgaricus	Firmicutes	NO	NO	NO	NO
Lactobacillus_delbrueckii_bulgaricus_ATCC_BAA-365	Firmicutes	NO	NO	NO	NO
Lactobacillus_fermentum_IFO_3956	Firmicutes	NO	NO	NO	NO
Lactobacillus_gasseri_ATCC_33323	Firmicutes	NO	NO	NO	YES
Lactobacillus_helveticus_DPC_4571	Firmicutes	NO	NO	NO	YES
Lactobacillus_johnsonii_FI9785	Firmicutes	NO	NO	NO	YES

Lactobacillus_johnsonii_NCC_533	Firmicutes	NO	NO	NO	YES
Lactobacillus_plantarum	Firmicutes	NO	NO	NO	NO
Lactobacillus_plantarum_JDM1	Firmicutes	NO	NO	NO	NO
Lactobacillus_reuteri_DSM_20016	Firmicutes	NO	NO	NO	YES
Lactobacillus_reuteri_F275_Kitasato	Firmicutes	NO	NO	NO	YES
Lactobacillus_rhamnosus_GG	Firmicutes	NO	NO	NO	NO
Lactobacillus_rhamnosus_Lc_705	Firmicutes	NO	NO	NO	NO
Lactobacillus_sakei_23K	Firmicutes	NO	NO	NO	NO
Lactobacillus_salivarius_UCC118	Firmicutes	NO	NO	NO	NO
Lactococcus_lactis	Firmicutes	NO	NO	NO	NO
Lactococcus_lactis_cremoris_MG1363	Firmicutes	NO	NO	NO	NO
Lactococcus_lactis_cremoris_SK11	Firmicutes	NO	NO	NO	NO
Lactococcus_lactis_KF147	Firmicutes	NO	NO	NO	NO
Laribacter_hongkongensis_HLHK9	Betaproteobacteria	YES	YES	YES	NO
Lawsonia_intracellularis_PHE_MN1-00	Deltaproteobacteria	NO	NO	NO	NO
Legionella_longbeachae_NSW150_uid46099	Gammaproteobacteria	YES	NO	YES	NO
Legionella_pneumophila_2300_99_Alcoy_uid48801	Gammaproteobacteria	YES	NO	YES	NO
Legionella_pneumophila_Corby	Gammaproteobacteria	YES	NO	YES	NO
Legionella_pneumophila_Lens	Gammaproteobacteria	YES	NO	YES	NO
Legionella_pneumophila_Paris	Gammaproteobacteria	YES	NO	YES	NO
Legionella_pneumophila_Philadelphia_1	Gammaproteobacteria	YES	NO	YES	NO
Leifsonia_xyli_xyli_CTCB0	Actinobacteria	YES	YES	NO	NO
Leptospira_biflexa_serovar_Patoc__Patoc_1__Ames_	Spirochaetes	YES	YES	YES	NO
Leptospira_biflexa_serovar_Patoc__Patoc_1__Paris_	Spirochaetes	YES	YES	YES	NO
Leptospira_borgpetersenii_serovar_Hardjo-bovis_JB197	Spirochaetes	YES	NO	YES	NO
Leptospira_borgpetersenii_serovar_Hardjo-bovis_L550	Spirochaetes	YES	NO	YES	NO
Leptospira_interrogans_serovar_Copenhageni	Spirochaetes	YES	NO	YES	NO
Leptospira_interrogans_serovar_Lai	Spirochaetes	YES	NO	YES	NO
Leptothrix_cholodnii_SP_6	Betaproteobacteria	YES	YES	YES	YES
Leptotrichia_buccalis_DSM_1135	Fusobacteria	NO	NO	NO	NO
Leuconostoc_citreum_KM20	Firmicutes	NO	NO	NO	NO
Leuconostoc_gasicomitatum_LMG_18811_uid50385	Firmicutes	NO	NO	NO	NO
Leuconostoc_kimchii_IMSNU_11154_uid48589	Firmicutes	NO	NO	NO	NO
Leuconostoc_mesenteroides_ATCC_8293	Firmicutes	NO	NO	NO	NO
Listeria_innocua	Firmicutes	NO	NO	NO	NO
Listeria_monocytogenes	Firmicutes	NO	NO	NO	NO
Listeria_monocytogenes_08_5923_uid43727	Firmicutes	NO	NO	NO	NO
Listeria_monocytogenes_4b_F2365	Firmicutes	NO	NO	NO	NO
Listeria_monocytogenes_Clip81459	Firmicutes	NO	NO	NO	NO
Listeria_monocytogenes_HCC23	Firmicutes	NO	NO	NO	NO
Listeria_monocytogenes_uid43671	Firmicutes	NO	NO	NO	NO
Listeria_seeligeri_serovar_1_2b_SLCC3954_uid46215	Firmicutes	NO	NO	NO	NO
Listeria_welshimeri_serovar_6b_SLCC5334	Firmicutes	NO	NO	NO	NO
Lysinibacillus_sphaericus_C3_41	Firmicutes	YES	NO	YES	NO
Macrococcus_caseolyticus_JCSC5402	Firmicutes	YES	NO	NO	NO

Magnetococcus_MC-1	Other_Bacteria	NO	YES	YES	NO
Magnetospirillum_magneticum_AMB-1	Alphaproteobacteria	YES	YES	YES	YES
Mannheimia_succiniciproducens_MBEL55E	Gammaproteobacteria	NO	NO	NO	NO
Maricaulis_maris_MCS10	Alphaproteobacteria	YES	YES	YES	NO
Marinobacter_aquaeolei_VT8	Gammaproteobacteria	YES	YES	YES	YES
Marinomonas_MWYL1	Gammaproteobacteria	NO	NO	NO	NO
Meiothermus_ruber_DSM_1279_uid46661	Deinococcus-Thermus	YES	YES	YES	NO
Meiothermus_silvanus_DSM_9946_uid49485	Deinococcus-Thermus	YES	YES	YES	NO
Mesoplasma_florum_L1	Firmicutes	NO	NO	NO	NO
Mesorhizobium_BNC1	Alphaproteobacteria	YES	YES	YES	NO
Mesorhizobium_loti	Alphaproteobacteria	YES	YES	YES	NO
Metallosphaera_sedula_DSM_5348	Crenarchaeota	YES	NO	NO	NO
Methanobacterium_thermoautotrophicum	Euryarchaeota	NO	NO	NO	NO
Methanobrevibacter_ruminantium_M1_uid45857	Euryarchaeota	NO	NO	NO	NO
Methanobrevibacter_smithii_ATCC_35061	Euryarchaeota	NO	NO	NO	NO
Methanocaldococcus_fervens_AG86	Euryarchaeota	NO	NO	NO	NO
Methanocaldococcus_FS406_22_uid42499	Euryarchaeota	NO	NO	NO	NO
Methanocaldococcus_infernus_ME_uid48803	Euryarchaeota	NO	NO	NO	NO
Methanocaldococcus_vulcanius_M7	Euryarchaeota	NO	NO	NO	NO
Methanocella_paludicola_SANAE	Euryarchaeota	NO	NO	NO	NO
Methanococcoides_burtonii_DSM_6242	Euryarchaeota	NO	NO	NO	NO
Methanococcus_aeolicus_Nankai-3	Euryarchaeota	NO	NO	NO	NO
Methanococcus_jannaschii	Euryarchaeota	NO	NO	NO	NO
Methanococcus_maripaludis_C5	Euryarchaeota	NO	NO	NO	NO
Methanococcus_maripaludis_C6	Euryarchaeota	NO	NO	NO	NO
Methanococcus_maripaludis_C7	Euryarchaeota	NO	NO	NO	NO
Methanococcus_maripaludis_S2	Euryarchaeota	NO	NO	NO	NO
