Comparative Genomics of Trace Element Dependence in Biology

Yan Zhang and Vadim N. Gladyshev

Biological trace elements are needed in small quantities but are used by all living organisms. A growing list of trace element-dependent proteins and trace element utilization pathways highlights the importance of these elements for life. In this minireview, we focus on recent advances in comparative genomics of trace elements and explore the evolutionary dynamics of the dependence of user proteins on these elements. Many zinc protein families evolved representatives that lack this metal, whereas selenocysteine in proteins is dynamically exchanged with cysteine. Several other elements, such as molybdenum and nickel, have a limited number of user protein families, but they are strictly dependent on these metals. Comparative genomics of trace elements provides a foundation for investigating the fundamental properties, functions, and evolutionary dynamics of trace element dependence in biology.

All organisms require a regular supply of nutrients. The major biological elements, carbon, hydrogen, oxygen, nitrogen, phosphorus, and sulfur, are present in large quantities and are the main constituents of nucleic acids, proteins, lipids, cell walls, and mechanical structures (1, 2). There are also chemical elements, most of which are metals, that are present in small quantities (known as biological trace elements), including cobalt, copper, iron, manganese, molybdenum, nickel, zinc, selenium, and several other elements. The roles of trace elements primarily involve providing proteins with unique catalytic, structural, electron transfer, and other properties. They are used in numerous pathways, and all organisms utilize at least some of these elements (3–5). Their deficiency or mutations in genes that handle these elements often result in abnormal development, metabolic abnormalities, and even death (6–8).

Among trace elements, some such as zinc and iron appear to be used by all organisms (9–11). Although organisms that survive under iron limitation have been reported (12, 13), it is unclear if they do not use this metal under iron-sufficient conditions. The number of proteins that utilize zinc or iron is large. For instance, >300 enzyme families require zinc for proper function (14). Metals are used by a wide range of proteins in the form of metal ions (e.g. manganese, copper, and nickel) or complex metal-containing cofactors (e.g. molybdopterin in the case of molybdenum and vitamin B12 in the case of cobalt) (15–17). Selenium is used mainly in the form of selenocysteine (Sec) residues in proteins (18).

From the many studies on the functions of trace elements, metalloproteins emerged as one of the most diverse sets of proteins (19). Some protein families are strictly dependent on certain metals for their function (e.g. copper in cytochrome c oxidase and nickel in urease), whereas other families include both metal-dependent and metal-independent forms or evolved to use alternative metals (e.g. glyoxalase I binds nickel in Escherichia coli but zinc in humans and yeast) (20). In addition, orthologs of selenoproteins exist in which Cys replaces Sec (21). These observations highlight complex evolutionary dynamics of the dependence of proteins on trace elements.

Recent advances in genome sequencing and analysis led to the development of an exciting new biological field, comparative genomics (22–24). It provides powerful tools for studying functions and evolutionary changes of genes, thereby offering new strategies for research in biology and medicine. Initial genome-wide studies identified a significant number of proteins that bind metals (25, 26) and suggested that information on the occurrence of metal-dependent/metal-independent members of protein families can assist in the understanding of utilization of these micronutrients.

In this minireview, we focus on recent advances in comparative genomics studies of trace elements, discuss user protein families (i.e. proteins that are known to use certain trace elements for their proper function) containing metal-dependent and metal-independent forms, and explore the evolutionary dynamics of the metal dependence of these families. In addition, the comparative genomics of interchanges between Sec- and Cys-containing proteins is discussed.

Zinc

Zinc is thought to be essential for all organisms and was suggested to be a key element in the origin of life (27). It is found in a great variety of enzymes, structural proteins, transcription factors, ribosomal proteins, etc. (28). Zinc uptake, storage, homeostasis, and user proteins have been thoroughly discussed in many articles and reviews (14, 28–33), and it is not our intention to review these subjects here. Instead, taking advantage of the abundance of zinc-binding protein families, we systematically analyze their zinc dependence.

