Bioinformatics of Metalloproteins and Metalloproteomes

Yan Zhang [1,2,3,*,†] and Junge Zheng [1,2,3,†]

[1]   Shenzhen Key Laboratory of Marine Bioresources and Ecology, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518055, China; junge.zheng@hotmail.com

[2]   Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen 518055, China

[3]   Shenzhen Bay Laboratory, Shenzhen 518055, China

*   Correspondence: zhangyan@szu.edu.cn; Tel.: +86-755-2692-2024

†   These authors contributed equally to this work.

check for updates

Abstract: Trace metals are inorganic elements that are required for all organisms in very low quantities. They serve as cofactors and activators of metalloproteins involved in a variety of key cellular processes. While substantial effort has been made in experimental characterization of metalloproteins and their functions, the application of bioinformatics in the research of metalloproteins and metalloproteomes is still limited. In the last few years, computational prediction and comparative genomics of metalloprotein genes have arisen, which provide significant insights into their distribution, function, and evolution in nature. This review aims to offer an overview of recent advances in bioinformatic analysis of metalloproteins, mainly focusing on metalloprotein prediction and the use of different metals across the tree of life. We describe current computational approaches for the identification of metalloprotein genes and metal-binding sites/patterns in proteins, and then introduce a set of related databases. Furthermore, we discuss the latest research progress in comparative genomics of several important metals in both prokaryotes and eukaryotes, which demonstrates divergent and dynamic evolutionary patterns of different metalloprotein families and metalloproteomes. Overall, bioinformatic studies of metalloproteins provide a foundation for systematic understanding of trace metal utilization in all three domains of life.

1. Introduction

Biological trace metals refer to those metals which are only needed in small quantities but are essential for normal development, survival, and reproduction of all living organisms [1,2]. These micronutrients include iron (Fe), zinc (Zn), copper (Cu), manganese (Mn), molybdenum (Mo), tungsten (W), nickel (Ni), cobalt (Co), chromium (Cr), vanadium (V), and several other elements, which play vital roles in a wide variety of biological and chemical events. Some metals act as critical cofactors or as important structural components for different enzymes, whereas others can accept or donate electrons in various redox reactions, or regulate biological processes by facilitating the binding of molecules to corresponding receptors [3,4]. The major metalloid, selenium (Se), is also involved in several pivotal cellular processes in both prokaryotes and eukaryotes [5]. Deficiency and excess of these micronutrients may cause abnormalities in development and metabolism or even death [6].

Among all metals, Fe and Zn are indispensable for all or almost all living organisms [7,8]. The utilization of other trace metals is highly diverse and scattered. Cells require metal ions as cofactors for the assembly and activation of metalloproteins. Most metals are directly incorporated into

their binding sites on these proteins, whereas a small number of metals must form metal-containing cofactors or complexes before their insertion into target proteins, such as molybdopterin (or named Mo cofactor, Moco) for Mo and cobalamin (vitamin B_{12}) for Co [9,10]. Metalloproteins comprise a numerous and diverse group within the proteomes of organisms, which not only catalyze a remarkably wide range of important reactions but also play important structural and regulatory functions [11]. It has been suggested that approximately one third of all proteins require metals for their biological roles and almost half of all enzymes must associate with one or more metal ions [12,13]. The number of metalloprotein families varies greatly depending upon which metal they use. For instance, Zn is estimated to be used by several hundred protein families, whereas less than ten protein families belong to Ni-dependent metalloenzymes [14,15]. In addition, Se exerts its biological functions mainly as the 21st amino acid selenocysteine (Sec, encoded by the codon TGA) in selenoproteins [16]. The presence of metalloproteins necessitates a tight regulation of metal metabolism and homeostasis to maintain the appropriate metal concentration in the cell while avoiding toxicity, which includes transport, delivery, storage, detoxification, and efflux machineries [17,18].

In recent years, due to the rapid development of high-throughput sequencing technologies, the availability of genomic sequences for a huge number of organisms has opened the door for systematic analysis of the occurrence and evolutionary trends of metal utilization in nature. The main achievements of these studies lie in in silico identification and comparative genomics of metalloproteins across the tree of life. A variety of bioinformatic algorithms have been developed for the detection of metalloprotein genes. Moreover, large-scale analysis of the metalloproteome (the complete set of metalloproteins) in different organisms may help us to better understand the utilization and function of trace metals in biology. This review focuses on current computational approaches used for metalloprotein gene prediction as well as relevant tools and resources. We will also introduce recent advances in comparative genomics involving several essential metals to present an integrated and holistic picture of the distribution and evolution of metalloproteins and metalloproteomes in all domains of life.

2. Identification of Metalloprotein Genes and Related Resources

2.1. Homology-Based Identification of Known Metalloprotein Genes

In the past several decades, a large number of metalloproteins and their functions have been recognized. Therefore, metalloprotein genes may be identified by sequence homology to previously characterized metalloproteins. In general, representative metalloprotein sequences can be used to scan either genomic (using TBLASTN) or protein databases (using BLASTP). Significant hits aligned with the query sequences are potential members of known metalloprotein families. It has been reported that many metalloprotein families either possess both metal-dependent (i.e., strictly dependent on certain metals for their function) and metal-independent (i.e., metal-coordinating residues are partially or completely lost) forms or have evolved to use alternative metals in different organisms [19,20]. Thus, metalloproteins should be further verified by examining the conservation of known metal-binding residues in the corresponding families.

The homology-based method may also be applied to identify selenoprotein genes in genomic datasets. Considering that selenoprotein genes have several highly specific sequence-structural features [16], additional steps are needed for their efficient and correct recognition. Representative selenoprotein sequences derived from previously reported selenoprotein families are used to search against the genomic database using TBLASTN or similar approaches. Homologs containing a putative in-frame TGA codon that aligns the Sec residue in the query are initially considered as selenoprotein gene candidates. As the mechanism underlying Sec biosynthesis and its insertion into selenoproteins has been thoroughly investigated in recent years [21,22], the presence of an essential *cis*-acting structure called Sec insertion sequence (SECIS) element which is located in either 3′-untranslated regions (eukaryotes and archaea) or immediately downstream Sec-TGA codon (bacteria)

of selenoprotein mRNAs should then be examined by SECIS prediction tools (see below for details). Furthermore, due to the fact that most known selenoproteins have homologs in which Sec is replaced with a cysteine (Cys) residue [23], all selenoprotein candidates should be searched against the NCBI non-redundant protein database using BLASTP for the presence of conserved Cys-containing homologs. A computational pipeline named Selenoprofiles has been developed to correctly annotate selenoprotein genes belonging to known families in genomic sequences based on homology searches [24].

2.2. Methods for Prediction of Metal-Binding Sites and Novel Metalloprotein Genes

In recent years, with the exponential increase in the number of completely sequenced genomes, there is an urgent need to develop bioinformatic algorithms allowing the prediction of new metalloprotein genes or even the search for entire sets of metalloproteins. A number of computational tools and methods have been developed for the prediction of either metal-binding sites in proteins (particularly for Zn and Fe) or metalloprotein genes in different organisms. A list of most of these tools is shown in Table 1.

Several early studies proposed a general protocol taking advantage of known amino acids present in the metal-binding region of metalloproteins and other related sequence-structural features [25–28]. It combines the concepts of metal-binding domains (sequence profiles) and metal-binding patterns (residues coordinating the metal ion in 3D structures of metalloproteins), and examined their occurrence and correlation in all proteins of organisms. This strategy has been widely used for identification of metalloprotein genes and metalloproteomes for several essential metals (such as Zn, Cu, and Fe) in a variety of organisms from the three domains of life, which provides important clues to protein function and protein-ion interaction [25–31]. Based on this approach, a software package called RDGB was developed to identify putative homologs of the proteins of interest (such as metalloproteins) in any genome [32].

Zincfinder is a software for the prediction of Zn-binding proteins based on the support vector machine (SVM) method [33]. This predictor identified some unprecedented Zn-binding sites which were further validated through homology modeling. Another SVM- and homology-based algorithm was also developed to improve the performance of Zn-binding site prediction, which could provide higher precision at different recall levels compared to Zincfinder [34].

TEMSP [35] is a structure-based tool to predict Zn-binding sites in proteins. This method may be a significant improvement over existing methods in predicting Zn-binding residues from protein structures with minimum overpredictions. Moreover, TEMSP can reliably predict the Zn-bound local structures and is helpful for functional inference of Zn-binding proteins.

Zincidentifier [36] is an integrative framework which combines sequence, structural, and graph-theoretic network features, followed by a two-step feature selection using a random forest algorithm. This method can not only be an effective tool for accurately predicting Zn-binding sites using structure data, but also provide information on new properties for characterizing Zn-binding sites.

ZincExplorer [37] is a new hybrid algorithm for effective prediction of Zn-binding sites in protein sequences, which is composed of SVM-, cluster-, and template-based predictors. Furthermore, it could infer the interdependent relationships of the predicted Zn-binding sites that bind to the same Zn atom.

ZincBinder [38] is a newly developed SVM-based method for prediction of Zn-binding sites in proteins based on sequence profile information. This tool showed better performance than some existing methods.

ZINCCLUSTER [39] is a novel tool for identification of Zn-binding sites of proteins based on primary sequences or 3D structures. It can predict amino acids interacting with Zn or other metal ions, which is based on the occurrence of significant triplets found in the active sites of metalloproteins. It appears to have higher prediction accuracy than other computational tools such as FINDSITE-metal and SeqCHED.

Table 1. Computational tools for metal-binding site and metalloprotein prediction.

Name	Website	Related Metals	Main Algorithm	Reported Performance	Ref.
RDGB	http://www.cerm.unifi.it/home/research/genomebrowsing.html	Zn, Cu, Fe and other metals	Integration of tools for retrieval of protein domains and genome analysis	Accuracy: 89.6%, precision: 85.9%	[32]
Zincfinder	http://zincfinder.dsi.unifi.it	Zn	[a] SVM	[b] AURPC: 0.590 (local predictor) and 0.633 (gating network)	[33]
Zincpred	http://www.fos.su.se/~{}nanjiang/zincpred/download/	Zn	SVM- and homology-based algorithm	AURPC: 0.723 (local predictor) and 0.701 (gating network)	[34]
TEMSP	http://netalign.ustc.edu.cn/temsp/	Zn	Structure-based algorithm with a range of geometric criteria	[c] AUC: 0.945	[35]
Zincidentifier	http://protein.cau.edu.cn/zincidentifier/	Zn	A two-step feature selection method based on random forest algorithm	AUC: 0.955, AURPC: 0.829	[36]
ZincExplorer	http://protein.cau.edu.cn/ZincExplorer/	Zn	A combination of SVM-, cluster- and template-based predictors	AURPC: 0.907	[37]
ZincBinder	http://proteininformatics.org/mkumar/znbinder/	Zn	SVM model trained on PSSM-based input feature	AUC: 0.91	[38]
ZINCCLUSTER	http://www.metalactive.in	Zn	SVM-based Ligand Finder and Cluster Finder algorithms	[d] MCC: 0.798, F1-score: 0.801	[39]
ZnMachine	http://bioinformatics.fzu.edu.cn/znMachine.html	Zn	A combination of several intensively-trained machine learning models	AUC: 0.933 (SVM) and 0.910 (neural network)	[40]
HemeBIND	http://mleg.cse.sc.edu/hemeBIND/	Fe (heme)	SVM	MCC: 0.504, F1-score: 56.87%	[41]
SCMHBP	http://iclab.life.nctu.edu.tw/SCMHBP/	Fe (heme)	Based on a newly-developed scoring card method for predicting heme-binding proteins	Accuracy: 85.90%	[42]
Isph	http://biodev.extra.cea.fr/isph	Fe (Fe-S)	A penalized linear model based on machine learning approach	Precision: 87.9%, recall: 80.1% (extended model)	[43]