Methanococcus_vannielii_SB	Euryarchaeota	NO	NO	NO	NO
Methanococcus_voltae_A3_uid49529	Euryarchaeota	NO	NO	NO	NO
Methanocorpusculum_labreanum_Z	Euryarchaeota	NO	NO	NO	NO
Methanoculleus_marisnigri_JR1	Euryarchaeota	NO	NO	NO	NO
Methanohalobium_evestigatum_Z_7303_uid49857	Euryarchaeota	NO	NO	NO	NO
Methanohalophilus_mahii_DSM_5219_uid47313	Euryarchaeota	NO	NO	NO	NO
Methanopyrus_kandleri	Euryarchaeota	NO	NO	NO	NO
Methanosaeta_thermophila_PT	Euryarchaeota	NO	NO	NO	NO
Methanosarcina_acetivorans	Euryarchaeota	NO	NO	NO	NO
Methanosarcina_barkeri_fusaro	Euryarchaeota	NO	NO	NO	NO
Methanosarcina_mazei	Euryarchaeota	NO	NO	NO	NO
Methanosphaera_stadtmanae	Euryarchaeota	NO	NO	NO	NO
Methanospirillum_hungatei_JF-1	Euryarchaeota	NO	NO	NO	NO
Methylacidiphilum_infernorum_V4	Other_Bacteria	YES	NO	YES	NO
Methylibium_petroleiphilum_PM1	Betaproteobacteria	YES	YES	YES	YES
Methylobacillus_flagellatus_KT	Betaproteobacteria	YES	YES	YES	NO
Methylobacterium_4_46	Alphaproteobacteria	YES	YES	YES	NO
Methylobacterium_chloromethanicum_CM4	Alphaproteobacteria	YES	YES	YES	NO

Methylobacterium_extorquens_AM1	Alphaproteobacteria	YES	YES	YES	NO
Methylobacterium_extorquens_DM4	Alphaproteobacteria	YES	YES	YES	NO
Methylobacterium_extorquens_PA1	Alphaproteobacteria	YES	YES	YES	NO
Methylobacterium_nodulans_ORS_2060	Alphaproteobacteria	YES	YES	YES	NO
Methylobacterium_populi_BJ001	Alphaproteobacteria	YES	YES	YES	NO
Methylobacterium_radiotolerans_JCM_2831	Alphaproteobacteria	YES	YES	YES	NO
Methylocella_silvestris_BL2	Alphaproteobacteria	YES	YES	YES	NO
Methylococcus_capsulatus_Bath	Alphaproteobacteria	YES	YES	YES	NO
Methylotenera_301_uid49469	Alphaproteobacteria	YES	YES	YES	YES
Methylotenera_mobilis_JLW8	Betaproteobacteria	YES	YES	YES	NO
Methylovorus_SIP3_4	Betaproteobacteria	YES	YES	YES	NO
Micrococcus_luteus_NCTC_2665	Actinobacteria	YES	YES	NO	NO
Microcystis_aeruginosa_NIES_843	Cyanobacteria	YES	NO	NO	YES
Mobiluncus_curtisii_ATCC_43063_uid49695	Actinobacteria	NO	NO	NO	NO
Moorella_thermoacetica_ATCC_39073	Firmicutes	NO	NO	NO	NO
Moraxella_catarrhalis_RH4_uid48809	Gammaproteobacteria	NO	NO	YES	NO
Mycobacterium_abscessus_ATCC_19977	Actinobacteria	YES	YES	NO	NO
Mycobacterium_avium_104	Actinobacteria	YES	NO	NO	NO
Mycobacterium_avium_paratuberculosis	Actinobacteria	YES	NO	NO	NO
Mycobacterium_bovis	Actinobacteria	YES	NO	NO	NO
Mycobacterium_bovis_BCG_Pasteur_1173P2	Actinobacteria	YES	NO	NO	NO
Mycobacterium_bovis_BCG_Tokyo_172	Actinobacteria	YES	NO	NO	NO
Mycobacterium_gilvum_PYR-GCK	Actinobacteria	YES	YES	NO	YES
Mycobacterium_JLS	Actinobacteria	YES	YES	NO	YES
Mycobacterium_KMS	Actinobacteria	YES	YES	NO	YES
Mycobacterium_leprae	Actinobacteria	YES	NO	NO	NO
Mycobacterium_leprae_Br4923	Actinobacteria	YES	NO	NO	NO
Mycobacterium_marinum_M	Actinobacteria	YES	NO	NO	YES
Mycobacterium_MCS	Actinobacteria	YES	YES	NO	YES
Mycobacterium_smegmatis_MC2_155	Actinobacteria	YES	YES	NO	NO
Mycobacterium_tuberculosis_CDC1551	Actinobacteria	YES	NO	NO	NO
Mycobacterium_tuberculosis_F11	Actinobacteria	YES	NO	NO	NO
Mycobacterium_tuberculosis_H37Ra	Actinobacteria	YES	NO	NO	NO
Mycobacterium_tuberculosis_H37Rv	Actinobacteria	YES	NO	NO	NO
Mycobacterium_tuberculosis_KZN_1435	Actinobacteria	YES	NO	NO	NO
Mycobacterium_ulcerans_Agy99	Actinobacteria	YES	NO	NO	NO
Mycobacterium_vanbaalenii_PYR-1	Actinobacteria	YES	YES	NO	YES
Mycoplasma_agalactiae_PG2	Firmicutes	NO	NO	NO	NO
Mycoplasma_agalactiae_uid46679	Firmicutes	NO	NO	NO	NO
Mycoplasma_arthritidis_158L3_1	Firmicutes	NO	NO	NO	NO
Mycoplasma_capricolum_ATCC_27343	Firmicutes	NO	NO	NO	NO
Mycoplasma_conjunctivae_HRC_581_uid32285	Firmicutes	NO	NO	NO	NO
Mycoplasma_crocodyli_MP145_uid47087	Firmicutes	NO	NO	NO	NO
Mycoplasma_gallisepticum	Firmicutes	NO	NO	NO	NO
Mycoplasma_genitalium	Firmicutes	NO	NO	NO	NO

Mycoplasma_hominis	Firmicutes	NO	NO	NO	NO
Mycoplasma_hyopneumoniae_232	Firmicutes	NO	NO	NO	NO
Mycoplasma_hyopneumoniae_7448	Firmicutes	NO	NO	NO	NO
Mycoplasma_hyopneumoniae_J	Firmicutes	NO	NO	NO	NO
Mycoplasma_mobile_163K	Firmicutes	NO	NO	NO	NO
Mycoplasma_mycoides	Firmicutes	NO	NO	NO	NO
Mycoplasma_penetrans	Firmicutes	NO	NO	NO	NO
Mycoplasma_pneumoniae	Firmicutes	NO	NO	NO	NO
Mycoplasma_pulmonis	Firmicutes	NO	NO	NO	NO
Mycoplasma_synoviae_53	Firmicutes	NO	NO	NO	NO
Myxococcus_xanthus_DK_1622	Deltaproteobacteria	YES	NO	YES	YES
Nakamurella_multipartita_DSM_44233	Actinobacteria	YES	YES	NO	NO
Nanoarchaeum_equitans	Nanoarchaeota	NO	NO	NO	NO
Natranaerobius_thermophilus_JW_NM_WN_LF	Firmicutes	NO	NO	NO	NO
Natrialba_magadii_ATCC_43099_uid46245	Euryarchaeota	YES	NO	YES	NO
Natronomonas_pharaonis	Euryarchaeota	YES	NO	YES	NO
Nautilia_profundicola_AmH	Epsilonproteobacteria	NO	NO	NO	NO
Neisseria_gonorrhoeae_FA_1090	Betaproteobacteria	NO	YES	YES	NO
