PART I

INTRODUCTION

1

NETWORKS IN BIOLOGY

BJÖRN H. JUNKER

1.1 INTRODUCTION

Our environment is a combination of tightly interlinked complex systems at various levels of magnitude. While the exact sciences of physics and chemistry describe our environment from subatomic level up to the molecular level, biology is carrying the burden to deal with an inexact and extremely complex universe that sometimes even seems lawless. Yet biological systems follow "laws" that physicists would rather refer to as "probabilities." By these laws, it is possible to describe biology at different detail levels with a certain precision.

The smallest biological detail level is the molecular level of DNA, RNA, proteins, and metabolites. All these molecules are ingredients of a cell, which in turn is a part of a tissue. Different tissues constitute the organs of an organism. Many organisms together form the ecosystem. Additionally, over time these organisms are subjected to evolution, which results in a certain phylogenetic relationship between them. At all these levels of detail, the relationships between the elements are of great interest. These relationships can be described as networks, in which the elements are the vertices (nodes, points) and the relationships are the edges (arcs, lines; see Chapter 2). Typical biological networks at the molecular level are gene regulation networks, signal transduction networks, protein interaction networks, and metabolic networks. An example of a biological network is given in Fig. 1.1. While parts of all these networks have been modeled since a long time, recent technological

Analysis of Biological Networks, Edited by Björn H. Junker and Falk Schreiber

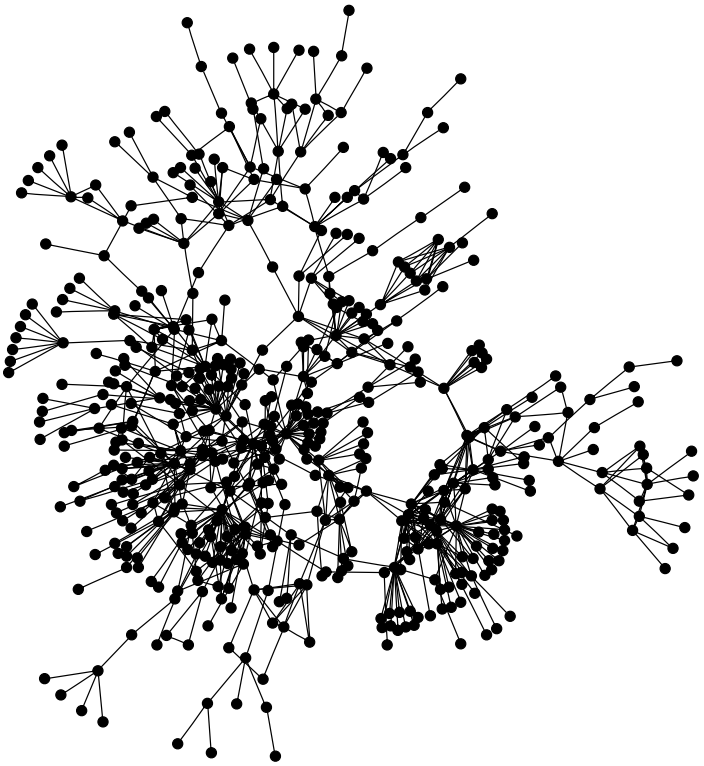

FIGURE 1.1 Example of a biological network. The largest strongly connected component (see Chapter 2) of the human protein interaction network is shown. The network is based on the complete data set for interaction of human proteins downloaded from the Database of Interacting Proteins (DIP, [35]) in January 2005.

advances have made it possible to elicit entire networks, or at least large proportions of them.

The next section contains a concise overview of basic biology and is especially aimed at readers who would like to refresh their knowledge of biology. Section 1.3 introduces the concept of systems biology. In Section 1.4, an overview is given about what findings have been made about different biological networks with modern network analysis methods.

