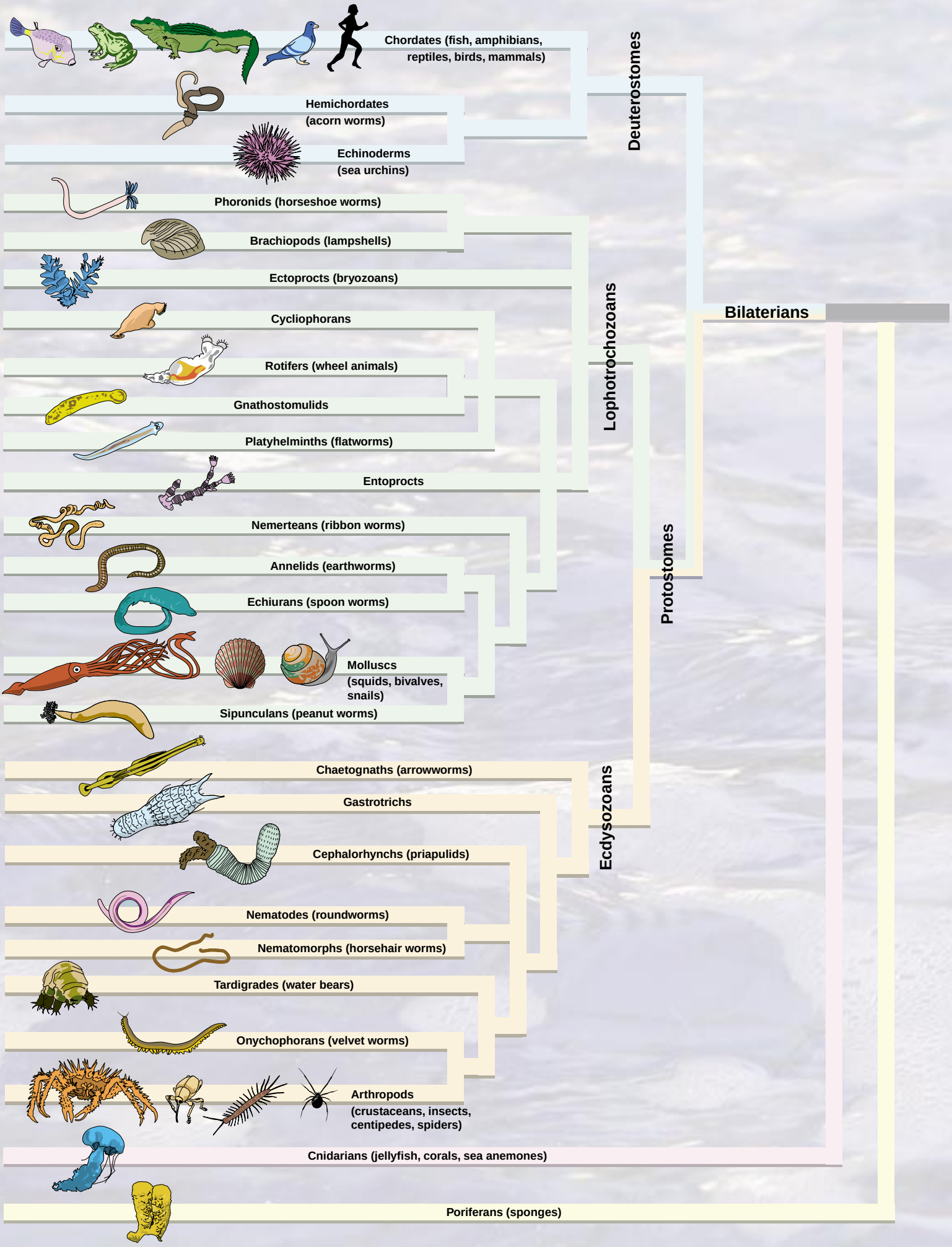

Chordates (fish, amphibians, reptiles, birds, mammals)
Hemichordates (acorn worms)
Echinoderms (sea urchins)
Phoronids (horseshoe worms)
Brachiopods (lampshells)
Ectoprocts (bryozoans)
Cycliophorans
Rotifers (wheel animals)
Gnathostomulids
Platyhelminths (flatworms)
Entoprocts
Nemerteans (ribbon worms)
Annelids (earthworms)
Echiurans (spoon worms)
Molluscs (squids, bivalves, snails)
Sipunculans (peanut worms)
Chaetognaths (arrowworms)
Gastrotrichs
Cephalorhynchs (priapulids)
Nematodes (roundworms)
Nematomorphs (horsehair worms)
Tardigrades (water bears)
Onychophorans (velvet worms)
Arthropods (crustaceans, insects, centipedes, spiders)
Cnidarians (jellyfish, corals, sea anemones)
Poriferans (sponges)
Deuterostomes
Lophotrochozoans
Ecdysozoans
Protostomes
Bilaterians

Animal Evolution: A View from the Genome

by Barbara Ellis

How did animals evolve, in the words of Darwin, "from so simple a beginning endless forms most beautiful and most wonderful"? How did the simple bodies of the first truly multicellular animals lead to six-legged insects, five-armed starfish, four-legged mammals, and legless snakes? Where did novelties such as insect wings and bird feathers come from? There's now a way, without having to rely on the patchy and often nonexistent fossil record, to trace back the origins of the different body plans and anatomical structures that give the animal kingdom its rich diversity. The evolutionary history of the animal kingdom is embedded in the genomes of the animals alive today, and can be studied in the laboratory.