Neisseria_gonorrhoeae_NCCP11945	Betaproteobacteria	YES	YES	YES	NO
Neisseria_meningitidis_053442	Betaproteobacteria	NO	YES	YES	NO
Neisseria_meningitidis_alpha14	Betaproteobacteria	NO	YES	YES	NO
Neisseria_meningitidis_FAM18	Betaproteobacteria	NO	YES	YES	NO
Neisseria_meningitidis_MC58	Betaproteobacteria	NO	YES	YES	NO
Neisseria_meningitidis_Z2491	Betaproteobacteria	NO	YES	YES	NO
Neorickettsia_risticii_Illinois	Alphaproteobacteria	YES	NO	YES	NO
Neorickettsia_sennetsu_Miyayama	Alphaproteobacteria	YES	NO	YES	NO
Nitratiruptor_SB155-2	Epsilonproteobacteria	YES	NO	NO	NO
Nitrobacter_hamburgensis_X14	Alphaproteobacteria	YES	YES	YES	NO
Nitrobacter_winogradskyi_Nb-255	Alphaproteobacteria	YES	YES	YES	NO
Nitrosococcus_halophilus_Nc4_uid46803	Gammaproteobacteria	YES	NO	YES	YES
Nitrosococcus_oceani_ATCC_19707	Gammaproteobacteria	YES	NO	YES	YES
Nitrosococcus_watsoni_C_113_uid50331	Gammaproteobacteria	YES	NO	YES	NO
Nitrosomonas_europaea	Betaproteobacteria	YES	NO	YES	NO
Nitrosomonas_eutropha_C71	Betaproteobacteria	YES	YES	YES	YES
Nitrosopumilus_maritimus_SCM1	Crenarchaeota	YES	NO	NO	NO
Nitrosospira_multiformis_ATCC_25196	Betaproteobacteria	YES	NO	YES	YES
Nocardia_farcinica_IFM10152	Actinobacteria	YES	YES	NO	NO
Nocardioides_JS614	Actinobacteria	YES	YES	YES	YES
Nocardiopsis_dassonvillei_DSM_43111_uid49483	Actinobacteria	YES	YES	YES	NO
Nostoc_punctiforme_PCC_73102	Cyanobacteria	YES	NO	NO	YES
Nostoc_sp	Cyanobacteria	YES	NO	NO	YES
Novosphingobium_aromaticivorans_DSM_12444	Alphaproteobacteria	YES	YES	YES	NO
Oceanobacillus_iheyensis	Firmicutes	YES	NO	YES	NO
Ochrobactrum_anthropi_ATCC_49188	Alphaproteobacteria	YES	YES	YES	NO
Oenococcus_oeni_PSU-1	Firmicutes	NO	NO	NO	NO

Oligotropha_carboxidovorans_OM5	Alphaproteobacteria	YES	YES	YES	NO
Onion_yellows_phytoplasma	Firmicutes	NO	NO	NO	NO
Opitutus_terrae_PB90_1	Other_Bacteria	YES	NO	YES	NO
Orientia_tsutsugamushi_Boryong	Alphaproteobacteria	YES	YES	YES	YES
Orientia_tsutsugamushi_Ikeda	Alphaproteobacteria	YES	YES	YES	YES
Paenibacillus_JDR_2	Firmicutes	YES	NO	YES	YES
Pantoea_ananatis_LMG_20103_uid46807	Gammaproteobacteria	NO	NO	NO	NO
Parabacteroides_distasonis_ATCC_8503	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Parachlamydia_sp_UWE25	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Paracoccus_denitrificans_PD1222	Alphaproteobacteria	YES	YES	YES	NO
Parvibaculum_lavamentivorans_DS-1	Alphaproteobacteria	YES	YES	YES	NO
Pasteurella_multocida	Gammaproteobacteria	NO	YES	NO	NO
Pectobacterium_carotovorum_PC1	Gammaproteobacteria	NO	NO	NO	NO
Pectobacterium_wasabiae_WPP163	Gammaproteobacteria	NO	NO	NO	NO
Pediococcus_pentosaceus_ATCC_25745	Firmicutes	NO	NO	NO	NO
Pedobacter_heparinus_DSM_2366	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Pelobacter_carbinolicus	Deltaproteobacteria	YES	NO	NO	NO
Pelobacter_propionicus_DSM_2379	Deltaproteobacteria	NO	NO	NO	NO
Pelodictyon_phaeoclathratiforme_BU_1	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Pelotomaculum_thermopropionicum_SI	Firmicutes	NO	NO	NO	NO
Persephonella_marina_EX_H1	Aquificae	YES	YES	YES	NO
Petrotoga_mobilis_SJ95	Thermotogae	NO	NO	NO	NO
Phenylobacterium_zucineum_HLK1	Alphaproteobacteria	YES	YES	YES	NO
Photobacterium_profundum_SS9	Gammaproteobacteria	YES	YES	YES	YES
Photorhabdus_asymbiotica	Gammaproteobacteria	NO	NO	NO	NO
Photorhabdus_luminescens	Gammaproteobacteria	NO	NO	NO	NO
Picrophilus_torridus_DSM_9790	Euryarchaeota	YES	NO	NO	NO
Pirellula_sp	Planctomycetes	YES	NO	YES	NO
Pirellula_staleyi_DSM_6068_uid43209	Planctomycetes	YES	NO	YES	NO
Planctomyces_limnophilus_DSM_3776_uid48643	Planctomycetes	YES	NO	YES	NO
Polaromonas_JS666	Betaproteobacteria	YES	YES	YES	NO
Polaromonas_naphthalenivorans_CJ2	Betaproteobacteria	YES	YES	YES	YES
Polynucleobacter_necessarius_asymbioticus_QLW_P1DMWA_1	Betaproteobacteria	YES	YES	YES	NO
Polynucleobacter_necessarius_STIR1	Betaproteobacteria	YES	YES	YES	NO
Porphyromonas_gingivalis_ATCC_33277	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Porphyromonas_gingivalis_W83	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Prevotella_ruminicola_23_uid47507	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Prochlorococcus_marinus_AS9601	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_CCMP1375	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_MED4	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_MIT_9211	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_MIT_9215	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_MIT_9301	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_MIT_9312	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_MIT_9515	Cyanobacteria	YES	NO	NO	NO

Prochlorococcus_marinus_MIT9313	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_NATL1A	Cyanobacteria	YES	NO	NO	NO
