characterize many carnivorous non-mammalian synapsids[24]. The molariform teeth at the back of the dentition of *Repenomamus* are small with blunt crowns; they probably played a minor role in food processing. Although mammals are considered definitive chewers within amniotes[25], the dental morphology and large pieces of prey in the stomach of *Repenomamus* suggest that chewing as a derived feature in mammals was probably not present in *Repenomamus*.

It is not easy to assess whether *Repenomamus* was a predator or scavenger. Scavengers are relatively rare among mammals—among extant carnivorous mammals, only two species of hyenas are habitual scavengers[12,26]. Compared to their hunting cousins, these hyenas have smaller second upper incisors and less jaw muscle leverage, which probably reflect their inability to capture and handle live prey. In contrast, the enlarged incisors and strong jaw muscles of *Repenomamus* are well shaped for catching prey, favouring it as a predator rather than a scavenger.

For fossil mammals, body size is one of the most important factors influencing life history strategy[27]. Early mammals or their close relatives, such as morganocodontids and kuehneotheriids in the Late Triassic to Early Jurassic periods, were small and considered to be nocturnal insectivores[2,3]; the same is true of most later Mesozoic mammals[28] (Fig. 4). The reason for the very small size of Mesozoic mammals is uncertain[5], but it has often been hypothesized that well-established larger (and presumably diurnal) reptilian carnivores and herbivores, particularly dinosaurs, prevented mammals from invading those niches[29]. *Repenomamus* extend significantly the upper limit of body size of Mesozoic mammals (Fig. 4) and are actually larger than several small dinosaurs, particularly dromaeosaurid dinosaurs, from the same fauna[11]. Larger animals can live longer and move faster, but they also need a larger food supply and broader home range[30]. Judging from their body size, *R. giganticus* could feed on larger prey and forage a wider area for food. These large Mesozoic mammals were probably carnivores that competed with dinosaurs for food and territory. □

Received 29 May; accepted 8 October 2004; doi:10.1038/nature03102.

