Cl concentrations in the Sajama ice core, and to a number of other pedological and geomorphological features indicative of long-term dry climates (*8*, *11–14*, *18*). This decline in human activity around the Altiplano paleolakes is seen in most caves, with early and late occupations separated by largely sterile mid-Holocene sediments. However, a few sites, including the caves of Tulan-67 and Tulan-68, show that people did not completely disappear from the area. All of the sites of sporadic occupation are located near wetlands in valleys, near large springs, or where lakes turned into wetlands and subsistence resources were locally still available despite a generally arid climate (*7*, *8*, *19*, *20*).

Archaeological data from surrounding areas suggest that the Silencio Arqueológico applies best to the most arid areas of the central Andes, where aridity thresholds for early societies were critical. In contrast, a weaker expression is to be expected in the more humid highlands of northern Chile (north of 20°S, such as Salar Huasco) and Peru (*21*). In northwest Argentina, the Silencio Arqueológico is found in four of the six known caves (*22*) [see review in (*23*)]. It is also found on the coast of Peru in sites that are associated with ephemeral streams (*24*). The southern limit in Chile and northwest Argentina has yet to be explored.

References and Notes

1. T. Dillehay, *Science* **245**, 1436 (1989).
2. D. J. Meltzer *et al.*, *Am. Antiq.* **62**, 659 (1997).
3. T. F. Lynch, C. M. Stevenson, *Quat. Res.* **37**, 117 (1992).
4. D. H. Sandweiss *et al.*, *Science* **281**, 1830 (1998).
5. L. Núñez, M. Grosjean, I. Cartajena, in *Interhemispheric Climate Linkages*, V. Markgraf, Ed. (Academic Press, San Diego, CA 2001), pp. 105–117.
6. M. A. Geyh, M. Grosjean, L. Núñez, U. Schotterer, *Quat. Res.* **52**, 143 (1999).
7. J. L. Betancourt, C. Latorre, J. A. Rech, J. Quade, K. Rylander, *Science* **289**, 1542 (2000).
8. M. Grosjean *et al.*, *Global Planet. Change* **28**, 35 (2001).
9. C. Latorre, J. L. Betancourt, K. A. Rylander, J. Quade, *Geol. Soc. Am. Bull.* **114**, 349 (2002).
10. Charcoal in layers containing triangular points has been ^{14}C dated at Tuina-1, Tuina-5, Tambillo-1, San Lorenzo-1, and Tuyajto-1 between 13,000 and 9000 cal yr B.P. (table S1 and fig. S1).
11. P. A. Baker *et al.*, *Science* **291**, 640 (2001).
12. G. O. Seltzer, S. Cross, P. Baker, R. Dunbar, S. Fritz, *Geology* **26**, 167 (1998).
13. L. G. Thompson *et al.*, *Science* **282**, 1858 (1998).
14. M. Grosjean, *Science* **292**, 2391 (2001).
15. E. P. Tonni, written communication.
16. M. T. Alberdi, written communication.
17. J. Fernandez *et al.*, *Geoarchaeology* **6**, 251 (1991).
18. The histogram of middens is processed from (*9*).
19. M. Grosjean, L. Núñez, I. Cartajena, B. Messerli, *Quat. Res.* **48**, 239 (1997).
20. The term Silencio Arqueológico describes the mid-Holocene collapse of human population at those archaeological sites of the Atacama Desert that are vulnerable to multicentennial or millennial-scale drought. The term Silencio Archaeológico does not conflict with the presence of humans at sites that are not susceptible to climate change, such as in spring and river oases that drain large (Pleistocene) aquifers or at sites where wetlands were created during the arid middle Holocene, such as Tulan-67, Tulan-68, and Laguna Miscanti.
21. M. Aldenderfer, *Science* **241**, 1828 (1988).
22. A mid-Holocene hiatus is found at Inca Cueva 4, Huachichocana 3, Pintocamayoc, and Yavi, whereas occupation continued at the oases of Susques and Quebrada Seca.
23. L. Núñez *et al.*, *Estud. Atacamenos* **17**, 125 (1999).
24. D. H. Sandweiss, K. A. Maasch, D. G. Anderson, *Science* **283**, 499 (1999).

