

PERSPECTIVES: GENOMICS

Mmu 16—Comparative Genomic Highlights

Neal G. Copeland, Nancy A. Jenkins, Stephen J. O'Brien

How similar are the mouse and human genomes to each other and to other mammalian genomes? What are the evolutionary forces that have shaped these genomes since they diverged nearly 90 million years ago? What makes

a person a human and not a mouse? The answers to these age-old questions form the cornerstone of modern comparative genomics and will determine the value of model organisms such as the mouse for understanding the functions of human genes. With the publication of the mouse chromosome 16 (Mmu 16) draft sequence by Mural *et al.* (1) on page 1661 of this issue, and a recent full sequence comparison of human chromosome 19 (Hsa 19) with related mouse sequences (2), the answers to some of these questions are coming into view.

The Mmu 16 draft sequence was generated with the technique of whole-genome shotgun sequence assembly, the approach used earlier for sequencing the *Drosophila* and human genomes (3, 4). The 5.3× DNA sequence coverage of Mmu 16 was derived from four mouse strains (A/J, DBA/2J, 129×1/SvJ, 129S1/SvImJ) chosen in part to complement the C57BL/6J sequence being generated by the public sequencing effort. The assembly produced 19,788 scaffolds (median size 4.5 megabases) that were ordered and mapped to mouse chromosomes using public genetic and radiation hybrid maps.

What have we learned from comparing whole chromosome sequence segments of mouse and human? First, we see a fine-grained affirmation of the well-established inference that the two mammalian species share around 200 homology segments. Homology segments are chromosome chunks that contain a linear stretch of the same gene homologs in two compared species (5–7). The two contiguous homologous gene arrays are also termed a “conserved synteny”. Mmu 16 contains seven human

homology segments, whereas Hsa 19 contains nine mouse homology segments.

These conserved syntenic segments are reorganized between the two species, but within the segments the homologous DNA sequence orders are strikingly parallel. Mural *et al.* (1) identified 11,822 “syntenic anchors” on Mmu 16, which are short stretches of DNA in mouse and human that show significant sequence match to each other but not to any other region in either genome. A remarkable 98.1% of the Mmu 16 syntenic anchors fall in the same syntenic chromosomal position relative to adjacent anchors in both species. More



than 50% of the anchors are located in runs of 128 anchors or more, in the same order and orientation in both genomes.

The close conservation of sequence order illustrates two important points. First, it increases confidence in the contig/scaffold scheme for genome assembly because the syntenic anchor orders are reinforced among independent mammalian lineages. Second, although the mouse genome contains two- to threefold as many rearrangements as the genomes of cats, cows, humans, dolphins, and other mammals (5, 6), the marked extent of “syntenic anchor” order conservation within the conserved segments implies that genome reorganization in the rodent lineage occurred but once. The reassortment is likely to have taken place before the 15-million-year-old divergence of mouse and rat—but after their predecessor’s divergence from the ancestors of primates and rodents around 90 million years ago (8,

9)—because the mouse and rat genomes are relatively concordant with each other (5, 10). A coarse view would suggest a fracturing of rodent and ancestral mammalian genomes. However, the fine-scale view afforded by the full genome sequence of the mouse demonstrates that the ancestral locus order and alignment are highly conserved within the syntenic segments.

Interestingly, 44% of Mmu 16 syntenic anchors are outside the limits of recognized genes, and only 34% of syntenic anchors overlap coding exons. Clearly, selective evolutionary forces are constraining rapid divergence of default neutral sequences in these nongenic regions, providing rather strong evidence for their involvement in the survival of these individual species.

