

# Gene Regulatory Networks and the Evolution of Animal Body Plans

Eric H. Davidson<sup>1\*</sup> and Douglas H. Erwin<sup>2</sup>

Development of the animal body plan is controlled by large gene regulatory networks (GRNs), and hence evolution of body plans must depend upon change in the architecture of developmental GRNs. However, these networks are composed of diverse components that evolve at different rates and in different ways. Because of the hierarchical organization of developmental GRNs, some kinds of change affect terminal properties of the body plan such as occur in speciation, whereas others affect major aspects of body plan morphology. A notable feature of the paleontological record of animal evolution is the establishment by the Early Cambrian of virtually all phylum-level body plans. We identify a class of GRN component, the “kernels” of the network, which, because of their developmental role and their particular internal structure, are most impervious to change. Conservation of phyletic body plans may have been due to the retention since pre-Cambrian time of GRN kernels, which underlie development of major body parts.

Large gene regulatory networks (GRNs) that determine the course of animal development are now being decoded experimentally [e.g., (1–8)]. These networks consist largely of the functional linkages among regulatory genes that produce transcription factors and their target cis-regulatory modules in other regulatory genes, together with genes that express spatially important signaling components. They have a modular structure, consisting of assemblies of multigenic subcircuits of various forms. Each such subcircuit performs a distinct regulatory function in the process of development (1, 9). GRN structure is inherently hierarchical, because each phase of development has beginnings, middle stages, and progressively more fine-scale terminal processes, so that network linkages operating earlier have more pleiotropic effects than those controlling terminal events. The earlier stages of formation of every body part involve specification of the domain of the developing organism that will become that part, followed by pattern formation, which determines its morphological structure. Only at the end of this process are deployed the differentiation gene batteries that encode the detailed functional properties of the body part (10).

The structure/function properties of developmental GRNs provide an approach to an old and general problem in animal evolution. What mechanisms account for the fact that has there

has been so little change in phylum- and superphylum-level body plans since the Early Cambrian [e.g., the Chengjiang fauna (11–13)] (Fig. 1), though on the other hand, great changes have subsequently occurred within phyla and classes (e.g., the advent of tetrapod vertebrates, insects, dinosaurs, modern forms of echinoids, and cephalopods)? Furthermore, continuous modification characterizes the process of speciation. Classic evolutionary theory, based on selection of small incremental changes, has sought explanations by extrapolation from observed patterns of adaptation. Macroevolutionary theories have largely invoked multi-level selection, among species and among clades. But neither class of explanation provides an explanation of evolution in terms of mechanistic changes in the genetic regulatory program for development of the body plan, where it must lie.

## Functional Properties of Diverse GRN Components

Change in the structure of the diverse kinds of subcircuits of which GRNs are constructed will have different consequences for the outcome of the developmental process, and therefore for evolution as well. Here we consider the following classes of GRN component: (i) evolutionarily inflexible subcircuits that perform essential upstream functions in building given body parts, which we term the “kernels” of the GRN; (ii) certain small subcircuits, the “plug-ins” of the GRN, that have been repeatedly coopted to diverse developmental purposes; (iii) switches that allow or disallow developmental subcircuits to function in a given context and so act as input/output (I/O) devices within the GRN; and (iv) differentiation gene batteries. These parts are illustrated in a real developmental

GRN, the sea urchin endomesodermal GRN, in fig. S1.

