



From Atoms to Traits

Charles Darwin saw that random variations in organisms provide fodder for evolution. Modern scientists are revealing how that diversity arises from changes to DNA and can add up to complex creatures or even cultures

• • • BY DAVID M. KINGSLEY



On a shelf in a library in Texas sits a small green volume, originally published 150 years ago and now generally recognized as one of the most important scientific books ever written. Its future success was not at all apparent when this first-edition copy of *On the Origin of Species* was printed, however. As Charles Darwin finished the proofs of his new work, he drew up a short list of important colleagues who should receive advance copies. He then anxiously awaited the verdicts of the leading thinkers of his time.

England's most famous living scientist in 1859 scribbled his reactions in notes found throughout that little green volume preserved at the University of Texas at Austin. Marked "from the author" on its frontispiece, it is the advance copy that Darwin sent to Sir John Herschel, one of his scientific heroes, whose own treatise on natural philosophy had first inspired Darwin to become a scientist. In the 1830s Herschel had memorably described the origin of species as a "mystery of mysteries" that might occur by natural processes. Darwin quoted Herschel's words in the very first paragraph of the book, which laid out the ingenious solution to the "mystery of mysteries" that Darwin was offering to both Herschel and the world.

Darwin's theory was at once sweeping and simple. He proposed that all living things on earth are descended from one or a few original forms. He did not presume to know how life it-

self first arose. Once life began, though, Darwin argued, organisms would slowly begin to change and diversify through a completely natural process: all living things vary; the differences are inherited. Those individuals with trait variants that are favorable in the environment they inhabit will thrive and produce more offspring than individuals with unfavorable variants. Advantageous traits will therefore accumulate over time by an inevitable process of "natural selection." To convince readers of the cumulative power of spontaneous variation and differential reproduction, Darwin pointed to the huge changes in size and form that had occurred in domesticated plants, pigeons and dogs after only a few centuries of selective breeding by humans.

Some of his scientific colleagues instantly saw the power of Darwin's argument. "How stupid of me not to have thought of that!" exclaimed Thomas Henry Huxley, after reading his own advance copy of Darwin's book. Unfortunately, the reaction of the man whose opinion Darwin said he valued "more than that of almost any other human being" was far less favorable. Herschel did not believe that useful new traits and species could arise from simple random variation, an idea he dismissed as the "law of higglety-pigglety." In his personal copy of *Origin of Species*, Herschel zeroed in on the fact that "favorable variations must 'occur' if anything is to be 'effected.'" Darwin actually knew nothing about

KEY CONCEPTS

- The idea that nature "selects" favorable variations in organisms was at the heart of Charles Darwin's theory of evolution, but how those variations arise was a mystery in Darwin's time.
- Random changes in DNA can give rise to changes in an organism's traits, providing a constant source of variation.
- Certain kinds of DNA changes can produce major differences in form and function, providing raw material for the evolution of new species and even new human cultures. —*The Editors*

the origin of the variant traits themselves, and Herschel felt that if Darwin could not explain the source of variation, he did not really have a theory sufficient to explain the origin of species.

In the 150 years since the debut of Darwin's theory, key questions about how traits are passed down to subsequent generations and how they undergo evolutionary change have been resolved by remarkable progress in the study of genes and genomes. Darwin's scientific descendants studying evolutionary biology today understand at least the basic molecular underpinnings of the beautiful diversity of plants and animals around us. Like Darwin's theory itself, the causes of variation are often simple, yet their effects are profound. And fittingly, these insights have come in a series of steps, many of them just in time for the successive 50-year anniversaries of Darwin's book.