An early comparative study was carried out to investigate the evolutionary history of ribosomal proteins that are encoded in all genomes and are generally well conserved (34). Members of
each examined ribosomal protein family were extracted from ~40 genomes (mostly bacteria, but mitochondrial ribosomal proteins were also analyzed). It was found that several ribosomal proteins, such as S14, S18, L31–L33, and L36, all of which bind zinc via four conserved Cys residues (sometimes replaced by His), also have a second zinc-independent form in which these metal-chelating residues are partially or completely degenerated. A typical event the authors observed was that genomes containing multiple copies of these ribosomal proteins encoded both zinc-dependent and zinc-independent forms. Further phylogenetic and genomic context analyses revealed that, in most cases, a duplication of an ancestral zinc-dependent form had taken place early during evolution, with subsequent alternative loss of zinc-dependent and zinc-independent forms in different lineages. A separate comparative genomics study analyzed Zur (a repressor of zinc transport) regulation in bacteria (35). In this study, an unexpected finding was that zinc repression was also detected in some paralogs of ribosomal protein genes (such as L36, L33, L31, and S14) in which zinc-binding motifs were disrupted. This finding suggested a mechanism for maintaining zinc availability under conditions of zinc starvation, as these non-zinc-binding paralogs were expressed under zinc-restricted conditions to partially replace the zinc-binding proteins, thereby freeing up some zinc for utilization by essential zinc-binding proteins. Similar results were recently reported by another comparative study in which a subset of proteins in the diverse COG0523 family of putative metal chaperones were found to play a predominant role in the response to zinc limitation based on the presence of the corresponding genes downstream from putative Zur-binding sites in all kingdoms of life (36).

On the basis of these analyses, we asked the question, “How widespread is the evolutionary strategy involving zinc-dependent and zinc-independent forms?” To address this question, we developed a strategy (Fig. 1) that analyzes zinc-containing proteins with defined ligands in the Protein Data Bank (PDB) data set against a non-redundant protein data base for the occurrence of homologs that lack metal utilization. For each family, a zinc-dependent form was designated if a homolog preserved three or four known ligands, and a zinc-independent form was defined by having fewer than three known ligands. One complication was that some zinc-binding proteins were characterized by multiple zinc sites. Thus, homologs that lost some zinc-binding ligands could still bind zinc if an additional zinc site was preserved. Each family that was found to possess zinc-independent forms was manually analyzed and classified into Group I, which had >10% zinc-independent forms, and Group II, which had strictly zinc-dependent proteins. It cannot be excluded that certain homologs of Group II zinc protein families could bind metals other than zinc, but thus far, there is no reliable computational approach to define metal specificity. A general representation of the occurrence of zinc-binding residues in members of the two groups is shown in supplemental Fig. S1. Approximately 20% of the zinc protein families in the PDB data set had a significant number of zinc-independent forms, e.g., methionine-R-sulfoxide reductase (37), several tRNA synthetases, ribosomal proteins, and various subunits of DNA polymerase III. Multiple alignments of some of these families are shown in Fig. 2.

We also analyzed the predicted zinc dependence of zinc protein families in hundreds of sequenced bacterial genomes. The majority of organisms containing these proteins possessed only single forms (supplemental Fig. S2), i.e., either zinc-dependent or zinc-independent. Phylogenetic analyses revealed a role of vertical inheritance and horizontal gene transfer in shaping the evolution of zinc protein families, which is consistent with previous observations (34). Overall, all these studies suggested a general picture of evolution of zinc utilization wherein many zinc protein families are strictly dependent on zinc, whereas some families yielded zinc-independent forms by disrupting ligands to the metal ion. These zinc-independent forms were already present in the ancestors of prokaryotes and eukaryotes and are widespread in modern organisms. Further studies are needed to better understand the nature of regulation of zinc-independent forms by zinc regulators such as Zur.

Although the loss of the ability to use zinc has taken place in diverse protein families and in many organisms, zinc remains one of the most widely used trace elements. A recent comparative genomics study predicted a set of human proteins that require zinc based on a set of zinc-binding patterns (38). Approximately 2800 human proteins (10% of the human proteome) were found, most of which contained typical zinc-binding motifs (Cys₄ or Cys₂-His₂). The majority of these zinc-binding proteins are involved in the regulation of gene expression. A smaller scale but more extensive study was carried out to analyze 57 organisms across the three domains of life for an overall view of zinc usage by these organisms (39). It was found that putative zinc-dependent metalloproteomes correlated with proteome size. A significant number of regulatory proteins that
bind zinc were identified in eukaryotes, consistent with the idea that, in higher organisms, zinc is not only important for the function of numerous proteins and enzymes, but it also regulates gene expression.

