

NEIL SHUBIN

*Author of **Your
Inner Fish***

**SOME
ASSEMBLY
REQUIRED**

"Neil Shubin
shows himself
to be a natural
storyteller and a
gifted scientific
communicator."

—*The Wall
Street Journal*

Decoding Four

Billion Years of Life,

from Ancient Fossils to DNA

*To the memory of my parents,
Seymour and Gloria Shubin*

FIRST VINTAGE BOOKS EDITION, AUGUST 2021

Copyright © 2020 by Neil Shubin

All rights reserved. Published in the United States by Vintage Books, a division of Penguin Random House LLC, New York, and distributed in Canada by Penguin Random House Canada Limited, Toronto. Originally published in hardcover in the United States by Pantheon Books, a division of Penguin Random House LLC, New York, in 2020.

Vintage and colophon are registered trademarks of
Penguin Random House LLC.

The Library of Congress has cataloged the Pantheon edition as follows:

Name: Shubin, Neil, author.

Title: Some assembly required : decoding four billion years of life,
from ancient fossils to DNA / Neil Shubin.

Description: First edition. | New York : Pantheon Books, 2020. |
Includes index.

Identifiers: LCCN 2019022792

Subjects: LCSH: Life—Origin. | Paleontology. | Evolution (Biology).

Classification: LCC QH325 .S544 2020 | DDC 576.8—dc23
LC record available at <https://lcn.loc.gov/2019022792>

Vintage Books Trade Paperback ISBN: 978-1-101-97268-7

eBook ISBN: 978-1-101-87134-8

Original illustrations by Kalliopi Monoyios

Author photograph © Tangled Bank Studio/Howard Hughes Medical Institute

Book design by Anna B. Knighton

www.vintagebooks.com

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Each week I'd arrive with a question and share a pot of tea with the great man as he held forth on the history of the field while offering spirited opinions on the ideas and personalities that shaped it. In advance of each visit, I'd scavenge the literature to generate a good subject to serve as fodder for his reminiscences. Transported in time and space by his stories, I felt incredibly fortunate to have such an amazing gig at the start of my own career.

One Thursday I came with a book, *The Material Basis of Evolution*, by the German scientist Richard Goldschmidt, a paperback reprint of a volume first published in 1940. Showing it to Mayr, I saw his face turn beet red as his eyes shot through me with an icy glare. He rose, stood still, and didn't so much as acknowledge my presence for an interval that felt interminable. I had crossed some hidden line and was quite certain that I could say farewell to my Thursday teas.

Mayr walked silently to an old wooden file cabinet and rifled through its contents. He returned with a yellowed reprint of one of Goldschmidt's papers and slapped the article onto the table, saying, "I wrote my book in response to the crap in the first sentence of a paragraph toward the end of this." Taking his cue, I thumbed through the paper until I hit page 96. It was unmistakable; on it were more angry marginalia than original text.

Three and a half decades had passed between the publication of Goldschmidt's article and Mayr's rage. How could a single sentence, let alone idea, evoke such passion and catalyze an 811-page book that itself launched entire research careers?

At issue was how changes in genes could bring about new inventions in the history of life. The conventional view was that inventions emerge gradually over time with small genetic changes at each step. This notion was supported by such a large body of theoretical and laboratory work that it was almost taken

6



Our Inner Battlefield

SEEDS FOR MY WORK were sown during a weekly ritual I performed while I was a graduate student in the 1980s. Every Thursday morning I would trudge up five flights of stairs to a large storage area in Harvard's Museum of Comparative Zoology. Home to the bird collection, the space had creaky wooden floors and twenty-foot-high ceilings. The walls were lined with cabinets and shelves filled with skeletons, feathers, and skins collected during expeditions of the nineteenth and twentieth centuries. The smell of the mothballs that protected the skins wafted through the air. History also permeated the place, both for ornithology and for science as a whole. That link to the past was what drew me: my pilgrimages were to meet with the eighty-year-old retired bird curator, Ernst Mayr.