Table 1. *Cont.*

Name	Website	Related Metals	Main Algorithm	Reported Performance	Ref.
MetalPredator	http://metalweb.cerm.unifi.it/tools/metalpredator/	Fe (Fe-S)	Integration of existing domain-based methodology with a new approach for discovering metal-binding motifs	Precision: 85.2%, recall: 88.6%	[44]
MetSite	N/A	Fe, Zn, Cu, Mn, Ca, Mg	Artificial neural network	Mean accuracy: 94.5%	[45]
FINDSITE-metal	http://cssb.biology.gatech.edu/findsite-metal/	Fe, Zn, Cu, Mn, Ni, Co, Ca, Mg	Integration of structure/evolutionary information and machine learning approach (SVM)	Overall accuracy: 70–90%	[46]
SeqCHED	http://ligin.weizmann.ac.il/seqched	Fe, Zn, Cu, Mn, Ni, Co, Ca, Mg	A modification of the CHED algorithm and machine learning filters (decision tree classifier and SVM)	Sensitivity: 84–85%, selectivity: 82–93% (stringent filtration)	[47]
MetalDetector	http://metaldetector.dsi.unifi.it/v2.0/	Transition metals that use cysteine and histidine as ligands	A combination of different machine learning algorithms (SVM-HMM, structured-output SVM)	Precision: 60–79%, recall: 71–88%	[48]
MIB	http://bioinfo.cmu.edu.tw/MIB/	Ca, Cu, Fe, Mg, Mn, Zn, Cd, Ni, Hg, Co	Fragment transformation method	Overall accuracy: 92.9–95.1%	[49]
SECISearch3 and Seblastian	http://seblastian.crg.eu/, http://gladyshevlab.org/SelenoproteinPredictionServer	Se	Homology-based RNA motif finding and selenoprotein gene detection approach	Precision: 81.48–100%, recall: 33.33–100%	[50]
SelGenAmic	N/A	Se	Selenoprotein gene assembly algorithm based on the GenAmic approach used by geneid	N/A	[51]
bSECISearch	http://genomics.unl.edu/bSECISearch/	Se	An algorithm for prediction of bacterial selenoprotein genes based on a concensus RNA structural model	True positive rate: 96.5%	[52]

[a] SVM: support vector machine. [b] AURPC: area under the recall-precision curve. [c] AUC: area under the curve. [d] MCC: Matthew's correlation coefficient.

ZnMachine [40] is a very recently reported method for high-throughput Zn-binding residue prediction by combining several well-designed machine learning models. These models were constructed using different types of sequence profiles and effective features derived from multiple sources and were then combined using a linear equation. This tool demonstrated competitive performance and could become a complementary approach to the existing methods.

HemeBIND [41] is the first specialized SVM-based algorithm used to predict heme (an Fe-porphyrin complex) binding sites by combining several sequence and structural attributes that have distinctly different distributions between heme binding and non-binding residues. A significant improvement in prediction performance has been shown when compared to individual structure-based and sequence-based classifiers alone.

SCMHBP [42] is a scoring card method based tool for predicting and analyzing heme-binding proteins using propensity scores of amino acids and dipeptides. This approach performed well when compared to some typical methods such as SVM, decision tree, and Bayes classifiers, which may improve our understanding of heme-binding proteins.

Previously, an integrative penalized linear model based on machine learning approach has been reported for the prediction of iron-sulfur (Fe-S) proteins, which demonstrated higher sensitivity and a good level of specificity when compared to motif-based approaches [43]. MetalPredator [44], another web-based tool to predict Fe-S proteins from protein sequences, was also developed to process complete proteomes with high-precision prediction, providing new resources for an extensive exploration of Fe-S proteomes on a large-scale.

Besides the above mentioned methods focusing on specific metals, several bioinformatic tools have been developed for general metal-binding site prediction. For example, MetSite [45] identifies metal-binding sites using sequence profiles and approximate structural information from predicted models; FINDSITE-metal [46] is a threading-based algorithm designed to detect metal-binding sites in modeled protein structures, which combines structure/evolutionary information with machine learning results; SeqCHED [47] is developed for predicting metal binding sites from protein sequences based on remote homology detection between template and target protein; MetalDetector [48] can predict metal binding sites for proteins having novel folds and does not rely on 3D structure similarity; MIB [49] is a web server built to predict different types of metal-binding residues using the fragment transformation method, which also provides metal ion-docking visualization functions to generate the predicted metal ion-bound 3D structure after prediction. Very recently, deep neural networks have also been applied for prediction of metal-binding abilities of certain amino acids such as histidine (His) and Cys residues, offering a powerful technique for large-scale functional genomic screening [53].

It should be admitted that in silico identification of metalloprotein genes is a quite challenging issue as current knowledge gained from experimental work makes it almost impossible to identify the complete metalloproteomes in most organisms. Since different bioinformatic methods have different advantages and disadvantages in capturing key aspects of metalloproteins, an alternative approach is to apply a variety of complementary tools for the prediction of metalloprotein genes. It has also been suggested that many metalloproteins in prokaryotes and humans have not been identified yet [54,55], and more efficient and accurate tools are needed to search for additional metalloproteins and metal-binding features.

Compared to metalloprotein prediction for which highly precise method is still lacking, computational identification of selenoprotein genes and the set of all selenoproteins (selenoproteome) in different organisms is easier and more reliable due to several specific features for Sec insertion machinery [21,22]. Several bioinformatic tools have been widely used for selenoprotein gene prediction in different kingdoms [56]. SECISearch3 is currently the most efficient and widely used method for identification of eukaryotic SECIS elements, and is a key component of Seblastian which is a new method for the identification of selenoprotein genes in eukaryotic genomes [50]. SelGenAmic is a gene assembly algorithm for detecting selenoprotein genes from eukaryotic genomes and has been used to identify selenoproteins in metazoans [51]. The method bSECISearch is used for the prediction

of selenoprotein genes in bacteria based on a predefined sequence/structural pattern and several other constraints [52]. Additional SECIS-independent algorithms were also developed, which employ Cys-containing proteins in large protein databases to search against nucleotide sequence datasets for selenoprotein genes [57]. Both SECIS-dependent and SECIS-independent approaches have shown excellent performance in identifying all or almost all selenoprotein genes in different genomes.

2.3. Metalloprotein Databases

Integration of information about metalloproteins from various resources (such as public nucleotide/protein databases and literatures) provides the basis for further understanding their utilization and extensive roles. In recent years, several metalloprotein databases have been built up, such as MDB, Metal-MACiE, MetalPDB, and some other databases. A list of these databases is shown in Table 2.

Table 2. Metalloprotein databases.

Name	Website	Main Content	Ref.
MDB	http://metallo.scripps.edu	Metalloproteins and metal-binding sites in protein structures	[58]
Metal-MACiE	http://www.ebi.ac.uk/thornton-srv/databases/Metal_MACiE/home.html	All metalloenzymes annotated in the MACiE database	[59]
dbTEU	http://gladyshevlab.bwh.harvard.edu/trace_element/	Transporters and metalloproteins for Cu, Mo, Co, Ni, and Se in more than 700 organisms	[60]
Mespeus	http://mespeus.bch.ed.ac.uk/MESPEUS_10/	Experimentally established geometry of metal and protein interactions	[61]
MetalPDB	http://metalweb.cerm.unifi.it	Metal-binding sites detected in the 3D structures of biological macromolecules	[62]
SelenoDB	http://www.selenodb.org	Selenoprotein genes in at least 58 animal genomes	[63]
ZincBind	http://zincbind.bioinf.org.uk	All known Zn-binding sites from PDB	[64]

MDB (the Metalloprotein Database and Browser) [58] is a web-accessible resource for exploring metalloproteins, which provides quantitative information on protein metal-binding sites from structures available at the Protein Data Bank (PDB). MDB also offers tools for the examination of patterns in metal-binding sites, and for the prediction of metal sites from new protein structures.

Metal-MACiE [59] is a publicly available database that contains the information on the roles of metals in catalytic mechanisms of metalloenzymes. This database can be used to improve our knowledge of the chemistry underlying metal-dependent catalysis.

dbTEU (DataBase of Trace Element Utilization) [60] is a manually curated protein database, which contains ~16,500 protein sequences of all known transporters and metalloproteins for several trace elements (such as Cu, Mo, Co, Ni, and Se) in a variety of organisms from all three domains of life. It also offers interactive tools for searching and browsing organisms, metalloprotein families, and other related information.

Mespeus [61] is a useful database for investigating the variety of metal ion-protein interactions. It lists experimentally established geometry of metal protein interactions with a user-friendly interface, and can also contribute to the modeling process.

MetalPDB [62] is a valuable resource of metal-binding sites detected in the 3D structures of biological macromolecules, which represents such sites in the form of Minimal Functional Sites.

This database also provides extensive statistical information on structural aspects associated with individual metals, giving a better understanding of the diversity in biochemical roles of metals.

SelenoDB [63] contains full annotations of selenoprotein genes in approximately 60 animal genomes, including alternative transcripts and a worldwide catalog of genetic variations in human selenoprotein genes. It is an important resource for medical and evolutionary studies on Se.

ZincBind [64] is a newly developed database of all known Zn-binding sites from PDB. It contains more than 16,000 unique Zn-binding sites, which are then organized into groups and families. It also has a friendly web interface for users to browse, search and download data of interest. This resource is useful to researchers working on Zn-binding site/protein prediction and modeling, and will be expanded with new data in future.