1.2 BIOLOGY 101

1.2.1 Biochemistry and Molecular Biology

The information about the assembly of an organism is stored in the desoxyribonucleic acid (DNA, see Fig. 1.2). DNA is a coiled ladder (*helix*) consisting of two sugar phosphate backbones enclosing pairs of the nucleotide bases adenine, cytosine, guanine, and thymine (A,C,G,T). The nucleotide A pairs only with T, whereas C pairs only with

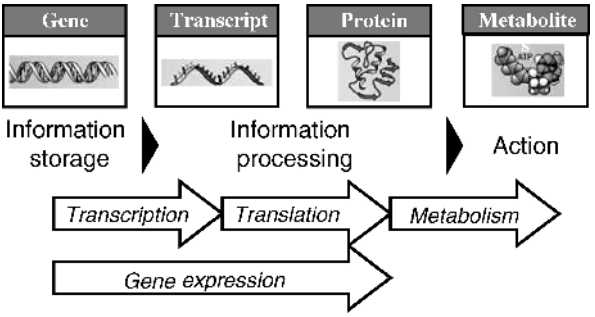

FIGURE 1.2 Information flow from genes to metabolites in cells.

G. While DNA constitutes the passive part of the cell's biochemistry, the active part is contributed by the proteins, as they catalyze reactions and are responsible for many other mechanisms in the cell. The process of information transmission from DNA to proteins is called *gene expression* (Fig. 1.2). This process can be divided into two main parts, *transcription* and *translation*. Transcription is a complicated, highly regulated process, in which a protein complex containing the RNA polymerase opens the DNA helix, reads one strand, and synthesizes a corresponding ribonucleic acid (RNA) like a blueprint. Transcription is initiated and terminated at certain signal sequences, which are called *promoter* and *terminator*, respectively. The corresponding RNA to a certain gene is called *transcript* (Fig. 1.2). In *eukaryotes* (see next section), the RNA then undergoes a process called *splicing*, in which the *introns* (noncoding regions) are excised so that only the *exons* (coding regions) remain. During translation, amino acid chains are synthesized from the (spliced) RNA by the ribosomes. The information of the RNA is read in triplets (*codons*), for which there are $4^3 = 64$ combinations. These are used to code both for 20 amino acids (sometimes more than one codon stands for one amino acid), as well as one start codon and three stop codons.

The structure of a protein is important for its functionality. The primary protein structure is simply the amino acid sequence, where as the secondary structure consists of regular three-dimensional patterns such as loops, helices, or sheets. Furthermore, the tertiary structure describes how these patterns are arranged in space to form a protein or a subunit thereof. Finally, the quaternary structure depicts how the different amino acid chains of the subunits are arranged to form an active protein complex. Proteins can play many different roles in the cell, for example, structural proteins that stabilize the cell's structure, transcription factors that regulate the process of transcription, or enzymes enzyme that catalyze reaction in which one metabolite is converted into another.

Metabolite is a term for all molecules of low molecular weight, such as sugars or amino acids. All the processes mentioned above are subject to tight regulation, which can take place at different levels. An environmental or internal signal (e.g., light, hormone) can be at first multiplied and processed through signal transduction chains. Then, a regulatory action can take place, for example, at the transcriptional level through activation or repression of gene expression, or at the protein level through posttranslational modification.

1.2.2 Cell Biology

Depending on the domain of life, the cells of an organism are organized in different ways. *Prokaryotes* such as bacteria are single cells that are not further subdivided. Their *genome*, the totality of the genes, is organized in one single circular chromosome. In contrast, the cells of *eukaryotes* are structured much more complex (see Fig. 1.3). Like prokaryotes, the cells are filled with the *cytoplasm*, but contrary to prokaryotes, additional *organelles* are separated from the cytoplasm through membranes. Organelles are, for example, *mitochondria* that produce chemical energy, the *endoplasmatic reticulum* that plays a role in protein synthesis, and the *nucleus* (Fig. 1.3). Plant cells are equipped with additional organelles, the *plastids*, which is an umbrella term for *chloroplasts* (responsible for photosynthesis), *chromoplasts* (pigment synthesis and storage), *amyloplasts* (starch synthesis and storage), and *vacuoles* that serve as storage organelle for metabolites.

Inside the nucleus, the genome is organized in several chromosomes, each of which is consisting of two *chromatides*, parallel coils that are connected near the middle to form an x-like structure. On the gene level, this means that a eukaryotic cell generally has at least two copies of every gene. Further on, most cells in most organisms are equipped with two sets of chromosomes, one from each parent.