Copper

Copper is used by several proteins and enzymes, such as cytochrome c oxidase (COX subunits I and II), copper/zinc superoxide dismutase, tyrosinase, and a large number of blue copper proteins and multicopper oxidases. Cys, His, and sometimes Met are the major copper-binding ligands in copper-binding proteins (40). Many previous studies have focused on copper transport, homeostasis, storage, and copper-binding proteins (cuproproteins) (40–43), whereas the evolution of copper utilization, especially copper dependence of various cuproprotein families, has thus far received less attention.

Several comparative genomics studies have been carried out to analyze copper utilization and to identify cuproproteins and cuproproteomes in organisms (44–47). These studies provided a first glance at copper utilization in the three domains of life and represent a new resource for studying the copper dependence of cuproprotein families. The majority of known cuproprotein families are strictly dependent on copper, including copper/zinc superoxide dismutase, blue copper proteins, and multicopper oxidases (46, 47). However, loss of copper ligands has been observed in members of several cuproprotein families, which was mostly accompanied by changes in the function of these proteins.

Cytochrome oxidase is one of the best studied cuproprotein families and acts as a terminal enzyme in the respiratory chain in aerobic organisms. Two major subgroups of this family include COX and quinol oxidase (QO) (48). In COX, both subunits I and II contain copper (CuA in COX II and CuB in COX I). However, QO subunit II has lost the CuA center (49, 50). Comparative analyses showed that COX is present in the majority of copper-utilizing organisms from bacteria to humans, whereas QO has been detected only in some bacteria (47), consistent with the idea that QO evolved from COX to become a new bacterial terminal respiratory oxidase that uses another substrate (quinol). At the same time, almost all QO-containing organisms also have COX.

Tyrosinases (or catechol/polyphenol oxidases) contain a type 3 copper center and are distributed in all three domains of life. These proteins are essential for pigmentation and are important factors in wound healing and primary immune response (51, 52). The active site of tyrosinase is characterized by a pair of antiferromagnetically coupled copper ions, named CuA and CuB (different from the CuA and CuB centers in COX), each coordinated by three His residues. In animals, two additional

The majority of known cuproprotein families are strictly dependent on copper, including copper/zinc superoxide dismutase, blue copper proteins, and multicopper oxidases (46, 47). However, loss of copper ligands has been observed in members of several cuproprotein families, which was mostly accompanied by changes in the function of these proteins.

Cytochrome oxidase is one of the best studied cuproprotein families and acts as a terminal enzyme in the respiratory chain in aerobic organisms. Two major subgroups of this family include COX and quinol oxidase (QO) (48). In COX, both subunits I and II contain copper (CuA in COX II and CuB in COX I). However, QO subunit II has lost the CuA center (49, 50). Comparative analyses showed that COX is present in the majority of copper-utilizing organisms from bacteria to humans, whereas QO has been detected only in some bacteria (47), consistent with the idea that QO evolved from COX to become a new bacterial terminal respiratory oxidase that uses another substrate (quinol). At the same time, almost all QO-containing organisms also have COX.

Tyrosinases (or catechol/polyphenol oxidases) contain a type 3 copper center and are distributed in all three domains of life. These proteins are essential for pigmentation and are important factors in wound healing and primary immune response (51, 52). The active site of tyrosinase is characterized by a pair of antiferromagnetically coupled copper ions, named CuA and CuB (different from the CuA and CuB centers in COX), each coordinated by three His residues. In animals, two additional
tyrosinase-related proteins (TRP1 and TRP2) that display significant homology to tyrosinase and that originated by the duplication of the ancestral tyrosinase gene were also detected (53, 54). TRP1 is a 5,6-dihydroxyindole-2-carboxylate oxidase, and TRP2 is a dopachrome tautomerase, both of which are involved in melanin biogenesis (55). TRP1 and TRP2 have the same six metal-binding His residues as tyrosinase. However, whereas TRP2 binds zinc, TRP1 binds an unknown metal that is not copper, zinc, or iron (56). Thus, the specific binding of different metals by these proteins may be responsible for their distinct catalytic functions in melanogenesis. It was suggested that mouse TRP2 contains a specific GPGGP motif, which is absent in TRP1 and tyrosinase (57); however, it is unclear how to distinguish copper-dependent and copper-independent forms in the tyrosinase family. We found that a conserved His is located immediately upstream of one of three His residues in the CuB site in all characterized tyrosinasines, but it is replaced by Leu in all TRP1 and TRP2 proteins (Fig. 3). Although this His has not been defined as a copper ligand based on known PDB structures, a previous study has shown that it is essential for copper binding in tyrosinase from Aspergillus oryzae (58). This additional His may play an important role in binding CuB and may help identify copper-dependent tyrosinasines in sequence databases. However, a recent study suggested the ability of the CuA site in mouse TRP1 to bind copper and sustain the typical tyrosinase enzymatic activities (59). Further experiments are needed to verify the metal specificity of the CuB site in TRP1. It is possible that tyrosinase may acquire copper inefficiently and subsequently lose it within the trans-Golgi network of melanocytes and then be reloaded with copper within melanosomes to catalyze melanin synthesis (60). This scheme would be consistent with the presence of an inactive “copper-independent” form of copper-dependent tyrosinase, suggesting an exquisite and complex spatial control of tyrosinase activity.