3. Comparative Genomics of Metalloproteins and Metalloproteomes

Comparative genomics provides a powerful tool for investigating evolutionary changes of genes, pathways and other characteristics along various lineages [65]. Based on comparative genomic approaches, we may better understand the use of metalloproteins and metal-dependent processes in living organisms. However, identification and quantification of complete metalloproteomes for most metals is currently impossible [66,67]. Even so, analysis of the majority of them in genomic databases can still greatly improve our understanding of the utilization and function of metals and their variations across species during evolution. In addition, analysis of genes involved in metal uptake, homeostasis, and metal-containing cofactor biosynthesis may assist in identification of metal utilization trait (i.e., the ability to use certain metal) [67–69]. A general procedure for comparative genomics of metal utilization is shown in Figure 1. In the following sections, we mainly focus on metalloproteins and discuss recent progress on comparative genomic analyses of metalloproteins for several important metals.
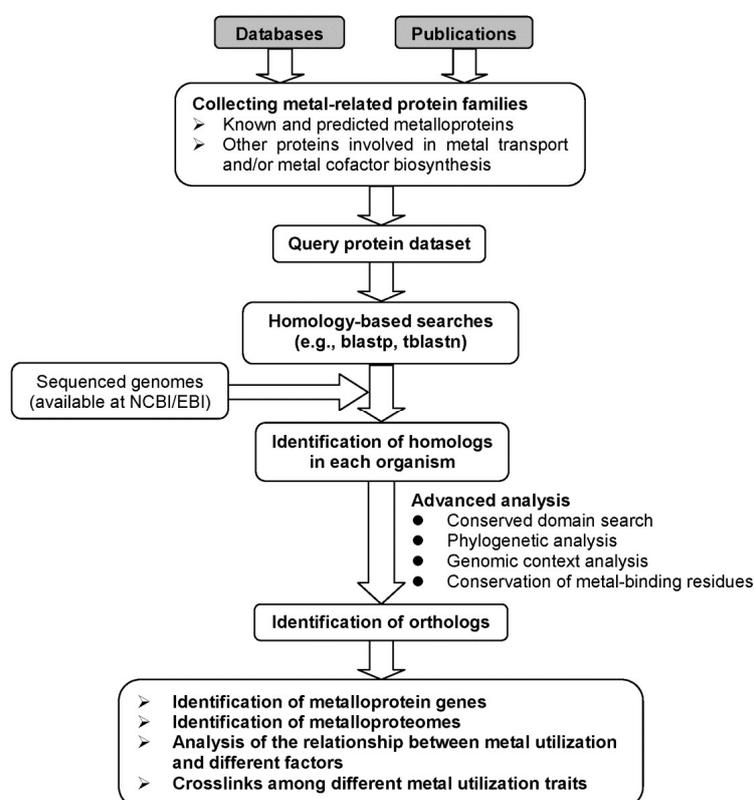
Figure 1. A general diagram for comparative genomic analysis of metal utilization.

3.1. Zinc and Iron

Zn and Fe are the two most commonly used trace metals in all organisms. A great number of proteins have been characterized or predicted to use one of the two metals (see references [25,26,30] and several web resources described above for a compiled list of Zn- or Fe-binding protein families). However, because of the widespread and complex use of the two metals, comparative analyses of the occurrence and evolutionary trends of their utilization are still very challenging to handle [69].

Zn is known to contribute to numerous biological processes in living systems, which is a key component present in hundreds of structural proteins, enzymes, transcription factors, and ribosomal proteins [14]. Previously, Zn proteomes (including both Zn-dependent proteins and some other proteins involved in Zn transport and homeostasis) have been predicted in a limited number of prokaryotic and eukaryotic organisms based on Zn-binding domains and patterns extracted from various databases [25,28,29,70]. In general, the number of Zn-binding proteins are positively correlated with the proteome size of an organism. Eukaryotes had a higher proportion (8%~10%) of Zn-binding proteins than prokaryotes (5%~6%). The majority of prokaryotic Zn proteins perform enzymatic catalysis (especially hydrolases) while the eukaryotic Zn proteome is mainly involved in both catalysis and transcription regulation of gene expression, suggesting that Zn-binding transcription factors have evolved to regulate more complex and diverse processes in higher organisms [28,29]. Another study analyzed two major groups of Zn proteins (Zn finger-containing proteins and Zn hydrolytic enzymes) in more than 800 organisms, which revealed that there is a correlation in their changes during evolution related to environmental change [71]. In recent years, comparative genomic approaches have been frequently used to investigate the distribution and diversification of certain Zn-dependent protein families, particularly Zn finger-containing transcription factors such as PRDM, Zic, and several other C2H2-Zn finger protein families, which provides a basis for further research on the origin, function, and evolutionary features of these proteins [72–75]. With an explosion in genomic resources and the rapidly expanding number of bioinformatic tools in the past decade, a more comprehensive analysis of Zn-dependent proteomes in all kingdoms of life is urgently needed.

Fe is the most abundant transition metal in cells and has a fundamental role in many metabolic processes, such as oxygen transport, electron transfer, nucleic acid synthesis, growth, and many important redox reactions [7]. Besides Fe ions, many proteins may use Fe in the form of heme or Fe-S clusters [7,11]. Due to the complexity and diversity of Fe utilization, it is very hard to identify the complete Fe-dependent proteomes; however, several studies have been performed aiming at the understanding of different groups of Fe-dependent proteins in various organisms. An early bioinformatic study investigated the occurrence of putative non-heme Fe-binding proteins in a small number of prokaryotes and eukaryotes, which demonstrated that extant organisms have inherited the majority of Fe proteome from the last universal common ancestor [26]. Compared to Zn proteome, the Fe proteome constituted a higher fraction of the proteome in archaea (7.1% on average) than in bacteria (3.9%) and in eukaryotes (1.1%). Another computational study compared the distribution of Fe-S proteins in more than 400 prokaryotic organisms with different life styles and found a strong relationship between environmental dioxygen levels and the usage of different Fe-S clusters [31]. Very recently, the complete human Fe proteome was systematically analyzed based on different types of Fe-containing cofactors [30]. About 2% of human genes encode Fe proteins (35%, 48%, and 17% for individual Fe ions, heme, and Fe-S clusters, respectively). Interestingly, genes encoding Fe proteins (especially Fe-S proteins) appeared to be more commonly related to pathologies than all other human genes, suggesting specific features of the physiological role of Fe. In addition, comparative genomic analyses were carried out for investigating Fe metabolism and homeostasis mechanisms in different organisms, such as cytosolic Fe-S cluster assembly machinery [76], heme biosynthesis and uptake machinery [77], and several other protein families involved in Fe transport and storage [78], which provide detailed insights into the composition and evolution of Fe metabolic network.

3.2. Copper

Cu is an important activator for several key enzymes participating in fundamental biological processes such as respiration, photosynthesis, and oxidative stress responses. A number of Cu-dependent proteins (cuproproteins) have been characterized in both prokaryotes and eukaryotes. The currently known cuproprotein families are shown in Table 3 (proteins involved in Cu transport and homeostasis are not included).

Table 3. List of known metalloprotein families for several metals in prokaryotes and eukaryotes.

Metal	Prokaryotes	Eukaryotes
Cu	Cytochrome c oxidase subunit I Cytochrome c oxidase subunit II Plastocyanin family Cu amine oxidase Cu-Zn superoxide dismutase Multicopper oxidase family Tyrosinase Azurin Rusticyanin Nitrosocyanin Nitrous oxide reductase Nitrite reductase NADH dehydrogenase 2 Particulate methane monooxygenase	Cytochrome c oxidase subunit I Cytochrome c oxidase subunit II Plastocyanin family Cu amine oxidase Cu-Zn superoxide dismutase Multicopper oxidase family Tyrosinase Galactose oxidase Hemocyanin Plantacyanin family Peptidylglycine α-hydroxylating monooxygenase Dopamine β-monooxygenase Cnx1G
Mo	Sulfite oxidase Xanthine oxidase Dimethylsulfoxide reductase MOSC-containing protein Fe-Mo-containing nitrogenase	Sulfite oxidase Xanthine oxidase MOSC-containing protein (mARC)
W	Aldehyde:ferredoxin oxidoreductase Certain members of dimethylsulfoxide reductase: Formate dehydrogenase and acetylene hydratase (obligately anaerobic bacteria) Formylmethanofuran dehydrogenase (methanogenic archaea)	N/A
Ni	Urease Ni-Fe hydrogenase Carbon monoxide dehydrogenase Superoxide dismutase SodN Acetyl-coenzyme A synthase/decarbonylase Methyl-coenzyme M reductase Lactate racemase	Urease
Co	Methylmalonyl-CoA mutase Isobutyryl-CoA mutase Ethylmalonyl-CoA mutase Glutamate mutase Methyleneglutarate mutase D-lysine 5,6-aminomutase Diol dehydratase Glycerol dehydratase Ethanolamine ammonia lyase B_{12}-dependent ribonucleotide reductase class II Methionine synthase Methyltetrahydromethanopterin:coenzyme M methyltransferase subunit A Other methyltransferases B_{12}-dependent reductive dehalogenase PceA/CprA LitR/CarH/CarA PpaA Epoxyqueuosine reductase	Methylmalonyl-CoA mutase B_{12}-dependent ribonucleotide reductase class II Methionine synthase

Comparative genomic studies have been previously carried out to analyze intrinsic features of different cuproteins or cuproteomes (the whole set of cuproteins) in various organisms [27,79–82]. Two early studies combined known Cu-binding domains and Cu-binding patterns to explore the occurrence of Cu proteins (including both cuproteins and some other proteins involved in Cu transport and homeostasis) in a limited number of sequenced genomes [27,79]. The proportion of Cu-binding proteins was small when compared to that of Zn or non-heme Fe proteins. Eukaryotes have expanded the Cu proteome inherited from the last common ancestor of all organisms by evolving new Cu domains and reusing old domains for novel functions.

