Ancient DNA Tells Story of Giant Eagle Evolution

DOI: 10.1371/journal.pbio.0030020

The recent discovery of a Hobbit-like hominid on the Indonesian island of Flores was startling in some respects—its rather modern existence, for one—but it represents a classic case of Darwinian evolution. For reasons that are not entirely clear, when animals make their way to isolated islands, they tend to evolve relatively quickly toward an outsized or pint-sized version of their mainland counterpart. Following this evolutionary script, the Flores woman, presumably a downsized version of *Homo erectus*, appears to have shared her island home with dwarf elephants and giant rats.

Perhaps the most famous example of an island giant—and, sadly, of species extinction—is the dodo, once found on the Indian Ocean island of Mauritius. When the dodo's ancestor (thought to be a migratory pigeon) settled on this island with abundant food, no competition from terrestrial mammals, and no predators, it could survive without flying, and thus was freed from the energetic and size constraints of flight. New Zealand also had avian giants, now extinct, including the flightless moa, an ostrich-like bird, and Haast's eagle (*Harpagornis moorei*), which had a wingspan up to 3 meters. Though Haast's eagle could fly—and presumably used its wings to launch brutal attacks on the hapless moa—its body mass (10–14 kilograms) pushed the limits for self-propelled flight.

As extreme evolutionary examples, these island birds can offer insights into the forces and events shaping evolutionary change. In a new study, Michael Bunce et al. compared ancient mitochondrial DNA extracted from Haast's eagle bones with DNA sequences of 16 living eagle species to better characterize the evolutionary history of the extinct giant raptor. Their results suggest the extinct raptor underwent a rapid evolutionary transformation that belies its kinship to some of the world's smallest eagle species.

The authors characterized the rates of sequence evolution within mitochondrial DNA to establish the evolutionary relationships between the different eagle species. Their analysis places Haast's eagle in the same evolutionary lineage as a group of small eagle species in the genus *Hieraaetus*. Surprisingly, the genetic distance separating the giant eagle and its more diminutive *Hieraaetus* cousins from their last common ancestor is relatively small.

DOI: 10.1371/journal.pbio.0030020.g001

Giant Haast's eagle attacking New Zealand moa (Art: John Megahan)

Without the fossils to directly determine divergence times, Bunce et al. relied on molecular dating techniques that use the rate of sequence evolution in the genes studied to establish the relative evolutionary ages of the eagles. Proposing a divergence date of roughly 0.7–1.8 million years ago, the authors acknowledge that while this is the "best available approximation of the 'true' date," additional molecular data could help refine the estimate. Whatever the date of divergence, the extinct giant eagle is clearly an anomaly among the eagles studied here. The increase in body size—by at least an order of magnitude in less than 2 million years—is particularly remarkable, Bunce et al. argue, since it occurred in a species still capable of flight.

The absence of mammalian competitors facilitated the evolution of much larger eagles and owls on Cuba and may have likewise precipitated the rapid morphological shift seen here. Haast's eagle, the authors write, "represents an extreme example of how freedom from competition on island ecosystems can rapidly influence morphological adaptation and speciation." Given its similarity to the smaller *Hieraaetus* species, the authors recommend reclassifying the New Zealand giant as *Hieraaetus moorei*. This study shows how quickly morphological changes can occur in vertebrate lineages within island ecosystems. Could it be that anthropologists might some day uncover evidence of a giant version of the Flores woman?

Bunce M, Szulkin M, Lerner HRL, Barnes I, Shapiro B, et al. (2005) Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle. DOI: 10.1371/journal.pbio.0030009

Tracking Orangutans from the Sky

DOI: 10.1371/journal.pbio.0030022

After apprenticing at her mother's side for some eight years—the first three clinging to her body—an orangutan is ready to make her own way in the forest canopy. The only great ape specializing in arboreal living, orangutans forage the treetops mostly for fruit, nuts, insects, leaves, and tree bark. They can recognize hundreds of species of edible fruit from trees and woody climbers and remember their location and fruiting season. Malay legend says orangutans (a variation on the Malay trope for "man of the woods") derived from humans who sought refuge from their species in the wilds of the forest. Today, with the last orangutan refuges shrinking drastically, the man of the woods has nowhere else to go.

Hundreds of thousands of orangutans once ranged throughout southeast Asia. Now just two orangutan species inhabit just two countries: Indonesia (Kalimantan, the southern part of Borneo and Sumatra Island) and Malaysia (Sabah and Sarawak, in northern Borneo). The Sumatran orangutan is listed as critically endangered; the Bornean, endangered. In Indonesian Borneo and Sumatra, logging operations clear an estimated 5 to 6 million forest acres a year, leaving the apes stranded in isolated stands of trees and the normally fire-resistant rainforest at sudden risk. Another force driving orangutan extinction in Indonesia is the poaching and illegal killing (mothers do not give up their babies without a fight) that secures orangutan babies for the exotic pet trade. In Sabah, Malaysia, the primary threat comes from clearing the forest for agriculture.

Conservation efforts depend, among other things, on having reliable data on population size, density, and distribution, but estimates of orangutan numbers in Sabah—which range from 2,000 to 20,000—are outdated. In a new study, Marc Ancrenaz and colleagues report an innovative method of directly estimating orangutan numbers from the number of nests detected during aerial surveys. (Orangutans are tough to spot directly, so researchers count the nests they sleep in at night.) Their survey, which covered the entire orangutan range throughout Sabah, estimates a total population of 11,000 orangutans—a drop of 35% in the past 20 years, based on a 1988 World Wildlife Federation report.

Counting orangutans from the ground can be very time-consuming, difficult work, especially when faced with the hip-deep muck and steep slopes of the rainforest floor. Though helicopters obviously cover greater distance and more remote territory than is possible by foot, they're generally used to survey animals in more open landscapes. By using ground survey data to refine their aerial survey results, Ancrenaz and colleagues could directly assess the distribution and size of orangutan populations throughout Sabah. (Sabah covers roughly 72,000 square kilometers.)

Over the course of two years, ground surveys—requiring 1,100 hours of field work—and aerial surveys—requiring just 72 hours—were conducted throughout all the major forests of Sabah. Commercial logging occurs in about 76% of all Sabah forests in commercial forest reserves. During the overflights, information was recorded on altitude, forest type, forest disturbance (on a scale from none to active exploitation), and signs of human activity.

The authors attribute the 35% decline in Sabah orangutan numbers primarily to habitat loss from agricultural development. The surveys revealed that lowland forests harbored the greatest density of nests and orangutans, with the densest populations found in several highly disturbed, fragmented forests along newly created palm oil plantations. These high orangutan densities could reflect an influx of refugees from recently destroyed forest habitat into areas that are still forested. In logged forests, higher population densities were found in old exploited or sustainably logged forests than in conventionally logged reserves.

While the authors acknowledge the density estimates could be more precise—better measures of nest decay and construction rates are needed—their survey reveals crucial information on orangutan numbers and distribution. Most orangutans in Sabah, including those making up one of the largest unfragmented populations in Borneo, live outside protected areas, in commercially exploited forests. These results suggest that orangutans may adapt better to degraded forests than previously thought—provided illegal hunting and agricultural conversion are controlled.

More field research will help quantify the impacts of human activity—from logging to stealing babies—on great ape ecology and survival, and determine whether exploited forests can support conservation. It may be, for example, that integrating agricultural fields with forested corridors could sustain orangutan populations over the long term. With time of the essence, these aerial surveys will speed that work, and help sustain orangutans' refuge in the treetops, above their human relatives.

