**General Trends in Trace Element Utilization Revealed by Comparative Genomic Analyses of Co, Cu, Mo, Ni, and Se**

Yan Zhang and Vadim N. Gladyshev

*From the Division of Genetics, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts 02115*

Trace elements are used by all organisms and provide proteins with unique coordination and catalytic and electron transfer properties. Although many trace element-containing proteins are well characterized, little is known about the general trends in trace element utilization. We carried out comparative genomic analyses of copper, molybdenum, nickel, cobalt (in the form of vitamin B_{12}), and selenium (in the form of selenocysteine) in 747 sequenced organisms at the following levels: (i) transporters and transport-related proteins, (ii) cofactor biosynthesis traits, and (iii) trace element-dependent proteins. Few organisms were found to utilize all five trace elements, whereas many symbionts, parasites, and yeasts used only one or none of these elements. Investigation of metalloproteomes and selenoproteomes revealed examples of increased utilization of proteins that use copper in land plants, cobalt in Dehalococcoides and Dicyostelium, and selenium in fish and algae, whereas nematodes were found to have great diversity of copper transporters. These analyses also characterized trace element metabolism in common model organisms and suggested new model organisms for experimental studies of individual trace elements. Mismatches in the occurrence of user proteins and corresponding transport systems revealed deficiencies in our understanding of trace element biology. Biological interactions among some trace elements were observed; however, such links were limited, and trace elements generally had unique utilization patterns. Finally, environmental factors, such as oxygen requirement and habitat, correlated with the utilization of certain trace elements. These data provide insights into the general features of utilization and evolution of trace elements in the three domains of life.

Biological trace elements refer to chemical elements required in minute quantities by an organism (1–3) and include chromium, cobalt, copper, iodine, iron, manganese, molybdenum, nickel, selenium, tungsten, vanadium, zinc, and probably several other elements. These trace elements function in different ways. Some are essential components of enzymes where they directly interact with substrates and often facilitate their conversion to products; some donate or accept electrons in reactions of reduction and oxidation; some structurally stabilize biological molecules; and some control biological processes by facilitating the binding of molecules to receptor sites on cell membranes (4).

Most biological trace elements are metals. Among them, Fe and Zn are thought to be the most abundant transition metal ions that are used by all organisms (5, 6). Other metals, such as Mn, Cu, Mo, Ni, and Co, are utilized by various metalloproteins in a wide range of organisms in all three domains of life. Additionally, Se, the major metalloid micronutrient, plays roles in various redox and metabolic processes (7–9).

The ability of the cell to maintain a specific trace element within a certain homeostatic range is mainly dependent on the processes of uptake, storage, and excretion. The relative importance of these processes varies among trace elements and among organisms. High affinity uptake systems for some trace elements have been characterized in both prokaryotes and eukaryotes. Among them, the ATP-binding cassette (ABC) transporters are the most frequently used uptake systems for metals, such as ZnuABC for Zn (10), MntABC for Mn (11), ModABC for Mo/W (12), and NikABCDE for Ni (13). Non-ABC transporters were also reported, such as ZupT for Zn and other divalent metal cations (14), MntH for Mn and Fe (15), CtaA and Ctr1 for Cu (16, 17), and NiCoT for Ni/Co (18). Besides high affinity transporters, metal ions could be transported via general cation influx systems, although the efficiency of such processes may be low (19, 20). Se-specific transporters have not been identified, and the uptake (in the form of selenate or selenite) is thought to be maintained up by the sulfate transport system (21). On the other hand, excessive uptake of certain metals (e.g. Cu) may result in metal overload and toxicity. Storage of trace elements in inactive sites or forms and excretion/export systems are essential mechanisms that prevent inappropriate amounts of reactive trace elements in the cell (e.g., metallothioneins for heavy metal binding/detoxification, and CopA transporter for Cu export (22–24)). In addition, release of a trace element from a storage site may be important to avoid deficiency.

The utilization of trace elements in cells is complex and not completely understood. Most metals are directly used as cofactors inserted into proteins involved in various metabolic pathways, whereas Mo and Co are mainly used in the form of molybdopterin cofactor (Moco) and cobalamin.
Comparative Genomics of Trace Elements

(vitamin B_{12}), respectively (25, 26). The number of metallo-protein families also varies from less than 10 (Ni-binding proteins) to more than 300 (Zn-binding proteins) (27–29). The use of Se is quite different from other trace elements as it is mainly used in the form of co-translationally inserted selenocysteine (Sec, the 21st amino acid, encoded by UGA codon), which is found in a number of selenoproteins in the three domains of life (30, 31).

In the past decade, a rapid increase in sequence information provided an opportunity to investigate the occurrence and evolutionary dynamics of numerous biochemical pathways that an organism utilizes, including trace element utilization. Genomewide comparative genomic approaches that study the relationship between genome structure and function across various groups of organisms have been used for the analysis of several trace elements, which advanced our understanding of general utilization and evolutionary trends in the use of trace elements in both prokaryotes and eukaryotes (32–36).

Recently, we have analyzed the utilization of five trace elements, Cu, Mo, Ni, Co, and Se, in a subset of sequenced organisms (37–41). These studies uncovered new features for each of these trace elements. In this study, we report an advanced comparative genomic analysis of these five trace elements in a common broad set of sequenced organisms. The main reason we are interested in these elements is that they are utilized by many organisms in all three domains of life but are characterized by a limited number of user proteins. Therefore, analyses of their utilization patterns provide important information about the function and evolution of trace elements. Our data combine the information from high affinity transport systems, user proteins, and other features, such as components involved in cofactor biosynthesis pathways and trace element homeostasis. We further used this information to examine interactions between elements and their common and unique features. We also propose new model organisms to study trace elements.

EXPERIMENTAL PROCEDURES

Genomic Sequence Resources—A total of 540 bacterial, 47 archaeal, and 160 eukaryotic organisms were analyzed (747 organisms, as of Nov. 2008). A list of fully sequenced prokaryotic and eukaryotic genomes can be found on the NCBI website (www.ncbi.nlm.nih.gov). Only one strain was selected for analysis for each species.