25. Grants from the National Geographic Society (5836-96), the Swiss National Science Foundation (21-57073), and Fondo Nacional de Desarrollo Científico y Tecnológico (1930022) and comments by J. P. Bradbury, B. Meggers, G. Seltzer, and D. Stanford are acknowledged.

Supporting Online Material

www.sciencemag.org/cgi/content/full/298/5594/821/DC1
Figs. S1 to S3
Tables S1 and S2

22 July 2002; accepted 9 September 2002

Network Motifs: Simple Building Blocks of Complex Networks

R. Milo,[1] S. Shen-Orr,[1] S. Itzkovitz,[1] N. Kashtan,[1] D. Chklovskii,[2] U. Alon[1]*

Complex networks are studied across many fields of science. To uncover their structural design principles, we defined "network motifs," patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks. We found such motifs in networks from biochemistry, neurobiology, ecology, and engineering. The motifs shared by ecological food webs were distinct from the motifs shared by the genetic networks of *Escherichia coli* and *Saccharomyces cerevisiae* or from those found in the World Wide Web. Similar motifs were found in networks that perform information processing, even though they describe elements as different as biomolecules within a cell and synaptic connections between neurons in *Caenorhabditis elegans*. Motifs may thus define universal classes of networks. This approach may uncover the basic building blocks of most networks.

Many of the complex networks that occur in nature have been shown to share global statistical features (*1–10*). These include the "small world" property (*1–9*) of short paths between any two nodes and highly clustered connections. In addition, in many natural networks, there are a few nodes with many more connections than the average node has. In these types of networks, termed "scale-free networks" (*4, 6*), the fraction of nodes having k edges, $p(k)$, decays as a power law $p(k) \sim k^{-\gamma}$ (where γ is often between 2 and 3). To go beyond these global features would require an understanding of the basic structural elements particular to each class of networks (*9*). To do this, we developed an algorithm for detecting network motifs: recurring, significant patterns of interconnections. A detailed application to a gene regulation network has been presented (*11*). Related methods were used to test hypotheses on social networks (*12, 13*). Here we generalize this approach to virtually any type of connectivity graph and find the striking appearance of

[1]Departments of Physics of Complex Systems and Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel 76100. [2]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.

*To whom correspondence should be addressed. E-mail: urialon@weizmann.ac.il

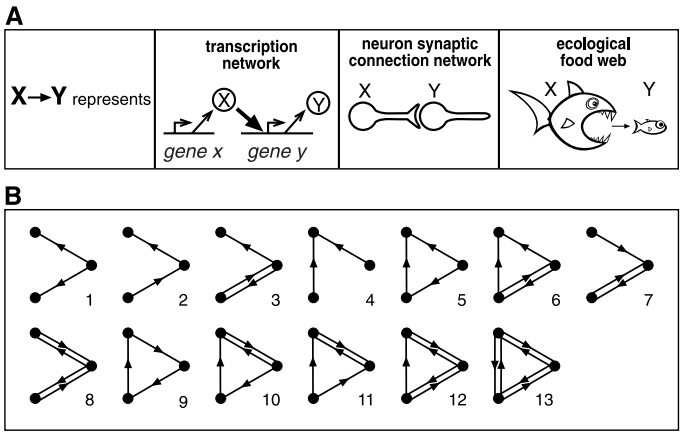

Fig. 1. (**A**) Examples of interactions represented by directed edges between nodes in some of the networks used for the present study. These networks go from the scale of biomolecules (transcription factor protein X binds regulatory DNA regions of a gene to regulate the production rate of protein Y), through cells (neuron X is synaptically connected to neuron Y), to organisms (X feeds on Y). (**B**) All 13 types of three-node connected subgraphs.

motifs in networks representing a broad range of natural phenomena.