By the mid-1980s, Mayr was among the last living members of a generation of geneticists, paleontologists, and taxonomists who had defined the field of evolutionary biology in the mid-twentieth century. Mayr's part in this scientific achievement was to write one of the classic books of this time, *Animals, Species and Evolution*, an immense tome that guided research for a generation of scientists on the formation of new species.

as axiomatic. The British statistician Sir Ronald A. Fisher derived it mathematically in the 1920s as he tried to link the emerging field of genetics with Darwinian evolution. Part of the logic is embedded in the idea that if you were to make a random change to a system, large changes are more likely to be bad, often catastrophically so, than smaller ones.

Take, for example, an airplane. Any random change that departs dramatically from the norm is almost certainly going to lead to an airplane that can't fly. Randomly changing the shape of the body; the position, form, or shape of the engines; or the configuration of the wings would likely lead to a grounded monstrosity. But small tweaks, such as to the color of the seats or minor alterations in size, are less likely to be dire. Indeed, they have more of a chance of increasing performance than large changes do, even marginally. This kind of thinking dominated the field of evolutionary biology for years, to the point that challenging it was akin to denying that gravity causes apples to fall from trees.

Goldschmidt, a refugee from Nazi Germany, entered academe in the United States having studied mutants for decades. With his move to North America, he crashed the party in the field of genetics, unconcerned by the status quo. Impressed by mutants with two heads or extra body segments, such as those Calvin Bridges was discovering, he thought that a major transformation could happen in a single step with a single dramatic mutation. The drama behind the idea is captured in one of Goldschmidt's most famous remarks, indeed the one that enraged Mayr so thoroughly: "The first bird hatched from a reptilian egg." No gradual change here—in his view, biological revolutions happened with a single mutation in one generation.

Goldschmidt's mutants were given a name: "hopeful monsters." They were monsters because they differed so dramati-

cally from the norm, and hopeful in that they were the seeds for an entire revolution in the history of life. In the world of plants, where changes in chromosomal numbers could bring about new species all at once, Goldschmidt's idea was not controversial. For animals, however, things were very different.

The assault on Goldschmidt's idea was immediate and fierce. The most salient criticisms challenged the chances that a hopeful monster could be viable and ultimately reproduce. First, the mutation would need to make viable and fertile offspring. It was well known by that time that most mutants, let alone dramatic ones, were either sterile or died before they could give rise to offspring. Even if a mutant were to survive and be fertile, its fate would still be unsure. It wouldn't do if only a single mutant were present in a population—it would need to find a mate that also had the mutation. For Goldschmidt's hopeful monster to give rise to a major revolution in a single step, a chain of unlikely events would have to happen: a major mutation would have to make a viable adult; it would have to happen in males and females simultaneously; and some of those individuals would need to be able to find each other, mate, and rear their own offspring, which themselves could reproduce.

By the time I studied biology in the 1970s, Goldschmidt's reputation remained something between a pariah and a heretic, as somebody who had dared publish a view so obviously wrong. Not only did he publish it, he seemed to relish his contrarian role, spending the final decades of his career defending hopeful monsters, often to public ridicule.

Mayr, Goldschmidt, and their contemporaries were debating one of the central issues of life's diversity—how major evolutionary changes happen. Although Goldschmidt's hopeful monsters were implausible, open questions remained. The issue was not with gradual change; biologists have long known that small

incremental genetic changes could lead to massive revolutions over the millions of years of geological time. A deeper puzzle emerges from the fossil record. Take, for example, the origin of a skeleton, one of the biggest events in our own species' history. For millions of years, wormlike ancestors lived with no bones inside their bodies. Bone has a characteristic structure, with highly organized layers of cells that manufacture the distinctive proteins and crystals that give the skeleton its rigidity and regulate its ways of growing. The origin of a skeleton allowed our ancestors to get large and have a rigid body to find prey, avoid predators, and move about. This invention arose because of the emergence of a new kind of cell, one that can produce the proteins needed to make skeletons, nourish them, and help them grow. But different kinds of tissues—whether skin, nerve, or bone—are made by cells that make hundreds of different proteins. Nerve cells are distinct from skeletal cells because numerous proteins give them the ability to conduct nervous impulses. These, of course, are lacking in the skeleton and the cells that build it. Likewise, cartilage, tendon, and bone are made from proteins that nerve cells do not produce. And the skeleton is only one example: the nearly 600-million-year history of animal life involved the origin of hundreds of new tissues, which enabled new ways of feeding, digesting, moving, and reproducing.