1. Bakker, R. T. Dinosaur physiology and the origin of mammals. *Evolution* **25,** 636–658 (1971).
2. Hopson, J. A. Endothermy, small size and the origin of mammalian reproduction. *Am. Nat.* **107,** 446–452 (1973).
3. Jerison, H. J. *Evolution of the Brain and Intelligence* (Academic, New York, 1973).
4. Crompton, A. W., Taylor, C. R. & Jagger, J. A. Evolution of homeothermy in mammals. *Nature* **272,** 333–336 (1978).
5. Lillegraven, J. A. in *Mesozoic Mammals: The First Two-thirds of Mammalian History* (eds Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A.) 1–6 (Univ. California Press, Berkeley, 1979).
6. Li, J.-L., Wang, Y., Wang, Y.-Q. & Li, C.-K. A new family of primitive mammal from the Mesozoic of western Liaoning, China [in Chinese]. *Chin. Sci. Bull.* **45,** 2545–2549 (2000).
7. Wang, Y.-Q., Hu, Y.-M., Meng, J. & Li, C.-K. An ossified Meckel's cartilage in two Cretaceous mammals and origin of the mammalian middle ear. *Science* **294,** 357–361 (2001).
8. Zhou, Z.-H., Barrett, P. M. & Hilton, J. An exceptionally preserved Lower Cretaceous ecosystem. *Nature* **421,** 807–814 (2003).
9. Li, C.-K., Wang, Y.-Q., Hu, Y.-M. & Meng, J. A new species of *Gobiconodon* from the Jehol Biota and its implication to the age of the fauna. *Chin. Sci. Bull.* **48,** 177–182 (2003).
10. Wang, S.-S., Wang, Y.-Q., Hu, H.-G. & Li, H.-M. The existing time of Sihetun vertebrate in western Liaoning, China—Evidence from U-Pb dating of zircon [in Chinese with English abstract]. *Chin. Sci. Bull.* **46,** 779–782 (2001).
11. Xu, X. & Wang, X.-L. A new dromaeosaur (Dinosauria: Theropoda) from the Early Cretaceous Yixian Formation of western Liaoning. *Vert. PalAsiat.* **42,** 111–119 (2004).
12. Nowak, R. M. *Walker's Mammals of the World* 6th edn (Johns Hopkins Univ. Press, Baltimore, 1999).
13. Silva, M. & Downing, J. A. *CRC Handbook of Mammalian Body Mass* (CRC Press, Boca Raton, 1995).
14. Jenkins, F. A. Jr Limb posture and locomotion in the Virginia opossum (*Didelphis marsupialis*) and in other non-cursorial mammals. *J. Zool.* **165,** 303–315 (1971).
15. Fischer, M. S., Schilling, N., Schmidt, M., Dieter Haarhaus, D. & Witte, H. Basic limb kinematics of small therian mammals. *J. Exp. Biol.* **205,** 1315–1338 (2002).
16. Wilson, R. W. Late Cretaceous (Fox Hills) multituberculates from the Red Owl Local Fauna of western South Dakota. *Dakoterra* **3,** 118–132 (1987).
17. Clemens, W. A., Wilson, G. P. & Molnar, R. E. An enigmatic (synapsid?) tooth from the Early Cretaceous of New South Wales, Australia. *J. Vert. Paleontol.* **23,** 232–237 (2003).
18. Jenkins, F. A. Jr & Schaff, C. R. The Early Cretaceous mammal *Gobiconodon* (Mammalia, Triconodonta) from the Cloverly Formation in Montana. *J. Vert. Paleontol.* **8,** 1–24 (1988).
19. Rougier, G. W. Vincelestes neuquenianus *Bonaparte (Mammalia, Theria), un Primitivo Mammifero del Cretaccico Inferior de la Cuenca Neuqina* PhD Thesis, Univ. Nacional de Buenos Aires, Buenos Aires (1993).
20. Alexander, R. McN., Jayes, A. S., Maloiy, G. M. O. & Wathuta, E. M. Allometry of limb bones of mammals from shrews (*Sorex*) to elephant (*Loxodonta*). *J. Zool.* **189,** 305–314 (1979).
21. Van Valkenburgh, B. in *Body Size in Mammalian Paleobiology: Estimation and Biological Implication* (eds Damuth, J. & MacFadden, B. J.) 181–205 (Cambridge Univ. Press, Cambridge, 1990).
22. Coombs, W. P. Jr Juvenile specimens of the ornithischian dinosaur *Psittacosaurus*. *Palaeontology* **25,** 89–107 (1982).
23. Carbone, C., Mace, G. M., Roberts, S. C. & Macdonald, D. W. Energetic constraints on the diet of terrestrial carnivores. *Nature* **402,** 286–288 (1999).
24. Van Valkenburgh, B. & Jenkins, I. Evolutionary patterns in the history of Permo-Triassic and Cenozoic synapsid predators. *Paleontol. Soc. Pap.* **8,** 267–288 (2002).
25. Reilly, S. M., McBrayer, L. D. & White, T. D. Prey processing in amniotes: biomechanical and behavioral patterns of food reduction. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **128,** 397–415 (2001).
26. Van Valkenburgh, B., Sacco, T. & Wang, X.-M. Pack hunting in Miocene Borophagine dogs: Evidence from craniodental morphology and body size. *Bull. Am. Mus. Nat. Hist.* **279,** 147–162 (2004).
27. Damuth, J. & MacFadden, B. J. in *Body Size in Mammalian Paleobiology: Estimation and Biological Implication* (eds Damuth, J. & MacFadden, B. J.) 1–10 (Cambridge Univ. Press, Cambridge, 1990).
28. Lillegraven, J. A., Kielan-Jaworowska, Z. & Clemens, W. A. (eds) *Mesozoic Mammals: The First Two-thirds of Mammalian History* (Univ. California Press, Berkeley, 1979).
29. Crompton, A. W. in *Comparative Physiology: Primitive Mammals* (eds Schmidt-Nielsen, K., Bolis, L. & Taylor, C. R.) 1–12 (Cambridge Univ. Press, Cambridge, 1980).
30. Eisenberg, J. F. in *Body Size in Mammalian Paleobiology: Estimation and Biological Implication* (eds Damuth, J. & MacFadden, B. J.) 25–38 (Cambridge Univ. Press, Cambridge, 1990).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank M.-M. Chang, Z.-H. Zhou, X.-L. Wang, X. Xu, F.-C. Zhang, Y. Wang, F. Jin and J.-Y. Zhang for help coordinating the research and fieldwork; X. Xu, X.-L. Wang, F.-C. Zhang, Z.-H. Zhou, and M. Norell for discussions on the research subject, and S.-H. Xie, S.-J. Li and A. Davidson for specimen preparation. This work was supported by funding from the Chinese Ministry of Science and Technology, the National Natural Science Foundation of China and the Chinese Academy of Sciences. Y.H. is also supported by a fellowship from the American Museum of Natural History, through the City University of New York.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y. H. (yhu@amnh.org).

The simplicity of metazoan cell lineages

Ricardo B. R. Azevedo[1], Rolf Lohaus[1,2], Volker Braun[2], Markus Gumbel[2], Muralikrishna Umamaheshwar[1], Paul-Michael Agapow[3], Wouter Houthoofd[4], Ute Platzer[2], Gaëtan Borgonie[4], Hans-Peter Meinzer[2] & Armand M. Leroi[3]

[1]*Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001, USA*
[2]*Division of Medical and Biological Informatics, German Cancer Research Center, D-69120, Heidelberg, Germany*
[3]*Department of Biology, Imperial College, Silwood Park, Ascot SL5 7PY, UK*
[4]*Department of Biology, Ghent University, B-9000 Ghent, Belgium*

Developmental processes are thought to be highly complex, but there have been few attempts to measure and compare such complexity across different groups of organisms[1–5]. Here we introduce a measure of biological complexity based on the similarity between developmental and computer programs[6–9]. We define the algorithmic complexity of a cell lineage as the length of the shortest description of the lineage based on its constituent sublineages[9–13]. We then use this measure to estimate the complexity of the embryonic lineages of four metazoan species from two different phyla. We find that these cell lineages are significantly simpler than would be expected by chance. Furthermore, evolutionary simulations show that the complexity of the embryonic lineages surveyed is near that of the simplest lineages evolvable, assuming strong developmental constraints on the spatial positions of cells and stabilizing selection on cell number. We propose that selection for decreased complexity has played a major role in moulding metazoan cell lineages.