This breakthrough has come from developmental biology—the study of how embryos develop—and it's created a huge upsurge of interest in the evolution of development. In fact, a completely new field of bioscience (colloquially referred to as evo-devo), in which evolutionary biologists, developmental biologists, paleontologists, and phylogeneticists share their expertise, is taking shape. Things are moving forward rapidly, and fascinating new insights into animal evolution are being published almost on a weekly basis.

To investigate the genetic changes that led to the evolution of different body plans, it's very important to have an accurate idea of the evolutionary relationships (phylogeny) of the animals alive today—to know who is descended from whom. Molecular phylogeny, such as comparing ribosomal DNA, has recently clarified a lot of the doubtful relationships. The "family tree" system of animal classification, with single-celled animals at the base and humans at the top, is out of favor nowadays, because it implies that evolution has a direction, and it's going our way. The branching diagram shown opposite, called a cladogram because it links clades (groups of animals who have all descended from a common ancestor), is a much more accurate way of showing relationships. Pairs of clades related by a common ancestor are linked by straight lines, so every branch in a cladogram is a "Y" (or a tuning-fork shape, as here). Most animals are now known to belong to a huge clade called the Bilateria, a name that refers to their unifying feature of bilateral symmetry; the body plan has a right-left axis, and a front-back axis and a top-bottom one, too. All bilaterian animals evolved from the same ancestral animal. The next most closely related clade to the Bilateria includes jellyfish, sea anemones, and corals (collectively called the Cnidaria), while sponges (the Porifera) are more distantly related.

Bilaterians are a step up in multicellular complexity from jellyfish and sponges, which have just two tissue layers—the ectoderm and the endoderm. The Bilateria have a third tissue layer, the mesoderm, between the ecto- and endoderm. And although cnidarians have some functional differences between layers of cells, only bilaterians have the complex 3-D arrays of cell types called organs.

The bilaterian lineage divided early on into two clades: the deuterostomes, which gave rise to the vertebrates and their cousins, and the protostomes. The latter divided again into the ecdysozoans and the lophotrochozoans. Insects, spiders, crustaceans, nematodes (roundworms) and the like are ecdysozoans, a name that derives from the fact that they all molt ("ecdysis"), while the lophotrochozoan clade embraces molluscs, earthworms, flatworms, and many lesser-known phyla (one of which, the Cycliophora, has only one member, discovered a few years ago on the mouthparts of the Norwegian lobster). Nowadays, each of these three great clades has a set of characteristics unique to that clade, but many—those inherited from the original bilaterians—are common to all three. By comparing what's unique and what's shared between descendants of common ancestors, it is now possible to work out the genetic changes that have given us the wonderful anatomical variety that we see all around us today.

This current picture of the evolutionary relationships among multicellular animals—i.e. the order in which they branched from common ancestors—is based mainly on ribosomal DNA analysis. The watery backdrop is the "gene pool" of the Beckman Institute.

A remarkably well-preserved upper Lower Cambrian fossil of a segmented worm, about 525 million years old, from the Chengjiang deposits in South China.

The first bilaterians evolved in the remote Precambrian, perhaps as much as 600 to 1,200 million years ago, from an ancestor shared with the cnidarians. Precambrian fossils are extremely rare, but with the current upsurge of interest in evolution, palaeontologists are searching worldwide for more, and they'll doubtless find them. Already, the 590- to 550-million-year-old Doushantuo deposits in southwest China have yielded some microscopic animal fossils bearing a striking resemblance to the embryos of modern bilaterians. It looks pretty certain now that the major evolutionary diversification of the bilaterians into the three primary clades also occurred in the Precambrian. The Cambrian period (545–490 million years ago) has an abundance of fossils, in striking contrast to the Precambrian, and they reveal a flamboyant blossoming of body plans and novel structures. During this era, almost all the major animal groups on earth today made their appearance, although one prominent group, the vertebrates, didn't appear until the Silurian, 100 million years later.

