



Disappearing act

The bizarre absence of certain gene classes in eukaryotes is key to understanding their evolution and complex links with prokaryotes.

James A. Lake

Eukaryotic evolution is something of a Gordian knot. Using single genes to unravel it won't work, as the genomes of eukaryotes (animals, plants, fungi and protists) are derived from those of several prokaryotes (eubacteria and archaeobacteria). So the focus has shifted towards analysing flows of gene populations, and even of entire prokaryote genomes, into eukaryotes. These more holistic studies are revealing the complex genetic and evolutionary connections between eukaryotes and prokaryotes.

Until recently, everyone assumed, based on a single ribosomal RNA gene, that eukaryotes descended from archaeobacteria — extremophilic prokaryotes distinct from 'true' bacteria, or eubacteria. Now we know that's not the case. More than two-thirds of the nuclear genes of the yeast *Saccharomyces cerevisiae*, for instance, are derived from eubacteria, and the balance from archaeobacteria. What we know of gene losses and gains also indicates that the eukaryotic genome probably resulted from the fusion of archaeobacterial and eubacterial genomes, effectively turning the tree of life into a ring of life. But how did evolution come up with the strange distribution of eubacterial and archaeobacterial genes we see in eukaryotes today?

In prokaryotes there are two major gene classes: operational and informational. Operational genes are involved mainly in day-to-day processes of cell maintenance, and code for amino-acid and nucleotide biosynthesis as well as related functions. Informational genes feature primarily in transcription, protein synthesis, DNA replication and other processes to convert information from DNA into proteins.

Because eukaryotes are derived from archaeobacteria and eubacteria, one might expect to find an archaeobacterial and a eubacterial copy of each nuclear gene. But strangely, archaeobacterial operational and eubacterial informational genes are almost completely absent from eukaryotes, even though the first eukaryote contained two sets of informational and operational genes.

This well-documented correlation between phylogenetic origin and gene disappearance is paradoxical because no one understands how these classes of genes left the scene. I call this correlation, which provides an important clue to the early evolution of eukaryotes, the Janus

paradox. Like the two faces of the Roman god Janus, thought to represent the Moon and the Sun, the phylogenetic origins of informational and operational genes in eukaryotes are as different as night and day. Finding a gene distribution such as this is the statistical equivalent of finding that a coin tossed at night (Janus's archaeobacterial face) always comes up heads (informational genes), and tossed during the day (Janus's eubacterial face) always comes up tails (operational genes).

But before we look at possible causes of the Janus paradox, we need to understand the interactions between operational and informational genes. In some ways, gene transfers between prokaryotes mimic patterns of telephone use. Some people only call their family and friends, for instance. This is similar to the pattern of transfer of informational genes between closely related prokaryotes. Others add distant associates to their basic phone lists — analogous to the broader transfer of operational genes between prokaryotes. The 'complexity hypothesis' attributes the reduced transfer rate seen in informational genes partly to the observation that their proteins are often deeply integrated into large complexes such as the ribosome. So to function, transferred informational genes must fit into complex pre-existing structures, effectively restricting their transfers to closely related prokaryotes (family and friends). Transferred operational genes, as members of smaller, less complex structures, fit in more easily, allowing them to function when transferred over larger phylogenetic distances (family, friends and associates).

Combining two genomes into one nucleus would not have been simple. We know, for example, that two sets of ribosomal RNA operons — genes controlled as a unit — cannot coexist in the same cell. Recent attempts to fuse a cyanobacterial genome into the genome of a host *Bacillus subtilis*, a common soil bacterium, were successful only when the cyanobacterial ribosomal RNA operons were removed from the fusion chromosome. We also

know from experiments on reconstituting ribosomal subunits that, consistent with the complex interactions within the ribosome, even a single damaged protein can completely inactivate protein synthesis.

This suggests a possible explanation for why the eubacterial informational genes disappeared. Because two types of ribosomal genes cannot exist in the same nucleus, the archaeobacterial ribosome may simply have been the lucky survivor when one of the components in the eubacterial ribosome was inactivated. Once this chance inactivation occurred, it was probably only a matter of time until all eubacterial informational genes were eliminated, given the extensive interactions between the ribosome and informational proteins.

Unfortunately, I have no good suggestion for why the archaeobacterial operational genes were eliminated. I hope that this will motivate some readers to think of hypotheses and experiments. It could be that somehow, the ready availability of operational genes within the eubacterially derived cellular organelles led to the preponderance of eubacterial operational genes.

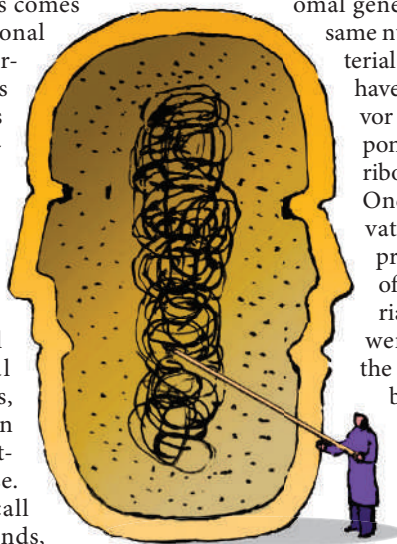
Whatever explanations of the Janus paradox are unearthed, it will be exciting to follow the quest. How the eukaryotic cell came to be is one of the greatest enigmas in biology. It is a story so complex that no single gene can tell it. Only entire genomes can.

James A. Lake is distinguished professor in the Departments of MCD Biology and Human Genetics, and in the Molecular Biology Institute, University of California, Los Angeles, 242 Boyer Hall, Los Angeles, California 90095, USA.

FURTHER READING

Martin, W. & Embley, T. M. *Nature* **431**, 134–136 (2004).
 Hey, J., Fitch, W. M. & Ayala, F. (eds) *Systematics and the Origin of Species On Ernst Mayr's 100th Anniversary* (The National Academies Press, Washington DC, 2005).
 Jain, R., Rivera, M. C. & Lake, J. A. *Proc. Natl. Acad. Sci. USA* **96**, 3801–3806 (1999).

For other essays in this series, see <http://nature.com/nature/focus/arts/connections/index.html>



J. KAPUSTA/IMAGES.COM

CONNECTIONS