In living organisms, there is a variety of different cell types responsible for various functions. A number of cells that perform a similar function constitute a *tissue*, examples for animal cells are epithelium and connective tissue, for plant cells epidermis or vascular tissue. A group of tissues that perform a specific function or a set of functions form an *organ*. Typical organs in animals are brain, lung, and liver. Typical organs in plants are leafs, stem, and seeds. All organs together constitute the entire organism.

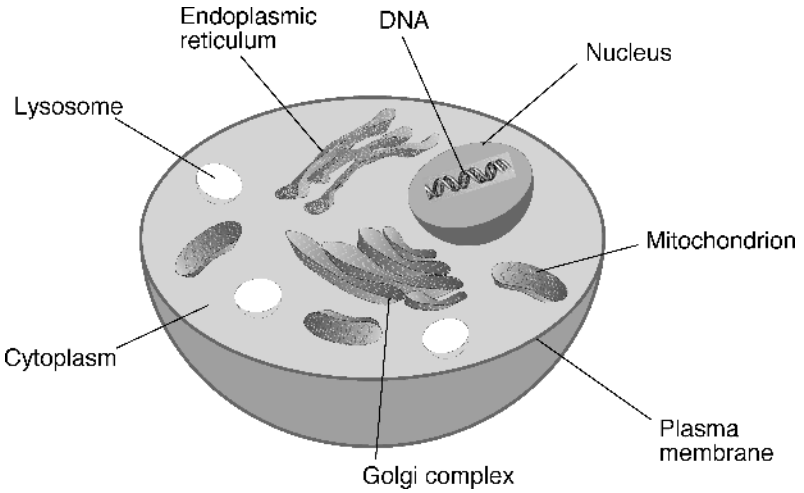

FIGURE 1.3 Schematic illustration of an animal cell with some organelles.

1.2.3 Ecology and Evolution

In the previous two sections, an overview was given about the biochemical and cellular composition of a single organism. In our environment, the organisms constantly interact with each other and are integrated components of the ecosystem. The influence of one organism on another is called a *biotic factor*. This influence might be the predator–prey relationship between two animals, or the relationship between a plant and an insect pollinating this plant. Further on, organisms are influenced also by *abiotic factors* such as climate and geology.

Organisms are subjected to evolution over large timescales. Evolution is the process by which populations of organisms acquire and pass on novel traits from generation to generation. The modern theory of evolution is based on the concept of natural selection, as first outlined in Darwin's 1859 book "*The Origin of Species*" [9]. Individual organisms that possess advantageous traits will be more likely to pass on their genes. In the 1930s, Darwin's theory was combined with Mendel's heredity laws to create the *modern synthesis*, which explains evolution as a change in frequency of alleles within a population between two generations. In modern times, sequence information from certain genes is used to derive evolutionary relationships between different organisms. From this data, phylogenetic trees can be constructed at different detail levels of the taxonomy (Fig. 1.4).

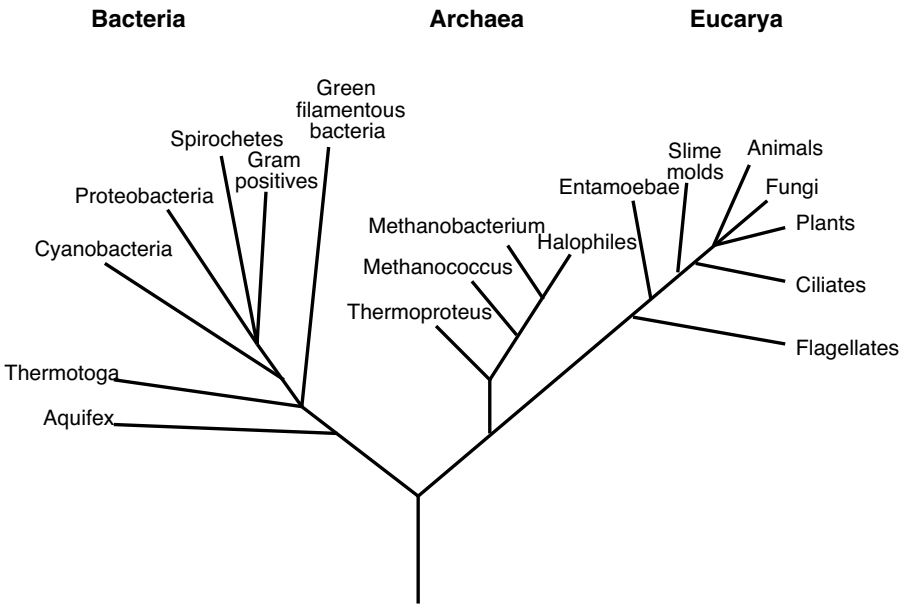

FIGURE 1.4 A speculative phylogenetic tree showing the separation of the three domains of life. Exemplary groups are shown, which represent different detail levels of the phylogeny.