Identification of Transporters, Cofactor Biosynthesis Pathways, and Metalloenzymes for Mo, Cu, Ni, and Co—In previous studies, we initially characterized the occurrence of utilization traits and metalloproteins separately for each of the four transition metals (39–41). A complete list of known high affinity transporters, cofactors, and metalloproteins is shown in supplemental Table S1. First, to identify metal-specific transporters, we used representative sequences of reported transporters as seeds to search for homologous sequences in organisms via TBLASTN (42) with an e-value <0.1. Distant homologs were further identified using iterative TBLASTN and PSI-BLAST (43) with default parameters. Orthologous proteins were then defined using the Clusters of Orthologous Groups (COG) Database and bidirectional best hits (44, 45). It should be noted that some transporters shown in supplemental Table S1 are similar to metal-unrelated homologs present within the same COG (e.g. NikABCDE belongs to a large family of peptide/nickel ABC transporters that also includes transporters involved in dipeptide and oligopeptide uptake (34)), and some are similar to each other (e.g. in prokaryotes, each Ni/Co transporter family could be divided into Ni-specific, Co-specific, and unclear subgroups (41)). Therefore, additional analyses, such as conservation of metal-binding ligands, gene neighborhood, and phylogenetic analysis, were further used to help identify orthologs from numerous homologs (40, 41). For example, if a transporter was either documented as a metal transporter or is located adjacent to genes encoding metal-dependent enzymes or cofactor (e.g. B_{12} or Moco) biosynthesis pathways, it was considered as a true transporter for the corresponding metal. Other orthologs of detected transporter families that did not have strong implication for their function were then predicted by phylogenetic analyses. However, it should be noted that the functional diversity of Ni/Co transporters is complex, and Ni- and Co-specific subgroups were previously found to be scattered in various branches of the tree of life (34, 41). Thus, when analyzing Ni/Co transporters, subgroups specific for Ni or Co were only predicted based on either previous reports or gene neighborhoods (i.e. if a transporter gene was located adjacent to genes encoding Ni-dependent enzymes, the repressor NikR or B_{12} biosynthesis proteins, it was considered as a predicted Ni- or Co-specific transporter). Other members of detected Ni/Co transporter families were considered as proteins with unassigned function (41).

The occurrence of Moco and vitamin B_{12} biosynthesis pathways was verified by the presence of most of the key components involved in the Moco or B_{12} biosynthetic pathway (25, 46). Members of metal-dependent protein families were identified using a similar strategy. In this study, X-dependent proteins (X represents a certain metal or a metal-containing cofactor) refer to strict X-binding proteins. We excluded proteins that may bind more than one metal in different organisms. For example, we only considered B_{12}-dependent enzymes as “Co-dependent” proteins because of the unspecificity of metal utilization and limited distribution of non-corrin Co-binding enzymes (47). Some proteins may contain several subunits, some of which are metal-independent. Therefore, only metal-binding domain-containing proteins were considered as user proteins. Finally, utilization of a metal was verified by the requirement for the presence of at least one metal-specific transporter, or a cofactor biosynthetic pathway (for Mo and Co), or at least one metal-dependent protein.

Identification of the Sec Utilization Trait and Selenoproteins—As described in previous studies (37, 48), we used E. coli Sec synthase, Sec-specific elongation factor, and selenophosphate synthetase, as well as human Sec synthase, Sec-specific elongation factor, and selenophosphate synthetase 2 sequences to characterize the Sec-decoding trait in prokaryotes and eukaryotes, respectively. Orthologous proteins were identified as described above. The presence of the Sec-decoding trait was verified by an additional requirement for the presence of at least one known selenoprotein gene. Representative sequences derived from previously reported selenoprotein families (see supplemental Table S1 for a list of prokaryotic and eukaryotic selenoprotein
families) were used to search against the genomic data base for additional selenoproteins with TBLASTN cutoff e-value 1.0. The presence of a putative Sec-encoding UGA codon and a 5′-untranslated region (in archaea and eukaryotes) or downstream (in bacteria) selenocysteine insertion sequence (SECIS) element, which is essential for Sec insertion into selenoproteins, were analyzed using SECISearch and bSECISearch programs, respectively (49, 50).

**Multiple Sequence Alignment and Phylogenetic Analysis**—To investigate the distribution of trace element utilization traits in various phyla, we adopted a phylogenetic tree developed by Ciccarelli et al. (51), which is based on concatenation of 31 orthologs occurring in 191 sequenced genomes. Multiple sequence alignments were performed using ClustalW (52) with default parameters. Ambiguous alignments in highly variable regions were excluded. The resulting multiple alignments were then checked for conservation of functional residues and manually edited if needed. Phylogenetic trees were constructed by PHYLIP programs (53). Pairwise distance matrices were calculated to estimate the expected amino acid replacements per position, and neighbor-joining trees were obtained. To evaluate robustness of the trees, we performed maximum likelihood with PHYML (54) using default parameters and likelihood test. If inconsistent topologies were obtained, a third program MrBayes (55), a Bayesian estimation of phylogeny, was used. The final phylogenetic tree was then manually refined for visualization purposes.

**Protein Sequence Resource**—Protein sequences of all transporters, transport-related proteins, metalloproteins, and selenoproteins collected in this study are available in the supplemental material.

**RESULTS AND DISCUSSION**

Previous analyses of utilization of individual trace elements, Cu, Mo, Ni, Co, and Se, showed that these five biological trace elements are widely utilized, but they are not used by all organisms (34, 37–41, 56). This is an ideal situation for applying comparative genomic approaches. In this study, we expanded the analyses of individual trace elements to more than 700 genomes and analyzed them for occurrence of the following: (i) transporters and supporting proteins (repressors and chaperones), (ii) cofactor biosynthesis machinery; and (iii) trace element-containing proteins. An overall view of utilization of Cu, Mo, Ni, Co, and Se in the three domains of life is shown in Fig. 1 and supplemental Figs. S1–S3.