Prochlorococcus_marinus_NATL2A	Cyanobacteria	YES	NO	NO	NO
Propionibacterium_acnes_KPA171202	Actinobacteria	YES	NO	NO	YES
Propionibacterium_acnes_SK137_uid48071	Actinobacteria	YES	NO	NO	YES
Propionibacterium_freudenreichii_shermanii_CIRM_BIA1_uid4	Actinobacteria	NO	NO	NO	YES
Prosthecochloris_aestuarii_DSM_271	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Prosthecochloris_vibrioformis_DSM_265	Bacteroidetes/Chlorobi	NO	NO	NO	NO
Proteus_mirabilis	Gammaproteobacteria	NO	NO	NO	NO
Pseudoalteromonas_atlantica_T6c	Gammaproteobacteria	YES	YES	YES	YES
Pseudoalteromonas_haloplanktis_TAC125	Gammaproteobacteria	YES	YES	YES	NO
Pseudomonas_aeruginosa	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_aeruginosa_LESB58	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_aeruginosa_PA7	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_aeruginosa_UCBPP-PA14	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_entomophila_L48	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_fluorescens_Pf0_1	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_fluorescens_Pf-5	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_fluorescens_SBW25	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_mendocina_ymp	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_putida_F1	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_putida_GB_1	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_putida_KT2440	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_putida_W619	Gammaproteobacteria	YES	YES	YES	YES
Pseudomonas_stutzeri_A1501	Gammaproteobacteria	YES	YES	YES	NO
Pseudomonas_syringae_phaseolicola_1448A	Gammaproteobacteria	NO	YES	YES	NO
Pseudomonas_syringae_pv_B728a	Gammaproteobacteria	NO	YES	YES	NO
Pseudomonas_syringae_tomato_DC3000	Gammaproteobacteria	NO	YES	YES	NO
Psychrobacter_arcticum_273-4	Gammaproteobacteria	NO	NO	NO	NO
Psychrobacter_cryohalolentis_K5	Gammaproteobacteria	NO	NO	NO	NO
Psychrobacter_PRwf-1	Gammaproteobacteria	NO	NO	NO	NO
Psychromonas_ingrahamii_37	Gammaproteobacteria	YES	YES	YES	YES
Pyrobaculum_aerophilum	Crenarchaeota	YES	YES	YES	NO
Pyrobaculum_arsenaticum_DSM_13514	Crenarchaeota	NO	YES	YES	NO
Pyrobaculum_calidifontis_JCM_11548	Crenarchaeota	YES	YES	YES	NO
Pyrobaculum_islandicum_DSM_4184	Crenarchaeota	NO	NO	NO	NO
Pyrococcus_abyssi	Euryarchaeota	NO	NO	NO	NO
Pyrococcus_furiosus	Euryarchaeota	NO	NO	NO	NO
Pyrococcus_horikoshii	Euryarchaeota	NO	NO	NO	NO
Ralstonia_eutropha_H16	Betaproteobacteria	YES	YES	YES	NO
Ralstonia_eutropha_JMP134	Betaproteobacteria	YES	YES	YES	NO
Ralstonia_pickettii_12D	Betaproteobacteria	YES	YES	YES	YES
Ralstonia_pickettii_12J	Betaproteobacteria	YES	YES	YES	YES
Ralstonia_solanacearum	Betaproteobacteria	YES	YES	YES	NO
Ralstonia_solanacearum_CFBP2957_uid50545	Betaproteobacteria	YES	YES	YES	NO

Renibacterium_salmoninarum_ATCC_33209	Actinobacteria	YES	NO	NO	NO
Rhizobium_etli_CFN_42	Alphaproteobacteria	YES	YES	YES	YES
Rhizobium_etli_CIAT_652	Alphaproteobacteria	YES	YES	YES	NO
Rhizobium_leguminosarum_bv__trifolii_WSM1325	Alphaproteobacteria	YES	YES	YES	NO
Rhizobium_leguminosarum_bv__trifolii_WSM2304	Alphaproteobacteria	YES	YES	YES	YES
Rhizobium_leguminosarum_bv_viciae_3841	Alphaproteobacteria	YES	YES	YES	YES
Rhizobium_NGR234	Alphaproteobacteria	YES	YES	YES	YES
Rhodobacter_capsulatus_SB_1003_uid47509	Alphaproteobacteria	YES	YES	YES	NO
Rhodobacter_sphaeroides_2_4_1	Alphaproteobacteria	YES	YES	YES	YES
Rhodobacter_sphaeroides_ATCC_17025	Alphaproteobacteria	YES	YES	YES	NO
Rhodobacter_sphaeroides_ATCC_17029	Alphaproteobacteria	YES	YES	YES	YES
Rhodobacter_sphaeroides_KD131	Alphaproteobacteria	YES	YES	YES	YES
Rhodococcus_erythropolis_PR4	Actinobacteria	YES	YES	NO	NO
Rhodococcus_jostii_RHA1	Actinobacteria	YES	YES	NO	YES
Rhodococcus_opacus_B4_uid13791	Actinobacteria	YES	YES	NO	YES
Rhodoferax_ferrireducens_T118	Betaproteobacteria	YES	YES	YES	YES
Rhodopseudomonas_palustris_BisA53	Alphaproteobacteria	YES	YES	YES	NO
Rhodopseudomonas_palustris_BisB18	Alphaproteobacteria	YES	YES	YES	NO
Rhodopseudomonas_palustris_BisB5	Alphaproteobacteria	YES	YES	YES	YES
Rhodopseudomonas_palustris_CGA009	Alphaproteobacteria	YES	YES	YES	NO
Rhodopseudomonas_palustris_HaA2	Alphaproteobacteria	YES	YES	YES	NO
Rhodopseudomonas_palustris_TIE_1	Alphaproteobacteria	YES	YES	YES	NO
Rhodospirillum_centenum_SW	Alphaproteobacteria	YES	YES	YES	NO
Rhodospirillum_rubrum_ATCC_11170	Alphaproteobacteria	NO	YES	YES	NO
Rhodothermus_marinus_DSM_4252	Bacteroidetes/Chlorobi	YES	YES	YES	NO