Some other studies provide more detailed information about cuproteins in the three domains of life [67,80,81]. Cytochrome c oxidase subunits I (COX I) and II (COX II) are the most widely distributed cuproteins in prokaryotes (Figure 2A). Multicopper oxidase (MCO), Cu-Zn superoxide dismutase (Cu-Zn SOD) and plastocyanin families were also found in many prokaryotes, whereas the occurrence of tyrosinase, nitrosocyanin, Cu amine oxidase, and particulate methane monooxygenase seems to be quite limited. Except for cuproteins that were exclusively present in individual kingdoms (e.g., azurin in bacteria and rusticyanin in archaea), significant difference in the distribution of most cuproteins was observed between archaea and bacteria (Figure 2A). On the other hand, only half of prokaryotic cuprotein families could be found in eukaryotic organisms, and the latter have the capacity to evolve new cuproteins such as galactose oxidase, hemocyanin, and plantacyanin. MCO, COX I, COX II, and Cu-Zn SOD were the most abundant cuprotein families in eukaryotes, while the distribution of some cuproteins appeared to be phylum-specific, e.g., hemocyanin in arthropods and plantacyanin in land plants. A recent comparative analysis of the presence of hemocyanin in different myriapod species suggests that these proteins have divergent evolutionary patterns in different myriapod taxa [82]. Further analysis of prokaryotic cuproteomes revealed that larger cuproteomes were mainly present in Alphaproteobacteria, Betaproteobacteria and Euryarchaeota/Halobacteriales. The largest bacterial and archaeal cuproteomes reported to date were detected in several *Sinorhizobium* species (*S. medicae* and *S. meliloti*, 22 cuprotein genes) and *Haloarcula marismortui* (25 cuprotein genes), respectively [81]. In eukaryotes, land plants possessed the largest cuproteomes, especially *Oryza sativa* containing 78 cuprotein genes). It is interesting that larger cuproteomes were mainly found in organisms living in oxygen-rich environments, which is consistent with the idea that proteins evolved to use Cu following the oxygenation of the Earth [80–82]. Because previous studies relied only on a limited number of organisms, future research is needed to update the distribution and evolution of cuproteins/cuproteomes using a much wider range of sequenced genomes belonging to different clades.
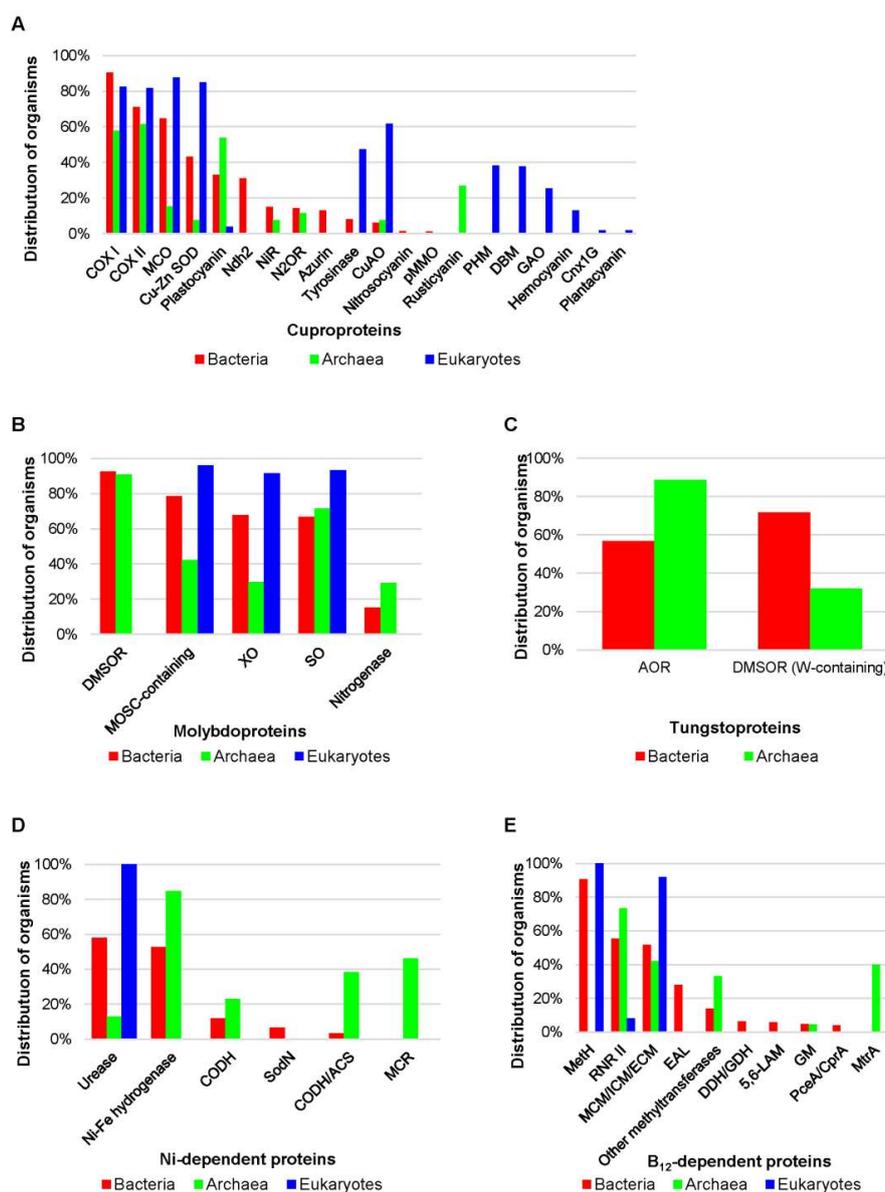
Figure 2. Distribution of metalloprotein families in the three domains of life. (**A**) Distribution of cuproproteins in Cu-utilizing organisms; (**B**) Distribution of molybdoproteins in Mo-utilizing organisms; (**C**) Distribution of tungstoproteins in W-utilizing organisms; (**D**) Distribution of Ni-dependent proteins in Ni-utilizing organisms; (**E**) Distribution of B_{12}-dependent proteins in Co-utilizing organisms. COX I, cytochrome c oxidase subunit I; COX II, cytochrome c oxidase subunit II; MCO, multicopper oxidase; Cu-Zn SOD, Cu-Zn superoxide dismutase; Ndh2, NADH dehydrogenase 2; NiR, nitrite reductase; N2OR, nitrous oxide reductase; CuAO, Cu amine oxidase; pMMO, particulate methane monooxygenase; PHM, peptidylglycine α-hydroxylating monooxygenase; DBM, dopamine β-monooxygenase; GAO, galactose oxidase; DMSOR, dimethylsulfoxide reductase; XO, xanthine oxidase; SO, sulfite oxidase; AOR, aldehyde:ferredoxin oxidoreductase; CODH, carbon monoxide dehydrogenase; SodN, Ni-containing superoxide dismutase; CODH/ACS, acetyl-coenzyme A synthase/decarbonylase; MCR, methyl-coenzyme M reductase; MetH, methionine synthase; RNR II, B_{12}-dependent ribonucleotide reductase class II; MCM, methylmalonyl-CoA mutase; ICM, isobutyryl-CoA mutase; ECM, ethylmalonyl-CoA mutase; EAL, ethanolamine ammonia lyase; DDH/GDH, diol/glycerol dehydratase; 5,6-LAM, D-lysine 5,6-aminomutase; GM, glutamate mutase; PceA/CprA, B_{12}-dependent reductive dehalogenase; MtrA, methyltetrahydromethanopterin:coenzyme M methyltransferase subunit A. Data used to generate this figure are available in the supplementary information of [81,83].

3.3. Molybdenum and Tungsten

Mo is required for the activity of a number of molybdoproteins that catalyze diverse reactions in the metabolism of carbon, nitrogen, and sulfur compounds [9]. With the exception of Fe-Mo-containing nitrogenase, Mo needs to be bound to a specific pyranopterin moiety to form Moco, an active compound at the active site of all molybdoproteins [84]. Some prokaryotes (such as hyperthermophilic archaea) use W to replace Mo, which is bound to the same pyranopterin to form tungstoproteins [85]. A list of known molybdoprotein and tungstoprotein families is shown in Table 3. Each family may contain a variety of enzymes [86,87]. It has been suggested that MOSC (Moco sulfurase C-terminal domain)-containing proteins are new members of the sulfite oxidase (SO) family due to similar structures for Mo-binding domains [86]; however, the lack of significant sequence similarity between them may challenge such an alternative classification approach [88].

Several comparative genomic studies have been conducted to explore the distribution and evolution of Mo utilization trait and molybdoproteins in all domains of life, which give preliminary indications of how this transition element is used by different organisms [81,89,90]. Very recently, the occurrence of all known molybdoprotein families in nearly 6000 sequenced prokaryotes and eukaryotes was analyzed, which presents a much more comprehensive view of the evolutionary trajectories of molybdoproteins in nature [83]. Dimethylsulfoxide reductase (DMSOR) is the most widespread molybdoprotein family in both archaea and bacteria, which was present in more than 90% Mo-utilizing organisms (Figure 2B). MOSC-containing protein, xanthine oxidase (XO), and SO families were also widespread in the majority of Mo-utilizing bacteria; however, most sequenced archaea do not have MOSC-containing protein and XO families. Several new domain fusions were detected for different members of DMSOR, SO, and XO in prokaryotes, providing valuable information for the inference of protein interactions and functions. The Fe-Mo-containing nitrogenase was only needed by a small number of bacteria and methanogenic archaea. On the other hand, MOSC-containing protein (or named mARC), SO, and XO are the three eukaryotic molybdeoprotein families, all of which were detected in almost all organisms that use Mo, indicating that they are all critical for maintaining the function of Mo in this kingdom. With regard to molybdoproteomes, many organisms in Actinobacteria and several subclasses of Proteobacteria were molybdoprotein-rich organisms (>20 molybdoprotein genes). To date, the largest molybdoproteome in bacteria was found in *Gordonibacter pamelaeae 7-10-1-b* (73 molybdoprotein genes, mostly belonging to the DMSOR family) [83]. In contrast, very few molybdoprotein-rich organisms were observed in archaea and eukaryotes. Further examination of the relationship between environmental factors and molybdoproteins revealed that the majority of molybdoprotein families and large molybdoproteomes are more frequently present in aerobic organisms, implying that oxygen has played a crucial role in the evolution of molybdoprotein genes [83].

Although it is still very difficult to distinguish between Mo and W utilization in different members of molybdoproteins due to quite similar physical-chemical and functional properties [91], it is worth mentioning that several attempts have been made to identify tungstoproteins from molybdoprotein families based on recent advances on tungstoproteins [83,92]. The currently known tungstoproteins include nearly all enzymes of the aldehyde:ferredoxin oxidoreductase (AOR) family and certain enzymes of the DMSOR family, including formate dehydrogenase and acetylene hydratase from strictly anaerobic bacteria and formylmethanofuran dehydrogenase from methanogenic archaea [92–94]. Preliminary analysis of tungstoproteins in prokaryotes revealed that AOR could be detected in the majority of W-utilizing prokaryotes while W-containing DMSOR proteins were present in most W-utilizing bacteria and a small number of archaea (mainly methanogens) (Figure 2C) [83]. These exploratory studies may provide the first global view of W utilization in prokaryotes.

3.4. Nickel and Cobalt

Ni is an essential cofactor for several enzymes that play critical roles in energy and nitrogen metabolism [15,95]. Some other Ni-containing proteins, such as glyoxalase I and acireductone dioxygenase, are not strictly Ni-dependent proteins which may bind alternative metals in different or

even same organisms [95]. Co is mainly used as a key component of cobalamin (or called vitamin B$_{12}$), which encompasses a group of closely related corrinoid compounds found in enzymes that mediate methyl transfer reactions, isomerase rearrangements, dehalogenation, and some other processes [96–98]. Moreover, Co is also detected in several non-corrin Co-containing enzymes in certain organisms, which may use other metals (such as Zn and Fe) to replace Co in many other organisms [99]. In this review, we only discuss strictly Ni-dependent and B$_{12}$-binding protein families which are shown in Table 3.

To our knowledge, only few comparative genomic studies have been conducted on Ni- or Co-dependent metalloproteins in a wide range of organisms from the three domains of life [81,100,101]. As prokaryotes use similar import systems for Ni and Co uptake [102,103], the utilization of the two trace metals could be highly correlated, which was supported by the observation that most prokaryotic organisms use both metals [81]. In bacteria, urease and methionine synthase (MetH) were the most frequently used Ni- and Co-dependent protein families, respectively (Figure 2D,E). However, they seemed to be rare or even absent in archaea, in which Ni-Fe hydrogenase and B$_{12}$-dependent ribonucleotide reductase class II (RNR II) were the most commonly used Ni and Co enzymes. Except for a small number of organisms (such as deltaproteobacteria and Methanosarcina species), most prokaryotes possessed no more than 5 Ni- and/or Co-dependent metalloprotein genes. The largest Ni-dependent proteome was previously reported in *Deltaproteobacterium MLMS-1* (16 Ni-binding protein genes, half were Ni-Fe hydrogenases) and the largest B$_{12}$-dependent proteome in *Dehalococcoides sp. CBDB1* (35 B$_{12}$-dependent protein genes, 32 were reductive dehalogenase CprA proteins) [81,100]. Another recent study analyzed the distribution of vitamin B$_{12}$ production pathway and a variety of B$_{12}$-dependent enzymes in over 11,000 bacterial species, which provides important information on B$_{12}$ utilization and its evolution in a much wider prokaryotic range [104]. Approximately 86% of the examined bacteria contained B$_{12}$-dependent enzyme families, most of which lacked the ability to synthesize B$_{12}$ and have to obtain this cofactor from exogenous sources. Proteobacteria and Bacteroidetes appeared to have larger numbers of B$_{12}$-dependent enzymes than others.