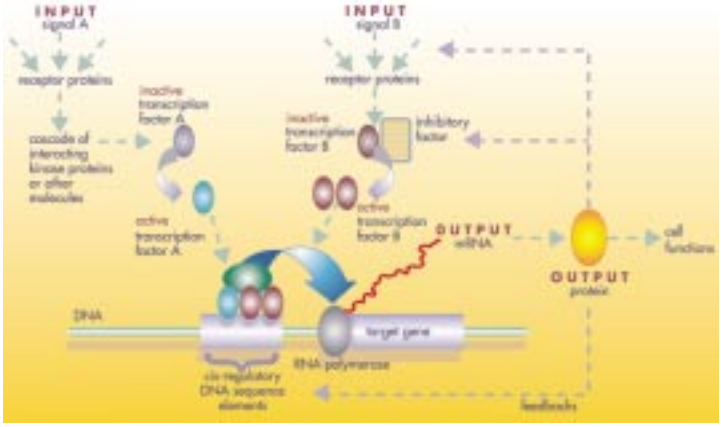
What made the bilaterians so much more successful than their cnidarian relatives, enabling them to spread out across the planet; adapt to life in seawater, freshwater, land and air; and grow as large as dinosaurs and whales? Their diversity and complexity are the result of having a larger complement of genes or gene families, a more sophisticated system of gene regulation, and, in particular, an "abstract patterning" mechanism for building body parts during development. In his new book, *Genomic Regulatory Systems: Development and Evolution*, Eric Davidson, the Norman Chandler Professor of Cell Biology, whose pioneering work on regulatory gene analysis contributed greatly to the current progress in understanding evolution, calls this abstract patterning mechanism "the secret of the bilaterians."

Development is a difficult task for a multicellular animal. It starts life as a single cell, which divides over and over again as quickly as it can into many, initially almost identical, cells, which then differentiate into specialized tissues and organs and anatomical structures such as limbs and wings. And every member of a species must develop correctly in the same way, each and every time. So why does one small, round cell develop into a sea urchin and another very similar one into a mammal? An obvious answer would be that the genes are different. But they're not: data from the genome sequencing projects, which have now provided full sequences for a variety of animals, has confirmed what was already becoming apparent from other research—bilaterians all have the same basic set of developmental genes. Some have duplicate copies and some have lost a few during evolution, but there's an astonishing commonality. Which presents an intriguing paradox: if the genes are the same, how come there are so many different types of animals?

"In development it is as if the wall, once erected, must then turn around and talk to the ceiling in order to place the windows in the right positions, and the ceiling must use the joint with the wall to decide where its wires will go."

At the molecular level, development involves the execution of a remarkable genetic program that regulates the construction of an organism. Of the thousands of genes in the genome, most are used at some time during development, and their deployment must be controlled accurately in space and time. The answer to the paradox lies not in the genes, but in the gene regulatory program, a program unique to each species. In some ways, writes Davidson in his book, this genetic program can be likened to an architect's blueprint for a large and complex building. Different buildings—perhaps a railway station and a cathedral—can be made from the same set of stones. It's the blueprint that dictates the different arrangement of the stones to make the different buildings. Similarly, different animals can be made from the same set of genes by following different blueprints. But animal blueprints also have to be interactive. "In development it is as if the wall, once erected, must then turn around and talk to the ceiling in order to place the windows in the right positions," he writes, "and the ceiling must use the joint with the wall to decide where its wires will go." Development also means a progressive increase in complexity; new populations of cells are generated, each of which reads out a genetic subprogram. And all the time, these populations are being instructed to expand to a given extent, through cell growth.

There's no "master gene" that coordinates development, the way a site foreman would oversee the implementation of an architectural blueprint. Each cell of the embryo has the same complete set of genes, derived from the fertilized egg cell. What makes one cell different from its neighbor depends on which genes are expressed, or turned on, to make proteins that have some function in the cell or transmit signals between cells, and which genes are blocked. Instead of a



Courtesy US Department of Energy Genomes To Life program, DOEGenomesToLife.org

An example of how a *cis*-regulatory element works. Signals A and B, which can be intra- or extracellular, activate transcription factors A and B along a signalling pathway. On reaching the *cis*-regulatory element, they help to initiate the synthesis of mRNA by RNA polymerase situated at the beginning of the target gene. The mRNA is translated into a functional protein (perhaps another transcription factor), which also provides feedback loops into the system.

single site foreman, the genes in each cell are controlled by short sequences of DNA called *cis*-regulatory elements. (*Cis*- means they're part of the DNA, as opposed to molecules that are *trans*-, not part of the DNA.)

With impressive foresight, way back in 1969, Davidson and colleague Roy Britten, now Distinguished Carnegie Senior Research Associate in Biology, Emeritus, proposed a theoretical model for such a system of genetic regulation. The underlying logic turned out to be more or less correct, but it wasn't possible to know for certain for another 30 years; only now are the tools of molecular biology (many of them developed by the Davidson group) good enough to detect such very small sequences of DNA and to analyze their function. A gene is thousands of base pairs long (a base pair is one "rung" of the DNA ladder), but the *cis*-regulatory elements have sequences of only a few hundred base pairs.

In essence, each *cis*-regulatory element functions like a tiny but very powerful biological computing device.