Rickettsia_africae_ESF_5	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_akari_Hartford	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_bellii_OSU_85-389	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_bellii_RML369-C	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_canadensis_McKiel	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_conorii	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_felis_URRWXCal2	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_massiliae_MTU5	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_peacockii_Rustic	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_prowazekii	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_rickettsii_Iowa	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_rickettsii_Sheila_Smith	Alphaproteobacteria	YES	YES	YES	YES
Rickettsia_typhi_wilmington	Alphaproteobacteria	NO	YES	YES	YES
Robiginitalea_biformata_HTCC2501	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Roseiflexus_castenholzii_DSM_13941	Chloroflexi	YES	YES	YES	NO
Roseiflexus_RS-1	Chloroflexi	YES	YES	YES	NO
Roseobacter_denitrificans_OCh_114	Alphaproteobacteria	YES	YES	YES	NO
Rothia_mucilaginosa	Actinobacteria	NO	NO	NO	YES
Rubrobacter_xylanophilus_DSM_9941	Actinobacteria	NO	NO	NO	NO
Ruegeria_pomeroyi_DSS_3	Alphaproteobacteria	YES	YES	YES	YES

Saccharomonospora_viridis_DSM_43017	Actinobacteria	YES	YES	NO	NO
Saccharophagus_degradans_2-40	Gammaproteobacteria	YES	YES	YES	NO
Saccharopolyspora_erythraea_NRRL_2338	Actinobacteria	YES	YES	NO	NO
Salinibacter_ruber_DSM_13855	Bacteroidetes/Chlorobi	YES	NO	YES	YES
Salinibacter_ruber_uid47323	Bacteroidetes/Chlorobi	YES	NO	YES	YES
Salinispora_arenicola_CNS-205	Actinobacteria	YES	YES	NO	NO
Salinispora_tropica_CNB-440	Actinobacteria	YES	YES	NO	NO
Salmonella_enterica_arizonae_serovar_62_z4_z23__RSK2980_u	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_Choleraesuis	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_Paratypi_ATCC_9150	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Agona_SL483	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Dublin_CT_02021853	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Enteritidis_P125109	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Gallinarum_287_91	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Heidelberg_SL476	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Newport_SL254	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Paratyphi_A_AKU_12601	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Paratyphi_B_SPB7	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Paratyphi_C_RKS4594	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Schwarzengrund_CVM19633	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_enterica_serovar_Typhi_Ty2	Gammaproteobacteria	NO	NO	NO	NO
Salmonella_typhi	Gammaproteobacteria	NO	NO	NO	NO
Sanguibacter_keddieii_DSM_10542_uid40845	Actinobacteria	YES	YES	NO	YES
Sebaldella_termitidis_ATCC_33386	Fusobacteria	NO	NO	NO	NO
Segniliparus_rotundus_DSM_44985_uid49049	Actinobacteria	YES	NO	NO	NO
Serratia_proteamaculans_568	Gammaproteobacteria	NO	NO	NO	NO
Shewanella_amazonensis_SB2B	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_ANA-3	Gammaproteobacteria	YES	YES	YES	YES
Shewanella_baltica_OS155	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_baltica_OS185	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_baltica_OS195	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_baltica_OS223	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_denitrificans_OS217	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_frigidimarina_NCIMB_400	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_halifaxensis_HAW_EB4	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_loihica_PV-4	Gammaproteobacteria	YES	YES	YES	YES
Shewanella_MR-4	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_MR-7	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_oneidensis	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_pealeana_ATCC_700345	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_piezotolerans_WP3	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_putrefaciens_CN-32	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_sediminis_HAW-EB3	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_violacea_DSS12_uid47085	Gammaproteobacteria	YES	YES	YES	NO
Shewanella_W3-18-1	Gammaproteobacteria	YES	YES	YES	NO

Shewanella_woodyi_ATCC_51908	Gammaproteobacteria	YES	YES	YES	NO
Shigella_boydii_CDC_3083_94	Gammaproteobacteria	NO	NO	NO	NO
Shigella_boydii_Sb227	Gammaproteobacteria	NO	NO	NO	NO
Shigella_dysenteriae	Gammaproteobacteria	NO	NO	NO	NO
Shigella_flexneri_2a	Gammaproteobacteria	NO	NO	NO	NO
Shigella_flexneri_2a_2457T	Gammaproteobacteria	NO	NO	NO	NO
Shigella_flexneri_5_8401	Gammaproteobacteria	NO	NO	NO	NO
Shigella_sonnei_Ss046	Gammaproteobacteria	NO	NO	NO	NO