1.3 SYSTEMS BIOLOGY

Biology is currently in the starting phase of a shift that will ultimately transform it into a precise science similar to physics and chemistry. The term "systems biology" is drawing more and more attention. While the origin of systems biology dates back to at least 1969 when Ludwig von Bertalanffy described his systems theory [37], it faced an explosion of interest in the new millennium. Hiroaki Kitano defined systems biology in his book "*Foundations of System Biology*" as "systems biology is a new field in biology that aims at system-level understanding of biological systems" [21]. That means the ultimate goal of systems biology is to understand entire biological systems by elucidating, modeling, and predicting the behavior of all components and interactions.

The central step toward a systems-level understanding of biology was to move away from *reductionist* to *wholist* approaches, sometimes also called bottom-up and top-down approaches, respectively [20]. Traditionally, reductionists look at *one* element of the system to find out the connections to neighbors, roles in all processes that the element is involved in, and mechanisms of action. In contrast, the wholist approach is to first make a snapshot of *all* elements at a certain level (genes, transcripts, proteins, and/or metabolites; see also Fig. 1.2). For this task, since the 1990s, many massively parallel experimental techniques have been developed. The entire set of components of one kind is described with terms ending *-ome* (genome, proteome), whereas the techniques to identify this set ends with *-omics* (genomics, proteomics). To date, more than hundred of these -omics technologies have been defined [1]. While some of them are just new words for old things, some others open an entirely new view on biological systems. The genomes of many organisms were sequenced, starting with *Escherichia coli* in 1997 [8], to reach 680 complete published genomes in November 2007 [2]. Recent technological developments will likely result in an exponential increase of this number [26]. Snapshots of the transcriptome (set of all RNA molecules of one biological sample, [10]) are routinely measured in laboratories all over the world. By the help of experimental techniques such as two-dimensional gels and mass spectrometry, the proteomes of several organisms can be determined [31]. Another recent development is metabolomics, in which a large number of metabolites are measured simultaneously in one sample [13]. With other "-omics" technologies, other "-omes" have been measured, such as the fluxome (the fluxes through metabolic pathways) or the interactome (the interactions between proteins and small molecules). Having established these high-throughput experimental techniques, scientists were confronted with the problem of how to make sense out of the wealth of generated data. One possible solution will be presented in the next section.

1.4 PROPERTIES OF BIOLOGICAL NETWORKS

Just as it is impossible to assemble an airplane by using a list of all parts, it seems impossible to gain any useful information out of the wealth of data generated with the

-omics methods detailed above. One particularly promising approach for the generation of hypothesis out of this data is network analysis, such promising that this entire book is dedicated to this area of research. While network analysis is not a new research field, it is noticeable that some fundamental properties of networks have been elucidated just at the change of the millennium. In 1998, Watts and Strogatz published a paper in which they illustrate that the neural network of the worm *Caenorhabditis elegans*, the power grid of the Western United States, and the collaboration graph of film actors have similar properties: they are highly clustered (densely connected subgraphs, see Chapter 2), yet they have small characteristic path lengths (see Chapter 2) [39]. The authors created the term *small world networks* for this phenomenon, by analogy with the popular small-world phenomenon [27], which states that any person on our planet links to any other person by a chain of on average six acquaintances. One year later, Barabási and Albert created a simple model for these networks, which they found to follow a scale-free power-law distribution and thus named them *scale-free networks* [5]. The consequence of this connectivity distribution is that many vertices have few links, while there are some that are highly connected. As a result, scale-free networks are very robust against failure, such as removal of arbitrary network elements [3]. To date it has been found that power grids, the Internet (routers and cables), the World Wide Web (webpages and links), protein interaction networks, metabolic networks, and many other networks follow these general rules [4]. However, the first obstacle for the application of these methods in biological research is the generation of networks out of the data sets determined with the -omics technologies. Because it is not possible to directly infer any networks from sequences, or from transcript, protein, or metabolite concentrations, additional information is needed, such as information about interactions. In the following sections, it will be briefly discussed which sources are available to derive biological networks, and which novel findings have been made investigating these networks.