We started with networks where the interactions between nodes are represented by directed edges (Fig. 1A). Each network was scanned for all possible n-node subgraphs (in the present study, $n = 3$ and 4), and the number of occurrences of each subgraph was recorded. Each network contains numerous types of n-node subgraphs (Fig. 1B). To focus on those that are likely to be important, we compared the real network to suitably randomized networks (*12–16*) and only selected patterns appearing in the real network at numbers significantly higher than those in the randomized networks (Fig. 2). For a stringent comparison, we used randomized networks that have the same single-node characteristics as does the real network: Each node in the randomized networks has the same number of incoming and outgoing edges as the corresponding node has in the real network. The comparison to this randomized ensemble accounts for patterns that appear only because of the single-node characteristics of the network (e.g., the presence of nodes with a large number of edges). Furthermore, the randomized networks used to calculate the significance of n-node subgraphs were generated to preserve the same number of appearances of all $(n-1)$-node subgraphs as in the real network (*17, 18*). This ensures that a high significance was not assigned to a pattern only because it has a highly significant subpattern. The "network motifs" are those patterns for which the probability P of appearing in a randomized network an equal or greater number of times than in the real network is lower than a cutoff value (here $P = 0.01$). Patterns that are functionally important but not statistically significant could exist, which would be missed by our approach.

We applied the algorithm to several networks from biochemistry (transcriptional gene regulation), ecology (food webs), neurobiology (neuron connectivity), and engineering (electronic circuits, World Wide Web). The network motifs found are shown in Table 1. Transcription networks are biochemical networks responsible for regulating the expression of genes in cells (*11, 19*). These are directed graphs, in which the nodes represent genes (Fig. 1A). Edges are directed from a gene that encodes for a transcription factor protein to a gene transcriptionally regulated by that transcription factor. We analyzed the two best characterized transcriptional regulation networks, corresponding to organisms from different kingdoms: a eukaryote (the yeast *Saccharomyces cerevisiae*) (*20*) and a bacterium (*Escherichia coli*) (*11, 19*). The two transcription networks show the same motifs: a three-node motif termed "feedforward loop" (*11*) and a four-node motif termed "bi-fan." These motifs appear numerous times in each network (Table 1), in nonhomologous gene systems that perform diverse biological functions. The number of times they appear is more than 10 standard deviations greater than their mean number of appearances in randomized networks. Only these subgraphs, of the 13 possible different three-node subgraphs (Fig. 1B) and 199 different four-node subgraphs, are significant and are therefore considered network motifs. Many other three- and four-node subgraphs recur throughout the networks, but at numbers that are less than the mean plus 2 standard deviations of their appearance in randomized networks.

We next applied the algorithm to ecosystem food webs (*21, 22*), in which nodes represent groups of species. Edges are directed from a node representing a predator to the node representing its prey. We analyzed data collected by different groups at seven distinct ecosystems (*22*), including both aquatic and terrestrial habitats. Each of the food webs displayed one or two three-node network motifs and one to five four-node network motifs. One can define the "consensus motifs" as the motifs shared by networks of a given type. Five of the seven food webs shared one three-node motif, and all seven shared one four-node motif (Table 1). In contrast to the three-node motif (termed "three chain"), the three-node feedforward loop was underrepresented in the food webs. This suggests that direct interactions between species at a separation of two layers [as in the case of omnivores (*23*)] are selected against. The biparallel motif indicates that two species that are prey of the same predator both tend to share the same prey. Both network motifs may thus represent general tendencies of food webs (*21, 22*).

We next studied the neuronal connectivity network of the nematode *Caenorhabditis elegans* (*24*). Nodes represent neurons (or neuron