In contrast to prokaryotes, the utilization of Ni and Co is quite restricted in eukaryotes, and very few organisms utilize both metals [81]. Only one Ni-dependent enzyme (urease) and three B$_{12}$-dependent enzymes (methylmalonyl-CoA mutase, RNR II, and MetH) have been reported in this kingdom (Figure 2D,E). Urease and MetH were present in all Ni- and Co-utilizing eukaryotes, respectively. Analysis of Ni- and Co-dependent metalloproteomes did not reveal organisms that contained many of these proteins. Interestingly, compared to the majority of unicellular organisms that lack B$_{12}$-binding proteins, *Dictyostelium discoideum* and several *Phytophthora* species contained all the three known eukaryotic B$_{12}$-dependent enzymes, implying a more important role of B$_{12}$ cofactor in these organisms [81]. In the future, it is necessary to perform more comprehensive surveys on the two metals using newly generated genomic resources.

3.5. Selenium

Se is a metalloid trace element, which is essential for normal physiological functions in humans, animals, and many other organisms [5,105]. It mainly occurs in the form of Sec, which is a key component of selenoproteins involved in numerous enzymatic reactions, such as redox homeostasis, thyroid hormone metabolism, anti-inflammatory actions, and reproduction [106,107]. The mechanism of Sec biosynthesis and its incorporation into proteins has been elucidated in both prokaryotes and eukaryotes [21,22]. So far, a significant number of selenoproteins have been reported in various organisms from bacteria to mammals, many of which were identified using reliable bioinformatic approaches [50–52,57,108,109]. Table 4 lists the majority of known and putative selenoproteins. Although the functions of most selenoproteins are not known and could only be inferred by sequence homology, it is very likely that most of them play important roles in antioxidation and detoxification [106].

Table 4. List of known selenoprotein families.

Prokaryotes	Eukaryotes
Known selenoproteins	*Known selenoproteins in mammals*
Formate dehydrogenase alpha subunit	Deiodinase (DIO) family: DIO1, DIO2, and DIO3
Selenophosphate synthetase	Glutathione peroxidase (GPX) family: GPX1, GPX2, GPX3, GPX4, and GPX6
Coenzyme F420-reducing hydrogenase alpha subunit	Thioredoxin reductase (TXNRD) family: TXNRD1, TXNRD2, and TXNRD3
Coenzyme F420-reducing hydrogenase delta subunit	Methionine sulfoxide reductase B1
Methylviologen-reducing hydrogenase alpha subunit	Selenoprotein F
Glycine reductase selenoprotein A	Selenoprotein H
Glycine reductase selenoprotein B	Selenoprotein I
Proline reductase	Selenoprotein K
Heterodisulfide reductase alpha subunit	Selenoprotein M
Methionine-S-sulfoxide reductase	Selenoprotein N
Peroxiredoxin (Prx)Thioredoxin (Trx)	Selenoprotein O
Glutaredoxin (Grx)	Selenoprotein P
Arsenite S-adenosylmethyltransferase	Selenoprotein S
	Selenoprotein T
Predicted selenoproteins:	Selenoprotein V
Thiol:disulfide isomerase-like protein	Selenoprotein W
Thiol:disulfide interchange protein	Selenophosphate synthetase 2
HesB-like	
Deiodinase-like	*Other known selenoproteins:*
Glutathione peroxidase-like	Methionine-S-sulfoxide reductase
Selenoprotein W-like	Protein disulfide isomerase
Fe-S oxidoreductase	Selenoprotein J
DsbA-like	Selenoprotein L
DsrE-like	Selenoprotein U
DsbG-like	Selenoprotein E
AhpD-like	SAM-dependent methyltransferase
Arsenate reductase	
Molybdopterin biosynthesis protein MoeB	*Predicted selenoproteins:*
Glutathione S-transferase	Prx-like protein
COG0737 UshA	Trx-fold protein
OsmC-like	Membrane selenoprotein MSP
Rhodanase-related protein	SelTryp
Sulfurtransferase COG2897	Other hypothetical proteins
Cation-transporting ATPase, E1-E2 family	
Methylated-DNA-protein-cysteine methyltransferase	
UGSC-containing protein	
CMD domain containing protein	
Organic mercuric lyase MerB2	
Predicted redox-active disulfide protein 2	
Prx-like/Trx-like/Grx-like and Trx-fold proteins	
Other hypothetical selenoproteins	

Previously, several computational and comparative genomic approaches have been carried out to investigate the distribution and evolution of Se metabolic pathways and selenoproteins in a large number of prokaryotic organisms and selected environmental samples [23,81,110–116], which provide detailed information on how this element is selectively used by proteins and organisms from different kingdoms. An early work analyzed the Sec biosynthetic pathway and known selenoproteins in several hundred bacterial and archaeal genomes, and found that only one-fourth of the examined organisms have selenoprotein genes. Most selenoprotein-rich organisms belong to Deltaproteobacteria and Clostridia [81]. Recently, a much more extensive evaluation has been conducted on Se metabolism and selenoproteins in bacteria by analyzing more than 5200 genomes, which generated the largest map of Se utilization in this kingdom [114]. More than 60 selenoprotein families/subfamilies could be

detected in bacteria. Formate dehydrogenase alpha subunit and selenophosphate synthetase were the most widespread bacterial selenoprotein families (Figure 3). A new selenoprotein-rich phylum Synergistetes and additional selenoprotein-rich organisms have also been identified. The largest bacterial selenoproteome was found in *Syntrophobacter fumaroxidans*, a syntrophic propionate-oxidizing deltaproteobacterium containing 39 selenoprotein genes [81,114]. Although both aerobic and anaerobic organisms could use Sec, the fact that most selenoprotein-rich organisms (78.3%) are obligate or facultative anaerobic suggests a somewhat stronger correlation between evolution of selenoprotein genes and low oxygen level [114].
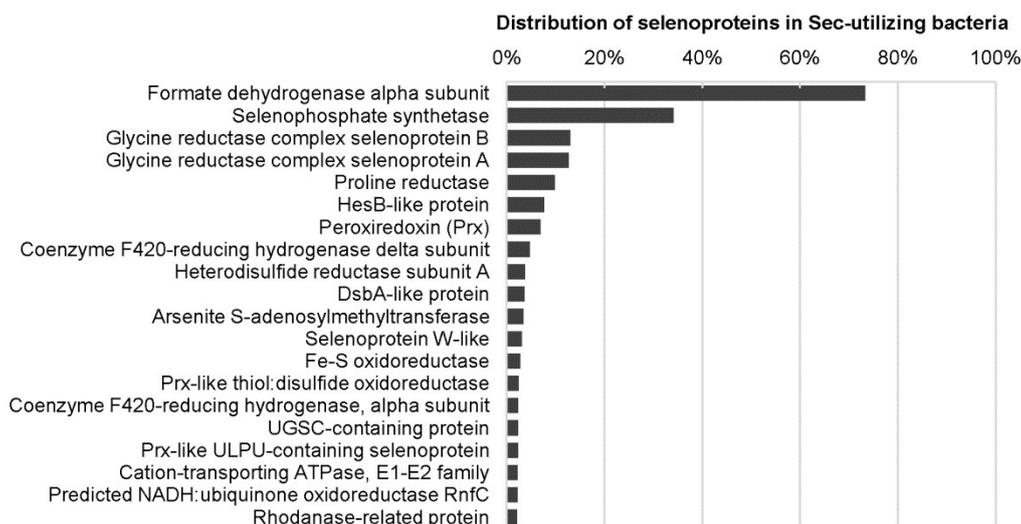
Figure 3. Distribution of the top 20 selenoproteins in Sec-utilizing bacteria. Data used to generate this figure can be found in [114].

In archaea, selenoprotein genes were only detected in a small number of organisms belonging to three phyla: Methanococcales, Methanopyrales, and Lokiarchaeota [81,116–118]. Compared to bacteria which contain a variety of known or predicted selenoprotein families, only nine selenoprotein families have been discovered in archaea, most of which are involved in methanogenesis [117]. The archaeal selenoproteomes show a relatively narrow size distribution (7~12 selenoproteins). Lokiarchaeota, a novel archaeal phylum and the closest archaeal relative to eukaryotes, was reported to have the largest archaeal selenoproteome (at least 12 selenoprotein genes) [118]. Further analysis of Lokiarchaeota selenoprotein genes suggests that this archaeon may serve as an intermediate form between the typical archaeal and eukaryotic Sec biosynthesis systems, providing new clues for the origin and evolution of the Sec utilization trait.

More efforts have been made to explore the distribution and evolution of selenoproteins in eukaryotes [81,119–124]. Several early comparative studies demonstrated that many selenoprotein families, such as glutathione peroxidases (GPXs), thioredoxin reductases (TXNRDs), and selenophosphate synthase 2 (SEPHS2) are shared between single-cell eukaryotes (such as green algae and many protists) and vertebrates, implying that the majority of eukaryotic selenoproteins originated from the ancestors of current eukaryotes and have been preserved throughout evolution [81,119–121]. However, massive and independent selenoprotein gene loss events (either loss of selenoprotein genes or replacement of Sec with Cys residue) were observed in different lineages such as fungi, land plants, nematodes, and some other organisms [119,120]. The size of eukaryotic selenoproteomes varies greatly between species. With the exception of mammals, aquatic organisms (such as algae and fish) generally have larger selenoproteomes than terrestrial ones (such as insects and nematodes). Although parallel loss of Sec utilization was observed in different groups of algae [122], the largest eukaryotic selenoproteome was described in the harmful pelagophyte alga *Aureococcus anophagefferens* (containing at least 59 selenoprotein genes) [125]. In animals, amphioxus was found to have the most

abundant and diverse selenoproteins (containing 40 selenoprotein genes) [121]. Further investigation of selenoproteins in sequenced vertebrates defined the ancestral vertebrate (28 selenoproteins) and mammalian (25 selenoproteins) selenoproteomes, and reconstructed their evolutionary history [120]. For example, mammalian TXNRD1 and TXNRD3 were found to have evolved from an ancestral glutaredoxin-domain-containing enzyme, and selenoprotein V and GPX6 appeared at the root of placental mammals by duplications of selenoprotein W and GPX3, respectively. By evaluating the potential forces for selenoprotein gain or loss and for substitutions between Sec and Cys residues in different vertebrate clades, it was proposed that the strength of natural selection on selenoprotein genes is distinct between land vertebrates and teleost fishes, suggesting that Se availability has shaped the evolution of vertebrates [124]. In addition, selenoprotein P (SELENOP), the only human selenoprotein with multiple Sec residues, has been suggested to function as a genetic marker of Se utilization in animals (i.e., its number of Sec residues correlates with the selenoproteome size) [126]. A recent study showed that SELENOP genes are present across metazoan lineages with highly variable numbers of Sec-TGA codons, ranging from a single Sec residue in certain insects to up to 132 in bivalve mollusks, implying a highly dynamic evolutionary process of this selenoprotein [127]. Very recently, it was also reported that Sec could be encoded by several early-branching fungal phyla, which provides new insights into the evolution of Sec utilization in fungi [128].