Hemocyanins are oxygen carrier proteins that are responsible for precise oxygen delivery from respiratory organs to tissues (61). Hemocyanin also belongs to the type 3 cuproprotein family and uses six His residues to bind two copper ions (62). Members of the hemocyanin family have been detected only in the hemolymph of animals in the phyla Arthropoda and Mol-
Molybdenum is used mainly in the form of molybdopterin (or molybdopterin cofactor) in molybdoenzymes, and nickel is found only in a limited number of enzymes (such as urease and nickel-iron hydrogenase) (16, 68). Recent comparative genomics studies have shown that molybdoenzymes are strictly dependent on this cofactor (47, 69). Similarly, nickel-dependent proteins are strictly dependent on this element (70). Although additional nickel-binding proteins were reported in some organisms (such as glyoxalase I) (20), it is unclear how to distinguish the metal specificily of these proteins. It would be helpful to identify sequence- and/or structure-based features of different forms of these protein families, which could then be used as signatures for further analyses.

Cobalt is found mainly in the corrin ring of vitamin B$_{12}$, a cofactor involved in methyl group transfer and rearrangement reactions (71). In addition, cobalt is also found in several non-corrin cobalt-containing enzymes, such as nitrile hydratase (NHase) and methionine aminopeptidase (72). Comparative genomics studies have shown that all vitamin B$_{12}$-dependent proteins contain a B$_{12}$-binding domain and are strictly dependent on this coenzyme (70, 73). It is very difficult to predict cobalt-dependent and cobalt-independent forms of non-corrin cobalt-binding protein families solely based on computational approaches. To date, only NHase is known to have different active site motifs for cobalt- and iron-binding forms (74, 75). We previously found that only several sequenced organisms (~4%) contain NHases, and most of them are cobalt-binding proteins. Phylogenetic analyses revealed that the iron-containing NHases might have evolved from cobalt-binding NHases (70).

**Selenium**

Selenium is an essential metalloid micronutrient in many organisms (including humans) and has antioxidant, redox regulatory, anti-inflammatory, chemopreventive, and other properties (76). In contrast to metals, selenium is cotranslationally inserted into proteins in the form of the Sec residue (18). The Sec biosynthesis and insertion into selenoproteins involve a complex molecular machinery that recodes UGA codons (normally functioning as stop signals) to serve as Sec codons (18, 68, 77, 78).

In recent years, much progress has been made in the discovery of new selenoproteins, most of which were identified by bioinformatics approaches and further verified by experiments (79–85). For example, 25 selenoprotein genes were found in humans and 24 in mice and rats (83). An interesting feature of these proteins is that almost all selenoproteins have widespread homologs in which Sec is replaced with Cys. Thus, selenium utilization can be examined by analyzing the evolutionary dynamics of selenium-dependent (selenoprotein) and selenium-independent (Cys-containing) forms of various selenoprotein families. Sec can greatly increase the catalytic activity of proteins compared with their Cys-containing homologs (86, 87). However, the question of why Sec is used very selectively in proteins and organisms has not been fully answered.