And here is the challenge: the origin of new tissues and cells from those of ancestors requires changes to hundreds of genes. How could new cells and tissues arise if a multitude of separate mutations must happen simultaneously throughout the genome? If the odds of one incremental mutation happening are relatively small, then imagine the impossibility of hundreds of them happening at once. This would be akin to winning the jackpot on not just one roulette wheel but every single wheel in a casino at the same time.

Pregnant with Meaning

It is hard to miss my University of Chicago colleague Vinny Lynch in the gym: sporting tattoos of a menagerie of species on his arms and legs, he stands out even among inked college students. Dragonflies and fish in a river scene populate his appendages.

The river scene is an homage to the Hudson River ecosystem that nurtured his childhood love of science. Growing up in a town along its banks, he developed a passion for the creatures that lived at the water's margin. Documenting, drawing, and reading about different animals transported him to another world. Unfortunately, his curiosity about life's diversity did not translate into success in school. He was a failure because, as he described it, he "didn't listen to lectures"; instead he stared out the window at birds and insects.

Fortunately, one biology teacher saw through his idyll and let him sit at the back of the class with books and field guides that she'd quiz him on later. This experience provided by one sage instructor propelled him to a career in biology. He has spent his life ever since exploring how animal diversity comes about: not just how animals live, eat, and move, but how over millions of years they arose from distant ancestors. And his specialty is applying high technology to these deep questions.

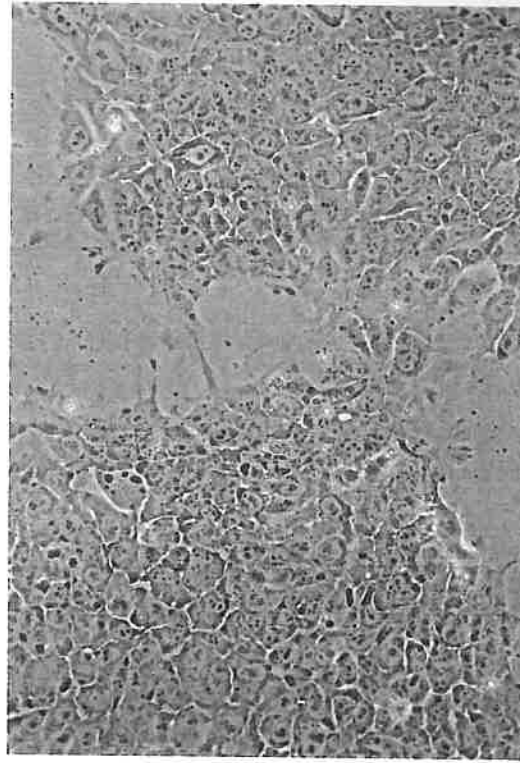
Progress in biology is as much about defining the right question as it is about finding an experimental system in which to explore it. T. H. Morgan found clues to genetics in flies. Barbara McClintock came to understand the working of genes in corn. Vinny Lynch is finding clues to the great revolutions in the history of life in decidua stromal cells.

Lynch's eyes widen as he describes decidua stromal cells.

When we first chatted about them, he gushed that they are some of the "most beautiful cells in the body." I'll admit it sounds impossibly nerdy, but once I saw them under the microscope, I came to agree. Most cells look like regular little dots under higher magnification. Not these. With big red bodies and rich connective tissue in between, they look almost lush, if you can apply that term to cells.

For Lynch, the beauty of decidual stromal cells is not only aesthetic but scientific. They are a window into the origin of one of the great inventions in the history of life: pregnancy. Most fish, birds, and reptiles, even very primitive mammals, hatch from eggs. They do not have the mammalian style of pregnancy, where the embryo develops within the mother and shares her blood supply. They also do not have decidual stromal cells.