Ancrenaz M, Gimenez O, Ambu L, Ancrenaz K, Andau P, et al. (2004) Aerial surveys give new estimates for orangutans in Sabah, Malaysia. DOI: 10.1371/journal.pbio.0030003

The Evolution of Self-Fertile Hermaphroditism: The Fog Is Clearing

DOI: 10.1371/journal.pbio.0030030

The nematode *Caenorhabditis elegans* is a little less lonely than the rest of us—it is a self-fertile hermaphrodite, which as a larva makes and stores sperm before switching to egg production for the remainder of its lifespan. (*C. elegans* also maintains some males at a low frequency, about 1 in 500, and the hermaphrodite's eggs can be fertilized by sperm either from males or themselves.) A sister species, *C. briggsae*, is also hermaphroditic, but phylogenetic evidence suggests the last common ancestor of the two species had a female/male mode of reproduction. This raises the question of how the sex determination mechanisms, which must have evolved independently, differ between the two species. In this issue, Sudhir Nayak, Johnathan Goree, and Tim Schedl show that a crucial difference lies in the activities of two genes.

In *C. elegans*, the early period of sperm production is controlled by multiple proteins, two of which are the focus of this study, the RNA-binding protein GLD-1 (encoded by the gene *gld-1*) and the F-box-containing protein FOG-2 (encoded by the gene *fog-2*). Together, they repress translation of a gene, *tra-2*, by binding to its messenger RNA. This allows another gene, *fem-3*, to transiently masculinize the larval germline to produce sperm.

Comparing the genomes of *C. elegans* and *C. briggsae*, Schedl and colleagues found they share 30 out of 31 sex

DOI: 10.1371/journal.pbio.0030030.g001

Wild-type *C. elegans* hermaphrodite stained to highlight the nuclei of all cells

determination genes, but not *fog-2*. More surprisingly, they found that the role of *gld-1* in sex determination is opposite in the two species. When *C. elegans* is deprived of *gld-1*, would-be hermaphrodites produce only oocytes. But when *C. briggsae* is deprived of *gld-1*, would-be hermaphrodites produce only sperm. Thus, the authors conclude, the control of hermaphrodite spermatogenesis is fundamentally different in the two species.

By further examining the *C. elegans* genome, the authors showed that *fog-2* arose from a gene duplication event after the *C. elegans*–*C. briggsae* split, which occurred approximately 100 million years ago. Since then, its final exon, which codes for the C-terminal end of the protein, has undergone rapid evolution. The authors also show that this is the "business end" of the protein for its interaction with GLD-1, suggesting that the divergence of *C. elegans* and *C. briggsae* sex determination pathways resulted, in part, from FOG-2's new interaction with GLD-1.

Exactly what the role of *fog-2* is in *C. elegans* is still unclear. The authors speculate that it may recruit additional factors onto the *gld-1/tra-2* mRNA complex, increasing efficiency of translation repression. Much remains to be discovered about *C. briggsae* sex determination as well. The authors suggest that additional genetic differences promoting self-fertility are likely to have accumulated since the two species diverged, which may act to strengthen the male–female germline switching signal. Investigation of this possibility may shed more light on how hermaphroditism operates in these two species, and how a developmental pathway controlling sex determination can evolve.

Nayak S, Goree J, Schedl T (2004) *fog-2* and the evolution of self-fertile hermaphroditism in *Caenorhabditis*. DOI: 10.1371/journal.pbio.0030006

Multiple *Campylobacter* Genomes Sequenced

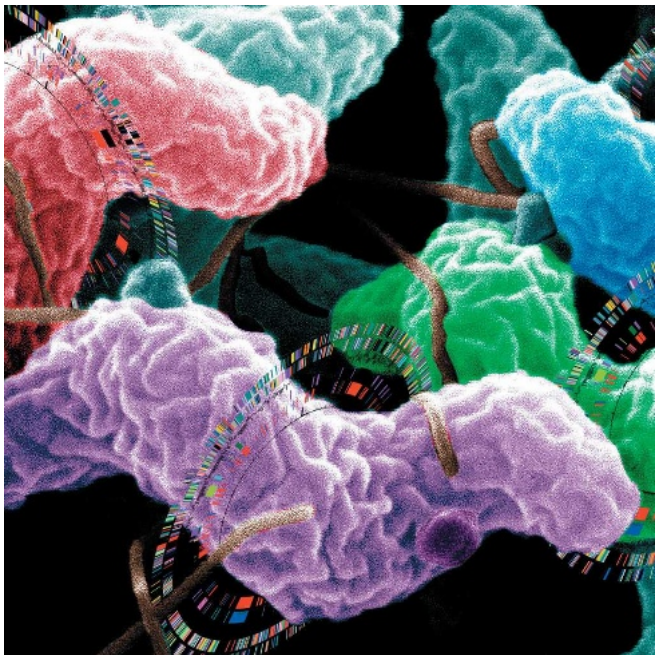
DOI: 10.1371/journal.pbio.0030040

In 1995, the first complete bacterial genome sequence was published. Now, nearly 200 bacterial genomes have been completed, and a new one hits the scientific press most weeks. This burgeoning industry is not just scientific "stamp collecting," however. Having all these genome sequences may provide useful clues about why some bacteria cause human disease, how to control their spread, and how to treat the infections caused by them. By comparing genome sequences, scientists can learn much more about what makes a bacteria tick than they can learn from a single sequence.

Derrick Fouts and his colleagues have taken this comparative approach with *Campylobacter*. Infection with a *Campylobacter* species is one of the most common causes of human bacterial gastroenteritis. In the US, 15 out of every 100,000 people are diagnosed with campylobacteriosis every year, and with many cases going unreported, up to 0.5% of the general population may unknowingly harbor *Campylobacter* in their gut annually. Diarrhea, cramps, abdominal pain, and fever develop within 2–5 days of picking up a pathogenic *Campylobacter* species, and in most people, the illness lasts for 7–10 days. But the infection can sometimes be fatal, and some individuals develop Guillain-Barré syndrome, in which the nerves that join the spinal cord and brain to the rest of the body are damaged, sometimes permanently.

Campylobacteriosis is usually caused by *C. jejuni*, a spiral-shaped bacterium normally found in cattle, swine, and birds, where it causes no problems. But the illness can also be caused by *C. coli* (also found in cattle, swine, and birds), *C. upsaliensis* (found in cats and dogs), and *C. lari* (present in seabirds in particular). Disease-causing bacteria generally get into people via contaminated food, often undercooked or poorly handled poultry, although contact with contaminated water, livestock, or household pets can also cause disease.

In 2000, *C. jejuni* was the first food-borne pathogen to be completely sequenced, but we still know little about how



DOI: 10.1371/journal.pbio.0030040.g001

Genome sequencing and comparison of four species of *Campylobacter*

Campylobacter species cause disease. In their search for clues, Derrick Fouts and coworkers have completely sequenced the genome of *C. jejuni* strain RM1221 (isolated from a chicken carcass) and compared it with the previously sequenced *C. jejuni* strain NCTC 11168 and with the unfinished sequences of *C. coli* strain RM2228 (a multi-drug-resistant chicken isolate), *C. lari* strain RM2100 (a clinical isolate), and *C. upsaliensis* strain RM3195 (taken from a patient with Guillain-Barré syndrome).