We have carried out comparative studies to analyze selenium utilization and evolution of selenoproteomes (88–91). To identify conversion events between Sec- and Cys-containing forms in selenoprotein families, we used the approach illustrated in Fig. 4A. If a single selenoprotein clustered with several closely related Cys-containing sequences, a Cys → Sec conversion (selenoprotein evolved from a Cys-containing protein) could be inferred. Similarly, if a Cys-containing form clustered with several closely related Sec-containing sequences, we inferred a Cys → Sec conversion (89). In prokaryotes, Cys → Sec appeared to be a general trend for evolution of selenoproteins, especially those in selenoprotein-rich organisms (Deltaproteobacteria and Clostridia). However, the observations that the size of most selenoproteomes was generally small (one to three selenoproteins) and that more than half of the examined organisms lacked selenoproteins were inconsistent with the Cys → Sec trend that should have resulted in widespread occurrence of selenoproteins. It appears that the evolution of new selenoproteins is balanced by selenoprotein loss events, even in closely related organisms in the same phylum. To verify this possibility, we developed the approach shown in Fig. 4B. We selected several sister and relatively distant species for each Sec-utilizing organism in the same phylum and individually analyzed each of their selenoproteins. If less than two sister species and at least two distantly related organisms contained either Sec- or Cys-containing orthologs, the corresponding selenoprotein family was predicted to be subject to family loss (or loss in progress) in the phylum. With this approach, selenoprotein gene loss events were observed in many sister species of selenoprotein-rich organisms, suggesting a dynamic balance between Sec acquisition and selenoprotein loss in these organisms. These findings may partially explain the discrepancy between catalytic advantages offered by Sec and its restricted use in nature.

The situation in eukaryotes appears to be more complex. Sec-containing forms of many mammalian selenoproteins were detected in unicellular eukaryotes, suggesting the ancient origin of these selenoproteins (90). It was proposed that many
ancestral selenoproteins were preserved in vertebrates, green algae, and a variety of protists. In contrast, land plants, fungi, nematodes, insects, and some unicellular eukaryotes manifested massive independent losses of selenoproteins or even the entire Sec utilization trait (90). In contrast to prokaryotes, fewer Cys → Sec and Sec → Cys conversion events were observed for eukaryotic selenoprotein families. The difficulties for such conversions are probably related to a complex Sec insertion machinery in eukaryotes. Interestingly, large selenoproteomes tend to occur in aquatic organisms, whereas the organisms that lack selenoproteins or have small selenoproteomes are mostly terrestrial (90). These results suggest that additional factors (such as stable environment, bioavailability of selenium, UV radiation, and oxygen content) may be involved in preserving selenoproteins in aquatic organisms or in the loss of these proteins in terrestrial organisms. Another comparative study focused on selenoprotein P (SelP; the only mammalian selenoprotein with multiple Sec residues) and selenoproteomes of SelP-containing organisms (91). The number of Sec residues in mammalian SelP varied by >2-fold and was lowest in rodents and primates. Compared with mammals, fish had more Sec residues in SelP and larger selenoproteomes. It was shown that several fish (such as *Danio rerio*) selenoproteins had Cys-containing homologs in mammals, whereas no fish had Cys-containing homologs of mammalian selenoproteins, suggesting unidirectional replacement of Sec in mammals (*i.e.* Sec (fish) → Cys (mammals)). At the same time, some fish selenoproteins that were also found in invertebrates had no homologs in mammals, suggesting loss of these proteins during evolution of mammals. Further analysis of the frequency of Sec → Cys and Cys → Sec changes in SelP indicated that the Sec → Cys transitions occurred 10 times more often than the Cys → Sec transitions (91). These data suggested that analyses of SelP, selenoproteomes, and Sec/Cys transitions may provide a genetic marker of selenium utilization in vertebrates.

The evolutionary dynamics of Sec- and Cys-containing forms and their conversions require improved statistical models for future studies. Recent analyses of the evolution of vertebrate selenoproteins indicated strict selection across Sec and Cys sites, which is consistent with a unique role of Sec in protein function and its low exchangeability with Cys (92, 93). It was also suggested that distinct biochemical properties of Sec, rather than the geographical distribution of selenium, oxygen levels, or Sec metabolic cost, play a role in driving adaptive changes in vertebrate selenoproteomes (92). In the future, additional genomic data and statistical models are needed to fully address these questions in both vertebrates and other organisms.