Variation Revealed

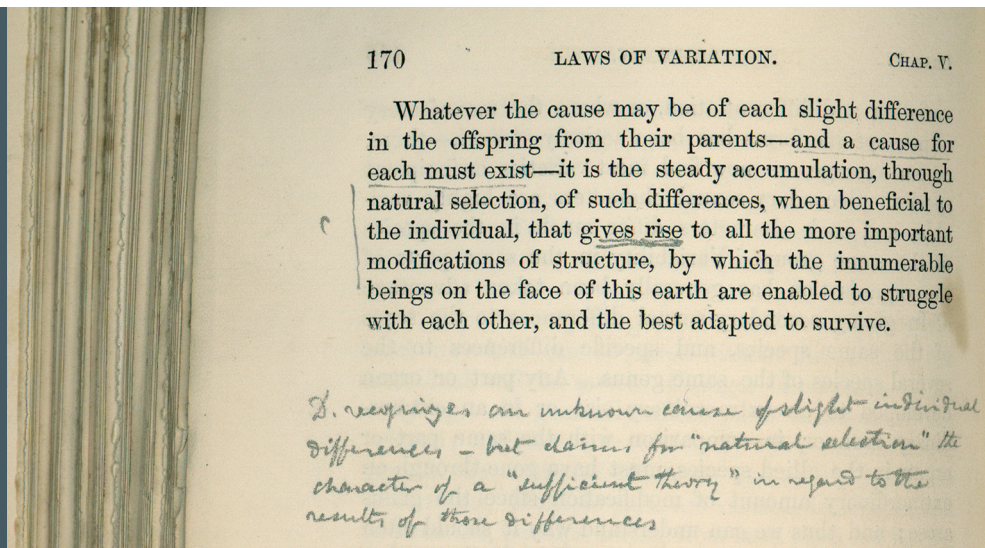
Darwin was not only unable to say where variants came from, he did not explain how those new traits could spread in subsequent generations. He believed in blending inheritance, the idea that offspring take on characteristics intermediate between their parents. But even Darwin recognized that the theory was problematic because if traits truly blended, then any rare new variant would be progressively diluted by generations of breeding with the great mass of individuals that did not share the trait.

Confusion about blending inheritance was swept away in 1900 by the rediscovery of Gregor

VARIATION in a pea plant, and in all life, stems from DNA.



SIR JOHN HERSCHEL, a prominent scientist of Charles Darwin's era, doubted the theory laid out in *On the Origin of Species*. Because Darwin could not explain the cause of trait variations, the idea that nature selected the advantageous variants seemed incomplete. In his personal copy of the book, Herschel wrote, "D. recognizes an unknown cause of slight individual differences—but claims for 'natural selection' the character of a 'sufficient theory' in regard to the results of those differences."



Whatever the cause may be of each slight difference in the offspring from their parents—and a cause for each must exist—it is the steady accumulation, through natural selection, of such differences, when beneficial to the individual, that gives rise to all the more important modifications of structure, by which the innumerable beings on the face of this earth are enabled to struggle with each other, and the best adapted to survive.

D. recognizes an unknown cause of slight individual differences - but claims for "natural selection" the character of a "sufficient theory" in regard to the results of those differences

Mendel's famous breeding experiments with peas, conducted in the 1850s and 1860s. Different pea plants in the Austrian monk's garden showed obvious morphological differences, such as tall versus short stems, wrinkled versus smooth seeds, and so forth. When true-breeding pea plants of contrasting types were crossed, the offspring usually resembled one of the two parents. With further crosses, both forms of a trait could reappear in undiluted form in future generations, however, demonstrating that the genetic information for alternative forms had not blended away. Mendel's experiments changed the general perception of heritable variants from ephemeral and blendable to discreet entities passed from parents to offspring, present even though they are not always visible.

Soon the inheritance patterns of Mendel's "genetic factors" were, intriguingly, found to be mirrored by the behavior of chromosomes in the cell nucleus. At the 50-year anniversary of *Origin of Species*, the origin of variants was still unknown, but genetic information was becoming a physical entity, and it was finally visible as threads inside the nucleus. By the 100th anniversary of the book's publication, hereditary information in chromosomes had already been traced to a large acidic polymer called deoxyribonucleic acid (DNA). James D. Watson and Francis Crick had proposed a structure for the DNA molecule in 1953, with stunning implications for our physical understanding of heredity and variation.

DNA is a long, two-stranded helix, with a backbone made of repetitive chains of sugar and phosphate. The two strands of the polymer are held together by the complementary pairing between four possible chemical bases: adenine, cy-

tosine, guanine and thymine (A, C, G, T), which also form the foundation of a simple genetic language. Just like the 26 letters in the English alphabet, the four chemical letters in the DNA alphabet can occur in any sequence along one strand of the helix, spelling out different instructions that are passed from parent to offspring.

The double-stranded helix provides a clear mechanism for copying genetic information as well. Cs always pair with Gs, and As pair with Ts across the middle of the DNA molecule, with these affinities determined by the complementary size, shape and bonding properties of the corresponding chemical groups. When the two strands of the DNA helix are separated, the sequence of letters in each strand can therefore be used as a template to rebuild the other strand.