Sideroxydans_lithotrophicus_ES_1_uid46801	Gammaproteobacteria	NO	YES	YES	NO
Silicibacter_TM1040	Alphaproteobacteria	YES	YES	YES	NO
Sinorhizobium_medicae_WSM419	Alphaproteobacteria	YES	YES	YES	NO
Sinorhizobium_meliloti	Alphaproteobacteria	YES	YES	YES	YES
Slackia_heliotrinireducens_DSM_20476	Actinobacteria	NO	NO	NO	NO
Sodalis_glossinidius_morsitans	Gammaproteobacteria	NO	NO	NO	NO
Solibacter_usitatus_Ellin6076	Acidobacteria	YES	NO	YES	NO
Sorangium_cellulosum__So_ce_56_	Deltaproteobacteria	YES	NO	YES	YES
Sphaerobacter_thermophilus_DSM_20745	Chloroflexi	YES	YES	YES	YES
Sphingobium_japonicum_UT26S_uid47077	Alphaproteobacteria	YES	YES	YES	NO
Sphingomonas_wittichii_RW1	Alphaproteobacteria	YES	YES	YES	NO
Sphingopyxis_alaskensis_RB2256	Alphaproteobacteria	YES	YES	YES	NO
Spirosoma_linguale_DSM_74_uid43413	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Stackebrandtia_nassauensis_DSM_44728_uid46663	Actinobacteria	YES	YES	NO	NO
Staphylococcus_aureus_aureus_MRSA252	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_aureus_MSSA476	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_COL	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_ED98	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_JH1	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_JH9	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_Mu3	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_Mu50	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_MW2	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_N315	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_NCTC_8325	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_Newman	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_RF122	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_USA300_FPR3757	Firmicutes	NO	NO	NO	NO
Staphylococcus_aureus_USA300_TCH1516	Firmicutes	NO	NO	NO	NO
Staphylococcus_carnosus_TM300	Firmicutes	NO	NO	NO	NO
Staphylococcus_epidermidis_ATCC_12228	Firmicutes	NO	NO	NO	NO
Staphylococcus_epidermidis_RP62A	Firmicutes	NO	NO	NO	NO
Staphylococcus_haemolyticus	Firmicutes	NO	NO	NO	NO
Staphylococcus_lugdunensis_HKU09_01_uid46233	Firmicutes	NO	NO	NO	NO
Staphylococcus_saprophyticus	Firmicutes	NO	NO	NO	NO
Staphylothermus_hellenicus_DSM_12710_uid45893	Crenarchaeota	NO	NO	NO	NO
Staphylothermus_marinus_F1	Crenarchaeota	NO	NO	NO	NO

Starkeya_novella_DSM_506_uid48815	Alphaproteobacteria	YES	YES	YES	YES
Stenotrophomonas_maltophilia_K279a	Gammaproteobacteria	YES	YES	YES	NO
Stenotrophomonas_maltophilia_R551_3	Gammaproteobacteria	YES	YES	YES	NO
Streptobacillus_moniliformis_DSM_12112	Fusobacteria	NO	NO	NO	NO
Streptococcus_agalactiae_2603	Firmicutes	NO	NO	NO	NO
Streptococcus_agalactiae_A909	Firmicutes	NO	NO	NO	NO
Streptococcus_agalactiae_NEM316	Firmicutes	NO	NO	NO	NO
Streptococcus_dysgalactiae_equisimilis_GGS_124	Firmicutes	NO	NO	NO	NO
Streptococcus_equi_4047	Firmicutes	NO	NO	NO	NO
Streptococcus_equi_zooepidemicus	Firmicutes	NO	NO	NO	NO
Streptococcus_equi_zooepidemicus_MGCS10565	Firmicutes	NO	NO	NO	NO
Streptococcus_gallolyticus_UCN34_uid46061	Firmicutes	NO	NO	NO	NO
Streptococcus_gordonii_Challis_substr_CH1	Firmicutes	NO	NO	NO	NO
Streptococcus_mitis_B6_uid46097	Firmicutes	NO	NO	NO	YES
Streptococcus_mutans	Firmicutes	NO	NO	NO	NO
Streptococcus_mutans_NN2025_uid46353	Firmicutes	NO	NO	NO	NO
Streptococcus_pneumoniae_70585	Firmicutes	NO	NO	NO	YES
Streptococcus_pneumoniae_ATCC_700669	Firmicutes	NO	NO	NO	YES
Streptococcus_pneumoniae_CGSP14	Firmicutes	NO	NO	NO	NO
Streptococcus_pneumoniae_D39	Firmicutes	NO	NO	NO	YES
Streptococcus_pneumoniae_G54	Firmicutes	NO	NO	NO	YES
Streptococcus_pneumoniae_Hungary19A_6	Firmicutes	NO	NO	NO	YES
Streptococcus_pneumoniae_JJA	Firmicutes	NO	NO	NO	YES
Streptococcus_pneumoniae_P1031	Firmicutes	NO	NO	NO	NO
Streptococcus_pneumoniae_R6	Firmicutes	NO	NO	NO	YES
Streptococcus_pneumoniae_Taiwan19F_14	Firmicutes	NO	NO	NO	NO
Streptococcus_pneumoniae_TCH8431_19A_uid49735	Firmicutes	NO	NO	NO	NO
Streptococcus_pneumoniae_TIGR4	Firmicutes	NO	NO	NO	YES
Streptococcus_pyogenes_M1_GAS	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_Manfredo	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_MGAS10270	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_MGAS10394	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_MGAS10750	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_MGAS2096	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_MGAS315	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_MGAS5005	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_MGAS6180	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_MGAS8232	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_NZ131	Firmicutes	NO	NO	NO	NO