The researchers describe numerous differences and

similarities between these different *Campylobacter* strains and species. For example, there are major structural differences between the genomes caused by the insertion of new stretches of DNA. Some of these pieces of DNA may carry genes that improve bacterial virulence or fitness, so their presence could help to explain the different biological behaviors of these strains. There are also major variations in the genes responsible for synthesis of molecules that are important for the interaction of *Campylobacter* with the environment. Such differences could underlie the host specificity of the different species.

Differences between the *Campylobacter* species in genes that are likely to be involved in aspects of bacterial virulence, such as adherence, motility, and toxin formation, are all detailed by Fouts et al., who also describe a new putative *Campylobacter* virulence locus. Further work is needed to relate these genomic differences to functional differences, but this detailed comparative genomic analysis provides the core blueprint for this important family of human pathogens. And in doing so, it lays the foundation for the development of new ways to monitor and control *Campylobacter* in the food chain and in human infection.

Fouts DE, Mongodin EF, Mandrell RE, Miller WG, Rasko DA, et al. (2005) Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. DOI: 10.1371/journal.pbio.0030015

Genome Sequencing: Using Models to Predict Who's Next

DOI: 10.1371/journal.pbio.0030025

It's hard to believe it was just ten years ago that scientists reported the first complete genome sequence of an organism, the bacterial pathogen *Haemophilus influenzae*. The list has grown considerably since then: add over 160 bacterial species (and counting), most major model organisms, and an ever-growing list of mammals—including, of course, humans. With 99% of our genome now fully sequenced, the Human Genome Project's next major goal is to identify all the functional elements contained in our 2.85 billion nucleotides. Such an effort is hardly trivial: producing the sequence of a mammalian-size genome can run from $10 to $50 million, the estimated price tag of the Cow Genome Project.

In an ideal world, any organism would be fair game for sequencing, but in the real world, sequencing resources are scarce. Comparing genome sequences turns out to be a great way to identify regions that have important functions, but comparative genomics studies would be far more efficient if scientists could figure out in advance which genomes would reveal the most information about a particular question. Taking up that challenge, computational biologist Sean Eddy reports a statistical model that predicts how many genomes, and at what evolutionary distance, are needed for effective comparative genomic analyses. In addition to confirming some working principles of comparative genomics, the model also reveals a surprisingly simple guideline for future studies.

Comparative genomics works by aligning sequences of different organisms to identify patterns that operate over both large and small distances. Aligning mouse chromosomes with human chromosomes, for example, shows that 99% of our protein-coding genes align with homologous sequences in mice. Underlying such analyses is the principle that DNA sequences that are highly conserved are likely to be functionally important. A common assumption is that adding more comparative genomes to the alignment helps distinguish functionally significant from irrelevant conserved sequences.

How do you go about creating an abstract model that captures what Eddy calls the "essential flavor of comparative genomic analysis"? His model puts aside the specific characteristics of individual organisms, genomic features, and analysis programs in favor of identifying higher-level patterns and scaling relationships, specifically between the number of genomes, evolutionary distance, and feature size (features include genetic elements like exons and transcription factors).

The model shows that the number of genomes required to identify conserved regions—that is, regions evolving under selection—scales inversely with the size of the feature being sought. Thus, to look for conserved sequences half as long, you need twice as many genomes, assuming a constant evolutionary distance and statistical power. For example, to identify a conserved human feature the size of a coding exon (about 50 nucleotides), it is sufficient to compare just the human and mouse genomes. But to identify conserved single nucleotides, you would need 55 comparative genomes at "mouse-like" evolutionary distances (roughly 75 million years).

Things get a little trickier when varying evolutionary distance. We can see a substitution only at a given point in time: we can't tell how many times a site has changed, for example, or whether it changed at some point and then changed back. But at short evolutionary distances—where it's safer to assume no sites have changed more than once—the evolutionary distance is roughly the same as the fraction of sites identified as changed, and evolutionary distance and the number of genomes needed scale inversely. Therefore, the closer the evolutionary distance, the more genomes needed: one would need seven times as many comparative genomes using human/baboon distances, for example, compared to human/mouse distances. So when it comes to using primate sequences to study the human genome, our most distant relatives (such as lemurs) offer far more comparative analysis power than our next of kin (chimps and bonobos).

While this model confirms the intuitive assumption that identifying smaller features requires more genomes, it reveals an inverse scaling relationship far more direct, and precise, than previously imagined. With the next phase of the Human Genome Project under way, Eddy's model offers valuable guidelines for identifying which genomes and how many might best meet this ambitious goal.

Sean R. Eddy (2005) A model of the statistical power of comparative genome sequence analysis. DOI: 10.1371/journal.pbio.0030010

Evolution of Sex Chromosomes: The Case of the White Campion

DOI: 10.1371/journal.pbio.0030028

There are many different sex-determining systems in plants and animals with separate sexes (dioecious species). In some species, environmental factors activate sex-determining genes that trigger expression of genes leading to male or female development. Other species have evolved specialized sex chromosomes. In the well-known X-Y system of mammals, individuals inheriting a Y chromosome become males, and XX individuals become females.

Sex chromosomes have arisen independently in many taxonomic groups. It is an interesting question whether the same mechanisms were involved each time. Similarities in sex chromosome evolution have been reported between birds and mammals (although in birds, females are the heterozygous sex). In a new study, Michael Nicolas and colleagues uncover striking parallels in the details of sex chromosome evolution between mammals and a far more distant group: plants.

Sex chromosomes are an oddity in flowering plants. They are limited to dioecious species, a subset of plants that carry male and female organs (stamens and carpels, respectively) on separate individuals (most flowering plants are hermaphrodites). The genus *Silene*, which includes the White Campion, includes both dioecious and hermaphrodite species. The authors focus on three dioecious species, *Silene dioica*, *S. latifolia*, and *S. diclinis*, which share an X-Y sex-determination system where Y specifies maleness.

The theory of sex chromosome evolution holds that sex chromosomes were once homologs (a pair of equivalent autosomes—the non-sex chromosomes) that evolved different morphology and gene content because they lost their ability to recombine. Suppression of recombination is thought to start around the sex-determining region, but may eventually affect much of the sex chromosomes. Recombination is a key genetic process in which two chromosomes pair and exchange their sequences. In the absence of recombination, the two chromosomes of a pair evolve separately.

In the case of mammals, whose sex chromosomes evolved about 320 million years ago, loss of recombination led to widely diverged X and Y chromosomes that pair only over a very small region, the pseudoautosomal region (PAR; because in this region the X and Y still behave like autosomes). The X and Y chromosomes of dioecious *Silene* species are morphologically distinct, like those of mammals, and they also have a PAR and a nonrecombining region. Nicolas and colleagues' results shed some light on how recombination suppression evolved on the *Silene* sex chromosomes.

DOI: 10.1371/journal.pbio.0030028.g001

Flowers of *Silene* species, clockwise from top left: male flowers of *S. latifolia* and *S. dioica,* hermaphrodite flower of *S. vulgaris,* male flower of *S. diclinis*

The authors studied four genes outside the PAR on the *Silene* X chromosomes that are also present on their Y chromosomes. They mapped the genes relative to the PAR and compared the nucleotide sequences of the X and Y version of each gene in each species. As expected of sequences that no longer recombine, the X and Y versions of each gene have diverged. Strikingly, the extent of nucleotide divergence increases with the gene's distance from the PAR.