Watson and Crick's DNA structure immediately suggested a possible physical basis for spontaneous variation. Physical damage or mistakes made in copying the DNA molecule prior to cell division might alter its normal sequence of letters. Mutations could take many different forms: substitution of a single letter for another at a particular position in the polymer, deletion of a block of letters, duplication or insertion of new letters, or inversion and translocation of the letters already present. Such changes were still theoretical at the time the structure was proposed. But as the 150th anniversary of Darwin's famous publication approaches, large-scale sequencing methods have made it possible to read entire genomes and to study genetic variation—the raw material for his proposed evolutionary process—with unprecedented detail.

By sequencing various organisms and their offspring, then looking for any spontaneous

THE AUTHOR



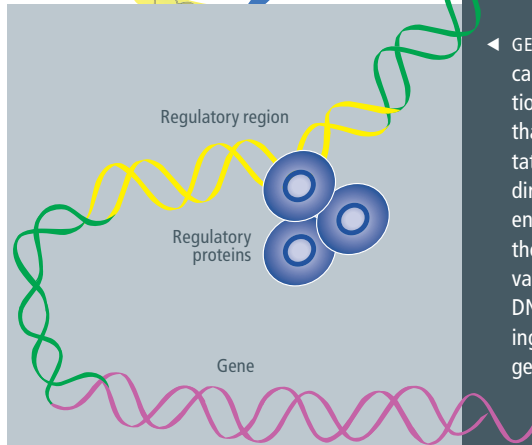
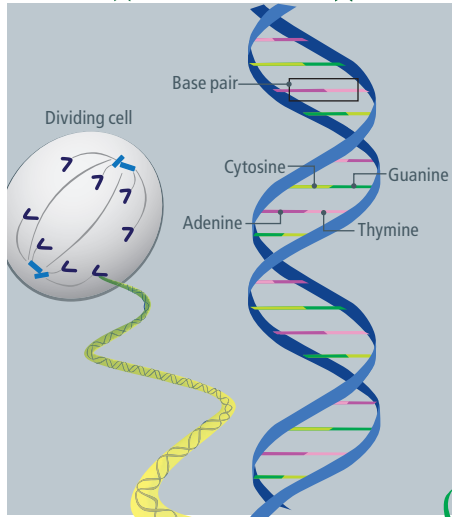
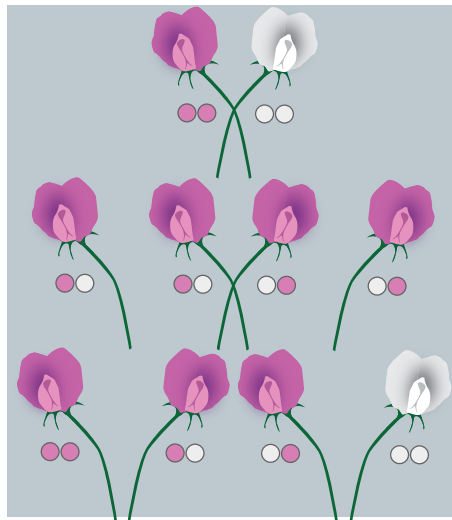
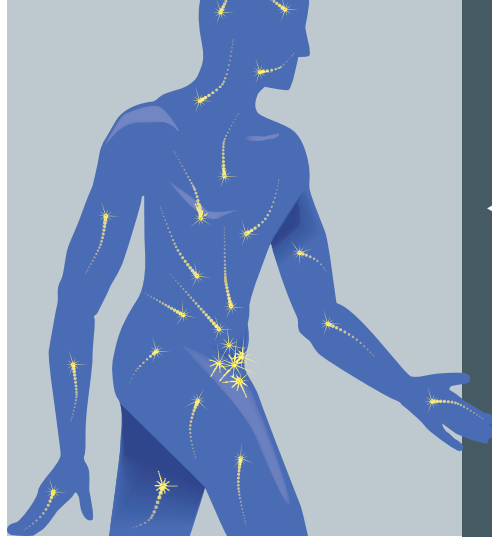
David M. Kingsley is professor of developmental biology at the Stanford University School of Medicine and a Howard Hughes Medical Institute investigator. His studies of how genes control the development and maintenance of skeletons and joints shed light on modern health problems as well as on the evolution of new animal forms over the course of millennia. Kingsley has shown how several basic genetic mechanisms work in diverse organisms to create new traits in natural populations.