Theoretically, comparative genomic approaches could be applied to study the metabolism of all trace metals and to identify the corresponding metalloproteins. However, due to limited knowledge about metal-binding sites and related properties for several other metals such as Mn, Cr, and V, metalloproteins that are strictly dependent on them remain poorly defined. For example, Mn is known to serve as a substitute for some other metals (e.g., Zn and Mg) in the active sites of numerous enzymes, resulting in the difficulty to distinguish Mn-dependent proteins from other metalloproteins [129]. Only a few proteins have been reported to bind Cr or V in certain organisms, including Cr-containing oligopeptide chromodulin and V-containing vanabins and haloperoxidases [130,131]; however, it is unclear whether these proteins are strictly dependent on the corresponding metal in other organisms. Therefore, comparative analysis of these metalloproteins seems to be a hard task and needs to be solved in the future.

4. Conclusions

Bioinformatics offers a powerful tool for studying metalloproteins, metalloproteomes, and their evolutionary trends in nature. Many of these studies have been devoted to the identification of metalloproteins based on known metal-binding motifs, patterns, and profiles. To date, it is still very difficult to identify complete metalloproteomes for almost all metals. Nevertheless, recent progress on bioinformatic research of metalloproteins, especially comparative genomics of several essential trace metals, has provided significant insights into the general principles of metal utilization and evolution across the three domains of life. In this review, we highlight recent studies that used various computational strategies and methods to predict metalloproteins and to investigate the distribution and evolution of metalloproteins/metalloproteomes in a wide range of organisms. In the future, with the dramatic increase in the number of sequenced genomes and improved computational techniques for identifying more metalloprotein genes, bioinformatics and comparative genomic approaches will play an even more important role in better understanding of metal utilization and function in biology.

References

1. Zoroddu, M.A.; Aaseth, J.; Crisponi, G.; Medici, S.; Peana, M.; Nurchi, V.M. The essential metals for humans: A brief overview. *J. Inorg. Biochem.* **2019**, *195*, 120–129. [CrossRef] [PubMed]

2. Freeland-Graves, J.H.; Sanjeevi, N.; Lee, J.J. Global perspectives on trace element requirements. *J. Trace Elem. Med. Biol.* **2015**, *31*, 135–141. [CrossRef]

3. Mertz, W. Review of the scientific basis for establishing the essentiality of trace elements. *Biol. Trace Elem. Res.* **1998**, *66*, 185–191. [CrossRef] [PubMed]

4. Nordberg, M.; Nordberg, G.F. Trace element research-historical and future aspects. *J. Trace Elem. Med. Biol.* **2016**, *38*, 46–52. [CrossRef] [PubMed]

5. Roman, M.; Jitaru, P.; Barbante, C. Selenium biochemistry and its role for human health. *Metallomics* **2014**, *6*, 25–54. [CrossRef]

6. Van Gossum, A.; Neve, J. Trace element deficiency and toxicity. *Curr. Opin. Clin. Nutr. Metab. Care* **1998**, *1*, 499–507. [CrossRef]

7. Sánchez, M.; Sabio, L.; Gálvez, N.; Capdevila, M.; Dominguez-Vera, J.M. Iron chemistry at the service of life. *IUBMB Life* **2017**, *69*, 382–388. [CrossRef]

8. King, J.C. Zinc: An essential but elusive nutrient. *Am. J. Clin. Nutr.* **2011**, *94*, 679Ss–684Ss. [CrossRef]

9. Magalon, A.; Mendel, R.R. Biosynthesis and insertion of the molybdenum cofactor. *EcoSal Plus* **2015**. [CrossRef]

10. Giedyk, M.; Goliszewska, K.; Gryko, D. Vitamin B$_{12}$ catalysed reactions. *Chem. Soc. Rev.* **2015**, *44*, 3391–3404. [CrossRef]

11. Degtyarenko, K. Bioinorganic motifs: Towards functional classification of metalloproteins. *Bioinformatics* **2000**, *16*, 851–864. [CrossRef] [PubMed]

12. Waldron, K.J.; Rutherford, J.C.; Ford, D.; Robinson, N.J. Metalloproteins and metal sensing. *Nature* **2009**, *460*, 823–830. [CrossRef]

13. Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G.L.; Thornton, J.M. Metal ions in biological catalysis: From enzyme databases to general principles. *J. Biol. Inorg. Chem.* **2008**, *13*, 1205–1218. [CrossRef] [PubMed]

14. Maret, W. Zinc and the zinc proteome. *Met. Ions Life Sci.* **2013**, *12*, 479–501. [PubMed]

15. Alfano, M.; Cavazza, C. Structure, function, and biosynthesis of nickel-dependent enzymes. *Protein Sci.* **2020**, *29*, 1071–1089. [CrossRef]

16. Zoidis, E.; Seremelis, I.; Kontopoulos, N.; Danezis, G.P. Selenium-dependent antioxidant enzymes: Actions and properties of selenoproteins. *Antioxidants (Basel)* **2018**, *7*, 66. [CrossRef]

17. Chandrangsu, P.; Rensing, C.; Helmann, J.D. Metal homeostasis and resistance in bacteria. *Nat. Rev. Microbiol.* **2017**, *15*, 338–350. [CrossRef]

18. Martinez-Finley, E.J.; Chakraborty, S.; Fretham, S.J.; Aschner, M. Cellular transport and homeostasis of essential and nonessential metals. *Metallomics* **2012**, *4*, 593–605. [CrossRef]

19. Zhang, Y.; Gladyshev, V.N. Comparative genomics of trace element dependence in biology. *J. Biol. Chem.* **2011**, *286*, 23623–23629. [CrossRef]

20. Sukdeo, N.; Clugston, S.L.; Daub, E.; Honek, J.F. Distinct classes of glyoxalase I: Metal specificity of the Yersinia pestis, Pseudomonas aeruginosa and Neisseria meningitidis enzymes. *Biochem. J.* **2004**, *384*, 111–117. [CrossRef]

21. Bulteau, A.L.; Chavatte, L. Update on selenoprotein biosynthesis. *Antioxid. Redox Signal.* **2015**, *23*, 775–794. [CrossRef]

22. Gonzalez-Flores, J.N.; Shetty, S.P.; Dubey, A.; Copeland, P.R. The molecular biology of selenocysteine. *Biomol. Concepts* **2013**, *4*, 349–365. [CrossRef]

23. Zhang, Y.; Romero, H.; Salinas, G.; Gladyshev, V.N. Dynamic evolution of selenocysteine utilization in bacteria: A balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome Biol.* **2006**, *7*, R94. [CrossRef]

24. Mariotti, M.; Guigó, R. Selenoprofiles: Profile-based scanning of eukaryotic genome sequences for selenoprotein genes. *Bioinformatics* **2010**, *26*, 2656–2663. [CrossRef] [PubMed]

25. Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. Zinc through the three domains of life. *J. Proteome Res.* **2006**, *5*, 3173–3178. [CrossRef] [PubMed]

26. Andreini, C.; Banci, L.; Bertini, I.; Elmi, S.; Rosato, A. Non-heme iron through the three domains of life. *Proteins* **2007**, *67*, 317–324. [CrossRef]

27. Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. Occurrence of copper proteins through the three domains of life: A bioinformatic approach. *J. Proteome Res.* **2008**, *7*, 209–216. [CrossRef]

28. Andreini, C.; Bertini, I.; Rosato, A. Metalloproteomes: A bioinformatic approach. *Acc. Chem. Res.* **2009**, *42*, 1471–1479. [CrossRef]

29. Andreini, C.; Bertini, I. A bioinformatics view of zinc enzymes. *J. Inorg. Biochem.* **2012**, *111*, 150–156. [CrossRef]

30. Andreini, C.; Putignano, V.; Rosato, A.; Banci, L. The human iron-proteome. *Metallomics* **2018**, *10*, 1223–1231. [CrossRef]

31. Andreini, C.; Rosato, A.; Banci, L. The relationship between environmental dioxygen and iron-sulfur proteins explored at the genome level. *PLoS ONE* **2017**, *12*, e0171279. [CrossRef]

32. 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. *J. Chem. Inf. Model.* **2011**, *51*, 730–738. [CrossRef]

33. Passerini, A.; Andreini, C.; Menchetti, S.; Rosato, A.; Frasconi, P. Predicting zinc binding at the proteome level. *BMC Bioinform.* **2007**, *8*, 39. [CrossRef] [PubMed]

34. Shu, N.; Zhou, T.; Hovmöller, S. Prediction of zinc-binding sites in proteins from sequence. *Bioinformatics* **2008**, *24*, 775–782. [CrossRef]

35. Zhao, W.; Xu, M.; Liang, Z.; Ding, B.; Niu, L.; Liu, H.; Teng, M. Structure-based de novo prediction of zinc-binding sites in proteins of unknown function. *Bioinformatics* **2011**, *27*, 1262–1268. [CrossRef] [PubMed]

36. Zheng, C.; Wang, M.; Takemoto, K.; Akutsu, T.; Zhang, Z.; Song, J. An integrative computational framework based on a two-step random forest algorithm improves prediction of zinc-binding sites in proteins. *PLoS ONE* **2012**, *7*, e49716. [CrossRef] [PubMed]

37. Chen, Z.; Wang, Y.; Zhai, Y.F.; Song, J.; Zhang, Z. ZincExplorer: An accurate hybrid method to improve the prediction of zinc-binding sites from protein sequences. *Mol. Biosyst.* **2013**, *9*, 2213–2222. [CrossRef]

38. Srivastava, A.; Kumar, M. Prediction of zinc binding sites in proteins using sequence derived information. *J. Biomol. Struct. Dyn.* **2018**, *36*, 4413–4423. [CrossRef]

39. Ajitha, M.; Sundar, K.; Arul Mugilan, S.; Arumugam, S. Development of METAL-ACTIVE SITE and ZINCCLUSTER tool to predict active site pockets. *Proteins* **2018**, *86*, 322–331. [CrossRef]

40. Yan, R.; Wang, X.; Tian, Y.; Xu, J.; Xu, X.; Lin, J. Prediction of zinc-binding sites using multiple sequence profiles and machine learning methods. *Mol. Omics* **2019**, *15*, 205–215. [CrossRef]

41. Liu, R.; Hu, J. HemeBIND: A novel method for heme binding residue prediction by combining structural and sequence information. *BMC Bioinform.* **2011**, *12*, 207. [CrossRef] [PubMed]

42. Liou, Y.F.; Charoenkwan, P.; Srinivasulu, Y.; Vasylenko, T.; Lai, S.C.; Lee, H.C.; Chen, Y.H.; Huang, H.L.; Ho, S.Y. SCMHBP: Prediction and analysis of heme binding proteins using propensity scores of dipeptides. *BMC Bioinform.* **2014**, *15* (Suppl. 16), S4. [CrossRef]

43. Estellon, J.; Ollagnier de Choudens, S.; Smadja, M.; Fontecave, M.; Vandenbrouck, Y. An integrative computational model for large-scale identification of metalloproteins in microbial genomes: A focus on iron-sulfur cluster proteins. *Metallomics* **2014**, *6*, 1913–1930. [CrossRef] [PubMed]