Cis-regulatory elements are usually adjacent to the gene they control but, just to make things more interesting, they can sometimes be several thousands of base pairs away along the chromosome. Although a few genes are controlled by just one *cis*-regulatory element, most are regulated by more than one, and some have a whole chain of them strung out along the chromosome. They're essentially "devices that make choices," says Davidson. Each *cis*-regulatory element has, on average, four to eight regulatory proteins, called transcription factors, associated with it. These proteins bring information to the *cis*-regulatory element from the world outside the cell nucleus, from other genes within it, and from neighboring cells (see diagram, above). When these transcription factors arrive (by diffusion) at the *cis*-regulatory element, they "dock" onto their own particular "landing bay," a very short sequence of DNA specific just to that transcription factor. Whether or not a transcription factor docks depends on its concentration and sometimes on its activation by other molecules, the cofactors. The *cis*-regulatory element "reads" the multiple inputs from the different transcription factors that dock—there could be one telling it what cell type it's going to become (muscle, nerve, or bone, for example), another to say where it is in relation to the other cells around it, yet another announcing that the cell is going to divide—and based on all this information the element produces a single output, an instruction that activates the gene or, as often as not, blocks it.

In essence, each *cis*-regulatory element functions like a tiny but very powerful biological computing device. The information-processing function of the *cis*-regulatory element is the link between the things that are happening in each cell and the response of the genes to them. And as the *cis*-regulatory elements are part of the DNA sequence, they're hardwired into the genome, and any changes in their sequence (such as by mutation, insertion or deletion of bases) are passed on to future generations—something that is of great significance in evolution.

As transcription factors are proteins, they're also encoded by genes, and these genes in turn are controlled by *cis*-regulatory elements. During development, certain genes that encode transcription factors play a very important role; they're known as the regulatory genes, and they choreograph the highly successful abstract patterning system of bilaterian development.

The first bilaterians probably developed in a way still seen in the embryos of many modern invertebrate marine animals. When a fertilized egg cell from such an animal starts to divide, the cells of the embryo get to know the cell type they're going to be as soon as they're born, and their "differentiation gene batteries"—sets of genes that are all expressed at the same time in a coordinated way so that the proteins they encode define the cell type—are turned on straight away. They're coordinated because their *cis*-regulatory

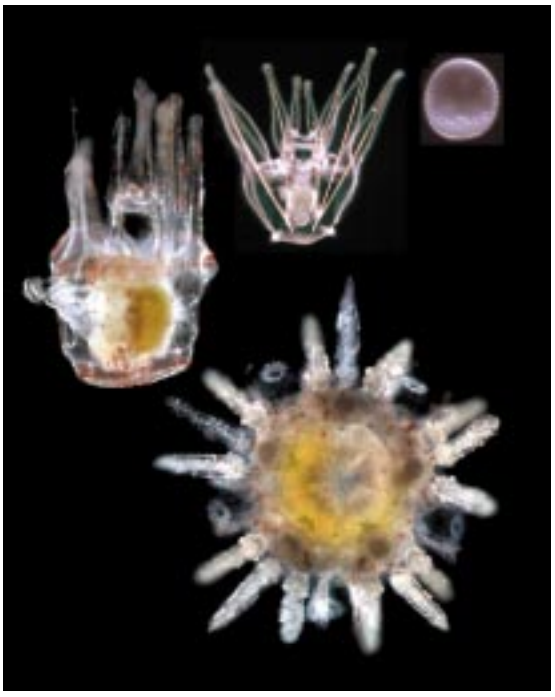
Davidson learned much about the ordering of complex perceptions from his father, leading American abstract expressionist painter Morris Davidson.

elements all respond to the same regulatory-gene transcription factor. This "direct cell-type specification" way of developing is an effective way of producing a free-swimming, self-feeding larval stage as quickly as possible, but it can only work when a small number of cells are involved, and seriously limits embryonic size to the product of about 10 cell-division cycles, or a few thousand cells.

If a more sophisticated system of development—abstract patterning or, to give it its full name, pattern formation by stepwise regional specification—hadn't evolved, bilaterians would never have grown any bigger or more complex than their jellyfish cousins. But as increasingly complex *cis*-regulatory control subcircuits were set up over time, linking regulatory genes encoding spatial transcription factors with those responsible for signalling pathways and growth control, larger embryos with new body structures could develop. Astonishingly, the subcircuits set up in those early days, more than 550 million years ago, are still used today by all bilaterians. Most types of invertebrate animals still use direct cell-type specification to get to the free-swimming larval stage, with pattern formation taking over after that to remodel the larva into an often very different adult form (as in the sea urchin, lower left). Interestingly, some groups that evolved after the appearance of the first bilaterian groups, such as the insects and the vertebrates, escaped this basic mechanism and devised their own ways of turning eggs into embryos.