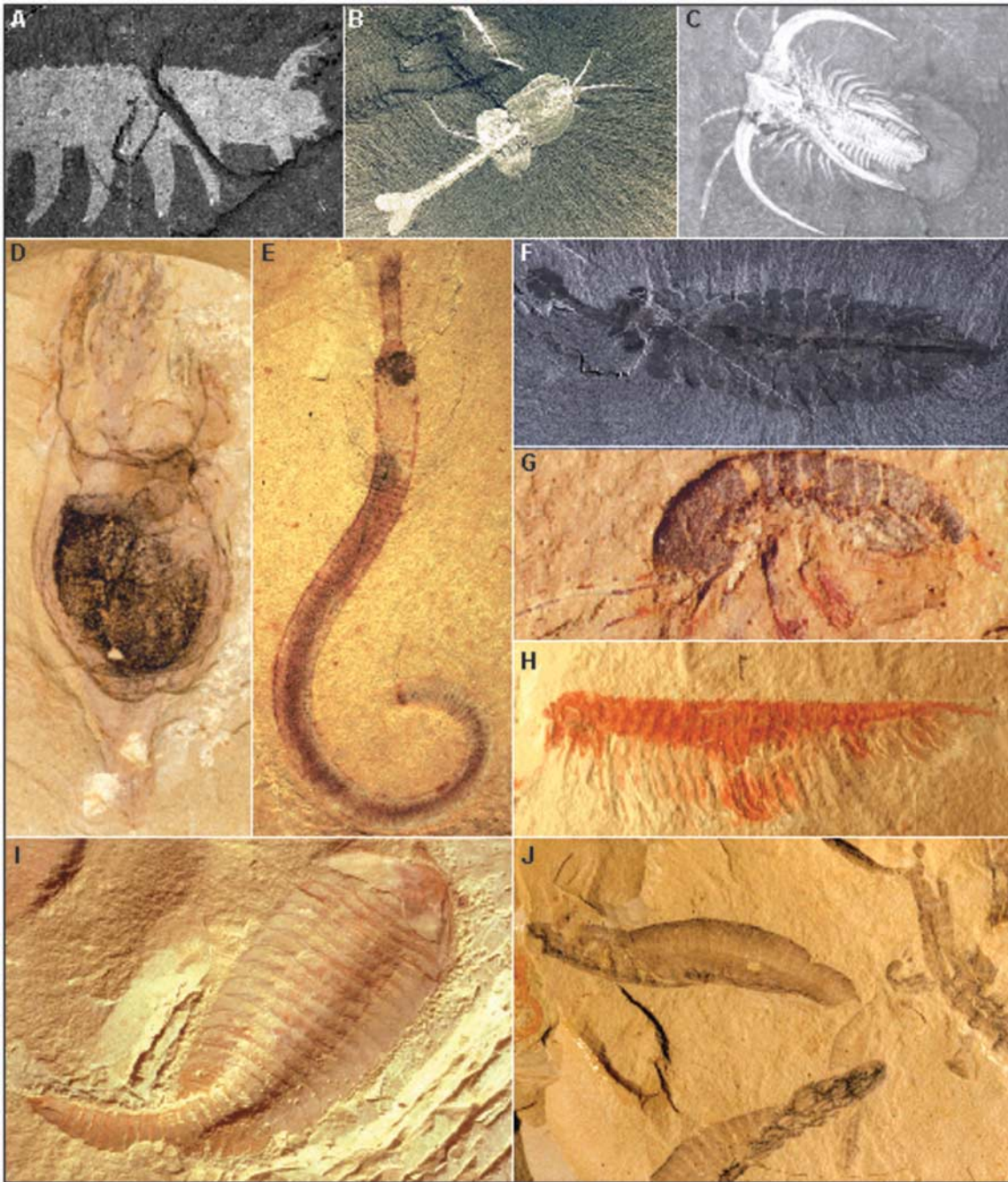
Five properties can be used to define GRN kernels. First, these are network subcircuits that consist of regulatory genes (i.e., genes encoding transcription factors). Second, they execute the developmental patterning functions required to specify the spatial domain of an embryo in which a given body part will form. Third, kernels are dedicated to given developmental functions and are not used elsewhere in development of the organism (though individual genes of the kernel are likely used in many different contexts). Fourth, they have a particular form of structure in that the products of multiple regulatory genes of the kernel are required for function of each of the participating cis-regulatory modules of the kernel (“recursive wiring”). Hence, the fifth property of the kernel is that interference with expression of any one kernel gene will destroy kernel function altogether and is likely to produce the catastrophic phenotype of lack of the body part. The result is extraordinary conservation of kernel architecture.

Two examples of kernels illustrate many of these points (Fig. 2). Both display detailed conservation of complex subcircuit architectural structure across immense periods of evolutionary time, and both are surrounded by other network linkages that are not conserved. The first (Fig. 2A) includes a gene regulatory feedback loop required for endoderm specification in echinoderms that has existed at least since divergence at the end of the Cambrian half a billion years ago (14) and could, of course, be much older. The second (Fig. 2B) is a heart-field specification kernel (15) that must be even more ancient, as it is used in both *Drosophila* and vertebrate development. These subcircuits operate to specify the cellular populations where, respectively, the gut and the heart will form and to set up the regulatory states on which subsequent developmental processes will depend.

In the echinoderm endoderm network, five of the six genes in the kernel (all except *delta*) encode DNA-recognizing transcription factors; that is, they are regulatory genes, and this is true of all the genes in the conserved circuitry in the heart network. In both kernels, the linkages are highly recursive. For example, in the endoderm kernel, the cis-regulatory module of the *otx* gene receives input from three of the five genes; the *foxa* gene, from three of the five; and the *gatae*, *foxa*, and *bra* genes from two of the same five genes; similarly, in the vertebrate heart network, the *nkx2.5*, *tbx*, *mef2c*, and *gata4* genes all receive inputs from multiple other genes of the kernel, as do the *tin*, *doc*, *mid*, *pnr*, and *mef2c* genes of the *Drosophila* network. It is also the case that loss of expression of any of these genes in either kernel has a catastrophic effect on development of the respective body parts (1, 14–16). There are a number of additional

<sup>1</sup>Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA. <sup>2</sup>Department of Paleobiology, MRC-121, National Museum of Natural History, Post Office Box 37012, Washington, DC 20013–7012, and Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA.