changes in the long chain of DNA letters passed from generation to generation, scientists have clearly shown that such mutations do occur fairly regularly. (Of course, only mutations that occur in germ cells would be passed to offspring and therefore detectable in this manner.) Absolute rates of mutation differ in different species but typically average 10^{-8} per nucleotide per generation for single base-pair substitutions. That frequency may sound low, but many plants and animals have very large genomes. In a multicellular animal with 100 million or even 10 billion base pairs in its genome, some spontaneous single base-pair changes are likely to occur every time hereditary information is passed down.

Particular types of substitutions are more likely than others, based on the chemical stability and structural properties of the DNA bases. In addition, some types of larger sequence changes occur much more frequently than the overall average rate of single base-pair substitutions. Stretches of DNA with eight or more identical letters in a row, known as homopolymers, are very prone to copying errors during the process of DNA replication, for example. So are regions known as microsatellites that consist of sequences of two, three or more nucleotides repeated over and over.

All these spontaneous changes within genomes add up to a lot of diversity, even within a single species, including our own. In a historic milestone, a reference sequence for the entire three-billion-base-pair human genome was completed in 2003, and four years later the nearly complete personal genome of Watson was published, making it possible to compare the two human sequences to each other and to that of Celera founder Craig Venter, whose genome sequence has also been made public. A side-by-side comparison of the three sequences offers several interesting revelations.

First, each individual's genome differs from the reference sequence by roughly 3.3 million single base-pair changes, which corresponds to variation in one of every 1,000 bases on average. Although deletions and insertions of larger DNA stretches and whole genes are not as frequent as single base-pair changes (a few hundred thousand instead of a few million events per genome), these events account for the majority of total bases that differ between genomes, with up to 15 million base pairs affected. Many entire genome *regions* have also recently been found to exist in different copy numbers between individuals, which reflects an unappreciated level of genome



Seeking Variation

◀ **GEMMULES:** Nine years after *Origin of Species*, Darwin put forth his theory that a new variant version of a trait might be passed from parent to offspring, and thus into the larger population, by “infinitely minute” particles that he called gemmules. Secreted by cells, the particles would carry the essence of the body parts from which they derived to the reproductive organs, to be absorbed by the germ cells.

◀ **MENDEL'S FACTORS:** Early 20th-century scientists rediscovered the ideas of Gregor Mendel, who experimented with pea plants during the 1850s and 1860s to derive detailed laws of inheritance. Mendel posited the existence of discrete factors carrying trait information and observed that each individual would carry two copies—one from each parent—of a given factor. Although both were present, only one of the copies would dominate and produce the visible trait.

◀ **THE DOUBLE HELIX:** The DNA molecule was already recognized as the vehicle for trait information when Francis Crick and James D. Watson discovered its structure in 1953. The paired strands joined by complementary chemical bases immediately suggested both an alphabet to convey the genetic message and a mechanism for it to change. Each time a cell divides, it makes a copy of its chromosomes, providing an opportunity for “typos” to be introduced into the sequence of bases.

◀ **GENE REGULATION;** Mendel's factors came to be known as genes, traditionally defined as stretches of DNA that encode a protein. Typos, or mutations, can alter or disable genes directly, but in the past decade scientists have also come to appreciate the importance of another source of variation: mutations that alter a DNA region responsible for regulating when and where in the body a gene is activated.