**General Patterns of Trace Element Utilization**—The requirement for trace elements in prokaryotes and at the same time their scattered occurrence in various phyla illustrate a dynamic nature of their utilization. In bacteria, Cu appeared to be used by more organisms (432 genomes; 80% of all sequenced) than other elements. The other three metals, Ni, Co (in the form of B12), and Mo, were also widely used, and approximately half of sequenced bacteria utilized all four metals (Fig. 1 and supplemental Fig. S1). However, in some phyla, all or almost all organisms lost the ability to utilize several metals, such as Chlamydiae and Mollicutes. Most organisms in these phyla were host-associated (either obligate intracellular symbionts or parasites), consistent with the idea that parasitic lifestyle may result in reduced utilization of metals. In contrast to the wide utilization of metals, Se (in the form of Sec) was utilized by a quarter of organisms mostly belonging to Deltaproteobacteria, Epsilonproteobacteria, and several other phyla. Some phyla in which the majority of organisms utilized the four metals, such as Cyanobacteria, appeared to have completely lost the ability to use Sec. Se utilization is thought to be an ancient trait that once was common to all or almost all organisms (37). Only 94 (17%) organisms were identified that use all five trace elements examined in this study.

Similar patterns of occurrence of trace element utilization were observed in archaea (Fig. 1 and supplemental Fig. S2). Essentially all archaea utilized Mo, Co, and Ni, whereas very few (six organisms in Methanococcales and Methanopyrales) used Sec. Only three Methanococcus species utilized all five trace elements. Cu utilization was only detected in half of sequenced archaea, probably because Cu is mainly used by aerobic organisms, whereas most sequenced archaea are anaerobes (39). It appears that the utilization of metals is consistent with ancient traits that have been and remain common to microbes, whereas Se utilization became scattered and diminished in prokaryotes.

In contrast to prokaryotes, metal utilization was highly variable in eukaryotes (Fig. 1 and supplemental Fig. S3). Almost all sequenced organisms utilized Cu, suggesting a uniformly essential nature of this metal in this domain of life. We also identified 105 (66%) Mo-utilizing organisms, including all animals, land plants, algae, certain fungi (such as Pezizomycotina), and stramenopiles. In contrast, other fungi (such as Saccharomyces and Schizosaccharomyces) and many protists (such as parasites and free-living ciliates) lacked the ability to use Mo as they had no known Mo transporters and Moco biosynthesis pathways nor molybdoenzymes. Our results are consistent with previous data (40) and show that although Mo utilization is widespread, many protozoa, especially parasites, have lost the ability to utilize this metal. Consistent with previous observations (41), the majority of Ni-utilizing organisms were fungi (except Saccharomyces), green algae, and land plants, whereas B12-utilizing organisms were mostly animals (except insects). Almost all unicellular eukaryotes lost the Co(B12) utilization trait, whereas higher eukaryotes, including animals, lost the Ni utilization trait. Approximately half of sequenced eukaryotes used Sec, including all animals (except some selenoproteinless insects such as Tribolium castaneum, Bombyx mori, and Drosophila willistoni (57, 58)), green algae, and some protozoa (including Kinetoplastida, Stramenopiles, and most Alveolata). Only nine organisms (mostly stramenopiles and green algae) could utilize all five trace elements.

**Occurrence of Transporters, Cofactor Biosynthesis Pathways, Metalloproteins, and Selenoproteins**—Initial genomic analyses of high affinity transporters, biosynthetic machinery, and sets of metalloproteins (metalloproteomes) and selenoproteins (selenoproteomes) have been carried out for individual trace elements and have used different sets of organisms (37–41). In this study, we extended these analyses to a common genomic dataset of more than 700 organisms from the three domains of life. We analyzed all known metalloproteins and high affinity transport systems for Cu, Mo, Ni, and Co and also analyzed all known selenoproteins and Moco, B12, and Sec biosynthesis.
Comparative Genomics of Trace Elements