Streptococcus_pyogenes_SSI-1	Firmicutes	NO	NO	NO	NO
Streptococcus_sanguinis_SK36	Firmicutes	NO	NO	NO	NO
Streptococcus_suis_05ZYH33	Firmicutes	NO	NO	NO	NO
Streptococcus_suis_98HAH33	Firmicutes	NO	NO	NO	NO
Streptococcus_suis_BM407	Firmicutes	NO	NO	NO	NO
Streptococcus_suis_P1_7	Firmicutes	NO	NO	NO	NO

Streptococcus_suis_SC84	Firmicutes	NO	NO	NO	NO
Streptococcus_thermophilus_CNRZ1066	Firmicutes	NO	NO	NO	NO
Streptococcus_thermophilus_LMD-9	Firmicutes	NO	NO	NO	NO
Streptococcus_thermophilus_LMG_18311	Firmicutes	NO	NO	NO	NO
Streptococcus_uberis_0140J	Firmicutes	NO	NO	NO	NO
Streptomyces_avermitilis	Actinobacteria	YES	YES	YES	NO
Streptomyces_coelicolor	Actinobacteria	YES	YES	YES	NO
Streptomyces_griseus_NBRC_13350	Actinobacteria	YES	YES	YES	NO
Streptomyces_scabiei_87_22_uid46531	Actinobacteria	YES	YES	YES	NO
Streptosporangium_roseum_DSM_43021	Actinobacteria	YES	YES	NO	NO
Sulfolobus_acidocaldarius_DSM_639	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_islandicus_L_D_8_5_uid43679	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_islandicus_L_S_2_15	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_islandicus_M_14_25	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_islandicus_M_16_27	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_islandicus_M_16_4	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_islandicus_Y_G_57_14	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_islandicus_Y_N_15_51	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_solfataricus	Crenarchaeota	YES	NO	NO	NO
Sulfolobus_tokodaii	Crenarchaeota	YES	NO	NO	NO
Sulfurihydrogenibium_azorense_Az_Fu1	Aquificae	NO	NO	YES	NO
Sulfurihydrogenibium_YO3AOP1	Aquificae	NO	NO	YES	NO
Sulfurospirillum_deleyianum_DSM_6946	Epsilonproteobacteria	NO	NO	NO	NO
Sulfurovum_NBC37-1	Epsilonproteobacteria	YES	NO	NO	NO
Symbiobacterium_thermophilum_IAM14863	Firmicutes	YES	NO	YES	YES
Synechococcus_CC9311	Cyanobacteria	YES	NO	NO	NO
Synechococcus_CC9605	Cyanobacteria	YES	NO	NO	NO
Synechococcus_CC9902	Cyanobacteria	YES	NO	NO	NO
Synechococcus_elongatus_PCC_6301	Cyanobacteria	YES	NO	NO	YES
Synechococcus_elongatus_PCC_7942	Cyanobacteria	YES	NO	NO	YES
Synechococcus_PCC_7002	Cyanobacteria	YES	NO	NO	YES
Synechococcus_RCC307	Cyanobacteria	YES	NO	NO	NO
Synechococcus_sp_WH8102	Cyanobacteria	YES	NO	NO	NO
Synechococcus_WH_7803	Cyanobacteria	YES	NO	NO	NO
Syntrophobacter_fumaroxidans_MPOB	Deltaproteobacteria	NO	NO	NO	NO
Syntrophomonas_wolfei_Goettingen	Other_Bacteria	NO	NO	NO	NO
Syntrophothermus_lipocalidus_DSM_12680_uid49527	Firmicutes	NO	NO	NO	NO
Syntrophus_aciditrophicus_SB	Deltaproteobacteria	NO	NO	NO	NO
Teredinibacter_turnerae_T7901	Gammaproteobacteria	YES	YES	YES	NO
Thauera_MZ1T	Betaproteobacteria	YES	YES	YES	YES
Thermanaerovibrio_acidaminovorans_DSM_6589	Other_Bacteria	NO	NO	NO	NO
Thermincola_JR_uid48823	Firmicutes	NO	NO	NO	YES
Thermoanaerobacter_italicus_Ab9_uid46241	Firmicutes	NO	NO	NO	NO
Thermoanaerobacter_mathranii_A3_uid49481	Firmicutes	NO	NO	NO	NO
Thermoanaerobacter_pseudethanolicus_ATCC_33223	Firmicutes	NO	NO	NO	NO

Thermoanaerobacter_tengcongensis	Firmicutes	NO	NO	NO	NO
Thermoanaerobacter_X514	Firmicutes	NO	NO	NO	NO
Thermobaculum_terrenum_ATCC_BAA_798	Other_Bacteria	YES	YES	YES	YES
Thermobifida_fusca_YX	Actinobacteria	YES	YES	YES	YES
Thermobispora_bispora_DSM_43833_uid48999	Actinobacteria	YES	YES	NO	NO
Thermococcus_gammatolerans_EJ3	Euryarchaeota	NO	NO	NO	NO
Thermococcus_kodakaraensis_KOD1	Euryarchaeota	NO	NO	NO	NO
Thermococcus_onnurineus_NA1	Euryarchaeota	NO	NO	NO	NO
Thermococcus_sibiricus_MM_739	Euryarchaeota	NO	NO	NO	NO
Thermocrinis_albus_DSM_14484_uid46231	Aquificae	YES	YES	YES	NO
Thermodesulfovibrio_yellowstonii_DSM_11347	Other_Bacteria	NO	NO	NO	NO
Thermofilum_pendens_Hrk_5	Crenarchaeota	NO	NO	NO	NO
Thermomicrobium_roseum_DSM_5159	Chloroflexi	YES	YES	YES	YES
Thermomonospora_curvata_DSM_43183	Actinobacteria	YES	YES	NO	NO
Thermoplasma_acidophilum	Euryarchaeota	NO	NO	NO	NO
Thermoplasma_volcanium	Euryarchaeota	NO	NO	NO	NO
Thermoproteus_neutrophilus_V24Sta	Crenarchaeota	NO	NO	NO	NO
Thermosipho_africanus_TCF52B	Thermotogae	NO	NO	NO	NO
Thermosipho_melanesiensis_BI429	Thermotogae	NO	NO	NO	NO
Thermosphaera_aggregans_DSM_11486_uid48993	Crenarchaeota	NO	NO	NO	NO
Thermosynechococcus_elongatus	Cyanobacteria	YES	NO	NO	NO
Thermotoga_lettingae_TMO	Thermotogae	NO	NO	NO	NO
Thermotoga_maritima	Thermotogae	NO	NO	NO	NO
Thermotoga_naphthophila_RKU_10	Thermotogae	NO	NO	NO	NO
Thermotoga_neapolitana_DSM_4359	Thermotogae	NO	NO	NO	NO
Thermotoga_petrophila_RKU-1	Thermotogae	NO	NO	NO	NO
Thermotoga_RQ2	Thermotogae	NO	NO	NO	NO
Thermus_thermophilus_HB27	Deinococcus-Thermus	YES	YES	YES	NO
Thermus_thermophilus_HB8	Deinococcus-Thermus	YES	YES	YES	NO