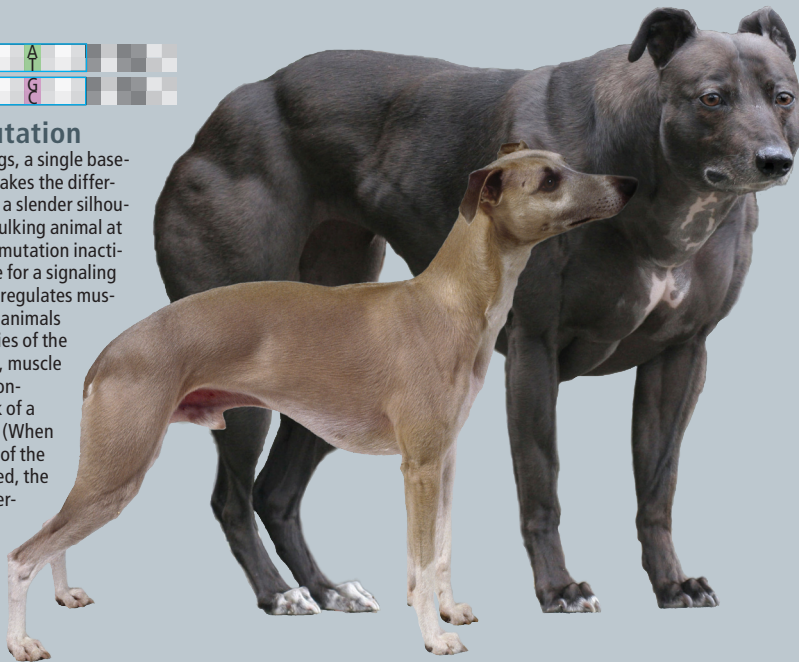
○ ○ ○ The Origin of Variation

Random changes to an organism's DNA can produce trait variations that are subtle or dramatic. The nature of the DNA mutation can also range from a simple substitution of one base pair for another to the duplication of entire genes or chromosomal regions. Examples below illustrate many of the ways spontaneous DNA alterations can give rise to diversity.



Point Mutation

In whippet dogs, a single base-pair change makes the difference between a slender silhouette and the hulking animal at the right. The mutation inactivates the gene for a signaling molecule that regulates muscle growth. In animals with both copies of the gene mutated, muscle growth is uncontrolled for lack of a "stop" signal. (When only one copy of the gene is disabled, the dogs are moderately more muscular and prized as racers.)



Insertion

In pea plants, an 800-base-pair sequence inserted into a gene produces peas that are wrinkled rather than smooth. The intruding DNA element disables a gene necessary for starch synthesis, altering the peas' sugar and water content. Such mobile elements are seen in the genomes of most multicellular organisms, including humans.



Gene Copy Number

Entire genes can be duplicated by copying errors during cell division, leading to differences between species and to variation among members of the same species. The genome of chimpanzees, which eat mostly green plants, normally contains just a single gene for the starch-digesting enzyme salivary amylase, whereas humans can carry up to 10 copies of the gene.

structural variation whose implications scientists are only beginning to explore. Finally, the sequence changes seen when comparing complete human genomes alter either the protein-encoding or regulatory information or the copy number of a substantial proportion of all 23,000 human genes, providing an abundant source of possible variation underlying many traits that differ between people.

The Molecular Basis of Traits

Herschel wanted an answer for how and why variants arose before he could accept Darwin's theory that natural selection acts on those traits, generating new living forms by completely natural processes. Today scientists know that spontaneous changes in DNA are the simple "why" of variation, but the answer to "how" those mutations translate into trait differences is more complex and makes for an active field of research with implications far beyond evolution studies.

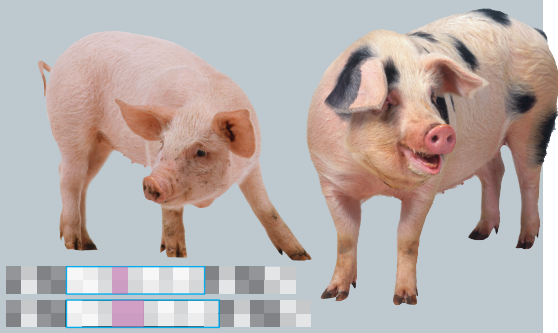
Biologists can now often connect the dots all the way from classic morphological and physiological traits in plants and animals to specific changes in the atoms of the DNA double helix.

They know, for example, that Mendel's tall and short pea plants differ by a single G to A substitution in a gene for the enzyme gibberellin oxidase. The so-called short variant of the gene changes a single amino acid in the enzyme, which reduces enzyme activity and causes a 95 percent drop in the production of a growth-stimulating hormone in the stems of the pea plants.

In contrast, Mendel's wrinkled seed trait results from the insertion of an 800-base-pair sequence in a gene for a starch-related enzyme. That inserted sequence interferes with the enzyme's production, reducing starch synthesis and producing changes in sugar and water content that lead to sweeter but wrinkly seeds. The inserted sequence also appears at multiple other locations in the pea genome, and it has all the hallmarks of a transposable element—a block of DNA code that can move from one place in the genome to another. Such "jumping" elements within genomes may be yet another common source of new genetic variants—either by inactivating genes or by creating new regulatory sequences that change gene activity patterns.