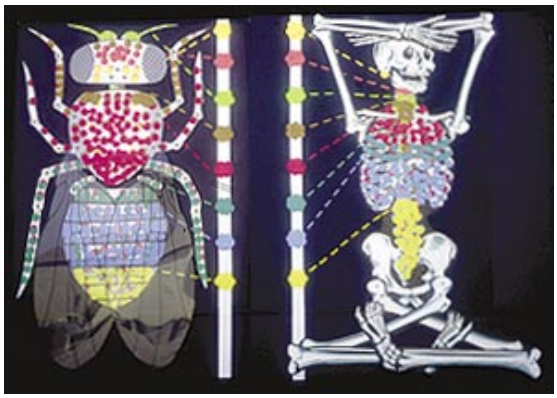
In the pattern-formation system of development, "a simple snapshot taken during developmental time in the animal will not resemble any parts of the structure that will finally emerge," writes Davidson. "Until the final stages, it'll look like abstract patterns." Early on, basic elements of the body plan such as the anterior-posterior axis and left-right symmetry are established. Later pattern-formation events define the spatial organization of the main parts of the body plan—head, tail, forelegs, hindlegs—then even later pattern-formation events define the detailed and smaller elements, such as the arrangement of the limb digits. Each stage involves the partitioning off of one group of

In the pattern-formation system of development, "a simple snapshot taken during developmental time in the animal will not resemble any parts of the structure that will finally emerge. Until the final stages, it'll look like abstract patterns."



C. Arenas-Mena *et al.*, *Development*, 2000, 127, 4631-4653. © Company of Biologists Ltd.

In the sea urchin, pattern formation remodels the larva into an adult form. From top (not to scale): embryo at the blastula stage; 8-arm larva, a small early rudiment of the adult body growing at the left-hand side of the stomach; metamorphosing larva with arm tissue contracting and tube feet of the 5-sided adult emerging from the side; one-week-old sea urchin juvenile.



Above: The *hox* genes are aligned in the same order on the chromosomes of fruit flies and humans. This diagram indicates roughly which body parts are patterned by which *hox* gene (courtesy Ed Lewis). Right: Spider from Corcovado National Park, Costa Rica (courtesy L. E. Gilbert, Integrative Biology, University of Texas at Austin).

cells into subgroups by the expression of regulatory genes encoding transcription factors. There can be a whole cascade of transcription factors, sometimes linked through signaling pathways, each of which controls the activity of other regulatory genes further downstream at the next level of regulation, which could again be genes encoding other spatially expressed transcription factors, and so on. Eventually, the gene batteries that make the differentiated tissues and organs are switched on, and it's only at this stage that an observer would start to see recognizable body parts emerging from the abstract picture. That's why it's called abstract patterning.

So the key players in the complex genetic program for pattern formation are the genes that encode the regulatory transcription factors and their *cis*-regulatory elements. Let's look at the best-known set of these, the *hox* cluster. In 1978, eight linked *hox* genes involved in the development of body segments in *Drosophila melanogaster*, aka the fruit fly, were discovered by Ed Lewis (PhD '42; now Morgan Professor of Biology, Emeritus). Lewis had been patiently working away on *Drosophila* in Kerckhoff Lab since 1939, and Davidson, who joined Caltech in 1970, and was already thinking about how development and evolution could be interlinked, feels he was fortunate to have been in the same department at that time. "Ed was always upstairs, and he used to say hey, come and look at this," he recalls. "I immediately realized that Ed's genes were some of the most interesting genes being worked on in the biology division." Lewis was awarded the Nobel Prize for his work in 1995 (see *E&S,* 1996, No. 1).

In the fruit fly, *hox* genes play an important role in the development of body segments. The key feature of these genes is order. They are ordered in the genome as two clusters in a long segment of the DNA on one of the chromosomes, and in space, the genes are expressed in the same general order along the body as that in which they lie along the chromosome. The first and second gene of one cluster is expressed in the head segments, then the third gene comes on a little farther posterior in the thorax, and so on. *Hox* genes have been found in every animal type looked at, and are always involved in anterior-posterior patterning of the body. "We were all surprised at that," Davidson recalls. No one had expected to find that humans had the same developmental genes as flies (left).

Even more surprising was finding that regulatory genes have been so highly conserved throughout evolution that they're sometimes even *interchangeable* between animals. Some fly *hox* genes have functioned well when transplanted into mice, and some mouse *hox* genes can replace those of flies. The *pax6* gene is particularly interesting. One of the important transcription factor-encoding regulatory genes, *pax6* is involved in development of the vertebrate eye. Its fruit-fly equivalent, *eyeless* (having these different gene names is confusing, but scientists had no idea, when they found and named them in their own particular lab animals, that they were dealing with the same genes) regulates development of compound insect eyes, with their numerous eyelets. Vertebrate and insect eyes are very different in construction, building materials, and the way they work. So what would happen if the *eyeless* gene of a fly was transferred into a mouse embryo? A mouse with fly eyes? No—the mouse develops a normal *mouse* eye, even using a fly gene. If a human *pax6* gene was transplanted into a spider, spider eyes would develop. No spiders would look out from their webs with six big, blue human eyes, unfortunately (or perhaps fortunately). Moreover, if *eyeless* or *pax6* genes are made to function in a different part of an embryo, the cells there form an extra, ectopic eye—frogs have grown extra frog eyes on their backs, flies have developed fly eyes on their legs (lower left).