Evolutionary biologists use sequence divergence as a clock: the longer two originally identical sequences have been isolated from one another, the more independent mutations they accumulate. The picture that emerges from the *Silene* data is one of a progressive suppression of recombination, gradually diminishing the PAR. A similar scenario has been proposed in mammals and birds. However, the authors estimate that the *Silene* sex chromosomes started diverging only 10 million years ago. The *Silene* chromosomes might therefore offer a better chance to observe recombination suppression in its early stages, and perhaps to get at its mechanisms.

The authors also report evidence for some degeneration of the *Silene* Y chromosome genes. Y degeneration is well documented in mammals, in which most X-linked genes have no Y-linked counterparts. Understanding X-Y divergence in *Silene* species may thus shed light on the evolution of sex chromosomes in vertebrates as well.

Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, et al. (2004) A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. DOI: 10.1371/journal.pbio.0030004

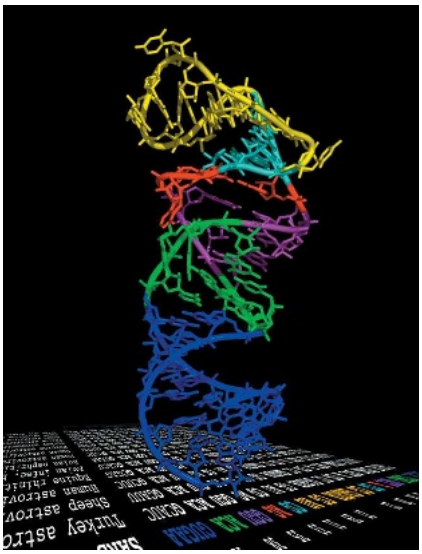
Structure of a Conserved RNA Element in the SARS Virus Genome Determined

DOI: 10.1371/journal.pbio.0030029

In February 2003, the first (and so far only) epidemic of severe acute respiratory syndrome (SARS) started in Guangdong Province, China. A respiratory illness that begins with a high temperature and can develop into life-threatening pneumonia, SARS is spread by close person-to-person contact. Before the end of the month, a Guangdong doctor had inadvertently taken the infection to Hong Kong. A woman staying in the same Hong Kong hotel as the doctor then carried the disease to Toronto. In March, the World Health Organization issued a global alert and warned against unnecessary travel to affected areas. Because of these and other containment efforts, 8,098 people became ill with SARS, rather than the predicted millions; 774 people died. The last case of the epidemic was reported in Taiwan in June 2003, and since then there have been only two cases in Singapore and nine in China.

By May 2003, a coronavirus had been identified as the cause of SARS, and the full genome sequence of this new human pathogen, which may have jumped from civet cats to people, had been published. From the viral genome, researchers have deduced the sequences and structures of the viral proteins, hoping to use this information to develop treatments and vaccines for SARS. But could the structure of the RNA genome itself also be a target for antiviral drugs?

The genome of the SARS virus is a single strand of RNA that folds into regular repeating patterns to form secondary structures such as helices. These then fold and bend in three dimensions to form complex tertiary structures. William Scott and colleagues have used X-ray crystallography to measure the exact positions of individual ribonucleotides and the interactions between them in a small segment of the SARS virus genome called the s2m element. This element sits at one end of the viral genome, and, as the researchers show, its sequence is highly conserved in related coronaviruses. Furthermore, unlike the rest of the SARS genome, which changes rapidly, the s2m element is absolutely conserved in SARS variants obtained from patients during the SARS epidemic. This strong sequence conservation indicates that the tertiary structure of s2m could be important for viral function, and when the researchers solved the three-dimensional crystal structure of the element, they found that it had a unique tertiary structure. In particular, there was a right-angle kink in its helical axis and a tunnel with a net negative charge.



DOI: 10.1371/journal.pbio.0030029.g001

Structure of a conserved RNA element within the SARS virus genome

The biological role of a new protein can often be deduced by comparing its shape with that of proteins with known functions. Scott and colleagues used this approach to hypothesize that the function of the s2m element involves interaction with a conserved host factor during the SARS life cycle. Finding a similar 90° kink in a region of ribosomal RNA that binds factors necessary for the initiation of protein synthesis, the researchers speculate that the SARS virus may use the s2m element to hijack its host cell's protein synthesis machinery. This and other putative roles need to be tested experimentally, but given that the s2m element is absent in the human genome, its unusual structural features could be an attractive target for the design of antiviral therapeutic agents.

Robertson MP, Igel H, Baertsch R, Haussler D, Ares M, et al. (2004) The structure of a rigorously conserved RNA element within the SARS virus genome. DOI: 10.1371/journal.pbio.0030005

Separating Wheat from Chaff in Plant Genomes

DOI: 10.1371/journal.pbio.0030039

Plant genome sizes span the modest—54 million base pairs (Mb) in the bitter cress *Cardamine amara*—to the enormous—124,000 Mb in the lily *Fritillaria assyriaca*. By comparison, fruitfly and human genomes have 180 Mb and 3,200 Mb, respectively. Genomes of important crops such as sorghum, soybean, maize, and wheat hover between 735 Mb and 16,900 Mb, and determining their complete sequences is daunting and costly.

Wide size variations do not necessarily reflect differences in gene content, but rather reflect the presence of repetitive sequence elements that do not generally code for genes. Repetitive elements account for at least 75% of the maize and sorghum genomes. In a new study, Joseph Bedell and his colleagues describe a way to filter away repetitive elements when sequencing the genome of sorghum (*Sorghum bicolor*), a staple crop in much of the developing world because of its resilience in arid climates.

The authors use an approach known as methylation filtration that has been employed before for pilot plant genome analyses. Here they present compelling evidence of the method's reliability when applied to large-scale genome sequencing. The approach is built on the observation that in plants, methylation—a chemical tagging of DNA with methyl groups—occurs at repetitive sequences to a much greater degree than at gene sequences. This provides an opportunity to concentrate sequencing efforts on the coding portion of the genome.

To eliminate repetitive sequences, the authors introduced small pieces of sorghum chromosomes into bacteria strains designed to specifically destroy DNA sequences that carry methyl groups. Using two independent assessments, they estimated that methylation filtration reduced the amount of sorghum DNA they would need to sequence by two thirds, from 735 Mb to approximately 250 Mb.

But were any genes lost in the filtration step? The authors compared their results to partial sequence information

generated previously from bacterial artificial chromosomes (BACs). BACs offer the most comprehensive representation of the genome because they contain large pieces of unmodified sorghum chromosomal DNA. Of the 148 genes identified on 14 sorghum BACs, 133 appeared in the filtered set. This means that the methylation filtration method captured at least 90% of the genes in the sorghum genome and 96% (131/137) if a repeat cluster of 11 known methylated genes is removed from the analysis.

Methylation filtration also compared favorably to shotgun sequencing, a method that reads the whole genome in small fragments that are progressively assembled into larger pieces by computer analysis. The authors reported that after sequencing 285 Mb of filtered sorghum DNA—approximately 1.15 times the length of the sorghum coding regions—they obtained on average 65% of the length of 96% of the genes. Theoretical calculations and simulation based on the genome of *Arabidopsis*—a plant model organism—predicted that shotgun approach would yield similar results (67% of the length of 96% of the genes) after sequencing the equivalent of 1.15 times its total length (rather than 1.15 times the length of just the coding regions). Thus, methylation filtration can provide as much information on coding sequences as the shotgun approach, with less investment in sequencing.