Biological systems are obviously complex in both structure and

composition. However, understanding how such complexity develops and evolves remains one of the great questions of biology[1–6,8,14]. One obstacle is the lack of measures of the overall complexity of biological systems that are also applicable across a wide range of taxa[2,5]. In addition, most studies of biological complexity have concentrated on the number of different parts in a system (for example, genes, cell types, species), rather than on how they interact or develop[2,3,5–8]. In fact, despite recurring claims that organismal development is complex, attempts to quantify this complexity have been rare[1–6,14]. For example, Sulston and colleagues concluded that the most striking finding about the embryonic cell lineage of the nematode *Caenorhabditis elegans* was its complexity[13]. Although the authors did not explicitly define lineage complexity, they were probably referring to the many 'perverse' cell-fate assignments present in the lineage, whereby cells belonging to a given organ or functional class arise from lineally unrelated cells[13]. In other words, the *C. elegans* embryonic lineage does not appear to follow any particular rules[15]. However, the assumption that the complexity of a cell lineage can be inferred from that of the resulting pattern of cell fates is questionable because simple developmental processes can produce complex morphological patterns[6,16]. Indeed, casual examination of metazoan cell lineages suggests that they show a high degree of modularity in which particular sublineages are used again and again[3,5,11–13,17].

How complex are animal cell lineages? Is lineage complexity under selection? If so, what are the selective forces that shape it? To answer these questions we propose a measure of cell lineage complexity and apply it to the embryonic lineages of four metazoan species. The complexity of a cell lineage is a function of three properties: the number of cell divisions that it contains, the number and distribution of cell fates that it gives rise to, and its topology or pattern of cell divisions[1,9,14]. To capture these properties, we define the complexity of a lineage as the length of its shortest algorithmic description, by analogy with Kolmogorov complexity[7–10,18].

We begin by coding the lineage as a series of unique 'rules', each corresponding to a cell division (Fig. 1a). These rules take the form: X → {Y,Z} ('cell X divides into cells Y and Z'), where X is an undifferentiated cell, and Y and Z may be undifferentiated and/or terminal cells of a particular fate (for example, neuronal). This initial list of rules provides a complete description of the patterns of cell division and cell fate specification in the lineage, ignoring planes of cell division (Fig. 1a). We then compress the initial description by successively collapsing equivalent rules until we obtain a set of reduced rules encoding a complete, non-redundant description of the lineage equivalent to the initial one[9] (Fig. 1b and Supplementary Methods). Lineage complexity (C) is then defined as the number of reduced rules in the shortest description of the lineage expressed as a proportion of the total number of cell divisions (that is, the maximum possible number of reduced rules for a lineage of the same size).

The reduced rules predicted by our algorithm estimate the minimum number of intermediate cell states required to generate a given distribution of terminal cell fates. We propose that these intermediate cell states correspond to discrete, stable patterns of gene expression, much like those of terminal cells[17,19,20]. Nested sequences of reduced rules constitute sublineages[11–13]. We expect that reduced rules, like sublineages, can be used in different developmental contexts, and may be deployed in new contexts as a result of simple genetic changes; therefore, reduced rules are examples of 'genetic process' developmental modules[17,21].

We next estimate C for the embryonic lineages of four metazoan species[13,22,23]: the free-living nematodes *C. elegans* (671 terminal cells), *Pellioditis marina* (638) and *Halicephalobus gingivalis* (175), and the ascidian *Halocynthia roretzi* (110) (Supplementary Methods). These lineages show complexities of 35%, 38%, 33% and 32%, respectively (Figs 2 and 3a). We then compared each real lineage to lineages with the same cell number and distribution of terminal cell fates but generated by random bifurcation[9] (Figs 2 and 3b). We found that real lineages were 26–45% simpler than the corresponding random lineages ($P < 0.0001$ for all species; Fig. 2 and Supplementary Fig. 1a).

Animal cell lineages might have evolved towards simpler forms in order to minimize the duration of development or the amount of genetic information required to specify them[13,23]. If so, are metazoan embryonic lineages as simple as they might be? To answer this question we used evolutionary simulations to search for lineages that had the same terminal cell number and fate distribution as the actual lineages but were simpler. At each generation, a population of 100 variant lineages was produced from a parent lineage and the simplest daughter lineage was allowed to found the next generation (Fig. 4 and Supplementary Fig. 2a). We observed that we could evolve lineages that were 10–18% simpler than the ancestral, real lineages within 20,000–50,000 generations (Figs 3c and 4 and Supplementary Fig. 1b). Thus, although metazoan lineages are simple, they are not as simple as they might be given the requirements of producing a certain number of cells with a particular distribution of fates.