\*To whom correspondence should be addressed. E-mail: davidson@caltech.edu



**Fig. 1.** Examples of Cambrian body plans from the Early Cambrian (~510 million years ago) Chengjiang Fauna of Yunnan Province, China (**D to I**) and the Middle Cambrian Burgess Shale Fauna of British Columbia, Canada (**A to C, J**). These fossils are the remains of animals all of which have body plans that can immediately be related to those of modern phyla, as indicated. For instance, the bilateral, anterior-posterior organization and position of the appendages in the arthropod examples resemble those of the modern counterparts; in addition, the chordate has a segmented dorsal muscular column and a notochord, as do modern chordates. (A) Onychophoran: *Aysheia pedunculata*; (B) arthropod: *Waptia fieldensis*; (C) arthropod: *Marrella splendens*; (D) possible ascidian: *Phlogites*; (E) priapulid: *Maotianshania cylindrica*; (F) pan-arthropod: *Opabinia regalis*; (G) arthropod: *Leancoilia illecebrosa*; (H) arthropod: *Jianfengia multisegmentalis*; (I) arthropod: *Fuxianjia protensa*; (J) chordate: *Haikouella lanceolata*; [(A to C)] and (F) are from D. H. Erwin, Smithsonian Institution; (D), (E), and [(G) to (J)] are courtesy of J.-Y. Chen, Nanjing Institute of Geology and Palaeontology, China (13).

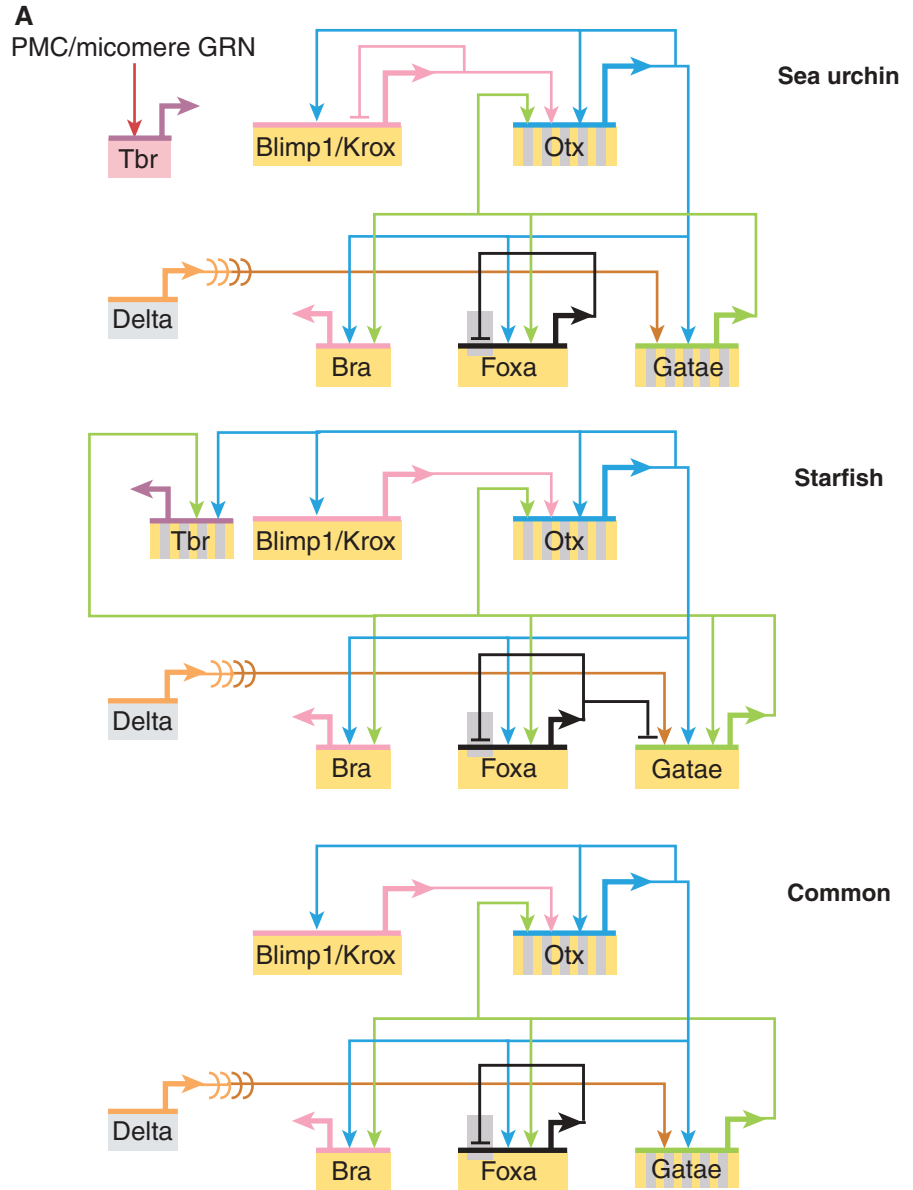
examples for which there is persuasive evidence for the existence of GRN kernels awaiting discovery of the direct genomic regulatory code. Prospective examples include kernels common to all members of a given phylum or superphylum required for the following: anterior to posterior (17–19) and midline to lateral (20, 21) specification of the nervous system (in deuterostomes and possibly across Bilateria); eye-field specification [in arthropods (22, 23)]; gut regionalization [in chordates (24, 25)]; development of immune systems [across Bilateria (26, 27)]; and regionalization of the hindbrain and specification of neural crest [in chordates (28, 29)].

