"Computer Genome" Is Full of Junk DNA

Random processes of gene duplications and deletions can generate multigene families and large quantities of vestigial sequences in eukaryotic genomes

A complete page in an April issue of the *Proceedings of the National Academy of Sciences* was filled with nothing but seemingly endless lines of dashes within which were interspersed a few letters and symbols. These days, when the pages of many biological journals burden the eye with huge and heroically generated gene sequences, this particular figure looks superficially like yet another. But it is not. It is a map of an entire genome, which comprises 83 genes and 20 times as much noncoding DNA.

But the really interesting thing about this genome is that it is entirely hypothetical, the product of the combined imagination of two University of California biologists and 6 hours of computing time on the campus's VAX computer.

William Loomis is a developmental biologist and Michael Gilpin a theoretical ecologist, and both are at San Diego. They recently joined forces to address the question of why there is so much DNA in the world. Specifically, why the nuclei of eukaryotic organisms contain something between 20 and 100 times as much DNA in their chromosomes as is apparently needed to encode the suite of proteins that are typically produced.

The product of gene cloning and sequencing in recent years has, of course, shown that most eukaryotic genes are interrupted with noncoding, intervening sequences (introns). Novel controlling elements, such as enhancers, have been discovered too. And some sequences appear to have something of a quasi-autonomous existence within the genome, and therefore have been termed selfish DNA. So this begins to answer the puzzle of the "excess" DNA. The issue that Loomis and Gilpin address, however, is the possibility that at least some of this DNA might be there simply as a result of chance.

Specifically, they ask whether the existence of multigene families and vestigial sequences might essentially be the product of random processes: no more, no less.

It is true, of course, that the members of some gene families have slightly different structures from each other and perform discrete functions. The globin family is a good example here. But in other cases, such as actin genes, the differences between family members seem to be more in timing of expression rather than in structure. And the number of genes in the actin family varies widely between species, for no very obvious reason. The slime mold *Dictyostelium discoideum* has 17 actin genes, for instance, whereas the yeast *Saccharomyces cerevisiae* has only one.

Many biologists are unhappy with the idea that much of the DNA might have no function.

Could it be, ask Loomis and Gilpin, that random processes are responsible for generating multigene families, which occasionally are functionally exploited but often are of no particular selective advantage?

Using a rather large and difficult simulation program, which includes rules for DNA duplication and deletion, Loomis and Gilpin indeed find that, given enough time, a single gene will blossom into a genome that contains many genes, some of which are members of multigene families and all of which are embedded in a very large proportion of dispensable sequences. One such "genome" was featured in the *PNAS* diagram.

Two types of simulation were done: one that contained a single gene type and concentrated on the dynamics of producing a genome of equilibrium size, and a second that allowed for the evolution of new genes and therefore produced a more complex genome.

In the first, the gene was represented by four units: a start unit (promoter), two coding units, and a stop unit (terminator), all of which was designated "#AA.". Duplications that included all four units produced a duplicate gene, whereas vestigial sequences resulted from partial duplications. With these and other rules built in, the simulation was allowed to run for up to 100,000 events, the result of which was a steady increase in the size of the genome, with an eventual plateau at a little over 4000 units. The number of genes fluctuated between one and ten, with extinction resulting when the sole gene in a genome was deleted.

The equilibrium size for the evolving genome is therefore seen to be quite large, and contains a very high proportion of noncoding sequences. The reason, explain Loomis and Gilpin, is that only at this point does the rate of nondeleterious deletions equal the rate of duplications.

This simulation clearly shows that eukaryotic genomes, which, in contrast with those of viruses and bacteria, are unaffected by considerations of packaging or speed of replication, will evolve to a large size, most of which sequences are vestigial in some way. But a genome with only one type of gene is not a very convincing model of reality. Loomis and Gilpin therefore modified the simulation to allow for the generation of new genes, which were assigned a selective advantage.

The result of the simulation after 10,000 events was a 6356-unit long genome, which contained 83 genes. A little over one third of the genes were present as two or more copies, and one of them had seven members in its family. This proportion of single-copy to multiple-copy genes is in fact very similar to what is seen in the typical eukaryotic genome. So too is the proportion—one twentieth—that the genes constitute of the entire genome.

These "data" on genome structure fit very nicely with predictions Loomis made 13 years ago in a short paper in *Developmental Biology*. Loomis, who had cut his professional teeth on the genetics and regulation of the *E. coli lac* operon, essentially said that if eukaryotic genomes contain just a few thousand genes, rather than the many hundreds of thousands that was commonly believed at the time, then their control would not have to be dramatically different from what had been seen in bacteria. Most of the DNA in eukaryotic genomes, he suggested, did nothing at all.

Many biologists were unhappy with the idea that much of the DNA might have no function, says Loomis. "There is a very strong feeling among these people that if a molecule, or any kind of biological structure, exists, then it must be serving some kind of selectively advantageous purpose. I disagree with this viewpoint very strongly." Loomis prefers to turn the question around. "We should ask, 'what is the selective advantage of getting rid of a particular structure?' This is not common thinking."

When he wrote the *Developmental Biology* paper, Loomis recalls, he "didn't have the foggiest idea how anyone would test what I was saying." That was in the days before

gene cloning and DNA sequencing. But the data that have since flowed from this technology, patchy though it is, is consistent with those initial predictions, he says. "Molecular maps covering 100 kb of DNA are characterized by islands of transcribed sequences in a sea of silent DNA," write Loomis and Gilpin. "But how do you summarize the empirical data?" asks Loomis. "You can take the globin region, you can take the chorion region and so on and so on. Each case is just a single case, and you need the sum of a hundred or so. No one has compiled that." Hence the importance of the computer-generated genomes.