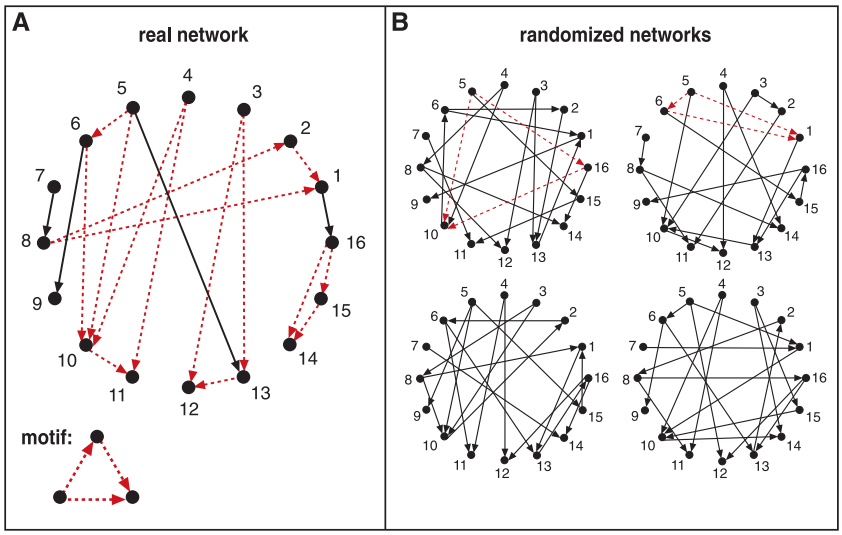

Fig. 2. Schematic view of network motif detection. Network motifs are patterns that recur much more frequently (**A**) in the real network than (**B**) in an ensemble of randomized networks. Each node in the randomized networks has the same number of incoming and outgoing edges as does the corresponding node in the real network. Red dashed lines indicate edges that participate in the feedforward loop motif, which occurs five times in the real network.

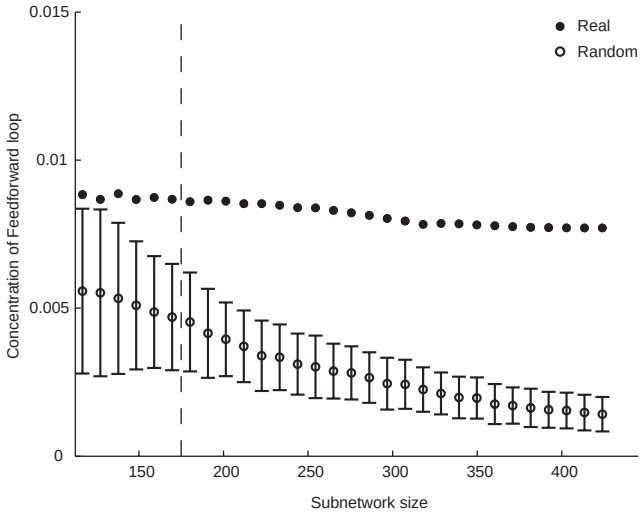

Fig. 3. Concentration C of the feedforward loop motif in real and randomized subnetworks of the *E. coli* transcription network (*11*). C is the number of appearances of the motif divided by the total number of appearances of all connected three-node subgraphs (Fig. 1B). Subnetworks of size S were generated by choosing a node at random and adding to it nodes connected by an incoming or outgoing edge, until S nodes were obtained, and then including all of the edges between these S nodes present in the full network. Each of the subnetworks was randomized (*17, 18*) (shown are mean and SD of 400 subnetworks of each size).

classes), and edges represent synaptic connections between the neurons. We found the feedforward loop motif in agreement with anatomical observations of triangular connectivity structures (*24*). The four-node motifs include the bi-fan and the bi-parallel (Table 1). Two of these motifs (feedforward loop and bi-fan) were also found in the transcriptional gene regulation networks. This similarity in motifs may point to a fundamental similarity in the design constraints of the two types of networks. Both networks function to carry information from sensory components (sensory neurons/transcription factors regulated by biochemical signals) to effectors (motor neurons/structural genes). The feedforward loop motif common to both types of networks may play a functional role in information processing. One possible function of this circuit is to activate output only if the input signal is persistent and to allow a rapid deactivation when the input goes off (*11*). Indeed, many of the input nodes in the neural feedforward loops are sensory neurons, which may require this type of information processing to reject transient input fluctuations that are inherent in a variable or noisy environment.

We also studied several technological networks. We analyzed the ISCAS89 benchmark set of sequential logic electronic circuits (*7, 25*). The nodes in these circuits represent logic gates and flip-flops. These nodes are linked by directed edges. We found that the motifs separate the circuits into classes that correspond to the circuit's functional description. In Table 1, we present two classes, consisting of five forward-logic chips and three digital fractional multipliers. The digital fractional multipliers share three motifs, including three- and four-node feedback loops. The forward logic chips share the feedforward loop, bi-fan, and bi-parallel motifs, which are similar to the motifs found in the genetic and neuronal information-processing networks. We found a different set of motifs in a network of directed hyperlinks between World Wide Web pages within a single domain (*4*). The World Wide Web motifs may reflect a design aimed at short paths between related pages. Application of our approach to nondirected networks shows distinct sets of motifs in networks of protein interactions and Internet router connections (*18*).