“Plug-ins” also consist of structurally conserved GRN subcircuits, but as they are used for many diverse developmental functions within and among species, these network subcircuits are not dedicated to formation of given body parts. Instead, they are inserted in many different networks where they provide inputs into a great variety of regulatory apparatus. The best examples are signal transduction systems, of which a small set, each affecting a confined repertoire of transcription factors, are used repeatedly, often acting as dominant spatial repressors in the absence of ligand and as facilitators of spatially confined expression in its presence (30). In Bilateria, Wnt (31), transforming growth

factor- $\beta$  (TGF- $\beta$ ) (32), fibroblast growth factor (33), Hedgehog (34), Notch (35), and epidermal growth factor (36) signaling systems are used for myriad purposes during development. Their deployment is very flexible, and even in homologous processes in related animals these plug-ins may be used differently (37). Consider, for example, the several dozen different TGF- $\beta$  genes in amniote vertebrates, expressed differentially in the (species-specific) terminal phases of development (32, 38). It follows that their connections into the network are evolutionarily very labile.

Differentiation gene batteries are defined as groups of protein-coding genes under common regulatory control, the products of which exe-

**Fig. 2.** Examples of putative GRN kernels. Networks were constructed and portrayed using BioTapestry software (55). **(A)** Endomesoderm specification kernel, common to sea urchin and starfish, the last common ancestor of which lived about half a billion years ago. The relevant area of the sea urchin network is shown at the top [(1, 9, 16); for currently updated version, details, and supporting data, see (56)]; the corresponding starfish network (14) is shown in the middle; and the network architecture, which has been exactly conserved since divergence—i.e., the kernel—is shown at the bottom. Horizontal lines denote cis-regulatory modules responsible for the pregastrular phase of expression considered, in endoderm (yellow), mesoderm (gray), or both endoderm and mesoderm (striped gray and yellow). The inputs into the cis-regulatory modules are denoted by vertical arrows and bars. The gray box surrounding the *foxa* input indicates that this repression occurs exclusively in mesoderm. **(B)** Possible heart specification kernels; assembled from many literature sources (15). Dashed lines show possible interactions. Some aspects of the GRN that may underlie heart specification in *Drosophila* are shown at the top; the approximately corresponding vertebrate relationships are shown in the middle; and shared linkages are shown at the bottom. Absence of a linkage simply means that this linkage is not known to exist, not that it is known not to exist. Many regulatory genes participate in vertebrate heart formation for which orthologous *Drosophila* functions have not been discovered, and the hearts themselves are of very different structure. However, as pointed out by many authors [see (7, 8, 57) for reviews of earlier references], a core set of regulatory genes are used in common and are now known to be linked in a similar way in a conserved subcircuit of the gene network architecture, as shown. The gray boxes represent in each case different ways that the same two nodes of the network are linked in *Drosophila* and vertebrates.