1.4.1 Networks on a Microscopic Scale

Biochemical networks have been under investigation for many decades. However, the efforts were until recently limited to the determination of the components of the networks, rather than addressing the design principles of its structure. The fundamental finding about all kinds of networks (as mentioned above) have also been investigated in biological networks, such as regulation networks, protein interaction networks, and metabolic networks.

Transcriptional regulation networks (or gene regulation networks) are controlling gene expression in cells. The expression of one gene can be controlled by the gene product of another gene. Thus, a directed graph (see Chapter 2) in which the vertices are genes and the directed edges represent control can be used to model these networks. Until recently, only fragments of these networks have been modeled, usually quantitatively, by assigning rate laws to every step. For example, quantitative models containing selected genes have greatly improved the understanding of morphogenesis of early embryos of the fruit fly *Drosophila melanogaster* [16]. Recent advances in

data collection and analysis made it possible to elucidate large-scale gene regulation networks [23]. It has been found that in this network type, certain motifs (small recurring patterns, see Chapter 5) such as feed-forward loops or single input modules are overrepresented when compared with randomly generated networks [23,36]. Through these investigations it was possible to define the "basic computational elements" of biological networks.

Signal transduction networks can be understood as gene regulation networks extended by signaling chains that contain different kinds of vertices and edges such as protein–protein interaction and phosphorylation. By quantitative modeling, emergent properties have been found in these networks such as integration of signals across multiple timescales, generation of distinct outputs depending on input strength and duration, and self-sustaining feedback-loops [7]. A more detailed explanation of gene regulation and signal transduction networks together with scientific results is given in Chapter 8.

Protein interaction networks are generated out of different types of large-scale experimental and computational approaches [38]. The different methods are resulting in significantly different networks, so that we can speak only of a network for a certain organism determined by using a certain method. The protein interaction network of the baker's yeast (*Saccharomyces cerevisiae*) as determined by systematic two-hybrid analyses was found to follow the laws of scale-free networks [17]. Furthermore, it has been shown that the most highly connected proteins in the cell are the most important for its survival [17]. In the network, this corresponds to the vertices with the highest number of connections (high degree centrality, see Chapter 4). In the same network, it has been shown that certain motifs are overrepresented [41] (see Chapter 5). Through comparison with orthologous networks from other higher eukaryotes, the authors found that these motifs are evolutionarily conserved. More details on protein interaction networks are given in Chapter 9.

Metabolic networks consist of metabolites that are converted into each other by enzymes. These networks have been determined through biochemical experiments over the last few decades, and they can be found in various kinds of biochemistry textbooks. A summary of biochemical pathways is given in the well-known Boehringer map [12]. Since few years, metabolic pathways have also been predicted from the genome of fully sequenced organisms. The KEGG database [19] is a public resource for these predicted pathways. In an early study, it was found that the large-scale structure of the core metabolic network from 43 organisms is identical, being dominated by the same highly connected substrates [18]. For the same set of metabolic networks, it has been stated later that they are organized into many small, highly connected topologic modules that combine in a hierarchical manner into larger, less cohesive units [34]. Several other studies have compared the structure of the metabolic networks of several organisms in order to derive information about their phylogenetic relationship [14,25,32]. While these first studies could not replicate the detail of phylogenetic studies based on sequence information, it was at least possible to deduce from the network whether an organism belongs to the domains of Archaea, Bacteria, or Eukaryotes (see also Fig. 1.4). A more detailed discussion of metabolic networks can be found in Chapter 10.