Pregnancy seems at once completely natural and utterly miraculous. Sperm maneuver through the uterus and fallopian



A beautiful cell: decidual stromal cells

tubes to ultimately find the egg. Then one sperm (in rare cases more) enters the egg and sets off a chain reaction of events. Sperm and egg genomes merge, and the two become a single cell. Over time that cell gives rise to a body made of trillions of cells all packed in the right place. A placenta and umbilicus form to connect the mother and the fetus housed in the protective womb. For the womb to hold the fetus, a suite of new structures has to be constructed.

Fertilization brings about a cascade of changes in the body of the mother as well. In the uterus, specialized cells form to connect the fetus to her, bringing their blood supplies in close proximity. These cells mask the fact that the fetus is an alien inside the mother, having a contribution of genes and proteins from the father. There is always the risk that the mother's immune system could go on a search-and-destroy mission for paternal proteins and kill the fetus. Specialized cells dampen those differences. The cell that does much of this magic, from buffering the mother's immune response to channeling nutrients to the fetus, is the decidual stromal cell.

The trigger that makes these cells and initiates many of the changes in the uterus is a spike in the hormone progesterone in the mother's bloodstream. On a monthly basis, progesterone rises in the mother's bloodstream, and the uterus prepares for pregnancy. When progesterone contacts cells of the uterus, it causes them to multiply and change, making the lining of the uterus, the endometrium, thicker. The rising progesterone levels cause a set of cells known as fibroblasts to change into decidual stromal cells. If pregnancy does not happen that month, the cells slough off. But if pregnancy is achieved, the ovaries start to make progesterone, the cells and the rich cellular medium that lines the uterus continue to grow, and the decidual stromal cells form and start to do their work.

Lynch's fascination with these cells derived from a scientific talk he attended in Texas while he was a graduate student at Yale University. A researcher, speaking about pregnancy, showed slides of decidual stromal cells. Lynch learned that these cells had a special property: you could make them in a dish. The researcher had found that when he took normal fibroblasts from anywhere in the body, put them in a petri dish, and added a cocktail of progesterone and some other chemicals, he could make normal decidual stromal cells. Unknown to Lynch at the time, and by sheer coincidence, all this work was being done at Yale in the building next door to his own.

Lynch quickly learned to make decidual stromal cells in the controlled environment of the lab. He now could probe their genomes to see how they had come about millions of years ago. He had at his disposal a very powerful new technology, one that makes use of incredibly fast gene sequencers. Using this technology, he could look at a cell, or an entire tissue, and see the sequence of every single gene that was active in it—all of them at once.

Think about what a technology like this can do. If the differences between cells arise from the genes active in each one, then identifying the constellation of genes turned on in different cells becomes a critical part of the quest to understand what makes cells distinct. Recall that a nerve cell differs from a bone cell because different genes are making different proteins inside each. Likewise, a decidual stromal cell is distinct from a fibroblast in the genes that are active within. Lynch could look at one cell and compare it to another to ask fundamental questions: What are the differences in gene activity between the two cells? Is it one gene that makes them different, or is it several acting together, and if so, which ones are they?

Lynch took fibroblasts, put them in the dish, hit them with

progesterone, and turned them into decidual stromal cells. Then he looked at which genes were activated. The result was as surprising as it was formidable. The origin of decidual stromal cells didn't involve a single gene, or even a handful of them, being activated. Rather, hundreds of genes were turned on at the same time.

Decidual stromal cells are unique to mammals—no other creature has a version of them. Their origin is a central part of the origin of pregnancy itself. But therein lies the problem. If the origin of this single kind of cell involved hundreds of genes being turned on at the same time, then how could pregnancy happen? It would require hundreds of mutations arising simultaneously across the entire genome.

For Lynch to answer his questions would require looking at each of those hundreds of genes that make decidual stromal cells.

To consider Lynch's next step, we need to pause and consider what would make genes turn on in order to transform a cell into a decidual stromal cell. Recall that there are molecular switches across the genome that, under the right circumstances, turn genes on and off. Most of these switches lie right next to the genes that they activate. Since progesterone is the trigger for the formation of decidual stromal cells, then we could reasonably assume that the switches would be responsive to it. The genetic switches would be tethered to a sequence that recognized progesterone. When progesterone was present, the switch would activate and the gene would make protein.