| Phyla            | Cu | Mo | Ni | Co (B_{12}) | Se (Sec) |
|------------------|----|----|----|-------------|----------|
| **Bacteria**     |    |    |    |             |          |
| Firmicutes/Mollicutes |  |    |    |             |          |
| Firmicutes/Lactobacillales |  |    |    |             |          |
| Firmicutes/Bacillales |  |    |    |             |          |
| Firmicutes/Clostridia |  |    |    |             |          |
| Chlamydia |  |    |    |             |          |
| Bacteroides |  |    |    |             |          |
| Chlorobi |  |    |    |             |          |
| Actinobacteria |  |    |    |             |          |
| Spirochaetes |  |    |    |             |          |
| Planctomycetes |  |    |    |             |          |
| Cyanobacteria |  |    |    |             |          |
| Chloroflexi |  |    |    |             |          |
| Deinococcus-Thermus |  |    |    |             |          |
| Thermomicrobia |  |    |    |             |          |
| Aquificae |  |    |    |             |          |
| Fusobacteria |  |    |    |             |          |
| Lentisphaeria |  |    |    |             |          |
| Verrucomicrobia |  |    |    |             |          |
| Candidate division TMT |  |    |    |             |          |
| Acidobacteria |  |    |    |             |          |
| Deltaproteobacteria |  |    |    |             |          |
| Epsilonproteobacteria |  |    |    |             |          |
| Alphaproteobacteria/Rickettsiales |  |    |    |             |          |
| Alphaproteobacteria/Other Bacteria |  |    |    |             |          |
| Alphaproteobacteria/Rhizobiales |  |    |    |             |          |
| Betaproteobacteria/Bordetella |  |    |    |             |          |
| Betaproteobacteria/Burkholderiales |  |    |    |             |          |
| Betaproteobacteria/Neisseriaceae |  |    |    |             |          |
| Betaproteobacteria/Other Bacteria |  |    |    |             |          |
| Gammaproteobacteria/Enterobacteriales |  |    |    |             |          |
| Gammaproteobacteria/Pasteurellaceae |  |    |    |             |          |
| Gammaproteobacteria/Vibrionales |  |    |    |             |          |
| Gammaproteobacteria/Pseudomonadales |  |    |    |             |          |
| Gammaproteobacteria/Xanthomonadales |  |    |    |             |          |
| Gammaproteobacteria/Other Bacteria |  |    |    |             |          |
| **Total** | 432 (80%) | 401 (74%) | 319 (59%) | 410 (76%) | 139 (26%) |
| **Archaean** |  |    |    |             |          |
| Nanoarchaeota |  |    |    |             |          |
| Crenarchaeota/Thermoproteales |  |    |    |             |          |
| Crenarchaeota/Desulfurococcales |  |    |    |             |          |
| Crenarchaeota/Sulfoproteales |  |    |    |             |          |
| Euryarchaeota/Thermoplasmatales |  |    |    |             |          |
| Euryarchaeota/Archaeglobales |  |    |    |             |          |
| Euryarchaeota/Haloarchaebacteria |  |    |    |             |          |
| Euryarchaeota/Methanomicrobiales |  |    |    |             |          |
| Euryarchaeota/Methanobacteriales |  |    |    |             |          |
| Euryarchaeota/Methanosarcinales |  |    |    |             |          |
| Euryarchaeota/Thermococcales |  |    |    |             |          |
| **Total** | 26 (55%) | 46 (98%) | 39 (83%) | 45 (96%) | 6 (13%) |
| **Eukaryotic** |  |    |    |             |          |
| Cryptophyta |  |    |    |             |          |
| Diplomonadida |  |    |    |             |          |
| Parabasalida |  |    |    |             |          |
| Kinetoplastida |  |    |    |             |          |
| Stramenopiles |  |    |    |             |          |
| Alveolata/Perkinsea |  |    |    |             |          |
| Alveolata/Alveolata |  |    |    |             |          |
| Alveolata/Ciliophora |  |    |    |             |          |
| Rhodophyta |  |    |    |             |          |
| Viridiplantae/Chlorophyta |  |    |    |             |          |
| Viridiplantae/Streptophyta |  |    |    |             |          |
| Dicystostelida |  |    |    |             |          |
| Entamoebidae |  |    |    |             |          |
| Fungi/Ascomycota/Pezizomycotina |  |    |    |             |          |
| Fungi/Ascomycota/Saccharomycotina |  |    |    |             |          |
| Fungi/Ascomycota/Schizosaccharomycetes |  |    |    |             |          |
| Fungi/Basidiomycota |  |    |    |             |          |
| Fungi/Micromycota |  |    |    |             |          |
| Fungi/Zygomycota |  |    |    |             |          |
| Metazoa/Pseudocoelemata/Nematoda |  |    |    |             |          |
| Metazoa/Coelemata/Others |  |    |    |             |          |
| Metazoa/Coelemata/Arthropoda |  |    |    |             |          |
| Metazoa/Coelemata/Chordata/Others |  |    |    |             |          |
| Metazoa/Coelemata/Mammals |  |    |    |             |          |
| **Total** | 154 (96%) | 105 (66%) | 51 (32%) | 49 (31%) | 76 (48%) |
pathways. Occurrence and composition of metalloproteomes and selenoproteomes in bacteria, archaea, and eukaryotes are shown in Figs. 2–4, respectively. A list of known transporters and user proteins for different trace elements is shown in supplemental Table S1.

Copper—The mechanisms of Cu trafficking in prokaryotes are not fully understood. Several Cu-specific transport and resistance proteins and systems are known, including CopA (or PacS), CusCBA/CFBA, CutC, PcoABCD, PcoE, and CueO (59–61). However, the function of some of these proteins, such as PcoABCD and PcoE, is unclear. Thus, in this study, we only analyzed those proteins that have been experimentally shown to be involved in Cu transport: CopA, CusCBA/CFBA, and CutC. It should be noted that although CtaA, a CopA homolog identified in several cyanobacteria (such as Synechococcus sp. PCC7942 and Synechocystis PCC 6803), was previously reported to be involved in Cu import from the periplasm (16, 62, 63), it is absent in other phyla. Therefore, close homologs of CopA/CtaA in other phyla were considered as CopA proteins that are involved in Cu export. In addition, occurrence of a CopA-associated Cu chaperone CopZ that directly delivers Cu to CopA (64) was also analyzed to help identify CopA orthologs. The Cu-associated Cus system belongs to a large group of CBA-type transport systems that are involved in export of different metal ions, xenobiotics, and drugs (65–67). This system usually contains three components encoded in the same operon. The inner membrane pump (CusA in the Cus system) belongs to the resistance nodulation cell division family. The other two components (CusB and CusC) are members of a membrane fusion protein family and an outer membrane factor family, respectively (60). It is difficult to distinguish CusA and CusC orthologs from Cu-unrelated homologs. However, CusB proteins appear to contain a conserved motif M_{12–14}M/C_{10–16}M in the N terminus, which may be involved in Cu binding and could be used

FIGURE 1. Overview of utilization of five trace elements in the three domains of life. The tree is based on a highly resolved tree of life, which was simplified to only show major branches. Phyla in which more than half of organisms use all five trace elements are shown in red (if having at least three organisms, shown in bold). Phyla in which no organisms use any of the five trace elements are shown in blue (if having at least three organisms, shown in bold). The length of the colored section of each bar represents a proportion of organisms that use the corresponding trace element among all sequenced organisms in this phylum. The number and percentage of organisms that utilize individual trace elements in each domain of life are also shown.

FIGURE 2. Metalloproteomes and selenoproteomes in bacteria. This figure shows the numbers of proteins that use the indicated trace elements in each sequenced bacterium. Organisms containing the largest eukaryotic metalloproteomes (for Cu, Mo, Ni, and Co) and selenoproteomes are indicated. Phyla in which the majority of sequenced organisms use all five trace elements are highlighted in red.
as a signature for CusB proteins. Multiple alignment of CusB and other homologs is shown in Fig. 5. Thus, based on the occurrence of \textit{cusB} genes and neighboring \textit{cusA}, \textit{cusC}, and/or \textit{cusF} genes, we could characterize the occurrence of the complete Cus system.

Distribution of each Cu transporter in various prokaryotic phyla is shown in supplemental Table S2. First, members of CopA family appeared to be the most widespread Cu exporters in prokaryotes. A total of 460 bacteria and 39 archaea were found to have CopA proteins. Occurrence of the other two transporters was relatively limited, especially the Cus system that was only detected in Gram-negative bacteria. Second, the number of organisms containing at least one Cu exporter (470 and 39 organisms for bacteria and archaea, respectively) was larger than that of Cu-utilizing organisms (432 and 26 organisms, respectively), suggesting that some organisms that lack Cu-dependent proteins do utilize Cu exporters (mostly CopA) to deplete the intracellular Cu levels. Third, some organisms were found to have multiple copies of certain Cu transporters, e.g. CopA. The distribution of Cu exporters in bacteria is shown in supplemental Fig. S4. The highest number of Cu transporters in bacteria was observed in \textit{Acidovorax} sp. JS42 and \textit{Ralstonia pickettii} (10 and 9 Cu exporters, respectively), both of which were isolated from highly contaminated environments. It is possible that these species need more efficient mechanisms to maintain cellular Cu homeostasis or protect against this metal. In archaea, CopA is the only identified Cu exporter whose occurrence was not unusual (supplemental Fig. S5).