1.4.2 Networks on a Macroscopic Scale

As stated before, networks are also present in the areas of biology dealing with larger space- or timescales. The interactions of different organisms can be depicted as ecological networks. Food webs have been under investigation since a long time. Qualitative food webs, which contain information only about predator–prey relationship, but no quantities, can be modeled as directed graphs (see Chapter 2). In this context, qualitative food webs are often called *static models*. However, they have not been the subject of many studies [11]. This is probably due to the fact that the available food webs are relatively small compared with biochemical networks, and thus not much new information can be gained out of the structure alone. Nevertheless, through comparison of 50 food webs of lakes, it was found that a relation exists between the number of species in a food web and the links per species [15]. Instead of investigating the structure of food webs alone, they are often modeled quantitatively with rate laws for every step (*dynamic models* [11]). Ecological networks other than food webs can be, for example, plant–pollinator interaction networks, which were found to exhibit an increased number of interactions per species upon increased diversity [28], analogous to the food webs mentioned above. More details on ecological networks are given in Chapter 12.

Phylogenetic networks describe the evolutionary relationships between organisms. Traditionally, they were presented as bifurcate or binary trees (see Chapter 2) [29]. The branchpoints of the tree represent points of separation of two species during evolution. However, recent studies suggest that population genealogies are often multifurcated (trees, see Chapter 2), or even containing reticulate relationships due to recombination events, which turns them into phylogenetic networks [33]. Recently, a network for the phylogenetic relationships between all groups of prokaryotes has been presented and termed the " net of life" [22]. A more detailed discussion of phylogenetic trees and networks can be found in Chapter 11. As mentioned in the previous section, this topic is linked to several biochemical networks through many studies that have been made to infer phylogeny especially from metabolic networks. Recently, it has been shown that bacterial metabolic networks evolve adaptively by horizontal gene transfer [30].

1.4.3 Other Biological Networks

Correlation networks have only been investigated for a relatively short time, and they represent an exception among biological networks. Their special feature is that these networks are not a direct result of experimental data, but they are determined by collecting large amounts of high-throughput data and calculating the correlations between all elements. So far this has been done for transcripts and metabolites. Barkai and coworkers compared large-scale gene expression data sets of six evolutionarily distant organisms [6]. They found that for all organisms the connectivity of the correlation network follows a power-law, highly connected genes tend to be essential and conserved, and the expression program is highly modular. Furthermore, transcript correlation networks have been used to identify hormone-related genes in plants [24]. Metabolite correlation networks have been constructed from pair-wise analysis of linear correlations between metabolites from profiling data [40]. It was found that

the connectivity distribution in these networks also follows the typical power-law for scale-free networks. More examples of correlation networks and their analysis are given in Chapter 13.

1.5 SUMMARY

Biology describes the processes of our environment from the molecular level to the level of the ecosystem. At all levels of detail, many of the respective processes can be modeled by networks. At the microscopic levels, these are gene regulation networks, signal transduction networks, protein interaction networks, and metabolic networks. At the macroscopic level, these are ecological and phylogenetic networks. All these networks have some special characteristics and are quite distinct from each other, but they also share common properties. Although the analysis of large-scale biological networks with modern tools has made significant progress in the last decade, this branch of science is still in its infancy.

1.6 EXERCISES

1. Describe the information flow within a cell, from DNA to metabolism. Name the processes.
2. What are the four levels of protein structure?
3. Describe the organization of a cell.
4. In a regular cell of most organisms, how many copies of each gene are present? Why?
5. Describe the term "systems biology" in your own words.
6. What are -omes and -omics?
7. Why is the measurement of a complete transcriptome not yielding a network?
8. Name at least four microscopic and two macroscopic networks in biology.
9. Why are correlation networks not intrinsic biological networks?

REFERENCES

1. -omes and -omics glossary and taxonomy. http://www.genomicglossaries.com/content/omes.asp.
2. Gold—genomes online database. http://www.genomesonline.org/.
3. R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–381, 2000.
4. A.-L. Barabási. Emergence of scaling in complex networks. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks*, pp. 69–84. Wiley-VCH, Weinheim (Germany) 2003.
5. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

