

# A forage in the junkyard

One of the main differences between the mouse and human genomes lies in the activity of 'junk' DNA sequences called retrotransposons. Carina Dennis considers what these sequences might be doing.

Wherever human garbage accumulates, mice will forage. But in genomics, the tables are being turned. For some geneticists, the most intriguing aspect of the mouse genome lies in its 'junk' — DNA sequences that don't code for proteins. The mouse, they believe, will prove a fertile hunting ground for researchers who want to understand why and how mammalian genomes accumulate this garbage, and whether it has any function.

Much of the junk is the work of transposons, genetic 'parasites' that have accumulated in mammalian genomes over millions of years by being copied into new genomic locations. Most are retrotransposons, which reproduce through an RNA intermediate, and so must use a reverse transcriptase enzyme to restore their original DNA sequence as they jump back into the genome.

The mouse genome is not especially full of such garbage — although at least 37.5% of it seems to be derived from retrotransposons<sup>1</sup>, the proportion is similar in the human genome<sup>2,3</sup>. But the activity of these retrotransposons differs greatly between the two genomes. In our evolutionary lineage, retrotransposon activity seems to have quietened down about 40 million years ago<sup>2</sup>, and less than 100 are actively jumping around in any individual's genome today<sup>4</sup>. But in the mouse, about 3,000 of these parasites are thought to be still on the move<sup>5</sup>, on occasion disrupting host genes and affecting the creature's genome in ways that geneticists are keen to investigate. "I believe that some of these elements have been the most dynamic forces to sculpt the genome," says Sandra Martin, who works on retrotransposons at the University of Colorado Health Sciences Center in Denver.

Among the most common types of retrotransposon are LINES — long interspersed elements — and their shorter counterparts, known as SINES. LINES can be up to 6,000 DNA base pairs long, and can enlist the host's gene transcription machinery to produce their own reverse transcriptase. SINES are typically a mere 300 or so base pairs long, and lack their own reverse transcriptase, so are thought to co-opt the enzyme produced by LINES to propagate themselves.



Say cheese: just one piece of 'junk' DNA can produce several coat colours in genetically identical mice.

The third major class, elements with long terminal repeats (LTRs), encode their own reverse transcriptase and bear similarities to retroviruses — indeed, they might be evolutionary relatives.

## Retro chic

LINES and SINES cover almost 20% and just over 8% of the mouse genome, respectively; LTRs account for about 10% (ref. 1). All have been accumulating over tens of millions of years. "They are ancient fossils," says Haig Kazazian, a geneticist at the University of Pennsylvania School of Medicine in Philadelphia.

Whenever a retrotransposon lands in a host gene, it may alter the gene's function or disable it completely. And because many retrotransposons in the mouse are still active, they are thought to be responsible for about 10% of naturally occurring mutations that cause a noticeable change in characteristics<sup>4</sup>.

Retrotransposons don't just exert their effects by jumping into individual genes. In August, teams led by John Moran of the University of Michigan Medical School in Ann Arbor and Jef Boeke of the Johns Hopkins University School of Medicine in Baltimore, Maryland, demonstrated, in

cultured human cells, that retrotransposons can also cause the deletion of large blocks of DNA when they insert into a chromosome<sup>6,7</sup>. Such deletions may have been important agents of evolutionary change.

Mutations and deletions caused by retrotransposons will, in many cases, be harmful. So most experts believe that the host genome is engaged in a constant war, fought through natural selection, to keep the genomic junkyard created by retrotransposons from spreading. The result is that the rubbish tends to accumulate in places where it doesn't get in the way and cause too much damage. Just as we might store our junk in a basement or closet, some corners of mammalian genomes are littered with retrotransposon remnants.

But other regions are spotless. Both the mouse and human genomes, for instance, seem to be meticulous about keeping the cluster of *Hox* genes that specify the body plan during embryonic development free from retrotransposons. Other similarly 'clean' regions might point researchers to genomic regions of crucial functional importance, suggests Dixie Mager, a geneticist at the British Columbia Cancer Research Centre's Terry Fox Laboratory in Vancouver, Canada. "Searching for regions where



**Garbage men:** Haig Kazazian (above) and Arian Smit are rooting through the rubbish dumps of the mouse genome for evolutionary clues.

retroelements are notably absent could offer insights," she suggests.

Strikingly, preliminary comparisons of the mouse and human genomes reveal that junk accumulates in similar places in both. Arian Smit of the Institute for Systems Biology in Seattle and his colleagues have compared blocks of sequence up to 500 kilobases long from mouse and human genomes, and found a similar pattern of density of retrotransposons, particularly SINEs, in corresponding regions<sup>1</sup>. Because many elements jumped into the mouse and human genomes long after the two species diverged, they must have independently come to rest in similar positions.