Previously, several bioinformatics studies examined the occurrence of Cu-binding proteins in model organisms on the basis of known Cu-binding patterns (56, 68). However, such approaches may be biased because not all Cu-binding motifs/ligands are known, and some Cu-binding proteins may bind other metals with the same ligands in the same or different organisms. Therefore, in this study, we only considered strict Cu-binding proteins as Cu-dependent proteins (supplemental
Table S1). Among known Cu-dependent proteins in prokaryotes, cytochrome c oxidase subunits I and II were the most frequently used Cu-binding proteins in both bacteria and archaea (supplemental Figs. S6A and S6B, details are shown in supplemental Table S2). Cytochrome c oxidase (or complex IV) is the terminal enzyme of the respiratory chains of mitochondria and many aerobic bacteria. It catalyzes the reduction of molecular oxygen to water and pumps an additional proton across the membrane for each proton consumed in the reaction. The resulting electrochemical gradient is used elsewhere, for instance in the synthesis of ATP (69). Other Cu-binding proteins, such as Cu,Zn superoxide dismutase, plastocyanin, and a variety of multicopper oxidases were also found in many prokaryotes. In contrast, the occurrence of particulate methane monooxygenase, nitrosocyanin, copper amine oxidase, and tyrosinase appeared to be very limited (varied from 5 to 35 organisms). In addition, some bacterial Cu-dependent protein families, including azurin, plastocyanin, NADH dehydrogenase 2, particulate methane monooxygenase, and tyrosinase were absent in archaea, whereas a blue copper protein, rusticyanin, was only detected in archaea.

Investigation of the whole set of Cu-dependent proteins (or cuproproteomes) showed a diverse distribution of Cu utilization. In bacteria (Fig. 2), large cuproproteomes were mainly observed in proteobacteria, especially Alphaproteobacteria/Rhizobiaceae among which two Sinorhizobium species (S. medicae and S. meliloti) contained the largest bacterial cuproproteomes (22 Cu-dependent proteins, half were cytochrome c oxidase I and cytochrome c oxidase II family proteins). The two organisms that contained the highest number of Cu exporters, Acidovorax sp. JS42 and R. pickettii, also had 14 and 19 Cu proteins (supplemental Fig. S7). In archaea (Fig. 3), large cuproproteomes were mainly found in Euryarchaeota/Haloarchaeales phylum, including Haloarcula marismortui that had the largest prokaryotic cuproproteome (25 Cu-dependent proteins; half are plastocyanin homologs). Thus, it appeared that
although bacteria and archaea have similar Cu-dependent protein families, occurrence of these proteins was mostly different.

We further extended this analysis to eukaryotes. Ctr1 and ATP7 (Ccc2 in *Saccharomyces cerevisiae*) are known as major Cu importer and exporter, respectively, from fungi to mammals (70). Besides, a bacterial CutC homolog was found to be involved in Cu export (71, 72). We found that all Cu-utilizing organisms had ATP7 and almost all (94%) had Ctr1 orthologs, revealing a good correspondence between the occurrence of Cu transporters (both importers and exporters) and Cu utilization (supplemental Table S3). CutC was detected in approximately half of the Cu-utilizing organisms, all of which contained single gene copies. Most organisms had 1–3 *ctr1* genes; however, nematodes possessed many of them, for example *Caenorhabditis elegans* (11 *ctr1* genes, supplemental Fig. S8), suggesting a complex Cu uptake system in these organisms. It is possible that these Ctr1 proteins are located in various membranes (i.e. plasma or organellar membrane) and cell types. The occurrence of Cu exporters varied from one to six genes, and three *Phytophthora* species, which are crop plant pathogens belonging to the genus *Oomycetes*, had relatively high numbers of exporters (i.e. *Phytophthora infestans* possessed six ATP7 proteins, supplemental Fig. S8).

Analysis of eukaryotic cuproproteomes revealed that land plants possessed the largest cuproproteomes (62 and 78 proteins in *Arabidopsis thaliana* and *Oryza sativa*, respectively; Fig. 4). Most of their Cu-binding proteins belonged to plantacyanin, copper amine oxidase, and multicopper oxidase families, suggesting important roles of these families in plant metabolism. Although nematodes had the highest number of *ctr1* genes as described above, only 15 Cu-dependent proteins were detected in these organisms (supplemental Fig. S9B). It is interesting that nematodes need so many Ctr1 transporters for their relatively few Cu user proteins. However, a possibility that additional Cu-dependent proteins are present in these organisms cannot be excluded.

**Molybdenum**—Previously, we carried out comparative genomics analyses to examine the occurrence and evolution of Mo utilization in 631 organisms from the three domains of life on the basis of Mo transport systems, Moco biosynthesis trait, and molybdoenzymes (40). In this study, we extended these analyses to include more than 100 newly sequenced organisms. The occurrence of Mo transporters, Moco biosynthesis trait, and Mo-dependent enzymes in prokaryotes and eukaryotes is shown in supplemental Tables S4 and S5. The only non-Moco-containing protein nitrogenase, which utilizes Fe-Mo as a cofactor (73), was also analyzed. It should be noted that because Mo utilization cannot be reliably distinguished from the utilization of W, the Mo data combined include Mo and W traits.
Prokaryotes mainly use ModABC (bacteria) and WtpABC (archaea) systems for molybdate uptake, which are often present in single copies in most species. Essentially all Mo-utilizing organisms have at least one known Mo transport system (supplemental Table S4). In eukaryotes, MOT1, the only known molybdate transporter, was detected in less than 40% Mo-utilizing organisms, which are land plants, green algae, pezizomy- 
cotina, and stramenopiles (supplemental Table S5). The absence of MOT1 in animals implied the presence of a currently unknown Mo transport system in these organisms. In bacteria, except for the aldehyde:ferredoxin oxidoreductase family (found in 59 organisms), Moco-containing enzymes showed widespread occurrence in Mo-utilizing organisms (93, 65, and 64% for dimethyl sulfoxide reductase, sulfite oxidase, and xanthine oxidase, respectively). Dimethyl sulfoxide reductase, the family used by most organisms, was largely represented by nitrate reductase and formate dehydrogenase. Many organisms possessed multiple Moco-containing protein families and several subfamilies within these families. On the other hand, only 77 organisms were found to possess nitrogenase, and almost all of them utilized Moco. In archaea, members of the dimethyl sulfoxide reductase family were found in 96% Mo-utilizing organisms. In contrast to bacteria, the aldehyde:ferredoxin oxidoreductase family was detected in 70% of Mo-utilizing archaea, whose occurrence was much higher than that of sulfite oxidase and xanthine oxidase families (52 and 30% respectively). Nitrogenase was only found in methanogenic archaea (supplemental Table S4).