Thioalkalivibrio_HL_EbGR7	Gammaproteobacteria	YES	YES	YES	NO
Thioalkalivibrio_K90mix_uid46181	Gammaproteobacteria	YES	YES	YES	NO
Thiobacillus_denitrificans_ATCC_25259	Betaproteobacteria	YES	YES	YES	NO
Thiomicrospira_crunogena_XCL-2	Gammaproteobacteria	NO	YES	YES	YES
Thiomicrospira_denitrificans_ATCC_33889	Gammaproteobacteria	YES	NO	NO	NO
Thiomonas_intermedia_K12_uid48825	Betaproteobacteria	YES	YES	YES	NO
Tolumonas_auensis_DSM_9187	Gammaproteobacteria	NO	NO	NO	YES
Treponema_denticola_ATCC_35405	Spirochaetes	NO	NO	NO	NO
Treponema_pallidum	Spirochaetes	NO	NO	NO	NO
Treponema_pallidum_SS14	Spirochaetes	NO	NO	NO	NO
Trichodesmium_erythraeum_IMS101	Cyanobacteria	YES	NO	NO	NO
Tropheryma_whipplei_TW08_27	Actinobacteria	YES	NO	NO	NO
Tropheryma_whipplei_Twist	Actinobacteria	YES	NO	NO	NO
Truepera_radiovictrix_DSM_17093_uid49533	Deinococcus-Thermus	YES	NO	YES	NO
Tsukamurella_paurometabola_DSM_20162_uid48829	Actinobacteria	YES	NO	NO	NO
uncultured_Termite_group_1_bacterium_phylotype_Rs_D17	Other_Bacteria	NO	NO	NO	NO

Ureaplasma_parvum_serovar_3_ATCC_27815	Firmicutes	NO	NO	NO	NO
Ureaplasma_urealyticum	Firmicutes	NO	NO	NO	NO
Ureaplasma_urealyticum_serovar_10_ATCC_33699	Firmicutes	NO	NO	NO	NO
Variovorax_paradoxus_S110	Betaproteobacteria	YES	YES	YES	NO
Veillonella_parvula_DSM_2008	Firmicutes	NO	NO	NO	NO
Verminephrobacter_eiseniae_EF01-2	Betaproteobacteria	YES	YES	YES	YES
Vibrio_cholerae	Gammaproteobacteria	NO	YES	YES	NO
Vibrio_cholerae_M66_2	Gammaproteobacteria	NO	YES	YES	NO
Vibrio_cholerae_MJ_1236	Gammaproteobacteria	NO	YES	YES	NO
Vibrio_cholerae_O395	Gammaproteobacteria	NO	YES	YES	NO
Vibrio_Ex25	Gammaproteobacteria	YES	YES	YES	NO
Vibrio_fischeri_ES114	Gammaproteobacteria	NO	YES	YES	NO
Vibrio_fischeri_MJ11	Gammaproteobacteria	NO	YES	YES	NO
Vibrio_harveyi_ATCC_BAA-1116	Gammaproteobacteria	YES	YES	YES	NO
Vibrio_parahaemolyticus	Gammaproteobacteria	YES	YES	YES	NO
Vibrio_splendidus_LGP32	Gammaproteobacteria	YES	YES	YES	NO
Vibrio_vulnificus_CMCP6	Gammaproteobacteria	YES	YES	YES	NO
Vibrio_vulnificus_YJ016	Gammaproteobacteria	YES	YES	YES	NO
Waddlia_chondrophila_WSU_86_1044_uid49531	Chlamydiae/Verrucomicrobia	NO	NO	NO	NO
Wigglesworthia_brevipalpis	Gammaproteobacteria	NO	NO	NO	NO
Wolbachia_endosymbiont_of_Brugia_malayi_TRS	Alphaproteobacteria	YES	NO	YES	NO
Wolbachia_endosymbiont_of_Culex_quinquefasciatus_Pel	Alphaproteobacteria	YES	NO	YES	NO
Wolbachia_endosymbiont_of_Drosophila_melanogaster	Alphaproteobacteria	YES	NO	YES	NO
Wolbachia_wRi	Alphaproteobacteria	YES	NO	YES	NO
Wolinella_succinogenes	Epsilonproteobacteria	YES	NO	NO	NO
Xanthobacter_autotrophicus_Py2	Alphaproteobacteria	YES	YES	YES	NO
Xanthomonas_albilineans	Gammaproteobacteria	YES	YES	YES	NO
Xanthomonas_campestris_8004	Gammaproteobacteria	YES	YES	YES	NO
Xanthomonas_campestris_ATCC_33913	Gammaproteobacteria	YES	YES	YES	NO
Xanthomonas_campestris_B100	Gammaproteobacteria	YES	YES	YES	NO
Xanthomonas_campestris_vesicatoria_85-10	Gammaproteobacteria	YES	YES	YES	NO
Xanthomonas_citri	Gammaproteobacteria	YES	YES	YES	NO
Xanthomonas_oryzae_KACC10331	Gammaproteobacteria	YES	YES	YES	NO
Xanthomonas_oryzae_MAFF_311018	Gammaproteobacteria	YES	YES	YES	NO
Xanthomonas_oryzae_PXO99A	Gammaproteobacteria	YES	YES	YES	NO
Xenorhabdus_bovienii_SS_2004_uid46345	Gammaproteobacteria	NO	NO	NO	NO
Xenorhabdus_nematophila_ATCC_19061_uid49133	Gammaproteobacteria	NO	NO	NO	NO
Xylanimonas_cellulosilytica_DSM_15894	Actinobacteria	YES	NO	NO	NO
Xylella_fastidiosa	Gammaproteobacteria	NO	NO	YES	NO
Xylella_fastidiosa_M12	Gammaproteobacteria	NO	NO	YES	NO
Xylella_fastidiosa_M23	Gammaproteobacteria	NO	NO	YES	NO
Xylella_fastidiosa_Temecula1	Gammaproteobacteria	NO	NO	YES	NO
Yersinia_enterocolitica_8081	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pestis_Antiqua	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pestis_biovar_Microtus_91001	Gammaproteobacteria	NO	NO	NO	NO

Yersinia_pestis_CO92	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pestis_KIM_10_uid288	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pestis_Nepal516	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pestis_Pestoides_F	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pestis_Z176003_uid47317	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pseudotuberculosis_IP_31758	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pseudotuberculosis_IP32953	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pseudotuberculosis_PB1_	Gammaproteobacteria	NO	NO	NO	NO
Yersinia_pseudotuberculosis_YPIII	Gammaproteobacteria	NO	NO	NO	NO
Zunongwangia_profunda_SM_A87_uid48073	Bacteroidetes/Chlorobi	YES	NO	YES	NO
Zymomonas_mobilis_NCIMB_11163	Alphaproteobacteria	NO	NO	NO	NO
Zymomonas_mobilis_ZM4	Alphaproteobacteria	NO	NO	NO	NO