44. Valasatava, Y.; Rosato, A.; Banci, L.; Andreini, C. MetalPredator: A web server to predict iron-sulfur cluster binding proteomes. *Bioinformatics* **2016**, *32*, 2850–2852. [CrossRef]

45. Sodhi, J.S.; Bryson, K.; McGuffin, L.J.; Ward, J.J.; Wernisch, L.; Jones, D.T. Predicting metal-binding site residues in low-resolution structural models. *J. Mol. Biol.* **2004**, *342*, 307–320. [CrossRef]

46. Brylinski, M.; Skolnick, J. FINDSITE-metal: Integrating evolutionary information and machine learning for structure-based metal-binding site prediction at the proteome level. *Proteins* **2011**, *79*, 735–751. [CrossRef] [PubMed]

47. Levy, R.; Edelman, M.; Sobolev, V. Prediction of 3D metal binding sites from translated gene sequences based on remote-homology templates. *Proteins* **2009**, *76*, 365–374. [CrossRef] [PubMed]

48. Passerini, A.; Lippi, M.; Frasconi, P. MetalDetector v2.0: Predicting the geometry of metal binding sites from protein sequence. *Nucleic Acids Res.* **2011**, *39*, W288–W292. [CrossRef]

49. Lin, Y.F.; Cheng, C.W.; Shih, C.S.; Hwang, J.K.; Yu, C.S.; Lu, C.H. MIB: Metal ion-binding site prediction and docking server. *J. Chem. Inf. Model.* **2016**, *56*, 2287–2291. [CrossRef]

50. Mariotti, M. SECISearch3 and seblastian: In-silico tools to predict SECIS elements and selenoproteins. *Methods Mol. Biol.* **2018**, *1661*, 3–16.

51. Jiang, L.; Liu, Q. SelGenAmic: An algorithm for selenoprotein gene assembly. *Methods Mol. Biol.* **2018**, *1661*, 29–39. [PubMed]

52. Zhang, Y.; Gladyshev, V.N. An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. *Bioinformatics* **2005**, *21*, 2580–2589. [CrossRef] [PubMed]

53. Haberal, İ.; Oğul, H. Prediction of protein metal binding sites using deep neural networks. *Mol. Inform.* **2019**, *38*, e1800169. [CrossRef] [PubMed]

54. Cvetkovic, A.; Menon, A.L.; Thorgersen, M.P.; Scott, J.W.; Poole, F.L., 2nd; Jenney, F.E., Jr.; Lancaster, W.A.; Praissman, J.L.; Shanmukh, S.; Vaccaro, B.J.; et al. Microbial metalloproteomes are largely uncharacterized. *Nature* **2010**, *466*, 779–782. [CrossRef]

55. Azia, A.; Levy, R.; Unger, R.; Edelman, M.; Sobolev, V. Genome-wide computational determination of the human metalloproteome. *Proteins* **2015**, *83*, 931–939. [CrossRef]

56. Santesmasses, D.; Mariotti, M.; Gladyshev, V.N. Bioinformatics of selenoproteins. *Antioxid. Redox Signal.* **2020**, in press. [CrossRef]

57. Kryukov, G.V.; Gladyshev, V.N. The prokaryotic selenoproteome. *EMBO Rep.* **2004**, *5*, 538–543. [CrossRef]

58. Castagnetto, J.M.; Hennessy, S.W.; Roberts, V.A.; Getzoff, E.D.; Tainer, J.A.; Pique, M.E. MDB: The metalloprotein database and browser at the scripps research institute. *Nucleic Acids Res.* **2002**, *30*, 379–382. [CrossRef]

59. Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G.L.; Thornton, J.M. Metal-MACiE: A database of metals involved in biological catalysis. *Bioinformatics* **2009**, *25*, 2088–2089. [CrossRef]

60. Zhang, Y.; Gladyshev, V.N. dbTEU: A protein database of trace element utilization. *Bioinformatics* **2010**, *26*, 700–702. [CrossRef]

61. Harding, M.M.; Hsin, K.Y. Mespeus—A database of metal interactions with proteins. *Methods Mol. Biol.* **2014**, *1091*, 333–342. [PubMed]

62. Putignano, V.; Rosato, A.; Banci, L.; Andreini, C. MetalPDB in 2018: A database of metal sites in biological macromolecular structures. *Nucleic Acids Res.* **2018**, *46*, D459–D464. [CrossRef] [PubMed]

63. Romagné, F.; Santesmasses, D.; White, L.; Sarangi, G.K.; Mariotti, M.; Hübler, R.; Weihmann, A.; Parra, G.; Gladyshev, V.N.; Guigó, R.; et al. SelenoDB 2.0: Annotation of selenoprotein genes in animals and their genetic diversity in humans. *Nucleic Acids Res.* **2014**, *42*, D437–D443. [CrossRef] [PubMed]

64. Ireland, S.M.; Martin, A.C.R. ZincBind-the database of zinc binding sites. *Database (Oxford)* **2019**, *2019*, baz006. [CrossRef]

65. Miller, W.; Makova, K.D.; Nekrutenko, A.; Hardison, R.C. Comparative genomics. *Annu. Rev. Genomics Hum. Genet.* **2004**, *5*, 15–56. [CrossRef]

66. Maret, W. Metalloproteomics, metalloproteomes, and the annotation of metalloproteins. *Metallomics* **2010**, *2*, 117–125. [CrossRef]

67. Zhang, Y.; Gladyshev, V.N. Comparative genomics of trace elements: Emerging dynamic view of trace element utilization and function. *Chem. Rev.* **2009**, *109*, 4828–4861. [CrossRef]

68. Tzou, W.S.; Chu, Y.; Lin, T.Y.; Hu, C.H.; Pai, T.W.; Liu, H.F.; Lin, H.J.; Cases, I.; Rojas, A.; Sanchez, M.; et al. Molecular evolution of multiple-level control of heme biosynthesis pathway in animal kingdom. *PLoS ONE* **2014**, *9*, e86718. [CrossRef]

69. Zhang, Y.; Ying, H.; Xu, Y. Comparative genomics and metagenomics of the metallomes. *Metallomics* **2019**, *11*, 1026–1043. [CrossRef]

70. Bertini, I.; Decaria, L.; Rosato, A. The annotation of full zinc proteomes. *J. Biol. Inorg. Chem.* **2010**, *15*, 1071–1078. [CrossRef]

71. Decaria, L.; Bertini, I.; Williams, R.J. Zinc proteomes, phylogenetics and evolution. *Metallomics* **2010**, *2*, 706–709. [CrossRef] [PubMed]

72. Aruga, J.; Hatayama, M. Comparative genomics of the Zic family genes. *Adv. Exp. Med. Biol.* **2018**, *1046*, 3–26. [PubMed]

73. Vervoort, M.; Meulemeester, D.; Béhague, J.; Kerner, P. Evolution of Prdm genes in animals: Insights from comparative genomics. *Mol. Biol. Evol.* **2016**, *33*, 679–696. [CrossRef] [PubMed]

74. Najafabadi, H.S.; Garton, M.; Weirauch, M.T.; Mnaimneh, S.; Yang, A.; Kim, P.M.; Hughes, T.R. Non-base-contacting residues enable kaleidoscopic evolution of metazoan C2H2 zinc finger DNA binding. *Genome Biol.* **2017**, *18*, 167. [CrossRef] [PubMed]

75. Salih, H.; Odongo, M.R.; Gong, W.; He, S.; Du, X. Genome-wide analysis of cotton C2H2-zinc finger transcription factor family and their expression analysis during fiber development. *BMC Plant. Biol.* **2019**, *19*, 400. [CrossRef] [PubMed]

76. Tsaousis, A.D.; Gentekaki, E.; Eme, L.; Gaston, D.; Roger, A.J. Evolution of the cytosolic iron-sulfur cluster assembly machinery in Blastocystis species and other microbial eukaryotes. *Eukaryot. Cell* **2014**, *13*, 143–153. [CrossRef]

77. Cavallaro, G.; Decaria, L.; Rosato, A. Genome-based analysis of heme biosynthesis and uptake in prokaryotic systems. *J. Proteome Res.* **2008**, *7*, 4946–4954. [CrossRef]

78. Hayrapetyan, H.; Siezen, R.; Abee, T.; Nierop Groot, M. Comparative genomics of iron-transporting systems in bacillus cereus strains and impact of iron sources on growth and biofilm formation. *Front. Microbiol.* **2016**, *7*, 842. [CrossRef]

79. Decaria, L.; Bertini, I.; Williams, R.J. Copper proteomes, phylogenetics and evolution. *Metallomics* **2011**, *3*, 56–60. [CrossRef]

80. Ridge, P.G.; Zhang, Y.; Gladyshev, V.N. Comparative genomic analyses of copper transporters and cuproproteomes reveal evolutionary dynamics of copper utilization and its link to oxygen. *PLoS ONE* **2008**, *3*, e1378. [CrossRef]

81. Zhang, Y.; Gladyshev, V.N. General trends in trace element utilization revealed by comparative genomic analyses of Co, Cu, Mo, Ni, and Se. *J. Biol. Chem.* **2010**, *285*, 3393–3405. [CrossRef] [PubMed]

82. Scherbaum, S.; Hellmann, N.; Fernández, R.; Pick, C.; Burmester, T. Diversity, evolution, and function of myriapod hemocyanins. *BMC Evol. Biol.* **2018**, *18*, 107. [CrossRef] [PubMed]

83. Peng, T.; Xu, Y.; Zhang, Y. Comparative genomics of molybdenum utilization in prokaryotes and eukaryotes. *BMC Genomics* **2018**, *19*, 691. [CrossRef] [PubMed]

84. Mendel, R.R. The molybdenum cofactor. *J. Biol. Chem.* **2013**, *288*, 13165–13172. [CrossRef] [PubMed]

85. Moura, J.J.; Bernhardt, P.V.; Maia, L.B.; Gonzalez, P.J. Molybdenum and tungsten enzymes: From biology to chemistry and back. *J. Biol. Inorg. Chem.* **2015**, *20*, 181–182. [CrossRef] [PubMed]

86. Hille, R.; Hall, J.; Basu, P. The mononuclear molybdenum enzymes. *Chem. Rev.* **2014**, *114*, 3963–4038. [CrossRef] [PubMed]

87. Leimkühler, S.; Iobbi-Nivol, C. Bacterial molybdoenzymes: Old enzymes for new purposes. *FEMS Microbiol. Rev.* **2016**, *40*, 1–18. [CrossRef]

88. Gladyshev, V.N.; Zhang, Y. Abundance, ubiquity and evolution of molybdoenzymes. In *Molybdenum and Tungsten Enzymes: Biochemistry*, 1st ed.; Hille, R., Schulzke, C., Kirk, M.L., Eds.; The Royal Society of Chemistry: Cambridge, UK, 2016; pp. 81–99.