Although only four molybdoenzyme superfamilies were detected in prokaryotes, investigation of molybdoenzyme sets (molybdoproteome) of each organism revealed many molybdo- proteins and variable occurrence of these proteins. Similar to the distribution of cuproproteins in bacteria, proteobacteria appeared to have larger molybdoproteomes than other organ- isms (Fig. 2). Surprisingly, the largest molybdoproteome was observed in Desulfotobacterium hafniense (Firmicutes/Clostridia). This organism contained at least 63 molybdoproteins, almost twice as many as other molybdoprotein-rich organisms (e.g. 35 and 31 molybdoproteins in Magnetospirillum magnetotacticum and Burkholderia xenovorans, respectively). Almost all molybdoproteins (~95%) in D. hafniense were members of the dimethyl sulfoxide reductase family. In archaea, relatively large molybdoproteomes were observed in Crenarchaeota/Sul- folobales (Fig. 3), and the majority of these proteins were members of the xanthine oxidase family.

In eukaryotes, Moco is the only known form of Mo utiliza- tion, and there are only two known molybdoenzyme families sulfite oxidase and xanthine oxidase (74). Essentially all organ- isms that possessed the Moco utilization trait had both sulfite oxidase and xanthine oxidase families. Land plants appeared to possess the largest molybdoproteomes in eukaryotes (10–11 molybdoproteins). In contrast, all sequenced saccharomy- cotina did not have molybdoenzymes. Although a small number of unsequenced yeast species, such as Candida nitratophilica and Pichia anomala, were reported to utilize Mo-containing assimilatory nitrate reductase (75, 76), the fact that homologs of this protein and Moco biosynthesis pathway were not detected in all sequenced yeast genomes strongly suggested the loss of Mo utilization in these organisms (40).

Nickel and Cobalt—As discussed previously (41), both metals were found to be widely used by prokaryotes; however, analyses of occurrence of Ni/Co transporters and metalloenzymes showed diversity among bacteria and archaea. Urease and B12-dependent methionine synthases were the most widespread Ni- and Co-containing proteins, respectively, in bacteria. In con- trast, Ni-Fe hydrogenase and B12-dependent ribonucleotide reductase were the most widespread Ni and Co users in archaea where urease and methionine synthase were very rare or even absent. Further analyses of Ni- or Co-dependent metallopro- teomes revealed that, except for Deltaproteobacteria and sev- eral Methanosarcina species, most prokaryotes contained small Ni- and Co-dependent metalloproteomes (1–4 proteins, Figs. 2 and 3). The largest Ni-dependent metalloproteome was observed in Deltaproteobacterium MLMS-1 (16 Ni-binding proteins) and the largest B12-dependent metalloproteome in Dehalococcoides sp. CBDB1 (35 B12-binding proteins).

A somewhat different Ni and Co utilization trend was observed in eukaryotes. Urease and methionine synthase were the most widespread Ni- and B12-dependent enzymes, respec- tively. Analysis of Ni- and Co-dependent metalloproteomes in eukaryotes did not reveal organisms that contained many such proteins (Fig. 4). Only single copies of urease and 1–3 B12-dependent proteins were detected in Ni- and Co-utilizing organisms, highlighting a restricted utilization of both met- als in eukaryotes. In contrast to the majority of unicellular organisms that lack B12 utilization, Dictyostelium discoi- deum and Phytophthora species contained all three known eukaryotic B12-dependent proteins as follows: methylmalonyl-CoA mutase, methionine synthase, and B12-dependent ribonucleotide reductase.

Selenium—Previous studies carried out initial analyses of the Sec utilization trait and selenoproteomes (37, 38). This study greatly extended these analyses by examining additional sequenced genomes and newly identified selenoproteins (supplemental Table S1)(77–79). This provides additional information on the utilization of Sec.

In bacteria, the size of selenoproteomes varied from 1 to 39 selenoproteins, and most selenoprotein-rich organisms (num- ber of selenoproteins ≥6) were anaerobic Deltaproteobacteria or Clostridia (Fig. 2), with the largest selenoproteome identified in Syntrophobacter fumaroxidans. Formate dehydrogenase remained the most widespread selenoprotein family in pro- karyotes with completely sequenced genomes (but metagenomics analyses of marine samples revealed a low abund- ance of this selenoprotein (78)), which was previously suggested to be responsible for maintaining the Sec-decoding trait (37). Archaeal selenoproteomes did not vary much (most often 7–10 selenoproteins), but there were only six identified Sec-utilizing archaea (Fig. 3).

Danio rerio was reported to have the largest selenoproteome among previously analyzed eukaryotes (36 selenoproteins (57)). Here, we analyzed 160 sequenced eukaryotic genomes. Among them, selenoprotein-rich organisms were found in animals (mostly in vertebrates; however, insects and nematodes had few selenoproteins), ciliates, stramenopiles, and green algae,
whereas several unicellular organisms (such as Apicomplexa and Kinetoplastida), possessed very few selenoproteins (Fig. 4). The majority of selenoprotein families occurred in both unicellular eukaryotes and mammals, suggesting that they originated at the base of the eukaryotic domain. It has been suggested that independent selenoprotein loss events happened in distant phyla (including land plants, fungi, nematodes, and some protists) and that aquatic organisms generally have large selenoproteomes than several groups of terrestrial organisms (38). Our results are consistent with these findings and emphasize the general trends for the utilization and evolution of Se in eukaryotes.