Why is this? One possibility is that the complexity of real cell

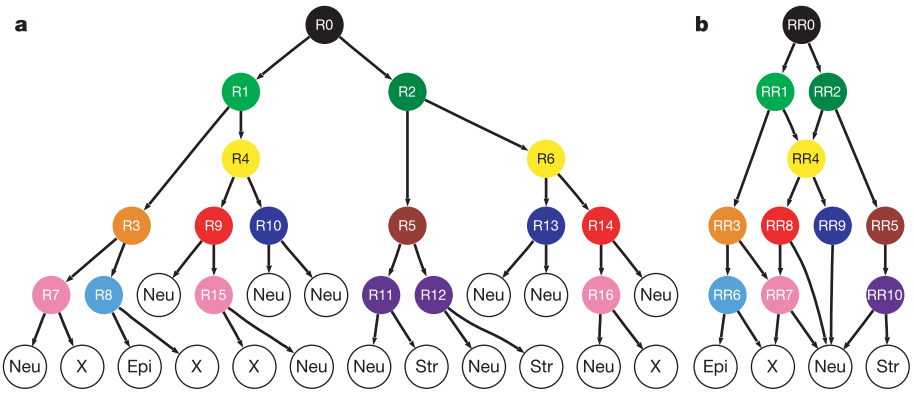

Figure 1 Example of the calculation of cell lineage complexity. **a**, The *C. elegans* ABarapp sublineage gives rise to 18 terminal cells of four different types (open circles): epidermal (Epi), neuron (Neu), structural (Str), and death (X). We begin by describing the cell lineage as a series of 17 rules, one for each cell division (solid circles): R0 → {R1,R2}, R1 → {R3,R4}, …, R16 → {Neu,X}. Solid circles of the same colour indicate equivalent rules, ignoring planes of cell division (for example, R7, R15 and R16). **b**, The minimum algorithmic description of the ABarapp sublineage consists of 11 reduced rules. Each reduced rule is represented by a solid circle labelled RR0–RR10, with a unique colour matching that of equivalent cell divisions (for example, RR7 → {Neu,X} corresponds to the initial rules R7, R15 and R16). The lineage complexity of ABarapp is calculated as the number of reduced rules divided by the total number of cell divisions: $C = 11/17 = 65\%$.

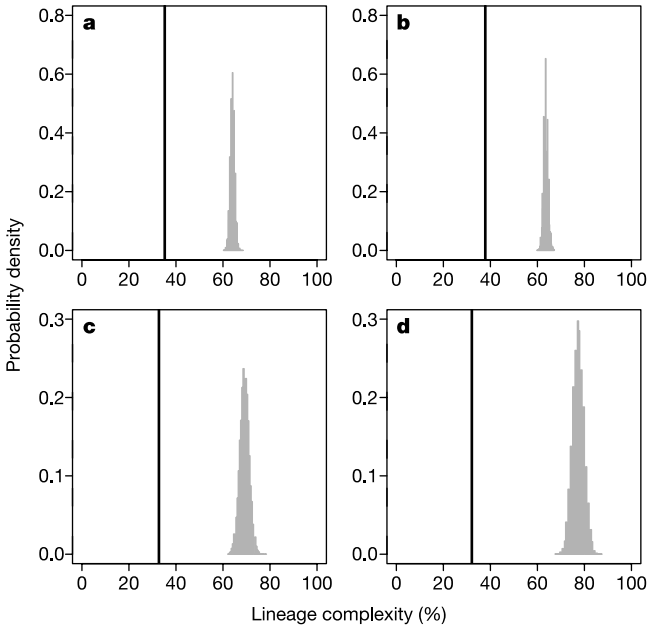

Figure 2 Metazoan embryonic cell lineages are simpler than expected by chance. **a**, *C. elegans* (complete embryonic lineage). **b**, *P. marina* (muscle-contraction stage lineage). **c**, *H. gingivalis* (muscle-contraction stage P_1 sublineage). **d**, *H. roretzi* (tissue-restricted stage lineage). Bold lines mark the lineage complexities (C) of the real lineages. Histograms show the distributions of C for 10,000 matching random lineages (a random bifurcation lineage with n cells was generated using ALES[9] by subjecting a founder cell to $n - 1$ rounds of cell division such that at each round all terminal cells have the same probability of dividing; cell states were randomly assigned to the terminal cells of the resulting lineage). Qualitatively similar results were obtained using other null models[9] (not shown).

lineages is a reflection of developmental constraints imposed by the spatial organization of cells in the embryo. Such constraints could occur if certain changes to the lineage topology or patterns of cell fate specification result in incorrect cell localization, and this in turn reduces the fitness of the organism. For example, in the four-cell stage *C. elegans* embryo the EMS blastomere must be exposed to a signal from its neighbouring sister cell P_2 in order to divide asymmetrically into MS and E, which give rise to mesoderm and gut, respectively[24]. However, if cell positions are altered such that the P_2 cell is in contact with the ABa and ABp blastomeres, but not with the EMS cell, then the gut does not form and the embryo dies[24]. In the species considered here, the spatial position of a cell in the embryo is largely determined by its position in the lineage diagram[13,15,22,23] (Supplementary Fig. 3 and Supplementary Movie). We simulated the effect of a spatial constraint on the evolution of lineage complexity by selecting the metazoan lineages for decreased complexity, while constraining the lineage positions of terminal cells (Fig. 4 and Supplementary Fig. 2b). We found that imposing a negligible selective constraint[25] on cell positions eliminated neutral drift[26], and that this reduced the selection response of C by 1.9–2.4%. In addition, as the strength of the constraint on cell positions increased, the magnitude of the selection response in cell lineage complexity decreased by a further 3.6–5.7% (Fig. 4 and Supplementary Fig. 1b). These results suggest that the metazoan lineages studied here are almost as simple as the simplest evolvable under strong constraints on the spatial positions of cells. Changes in patterns of cell migration might alleviate the effects of the spatial constraint. This might explain why the *H. gingivalis* lineage is 5.6% and 7.9% simpler than comparable *C. elegans* and *P. marina* muscle-contraction P_1 sublineages (Supplementary Methods), respectively, and shows greater levels of cell migration than either of these species[23,27].