The explanation for these unexpected results is that *pax6* is a regulatory gene active in a growing field of undifferentiated cells near the top of the embryonic patterning cascade mentioned earlier. It encodes a transcription factor that sets up a train of events leading to the formation and patterning of an entire structure (the eye), but it doesn't control the actual *construction* of the eye (such as the lens, the cornea, the optical pigment), which is done by batteries of genes further along in the development program. Mice always grow mouse eyes because the fly's *eyeless* gene activates the mouse's own eye-differentiation gene batteries, which go on to make all the parts for a mouse eye. To explain the ectopic eyes, think of the undifferentiated field of cells as a clean slate, prepared to respond to any of a number of regulatory genes that start a differentiation program. Inducing *pax6* to run in undifferentiated back cells of a frog embryo, or leg cells of a fly embryo, activates the eye-differentiation gene batteries in those cells. In fact, *pax6* works at the terminal differentiation

Czerny, T. *et al., Mol. Cell* 3, 1999, 297-307. © Elsevier Science

Ectopic eye induced on the leg of a fruit fly by forcing the expression of a fly homolog of the *pax6* gene in the embryonic leg.

The eyes of squids, flatworms, flies (top row) and vertebrates (bottom row: trumpet fish, human, heron) use different optical principles and visual pigments and are constructed from different materials, but their development is always initiated by the same *pax6* gene. (Animal photos courtesy BioMEDIA ASSOCIATES, www.ebiomedia.com.)

stages of eye formation as well, because over time, regulatory genes can gain extra functions in new areas and at several different levels of the cascade, and this can lead to the creation of new structures that, when preserved by natural selection, contribute to new animal forms.

How do pattern-formation systems reinvent themselves, and animal forms change, during development? One of the most important ways is by cooption, which is when a regulatory gene gains control of a new target site downstream, or controls the same apparatus but in a new area of the developing animal. The hypothetical regulatory gene followed through three different stages of evolution in the box on the opposite page shows how this could happen. Over time, the system becomes more and more complicated as the downstream effects of the gene affect more gene batteries; but all the while, the gene is still used for its ancestral function—to start development of the structure in the embryo. Cooption is rather like walking, writes Davidson. "One linkage, upstream or downstream, stays where it was last put and bears functional weight, while the other moves; and then, if its move is useful, it may serve as the functional anchor while the first changes. After a few such 'steps', all the linkages surrounding a given phase of activity of a regulatory gene may be different from the ancestral stage."

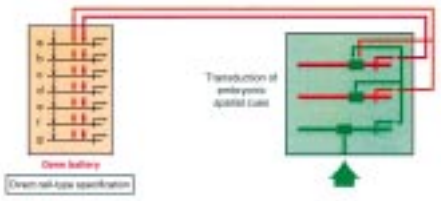
Normally, mutations to any of the genes active in the early stages of embryogenesis are just too disruptive to be survivable, and the mutation dies with the embryo. But if the cooptive change was such that the gene carried on doing its old job in addition to the new one, a viable, but somewhat different, animal could result, one that might survive to adulthood and pass this cooption on to its offspring. Small genetic changes in *cis*-regulatory control happen continuously in all animals. When they're at a downstream level of development fairly close to the final differentiation stages, they cause small differences between animals of the same species, the sort that breeders take advantage of. However, if *cis*-regulatory control of particularly significant *upstream* regulatory genes changed, there could be far more significant changes. A duplicated subset of the vertebrate *hox* cluster, for instance, was coopted to patterning limb development in vertebrates about 350 million years ago, a serendipitous evolutionary change that resulted in paired limbs—and enabled vertebrates to swim, walk, run, and fly their way all over the world.

If a *cis*-regulatory module controlling the timing of cell division gained a downstream gene controlling commitment—that moment in a cell's life when it stops developing and resigns itself to being an adult cell type forever—areas of the body could grow bigger or smaller. Let's imagine that this happened in the nose of a developing tapir, and that the program that determines the number of cell divisions changed from six to 10 cycles. There would be a 60-fold increase in size, and a baby tapir would be born with a bigger nose. This

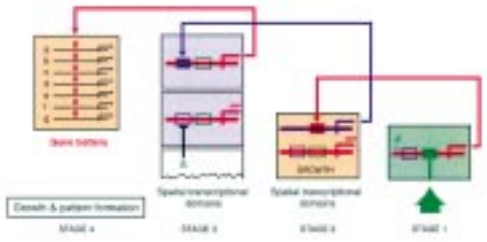
would be inherited by its offspring, so eventually lots of long-nosed tapirs would be running around. And if there was an evolutionary advantage in having such an extended nose, or if it increased the tapir's breeding success, a new species might eventually arise. Could this explain how the elephant got its trunk? It's far too simplistic a way of looking at speciation, of course,

COOPTION

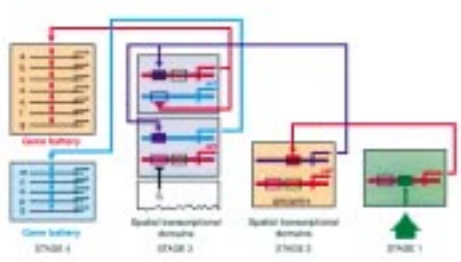
Three evolutionary stages of an imaginary pattern-formation system for a body part, showing how a simple system can gain complexity by cooption. The colored boxes are transcriptional domains, groups of cells whose state depends on the product of a gene (represented here as a thick horizontal line) of the same color. The short bent arrows indicate gene activation (transcription).