None of the network motifs shared by the food webs matched the motifs found in the gene regulation networks or the World Wide Web. Only one of the food web consensus motifs also appeared in the neuronal network. Different motif sets were found in electronic circuits with different functions. This suggests that motifs can define broad classes of networks, each with specific types of elementary structures. The motifs reflect the underlying processes that generated each type of network; for example, food webs evolve to allow a flow of energy from the bottom to the top of food chains, whereas gene regulation and neuron networks evolve to process information. Information processing seems to give rise to significantly different structures than does energy flow.

We further characterized the statistical significance of the motifs as a function of network size, by considering pieces of various sizes (subnetworks) of the full network. The concentration of motifs in the subnetworks is about the same as that in the full network (Fig. 3). In contrast, the concentration of the corresponding subgraphs in the randomized versions of the subnetworks decreases sharply with size. In analogy with statistical physics, the number of appearances of each motif in the real networks

Table 1. Network motifs found in biological and technological networks. The numbers of nodes and edges for each network are shown. For each motif, the numbers of appearances in the real network (N_{real}) and in the randomized networks ($N_{rand} \pm$ SD, all values rounded) (*17, 18*) are shown. The *P* value of all motifs is $P < 0.01$, as determined by comparison to 1000 randomized networks (100 in the case of the World Wide Web). As a qualitative measure of statistical significance, the *Z* score $= (N_{real} - N_{rand})/SD$ is shown. NS, not significant. Shown are motifs that occur at least $U = 4$ times with completely different sets of nodes. The networks are as follows (*18*): transcription interactions between regulatory proteins and genes in the bacterium *E. coli* (*11*) and the yeast *S. cerevisiae* (*20*); synaptic connections between neurons in *C. elegans*, including neurons connected by at least five synapses (*24*); trophic interactions in ecological food webs (*22*), representing pelagic and benthic species (Little Rock Lake), birds, fishes, invertebrates (Ythan Estuary), primarily larger fishes (Chesapeake Bay), lizards (St. Martin Island), primarily invertebrates (Skipwith Pond), pelagic lake species (Bridge Brook Lake), and diverse desert taxa (Coachella Valley); electronic sequential logic circuits parsed from the ISCAS89 benchmark set (*7, 25*), where nodes represent logic gates and flip-flops (presented are all five partial scans of forward-logic chips and three digital fractional multipliers in the benchmark set); and World Wide Web hyperlinks between Web pages in a single domain (*4*) (only three-node motifs are shown). e, multiplied by the power of 10 (e.g., 1.46e6 $= 1.46 \times 10^6$).

Network	Nodes	Edges	Feed-forward loop N_{real}	Feed-forward loop $N_{rand} \pm$ SD	Feed-forward loop *Z* score	Bi-fan N_{real}	Bi-fan $N_{rand} \pm$ SD	Bi-fan Z score	Bi-parallel N_{real}	Bi-parallel $N_{rand} \pm$ SD	Bi-parallel *Z* score
Gene regulation (transcription)					**Feed-forward loop**			**Bi-fan**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs					**Three chain**			**Bi-parallel**			
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)					**Three-node feedback loop**			**Bi-fan**			**Four-node feedback loop**
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web					**Feedback with two mutual dyads**			**Fully connected triad**			**Uplinked mutual dyad**
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4±4e2	15,000	1.2e6	1e4 ± 2e2	5000

*Has additional four-node motif: (X→Z, W; Y→Z, W; Z→W), $N_{real} = 150$, $N_{rand} = 85 \pm 15$, $Z = 4$. †Has additional four-node motif: (X→Y, Z; Y→Z; Z→W), $N_{real} = 204$, $N_{rand} = 80 \pm 20$, $Z = 6$. The three-node pattern (X→Y, Z; Y→Z; Z→Y) also occurs significantly more than at random. It is not a motif by the present definition because it does not appear with completely distinct sets of nodes more than $U = 4$ times. ‡Has additional four-node motif: (X→Y; Y→Z, W; Z→X; W→X), $N_{real} = 914$, $N_{rand} = 500 \pm 70$, $Z = 6$. §Has two additional three-node motifs: (X→Y, Z; Y→Z; Z→Y), $N_{real} = 3e5$, $N_{rand} = 1.4e3 \pm 6e1$, $Z = 6000$, and (X→Y, Z; Y→Z), $N_{real} = 5e5$, $N_{rand} = 9e4 \pm 1.5e3$, $Z = 250$.