89. Llamas, A.; Tejada-Jiménez, M.; Fernández, E.; Galván, A. Molybdenum metabolism in the alga Chlamydomonas stands at the crossroad of those in Arabidopsis and humans. *Metallomics* **2011**, *3*, 578–590. [CrossRef]

90. Zhang, Y.; Rump, S.; Gladyshev, V.N. Comparative genomics and evolution of molybdenum utilization. *Coord. Chem. Rev.* **2011**, *255*, 1206–1217. [CrossRef]

91. Bevers, L.; Hagedoorn, P.; Hagen, W. The bioinorganic chemistry of tungsten. *Coord. Chem. Rev.* **2009**, *253*, 269–290. [CrossRef]

92. Pushie, M.J.; Cotelesage, J.J.; George, G.N. Molybdenum and tungsten oxygen transferases–and functional diversity within a common active site motif. *Metallomics* **2014**, *6*, 15–24. [CrossRef] [PubMed]

93. Niks, D.; Hille, R. Molybdenum- and tungsten-containing formate dehydrogenases and formylmethanofuran dehydrogenases: Structure, mechanism, and cofactor insertion. *Protein Sci.* **2019**, *28*, 111–122. [CrossRef] [PubMed]

94. Boll, M.; Einsle, O.; Ermler, U.; Kroneck, P.M.; Ullmann, G.M. Structure and function of the unusual tungsten enzymes acetylene hydratase and class II benzoyl-coenzyme a reductase. *J. Mol. Microbiol. Biotechnol.* **2016**, *26*, 119–137. [CrossRef]

95. Boer, J.L.; Mulrooney, S.B.; Hausinger, R.P. Nickel-dependent metalloenzymes. *Arch. Biochem. Biophys.* **2014**, *544*, 142–152. [CrossRef]

96. Kräutler, B. Vitamin B_{12}: Chemistry and biochemistry. *Biochem. Soc. Trans.* **2005**, *33*, 806–810. [CrossRef]

97. Smith, A.D.; Warren, M.J.; Refsum, H. Vitamin B_{12}. *Adv. Food Nutr. Res.* **2018**, *83*, 215–279.

98. Takano, H.; Mise, K.; Hagiwara, K.; Hirata, N.; Watanabe, S.; Toriyabe, M.; Shiratori-Takano, H.; Ueda, K. Role and function of LitR, an adenosyl B$_{12}$-bound light-sensitive regulator of bacillus megaterium QM B1551, in regulation of carotenoid production. *J. Bacteriol.* **2015**, *197*, 2301–2315. [CrossRef]

99. Kobayashi, M.; Shimizu, S. Cobalt proteins. *Eur. J. Biochem.* **1999**, *261*, 1–9. [CrossRef] [PubMed]

100. Zhang, Y.; Rodionov, D.A.; Gelfand, M.S.; Gladyshev, V.N. Comparative genomic analyses of nickel, cobalt and vitamin B$_{12}$ utilization. *BMC Genomics* **2009**, *10*, 78. [CrossRef]

101. Zheng, K.; Ngo, P.D.; Owens, V.L.; Yang, X.P.; Mansoorabadi, S.O. The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea. *Science* **2016**, *354*, 339–342. [CrossRef]

102. Mulrooney, S.B.; Hausinger, R.P. Nickel uptake and utilization by microorganisms. *FEMS Microbiol. Rev.* **2003**, *27*, 239–261. [CrossRef]

103. Eitinger, T.; Suhr, J.; Moore, L.; Smith, J.A. Secondary transporters for nickel and cobalt ions: Theme and variations. *BioMetals* **2005**, *18*, 399–405. [CrossRef] [PubMed]

104. Shelton, A.N.; Seth, E.C.; Mok, K.C.; Han, A.W.; Jackson, S.N.; Haft, D.R.; Taga, M.E. Uneven distribution of cobamide biosynthesis and dependence in bacteria predicted by comparative genomics. *ISME J.* **2019**, *13*, 789–804. [CrossRef] [PubMed]

105. Ying, H.; Zhang, Y. Systems biology of selenium and complex disease. *Biol. Trace Elem. Res.* **2019**, *192*, 38–50. [CrossRef]

106. Steinbrenner, H.; Speckmann, B.; Klotz, L.O. Selenoproteins: Antioxidant selenoenzymes and beyond. *Arch. Biochem. Biophys.* **2016**, *595*, 113–119. [CrossRef]

107. Hariharan, S.; Dharmaraj, S. Selenium and selenoproteins: It's role in regulation of inflammation. *Inflammopharmacology* **2020**, *28*, 667–695. [CrossRef]

108. Li, M.; Huang, Y.; Xiao, Y. A method for identification of selenoprotein genes in archaeal genomes. *Genom. Proteom. Bioinform.* **2009**, *7*, 62–70. [CrossRef]

109. Santesmasses, D.; Mariotti, M.; Guigó, R. Selenoprofiles: A computational pipeline for annotation of selenoproteins. *Methods Mol. Biol.* **2018**, *1661*, 17–28.

110. Castellano, S.; Lobanov, A.V.; Chapple, C.; Novoselov, S.V.; Albrecht, M.; Hua, D.; Lescure, A.; Lengauer, T.; Krol, A.; Gladyshev, V.N.; et al. Diversity and functional plasticity of eukaryotic selenoproteins: Identification and characterization of the SelJ family. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16188–16193. [CrossRef]

111. Zhang, Y.; Gladyshev, V.N. Trends in selenium utilization in marine microbial world revealed through the analysis of the global ocean sampling (GOS) project. *PLoS Genet.* **2008**, *4*, e1000095. [CrossRef]

112. Lin, J.; Peng, T.; Jiang, L.; Ni, J.Z.; Liu, Q.; Chen, L.; Zhang, Y. Comparative genomics reveals new candidate genes involved in selenium metabolism in prokaryotes. *Genome Biol. Evol.* **2015**, *7*, 664–676. [CrossRef]

113. Cravedi, P.; Mori, G.; Fischer, F.; Percudani, R. Evolution of the selenoproteome in helicobacter pylori and epsilonproteobacteria. *Genome Biol. Evol.* **2015**, *7*, 2692–2704. [PubMed]

114. Peng, T.; Lin, J.; Xu, Y.Z.; Zhang, Y. Comparative genomics reveals new evolutionary and ecological patterns of selenium utilization in bacteria. *ISME J.* **2016**, *10*, 2048–2059. [CrossRef]

115. Miller, W.G.; Yee, E.; Lopes, B.S.; Chapman, M.H.; Huynh, S.; Bono, J.L.; Parker, C.T.; Strachan, N.J.C.; Forbes, K.J. Comparative genomic analysis identifies a campylobacter clade deficient in selenium metabolism. *Genome Biol. Evol.* **2017**, *9*, 1843–1858. [CrossRef] [PubMed]

116. Santesmasses, D.; Mariotti, M.; Guigó, R. Computational identification of the selenocysteine tRNA (tRNASec) in genomes. *PLoS Comput. Biol.* **2017**, *13*, e1005383. [CrossRef] [PubMed]

117. Rother, M.; Quitzke, V. Selenoprotein synthesis and regulation in Archaea. *Biochim. Biophys. Acta Gen. Subj.* **2018**, *1862*, 2451–2462. [CrossRef] [PubMed]

118. Mariotti, M.; Lobanov, A.V.; Manta, B.; Santesmasses, D.; Bofill, A.; Guigó, R.; Gabaldón, T.; Gladyshev, V.N. Lokiarchaeota marks the transition between the archaeal and eukaryotic selenocysteine encoding systems. *Mol. Biol. Evol.* **2016**, *33*, 2441–2453. [CrossRef] [PubMed]

119. Lobanov, A.V.; Fomenko, D.E.; Zhang, Y.; Sengupta, A.; Hatfield, D.L.; Gladyshev, V.N. Evolutionary dynamics of eukaryotic selenoproteomes: Large selenoproteomes may associate with aquatic life and small with terrestrial life. *Genome Biol.* **2007**, *8*, R198. [CrossRef]

120. Mariotti, M.; Ridge, P.G.; Zhang, Y.; Lobanov, A.V.; Pringle, T.H.; Guigo, R.; Hatfield, D.L.; Gladyshev, V.N. Composition and evolution of the vertebrate and mammalian selenoproteomes. *PLoS ONE* **2012**, *7*, e33066. [CrossRef]

121. Jiang, L.; Ni, J.; Liu, Q. Evolution of selenoproteins in the metazoan. *BMC Genomics* **2012**, *13*, 446. [CrossRef]

122. Liang, H.; Wei, T.; Xu, Y.; Li, L.; Kumar Sahu, S.; Wang, H.; Li, H.; Fu, X.; Zhang, G.; Melkonian, M.; et al. Phylogenomics provides new insights into gains and losses of selenoproteins among archaeplastida. *Int. J. Mol. Sci.* **2019**, *20*, 3020. [CrossRef]

123. Mariotti, M.; Santesmasses, D.; Capella-Gutierrez, S.; Mateo, A.; Arnan, C.; Johnson, R.; D'Aniello, S.; Yim, S.H.; Gladyshev, V.N.; Serras, F.; et al. Evolution of selenophosphate synthetases: Emergence and relocation of function through independent duplications and recurrent subfunctionalization. *Genome Res.* **2015**, *25*, 1256–1267. [CrossRef] [PubMed]

124. Sarangi, G.K.; Romagné, F.; Castellano, S. Distinct patterns of selection in selenium-dependent genes between land and aquatic vertebrates. *Mol. Biol. Evol.* **2018**, *35*, 1744–1756. [CrossRef] [PubMed]

125. Gobler, C.J.; Berry, D.L.; Dyhrman, S.T.; Wilhelm, S.W.; Salamov, A.; Lobanov, A.V.; Zhang, Y.; Collier, J.L.; Wurch, L.L.; Kustka, A.B.; et al. Niche of harmful alga Aureococcus anophagefferens revealed through ecogenomics. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 4352–4357. [CrossRef]

126. Lobanov, A.V.; Hatfield, D.L.; Gladyshev, V.N. Reduced reliance on the trace element selenium during evolution of mammals. *Genome Biol.* **2008**, *9*, R62. [CrossRef]

127. Baclaocos, J.; Santesmasses, D.; Mariotti, M.; Bierła, K.; Vetick, M.B.; Lynch, S.; McAllen, R.; Mackrill, J.J.; Loughran, G.; Guigó, R.; et al. Processive recoding and metazoan evolution of selenoprotein P: Up to 132 UGAs in molluscs. *J. Mol. Biol.* **2019**, *431*, 4381–4407. [CrossRef]

128. Mariotti, M.; Salinas, G.; Gabaldón, T.; Gladyshev, V.N. Utilization of selenocysteine in early-branching fungal phyla. *Nat. Microbiol.* **2019**, *4*, 759–765. [CrossRef]

129. Khrustalev, V.V.; Barkovsky, E.V.; Khrustaleva, T.A. Magnesium and manganese binding sites on proteins have the same predominant motif of secondary structure. *J. Theor. Biol.* **2016**, *395*, 174–185. [CrossRef]

130. Vincent, J.B. The biochemistry of chromium. *J. Nutr.* **2000**, *130*, 715–718. [CrossRef]

131. Ueki, T.; Adachi, T.; Kawano, S.; Aoshima, M.; Yamaguchi, N.; Kanamori, K.; Michibata, H. Vanadium-binding proteins (vanabins) from a vanadium-rich ascidian Ascidia sydneiensis samea. *Biochim. Biophys. Acta* **2003**, *1626*, 43–50. [CrossRef]