The existence of spatial constraints is not, however, the only reason that cell lineages do not evolve towards even greater simplicity. The selection responses of populations of lineages selected for increased simplicity repeatedly formed plateaus (Fig. 4 and Supplementary Fig. 1b). In no case were the plateaus caused by convergence on the simplest possible cell lineages because it is easy to construct lineages with the same cellular composition as the real ones, but that are far simpler than the simplest lineages achieved in our simulations. For example, we have derived an artificial *C. elegans* lineage with $C = 4.6\%$ (Supplementary Fig. 4), compared with 35% for the real lineage, and 21–23% for the simplest evolved lineages (Fig. 4a). Prolonging our simulation runs should lead to a further reduction in the complexity of the artificial *C. elegans*

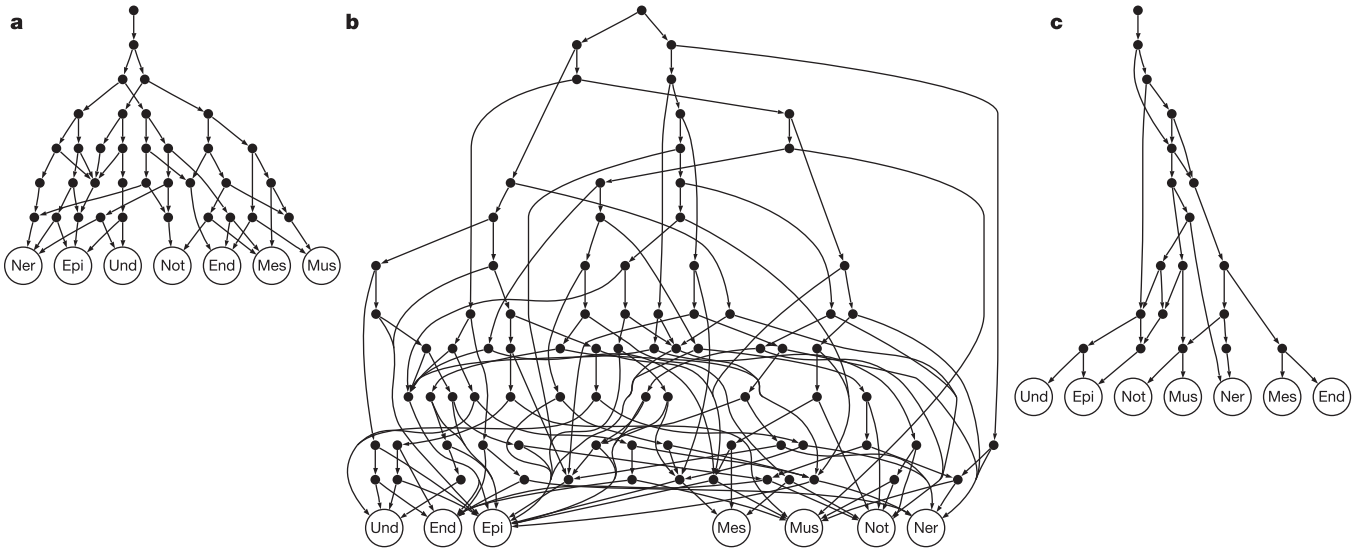

Figure 3 The simplicity of the ascidian cell lineage. Shortest algorithmic descriptions of three lineages capable of generating the cells in the *H. roretzi* tissue-restricted stage embryo. **a**, The real lineage has a complexity of $C = 32\%$. **b**, A random bifurcation lineage with over twice the complexity of the real one ($C = 76\%$; Fig. 2d). **c**, The simplest lineage evolved from the *H. roretzi* lineage by selection for low complexity is approximately half as complex as the real one ($C = 17\%$; Fig. 4d). Solid circles represent the reduced rules required to generate the different terminal cell states (open circles): endoderm (End), epidermis (Epi), mesenchyme (Mes), muscle (Mus), nervous system (Ner), notochord (Not) and undifferentiated (Und).

lineage, but it is highly unlikely ever to reach 4.6% because the evolvability at the end of the evolutionary simulations is extremely low (Fig. 4a). Before selection, the probability that a 'mutation' will simplify the *C. elegans* lineage is 0.76% (Supplementary Fig. 5), but it declines to 0.00012 ± 0.00015% after 50,000 generations of selection for low complexity without constraints on cell position (1,000,000 offspring; mean and 95% confidence intervals based on ten replicates). These results suggest that the simplest lineages are mutationally inaccessible in our simulations[28]. Furthermore, cell lineages evolved under the spatial constraint appear to be driven into regions of lineage space from which simpler lineages are even less accessible (Supplementary Fig. 6). Results from more elaborate models of lineage evolution (M.U., R.L. & R.B.R.A., unpublished results) suggest that these generative constraints[25] on the evolution of lineage complexity are caused by the restriction of the lineage 'search space' to cell lineages with the same size and cell fate distribution as the ancestral lineage (Supplementary Fig. 2). This simplification, although unrealistic[12,23,29], seems nevertheless to provide a reasonable approximation to evolutionary models with an unrestricted search space that incorporate strong stabilizing selection on terminal cell number and fate distribution.