This insight gave Lynch the clues he needed to probe the genome. He could look for the telltale signature of genetic switches that had, as part of their sequence, a region that recognized progesterone. This region would have a sequence that the hormone could bind to, so with any luck he could find them in a comparison of his genes within computer databases.

And that was exactly what he found. Almost all of the hundreds of genes that make decidua stromal cells had a switch that responded to progesterone. This discovery, while interesting, did little to answer the question that got Lynch into all this in the first place. Somehow, during the origin of pregnancy, hundreds of genes had to become active in response to progesterone. Since hundreds of genes are turned on in response to progesterone, hundreds of switches that respond to progesterone had to exist across the genome, near each of the genes that is activated by the hormone. This was no simple mutation of DNA, like changing a single letter in the code. Lynch was seeing a batch of letters that had to change simultaneously in hundreds of places across a genome to make decidua stromal cells. The implausible just got impossible.

As each new experiment made the origin of the cells ever more unlikely, Lynch returned to the structure of the genetic switches themselves. Perhaps something they all shared would offer an explanation? Looking in detail at the sequences, he used a computer algorithm to see if there was any shared pattern. A simple gene sequence emerged, one that was shared by virtually all the switches. Running the sequence across a huge database of all known sequences, he found the answer: each genetic switch had the telltale signature of a jumping gene, the type of gene that McClintock found first in corn. These genes, as we saw earlier, make copies of themselves to insert all over the genome. McClintock had seen them as great disrupters—that is, when they hop and insert into another gene, they can disrupt the function of that gene and make a pathology. Lynch saw something else.

This simple linkage made possible a complex, seemingly impossible invention. Hundreds of genes did not have to mutate

independently. Lynch saw that a mutation happened in a single jumping gene, turning a regular sequence into a switch that responded to progesterone. Then the mutation spread across the genome as the jumping gene with the switch duplicated, jumped, and landed in new places. Jumping genes distributed switches all over the genome very rapidly. When they landed next to a gene, that gene could now be turned on in response to progesterone. In this way, hundreds of genes gained the ability to be turned on during pregnancy. A genetic change, involving the coordination of hundreds of genes, could occur not by hundreds of independent mutations but by jumping genes carrying a single mutation throughout the genome. In this way, genetic changes could spread very quickly as genes jump, make copies of themselves, and land in different places.

Jumping genes are the ultimate selfish elements—they can duplicate and jump to spread and multiply across the genome. Lynch was finding that, on occasion, jumping genes can carry useful mutations that do dramatically new things.

There is a war going on inside the genome, between jumping genes and the rest of our DNA. That tension between a selfish gene and the forces that strive to control it occurs in genomes every day. It turns out that DNA has hidden mechanisms to control jumping genes. One of them involves a small DNA sequence that functions like a hunter-killer, able to silence jumping genes by attaching to the part of the gene that makes it jump, then literally bundling it up in protein so it cannot jump around. Neutered in this way, the gene doesn't jump; it stays put. This silencing mechanism can control jumping genes and stop them from taking control to the point of disrupting the workings of the genome. It may also serve to domesticate jumping genes. If a jumping gene contains a potentially useful sequence, the hunter-

killer DNA can neuter the jumping ability and make it stay put to play a new role. It can silence the jumping part but keep the helpful mutation.

That is what Lynch found with his switches: each of the switches that made decidua stromal cells had a special sequence that looked for all the world like it originally came from a jumping gene. But the gene had one difference: a small stretch of DNA was missing, and not just any DNA—the DNA that caused the gene to jump. It was as if the code had been hacked to stop the gene from jumping and keep it in place to do its work of making decidua stromal cells. With its springs clipped, the no-longer-jumping gene was put to work where it landed.

What Lynch saw in pregnancy is a window into a much larger world. Genomes are at war with themselves: between jumping genes and the forces that try to contain them. Out of this struggle comes invention, where a single mutation can spread across the genome and, over time, bring about a revolution.

