direction. NGC 6397 has been studied intensively from the ground, and in 1992, Canadian astronomers Harvey Richer and Gregory Fahlman published a report in *Nature* on that cluster, claiming that the numbers of stars increased sharply and inversely with stellar mass. (They still concluded that red dwarfs were not abundant enough to solve the galactic dark matter problem, however.)

When Paresce, Guido De Marchi of STScI, and the rest of their colleagues analyzed the HST image of NGC 6397, however, the data told a different story. Down to a certain level of brightness, the number of red dwarfs rose. Beyond that point, however, the number of fainter stars quickly dropped to zero. Paresce's group converted each star's brightness to an estimated mass and found that most of the red dwarfs in their picture were about 0.2 solar masses. At lower masses (down to about 0.08 solar masses, the smallest star that can sustain hydrogen fusion), it appears that red dwarfs are rarely made, says Paresce, whose team will publish their work in the 10 February issue of *Astrophysical Journal*.

Paresce attributes the stark conflict with the earlier results to the nearsightedness of ground-based telescopes. Those instruments, he says, have trouble picking faint stars out of crowded fields and discriminating between distant galaxies and blurry stars. To correct for this myopia, ground-based astronomers modify their star counts based on knowledge of the equipment's limits. Those corrections have always been suspect, argues Paresce.

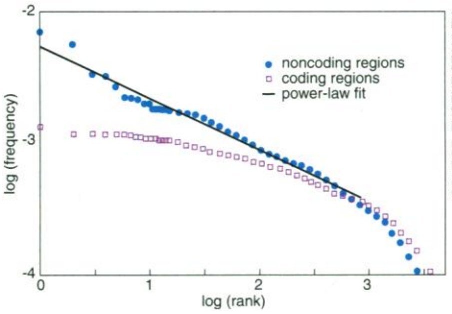
Astronomers are now starting to wonder how far the work on NGC 6397 can be generalized. Does it apply to the rest of the galaxy and the universe? Researchers generally classify stars into two categories, Population I and Population II. Globular clusters contain Population II stars, which are much older than those of Population I and contain fewer metal elements. While very low mass Population II stars may be rare, that might not be true for Population I, says STScI's Henry. His own work on much closer stars, most of which are Population I, shows the number of stars increasing with lower mass right down to the hydrogen-burning limit of 0.08 solar masses. The conflicting data, says Henry, could hint that Population I and II stars form in slightly different ways. "It could be that you just can't make low mass stars with low metallicity," he says.

Even as stellar theorists try to incorporate these latest results into their models of how pockets of gases collapse to form stars, cosmologists continue to scratch their heads about the mystery of dark matter. Not only do the new images suggest that faint red dwarfs are not the solution; they cast doubt on related notions, such as the idea that dark matter consists of hordes of brown dwarfs, invisible starlike objects with masses below the hydrogen-burning limit. It would be "very odd," argues Bruce Margon, an astronomer at the University of Washington, if the universe rarely made objects between 0.2 and 0.08 solar masses, but instead made lots of even smaller starlike objects. Yet the surprising absence of red dwarfs suggests that astronomers should continue to expect the unexpected from the universe.

–John Travis

MATHEMATICAL BIOLOGY

Hints of a Language in Junk DNA



Line of evidence. Plotting frequency against rank for arbitrary "words" in noncoding yeast DNA yields the linear plot found in human languages.

SOURCE: EUGENE STANLEY

As molecular biologists unraveled the mysteries of the genes, they were startled to find that only a small fraction of DNA actually carries the code for making proteins. Fully 97% of the DNA in human cells looks like meaningless filler, and researchers have struggled to explain why it is there. Physicist Eugene Stanley of Boston University thinks his group has found a clue: Even if it doesn't encode proteins, says Stanley, the "junk" DNA may carry some kind of message.

Using statistical techniques borrowed from linguistics, Stanley says he and his colleagues Rosario Mantegna, Sergey Buldyrev, and Shlomo Havlin, along with C.-K. Peng, Ary Goldberger, and Michael Simons from Harvard Medical School, "have shown fairly clearly that the 'junk' has all the features of a language." The analysis, which will appear in the 5 December issue of *Physical Review Letters*, says nothing about how that language works. But biologists, already pursuing other hints of functions for junk DNA (*Science*, 4 February, p. 608), are intrigued. "It sounds extraordinarily interesting," says University of Cambridge geneticist Lawrence Hurst, "though it might be hard to tell what it means."

The group first looked into the junk DNA problem several years ago by applying statistical techniques to the sequences of nucleotide bases in the junk. The analysis seemed to show that junk sequences could not be completely random, because the chance of finding a particular base at a given point in a sequence showed a weak, statistical dependence on nearby bases. No one knew what these correlations meant, but Stanley took them as "a hint that something interesting was lurking in the junk."

In an attempt to find out what that something might be, the Boston researchers applied two standard linguistic tests to junk sequences from a variety of organisms. One test was developed 45 years ago, when linguist George Zipf took texts in several languages and ranked the words in order of frequency. He then created a histogram, plotting the log of the actual frequency of occurrence against the log of the rank. Surprisingly, this procedure yields a linear plot with a slope of –1 for every human language. "This is really remarkable," says Stanley. "There's no rhyme or reason why that should be true."

When the group arbitrarily divided up their samples of junk into "words" between 3 and 8 bases long and applied the Zipf test, the telltale linear plot emerged. They then tried a second test developed by information theorist Claude Shannon in the 1950s to quantify the "redundancy" of any string of characters. Languages are redundant sequences, explains Stanley: You can fill in a typographical error by noting nearby characters. A random sequence, in contrast, has no redundancy. Applied to junk DNA, Shannon's formula revealed a surprising amount of redundancy—another sign that something was written in these mysterious stretches.

The familiar coding regions of the genes fail both tests—an expected result, Stanley says, because the "language" of the genes lacks key features of ordinary languages. "The coding part has no grammar—each triplet [of bases] corresponds to an amino acid [in a protein]. There's no higher structure to it." In contrast, junk DNA's similarity to ordinary languages may imply that it carries different kinds of messages, says Stanley.

Although these findings are surprising, the idea that junk DNA has a function is not. "I think that what we call 'junk DNA' will have a number of uses," says Harvard biologist Walter Gilbert. He notes that, among other recent hints, Leroy Hood of the University of Washington and Ben Koop of Canada's University of Victoria recently found a long stretch of "junk" that is common to both humans and mice, implying that it and perhaps other junk sequences have some important function in the cell. Among the possibilities under discussion, he says, are roles in protecting and repairing the genetic material and in regulating genes. "I think the junk is like the stuff in a junk shop," Gilbert says. "You can find lovely things in it."

–Faye Flam