DOI: 10.1371/journal.pbio.0030039.g001

A field of hybrid sorghum

Methylation filtration does not yield a complete genome map, but it offers quicker, more affordable access to genes than most commonly used sequencing approaches. Sorghum is closely related to maize and sugar cane, and more distantly to rice. The availability of its genome sequence offers the chance for more in-depth experiments into the evolution of the grass family, and promises important insights into the genetic control of drought resistance.

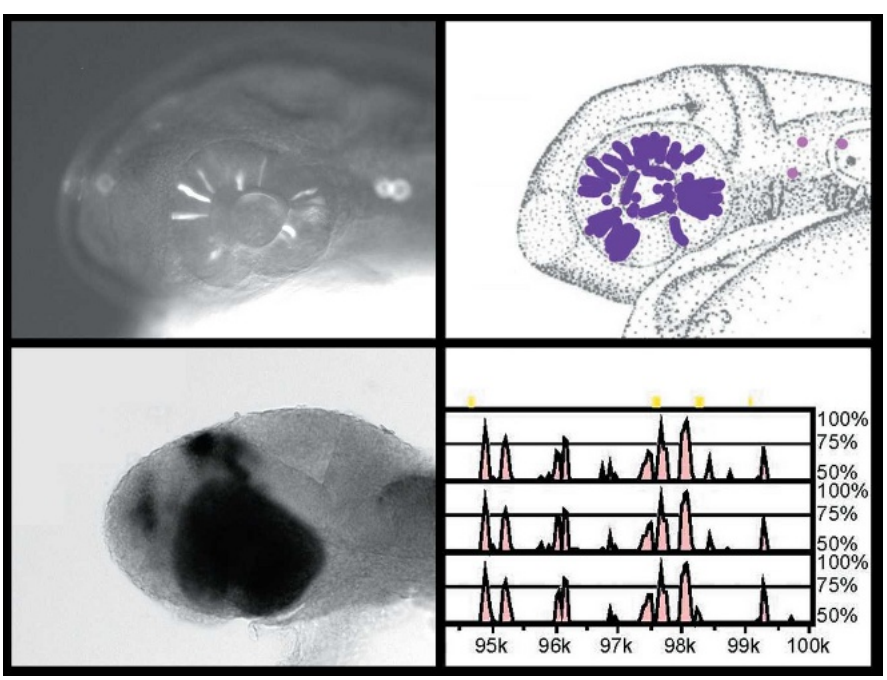
Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D, et al. (2005) Sorghum genome sequencing by methylation filtration. DOI: 10.1371/journal.pbio.0030013

Unexpressed but Indispensable—The DNA Sequences That Control Development

DOI: 10.1371/journal.pbio.0030019

Amidst the hoopla over the exact number of genes we have in our genome—more than a fruitfly, fewer than a rice plant—a more fundamental genetic truth has often been obscured. The expression of 20,000–30,000 genes is under the control of an uncounted host of non-coding sequences, which bind transcription factors and thereby regulate when and where genes are expressed. Unlike coding sequences, whose signatures are easy to spot, the characteristic features of non-coding regulatory elements are largely unknown, making their discovery by simple sequence analysis difficult. In this issue, Greg Elgar and colleagues attack this problem by comparing the non-coding sequences of the human and the pufferfish.

Since the last common ancestor of these two species existed 450 million years ago, the authors reasoned that non-coding sequences conserved between them are likely to be fundamental to vertebrate development. Through sequence alignment with increasingly strict criteria, they identified 1,373 highly conserved non-coding elements (CNEs), with an average length of about 200 base pairs. The average sequence match is 84%: not perfect, but much higher than for coding regions shared by humans and pufferfish. A quick check showed that virtually all the sequences also occurred in rodents, chickens, and zebrafish, but not in the nematode, fruitfly, or even the sea squirt, a primitive non-vertebrate chordate.



DOI: 10.1371/journal.pbio.0030019.g001

Highly conserved vertebrate non-coding elements direct tissue-specific reporter gene expression

CNEs are not spread uniformly throughout the genome. Instead, they are bunched together in fewer than 200 clusters, most of them in close proximity to genes implicated in transcriptional regulation or development. This

clustering of CNEs suggests they may not only attract transcription factors, but may also influence the local topology of the DNA, thereby controlling access to their associated gene. Several clusters also appear in regions without any known genes—the identification of these clusters might lead to the discovery of new developmentally significant genes.

While "in silico" discoveries such as this can be the jumping-off point for whole new areas of investigation, their validity must be tested "in aqua," in the wet biology of real organisms. For this Elgar and colleagues chose the zebrafish, because its transparent embryo is ideal for observing developmental events. They injected individual CNEs into embryos, along with a green fluorescent protein (GFP) reporter. By day two of development, 23 out of 25 CNEs injected had upregulated GFP expression, indicating interaction of these sequences with endogenous transcription factors. Different CNEs caused different regional patterns of expression, in keeping with their presumed roles in distinct developmental processes.

The discovery of these developmentally important sequences opens several avenues of new research. For example, analyzing the sequence and location of these CNEs may help point the way to other non-coding elements that remain undiscovered. It is also likely that mutations in these critical sequences cause human diseases. Studying how such mutations drive development astray may lead to better understanding not only of these diseases, which are likely to be rare, but also of normal human development.

Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, et al. (2004) Highly conserved non-coding sequences are associated with vertebrate development. DOI: 10.1371/journal.pbio.0030007

Computation Provides a Virtual Recording of Auditory Signaling

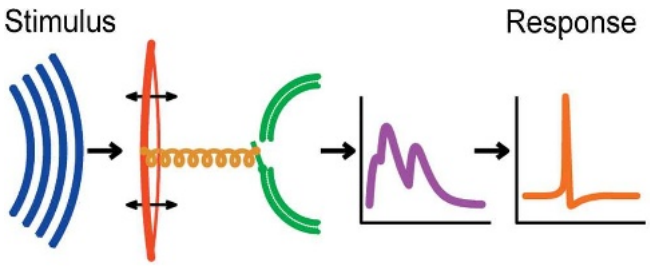
DOI: 10.1371/journal.pbio.0030026

A small rodent rustles through a field in the still night, making just enough noise to betray its location to a circling barn owl. A female frog sits on the bank of a pond amid a cacophony of courting bullfrogs, immune to the mating calls of all but her own species. Thanks to a sophisticated sensory processing system, animals can cut through a vast array of ambient auditory stimuli to extract meaningful information that allows them to tell where a sound came from, for example, or whether they should respond to a particular mating call.

An acoustic stimulus arrives at the ear as sound energy in the form of air pressure fluctuations. The sound signal triggers oscillations in mechanical resonators such as the eardrum and hair sensilla. These oscillations convert sound energy into mechanical energy, opening ion channels in auditory receptor cells and producing electrical currents that change the neuron's membrane potential. This, in turn, produces the action potential that carries the sound signal to the brain. This multi-step signal transduction process takes less than a millisecond, but exactly how it occurs at this time scale remains obscure. Direct measurements of the individual steps can't be made without destroying the mechanical structure; consequently, most measurements are taken downstream of the mechanical oscillations at locations like the auditory nerve. Likewise, the temporal resolution of most stimulus–response trials is far too imprecise to analyze processing at the sub-millisecond level.