These shifts are a far cry from Goldschmidt's hopeful monsters. A revolutionary mutation doesn't have to arise in a single step. An incremental change can arise in one place in the genome and, if tethered to a jumping gene, spread and be amplified over time in subsequent generations.

But the war inside the genome extends even wider. And pregnancy, again, reveals how.

Hacking the Hackers

In the placenta, right at the boundary between the fetus and the mother, one protein has a very special role to play. Syncytin sits at this interface and serves as a molecular traffic cop as the mother and fetus exchange nutrients and waste products. A num-

ber of observations show that this protein is vital for the health of the embryo. When a group of scientists made a mouse with a defective syncytin gene, the mice grew and lived normally, but they couldn't reproduce. After fertilization, the placenta would fail to form, and the embryo would not survive. Lacking syncytin, the mother could not make a functional placenta, and the embryo had no way of obtaining nutrients. Defects in syncytin also cause a wide range of problems in pregnancy in people. Women with preeclampsia have a defective syncytin gene; they make the protein, but it cannot do its job well. This sets off a chain reaction in the placenta that leads to dangerously high blood pressure.

A biochemistry laboratory in France began to look at the structure of the protein by exploring the sequence of DNA that makes it. As we saw with Lynch's work, once a gene is sequenced, the code can be run on a computer and compared to databases containing other genes in living creatures. These pattern-recognition packages cross-check the entire gene as well as small stretches of it for any similarities to other genes that have been sequenced. Over the past few decades, databases have been filled with millions of sequences of proteins and genes for everything from microbes to elephants. These comparisons have revealed that many genes are part of the duplicated gene families that we saw in Chapter 5. In the case of syncytin, the researchers were looking for similarities to other proteins that might give clues to how syncytin works during pregnancy.

The searches were revealing a puzzle. The database hunt showed that syncytin had no similarities to proteins in any other animals. It didn't look like anything in plants or bacteria either. The computer match was as bewildering as it was surprising: the sequence of syncytin looked for all the world like a virus and was identical in places to HIV, the virus that causes AIDS. Why

would a virus like this have any similarity to a protein in mammals, let alone one that is an essential part of pregnancy?

Before exploring syncytin, the researchers needed to become experts on viruses. Viruses are devious molecular parasites. They have genomes stripped of everything but the machinery needed for infection and reproduction. They invade cells, enter the nucleus, and enter the genome itself. Once in the DNA, they take over and use the host's genome to make copies of themselves and produce viral proteins instead of those of the host. With this infection, a single host cell becomes a factory to make millions of viruses. For a virus like HIV to spread from one cell to the next, it makes a protein that causes the host's cells to stick together. The protein brings the cells together and makes pathways for the virus to move from cell to cell. To do this, the protein sits at the interface between the cells and controls the traffic between them. Does this sound familiar? It should, because syncytin does the same thing in the human placenta. Syncytin brings cells together in the placenta and controls the traffic of molecules between the fetal and the maternal cells.

The more they looked, the more the team found that syncytin is essentially a viral protein that has lost its ability to infect other cells. This similarity between a mammalian protein and a virus led to a new idea. At some point in the distant past, a virus invaded our ancestors' genome. That virus contained a version of syncytin. Instead of commandeering our ancestors' genome to make endless copies of itself, the virus became neutered, lost its ability to infect, and then was put to work by a new master. Our genome is in a continual war with viruses. In this case, by mechanisms we have yet to understand, the infectious part of the virus was knocked out, and the virus was put to use making syncytin for the placenta. Viruses brought the protein to the

genome, and the attacker's genome was hacked to be useful for the host.

The scientists then looked at the structure of syncytin in different mammals and found that the version in mice is different from that in primates. Comparing the databases, they saw that different viral invasions are responsible for the syncytins in different mammals. The primate version arose when a virus entered the common ancestor of all living primates. The syncytin of rodents and other mammals came about from a different event, leading to their versions of syncytin. The end result is that primates, rodents, and other mammals have different syncytins derived from different invaders.

Our DNA is not entirely an inheritance from ancestors. Viral invaders have inserted themselves and been put to work: our ancestors' battles with them have been one of the many roots of invention.