appears to be an extensive variable (i.e., one that grows linearly with the system size). These variables are nonextensive in the randomized networks. The existence of such variables may be a unifying property of evolved or designed systems. The decrease of the concentration C with randomized network size S (Fig. 3) qualitatively agrees with exact results (*2*, *26*) on Erdos-Renyi random graphs (random graphs that preserve only the number of nodes and edges of the real network) in which $C \sim 1/S$. In general, the larger the network is, the more significant the motifs tend to become. This trend can also be seen in Table 1 by comparing networks of different sizes. The network motif detection algorithm appears to be effective even for rather small networks (on the order of 100 edges). This is because three- or four-node subgraphs occur in large numbers even in small networks. Furthermore, our approach is not sensitive to data errors; for example, the sets of significant network motifs do not change in any of the networks upon addition, removal, or rearrangement of 20% of the edges at random.

In information-processing networks, the motifs may have specific functions as elementary computational circuits (*11*). More generally, they may be interpreted as structures that arise because of the special constraints under which the network has evolved (*27*). It is of value to detect and understand network motifs in order to gain insight into their dynamical behavior and to define classes of networks and network homologies. Our approach can be readily generalized to any type of network, including those with multiple "colors" of edges or nodes. It would be fascinating to see what types of motifs occur in other networks and to understand the processes that yield given motifs during network evolution.

References and Notes

1. S. H. Strogatz, *Nature* **410**, 268 (2001).
2. B. Bollobas, *Random Graphs* (Academic, London, 1985).
3. D. Watts, S. Strogatz, *Nature* **393**, 440 (1998).
4. A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999).
5. M. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 404 (2001).
6. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A. L. Barabasi, *Nature* **407**, 651 (2000).
7. R. F. Cancho, C. Janssen, R. V. Sole, *Phys. Rev. E* **64**, 046119 (2001).
8. R. F. Cancho, R. V. Sole, *Proc. R. Soc. London Ser. B* **268**, 2261 (2001).
9. L. Amaral, A. Scala, M. Barthelemy, H. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149 (2000).
10. B. Huberman, L. Adamic, *Nature* **401**, 131 (1999).
11. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, *Nature Genet.* **31**, 64 (2002).
12. P. Holland, S. Leinhardt, in *Sociological Methodology*, D. Heise, Ed. (Jossey-Bass, San Francisco, 1975), pp. 1–45.
13. S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge Univ. Press, New York, 1994).
14. N. Guelzim, S. Bottani, P. Bourgine, F. Kepes, *Nature Genet.* **31**, 60 (2002).
15. M. Newman, S. Strogatz, D. Watts, *Phys. Rev. E* **64**, 6118 (2001).
16. S. Maslov, K. Sneppen, *Science* **296**, 910 (2002).
17. The randomized networks used for detecting three-node motifs preserve the numbers of incoming, outgoing, and double edges with both incoming and outgoing arrows for each node. The randomized networks used for detecting four-node motifs preserve the above characteristics as well as the numbers of all 13 three-node subgraphs as in the real network. Algorithms for constructing these randomized network ensembles are described (*18*). Additional information is available at www.weizmann.ac.il/mcb/UriAlon.
18. Methods are available as supporting material on *Science* Online.
19. D. Thieffry, A. M. Huerta, E. Perez-Rueda, J. Collado-Vides, *Bioessays* **20**, 433 (1998).
20. M. C. Costanzo *et al.*, *Nucleic Acids Res.* **29**, 75 (2001).
21. J. Cohen, F. Briand, C. Newman, *Community Food Webs: Data and Theory* (Springer, Berlin, 1990).
22. R. Williams, N. Martinez, *Nature* **404**, 180 (2000).
23. S. Pimm, J. Lawton, J. Cohen, *Nature* **350**, 669 (1991).
24. J. White, E. Southgate, J. Thomson, S. Brenner, *Philos. Trans. R. Soc. London Ser. B* **314**, 1 (1986).
25. F. Brglez, D. Bryan, K. Kozminski, *Proc. IEEE Int. Symp. Circuits Syst.*, 1929 (1989).
26. In Erdos-Renyi randomized networks with a fixed connectivity (*2*), the concentration of a subgraph with n nodes and k edges scales with network size as $C \sim S^{n-k-1}$ (thus, $C \sim 1/S$ for the feedforward loop of Fig. 3 where $n = k = 3$). The sole exception in Table 1 in which C should not vanish at large S is the three-chain pattern in food webs where $n = 3$ and $k = 2$.
27. D. Callaway, J. Hopcroft, J. Kleinberg, M. Newman, S. Strogatz, *Phys. Rev. E* **64**, 041902 (2001).