6. S. Bergmann, J. Ihmels, and N. Barkai. Similarities and differences in genomewide expression data of six organisms. *PLoS Biology*, 2:85–93, 2004.
7. U. S. Bhalla and R. Iyengar. Emergent properties of networks of biological pathways. *Science*, 283:381–387, 1999.
8. F. R. Blattner, G. Plunket, III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. The complete genome sequence of *Escherichia coli* k-12. *Science*, 277:1453–1462, 1997.
9. C. Darwin. *On the Origin of Species by Means of Natural Selection*. John Murray, London, 1859.
10. J. L. DeRisi, V. R. Iyer, and P. O. Brown. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278:680–686, 1997.
11. B. Drossel and A. J. McKane. Modelling food webs. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks*, pp. 218–247. Wiley-VCH, Weinheim (Germany) 2003.
12. G. Michal. *Biochemical Pathways (wall charts)*. Boehringer, Mannheim, Basle Switzerland, 1993.
13. O. Fiehn, J. Kopka, P. Dormann, T. Altmann, R. N. Trethewey, and L. Willmitzer. Metabolite profiling for plant functional genomics. *Nature Biotechnology*, 18:1157–1161, 2000.
14. C. V. Forst and K. Schulten. Phylogenetic analysis of metabolic pathways. *Journal of Molecular Evolution*, 52:471–489, 2001.
15. K. Havens. Scale and structure in natural food webs. *Science*, 257:1107–1109, 1992.
16. J. Jaeger, S. Surkova, M. Blagov, H. Janssens, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz. Dynamic control of positional information in the early *Drosophila* embryo. *Nature*, 430:368–371, 2004.
17. H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41–42, 2001.
18. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 107:651–654, 2000.
19. M. Kanehisa and S. Goto. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28:27–30, 2000.
20. F. Katagiri. Attacking complex problems with the power of systems biology. *Plant Physiology*, 132:417–419, 2003.
21. H. Kitano. *Foundations of Systems Biology*. The MIT Press, Cambridge, MA, 2001.
22. V. Kunin, L. Goldovsky, N. Darzentas, and C. A. Ouzounis. The net of life: Reconstructing the microbial phylogenetic network. *Genome Research*, 15:954–959, 2005.
23. T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, W. Fraenkel, D. K. Gifford, and R. A. Young. Transcriptional regulatory networks in *Saccharomyces cerivisiae*. *Science*, 298:799–804, 2002.
24. J. Lisso, D. Steinhauser, T. Altmann, J. Kopka, and C. Müssig. Identification of brassinoid-related genes by means of transcript co-response analyses. *Nucleic Acids Research*, 33:2685–2696, 2005.

25. H.-W. Ma and A.-P. Zeng. Phylogenetic comparison of metabolic capacities of organisms at genome level. *Molecular Phylogenetics and Evolution*, 31:204–213, 2004.
26. M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P., Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437:376–380, 2005.
27. S. Milgram. The small-world problem. *Psychology Today*, 1:61–67, 1967.
28. J. M. Olesen and P. Jordano. Geographic patterns in plant-pollinator mutualistic networks. *Ecology*, 83:2416–2424, 2002.
29. N. R. Pace. A molecular view of microbial diversity and the biosphere. *Science*, 276:734–740, 1997.
30. C. Pál, B. Papp, and M. J. Lercher. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature Genetics*, 37:1372–1375, 2005.
31. A. Pandey and M. Mann. Proteomics to study genes and genomes. *Nature*, 405:837–846, 2000.
32. J. Podani, Z. N. Oltvai, H. Jeong, B. Tombor, A.-L. Barabási, and E. Szathmáry. Comparable systems-level organization of archaea and eukaryotes. *Nature Genetics*, 29:54–56, 2001.
33. D. Posada and K. A. Crandall. Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, 16:37–45, 2001.
34. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297:1551–1555, 2002.
35. L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, and D. Eisenberg. The database of interacting proteins: 2004 update. *Nucleic Acids Research*, 32:D449–D451, 2004.
36. S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, 31:64–68, 2002.
37. L. von Bertalanffy. *General Systems Theory: Foundations, Development, Applications*. George Braziller, New York (NY, USA) 1969.
38. C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417:309–403, 2002.
39. D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.
40. W. Weckwerth, M. E. Loureiro, K. Wenzel, and O. Fiehn. Differential metabolic networks unravel the effect of silent plant phenotypes. *Proceedings of the National Academy of Sciences USA*, 101:7809–7814, 2004.
41. S. Wuchty, Z. N. Oltvai, and A.-L. Barabási. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature Genetics*, 35:176–179, 2003.