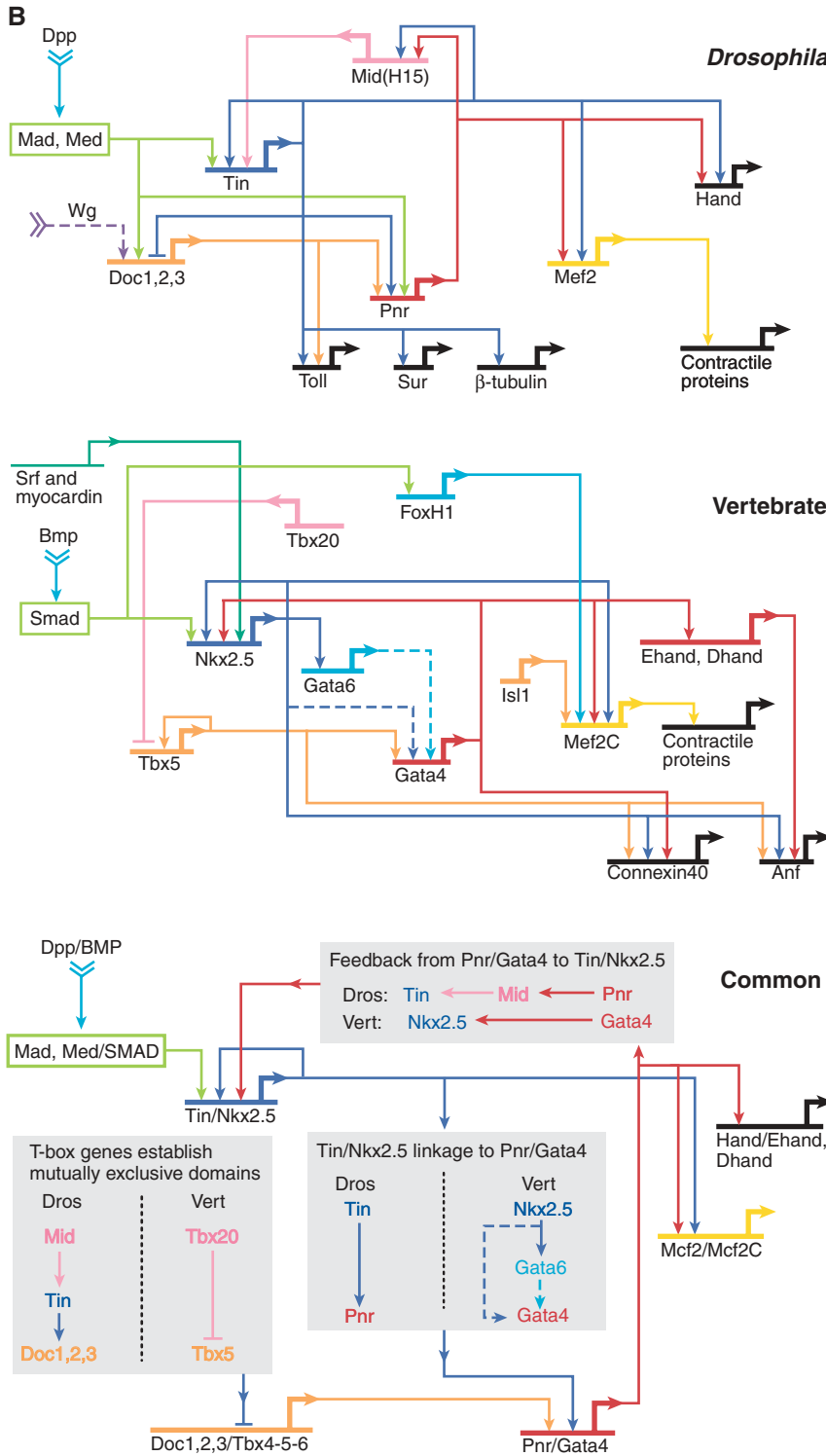


cute cell type-specific functions. Such functions contrast with those of kernels and plug-ins, the significance of which is entirely regulatory. Differentiation gene batteries build muscle cells and make skeletal biominerals, skin, synaptic transmission systems, etc. The structure of differentiation gene batteries has been discussed at the network level elsewhere (10); as an example, see the skeletogenic and pigment cell differentiation gene batteries in fig. S1. Differentiation gene batteries display inherent evolutionary lability and undergo continuous renovation. Numerous examples can be found in studies of speciation [e.g., (39)]. Internal changes occur in differentiation gene batteries in various ways: Any of the tens or hundreds of structural genes constituting their working components may alter functionally by incremental changes in their protein-coding sequences; new genes may be added to them if they acquire cis-regulatory

modules targeting members of the given small set of transcriptional regulators to which that battery responds; or similarly, they may lose genes. But differentiation gene batteries reside at the periphery of developmental GRNs (40), because their outputs terminate the network. They are expressed in the final stages of given developmental processes. They do not regulate other genes, and they do not control the progressive formation of spatial patterns of gene expression that underlies the building of the body plan; in short, they do not make body parts. They receive rather than generate developmental instructions.