Meanwhile, emerging evidence suggests that the host genome sometimes puts its resident retrotransposons to good use — much as we might dust off old junk and recycle it for a new purpose. Mart Speak of Tallinn Technical University in Estonia, for instance, has shown that LINEs can alter the expression of neighbouring host genes<sup>8</sup>. Again, most such changes are likely to be harmful, but in some cases, they may regulate gene activity in ways tolerated by the host.

"Young elements are interesting because they are still active and moving around the genome and possibly disrupting genes," says Mager. "But older elements are also interest-

ing because they have had time to become absorbed by the genome and possibly assume a regulatory function."

LINEs, for instance, seem to accumulate on the sex chromosomes of both human and mouse<sup>1-3</sup>, and are particularly common in the regions of the X chromosome involved in shutting down gene expression in one of the two copies of the chromosome present in female mammals<sup>9,10</sup>. Perhaps, some researchers speculate, LINEs have been recruited as part of the machinery responsible for this inactivation. But other experts think this unlikely, arguing that any regulatory role for retrotransposons will have arisen purely by chance. "There is no evidence of active recruitment," asserts Anthony Furano of the National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland.

### Line dancing

Irrespective of whether the mammalian genome has made use of some retrotransposons, researchers may be able to manipulate them to their advantage in the future. Kazazian and his colleagues have created a mouse model to study the movement of a human LINE, in which an active retrotransposon can be passed onto the next generation, inserting into a new genomic position as it does so<sup>11</sup>. This will allow researchers to study retrotransposition and its consequences in a controlled way in living mammals for the first time.

Combine the potential of this model with the mouse's extensive evolutionary legacy of past retrotransposon activity, add the fact

that many of its retrotransposons are still moving around, and mice should become the ideal organisms in which to study the influence of retrotransposons on mammalian genomes. But it may be a couple more years before the full harvest can be reaped.

The current draft of the mouse genome was produced by a technique called whole-genome shotgun sequencing (see page 460). This method is notoriously poor at rendering the correct sequence of large expanses of junk DNA that contain the repetitive products of retrotransposon activity. Of the 3,000 potentially active retrotransposons thought to be in the mouse genome, only 400 — of which just 12 have their entire coding region intact — could be found in the assembled sequence<sup>1</sup>. "It is disappointing, as important information has been lost," says Kazazian.

But the public consortium responsible for the draft mouse genome plans to finish the job by 2005, using a map-based method that should fill in this missing information. For retrotransposon aficionados, that holds the exciting prospect of being able to compare the influence of these elements on human and mouse genomes — and perhaps other species. A comparative sequencing effort along these lines is being spearheaded by Eric Green of the National Human Genome Research Institute in Bethesda, Maryland, who is generating large quantities of high-quality sequences from more than two dozen vertebrate species.

Other researchers want to compare retrotransposons between strains of mice, and even different mice within the same strain. Researchers led by Emma Whitelaw at the University of Sydney in Australia have already shown that a single retrotransposon can cause heritable differences in coat colour between mice that are genetically identical<sup>12</sup>. Retrotransposons, it seems, may help to explain not only differences between species, but also a spectrum of traits between individuals — not a bad haul from a forage in the genome's junkyard. ■

**Carina Dennis is Nature's Australasian correspondent.**

1. Mouse Genome Sequencing Consortium *Nature* **420**, 520–562 (2002).
2. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
3. Venter, J. C. et al. *Science* **291**, 1304–1351 (2001).
4. Ostertag, E. M. & Kazazian, H. H. Jr *Annu. Rev. Genet.* **35**, 501–538 (2001).
5. Goodier, J. L., Ostertag, E. M., Du, K. & Kazazian, H. H. Jr *Genome Res.* **11**, 1677–1685 (2001).
6. Gilbert, N., Lutz-Prigge, S. & Moran, J. V. *Cell* **110**, 315–325 (2002).
7. Symer, D. E. et al. *Cell* **110**, 327–338 (2002).
8. Speak, M. *Mol. Cell. Biol.* **21**, 1973–1985 (2001).
9. Bailer, J. A., Carrel, L., Chakravarti, A. & Eichler, E. E. *Proc. Natl. Acad. Sci. USA* **97**, 6634–6639 (2000).
10. Lyon, M. F. *Cytogenet. Cell Genet.* **80**, 133–137 (1998).
11. Ostertag, E. M. et al. *Nature Genet.* **32**, 655–660 (2002).
12. Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K. & Whitelaw, E. *Nature Genet.* **23**, 314–318 (1999).