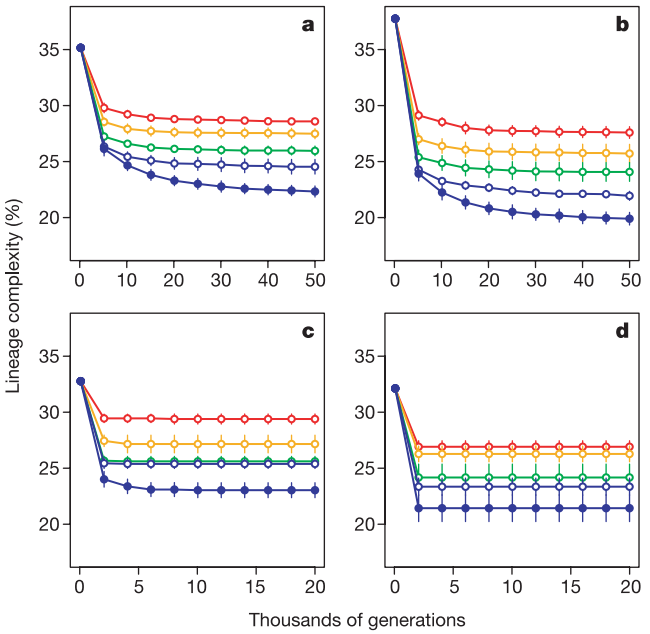

Figure 4 Metazoan cell lineages are not as simple as they could be. **a**–**d**, Responses to selection for decreased lineage complexity (C) of the lineages listed in Fig. 2. Each generation, 100 variant lineages were generated by allowing the exchange of a pair of randomly selected sublineages or terminal cells (Supplementary Fig. 2). The fitness of a lineage was defined as $W = 1/[C(D+1)^k]$, where C is the complexity of the current lineage, k is the strength of the selective constraint on the lineage positions of cells,

$$D = 2\left[\sum_{i=1}^{n}(L_i' - L_i)^2\right] / \left(\sum_{i=1}^{n} L_i^2\right)$$

is a measure of the deviation in lineage positions relative to the parent lineage, L_i and L_i' are the lineage positions of the ith cell in the parent and current lineages, respectively, and n is the total number of terminal cells. The offspring lineage with the highest value of W (or the parent lineage, if no offspring had a fitness equal to or higher than that of the parent) was selected to found the next generation. This procedure was iterated for 20,000 or 50,000 generations. Plots show the mean selection responses of C (and 95% confidence intervals) in ten replicate experiments, taken every 2,000 or 5,000 generations. Each lineage was subject to directional selection to reduce C, either without ($k = 0$, blue closed circles) or with ($k > 0$, open circles) a selective constraint on the lineage positions of cells. Spatial constraints of varying strengths were simulated: negligible ($k = 10^{-10}$, open blue), weak ($k = 1$, green), moderate ($k = 10$, orange) and strong constraints ($k = 100$, red). The ascidian lineage (**d**) was only evolved on one side (55 cells), so as not to break the bilateral symmetry[22]. The simulations were carried out using LES (Lineage Evolution System; Supplementary Methods).

It is widely believed that morphological complexity tends to increase in evolution[1–4,14]. For example, Valentine and co-workers[30] have estimated that the maximum in one correlate of cell lineage complexity (Supplementary Fig. 1)—the number of terminal cell types—has increased at an average rate of 0.3 per million years in metazoans. Our results, however, suggest that certain animals generate morphological complexity while actively maintaining simple, highly modular cell lineages. There may be several reasons for this. Simpler lineages might develop faster. For example, the *P. marina* lineage is 28% slower and 4.4% more complex than a comparable *C. elegans* muscle-contraction lineage[23] (Supplementary Methods). Indeed, developmental rate could be viewed as the biological analogue of another measure of algorithmic complexity—logical depth or execution time[18]. In addition, the quantity $1/C$ measures the average number of times a reduced rule is used during development, suggesting that the specification of simpler cell lineages might require less genetic information, and thus be more efficient[1,13].

Thus, although we do not yet fully understand the selective forces that influence the evolution of cell lineages, we provide here a method for estimating and comparing cell lineage complexity in different organisms. We furthermore demonstrate that some metazoan embryonic lineages are simpler than they appear. Finally, we suggest that these metazoan cell lineages could not be much simpler than they are, given the necessity of placing precise numbers of cells in particular positions in developing embryos. □

Received 6 August; accepted 9 November 2004; doi:10.1038/nature03178.