Saccharomycotina Lost the Ability to Utilize Mo, Ni, Co (B12), and Se—It is interesting that, compared with other fungi, many saccharomycotina (including the most utilized eukaryotic model organism, S. cerevisiae, as well as closely related organisms such as Candida, Kluyveromyces, and Pichia) do not possess signatures of Mo, Ni, Co, and Se utilization. Neither high affinity transporters nor user proteins that utilize these micronutrients were detected in their genomic sequences. These observations strongly suggest that these four trace elements are not used by saccharomycotina and that perhaps alternative metabolic pathways have evolved in these organisms. For example, both B12-independent methionine synthesis and urease-independent urea degradation pathways were identified in S. cerevisiae and Candida albicans (80, 81). In addition, S. cerevisiae lacked any of the known Moco-containing enzymes; hence, it should not require Mo as a mineral supplement. In fact, high molybdate levels are inhibitory for this organism. Recently, an efficient Hsp90/Cdc37p chaperone system was identified as a determinant of molybdate resistance in S. cerevisiae (82). Similar situations may be expected in other organisms that do not utilize certain trace elements.

New Model Organisms to Study Trace Element Utilization—From our analyses, several candidate model organisms emerged for studying trace elements in eukaryotes. A list of these model organisms is shown in Table 1. For example, land plants (e.g. A. thaliana and O. sativa) have the largest number of Mo- and Cu-containing proteins among sequenced eukaryotes. An earlier bioinformatics study that analyzed the occurrence of putative Cu-binding proteins in several sequenced eukaryotes based on Cu-binding patterns derived from the Protein Data Bank showed that A. thaliana had the largest proteome fraction of Cu-binding proteins among analyzed eukaryotes (56). In our work, rice (O. sativa) appeared to contain even more Cu-dependent proteins than A. thaliana. Both plants are promising models for studying Mo utilization in eukaryotes.

The largest number of Cu importer Ctrl and Cu exporter ATP7 proteins was observed in C. elegans and several Phytophthora species, respectively. However, they do not have a similarly high proportion of Cu-dependent proteins (Supplemental Fig. S9B). Thus, these organisms, especially C. elegans with its powerful genetic tools, are very promising models for investigation of Cu transport/homeostasis in animals. Currently, this resource is essentially untapped.

Previously, most studies on B12 metabolism in eukaryotes focused on mammals because there have been no unicellular B12-dependent organisms amenable to genetic analyses. We found that several unicellular eukaryotes contained the B12-dependent proteins, most notably D. discoideum, a social amoeba that is a powerful model organism for studying cellular and developmental processes (83). Therefore, investigation of B12 utilization in D. discoideum may provide important information with regard to the role of B12 in eukaryotes.

Among eukaryotes investigated in our study, zebrafish had the largest eukaryotic selenoproteome. It should be an excellent model organism to study Se utilization in eukaryotes, but only a few previous studies used it. Se is required during development in mammals, and several selenoproteins are essential in these organisms due to their roles in development. Because of similarities in selenoproteomes, the use of zebrafish could help characterize these essential pathways.

In prokaryotes, D. hafniense emerged as a promising model for the analyses of multiple trace elements. This organism belongs to Firmicutes/Clostridia and can utilize all five trace elements. In contrast to closely related species, many metal-containing proteins (e.g. 63 Moco-containing proteins, 10 Ni-containing proteins, and 21 B12-containing proteins) are present in D. hafniense, suggesting prominent roles of these metals in this organism. Further studies are needed to identify pathways in which these trace elements are utilized.

Metabolic and Evolutionary Interactions among Trace Elements—The availability of detailed information on the occurrence and utilization of the five trace elements analyzed in this study offered us an opportunity to systematically examine biological interactions among these elements. Some interactions have already been reported, including the following: (i) common transport systems for Ni and Co uptake in prokaryotes (34, 41); (ii) identification of the Sec utilization trait as a subset of the Mo trait in prokaryotes (40); and (iii) identification of Cnx1G, a Cu-containing enzyme involved in Moco biosynthesis in plants (84). In addition, in some organisms, certain proteins could bind multiple trace elements, such as Ni–Fe–Se-containing hydrogenases in Sec-decoding archaea (85).

Our analyses of metalloproteomes (for Cu, Mo, Ni, and Co) and selenoproteomes in prokaryotes and eukaryotes (Figs. 2–4) revealed that certain phyla, such as Deltaproteobacteria (bacteria) and Stramenopiles (eukaryotes), use four or all five examined trace elements. In addition, high content of trace element-containing proteins suggested important roles for micronutrients in these organisms, providing potentially useful models for investigation of the function of trace elements. In several other bacterial phyla (such as Chlorobi and Gammapro-
Prokaryotes use common transporter families for Ni and Co uptake, despite the fact that the subtypes of various transporters show different ion preferences ranging from Ni-specific to unbiased transport of both ions to Co-specific (41). Thus, significant overlap between the two utilization traits was observed in bacteria and archaea (supplemental Fig. S10A). In contrast, a different distribution of Ni- and Co-utilizing organisms in eukaryotes suggested an independent relationship between them in this domain of life. In addition, as noted previously, Sec-utilizing organisms were essentially a subset of Mo-utilizing organisms in prokaryotes (supplemental Fig. S10B), most likely because of formate dehydrogenase, which is not only a widespread Mo enzyme but is also the major selenoprotein family in these organisms (40). We did not detect significant overlaps between other trace elements. In general, trace elements analyzed in our study had different patterns of occurrence, suggesting that the pathways of their utilization are independent of each other.