Cis-regulatory linkages that may be considered as I/O switches regulating other network subcircuits appear to be responsible for many kinds of evolutionary change in developmental process. For example, a common form of variation, which must be trivial at the regulatory

level because it occurs even within genera and species, is in size of homologous body parts. We can easily imagine that this parameter depends only on the input linkage between a regulatory gene of the network controlling the patterning of the body part, and a cell cycle cassette; indeed, such linkages are explicitly known, for instance, in the gene network regulating pituitary development where the target is the cell cycle control genes (41). Here, the *pitx2* regulatory gene specifically activates the cell cycle control genes *cyclin D1*, *cyclin D2*, and *c-myc*. Many *hox* gene functions are also in this class: They act to permit or prohibit patterning subcircuits from acting in given regions of an animal. Examples include the direct repressive effects of the *Ubx* gene product on expression of wing-patterning genes in the *Drosophila* haltere (42–44); the role of group 10 and 11 *hox* genes in specification of vertebral morphol-



ogy in vertebrates (45); and, in beetles, the function of Ubx to allow the wing-pattern network to operate in the forewings, preventing expression of a different program expressed normally in the hindwing (46).

#### Predicted Evolutionary Consequences of Changes in GRN Architecture

Viewed in this way, it is apparent that the effects of changes in different component classes will

be qualitatively distinct, causing disparate kinds of effect on body plan and on adaptive organismic functionality. Furthermore, there emerges a relation between the network-component class in which changes might occur and the taxonomic level of morphogenetic effects (Fig. 3).

The most frequent and least constrained kinds of change will occur in the peripheral regions of the GRN, i.e., within differentiation gene batteries themselves and the apparatus that

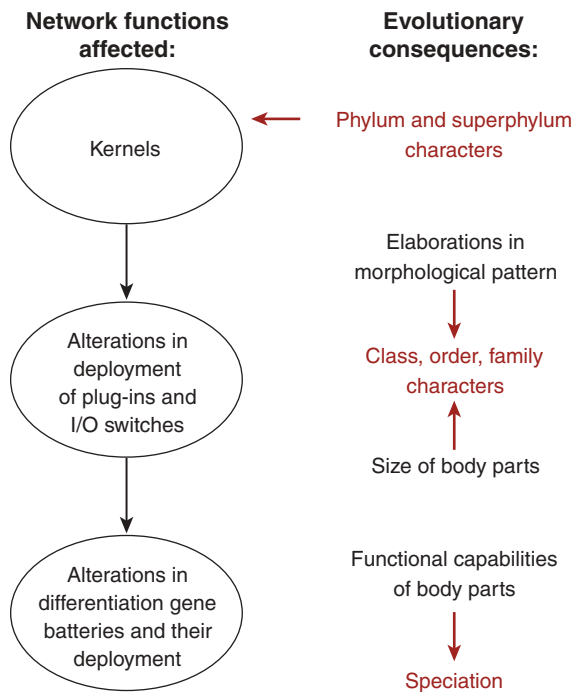
controls their deployment. This is for the simple reason that there are no downstream consequences in cis-regulatory wiring elsewhere in the network if peripheral input linkages change, as will commonly result from change in more internal locations. Such peripheral, small changes are just what is observed in the countless processes of speciation. They account for many adaptive properties of the organism, for instance, different properties of the integument, the repertoire of digestive enzymes, the positioning of peripheral sensory elements, etc.

At the other extreme (Fig. 3) are the kernels of the network. They operate the peculiarly crucial step of specifying the domain for each body part in the spatial coordinate system of the postgastrular embryo. We think that change in them is prohibited on pain of developmental catastrophe, both because of their internal recursive wiring and because of their roles high in the developmental network hierarchy. We predict that when sufficient comparative network data are available, there will be found conserved network kernels similar in complexity and character to those of Fig. 2, which program the initial stages of development of every phylum-specific body part and perhaps of superphylum and pan-bilateria body parts as well. It would follow that these kernels must have been assembled during the initial diversification of the Bilateria and have retained their internal character since. Critically, these kernels would have formed through the same processes of evolution as affect the other components, but once formed and operating to specify particular body parts, they would have become refractory to subsequent change. Molecular phylogeny places this evolutionary stage in the late Neoproterozoic when Bilateria begin to appear in the fossil record (47–51), between the end of the Marinoan glaciation at about 630 million years ago and the beginning of the Cambrian. Therefore the mechanistic explanation for the surprising fact that essentially no major phylum-level body parts have evolved since the Cambrian may lie in the internal structural and functional properties of GRN kernels: Once they were assembled, they could not be disassembled or basically rewired, only built on to.

Between the periphery of developmental GRNs and their kernels lies the bulk of the network architecture. Here we see skeins of special cross-regulatory circuitry, plug-ins, and I/O connections; and here is where have occurred the changes in network architecture that account for the evolutionary novelties attested in the fossil record of animals.