1. Bonner, J. T. *The Evolution of Complexity* (Princeton Univ. Press, Princeton, 1988).
2. McShea, D. W. Metazoan complexity and evolution: Is there a trend? *Evolution* **50,** 477–492 (1996).
3. Carroll, S. B. Chance and necessity: The evolution of morphological complexity and diversity. *Nature* **409,** 1102–1109 (2001).
4. Arthur, W. The emerging conceptual framework of evolutionary developmental biology. *Nature* **415,** 757–764 (2002).
5. Minelli, A. *The Development of Animal Form: Ontogeny, Morphology, and Evolution* (Cambridge Univ. Press, Cambridge, UK, 2003).
6. Apter, M. J. & Wolpert, L. Cybernetics and development. I. Information theory. *J. Theor. Biol.* **8,** 244–257 (1965).
7. Atlan, H. & Koppel, M. The cellular computer DNA: Program or data. *Bull. Math. Biol.* **52,** 335–348 (1990).
8. Szathmary, E., Jordan, F. & Pal, C. Molecular biology and evolution: Can genes explain biological complexity? *Science* **292,** 1315–1316 (2001).
9. Braun, V. *et al.* ALES: Cell lineage analysis and mapping of developmental events. *Bioinformatics* **19,** 851–858 (2003).
10. Papentin, F. On order and complexity. I. General considerations. *J. Theor. Biol.* **87,** 421–456 (1980).
11. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **56,** 110–156 (1977).
12. Sternberg, P. W. & Horvitz, H. R. Postembryonic nongonadal cell lineages of the nematode *Panagrellus redivivus*: Description and comparison with those of *Caenorhabditis elegans*. *Dev. Biol.* **93,** 181–205 (1982).
13. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100,** 64–119 (1983).
14. Brooks, D. R. & Wiley, E. O. *Evolution as Entropy* 2nd edn (Univ. Chicago Press, Chicago, 1988).
15. Schnabel, R., Hutter, H., Moerman, D. & Schnabel, H. Assessing normal embryogenesis in *Caenorhabditis elegans* using a 4D microscope: Variability of development and regional specification. *Dev. Biol.* **184,** 234–265 (1997).
16. Goodwin, B. C., Kauffman, S. & Murray, J. D. Is morphogenesis an intrinsically robust process? *J. Theor. Biol.* **163,** 135–144 (1993).
17. Raff, R. A. *The Shape of Life* (Univ. Chicago Press, Chicago, 1996).
18. Bennett, C. H. in *Complexity, Entropy and the Physics of Information* (ed. Zurek, W. H.) 137–148 (Addison-Wesley, Redwood City, 1990).
19. Kauffman, S. A. *The Origins of Order* (Oxford Univ. Press, Oxford, 1993).
20. Geard, N. & Wiles, J. A gene network model for developing cell lineages. *Artif. Life* (in the press).
21. Wagner, G. P. & Mezey, J. G. in *Modularity in Development and Evolution* (eds Schlosser, G. & Wagner, G. P.) 338–358 (Univ. Chicago Press, Chicago, 2004).
22. Nishida, H. Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Dev. Biol.* **121,** 526–541 (1987).
23. Houthoofd, W. *et al.* Embryonic cell lineage of the marine nematode *Pellioditis marina*. *Dev. Biol.* **258,** 57–69 (2003).
24. Goldstein, B. Induction of gut in *Caenorhabditis elegans* embryos. *Nature* **357,** 255–257 (1992).
25. Richardson, M. K. & Chipman, A. D. Developmental constraints in a comparative framework: a test case using variations in phalanx number during amniote evolution. *J. Exp. Zool. B (Mol. Dev. Evol.)* **296,** 8–22 (2003).
26. Fontana, W. & Schuster, P. Continuity in evolution: On the nature of transitions. *Science* **280,** 1451–1455 (1998).

27. Borgonie, G., Jacobsen, K. & Coomans, A. Embryonic lineage evolution in nematodes. *Nematology* **2,** 65–69 (2000).
28. Stadler, B. M., Stadler, P. F., Wagner, G. P. & Fontana, W. The topology of the possible: formal spaces underlying patterns of evolutionary change. *J. Theor. Biol.* **213,** 241–274 (2001).
29. Sommer, R. J., Carta, L. K. & Sternberg, P. W. The evolution of cell lineage in nematodes. *Development* (Suppl.), 85–95 (1994).
30. Valentine, J. W., Collins, A. G. & Meyer, C. P. Morphological complexity increase in metazoans. *Paleobiology* **20,** 131–142 (1994).

Supplementary Information accompanies the paper on **www.nature.com/nature**

Acknowledgements We thank S. Emmons, Y. Fofanov, D. Graur, D. Portman, T. Shin, S. Srinivasan and M. Travisano for discussions. Z. Altun and D. Hall gave advice on the classification of *C. elegans* cells. The Sun Microsystems Center of Excellence in the Geosciences at the University of Houston provided access to high-performance computing resources. The Foundation for Science and Technology (Portugal), European Molecular Biology Organization, Biotechnology and Biological Sciences Research Council (UK), and the University of Houston provided financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.B.R.A. (razevedo@uh.edu).

Unexpected complexity of the *Wnt* gene family in a sea anemone

Arne Kusserow[1], Kevin Pang[2], Carsten Sturm[1], Martina Hrouda[3], Jan Lentfer[1], Heiko A. Schmidt[4], Ulrich Technau[1]*, Arndt von Haeseler[4,5], Bert Hobmayer[3], Mark Q. Martindale[2] & Thomas W. Holstein[1,6]

[1]*Institute of Zoology, Darmstadt University of Technology, Schnittspahnstrasse 10, D-64287 Darmstadt, Germany*
[2]*Kewalo Marine Lab PBRC, University of Hawaii, 41 Ahui Street, Hawaii 96813 Honolulu, USA*
[3]*Institute of Zoology and Limnology, Center of Molecular Biosciences Innsbruck, University of Innsbruck, Technikerstrasse 25, A-6020 Innsbruck, Austria*
[4]*John von Neumann-Institut für Computing (NIC), FZ Jülich, D-52425 Jülich, Germany*
[5]*Institut für Bioinformatik, Heinrich-Heine-Universität Düsseldorf, Universitätsstrasse 1 / Geb. 25.13.02, D-40225 Düsseldorf, Germany*
[6]*Department of Molecular Evolution and Genomics, University of Heidelberg, Im Neuenheimer Feld 230, D-69120 Heidelberg, Germany*