One of the few generalizations evolutionary

WALLY EBERHART (peas); ZUMA LIVE/NEWS.COM (mirrored whippet); TRACY MORGAN (whippet); FRANS LANTING/Corbis (chimp); ALTRENDU IMAGES (boy)



Duplication

Sequences containing the same base pair repeated eight or more times, known as homopolymers, are highly prone to copying errors. In pigs, the gain of two additional C-G pairs in such a sequence inactivates a gene for a signal receptor in pigment cells, producing light-colored coats. Copying mistakes within individual cells can also cause the duplicated sequence to lose bases, restoring the gene's function and producing dark patches on the body.



Regulatory Changes

Mutations in the DNA that controls when and where genes are activated can produce profound trait changes by altering the formation of entire body parts during the organism's development. Changes in the regulatory regions of a single gene that controls patterns of cell division during stem development account for much of the shape difference between the bushy teosinte plant (top) and its descendant, the tall modern cornstalk.



differences such as these between individuals might arise by natural processes, but bigger structural differences between species could not have done so. Many small changes can add up to big ones, however. In addition, certain genes have powerful effects on cell proliferation and cell differentiation during embryonic development, and changes in those control genes can produce dramatic changes in the size, shape and number of body parts. A subspecialty within evolutionary biology that has come to be known as evo-devo concentrates on studying the effects of changes in important developmental genes and the role they play in evolution.

The potent influence of such genes is illustrated by the modern maize plant, which looks completely different from a wild, weedy ancestor called teosinte in Central America. Many of the major structural differences between maize and teosinte map to a few key chromosome regions. Mutations in a regulatory area of a single gene that controls patterns of cell division during plant stem development account for much of the difference between an overall bush shape and a single, central stalk. Changes in a second gene that is active during seed development help to transform the stony, mineral-encased seeds of teosinte into the softer, more exposed kernels of maize. Ancient Mesoamerican farmers developed maize from teosinte without any direct knowledge of DNA, genetics or development, of course. But by mating plants with desirable properties, they unwittingly selected spontaneous variants in key developmental control genes and thereby converted a bushy weed into a completely different looking plant that is useful for human agriculture in relatively few steps.

Similar principles underlie the evolution of new body forms in completely wild populations of stickleback fish. When the last Ice Age ended 10,000 years ago, migratory populations of ocean fish colonized countless newly formed lakes and streams in North America, Europe and Asia. These populations have since had approximately 10,000 generations to adapt to the new food sources, new predators, and new water colors, temperatures and salt concentrations found in the freshwater environments. Today many freshwater stickleback species show structural differences that are greater than those seen between different genera of fish, including 30-fold changes in the number or size of their bony plates, the presence or absence of entire fins, and major changes in jaw and body shape, tooth structures, defensive spines and body color.

STICKLEBACKS ADAPT

In just 10,000 generations, three-spined stickleback fish have evolved myriad forms to suit diverse environments. Mutations affecting the activity of three developmental-control genes have produced striking anatomical changes, including the complete loss of pelvic hind fins, large differences in bony armor and much lighter skin color. In each fish pair shown below, a typical marine ancestor is on top and an evolved freshwater stickleback is underneath.



biologists can make about the nature of variation is that one usually cannot tell just by looking what the underlying genetic source of a trait variant is going to be. Darwin wrote extensively about dramatic morphological differences present in pigeons, dogs and other domesticated animals, for example. Today we know that the interesting traits in domesticated animals are based on many different types of DNA sequence change.

The difference between black and yellow color in Labrador retrievers stems, for instance, from a single base change that inactivates a signal receptor in the pigment cells of yellow dogs. Increased muscle size and improved racing performance in whippet dogs have also been traced to a single base-pair change, which inactivates a signal that normally suppresses muscle growth. In contrast, the special dorsal stripe of hair in Rhodesian ridgeback dogs comes from the duplication of a 133,000-base-pair region containing three genes that encode a growth factor for fibroblast cells, which amps up production of the growth factor.