#### Reinterpreting the Evolutionary Record

We propose that architectural changes in animal body plans have been produced over the past 600 million years by changes in GRNs of at least three general classes, with extremely different developmental consequences and rates of occurrence. This challenges the generally time-



**Fig. 3.** Diverse kinds of change in GRNs and their diverse evolutionary consequences. The left column shows changes in network components; the right column shows evolutionary consequences expected, which differ in their taxonomic level (red).

homogeneous view of most evolutionary biologists. Current microevolutionary thinking assumes that observed types of genetic change (from single base substitutions to gene duplications) are sufficient to explain all evolutionary events, past and present. Such changes are considered as having occurred during evolution in a temporally homogeneous way. Microevolution does intersect with mechanisms of GRN change at the level of change within cis-regulatory modules. But attempting to explain an aspect of animal evolution that depends on one kind of network alteration by adducing evidence from an aspect that depends on another can be fundamentally misleading. Comparative molecular dissection of GRNs should allow identification of the evolutionary point of origin of each sub-circuit and linkage in the network, and hence each morphological character of the body plan.

If the early assembly of kernels underlies the phyletic conservation of body parts since the Cambrian, then the position in GRNs of subsequent adaptational change is forced to lower levels in the network hierarchy. The result is what has been termed developmental or phylogenetic constraints (52–54). The different levels of change that have occurred in evolution are imperfectly reflected at different levels of Linnean classification, and we think that these inhomogeneous events have been caused by architectural alterations in different locations in

the underlying GRNs. Following the early assembly of kernels, the varying effects of plug-in redeployment, changes in I/O linkages, and piecemeal alterations in differentiation gene batteries provide a basis for mechanistic analysis of subphyletic animal evolution. To the extent that kernel formation underlies critical morphological innovations, some kernels must indirectly be responsible for major events in Neoproterozoic niche construction. Motility, predation, digestion, and other canonical features of the Bilateria followed from the evolutionary appearance of the genetic programs for the respective body parts. These innovations became an engine of change that irreversibly altered the Earth's environment and, thus, the probability of success of subsequent evolutionary changes. We believe that experimental examination of the conserved kernels of extant developmental GRNs will illuminate the widely discussed but poorly understood problem of the origination of animal body plans in the late Neoproterozoic and Cambrian and their remarkable subsequent stability.

#### References and Notes

1. M. Levine, E. H. Davidson, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4936 (2005).
2. A. Ochoa-Espinosa *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4960 (2005).
3. T. Koide, T. Hayata, K. Cho, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4943 (2005).
4. H. Singh, K. Medina, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4949 (2005).
5. T. Inoue, M. Wang, T. Ririe, J. Fernandes, P. Sternberg, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4972 (2005).
6. M. F. Maduro, J. H. Rothman, *Dev. Biol.* **246**, 68 (2002).
7. R. M. Cripps, E. N. Olson, *Dev. Biol.* **246**, 14 (2002).
8. I. Reim, J. P. Mohler, M. Frasch, *Mech. Dev.* **122**, 1056 (2005).
9. P. Oliveri, E. H. Davidson, *Curr. Opin. Genet. Dev.* **14**, 351 (2004).
10. E. H. Davidson, *Gene Regulatory Systems. Development and Evolution* (Academic Press, San Diego, 2001).
11. J. W. Valentine, *On the Origin of Phyla* (Univ. of Chicago Press, Chicago, 2004).
12. D. H. Erwin, J. W. Valentine, J. J. Sepkoski Jr., *Evolution Int. J. Org. Evolution* **41**, 1177 (1987).
13. J.-Y. Chen, *The Dawn of Animal World* (Jiangsu Science and Technology Publishing House, Nanjing, 2004).
14. V. F. Hinman, A. Nguyen, R. A. Cameron, E. H. Davidson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13356 (2003).
15. For sources of each linkage shown, see (8, 9) and supporting online material.
16. E. H. Davidson *et al.*, *Science* **295**, 1669 (2002).
17. C. J. Lowe *et al.*, *Cell* **113**, 853 (2003).
18. R. Lichtneckert, H. Reichert, *Heredity* **94**, 465 (2005).
19. F. Hirth, L. Kammermeier, E. Frei, U. Walldorf, M. Reichert, *Development* **130**, 2365 (2003).