It is of course very difficult to prove that a structure or a sequence of DNA has no function. "People will always say, ah, but you haven't looked under the right conditions," says Loomis. In the case of multigene families, the best data come from mutation experiments. Nematodes, for instance, have two acetylcholinesterase genes, both of which have to be inactivated before the animal is paralyzed. "Knock either one out, and the worm is fine, which tells you that the fact that there are two genes in this family is of no particular functional significance, probably," says Loomis. But this kind of work is exceedingly hard to do, and so the data coming from it will be limited.

Loomis and Gilpin have by now generated many complex genomes using their simulation program. "I considered at one point following in detail the history of different sections," says Loomis. "I would have been able to say, here's where the duplication occurred, here's where the deletion occurred and so on. It would have been a clear evolutionary tree." He didn't do it, because he realized there would be no real information to be gained. "Every simulation is different and therefore any given simulation is rather meaningless: each is like a different planet."

Although the simulation data encourage Loomis to believe that his earlier predictions are correct, this recent work should not be seen as answering every question about eukaryotic genomes, he stresses. "We are simply explaining one aspect of genomes: the outcome of random duplications and deletions. For instance, multigene families can appear as a consequence of random duplications and deletions, and have no necessary selectively advantageous function. Large quantities of dispensable sequences will accumulate in the genome before its size stabilizes. We are not trying to explain anything else." ■ ROGER LEWIN

ADDITIONAL READING

W. F. Loomis and M. E. Giplin, "Multigene families and vestigial sequences," *Proc. Natl. Acad. Sci. U.S.A* **83**, 2143 (1986).

Briefing:

A Solution to the Solar Neutrino Puzzle?

Two Soviet physicsts have offered what seems to be the most natural and plausible explanation yet for the mystery of the missing solar neutrinos. Their mechanism requires no exotic new particles, no unobserved new forces, and no modifications to the standard model of the solar interior.

Instead, S. P. Mikheyev and A. Yu. Smirnov of the Institute for Nuclear Research, Academy of Sciences, Moscow, have pointed out a previously unrecognized effect caused by the conventional weak interactions. Simply put, electron-type neutrinos emitted in the core of the sun are changed into muon-type neutrinos on their way out. These transformed particles then escape detection on Earth.

Although Mikheyev and Smirnov actually announced their result at a meeting last year in Finland, it was not widely appreciated until this spring, when Cornell University physicist Hans A. Bethe called attention to it in a paper published in *Physical Review Letters*. "I think this is the first explanation [of the solar neutrino problem] that could be right," Bethe says, echoing a perception now common among his colleagues. As University of Washington physicist Wick Haxton puts it, "Looking back, it's almost unbelievable that this mechanism was overlooked for so long."

Indeed, the solar neutrino problem is now nearly two decades old. According to the standard argument, nuclear reactions in the core of the sun will produce neutrinos at a certain, calculable rate. These neutrinos will then stream freely through the sun's outer layers and will be detectable on Earth. However, the standard argument is clearly going wrong somewhere: a solar neutrino detector developed by Brookhaven National Laboratory's Raymond Davis has operated since 1968 in South Dakota's Homestake gold mine, and has consistently measured a neutrino flux of only one-third the predicted value.

The theorists are thus left with two alternatives. Either the neutrinos are not being produced at the predicted rate—and it is hard to think of a plausible reason why not, since the standard model of the sun is based on well-understood nuclear physics and has been very successful in relating the mass and composition of the sun to its luminosity and lifetime—or else the particles are somehow getting lost on their way to South Dakota. More precisely, since the Homestake detector is sensitive only to electron neutrinos produced by certain high-energy reactions, it is the high-energy electron neutrinos that are getting lost. The question is, Where?

The answer given by Mikheyev and Smirnov starts from the fact that any neutrino traveling through ordinary matter has a slight chance of being scattered by the weak interactions. In the case of the muon- and tau-type neutrinos this effect is negligible. However, as Lincoln Wolfenstein of Carnegie-Mellon University first pointed out in 1978, the implications for an electron neutrino are quite different: the particle behaves as if its mass had been increased by a tiny fraction proportional to the density of the surrounding matter.

What Mikheyev and Smirnov realized is that this tiny effect can have large consequences at the center of the sun, where the density is more than 130 grams per cubic centimeter. In those regions an electron neutrino might actually be more massive than its cousin, the muon neutrino; moreover, as the electron neutrino propagated outward to regions of lower density and lower mass it would actually *become* a muon neutrino—and thus be rendered unobservable in the Homestake detector.

This mechanism obviously depends upon neutrinos having a small mass to begin with. It also requires a certain amount of mixing between the electron and muon neutrinos—that is, a certain probability that one type of neutrino can transform itself into the other as it moves along. While neither of these phenomena have been observed in the laboratory, both are predicted by the grand unified theories of particle interaction. Indeed, by requiring that the Mikheyev and Smirnov theory agree with the data from the Homestake detector, Bethe and others have estimated that the mass of the muon neutrino is less that 0.008 electron volts, and that the probability of mixing is less than 1 percent. Both figures are right in line with the results of the grand unified theories.

Unfortunately, it will be very difficult to detect such small effects in laboratory experiments. However, some predictions of the Mikheyev-Smirnov theory could be tested by a solar neutrino detector made of gallium—a project often proposed, and never yet funded. "It's deserved support for many years," says Paul Langaker of the University of Pennsylvania. "And now it's even more important." ■ M. MITCHELL WALDROP

ADDITIONAL READING

S. P. Mikheyev and A. Yu. Smirnov, in *Proceedings of the Tenth International Workshop on Weak Interactions*, Savonlinna, Finland, 16–25 June 1985 (unpublished).

H. A. Bethe, "Possible explanation of the solar neutrino puzzle," *Phys. Rev. Lett.* **56**, 1305 (1986).