28. We thank S. Maslov and K. Sneppen for valuable discussions. We thank J. Collado-Vides, N. Martinez, R. Govindan, R. Durbin, L. Amaral, R. Cancho, S. Maslov, and K. Sneppen for kindly providing data, as well as D. Alon, E. Domany, M. Elowitz, I. Kanter, O. Hobart, M. Naor, D. Mukamel, A. Murray, S. Quake, R. Raz, M. Reigl, M. Surette, K. Sneppen, P. Sternberg, E. Winfree, and all members of our lab for comments. We thank Caltech and the Aspen Center for Physics for their hospitality during part of this work. We acknowledge support from the Israel Science Foundation, the Human Frontier Science Program, and the Minerva Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/298/5594/824/DC1
Methods
Table S1

1 May 2002; accepted 10 September 2002

Progression of Vertebrate Limb Development Through SHH-Mediated Counteraction of GLI3

Pascal te Welscher,[1] Aimée Zuniga,[1] Sanne Kuijper,[2] Thijs Drenth,[1] Hans J. Goedemans,[1] Frits Meijlink,[2] Rolf Zeller[1]*

Distal limb development and specification of digit identities in tetrapods are under the control of a mesenchymal organizer called the polarizing region. *Sonic Hedgehog* (SHH) is the morphogenetic signal produced by the polarizing region in the posterior limb bud. Ectopic anterior SHH signaling induces digit duplications and has been suspected as a major cause underlying congenital malformations that result in digit polydactyly. Here, we report that the polydactyly of *Gli3*-deficient mice arises independently of SHH signaling. Disruption of one or both *Gli3* alleles in mouse embryos lacking *Shh* progressively restores limb distal development and digit formation. Our genetic analysis indicates that SHH signaling counteracts GLI3-mediated repression of key regulator genes, cell survival, and distal progression of limb bud development.

The *Hedgehog* (*Hh*) signaling pathway controls many key developmental processes during animal embryogenesis (*1*). In *Drosophila* embryos, all known functions of Hh signaling are mediated by the transcriptional effector Cubitus interruptus (Ci) (*2*). Several homologs of *Hh* and *Ci* have been identified in higher vertebrates. In particular, *Sonic Hedgehog* (SHH) and the *Ci* homolog GLI3 are required for vertebrate limb development (*3*–*6*). GLI3 acts first during the initiation of limb bud development and before the activation of SHH signaling in posterior restriction of the basic helix-loop-helix transcription factor dHAND. dHAND in turn prevents *Gli3* expression from spreading posteriorly (Fig. 1A, panel 1) (*7*). In addition, GLI3 restricts the SHH-independent early expression of *5′HoxD* genes and *Gremlin* to the posterior mesenchyme (*8*). Subsequently, dHAND functions in the activation of *Shh* expression (*9*). Limb bud morphogenesis is then controlled by reciprocal interactions of two signaling centers (Fig. 1A, panel 2): the polarizing region, an instructive organizer located in the posterior limb bud mesenchyme, and the apical ectodermal ridge (AER). SHH signaling by the polarizing region in combination with bone morphogenetic proteins

[1]Department of Developmental Biology, Faculty of Biology, Utrecht University, Padualaan 8, NL-3584 CH Utrecht, Netherlands. [2]Hubrecht Laboratorium, Uppsalalaan 8, NL-3584 CT Utrecht, Netherlands.

*To whom correspondence should be addressed. E-mail: R.Zeller@bio.uu.nl