Zombie Memories

When Jason Shepherd was a child growing up in New Zealand and South Africa, he so pestered his mother with questions that she finally told him he needed to become a scientist to find his own answers. By the time he graduated from high school, he had decided to enter medicine. He began a crash program to give him both premed and medical training in a short few years. In the first year of the program, he encountered Oliver Sacks's classic *The Man Who Mistook His Wife for a Hat*. That single book changed his life. Inspired by Sacks, he left the program and launched a new career studying the molecules and cells that make our brain work. His quest, as he describes it, became to

find out what makes us human. Memory, and its loss, became Shepherd's scientific quarry. Our ability to recollect the past defines much of how we learn, relate to others, and function in the world. This is no esoteric subject. One of the great challenges we face as a society is neurodegenerative disease. As we lengthen our lifespan, the aging brain serves as an ever more critical barrier. Loss of memory and cognitive function are scourges with emotional, social, and financial tolls that are incalculably large.

In Shepherd's senior year in college, as he was looking for a paper topic for a course in neurobiology, he ran across an article on a gene called *Arc* that appeared to be involved in making memories. In mice, *Arc* is turned on as creatures learn. Moreover, it is active in the brain in the spaces between the different nerve cells. *Arc* seemed to fit the bill of a gene important in memory.

A few years after Shepherd's college assignment, technology had evolved to the point where researchers were able to make mice lacking the *Arc* gene. The mice survived but had a number of defects. When offered a maze with cheese in the center, they could solve the maze, but they could not remember its structure the next day. This is something that mice with normal memories can often do. In test after test, the mice revealed a specific deficit in forming memories. Mutations of *Arc* in humans are known to be associated with a range of neurodegenerative disorders, from Alzheimer's to schizophrenia.

Memory and the *Arc* gene became the focus of Shepherd's career. He went to graduate school to study *Arc* with one of the biologists who had first explored its role in behavior. Then, after graduating, he did his postdoc training with the scientist who discovered where the *Arc* gene lies in the genome. Shepherd had *Arc* on the brain both literally and figuratively.

Building his own lab as an independent scientist at the University of Utah, Shepherd devised experiments to understand

how the protein of *Arc* works. Clearly it is involved in conveying signals from one nerve cell to the next, and that signal is important in memory and learning. He would find answers to his questions by purifying the protein and then analyzing its structure.

Purifying a protein involves a number of steps to strip away everything in a cell but the protein of interest. The process begins with chemically macerating the tissue—in this case, brains—into fluids, then treating them successively to isolate the desired protein from all the others present. The protein soup is run through a series of tubes with each pulling out different contaminants. In one of the final steps, the fluid is run through a glass column packed with a special gel. The gel removes the final contaminants and other proteins, and the fluid that makes it through contains only the purified protein. Shepherd went through each step, getting small amounts of liquid to process along the way. He poured the fluid into the last glass column and—nothing. Nothing came out of the column. He changed the gel to a fresh batch. Again nothing came through. Clearly something was clogging it. The team tried new columns, but the tubes were still clogged. They tinkered with concentrations of different fluids. The clogs remained.

Shepherd's lab technician had a hunch. Maybe there was something special about the *Arc* protein that clogged the columns. Instead of being an artifact, perhaps this was saying something about the structure of the *Arc* molecule itself. Shepherd and his assistant took the clogged fluids to an electron microscope, where they could see the structure of the proteins on a computer screen at ultrahigh magnification. The structure was so surprising that Shepherd exclaimed, upon seeing it, "What the hell is going on?"

Arc was forming hollow spheres, and these spheres were so big that they got stuck in the spaces inside the gel filter. He had

seen versions of these spheres before, in his premedical training. The structure of the spheres was identical to those made by some viruses as they move from cell to cell to infect them.

Shepherd works in the research wing of the University of Utah Medical Center, so he went across the building to visit a team that studies the virus that causes HIV. HIV moves from cell to cell by forming a protein capsule that conveys its genetic information. Showing the microscopic images to the virology team, Shepherd left it to the scientists to figure out what the curious spheres were. The HIV researchers thought they were from a virus like HIV. They couldn't find any difference between the *Arc* capsule and those made by the HIV virus. Both were made of four different chains of proteins, and both had the same molecular structure, even down to the atomic architecture of the bends and folds. Much like anatomists studying and naming bones, biochemists have their names for structures as well. A bend in the molecular structure known as zinc knuckle is one characteristic of HIV. *Arc* had that too.