Stage 1, above, shows a simple, direct cell-type specification network. A green gene giving spatial cues from the embryo activates orange and red genes; all are regulatory genes encoding transcription factors. The gene battery encodes proteins used for some differentiated cell type and has *cis*-regulatory elements (a–g) that respond to input from transcription factors encoded by the orange and red genes.



In stage 2, a pattern-formation system has evolved. Focusing only on the red gene, we see it now activates a new, purple regulatory gene and a growth circuit. The purple gene product, another transcription factor, activates a second *cis*-regulatory element acquired by the red gene (which can be repressed by spatial signals from the embryo). The red gene is now activated by *cis*-regulatory interactions of the purple gene product with the purple *cis*-regulatory element of the red gene. It then activates the gene battery.



In stage 3, the pattern-formation process is even more elaborate. A new blue regulatory gene has been coopted by introduction into its *cis*-regulatory system of a *cis*-regulatory element that responds to the purple transcription factor (purple solid box). This blue gene activates a new, blue gene battery, which works in a different area of the embryonic structure being formed, thus increasing the complexity of this body part. The red gene now controls both the ancestral gene battery and the new one. But at all three stages of evolution, the green gene starts activation of the red gene, and the red gene still activates the gene battery that starts development of the body part.

but it shows how the evolution of developmental programs could play a role.

The Davidson lab at Caltech is at an advanced stage of mapping the entire network architecture of the *cis*-regulatory elements that control just 50 to 60 genes involved in the formation of the endomesoderm—the cell layers that produce most of the internal organs and tissues—in the sea urchin embryo, and of finding out how they're linked to one another by the regulatory transcription factors. It's an ambitious task, with layers and layers of complexity to unravel, and no one has dared attempt it before. But what they've found so far, Davidson says, is "extraordinarily interesting and illuminating," and he's optimistic: "Pretty soon I think we will understand the network. It's the evolutionary history of the animal, its heritage—it tells each gene what inputs it'll listen to throughout the life cycle. The *cis*-regulatory elements that control each gene enable it to respond to what it will encounter in every cell, every time, for the life cycle of the animal. That's what is hardwired into the genome. The network gives us a map of all these connections."

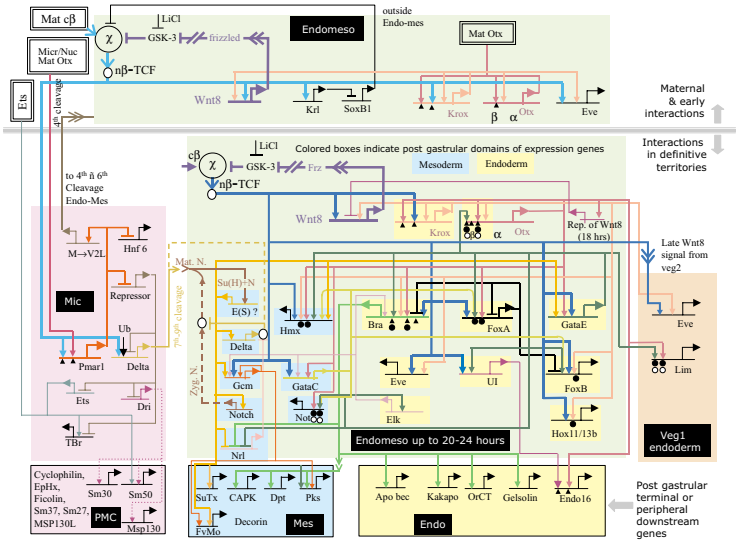
To investigate *cis*-regulatory elements involved in embryonic development and pattern formation requires fertilized eggs or one-cell embryos, because genes have to be injected into them to see what effect they have on their development. The beloved lab animal of the Davidson group, the California purple sea urchin, *Strongylocentrotus purpuratus*, provides them with an unlimited supply. "Years ago when I came to Caltech," Davidson recalls, "we built a huge egg-to-egg culture system at Caltech's Pacific outpost, the Kerckhoff Marine Laboratory in Corona del Mar, and we found sea urchins to stock it by diving for them." Once one of the regularly working scuba divers himself (see *E&S,* 1987, No. 4), he now mainly uses contract divers to do the work. The sea urchins live about 30 to 60 feet down in the

coastal waters, and have few natural enemies except the occasional fish, sea otters and fishermen supplying Japanese restaurants. For molecular developmental biologists, the sea urchin embryo has many virtues, including transparency, incredible fecundity, high tolerance for micromanipulation, easy gene transfer, and a simple embryology. Best of all, it grows into a larva that swims, feeds, and looks after itself in a matter of days. "Sea urchins are great for *cis*-regulatory analysis," Davidson says. "You do something with the eggs one day, and you get results the next. There's no need to wait until they grow up and have offspring."