Given these experimental limitations, Tim Gollisch and Andreas Herz turned to computational methods and showed that it's possible to reveal the individual steps of complex signal processing by analyzing the output activity alone. Using grasshopper auditory receptors as models, the authors identified the individual signal-processing steps from eardrum vibrations to electrical potential within a sub-millisecond time frame and propose a model for auditory signaling.

The crucial step in their study is the search for those sets of inputs (stimuli) that would yield a given *fixed* output (response). To get the parameters to describe the final output, the authors generated a sound stimulus (two short clicks) and recorded axon responses of receptor neurons in a grasshopper auditory nerve. From these recordings, they defined the fixed output as the probability of a receptor neuron firing a single action potential. They then asked how the various parameters, which were



DOI: 10.1371/journal.pbio.0030026.g001

A schematic representation of auditory signaling

associated with different time scales, could produce the same predefined firing probability.

By varying the stimulus parameters and comparing the obtained values within their mathematical framework—and making certain assumptions, for example, that the steps signal through a "feedforward" process—they could then tease out the individual processing steps that contribute to the desired output within the required time frame. With this approach, Gollisch and Herz disentangled individual steps of two consecutive integration processes—which they conclude are the mechanical resonance of the eardrum and the electrical integration of the receptor neuron—down to the microsecond level. Surprisingly, this fine temporal resolution is achieved even though the neuron's action potentials jitter by about one millisecond.

Thus, using just the final output, this approach can extract the temporal details of the individual processes that contribute to the chain of auditory transduction events. While this method is best-suited for deconstructing unidirectional pathways, the authors suggest it could also help separate "feedforward" from feedback signaling components, especially when feedback is triggered by the final steps. But since many sensory systems share the same basic signal-processing steps, this method is likely applicable to a broad range of problems.

Gollisch T, Herz AVM (2005) Disentangling sub-millisecond processes within an auditory transduction chain. DOI: 10.1371/journal.pbio.0030008

Whisker Velocity Patterns Tell Rats What They're Feeling

DOI: 10.1371/journal.pbio.0030036

Whiskers don't fossilize, so it's hard to say when they first evolved. But it's quite likely they emerged along with mammals, over 200 million years ago. To elude the eye (and feet) of ungainly dinosaurs, it's thought these shrew-like prototypes foraged at night and sought refuge underground, where the sensory advantages of whiskers would come in handy. Nocturnal animals use whiskers much like the blind use walking sticks: to navigate their surroundings, explore close objects, and avoid running into things.

Whiskers, or vibrissae, connect to nerves, blood vessels, and muscles. These special connections allow rats, for example, to actively "whisk" the surface of objects and discern fine differences in texture, just as we move our fingertips along a surface to pick up details. In the wild, whisking helps rats navigate unfamiliar terrain to find food. But how does the brain know what the animal is touching?

Rat whiskers scan surfaces in a rhythmic motion that excites sensory receptor cells embedded in their whisker pad. Receptors in each whisker shaft are innervated by several hundred "first-order neurons" that relay sensory signals to second-order neurons in the brain stem, then on to third-order neurons in the thalamus, and finally on to the cortex, where sensory stimuli are integrated in cell clusters called barrels.

Ehsan Arabzadeh, Erik Zorzin, and Mathew Diamond work with rats to investigate how sensory receptors extract fundamental features from complex and diverse stimuli to encode texture. Not much is known about how receptor and cortical neurons respond to active whisking along irregular surfaces, though responses to simple stimuli (like sinusoidal vibrations) suggest that neurons might represent texture by encoding kinetic features of whisker vibrations, in particular, velocity. In a new study, Diamond and colleagues investigate the connection between textures, whisker vibrations, and neural codes: do distinct textures produce distinct vibrations? If so, how are these vibrations encoded and reported?

The authors first collected kinetic data of whiskers moving across different textured surfaces. Stimulating cranial nerve VII of anesthetized rats (the motor nerve) generated whisking movements akin to those seen in conscious rats; the kinetics of these movements and the vibrations of the whisker shafts were measured under different conditions, including no contact with objects ("free whisk"), contact with smooth objects, and contact with various grades of sandpaper. These vibrations were then "played back" to other rats, while measuring the neuronal activity at two critical stages in the sensory pathway: the first-order neurons that innervate the whiskers and the barrel cortex neurons that integrate the incoming signal.

DOI: 10.1371/journal.pbio.0030036.g001

Timing of neuronal activity captures sensory information

Altogether, the authors collected a neural dataset consisting of first-order recordings, barrel cortical cluster recordings, and simultaneous paired recordings from both sites, all in response to playback of the library of texture-related vibrations. This approach afforded the opportunity to directly compare encoding of information at both levels in the sensory pathway. These recordings show, the authors argue, that temporally distinct firing patterns in the trigeminal ganglion (the cell bodies of the first-order neurons) and cortex captured the kinetic features of the texture-induced vibrations. Each texture's "kinetic signature" is encoded by a characteristic, temporally precise firing pattern associated with whisker movement. Compared to free whisking, coarse sandpaper produced irregular bursts of high and low velocity, and both first-order and cortical neurons fired far more impulses for coarse sandpaper than for free whisks. The authors then used stimuli consisting of random velocities to uncover the "tuning curves" of neurons, and simulations showed that these neuronal tuning curves could perfectly predict the real neural responses to textures.

Noting the close match between the simulated and natural responses, Diamond and colleagues conclude that the texture-induced firing patterns observed in the first-order and cortical neurons suggest that neurons selectively encode elemental kinetic features—namely, high velocity—to tell rats what they're whisking. This selectivity allows even a single whisker to transmit significant bits of texture-specific information to the brain. Interesting as rat whisking may be, these findings have relevance beyond the world of whiskered beings, shedding light on the underlying neural processes that translate touch into recognition.

Arabzadeh E, Zorzin E, Diamond ME (2005) Neuronal encoding of texture in the whisker sensory pathway. DOI: 10.1371/journal.pbio.0030017

Hirsute or Hairless? Two Proteins May Spell the Difference

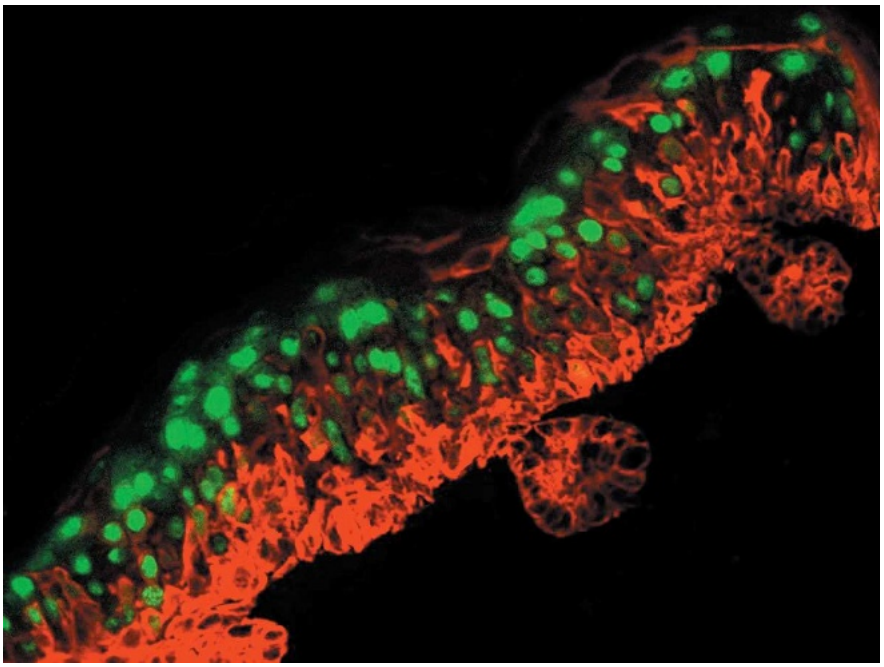
DOI: 10.1371/journal.pbio.0030035

If you're a cat fancier, you're well aware that hair follicles are expendable. The product of a spontaneous mutation that caught a cat breeder's eye, *le chat nu*, would quickly succumb in the wild—its winter coat consists of little more than a ridge of fur down the midback and tail—and needs special care to thrive as a pet. Hairless animals in the lab, on the other hand, can be very instructive. Understanding how hair develops sheds light on the fundamental processes that generate a wide range of tissues and organs, including the lungs, cornea, and mammary glands.