20. D. Arendt, K. Nübler-Jung, *Development* **126**, 2309 (1999).
21. R. A. Cornell, T. Von Ohlen, *Curr. Opin. Neurobiol.* **10**, 63 (2000).
22. C. Punzo, M. Seimiya, S. Flister, W. J. Gehring, S. Plaza, *Development* **129**, 625 (2002).
23. S. J. Silver, I. Rebay, *Development* **132**, 3 (2005).
24. B. A. Afouda, A. Cia-Uitz, R. Patient, *Development* **132**, 763 (2005).
25. R. A. Shivdasani, *Dev. Biol.* **249**, 191 (2002).
26. J. A. Hoffmann, F. C. Kafatos, C. A. Janeway Jr., R. A. B. Ezekowitz, *Science* **284**, 1313 (1999).
27. E. V. Rothenberg, E. H. Davidson, in *Innate Immunity*, R. A. B. Ezekowitz, J. A. Hoffman, Eds. (Humana, Totowa, NJ, 2003), pp. 61–88.
28. H. Wada, P. W. H. Holland, S. Sato, H. Yamamoto, N. Satoh, *Dev. Biol.* **187**, 240 (1997).
29. D. Meulemans, M. Bronner-Fraser, *Dev. Cell* **7**, 291 (2004).
30. S. Barolo, J. W. Posakony, *Genes Dev.* **16**, 1167 (2002).
31. K. M. Cadigan, R. Nusse, *Genes Dev.* **11**, 3286 (1997).
32. D. M. Kingsley, *Genes Dev.* **8**, 133 (1994).
33. G. Szebenyi, J. F. Fallon, *Int. Rev. Cytol.* **185**, 45 (1999).
34. A. P. McMahon, P. W. Ingham, C. J. Tabin, *Curr. Top. Dev. Biol.* **53**, 1 (2003).
35. S. Artavanis-Tsakonis, K. Matsuno, M. E. Fortini, *Science* **268**, 225 (1995).
36. B.-Z. Shilo, *Exp. Cell Res.* **284**, 140 (2003).
37. M. Tokuoka, K. S. Imai, Y. Satou, N. Satoh, *Dev. Biol.* **274**, 211 (2004).
38. B. L. M. Hogan, *Genes Dev.* **10**, 1580 (1996).
39. J. C. Avise, *Phylogeography: The History and Formation of Species* (Harvard Univ. Press, Cambridge, MA, 2000).
40. H. Bolouri, E. H. Davidson, *Dev. Biol.* **246**, 2 (2002).
41. X. Zhu, M. G. Rosenfeld, *Curr. Opin. Genet. Dev.* **14**, 567 (2004).
42. S. D. Weatherbee, G. Halder, J. Kim, A. Hudson, S. Carroll, *Genes Dev.* **12**, 1474 (1998).
43. B. M. Hersch, S. B. Carroll, *Development* **132**, 1567 (2005).
44. R. Galant, C. M. Walsh, S. B. Carroll, *Development* **129**, 3115 (2002).
45. D. M. Wellik, M. R. Capecchi, *Science* **301**, 363 (2003).
46. Y. Tomoyasu, S. R. Wheller, R. E. Denell, *Nature* **433**, 643 (2005).
47. K. J. Peterson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6536 (2004).
48. S. Aris-Brosou, Z. Yang, *Syst. Biol.* **51**, 703 (2002).
49. E. J. Douzery, E. A. Snell, E. Bapteste, F. Delsuc, H. Philippe, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15386 (2004).
50. D. H. Erwin, E. H. Davidson, *Development* **129**, 3021 (2002).
51. D. Condon *et al.*, *Science* **308**, 95 (2005).
52. J. Maynard Smith *et al.*, *Q. Rev. Biol.* **60**, 265 (1985).
53. K. Schwenk, *Zoology* **98**, 251 (1994).
54. P. Alberch, in *Evolution and Development*, J. T. Bonner, Ed. (Springer-Verlag, Berlin, 1982), pp. 312–332.
55. W. J. R. Longabaugh, E. H. Davidson, H. Bolouri, *Dev. Biol.* **283**, 1 (2005).
56. <http://supp.caltech.edu/endomes>
57. I. Reim, M. Frasch, *Development* **132**, 4911 (2005).
58. E.H.D. acknowledges support from NASA grant NAG2-1587, U.S. Department of Energy grant DE-FG02-03ER63584, and NIH grant HD37105. D.H.E. acknowledges the Packard Foundation and the Santa Fe Institute. We thank A. H. Knoll, D. Krakauer, E. V. Rothenberg, M. Levine, and four anonymous reviewers for useful comments on earlier drafts of the manuscript.

#### Supporting Online Material

[www.sciencemag.org/cgi/content/full/311/5762/796/DC1](http://www.sciencemag.org/cgi/content/full/311/5762/796/DC1)

SOM Text

Fig. S1

References and Notes

10.1126/science.1113832