Identification of *A. thaliana* Cnx1G provided a molecular link between Mo and Cu metabolism in plants (84). However, experiments in *E. coli* and *Rhodobacter sphaeroides* showed that Cu was not strictly required for Moco biosynthesis in these two organisms (86). In this study, we analyzed Cnx1G orthologs in various organisms and found that His-618 (numbering based on *A. thaliana* Cnx1G sequence), a candidate Cu-binding ligand (84), was not conserved. Although Cu-binding function could not be excluded for many Cnx1G proteins that lack His-618, it is possible that whereas Cu may be utilized during Moco biosynthesis in some organisms such as plants, it does not appear to be strictly required for Moco biosynthesis in many other organisms. Similarly, the occurrence of other multielement-containing proteins was restricted. For example, the Ni-Fe-Se-containing hydrogenases were only detected in Sec-containing archaea and several Deltaproteobacteria (37). These data further stress independence in trace element utilization.
We examined the possibility that the utilization of some trace elements may be affected by certain environmental factors. First, as observed previously, Cu was mainly used by aerobic organisms, whereas organisms possessing the Sec-decoding trait favored anaerobic conditions (37, 39), suggesting conflicting roles that oxygen played during evolution of utilization of Cu and Se (Fig. 6A). Second, similar to Ni and Co utilization in prokaryotes (41), the majority of bacteria that lack the Mo utilization trait were host-associated, especially obligate intracellular symbionts and parasites, whereas the majority of extracellular symbionts utilized Ni, Co, and Mo (Fig. 6B). Our data imply that the host-associated lifestyle may result in the loss of utilization of trace elements (e.g. Ni, Co, and Mo), perhaps due to limited resources and genome streamlining. Most intracellular parasites or symbionts had much smaller genomes (e.g. <2 Mbp) and lower G + C content (i.e. G + C <40%) than other organisms (41). It is possible that these metal utilization traits are dispensable for intracellular organisms and hence have been lost due to evolutionary pressure on genome size, although these organisms may still depend on Mo-, Ni-, or Co-dependent proteins of the host. In contrast, these metal utilization traits mostly remained intact in extracellular symbionts. Other factors, such as Gram strain and pH, appeared to have no significant effect on the evolution of trace element utilization.

A similar situation was observed in eukaryotes. Except for Cu and Se that were detected in both parasites and non-parasites, most parasitic organisms did not use Mo, Ni, and Co, which is consistent with what we observed in prokaryotes (Fig. 7). Thus, it appeared that these metals may have become unnecessary for parasites because of reduced availability or dependence on the corresponding pathways of the host. On the other hand, the lack of utilization of most trace elements analyzed in saccharomycotina (see above) suggested that unknown factors may have also affected the utilization of multiple trace elements.

Previous studies showed that the size of cuproproteomes and selenoproteomes correlates with oxygen utilization and/or aquatic environment (37, 38, 39). For example, in prokaryotes, larger cuproproteomes were mainly found in aerobic organisms, whereas larger selenoproteomes were in anaerobic organisms. In eukaryotes, aquatic organisms generally have large selenoproteomes (38). However, no significant correlation was observed between different factors and the size of Mo-, Ni-, and Co-dependent metalloproteomes (data not shown). A future challenge would be to discover additional trends that influence trace element utilization in the three domains of life.

In this study, we carried out comprehensive comparative genomic analyses of the utilization of the following five trace elements: Cu, Mo, Ni, Co, and Se. We extended previous smaller scale genomic analyses of individual trace elements to 747 fully sequenced genomes in the three domains of life by analyzing the occurrence of transporters, cofactor biosynthesis pathways, metalloproteomes, and selenoproteomes. We found that few organisms could utilize all five trace elements. Although a few biological interactions were observed, trace elements mostly had unique patterns of occurrence, and their utilization was generally independent of each other. In prokaryotes, the Cu utilization trait and large cuproproteomes were mainly observed in aerobic organisms, whereas the Sec utilization trait and large selenoproteomes were mainly observed in anaerobic organisms. Most parasites and intracellular symbionts lacked the Mo, Ni, and Co utilization traits. In eukaryotes, many parasites and saccharomycotina lost the ability to use four or all five elements. Overall, our data provide important insights into the general trends of utilization and evolution of trace elements in the three domains of life, characterize these processes in common model organisms, and identify candidate model organisms for future experimental analyses.

REFERENCES

1. Mertz, W. (1981) Science 213, 1322–1328
2. Van Gossum, A., and Neve, J. (1998) Curr. Opin. Clin. Nutr. Metab. Care 1, 499–507
3. Goldhaber, S. B. (2003) Regul. Toxicol. Pharmacol. 38, 232–242
4. Mertz, W. (1998) Biol. Trace Elem. Res. 66, 185–191
5. Frassinetti, S., Bronzetti, G., Caltavuturo, L., Cini, M., and Croce, C. D. (2006) J. Environ. Pathol. Toxicol. Oncol. 25, 597–610
6. Galaris, D., and Pantopoulos, K. (2008) Crit. Rev. Clin. Lab. Sci. 45, 1–23
7. Foster, L. H., and Sumar, S. (1997) Crit. Rev. Food Sci. Nutr. 37, 211–228
8. Burk, R. F. (2002) Nutr. Clin. Care 5, 75–79
9. Moghadzazadeh, B., and Beggs, A. H. (2006) Physiology 21, 307–315
10. Patzer, S. I., and Hanke, K. (1998) Mol. Microbiol. 28, 1199–1210
11. Bartsevich, V. V., and Pakrasi, H. B. (1996) J. Biol. Chem. 271, 26057–26061
12. Maupin-Furlow, J. A., Rosentel, J. K., Lee, J. H., Deppenmeier, U., Gunsalus, R. P., and Shanmugam, K. T. (1995) J. Bacteriol. 177, 4851–4856
13. Eitinger, T., and Mandrand-Berthelot, M. A. (2000) Arch. Microbiol. 173, 1–9
14. Grass, G., Franke, S., Taude, N., Nies, D. H., Kurcharski, L. M., Maguire, M. E., and Rensing, C. (2005) J. Bacteriol. 187, 1604–1611
15. Makui, H., Roig, E., Cole, S. T., Helmann, J. D., Gros, P., and Cellier, M. F. (2000) Mol. Microbiol. 35, 1065–1078
16. Phung, L. T., Ajlani, G., and Haselkorn, R. (1994) Proc. Natl. Acad. Sci. U.S.A. 91, 9651–9654
17. Dancis, A., Yuan, D. S., Haile, D., Askwith, C., Eide, D., Moehle, C., Kaplan,