How complex, three-dimensional structures emerge from single sheets of cells is a fundamental question in developmental biology. The dispensability of hair follicles makes them the perfect model system for studying this question—specifically, how structures and organs develop from buds. In a new study, Elaine Fuchs and colleagues use a three-pronged approach—involving gene expression analysis, transgenic mice, and cell cultures—to study how epithelial buds, the precursors of hair follicles, form. Their experiments point to two key actors in a signaling pathway that molds a targeted cluster of cells into a hair bud.

During the budding process, overlapping signaling pathways from two adjacent embryonic cell layers—the epithelium and the mesenchyme—direct morphogenesis. The mesenchymal cells initiate the cell-to-cell "crosstalk" that controls bud formation by first directing a small cluster of epithelial cells to form a placode, the pouch that forms hair plugs. The placode in turn directs underlying mesenchymal cells to form the base of the hair follicle, called the dermal papilla, and both structures contribute to the mature hair follicle. During development, cells are constantly bombarded with external signals. The trick is figuring out which signals trigger the transcriptional and behavioral properties in cells that spur bud formation.

In previous experiments, Fuchs and colleagues showed that reducing expression of E-cadherin—a membrane protein that forms the adhesive junctions between epidermal cells—is essential for allowing the cell remodeling required for bud formation. Here, the authors analyze the timing of external signals against the response of targeted cells to determine how targeted cells translate signals into changes in cell adhesion and remodeling, proliferation, and differentiation—the agents of most types of organogenesis.



DOI: 10.1371/journal.pbio.0030035.g001

Transgenic epidermis expressing Snail (red) results in expanded keratin 1 expression (green)

Since Snail, a protein that impedes the transcription of a subset of genes, functions in many developmental processes requiring epithelial remodeling, the authors reasoned it might do the same in hair bud formation. Working with developing mouse embryos, they saw a spike in Snail expression on embryonic day 17.5, coinciding with hair bud formation, enhanced cell proliferation, and the down-regulation of E-cadherin. Artificially sustaining Snail expression in the skin of transgenic mice caused abnormal levels of cell proliferation in the epidermis and reduced cell adhesion.

Working with skin keratinocytes, precursors of hair fibers, Fuchs and colleagues explored several signaling proteins known to be involved in bud formation as possible activators of Snail expression. When the authors treated keratinocytes with small amounts of one stimulator, TGF-β2, they saw "rapid and transient induction of Snail." Snail proteins were absent from 17.5-day-old knockout mice lacking TGF-β2 but not from their nonmutant littermates. Conversely, transgenic mice with elevated TGF-β2 signaling activity displayed ectopic expression of Snail. Knockout mice lacking TGF-β2 also showed higher levels of E-cadherin—normally down-regulated by Snail—than their nonmutant littermates.

Altogether, these findings suggest that TGF-β2 signaling transiently induces Snail, which in turn down-regulates E-cadherin and activates a proliferation pathway in the developing bud. Reduced E-cadherin, the authors conclude, appears to contribute to Snail-mediated enhanced proliferation by allowing proteins normally sequestered at the membrane to operate in a proliferation pathway after the number of cellular junctions diminishes. By identifying which molecules are active in specific cell types at specific developmental stages, this study lays the foundation for dissecting the mechanisms that connect two key processes—intercellular remodeling and proliferation—in epithelial development. And since the consequences of TGF-β2 activity seen here in the hair bud more closely resemble certain types of skin cancer progression than skin development, a mechanistic understanding of hair follicle development promises to shed light on how skin cancer develops as well.

Jamora C, Lee P, Kocieniewski P, Azhar M, Hosokawa R, et al. (2004) A signaling pathway involving TGF-β2 and Snail in hair follicle morphogenesis. DOI: 10.1371/journal.pbio.0030011

Gene Signatures Predict Interferon Response for MS Patients

DOI: 10.1371/journal.pbio.0030018

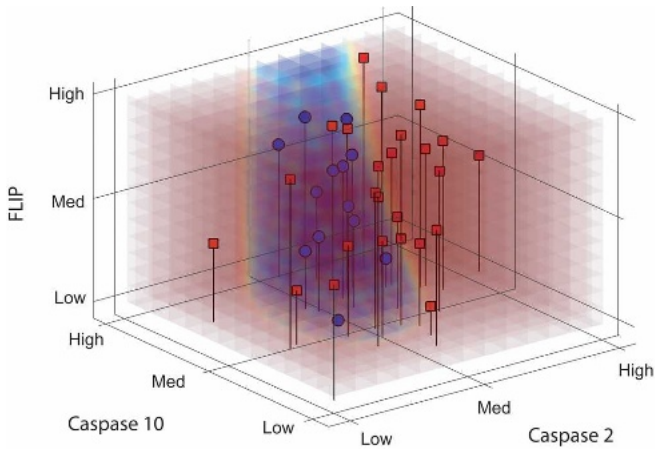
Multiple sclerosis (MS) can be an unpredictable disease. It develops when the body's immune system attacks healthy nerve cells and disrupts normal nerve signaling. Patients experience a wide range of symptoms—including tingling, paralysis, pain, fatigue, and blurred vision—that can appear independently or in combination, sporadically or persistently. Although symptoms appear in no particular order, flare-ups are common in the majority of patients.

MS flare-ups are commonly treated with beta-interferon. Adverse effects are not uncommon, and, more importantly, a sizable proportion of patients show a reduced response, or no response at all. Given the variability of the disease and treatment response, being able to predict how a particular patient is likely to respond to interferon would help doctors decide how close to monitor the patient or even whether to consider alternative treatments. In a new study, Sergio Baranzini et al. describe a computational model that can predict a patient's therapeutic response to interferon based on their gene expression profiles.

Immune cells typically secrete interferons to fend off viruses and other pathogens. Interferons stem viral infection by inhibiting cell division in neighboring cells—thus preventing the virus from reproducing—and triggering pathways that kill the infected cells. It's thought that interferon therapy may relieve symptoms associated with MS by correcting imbalances in the immune system that lead to disease. Interferon therapy produces changes in the gene expression profile of targeted cells—that is, it inhibits or activates certain genes—which in turn alters the cells' activity.