Inspection of the annotated genes tells a similar story, but with an interesting twist. Mmu 16 contains 731 genes of medium to high confidence on the basis of gene-annotation algorithms. Homologous genes are distributed among the seven conserved syntenic blocks in the human genome. Apart from one 363-kilobase (kb) syntenic block containing three genes on Hsa 12, all of the other syntenic blocks have been described previously. Most Mmu 16 genes (509) have human orthologs in the expected syntenic position, but 222 (30%) do not. Of these exceptions, 44 are gene paralogs mostly derived by rodent or primate lineage-specific tandem gene duplication, whereas 164 genes have human homologs that violate syntenic expectations. (Two genes are said to be orthologs if they are derived from a speciation event, but paralogs if they are derived from a duplication event.) Several are rearranged by small interstitial inversions—two inversions of the Mmu 16/Hsa 21 and Mmu 16/Hsa 22 syntenic regions involving 8% of the genes and one gene in the Mmu 16/Hsa 16 syntenic region—but the others are rather puzzling.

Fourteen Mmu 16 genes have no known human homologs, whereas 21 human genes in the compared regions are unique to humans. Extrapolating across the entire genomes and presuming a 90-million-year interval since mouse and human shared an ancestor (8, 9), this means that one new gene arose or disappeared on average every 192,000 years. These preliminary estimates will surely become more precise as we inspect genome sequences from additional chromosomes and from other mammals in the future.

As with human chromosomes, gene density varies considerably along the

Enhanced online at
www.sciencemag.org/cgi/
content/full/296/5573/1617

N. G. Copeland and N. A. Jenkins are in the Mouse Cancer Genetics Program, and S. J. O'Brien is in the Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD, 21702, USA. E-mail: copeland@ncicfcrf.gov

CREDIT: JOE SUTLIFF

length of Mmu 16. One 6-Mb region contains only 7 genes; another 1.1-Mb region contains 17 genes. In general, mouse genes tend to be smaller than their human counterparts. This is largely attributable to differences in the amount of SINE and LINE sequences in these two genomes. In human, these large repetitive-element families account for 46% of base pairs, whereas in mouse they account for only 36%.

In contrast to the high degree of conservation observed for single-copy genes, those located within tandem gene clusters differ extensively in their number, coding potential, and organization between the two species. A good example is the zinc finger (*ZNF*) genes located on Hsa 19. This human chromosome carries 262 C2H2 *ZNF* genes, dispersed among 11 different syntenic clusters. Most clusters contain closely related gene sequences that appear to have arisen by tandem duplication of ancestral copies. Many related mouse clusters, however, contain very different complements of *ZNF* genes, and gene analysis suggests that different founder genes were duplicated, lost, and selected independently in each conserved cluster. Most duplicated genes retain their coding capacity, suggesting that they have nonredundant adaptive functions that complement those of the ancestral parental genes. Because *ZNF* genes are important regulators of gene expression, these species-specific amplifications and deletions almost certainly helped to shape the evolution of these two species. Similar results were also reported for olfactory and putative pheromone receptors genes and could easily account for differences in the

way humans and mice taste their food and attract sex partners.

The mosaic organization of mammalian genomes is likely to be due principally to lineage-specific rearrangements of these genomes over their evolutionary history. Evidence for these rearrangements can be seen in gene density changes—SINE + LINE density, and G + C content—in sequences located at the boundary of the rearranged syntenic segments. As pointed out by Mural and colleagues (1), these sequence differences could easily be explained by the breaking and joining of ancestral chromosomal regions with very disparate properties. Several syntenic breakpoints are located in clustered gene families, with the break splitting closely related family members. Mouse breakpoint clones also tend to be L1 sequence-rich, showing a twofold increase over the L1 repeat content of other mouse DNA. These repeated sequences might have been the driving force behind the genomic rearrangements; repeated sequences have been proposed to drive the genomic rearrangements documented in several human diseases.