It became clear that the *Arc* protein was virtually identical to viruses like HIV. And both molecules functioned in the exact same way—they conveyed small bits of genetic material from one cell to the next. Syncytin, as we have seen, is also HIV-like, albeit in different ways.

Working with geneticists, Shepherd's team mapped the structure of *Arc* DNA and scoured the genome databases of the animal kingdom for other creatures that have it. In tracing the structure and distribution of the gene, a story of ancient infections emerges. All land-living animals have the *Arc* gene; fish do not. This means that about 375 million years ago a virus entered the genome of the common ancestor of all land-living animals. I like to think that it was a close relative of *Tiktaalik* that got the first infection. Once the virus joined the host, it carried with it

the ability to make a special protein, a version of *Arc*. Normally the protein would be used to allow the virus to move from cell to cell and spread. But in this case, because of where it entered the fish's genome, it made that protein become active in brains and enhanced memories. The individuals with the virus were the recipients of a biological gift. The virus was hacked, neutered, and domesticated for a new function in brains. Our ability to read, write, and remember the moments of our lives is due to an ancient viral infection that happened when fish took their first steps on land.

Excited to present his results, Shepherd went to a conference on neuroscience and behavior. Before he spoke, he heard a scientist who works on fruit flies give her talk. She showed that the flies have *Arc*. Fly *Arc*, like ours, is active in spaces between neurons. Moreover, fly *Arc* forms hollow capsules that convey molecules from one nerve cell to the next. But fly *Arc* looks like a different virus from the one in land-living animals. Theirs came from a separate encounter with viruses.

How does a genome domesticate a virus and put it to work rather than allow it to infect? The answer is not clear, but there are many different ways this might happen. Think of the fate of both a virus and a host under a few different circumstances. If the virus is very infectious, the host will die and the virus will not pass from generation to generation. If the virus is relatively benign, or beneficial, it will enter the genome and reside there. If it makes it to the genome of a sperm or egg, the virus will pass its genome on to offspring. Over time, if the virus has a very beneficial effect, say by making creatures with more efficient placentas or better memories, natural selection can hone it to stay put and do its job ever more efficiently.

The genome is the stuff of B movies, like a graveyard filled with ghosts. Bits and pieces of ancient viral fragments lie every-

where—by some estimates, 8 percent of our genome is composed of dead viruses, more than a hundred thousand of them at last count. Some of these fossil viruses have kept a function, to make proteins useful in pregnancy, memory, and countless other activities discovered in the past five years. Others sit like corpses where they attached to the genome only to be extinguished and decay.

A struggle is going on inside genomes. Some bits of genetic material exist to make ever more copies of themselves. They can be foreign invaders, such as viruses that enter the genome to commandeer it. They can also be innate parts of our genome, such as jumping genes that proliferate and insert everywhere. Occasionally, when these selfish genetic elements land in a special place, they can be put to use to make new tissues, such as the endometrium, or to allow for new functions, such as memory and cognition. Genetic mutations can spread far and wide across the genome in a small number of generations. And if viruses occupy different species, similar genetic changes can arise in different kinds of creatures independently.

My Thursday teas with Mayr continued for another two years after the Goldschmidt faux pas. During those later meetings, I discovered that Mayr had grudging respect for Goldschmidt's attempt to unite experiments in genetics and developmental biology with the major events in the fossil record. By the mid-1980s, he knew that a revolution was coming from molecular biology and so encouraged the graduate students in his orbit to keep current in that area of research.

As Lillian Hellman might have said in this context, nothing ever begins when, *or where*, you think it did. Genomes are not static strands; they are ever twisting and turning while viruses

attack and other genes jump. Genetic mutations can spread across the genome and among different species. Changes to the genome can be rapid, similar genetic changes can happen independently in different creatures, and the genomes of different species can blend and merge to forge biological inventions.