Blood samples were taken from 52 patients with relapsing-remitting MS (marked by acute flare-ups followed by partial or full recovery), and their RNA was isolated from a class of immune cells called peripheral blood mononuclear cells. After patients started interferon therapy, blood was taken at specific time points over the course of two years. Baranzini et al. measured the expression level of 70 genes—including a number involved in interferon interactions and immune regulation—at each time point.

The authors used statistical analyses to search for gene expression profiles that were associated with patients' therapeutic outcomes. They looked for patterns in analyses of single genes, gene pairs, and gene triplets, and found their model's predictive accuracy increased with gene number. They also looked for genes that showed different expression patterns over the two years based on patient response, time passed, and patient response over time. These analyses identified genes that increased activity independently of clinical response (interferon can activate genes that have no effect on disease), as well as genes that were associated with a good or poor response. Some of these genes were also the best predictors of patient response before therapy was started.



DOI: 10.1371/journal.pbio.0030018.g001

Expression levels of three genes in beta-interferon responders (red) and non-responders (blue)

This approach can predict the probability of a good or poor clinical response with up to 86% accuracy. Baranzini et al. offer hypotheses to explain how the observed gene activity might produce the differential responses to therapy—for example, a poor response may stem from downstream signaling events rather than from problems with drug metabolism. But the authors caution that the mechanisms connecting these genetic signatures to specific outcomes—and the mechanisms that produce a positive interferon response—have yet to be established. For now, these patterns should be thought of as markers. Still, these results suggest that doctors could one day tailor MS patients' treatments to their molecular profile, and perhaps take some of the uncertainty out of this capricious disease.

Baranzini SE, Mousavi P, Rio J, Caillier SJ, Stillman A, et al. (2004) Transcription-based prediction of response to IFNβ using supervised computational methods. DOI: 10.1371/journal.pbio.0030002

Dispersal or Drift? More to Plant Biodiversity Than Meets the Eye

DOI: 10.1371/journal.pbio.0030043

Over 250 million years ago (mya), all the continents of Earth formed a single land mass called Pangaea. Some 50 million years later, this supercontinent began to split in two, forming Laurasia—now North America, Asia, and Europe—and Gondwana—present-day Antarctica, Australia, South America, Africa, and India. After another 50 million years, Gondwana, too, broke up. At the end of the Cretaceous period, New Zealand split off (about 80 mya), then South America and Australia separated from Antarctica (about 35 mya). Fairy-tale quality aside, the story of continental drift fits comfortably with the geological and fossil record and feeds our understanding of current distributions of plant biodiversity.

Although we know how and when Pangaea broke apart, the distribution of fossils of the same species on many different continents, separated by vast ocean waters, challenges us to explain how they got there. Plant life on New Zealand, for example, shares striking similarities to that on other Southern Hemisphere land masses, but scientists have yet to agree on how this came to pass. In particular, one genus, *Nothofagus*—the southern beech tree, a plant whose 80-million-year-old fossil history goes back to the days of Gondwana—has polarized views on the nature of Southern Hemisphere biogeography.

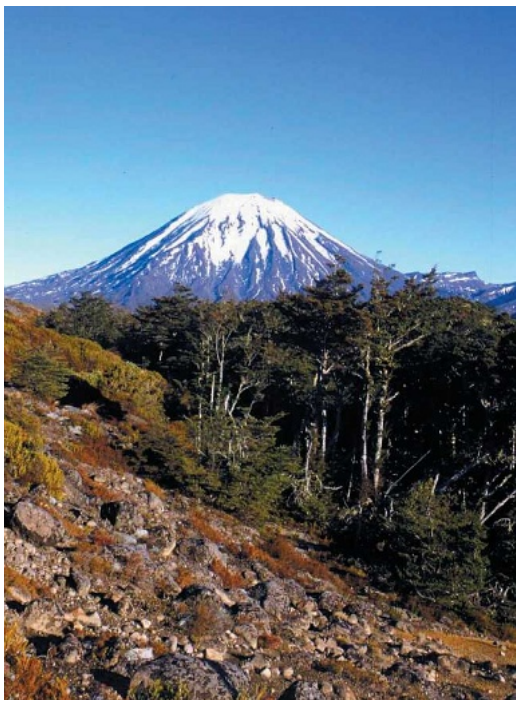
One theory suggests that geographic barriers (New Zealand and Australia are

separated by the Tasman Sea) would have prevented species expansion after the break-up of the continents, so similar contemporary species must have already existed in both places before New Zealand broke away from Gondwana. In this scenario, called vicariance, ancestors of existing lineages drifted with the repositioned land masses. Another hypothesis, born of existing distributions and fossil data, suggests that long-range oceanic dispersal is more likely. But since *Nothofagus* seeds are not considered ocean-worthy vessels, many believe vicariance is the only possible explanation.

Peter Lockhart and colleagues argue that a clear picture of the divergence dates of various southern beech species could help clarify the relative contributions of vicariance versus dispersal. But they would need significant lengths of DNA sequences to reliably characterize the evolutionary history of each species.

Consequently, Lockhart and colleagues analyzed a 7.2-kilobase fragment of the chloroplast genome (which typically ranges from 110,000 bp to 160,000 bp) for 11 species of three *Nothofagus* subgenera—*Lophozonia*, *Fuscospora*, and *Nothofagus*—from South America, Australia, and New Zealand. Reconstructing the trees' evolutionary relationships (phylogeny) based on analyses of their chloroplast sequences, the authors discovered a nuanced evolutionary history that supports vicariance for some species and dispersal for others.

Assuming that beech was present throughout Gondwana (which fossil data support), the sequence of the Gondwana breakup should be reflected in the beech's phylogeny. New Zealand beeches should be more distantly related to both Australian and South American species, because of the greater period of separation—65 million years compared to 30 million years. Yet Australian and New Zealand beeches are more closely related to each other than to South American species, which reflects more recent relationships. Given that fossils of all beech subgenera extend back to the New Zealand Cretaceous period, the dating of splits and the nature of the relationships indicate extinction of beech lineages within current subgenera in New Zealand, and possibly in Australia and South America.



DOI: 10.1371/journal.pbio.0030043.g001

***Nothofagus*, the southern beech, on the slopes of Mt. Ruapehu in New Zealand (Photo: Peter Lockhart)**

Lockhart and colleagues' analyses suggest that the relationships of the Australian and New Zealand *Lophozonia* and *Fuscospora* species are too recent to have roots in Gondwana, indicating a role for transoceanic dispersal. The evolutionary relationship between the Australasian and South American *Fuscospora* lineages, however, is consistent with vicariance. These divergence results, the authors conclude, indicate that current distributions of *Nothofagus* cannot be explained solely by continental drift (followed by extinction of some species) and that contemporary New Zealand *Nothofagus* species are not direct descendants of the beeches thought to have reached the island after the split from Antarctica.

Taken together, the results highlight the need for caution in evaluating fossil evidence. The fossil record doesn't necessarily capture when a species first appeared, and a continuous fossil presence can mask extinctions and reinvasions. The authors conclude that their molecular data make the case for investigating possible mechanisms of long-range dispersal—especially the dispersal properties of *Nothofagus* seeds—and stresses the need to consider more complex hypotheses to explain something as dynamic and complex as the evolutionary history of biodiversity.

Knapp M, Stöckler K, Havell D, Delsuc F, Sebastiani F, et al. (2005) Relaxed molecular clock provides evidence for long-distance dispersal of *Nothofagus* (southern beech). DOI: 10.1371/journal.pbio.0030014