As provocative and fascinating as these inferences are, they are only the harbinger of what is yet to come when the public sequencing project discloses a finished, more thoroughly curated, sequence of mouse and man. (Celera has deposited the Mmu 16 sequence at GenBank, but the remaining mouse sequence is proprietary, requiring hefty fees for inspection and analysis.) The prospect of whole-genome sequencing for other mammals (rat, chimpanzee, macaque, cattle, pig, dog, and cat are likely candidates) offers an unprecedented opportunity to ad-

dress a variety of genomic mysteries, hitherto restricted to speculation and learned guesswork. What is the nature of and the selective pressure responsible for the high incidence of conserved syntenic anchors outside coding gene limits, estimated here as 44%? What are the evolutionary forces that drive and maintain the chromosomal exchanges, translocations, and internal inversions that punctuate the genomes of modern mammals? In lineages with highly reshuffled chromosomes (rodents, bears, chimps, owl monkeys, squirrel monkeys, muntjaks, and others) (6, 8), which events favor the burst of these rare genomic reorganizations? How do new genes arise and others disappear in species genomes? Do these events actually matter in species adaptation and survival? As whole genome sequences become interpreted against the mammalian evolutionary background and dynamic genome tinkering is revealed, we shall be able to view what has happened in our evolutionary past, what matters to our future, how modern genomes and developmental adaptations were sculpted. Our genomes hold the gene-script for specifying modern species, including ourselves, and are now beginning to reveal a rich new perspective of how they came to be.

References

1. R. J. Mural et al., *Science* **296**, 1661 (2002).
2. P. Dehal et al., *Science* **293**, 104 (2001).
3. J. C. Venter et al., *Science* **291**, 1304 (2001).
4. E. W. Myers et al., *Science* **287**, 2196 (2000).
5. S. J. O'Brien et al., *Science* **286**, 458 (1999).
6. R. Stanyon et al., *Cytogenet. Cell Genet.* **84**, 150 (1999).
7. J. Nadeau and B. Taylor, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 814 (1984).
8. W. J. Murphy et al., *Science* **294**, 2348 (2001).
9. E. Eizirik et al., *J. Hered.* **92**, 212 (2001).
10. W. J. Murphy et al., *Genome Biol.* **2**, 5 (2001).

PERSPECTIVES: CLIMATE CHANGE

Carbonate Mysteries

Henry Elderfield

The pH of seawater is controlled by dissolved calcium carbonate, which provides a buffer against pH perturbations—natural or manmade—that strive to change it. One such perturbation is atmospheric carbon dioxide (CO_2), an acidic gas that is exchanged between atmosphere and ocean.

During glacial cycles, concentrations of CO_2 in the atmosphere vary widely, putting the carbonate buffer through a tough test. A recent study suggests that it coped more efficiently than previously thought (1). A second study reaches the

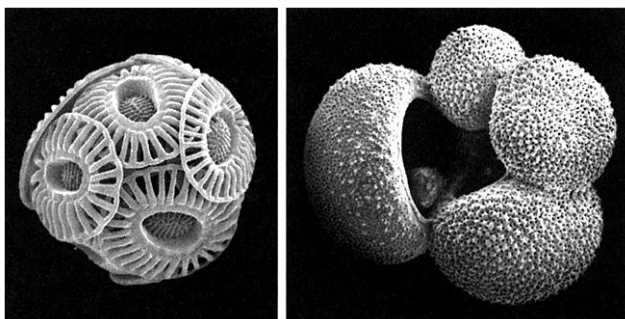
opposite conclusion (2). Other work, on the short time scale of a laboratory experiment, suggests that changes in buffer

strength may be balanced by changes in biological precipitation of calcium carbonate from seawater (3).

Who is right? The answer has considerable implications for the ability of the oceans to sequester anthropogenic CO_2 from the atmosphere.

The seawater carbonate system works just like a buffer used in a chemistry laboratory. It counters changes in pH by altering the proportions of the abundant weak acids bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}).

The oceans are currently buffered to a slightly alkaline pH of 7.6 to 8.2. The buffer works well as long as relatively small amounts of acid or base are added. However, all chemical buffers are only effi-



Scanning electron microscope images of marine calcifying organisms. (Left) The coccolithophore *Emiliania huxleyi* (diameter 5 μm) and (right) the foraminifer *Globigerina bulloides* (diameter 400 μm).

The author is in the Department of Earth Sciences, University of Cambridge, Cambridge CB2 3EQ, UK. E-mail: he101@esc.cam.ac.uk

CREDIT: LEFT PANEL, PATRIZIA LIVERI; RIGHT PANEL, H. ELDERFIELD