

 GENE EXPRESSION

## Genomic space-savers

Genomes use several strategies to maximize the packaging of genetic information, perhaps the best known being the use of overlapping ORFs. A systematic study of the nature and extent of information packaging devices shows that they are found extensively in genomes as diverse as viruses and plants.

The authors studied 700 genomes to identify coding sequences that contain additional information. They looked at whether 6- or 7-bp sequences are enriched or depleted in the test genomes compared with randomized versions of each genome sequence. The enrichment or depletion of a short sequence in a real genome would indicate pressure to evolve sequences containing a certain code.

The authors found that the genomes of all phyla encode extensive overlapping information, although the information content of eukaryotes is lower than that of bacteria. The biological basis for the overlapping codes is varied. For example, mononucleotide repeats are depleted in the ORFs of all phyla (to minimize DNA replication slippage) and codes for restriction enzyme sites are depleted in bacteria that encode the recognizing restriction enzymes (to avoid aberrant digestion). Similarly, weaker sequence-dependent folding of RNA that is seen at translation start sites might correlate with higher translation efficiency.

Previous studies have revealed examples of overlapping codes, but this is the first study to have systematically characterized this phenomenon globally. The conclusions reached here could be explored further within genomic subregions and by using longer sequences, and these findings could eventually be used in synthetic biology applications.

Tanita Casci

**ORIGINAL RESEARCH PAPER** Itzkovitz, S., Hodis, E. & Segal, E. Overlapping codes within protein-coding sequences. *Genome Res.* 14 Sep 2010 (doi:10.1101/gr.105072.110)