* Present address: Sars Centre, Thormøhlensgt. 55, N-5008 Bergen, Norway

The *Wnt* gene family encodes secreted signalling molecules that control cell fate in animal development and human diseases[1]. Despite its significance, the evolution of this metazoan-specific protein family is unclear. In vertebrates, twelve Wnt subfamilies were defined, of which only six have counterparts in Ecdysozoa (for example, *Drosophila* and *Caenorhabditis*)[2]. Here, we report the isolation of twelve *Wnt* genes from the sea anemone *Nematostella vectensis*[3], a species representing the basal group[4] within cnidarians. Cnidarians are diploblastic animals and the sister-group to bilaterian metazoans[5]. Phylogenetic analyses of *N. vectensis Wnt* genes reveal a thus far unpredicted ancestral diversity within the Wnt family[2,6,7]. Cnidarians and bilaterians have at least eleven of the twelve known *Wnt* gene subfamilies in common; five subfamilies appear to be lost in the protostome lineage. Expression patterns of *Wnt* genes during *N. vectensis* embryogenesis indicate distinct roles of Wnts in gastrulation, resulting in serial overlapping expression domains along the primary axis of the planula larva. This unexpectedly complex inventory of Wnt family signalling factors evolved in early multicellular animals about 650 million years (Myr) ago, predating the Cambrian explosion by at least 100 Myr (refs 5, 8). It emphasizes the crucial function of *Wnt* genes in the diversification of eumetazoan body plans[9].

We isolated twelve *Wnt* genes from *N. vectensis*, yet only one orthologue (*Wnt3*) was identified from the freshwater polyp *Hydra magnipapillata*[6]. Alignments of these cnidarian sequences were made using representatives in known databases from all three major metazoan clades: that is, deuterostomes (including all human sequences), ecdysozoans, and lophotrochozoans (Supplementary Tables S1 and S2). Phylogenetic analyses were based on three different phylogenetic methods: that is, the maximum parsimony (MP) and maximum likelihood (ML, TREE-PUZZLE and IQPNNI) approaches (Supplementary Figs S1–S3) and bayesian phylogenetic inference (Fig. 1). All approaches generated twelve *Wnt* gene subfamilies identified as *WntA* and *Wnt1–11*. Cnidarians possess orthologues of eleven of the twelve *Wnt* subfamilies, *WntA*, *Wnt1–8,* and *Wnt10–11* (Table 1). Only *Wnt9* was not found in cnidarians. It remains unclear whether we failed to identify this gene in *N. vectensis* or whether *Wnt9* has been lost in cnidarian evolution. The sea anemone *NvWnt* subfamilies *NvWnt7* and *NvWnt8* exhibit two paralogous genes which share no orthology with the same *Wnt* subfamilies in mammalians (Fig. 1). Therefore, they represent cnidarian or anthozoan specific duplications.

Thus at least eleven of twelve *Wnt* gene subfamilies must have already been present before the divergence of bilaterians and cnidarians. They constituted the *Wnt* repertoire of the last common ancestor of bilaterians and cnidarians, the *Ur-Eumetazoa* (see Table 1). Our comparison also indicates the existence of only seven *Wnt* gene subfamilies (*WntA, -1*, *-5–7* and *-9–10*) in insects and only five *Wnt* genes in *Caenorhabditis elegans* (Table 1). Full genome sequences are available from these three species (*C. elegans*, *Drosophila melanogaster* and *Anopheles gambiae)* so it is highly unlikely that we missed *Wnt* orthologues from ecdysozoans in our analysis. In lophotrochozoans, the second major protostomian clade, *Wnt* gene subfamilies *Wnt3*, *-6*, *-8*, and *-11* have not been reported yet[2,10]. Thus it remains to be clarified which *Wnt* gene subfamilies existed at the protostome–deuterostome divergence. In turn, our data reveal that only one *Wnt* gene subfamily (*WntA*) was lost during the evolution of deuterostomes (Table 1).

Although the *Wnt* gene subfamilies are statistically well supported, there is not enough phylogenetic resolution to distinguish reliable relationships among all *Wnt* subfamilies. Nonetheless, there is a clustering of the *Wnt1*, *-6*, *-10*, *-9* and *-3* subfamilies in the phylogenetic data (Fig. 1), which is also supported by human and fly genome data[11]. In the *D. melanogaster* genome, *DmWnt1* (*Wg*), *DmWnt6* and *DmWnt10* are positioned immediately adjacent to each other on the second chromosome and transcribed in the same orientation. This order is conserved in the mammalian genome, where also *Wnt3A* and *-9A* and *Wnt3* and *-9B* are closely linked[11]. Thus, *Wnt* genes *Wnt1*, *-6*, *-10*, *-9* and *-3* might represent an ancestral cluster of *Wnt* genes that originated in the evolution of the common ancestor of cnidarians and bilaterians. No *Wnt* genes have been described so far from unicellular eukaryotes, from cellular slime moulds (*Dictyostelium discoideum*) or from choanoflagellates[12], unicellular and colonial Protozoa that are closely related to Metazoa. At present no data are available from sponges, which probably diverged before the origin of the eumetazoan ancestor, but we presume that the appearance of *Wnt* genes itself was linked to the origin and evolution of multi-cellular animals from single-cell (protozoan) ancestors.

To analyse the possible function of different *Wnt* genes in *N. vectensis* embryogenesis, *Wnt* gene expression for ten genes was assayed by *in situ* hybridization from the early blastula through to newly settled polyps forming their first tentacles (Fig. 2). Each *Wnt* gene displayed a distinct expression pattern during early embryogenesis. Most of the *N. vectensis Wnt* genes are expressed along the primary body axis, where they are restricted to the blastopore during gastrulation and to the oral region of planula or polyps