To map an entire regulatory system, the Davidson lab has developed a new set of technologies for finding genes expressed in the endoderm or mesoderm at different times—they've found hundreds, which they're now sorting through to find the ones most central to the process. Then they intend to analyze the way in which these genes are regulated; that is, how they're connected in the network through their *cis*-regulatory regions. The *cis*-regulatory elements are notoriously difficult to find; the very short stretches of bases for each one are hidden in millions of base pairs of no apparent function, the so-called "junk DNA", but that's another story. To locate the regulatory elements, they're enlisting the help of evolutionary conservation: when the nucleotide sequences around these genes are compared among different species of sea urchins whose common ancestor is millions of years old, only regions that have a use remain unchanged. All parts of the DNA sequence that don't bind proteins can change over this long a time, so the short, unchanged segments will stand out in these comparisons. And these identical little patches that are the same between the different species have turned out to be the *cis*-regulatory elements—a very interesting finding. *Cis*-regulatory analysis comes next—this is really what the Davidson group is best known for. A short fragment of DNA containing the *cis*-regulatory element is isolated, attached to another piece of DNA that encodes a traceable protein (usually colored or fluorescent), and injected back into the embryo. The cells of the developing embryo in which the protein appears are the ones in which the gene regulated by that *cis*-regulatory element is active. This way, the network connections can be checked. The final stage is to "knock out" genes to find the effect on all the other genes in the system. All the results are fed into an impressive computational model, the "wiring diagram" (shown opposite), that's updated every week on the lab Web page (www.its.caltech.edu/~mirsky/endomeso.htm) as the results come in. Eventually, it will show where and when each gene is expressed throughout the various stages of endomesoderm development. It's going to be a lot of work, but once complete regulatory systems have been mapped for key members of the different animal clades, their similarities and differences will reveal the precise role of development in the evolutionary history of animals, something no one would have believed possible just a few years ago.

Soon, biologists will be able to work back, like comparative linguists who reconstruct extinct protolanguages from languages still spoken today, to the last common ancestor of all the Bilateria. As Davidson writes: "Although the ancestors of modern animals are extinct, the evidence of how they worked is still swimming, walking, flying around outside, in the form of the DNA of the modern bilaterians." And when groups of animals become extinct, what the planet is actually losing is their specific developmental-gene regulatory

Purple sea urchins at Kerckhoff Marine Lab during a winter harvesting campaign for rare nucleoproteins such as transcription factors, left, showing some of the 1,500 males and 1,500 females being spawned. A gravid female deposits about 10 million eggs (top left), so a total of 15 billion eggs can be collected, which are poured into 4-liter beakers (top right) and mixed with sperm from the males. Growing the fertilized embryos for 24 hours to the 200-cell stage provides 3 trillion nuclei from which workable amounts of nucleoprotein can be extracted.



The full regulatory gene network for endomesoderm specification mapped so far in the sea urchin embryo, showing all the linkages functional in different places and at different stages of the developmental process. Each short horizontal line represents the *cis*-regulatory element responsible for expression of a gene, and a short bent arrow extending from it indicates gene transcription. The colored lines connect transcription factors from the gene that encodes them to the *cis*-regulatory element or elements that they affect. This "wiring diagram" will get increasingly complex as more gene interactions are mapped.

networks—while the genes live on in other species.

Those ancient regulatory genes conserved in common by the three great clades, it is argued, must have been present in the original bilaterian. These include *hox, pax, orthodenticle* (for the nervous system), and quite a few others, so perhaps our ancestor swimming in Precambrian seas amongst the jellyfish was a small animal with a head end and a tail end, bilateral symmetry, a gut, nervous system, photoreceptor organs, and possibly some outgrowths or appendages. It's still only a blurred image, but it will get clearer as more regulatory gene networks are mapped.

Could we rewind evolution to restore extinct regulatory networks? Already, a team at the University of Southern California has succeeded in hatching chicks with tooth buds in their beaks; birds lost their teeth at least 60 million years ago. Other teams have had some success in giving snakes back their legs, and regenerating the eyes of eyeless cave fish. Could this be the way to reconstruct extinct animals? Admittedly it's a long way from a chicken with teeth to a complete dinosaur, but it's food for thought.

More than 600 million years of evolutionary experimentation have put the regulatory genes of the original bilaterians to many new uses in different areas of a developing embryo, to give us the rich diversity of animal life that we see all around us today. And the DNA in every cell of every animal alive today carries within it a forensic record of the changes that have happened over those millions of years. Deciphering it will keep the Davidson group and others busy for years, but we'll have a much better understanding of who we are and where we came from by the time the 200th anniversary of Darwin's birth comes around in 2009. □

Right: Could extinct animals eventually be recreated by restoring their lost regulatory networks in modern descendants? Artwork courtesy Chris Draper.

Davidson's book, Genomic Regulatory Systems: Development and Evolution *is published by Academic Press, 2001. For an introduction to the subject, try* From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design *by Sean Carroll, Jennifer Grenier, and Scott Weatherbee, published by Blackwell Science, 2001.*

PICTURE CREDITS:
42-43 – Doug Cummings