Conclusions

Comparative genomics provides a powerful tool for studying trace element utilization and its evolutionary trends. Most of these studies used strategies based on either identification of metalloproteomes using metal-binding motifs or investigation of trace element utilization traits (*e.g.* transporters, cofactor biosynthesis pathways, and user proteins). Currently, it is difficult to identify complete metalloproteomes for most metals as there is no reliable method to predict all metal-binding proteins. A very recent study that was based on liquid chromatography, high-throughput tandem mass spectrometry, and inductively coupled plasma mass spectrometry analyses showed that metalloproteomes are much more extensive and diverse than previously recognized (67). Nevertheless, comparative studies have provided significant advances in our understanding of the general principles of utilization of trace elements across the three domains of life. These studies offer new insights and help to understand the evolutionary dynamics of trace element dependence in proteins and organisms. In this minireview, we have discussed recent comparative genomics advances for studying the metal dependence of metal-binding protein families and the selenium dependence of selenoprotein families. Most metal-independent proteins evolved from original metal-dependent forms that use zinc and copper, whereas molybdenum, nickel, and vitamin B12 are uniquely used and cannot be easily replaced or lost while preserving protein function. On the other hand, almost all selenoproteins have Cys-containing homologs, and analyses of Sec/Cys conversion events provide a great tool for examining the evolution of selenium utilization. In the next few years, with the increased availability of sequenced genomes and improved tools for their analyses, comparative genomics should play a significant role in the better understanding of trace element utilization and evolution.

REFERENCES

1. Prentice, A. M. (2005) *Public Health Nutr.* 8, 932–939
2. Swinburn, B. A., and Ravussin, E. (1994) *Baillieres Clin. Endocrinol. Metab.* 8, 527–548
3. Fraga, C. G. (2005) *Mol. Aspects Med.* 26, 235–244
4. Baran, E. J. (2004) *Mini-Rev. Med. Chem.* 4, 1–9
5. Mertz, W. (1998) *Biol. Trace Elem. Res.* 66, 185–191
6. Keen, C. L., Clegg, M. S., Hanna, L. A., Lanoue, L., Rogers, J. M., Daston, G. P., Oteiza, P., and Uriu-Adams, J. Y. (2003) *J. Nutr.* 133, 15975–16055
7. Ames, B. N. (2006) *Proc. Natl. Acad. Sci. U.S.A.* 103, 17589–17594
8. Ames, B. N. (2001) *Mutat. Res.* 475, 7–20
9. Galaris, D., and Pantopoulos, K. (2008) *Crit. Rev. Clin. Lab. Sci.* 45, 1–23
10. Frassinetti, S., Bronzetti, G., Calvatutturo, L., Cini, M., and Croce, C. D. (2006) *Environ. Pathol. Toxicol. Oncol.* 25, 597–610
11. Salgueiro, J., Zubillaga, M., Lysionek, A., Sarabia, M. I., Calmanovici, G., Caro, R., De Paoli, T., Hager, A., Weill, R., and Bocci, J. (1999) *Acta Physiol. Pharmacol. Ther. Latinaom.* 49, 1–12
12. Posey, J. E., and Gherardini, F. C. (2000) *Science* 288, 1651–1653
13. Archibald, F. (1983) *FEBS Microbiol. Lett.* 19, 29–32
14. Coleman, J. E. (1992) *Annu. Rev. Biochem.* 61, 897–946
15. Mendel, R. R., Smith, A. G., Marquet, A., and Warren, M. J. (2007) *Nat. Prod. Rep.* 24, 963–971
16. Mendel, R. R., and Bittner, F. (2006) *Biochim. Biophys. Acta* 1763, 621–635
17. Krammer, T. (2005) *Biochem. Soc. Trans.* 33, 806–810
18. Böck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Vereyk, B., and Zinoni, F. (1991) *J. Mol. Microbiol.* 5, 515–520
19. Degtjareiko, K. (2000) *Bioinformatics* 16, 851–864
20. Sukdeo, N., Clugston, S. L., Daub, E., and Honek, J. F. (2004) *Biochem. J.* 384, 111–117
21. Gladyshev, V. N. (2007) in *Selenium: Its Molecular Biology and Role in Human Health* (Hatfield, D. L., Berry, M. J., and Gladyshev, Y. N., eds) 2nd Ed., p. 83, Springer, New York
22. Hardison, R. C. (2003) *Proc. Biol. 1, E58
23. Filipski, A., and Kumar, S. (2005) in *The Evolution of the Genome* (Gregory, T. R., ed) pp. 521–583, Elsevier, San Diego, CA
24. Gregory, T. R., and DeSalle, R. (2005) in *The Evolution of the Genome* (Gregory, T. R., ed) pp. 585–675, Elsevier, San Diego, CA
25. Shi, W., Zhan, C., Ignatov, A., Manjasetty, B. A., Marinkovic, N., Sullivan,