Modern-day critics of Darwin and evolutionary theory have often suggested that small

BILL LING (pig); PETER ANDERSON (spotted pig); JERRY FRIEDMAN (teosinte); PAUL DAMIEN (National Geographic, cornstalks); MICHAEL SHAPIRO, PAMELA COLOSIMO, FRANK CHAN, CRAIG MILLER AND DAVID M. KINGSLEY (Stanford University) (sticklebacks)



Skin Color

Changes to at least three genes have been linked to the trait of light-colored skin, which is believed to have arisen in northern Europe. Light skin more easily absorbs ultraviolet rays needed to produce vitamin D, an advantage where sunlight is limited.



Starch Digestion

People vary widely in the number of copies they carry of a gene for the starch-digesting enzyme salivary amylase. Members of cultures with starch-rich diets tend to have higher numbers of the amylase gene and high levels of the enzyme in their saliva.

PATRICK GIARDINO/Corbis (women); TRAVEL INK (man with rice pot)

readily makes vitamin D from limited sunlight.

Similarly, strong molecular “signatures of selection” have been found around a gene that controls the ability to digest lactose, the predominant sugar in milk. Humans are mammals, nurse their young and produce an intestinal enzyme that breaks lactose into the simpler sugars glucose and galactose. Humans are also unique among mammals in continuing to use the milk of other animals as a significant source of nutrition well beyond childhood. This cultural innovation has occurred independently in groups in Europe, Africa and the Middle East, using milk derived from cattle, goats and camels.

An ability to digest milk in adulthood depends on a mutant form of the intestinal lactase gene, which in most mammals and most human groups, is active only during the infant nursing period. In humans from populations with a long history of dairy herding, however, a mutant form of the lactase gene continues to be active in adulthood. This genetic innovation has been linked to single base-pair changes in the regula-

tory DNA regions that control the gene, but different lactose-tolerant populations have different mutations in the key region—a striking example of the repeated evolution of a similar trait by independent changes affecting one gene.

Another example of a recent nutrition-related adaptation in humans involves the multiplication of a complete gene. Whereas chimpanzees have only one copy of the gene for salivary amylase, an enzyme that digests starch in food, humans show marked variation in the number of amylase gene copies they carry. In some individuals, duplications of the gene have produced as many as 10 copies along a single chromosome. People from cultures that eat diets rich in starch, such as rice, have higher average amylase gene copy numbers and higher amylase enzyme levels in their saliva than do people from cultures that rely on hunting and fishing.

Dairy herding and agriculture both arose in the past 10,000 years. Although that only corresponds to just 400 or so human generations, major new sources of nutrition are clearly already leading to the accumulation of novel genetic variants in populations that exploit those food sources.

Herschel’s most persistent objection to Darwin’s theory was his feeling that useful new traits could never appear from simple random variation. In published comments and letters, he argued that such characteristics would always require “mind, plan, design, to the plain and obvious exclusion of the haphazard view of the subject and the casual concourse of atoms.” Herschel was correct to point out that the origin of variation was still a mystery in 1859. After 150 years of additional research, however, we can now catalogue a variety of spontaneous DNA sequence variants that occur every time a complex genome is passed from parents to offspring.

Only a tiny fraction of these changes are likely to improve, rather than degrade, the original hereditary information and the trait that derives from it. Nevertheless, sweeter peas, bigger muscles, faster running ability or improved ability to digest new foods have all arisen from simple new arrangements of atoms in the DNA sequence of peas, dogs and humans. Thus, the “casual concourse of atoms” clearly can generate interesting new traits. And the intrinsic variability of living organisms continues to provide the raw material by which, in Darwin’s famous words at the end of his small green book, “endless forms most beautiful and most wonderful, have been, and are being evolved.” ■

MORE TO EXPLORE

cis-Regulatory Changes in *Kit Ligand* Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans. Craig T. Miller et al. in *Cell*, Vol. 131, No. 6, pages 1179–1189; December 14, 2007.

Independent Introduction of Two Lactase-Persistence Alleles into Human Populations Reflects Different History of Adaptation to Milk Culture. Nabil S. Enattah et al. in *American Journal of Human Genetics*, Vol. 82, No. 1, pages 57–72; 2008.

Evolution: Constant Change and Common Threads. HHMI 2005 Holiday Lectures on Science. Sean B. Carroll and David M. Kingsley. Webcast or DVD available at www.hhmi.org/biointeractive/evolution